



# Development and validation of a clinical-radiomics nomogram for predicting a poor outcome and 30-day mortality after a spontaneous intracerebral hemorrhage

Yuanliang Xie<sup>1#^</sup>, Faxiang Chen<sup>1#</sup>, Hui Li<sup>1</sup>, Yan Wu<sup>1</sup>, Hua Fu<sup>2</sup>, Qing Zhong<sup>1</sup>, Jun Chen<sup>3</sup>, Xiang Wang<sup>1</sup>

<sup>1</sup>Department of Radiology, the Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; <sup>2</sup>Department of Radiology, the Fifth Affiliated Hospital of Nanchang University, Fuzhou, China; <sup>3</sup>GE Healthcare, Wuhan, China

*Contributions:* (I) Conception and design: Y Xie, X Wang; (II) Administrative support: X Wang; (III) Provision of study materials or patients: F Chen, H Li, Y Wu; (IV) Collection and assembly of data: Q Zhong, H Fu; (V) Data analysis and interpretation: Y Xie, J Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

*Correspondence to:* Xiang Wang. Department of Radiology, the Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, 26 Shengli Street, Jiang'an District, Wuhan 430014, China. Email: wangxiang\_1971@163.com.

**Background:** Noncontrast computed tomography (NCCT) is often performed for patients with a suspected spontaneous intracerebral hemorrhage (ICH) at the time of admission. Both clinical and radiomic features on the initial NCCT can predict the outcomes of those with ICH, but satisfactory model performance remains challenging.

**Methods:** A total of 258 acute ICH patients from the Central Hospital of Wuhan (CHW) between January 2018 and December 2020 were retrospectively assigned to training and internal validation cohorts at a ratio of 7:3. An independent external testing cohort of 87 patients from January 2021 to July 2021 from the Fifth Affiliated Hospital of Nanchang University (FAHNU) was also used. Based on the least absolute shrinkage and selection operator (LASSO) algorithm, radiomics (rad)-scores were generated from 9 quantitative features on the initial NCCT images. Three models (radiomics, clinical, and hybrid) were established using stepwise logistic regression analysis. The Akaike information criterion and the likelihood ratio test were used to compare the goodness of fit of the three models. Receiver operating characteristic (ROC) curve analysis was performed and bar charts were constructed to evaluate the discrimination of constructed model for predicting a poor outcome following ICH.

**Results:** The three cohorts had similar baseline clinical characteristics, including demographic features and outcomes. In the clinical model, hematoma expansion [2.457 (0.297, 2.633);  $P=0.014$ ], intracerebral ventricular hemorrhage [2.374 (0.180, 1.882);  $P=0.018$ ], and location [-2.268 (-2.578, -0.188);  $P=0.023$ ] were independently associated with a poor clinical outcome. In the hybrid model, location [-2.291 (-2.925, -0.228);  $P=0.022$ ], and rad-score [5.255 (0.680, 11.460);  $P<0.001$ ] were independently associated with a poor outcome. The hybrid model achieved satisfactory discriminability, with areas under curve (AUCs) of 0.892 [95% confidence interval (CI): 0.847 to 0.937], 0.893 (95% CI: 0.820 to 0.966), and 0.838 (95% CI: 0.755 to 0.920) in the training, internal validation, and external testing cohorts, respectively. The hybrid model also achieved good discriminability in the prediction of 30-day mortality, with AUCs of 0.840, 0.823, and 0.883 in the training, internal validation, and external testing cohorts, respectively. The rad-score [2.861 (1.940, 4.220);  $P<0.001$ ] was the predominant risk factor associated with 30-day mortality.

<sup>^</sup> ORCID: 0000-0002-4190-794x.

**Conclusions:** Radiomic analysis based on initial NCCT scans showed added value in predicting a poor outcome after ICH. A clinical-radiomics model yielded improved accuracy in predicting a poor outcome and 30-day death following ICH compared with radiomics alone.

**Keywords:** Intracerebral hemorrhage (ICH); computed tomography (CT); radiomics; nomogram; outcome

Submitted Feb 11, 2022. Accepted for publication Jul 07, 2022.

doi: 10.21037/qims-22-128

View this article at: <https://dx.doi.org/10.21037/qims-22-128>

## Introduction

Spontaneous intracerebral hemorrhage (ICH) is a life-threatening stroke, with in-hospital and one-year mortality rates that exceed 32% and 45%, respectively (1). Baseline hematoma size, intraventricular extension, hematoma expansion (HE), Glasgow coma scale (GCS), and age are independent predictors of a poor outcome and mortality following ICH (2,3). Hematoma volume is the most important determinant of brain tissue damage via mechanical extrusion and secondary injury due to the presence of intraparenchymal blood. HE occurs in approximately 30% of patients experience HE over the first 24 h following an ICH, resulting in neurologic deterioration and poor 30-day and long-term outcomes (4); HE is another absolute predictor of a poor outcome. Timely and accurate identification of HE following an ICH is critical to facilitate immediate intervention or surgical management, whereas reliable exclusion of HE is also important for individualized management.

Most previous studies have suggested that spot sign-based enhanced computed tomography (CT) imaging, characterized as foci of enhancement within the hematoma, is a promising predictor of an HE and poor outcome following ICH (5,6). However, enhanced CT scans may not always be possible due to the patient's clinical condition, such as a reduced GCS score, the availability of iodine contrast agents, increased radiation exposure, and the increased time needed to perform the procedure. In consideration of these limitations, CT angiography (CTA) or multi-phase enhanced CT scans are not part of the routine diagnostic workup for ICH; they are recommended based on second-level evidence (Class IIb; Level B) in the American Heart Association/American Stroke Association (AHA/ASA) guidelines (7). Noncontrast CT (NCCT) is the first-line diagnostic method identified by these criteria and is globally considered the gold standard for diagnosing an ICH. Several signs shown on NCCT, such as hypodensities

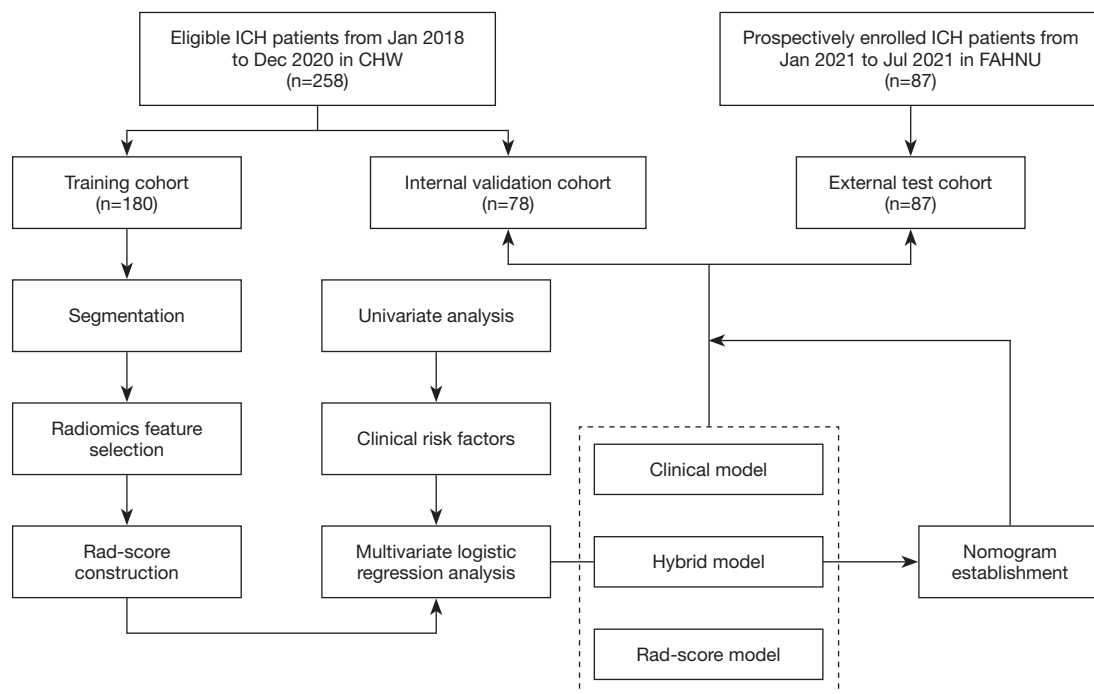
within the hematoma, irregular HE shape, heterogeneous density, and the swirl sign, blend sign, black hole sign, and an island sign, have been recently validated as predictors of HE expansion (3,8). However, the application of single or multiple radiographic signs alone in the early diagnosis of HE remains challenging due to inherent interpreter differences (9).

Radiomics, a noninvasive method for objectively assessing the heterogeneity of extracted quantitative features from biomedical images in a reproducible and high-throughput manner, can be used to support clinical decision-making (10). Radiomic features include morphology, texture, and high-level statistical features and permit the accurate description of hematoma geometry and heterogeneity. Radiomics has been gradually explored by early cancer researchers in other fields. In our previous study, the extracted texture features of NCCT images were able to predict early HE (11). A similar result was reported in a retrospective study with 251 ICH patients (12). Clinical data, such as neutrophil-to-lymphocyte ratio (NLR) and serum calcium, can predict 30-day mortality following acute ICH (13,14). However, the factors contributing to 30-day mortality and poor clinical outcomes following ICH are complicated. Whether radiomics combined with clinical information can yield an additional predictive benefit is still unknown. Therefore, we aimed to establish a hybrid model consisting of clinical and radiomics features with a developing cohort and to validate it on internal and external cohorts for predicting a poor outcome and 30-day mortality following ICH. We present the following article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-128/rc>).

## Methods

### Patients

The study was conducted in accordance with the



**Figure 1** Study workflow. ICH, intracerebral hemorrhage; rad-score, radiomics score; CHW, the Central Hospital of Wuhan; FAHNU, the Fifth Affiliated Hospital of Nanchang University.

Declaration of Helsinki (as revised in 2013). The study was approved by the Medical Ethics Committee of the Central Hospital of Wuhan (CHW; No. 2021-36). Consecutive hospitalized patients with acute ICH between January 2018 and December 2020 at the CHW were included in either a training or an internal validation cohort. Informed consent for this retrospective research was waived. Patients with an acute ICH from 1 January 2021 to 31 July 2021 from the Fifth Affiliated Hospital of Nanchang University (FAHNU) were prospectively enrolled in an external test cohort, and informed consent was obtained from every patient or their family members (Figure S1; Appendix 1). The inclusion criteria for this study were as follows: (I) age  $\geq 18$  years or above; (II) time to initial CT and/or CTA less than 24 h from symptom onset; and (III) one or more CT follow-ups within 72 h and available for HE assessment. The exclusion criteria were as follows: (I) isolated subarachnoid hemorrhage, ventricular hemorrhage, or subdural/epidural hemorrhage; (II) secondary causes, such as trauma, hemorrhagic transformation of ischemic infarcts, tumor, infection, vasculitis, or vascular malformation; (III) surgical removal of HE within 72 h of symptom onset; (IV) acquisition thickness  $\geq 1.5$  mm; (V) poor image quality; or (VI) unavailable image data. All patients were

treated according to the Guidelines for the Management of Spontaneous Intracerebral Hemorrhage of the AHA (Ver. 2015) (7). The study workflow is presented in Figure 1.

### CT imaging and radiographic interpretation

All CT images were acquired on multi-detector CT scanners (Lightspeed 16, GE Healthcare, Chicago, IL, USA; iCT, Philips Medical Systems, Best, the Netherlands; Somatom Definition AS, Siemens Healthcare, Erlangen, Germany) in Digital Imaging and Communications in Medicine format with a cut thickness of 1–1.25 mm, an auto-tube current of 239–273 mAs, a tube voltage of 120 KV, a field of view (FOV) of 25 cm, and a matrix size of 512×512 pixels.

All CT scans were reviewed by one specialized neuroradiologist (HL, with 11 years of experience with neuroradiology interpretation) and one resident (QZ, with 3 years of experience with radiology interpretation) who were both blinded to clinical data. The ICH volumes on NCCT at baseline and follow-up (24 h) were calculated using volumetric studies aggregated from manual segmentation by computerized planimetry software (ITK-Snap; <http://www.itksnap.org/pmwiki/pmwiki.php>) and

verified by the two blinded readers. The hematoma site was divided into two subgroups: deep (including basal ganglia, thalamus, corpus callosum, brainstem, and cerebellum) and lobar. An HE was defined as more than 6 mL or 33% growth compared to the initial ICH volume, including the development of an intraventricular hemorrhage (IVH) after the initial CT scan. The CTA spot signs and signs on NCCT, such as hypodensity, blend sign, irregularity, and satellite, and island signs, were evaluated by the two readers following the methods used in a previous study (15). Disagreements were resolved by a third reader (WX, with 13 years of experience with neuroradiology interpretation).

### *Clinical data and outcomes*

The following clinical data were collected: gender, age, medical history (hypertension, diabetes mellitus, anticoagulant use, anti-platelet use, coronary artery disease, history of stroke, alcohol consumption, hepatic insufficiency (B and C of Child-Pugh grade), renal insufficiency (serum creatinine  $\geq 451$  mol/L), initial GCS score, time to baseline NCCT, time to follow-up CT, blood pressure at admission, blood lipids, blood glucose, NLR, serum calcium concentration, hematoma location, and hematoma extension into the ventricles. A modified Rankin scale (mRS) was measured by a designated neurologist (CW, with 17 years of experience in neurosurgery) who was blinded to the primary outcome at discharge. A poor outcome was defined as an mRS grade of 4–6 (mRS 0–3 = favorable; mRS 4–5 = moderate-severe disability; mRS 6 = deceased).

### *Radiomic protocol*

All CT images were initially resampled into voxel sizes of  $1 \times 1 \times 1$  mm<sup>3</sup> using linear interpolation in A.K. software (artificial intelligence kit; A.K.3.1.0.R, GE Healthcare, Shanghai, China) to reduce heterogeneity between images obtained from different CT scanners, with thicknesses of 1.0–1.25 mm.

We then performed hematoma segmentation and extracted radiomics features. An experienced radiologist (LH, with 11 years of neuroradiology experience), who was blinded to clinical information, manually delineated regions of interest (ROIs) along the edge of the hematoma slice by slice in multiple successive slices, and then analyzed the data. To improve the contrast and interobserver agreement on the interface between the hematoma and the brain parenchyma, a relatively narrow window width (60–70 HU)

combined with a flexible window level (30–40 HU) was used. A semi-automatic segmentation method using a CT threshold was applied to identify hematomas. Another radiologist (YW, with 20 years of experience) reevaluated 30 patients from the developmental cohort using stratified sampling. We assessed the feature stability between the 30 matched ROIs identified by the two readers using the intraclass correlation coefficient (ICC). An ICC greater than 0.70 indicated high feature stability. Features with an ICC below 0.70 were excluded.

The CT images were analyzed to extract 1,072 radiomics features per patient. In these radiomics features, there were 7 distinct groups of features: shape features, first order features, gray level co-occurrence matrixes, gray level dependence matrixes, gray level run length matrixes, gray level size zone matrixes, and neighborhood ray tone difference matrixes. Quantitative radiomics features were extracted from three types of images: the original image, the Laplacian of Gaussian (LoG) image, and the Wavelet image, which were generated through eight decompositions after wavelet filtering. Applying the High (H) or Low (L) pass filters in three dimensions yielded eight combinations: LHL, HHL, HLL, HHH, HLH, LHH, LLH, and LLL. By applying an LoG filter with a sequence of sigma values, LoG images were generated. Images with a low sigma emphasized fine textures, and those with a high sigma emphasized coarse textures. In this study, sigmas of 2, 3, and 4 were used.

### *Establishment of the radiomics score*

A three-step procedure was performed to reduce the dimensionality of the radiomics features. First, we excluded radiomics features with a variance of less than 1.0. Statistically significant features ( $P < 0.05$ ) were identified using the Student's *t*-test or the Mann-Whitney U test. The least absolute shrinkage and selection operator (LASSO) was used to identify the optimized subset of features for selecting potential radiomics predictors in the training cohort (Figure S2A). A 10-fold cross-validation was used to avoid over-fitting (Figure S2B,S2C). Features with nonzero coefficients were used to construct the radiomics score (rad score): rad score =  $(\sum \beta_i * X_i) + \text{Intercept}$  ( $i=0, 1, 2, 3, \dots$ ) where  $X_i$  represented the  $i^{\text{th}}$  selected feature and  $\beta_i$  was its coefficient.

### *Clinical and hybrid models*

Clinical characteristics were compared between mRS

0–3 and mRS 4–6 and survivor and non-survivor groups. Clinical data included the following: age, gender, smoking, drinking, comorbidities, warfarin use, GCS score at admission, time to baseline NCCT, blood pressure at admission, NLR, prothrombin time, serum calcium concentration, and hematoma broken into ventricular versus non-ventricular. A recent study showed that a nomogram derived from NCCT signs and clinical factors could be applied for the risk stratification of HE (16). However, these NCCT signs were not selected as risk factors during the model construction because the radiomics features on NCCT were used in our study. During the development of clinical models, significant variables were selected for a stepwise multivariate logistic regression analysis with the Akaike information criterion (AIC) and likelihood ratio test (LRT) serving as the stopping rule. Based on individual data from the training cohort and binary logistic regression estimates, we determined the probability of poor outcomes and 30-day mortality based on the clinical model. In the training, validation, and independent test cohorts, the exact same multivariable regression formula was applied to calculate the predictive probability of a poor outcome and death within 30 days. A stepwise logistic regression analysis was then used to develop a hybrid model by combining the rad score with the clinical risk factors identified. The AIC and LRT were also used as the terminal rules during model building. We calculated the probability of a poor outcome and 30-day mortality for the three cohorts based on the multivariate logistic regression model estimates.

#### ***Model construction, calibration, and validation***

All participants were randomly divided into training and validation cohorts according to a 7:3 ratio. Three models, a radiomics model (rad-score-based), a clinical model (clinical-factor-based) and a hybrid model (clinical-radiomics score-based), were established in the training cohort.

#### **Discrimination**

Receiver operating curves (ROCs) were used to assess the model's discrimination capability for a poor outcome and 30-day-death. The bar charts were plotted to display the discrimination performance. Further, accuracy, precision, sensitivity, specificity and AIC, and LRT were used for evaluating the constructed models.

#### **Calibration**

Calibration curves were plotted in both the training test and

independent validation cohorts to explore the agreement between the observed outcome and predicted probabilities of the models. The Hosmer-Lemeshow test was used to determine the goodness of fit of the models, and a P value of more than 0.05 was considered well-calibrated.

#### **Clinical applications**

Decision curve analysis (DCA) was used to assess the clinical usefulness of built models by quantifying the net benefits at different threshold probabilities in the three cohorts. A nomogram was formulated based on radiomics score and clinical factors by multivariable logistic regression.

#### ***Statistical analysis***

The software R version 3.5.3 (<https://www.R-project.org/>; The R Foundation for Statistical Computing, Vienna, Austria) and SPSS 25.0 (IBM Corp., Armonk, NY, USA) were used to perform statistical analyses. The missing variables were handled by single imputation using an expectation-maximization algorithm. Categorical variables were expressed as frequency (percentage), and continuous variables were presented as mean  $\pm$  standard deviation (SD). Categorical variables were analyzed using a  $\chi^2$  or Fisher's exact test. The Kolomogorov-Smirnov method was used to test the normality of all measurement data. An independent sample t-test or Mann-Whitney U test was used to measure statistical differences between the mRS 0–3 and mRS 4–6 groups and the 30-day-death and survivor groups. Independent predictors of a poor outcome and 30-day death were identified using logistic regression analysis, and ROC curve analysis was performed and compared for statistically significant variables. The DeLong test was used to compare the discrimination of the three models. An area under the curve (AUC) of more than 0.75 was considered good discriminability for a poor outcome or 30-day-death. A P value <0.05 was considered statistically significant.

## **Results**

### ***Demographic and clinical characteristics***

Patients with intracerebral hemorrhage (n=470) were screened. A total of 258 of 354 patients with ICH were included in the training and internal validation cohorts, and 87 patients were enrolled in the external test cohort. Emergency surgical treatment was performed on 94 (19.8%,

**Table 1** Baseline clinical characteristics of intracerebral hemorrhage patients in the training, internal validation, and external test cohorts

| Characteristics            | Training cohort (n=180) | Internal validation cohort (n=78) | External test cohort (n=87) | P value |
|----------------------------|-------------------------|-----------------------------------|-----------------------------|---------|
| Male gender (%)            | 123 (68.3)              | 49 (62.8)                         | 66 (75.9)                   | 0.786   |
| Age (years)                | 59.5±11.9               | 60.9±12.4                         | 59.5±13.1                   | 0.777   |
| Time to baseline NCCT (h)  | 3.0 (2.0, 7.0)          | 3.0 (1.375, 8.0)                  | 1.0 (1.0, 4.0)              | <0.001  |
| Initial GCS score (>8) (%) | 47 (26.1)               | 19 (24.4)                         | 26 (29.9)                   | 0.704   |
| IVH (%)                    | 72 (40.0)               | 29 (37.2)                         | 37 (42.5)                   | 0.783   |
| ICH location               |                         |                                   |                             | 0.056   |
| Deep (%)                   | 159 (88.3)              | 65 (83.3)                         | 67 (77.0)                   |         |
| Lobar (%)                  | 21 (11.7)               | 13 (16.7)                         | 20 (23.0)                   |         |
| HE (%)                     | 43 (23.9)               | 21 (26.9)                         | 20 (23.0)                   | 0.823   |
| NLR                        | 6.55 (3.33,12.52)       | 6.03 (3.12, 9.37)                 | 3.50 (1.97, 6.74)           | 0.038   |
| SBP (mmHg)                 | 173.2±27.3              | 170.2±29.8                        | 174.1±32.8                  | 0.079   |
| 30-day mortality (%)       | 21 (11.7)               | 8 (10.3)                          | 9 (10.3)                    | 0.921   |
| mRS 4–6 (%)                | 116 (64.4)              | 50 (64.1)                         | 51 (58.6)                   | 0.633   |

Data were presented as mean ± SD, median (interquartile range) or n (%) unless otherwise stated. NCCT, noncontrast computed tomography; GCS, Glasgow coma scale; IVH, intraventricular hemorrhage; ICH, intracerebral hemorrhage; HE, hematoma expansion; NLR, neutrophils to lymphocyte ratio; SBP, systolic blood pressure; mRS, modified ranking score.

94/470) patients in the derivation cohort and 23 (18.3%, 23/126) patients in the external test cohort; these patients were excluded from further analysis. Out of 258 patients in the retrospective cohort, 21 and 8 deaths in the training and internal validation cohorts occurred in the first 30 days after an ICH, respectively. Deep ICHs occurred in 224 cases (86.8%), and 34 were lobar. In the external testing cohort, 9 deaths occurred within 30-day of the ICH, and 67 (77.0%, 67/87) were deep. A total of 166 (64.3%, 166/258) and 51 (58.6%, 166/258) patients had poor outcomes mRS4–6 in the developmental and independent test cohorts, respectively. The detection rate of HE was similar rates between the developmental and external testing cohorts, and HE occurred more frequently in deep ICH than in lobar. Although baseline NCCT was performed earlier in the external testing cohort than in the training or internal validation cohorts, there was no significant difference in the time to baseline NCCT between the HE and non-HE subgroups ( $Z=1.503$ ,  $P=0.133$ ). There were also no significant differences in ICH location, death, HE, poor outcome (mRS 4–6), or mortality rate between these three cohorts, indicating that there was good homology between the three cohorts for comparative analysis (Table 1). In addition, 40.7% (105/258) and 60.9% (53/87) of patients underwent brain CTA within 24 h. Comparisons of the

clinical characteristics and radiographic findings of patients with confirmed ICH between mRS 4–6 vs. mRS 0–3 and 30-day mortality versus survival are shown in Table 2.

#### **Rad score construction and model establishment**

A total of 30 patients were randomly selected from the developmental cohort for a consistency analysis of radiomic features. Of the 1,072 quantitative radiomic features, ICCs were above 0.7 for all except for 82 texture features, which were excluded from establishing the rad-score, indicating good interobserver agreement. Based on the training cohort, nine features were introduced into the rad-score formula (Appendix 1).

#### **Predictive performance of the clinical, rad score, and hybrid models**

Univariate logistic regression analysis identified age, initial GCS score  $\leq 8$ , time to baseline NCCT, deep ICH, baseline ICH volume, IVH, HE,  $NLR > 6$ , and systolic blood pressure (SBP) as significant predictors of a poor outcome and 30-day mortality. As shown in Table 3, with results reported as odds ratio [95% confidence interval (CI)], HE [2.457 (0.297, 2.633);  $P=0.014$ ], IVH [2.374

**Table 2** Clinical characteristics and radiographic findings of patients with a confirmed intracerebral hemorrhage

| Variables                 | Training and internal validation cohorts |                        |         |                               |                      |         | External validation cohort |                       |         |                               |                     |         |
|---------------------------|--|------------------------|---------|-------------------------------|----------------------|---------|----------------------------|-----------------------|---------|-------------------------------|---------------------|---------|
|                           | 30-day death                             |                        |         | Composite unfavorable outcome |                      |         | 30-day death               |                       |         | Composite unfavorable outcome |                     |         |
|                           | Survivor<br>(n=229)                      | Non-survivor<br>(n=29) | P value | mRS 0-3<br>(n=92)             | mRS 4-6<br>(n=166)   | P value | Survivor<br>(n=78)         | Non-survivor<br>(n=9) | P value | mRS 0-3<br>(n=36)             | mRS 4-6<br>(n=51)   | P value |
| Male, n (%)               | 153 (66.8)                               | 19 (65.5)              | 0.889   | 60 (23.3)                     | 112 (43.4)           | 0.783   | 60 (76.9)                  | 6 (66.7)              | 0.681   | 23 (63.9)                     | 43 (84.3)           | 0.053   |
| Age (years)               | 59.6±12.2                                | 62.7±10.5              |         | 62±10.7                       | 58.7±12.6            |         | 58.7±11.9                  | 59.5±13.1             |         | 57.9±12.5                     | 60.6±13.5           |         |
| <60                       | 108 (47.2)                               | 10 (34.5)              | 0.359   | 36 (14.0)                     | 82 (31.8)            | 0.024   | 42 (53.8)                  | 3 (33.3)              | 0.087   | 20 (55.6)                     | 25 (49.0)           | 0.334   |
| ≥60                       | 121 (52.8)                               | 19 (65.5)              | 0.197   | 56 (21.7)                     | 84 (32.6)            | 0.073   | 36 (46.2)                  | 6 (66.7)              | 0.304   | 16 (44.4)                     | 26 (51.0)           | 0.702   |
| Smoking (%)               | 86 (37.6)                                | 13 (44.8)              | 0.448   | 31 (12.0)                     | 68 (26.4)            | 0.155   | 26 (33.3)                  | 2 (22.2)              | 0.712   | 10 (27.8)                     | 18 (35.3)           | 0.494   |
| Alcohol consumption (%)   | 48 (21.0)                                | 8 (27.6)               | 0.415   | 17 (6.6)                      | 39 (15.1)            | 0.219   | 16 (20.5)                  | 1 (11.1)              | 0.682   | 6 (16.7)                      | 11 (21.6)           | 0.784   |
| Comorbidities, n (%)      |  |                        |         |                               |                      |         |                            |                       |         |                               |                     |         |
| Hypertension              | 224 (97.8)                               | 28 (96.6)              | 0.670   | 90 (34.9)                     | 162 (62.8)           | 0.635   | 76 (97.4)                  | 9 (100.0)             | 0.627   | 35 (97.2)                     | 50 (98.1)           | 0.802   |
| Diabetes                  | 38 (16.6)                                | 8 (27.6)               | 0.145   | 16 (6.2)                      | 30 (11.6)            | 0.517   | 4 (5.1)                    | 2 (22.2)              | 0.115   | 16 (44.4)                     | 30 (58.8)           | 0.517   |
| Hyperlipidemia            | 86 (37.6)                                | 15 (51.7)              | 0.141   | 36 (14.0)                     | 65 (25.2)            | 0.552   | 9 (11.5)                   | 1 (11.1)              | 1.000   | 3 (8.3)                       | 7 (13.7)            | 0.513   |
| Cerebrovascular disease   | 1 (0.4)                                  | 4 (13.8)               | <0.001  | 0 (0)                         | 5 (1.9)              | 0.108   | 4 (5.1)                    | 1 (11.1)              | 0.429   | 2 (5.6)                       | 3 (5.9)             | 1.000   |
| Heart failure             | 3 (1.2)                                  | 0 (-)                  | -       | 1 (0.4)                       | 2 (0.8)              | 0.710   | 1 (1.3)                    | 0 (-)                 | -       | 1 (2.8)                       | 0                   | -       |
| Renal insufficiency       | 10 (4.4)                                 | 5 (17.2)               | 0.005   | 5 (1.9)                       | 10 (3.9)             | 0.543   | 5 (6.4)                    | 2 (22.2)              | 0.152   | 2 (5.6)                       | 5 (9.8)             | 0.695   |
| Hepatic insufficiency     | 6 (2.6)                                  | 1 (3.4)                | 0.796   | 2 (0.8)                       | 5 (1.9)              | 0.517   | 1 (1.3)                    | 1 (11.1)              | 0.197   | 1 (2.8)                       | 1 (2.0)             | 1.000   |
| Warfarin use, n (%)       | 50 (21.8)                                | 10 (34.5)              | 0.129   | 17 (18.5)                     | 43 (25.9)            | 0.176   | 19 (24.4)                  | 3 (33.3)              | 0.686   | 5 (13.9)                      | 17 (33.3)           | 0.048   |
| Initial GCS score         |  |                        | <0.001  |                               |                      | <0.001  |                            |                       | <0.001  |                               |                     | <0.001  |
| ≤8                        | 43 (18.8)                                | 23 (79.3)              |         | 4 (4.3)                       | 62 (37.3)            |         | 18 (23.1)                  | 8 (88.9)              |         | 2 (5.6)                       | 24 (47.1)           |         |
| >8                        | 186 (81.2)                               | 6 (20.7)               |         | 88 (95.7)                     | 104 (62.7)           |         | 60 (76.9)                  | 1 (11.1)              |         | 34 (94.4)                     | 27 (52.9)           |         |
| Time to baseline NCCT (h) | 3.0 (2.0, 8.0)                           | 2.0 (1.0, 4.0)         | 0.040   | 5.0 (2.0, 20.0)               | 3.0 (1.0, 5.0)       | <0.001  | 1.25 (1.0, 4.25)           | 1.0 (0.85, 3.25)      | 0.382   | 4.0 (1.0, 12.0)               | 1.0 (1.0, 2.0)      | <0.001  |
| Baseline ICH volume (mL)  | 14.8<br>(5.6, 29.0)                      | 25.0<br>(13.7, 60.2)   |         | 5.7<br>(2.2, 15.4)            | 21.6<br>(11.1, 43.8) |         | 13.1<br>(5.2, 30.6)        | 42.3<br>(27.6, 69.3)  |         | 9.6<br>(3.3, 34.0)            | 19.4<br>(9.5, 42.3) |         |
| <30                       | 172 (75.5)                               | 15 (51.7)              | 0.007   | 85 (92.4)                     | 103 (62.0)           | <0.001  | 59 (75.6)                  | 2 (22.2)              | 0.003   | 32 (88.9)                     | 29 (56.9)           | 0.001   |
| ≥30                       | 56 (24.5)                                | 14 (48.3)              | 0.007   | 7 (7.6)                       | 63 (38.0)            | <0.001  | 19 (24.4)                  | 7 (77.8)              | 0.001   | 4 (11.1)                      | 22 (43.1)           |         |

**Table 2** (continued)

Table 2 (continued)

| Variables                           | Training and internal validation cohorts |                     |                               |                | External validation cohort |         |                               |                    |         |                |                |         |
|-------------------------------------|--|---------------------|-------------------------------|----------------|----------------------------|---------|-------------------------------|--------------------|---------|----------------|----------------|---------|
|                                     | 30-day death                             |                     | Composite unfavorable outcome |                | 30-day death               |         | Composite unfavorable outcome |                    |         |                |                |         |
|                                     | Survivor (n=229)                         | Non-survivor (n=29) | P value                       | mRS 0-3 (n=92) | mRS 4-6 (n=166)            | P value | Survivor (n=78)               | Non-survivor (n=9) | P value | mRS 0-3 (n=36) | mRS 4-6 (n=51) | P value |
| ICH location, n (%)                 |  |                     | 0.143                         |                |                            | <0.001  |                               |                    | 0.424   |                |                | 0.004   |
| Deep                                | 196 (85.6)                               | 28 (96.6)           |                               | 69 (75.0)      | 155 (93.4)                 |         | 61 (78.2)                     | 6 (66.7)           |         | 22 (61.1)      | 45 (88.2)      |         |
| Lobar                               | 33 (14.4)                                | 1 (3.4)             |                               | 23 (25.0)      | 11 (6.6)                   |         | 17 (21.8)                     | 3 (33.3)           |         | 14 (38.9)      | 6 (11.8)       |         |
| IVH, n (%)                          | 78 (34.1)                                | 23 (79.3)           | <0.001                        | 15 (16.3)      | 86 (51.8)                  | <0.001  | 31 (39.7)                     | 6 (66.7)           | 0.161   | 7 (19.4)       | 30 (58.8)      | 0.001   |
| HE in 24 h                          | 48 (21.0)                                | 16 (55.2)           | <0.001                        | 5 (5.4)        | 59 (35.5)                  | <0.001  | 14 (17.9)                     | 6 (66.7)           | 0.004   | 1 (2.8)        | 19 (37.3)      | <0.001  |
| NLR <sup>x</sup>                    | 6.0 (3.2, 10.8)                          | 12.8 (6.2, 16.5)    | 0.001                         | 4.0 (2.9, 7.2) | 7.9 (4.4, 14.3)            | <0.001  | 3.6 (2.1, 7.3)                | 3.3 (1.9, 5.0)     | 0.549   | 3.8 (2.4, 8.4) | 3.2 (2.0, 5.7) | 0.459   |
| Serum calcium (mmol/L) <sup>y</sup> | 2.31±0.14                                | 2.22±0.15           | 0.004                         | 2.3±0.14       | 2.29±0.15                  | 0.036   | 2.31±0.14                     | 2.36±0.12          | 0.320   | 2.31±0.12      | 2.33±0.15      | 0.525   |
| SBP (mmHg)                          | 172±2.7                                  | 178±30.5            | 0.294                         | 166±28.4       | 176±27.5                   | 0.013   | 172±32                        | 189±39             | 0.139   | 163 ± 29       | 181 ± 33       | 0.009   |
| Radiological signs, n (%)           |  |                     |                               |                |                            |         |                               |                    |         |                |                |         |
| Blend sign                          | 39 (17.3)                                | 8 (28.6)            | 0.146                         | 8 (8.7)        | 39 (24.1)                  | 0.002   | 15 (19.2)                     | 2 (22.2)           | 0.658   | 4 (11.1)       | 13 (25.5)      | 0.103   |
| Black hole sign                     | 18 (8.0)                                 | 5 (17.9)            | 0.151                         | 5 (5.4)        | 18 (11.1)                  | 0.173   | 8 (10.3)                      | 1 (11.1)           | 0.608   | 2 (5.6)        | 7 (13.7)       | 0.291   |
| Satellite or island signs           | 59 (26.1)                                | 11 (39.3)           | 0.141                         | 10 (10.9)      | 60 (37.0)                  | <0.001  | 8 (10.3)                      | 3 (33.3)           | 0.064   | 2 (5.6)        | 9 (17.6)       | 0.108   |
| Spot sign on CTA <sup>z</sup>       | 2 (1.4)                                  | 1 (8.3)             | 0.219                         | 0              | 3 (3.1)                    | -       | 2 (2.6)                       | 2 (22.2)           | 0.013   | 0              | 4 (7.8)        | -       |

Data were presented as mean ± SD, median (interquartile range) or n (%), unless otherwise stated. x, missing data in 2/258 (0.8%) cases; y, missing data in 1/258 (0.4%) case; z, missing data in 105/258 (40.7%) cases. GCS, glasgow coma scale; HE, hematoma expansion; ICH, intracerebral hemorrhage; IVH, intraventricular hemorrhage; NCCT, noncontrast computed tomography; SD, standard deviation. SBP, systolic blood pressure; NLR, neutrophils to lymphocyte ratio; mRS, modified ranking score; CTA, computed tomography angiography.



**Table 3** Comparison of the three prediction models based on stepwise multivariate analyses for prediction of a poor outcome after an ICH

| Models          | Adjusted OR (95% CI)    | P value | AIC     | LRT ( $\chi^2$ ) |
|-----------------|-------------------------|---------|---------|------------------|
| Clinical model  |                         |         | 190.610 | 0.263            |
| Location (deep) | -2.268 (-2.578, -0.188) | 0.023   |         |                  |
| HE              | 2.457 (0.297, 2.633)    | 0.014   |         |                  |
| IVH             | 2.374 (0.180, 1.882)    | 0.018   |         |                  |
| Radiomics model |                         |         | 153.095 | 0.364            |
| Rad score       | 3.049 (2.132, 4.360)    | <0.001  |         |                  |
| Hybrid model    |                         |         | 143.069 | 0.449            |
| Location (deep) | -2.291 (-2.925, -0.228) | 0.022   |         |                  |
| Rad score       | 5.255 (0.680, 11.460)   | <0.001  |         |                  |
| IVH             | 1.889 (-0.035, 1.897)   | 0.059   |         |                  |
| HE              | 1.478 (-0.351, 2.503)   | 0.139   |         |                  |

ICH, intracerebral hemorrhage; HE, hematoma expansion; IVH, intraventricular hemorrhage; OR, odds ratio; CI, confidence interval; AIC, Akaike information criterion; LRT, Likelihood ratio test.

**Table 4** Performance of the three models in the prediction of the outcome following an intracerebral hemorrhage

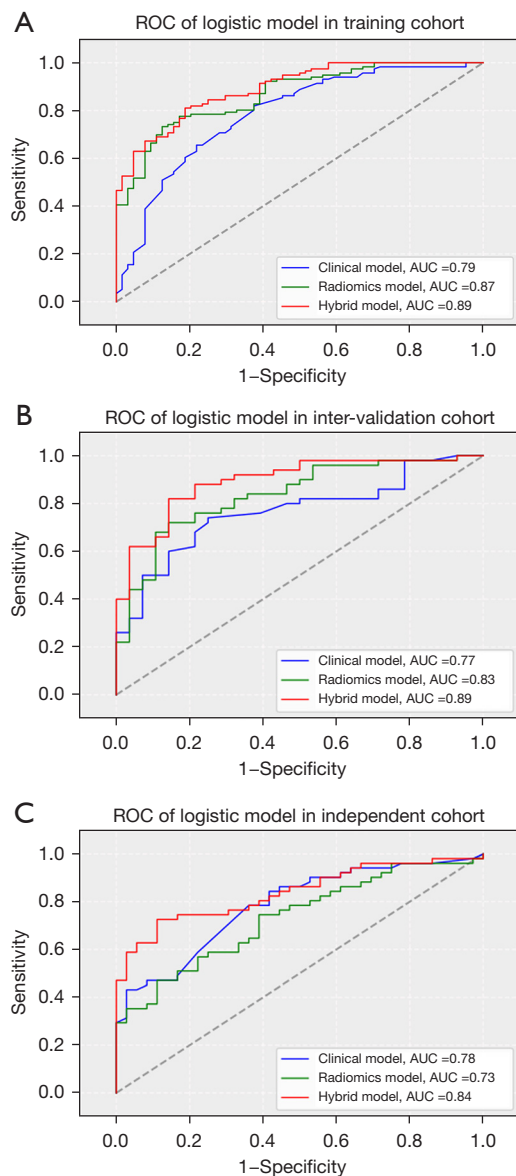
| Cohorts                    | Poor outcome (mRS4-6) |             |             | 30-day mortality |             |             |
|----------------------------|-----------------------|-------------|-------------|------------------|-------------|-------------|
|                            | AUC (95% CI)          | Sensitivity | Specificity | AUC (95% CI)     | Sensitivity | Specificity |
| Training cohort            |                       |             |             |                  |             |             |
| Clinical model             | 0.785 (0.714-0.857)   | 0.871       | 0.516       | 0.831            | 0.762       | 0.730       |
| Radiomics model            | 0.867 (0.815-0.918)   | 0.750       | 0.828       | 0.766            | 0.571       | 0.786       |
| Hybrid model               | 0.892 (0.847-0.937)   | 0.862       | 0.672       | 0.840            | 0.238       | 0.987       |
| Internal validation cohort |                       |             |             |                  |             |             |
| Clinical model             | 0.766 (0.659-0.872)   | 0.820       | 0.500       | 0.809            | 0.778       | 0.739       |
| Radiomics model            | 0.834 (0.742-0.927)   | 0.620       | 0.893       | 0.775            | 0.556       | 0.898       |
| Hybrid model               | 0.893 (0.820-0.966)   | 0.820       | 0.857       | 0.823            | 0.222       | 1.000       |
| External testing cohort    |                       |             |             |                  |             |             |
| Clinical model             | 0.783 (0.689-0.879)   | 0.902       | 0.389       | 0.880            | 1.000       | 0.705       |
| Radiomics model            | 0.731 (0.627-0.836)   | 0.784       | 0.528       | 0.749            | 0.667       | 0.769       |
| Hybrid model               | 0.838 (0.755-0.920)   | 0.863       | 0.528       | 0.883            | 0.111       | 0.987       |

mRS, modified ranking score; AUC, area under the curve; CI, confidence interval; HE, hematoma expansion.

(0.180, 1.882);  $P=0.018$ ], and location [-2.268 (-2.578, -0.188);  $P=0.023$ ] were independently associated with a poor outcome following ICH in the clinical model. In the hybrid model, location [-2.291 (-2.925, -0.228);  $P=0.022$ ] and rad-score [5.255 (0.680, 11.460);  $P<0.001$ ] were independently associated with a poor outcome. The hybrid model (AIC=143.069,  $\chi^2=0.449$ ) had the lowest AIC

and the highest LRT chi-square values compared with the radiomics model (AIC=153.095,  $\chi^2=0.364$ ) and the clinical model (AIC=190.610,  $\chi^2=0.263$ ). A bar chart was used to intuitively display the discriminability of the rad score, as shown in [Figure S3](#).

The performance of the three models for predicting a poor outcome is shown in [Table 4](#). The hybrid model



**Figure 2** ROC curves of the clinical, radiomic, and hybrid models. (A) training cohort, (B) internal validation cohort, and (C) external testing cohort. The ROC was based on the confusion matrix, and the DeLong test was used to compare the discriminability of the three models. The hybrid model (red curve) had the highest AUC for the prediction of a poor outcome in all three cohorts. However, this was not statistically different from the AUCs of the clinical (blue curve), radiomic (green curve), and hybrid models (red curve) in the external testing cohort. ROC, receiver operating characteristic; AUC, area under the curve.

achieved satisfactory discrimination, with AUCs of 0.892 (95% CI: 0.847 to 0.937), 0.893 (95% CI: 0.820 to 0.966), and 0.838 (95% CI: 0.755 to 0.920) in the training, internal validation, and external testing cohorts, respectively (Figure 2). The hybrid model yielded the highest AUCs for poor outcome in both the training (hybrid *vs.* clinical,  $z=3.116$ ,  $P=0.0018$ ; hybrid *vs.* radiomics,  $z=1.770$ ,  $P=0.077$ ) and internal validation cohorts (hybrid *vs.* clinical,  $z=2.162$ ,  $P=0.031$ ; hybrid *vs.* radiomics,  $z=2.799$ ,  $P=0.005$ ), meanwhile, the radiomics model had the lowest predictive ability in the external testing cohort (hybrid *vs.* radiomics,  $z=3.904$ ,  $P<0.001$ ). There was no significant difference in the ROC curves of the three models in the external testing cohort (DeLong test,  $P=0.819$ ), although a relatively low specificity for detecting a poor outcome was calculated.

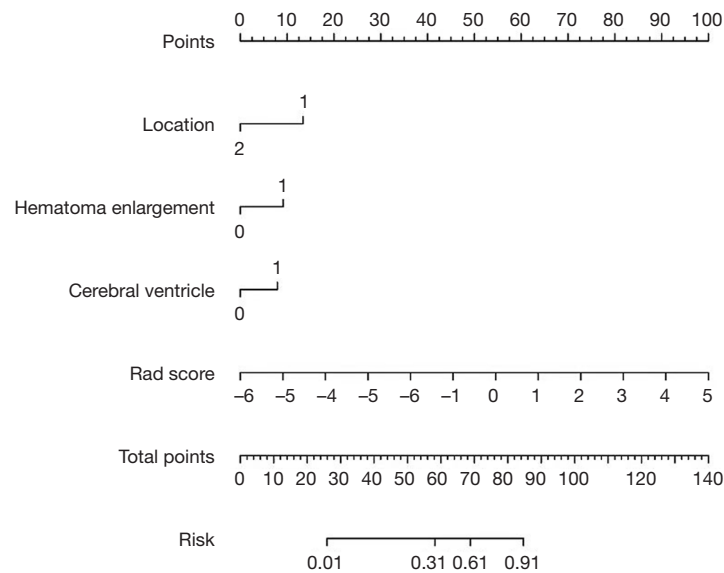
For the prediction of 30-day mortality, the hybrid model also achieved good discriminability, with AUCs of 0.840, 0.823, and 0.883 in the training, internal validation, and external testing cohorts, respectively. The rad score (2.861, 1.940, 4.220;  $P<0.001$ ) was the predominant risk factor associated with 30-day mortality.

### Nomogram

Based on these independent risk factors, a nomogram was established to predict a poor outcome after ICH (Figure 3). The rad score comprised most of the scoring system compared with other factors, including location, HE, and IVH, indicating a predominant role of quantitative radiomic parameters in predicting a poor outcome. Calibration and DCA showed favorable agreement on the probability of a poor outcome between nomogram estimation and actual observation in both the internal validation and external datasets (Appendix 1, Figure S4).

### Discussion

Herein, we developed and validated a hybrid model nomogram for predicting a poor outcome following the externally validated ICH. The AUCs were 0.905 (0.868–0.940), 0.886 (0.819–0.947), and 0.861 (0.795–0.922) in the training, internal validation, and external testing cohorts, respectively, which showed that our nomogram can be easily translated into routine clinical practice.



**Figure 3** Hybrid nomogram for predicting a poor outcome in patients with an ICH. This nomogram was developed using the rad-score and clinical parameters via multivariate logistic regression analysis of the training cohort. The range in rad-score was  $-6$  to  $5$ . the rad-score comprised most of the scoring system compared with other factors, including location (deep *vs.* lobar), hematoma enlargement, and cerebral-ventricle. ICH, intracerebral hemorrhage; rad-score, radiomics score; cerebral ventricle, intracerebral ventricular hemorrhage.

Radiomics based on initial NCCT showed added value for predicting a poor outcome after ICH compared with some traditional predictive models based on clinical parameters (17–20). Moreover, the hybrid model was more accurate at predicting a poor outcome and 30-day mortality in patients with deep ICH.

Our clinical model identified deep ICH, HE, and IVH as independent risk factors for a poor outcome after ICH. Typically, ICH occurs in deep locations such as the basal ganglia, thalamus, brain stem, internal capsule, and corpus callosum. This can lead to more serious damage to white matter fiber due to mass effect and sequential neurodegeneration, characterized by extreme loss of muscle strength. Most often, IVH occurs after a deep ICH, especially in the thalamus and caudate, which leads to worse short/long-term prognoses and a mortality of more than 50% (21). Witsch *et al.* (22) reported that 19 of 282 ICH patients developed a delayed IVH, although this did not appear to portend a worse outcome. In accordance with published literature, our results have shown that HE is an independent risk factor for a poor outcome/30-day mortality in ICH patients, which supports that HE is a critical target for preventing deterioration during the acute phase of an ICH (23,24).

The rad-score, based on NCCT radiomic features,

was the predominant risk factor in the hybrid model for predicting a poor outcome and 30-day mortality after ICH. Previous studies have identified several radiological signs, such as the blend sign, swirl sign, black hole sign, hypodensity, density heterogeneity and irregular shape based on NCCT, and the spot sign shown on CTA that can identify an early HE after ICH (6,25). The spot sign strongly predicts HE in various studies (26). However, CTA is still not widely accepted due to contraindications for the use of an iodine-based contrast agent. A predictive model that combines radiomics features with clinical characteristics can better discriminate early HE when compared to the radiological, clinical-only, or clinical-radiological features (27,28). The presented rad-score includes nine features that define the nature of the ICH's size, shape, and heterogeneity, which can permit a more objective prediction of HE in comparison with visual radiological signs. This rad-score for the prediction of poor outcome is different from a rad-score previously proposed for the prediction of hematoma expansion (29), including some morphology parameters (original\_shape\_Maximum2DDiameterColum, original\_shape\_MinorAxisLength), means that the volume, and shape of baseline hematoma may be more relevant to the prognosis of ICH.

For prediction of a poor outcome, baseline ICH volume

and initial GCS score are often considered risk factors for an adverse outcome after ICH (2,30). In our training cohort, a median of hematoma volume of 21.6 mL was related to mRS 4–6. This may imply three hypotheses: first, it may not be safe if the baseline hematoma volume is less than 30 mL or even 20 mL, and early HE should be identified and monitored as soon as possible; second, GCS and National Institutes of Health Stroke Scale (NIHSS) are important clinical scale tools for evaluating early-stage ICH and prognosticating medium and long-term outcomes (30); third, as mentioned above, the treatment strategy should be based on ICH location, as ICH that occurs in the brain stem, internal capsule, and thalamus might lead to worse outcomes even if its volume is less than 20 mL (31). Due to sample size limitations, a baseline critical ICH volume for every ICH location is not proposed in this study. Our rad-score's role in the prediction of a poor outcome following ICH may be due to its inherent feature involved ICH volume information. In our study, ICH patients from 2018 to 2021 were analyzed according to the Guidelines for the Management of Spontaneous Intracerebral Hemorrhage of the AHA (Ver. 2015), which minimizes the effects of various therapy strategies. In addition, perihemorrhagic edema correlates with functional outcome, although it usually peaks three days after ICH (32); therefore, it cannot be estimated on the first CT scan.

Although the AUC of the hybrid model was not significantly higher than those of the other two models, our nomogram combining clinical and NCCT radiomic features on admission had a favorable performance when validated using the external testing dataset. However, the relatively smaller external cohort size might lead to bias. In addition, the hybrid model could stratify ICH patients into low-, medium- and high-risk groups for developing a poor outcome or mortality. With its higher prediction performance and simplicity than radiologic scores alone, we propose that the hybrid model could be used as a preliminary screening and triage tool at the time of hospital admission to identify those at risk of a poor outcome. Further, the model could be used to select and/or stratify patients in clinical trials to homogenize the patient sample.

Our study had several limitations. First, selection bias was unavoidable due to the limited and unbalanced sample size; a high ratio in deep hemorrhage negatively affects the generalizability of the hybrid model. Second, the nature of this study was retrospective in design, and its sample size was small. A larger prospective multi-center study is needed to provide more insight into this issue. Third, accuracy

decreases with an irregular, hypodense hematoma (<50 HU). Therefore, manual volumetric segmentation with isotropic images might be the best choice if time is not a consideration. However, an auto-segmentation method based on a deep-learning technique would be more efficient and practical (33).

## Conclusions

We developed a radiomics clinical nomogram for predicting a poor outcome following an ICH. Internal and external validation of the nomogram confirmed the accuracy of this model. We found that rad-score-based NCCT combined with IVH may accurately predict a poor outcome following an ICH. Further studies using an auto-segmentation method are needed to determine whether the nomogram could be applied to other patient cohorts.

## Acknowledgments

*Funding:* This work was supported by a grant from the Wuhan Municipal Health Commission of Hubei Province, China (No. WX21B15).

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-22-128/rc>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-128/coif>). JC is employed by GE and only provided technical support during the research. The other authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Medical Ethics Committee of the Central Hospital of Wuhan (No. 2021-36). The informed consent of Central Hospital of Wuhan (CHW) for this retrospective research was waived, and the written informed consent was provided by all participants of the Fifth Affiliated Hospital of Nanchang

University (FAHNU) before their participation in the study.

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Fernando SM, Qureshi D, Talarico R, Tanuseputro P, Dowlatshahi D, Sood MM, Smith EE, Hill MD, McCredie VA, Scales DC, English SW, Rochweg B, Kyeremanteng K. Intracerebral Hemorrhage Incidence, Mortality, and Association With Oral Anticoagulation Use: A Population Study. *Stroke* 2021;52:1673-81.
2. Chen HS, Hsieh CF, Chau TT, Yang CD, Chen YW. Risk factors of in-hospital mortality of intracerebral hemorrhage and comparison of ICH scores in a Taiwanese population. *Eur Neurol* 2011;66:59-63.
3. Morotti A, Arba F, Boulouis G, Charidimou A. Noncontrast CT markers of intracerebral hemorrhage expansion and poor outcome: A meta-analysis. *Neurology* 2020;95:632-43.
4. Flibotte JJ, Hagan N, O'Donnell J, Greenberg SM, Rosand J. Warfarin, hematoma expansion, and outcome of intracerebral hemorrhage. *Neurology* 2004;63:1059-64.
5. Radmanesh F, Falcone GJ, Anderson CD, Battey TW, Ayres AM, Vashkevich A, McNamara KA, Schwab K, Romero JM, Viswanathan A, Greenberg SM, Goldstein JN, Rosand J, Brouwers HB. Risk factors for computed tomography angiography spot sign in deep and lobar intracerebral hemorrhage are shared. *Stroke* 2014;45:1833-5.
6. Rodriguez-Luna D, Coscojuela P, Rodriguez-Villatoro N, Juega JM, Boned S, Muchada M, Pagola J, Rubiera M, Ribo M, Tomasello A, Demchuk AM, Goyal M, Molina CA. Multiphase CT Angiography Improves Prediction of Intracerebral Hemorrhage Expansion. *Radiology* 2017;285:932-40.
7. Hemphill JC 3rd, Greenberg SM, Anderson CS, Becker K, Bendok BR, Cushman M, Fung GL, Goldstein JN, Macdonald RL, Mitchell PH, Scott PA, Selim MH, Woo D; Council on Clinical Cardiology. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke* 2015;46:2032-60.
8. Du C, Liu B, Yang M, Zhang Q, Ma Q, Ruili R. Prediction of Poor Outcome in Intracerebral Hemorrhage Based on Computed Tomography Markers. *Cerebrovasc Dis* 2020;49:556-62.
9. Dowlatshahi D, Morotti A, Al-Ajlan FS, Boulouis G, Warren AD, Petrcich W, Aviv RI, Demchuk AM, Goldstein JN. Interrater and Intrarater Measurement Reliability of Noncontrast Computed Tomography Predictors of Intracerebral Hemorrhage Expansion. *Stroke* 2019;50:1260-2.
10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441-6.
11. Li H, Xie Y, Wang X, Chen F, Sun J, Jiang X. Radiomics features on non-contrast computed tomography predict early enlargement of spontaneous intracerebral hemorrhage. *Clin Neurol Neurosurg* 2019;185:105491.
12. Xie H, Ma S, Wang X, Zhang X. Noncontrast computer tomography-based radiomics model for predicting intracerebral hemorrhage expansion: preliminary findings and comparison with conventional radiological model. *Eur Radiol* 2020;30:87-98.
13. Lattanzi S, Cagnetti C, Rinaldi C, Angelocola S, Provinciali L, Silvestrini M. Neutrophil-to-lymphocyte ratio improves outcome prediction of acute intracerebral hemorrhage. *J Neurol Sci* 2018;387:98-102.
14. Chen W, Wang L, Chen J. Considering Blood Pressure Level in the Association Between Serum Calcium Level and the Size and Expansion in Patients With Intracerebral Hemorrhage. *JAMA Neurol* 2017;74:483.
15. Law ZK, Ali A, Krishnan K, Bischoff A, Appleton JP, Scutt P, Woodhouse L, Pszczolkowski S, Cala LA, Dineen RA, England TJ, Ozturk S, Roffe C, Bereczki D, Ciccone A, Christensen H, Ovesen C, Bath PM, Sprigg N; TICH-2 Investigators. Noncontrast Computed Tomography Signs as Predictors of Hematoma Expansion, Clinical Outcome, and Response to Tranexamic Acid in Acute Intracerebral Hemorrhage. *Stroke* 2020;51:121-8.
16. Zhang X, Gao Q, Chen K, Wu Q, Chen B, Zeng S, Fang X. A predictive nomogram for intracerebral hematoma expansion based on non-contrast computed tomography and clinical features. *Neuroradiology* 2022;64:1547-56.

17. Miyahara M, Noda R, Yamaguchi S, Tamai Y, Inoue M, Okamoto K, Hara T. New Prediction Score for Hematoma Expansion and Neurological Deterioration after Spontaneous Intracerebral Hemorrhage: A Hospital-Based Retrospective Cohort Study. *J Stroke Cerebrovasc Dis* 2018;27:2543-50.
18. Safatli DA, Günther A, Schlattmann P, Schwarz F, Kalff R, Ewald C. Predictors of 30-day mortality in patients with spontaneous primary intracerebral hemorrhage. *Surg Neurol Int* 2016;7:S510-7.
19. Widyadharma IPE, Krishna A, Soejitno A, Laksmidewi AAAP, Tini K, Putra IBK, Budiarsa IGN, Indrayani IAS. Modified ICH score was superior to original ICH score for assessment of 30-day mortality and good outcome of non-traumatic intracerebral hemorrhage. *Clin Neurol Neurosurg* 2021;209:106913.
20. He XW, Chen MD, Du CN, Zhao K, Yang MF, Ma QF. A novel model for predicting the outcome of intracerebral hemorrhage: Based on 1186 Patients. *J Stroke Cerebrovasc Dis* 2020;29:104867.
21. Hanley DF, Lane K, McBee N, Ziai W, Tuhirim S, Lees KR, et al. Thrombolytic removal of intraventricular haemorrhage in treatment of severe stroke: results of the randomised, multicentre, multiregion, placebo-controlled CLEAR III trial. *Lancet* 2017;389:603-11.
22. Witsch J, Bruce E, Meyers E, Velazquez A, Schmidt JM, Suwatcharangkoon S, Agarwal S, Park S, Falo MC, Connolly ES, Claassen J. Intraventricular hemorrhage expansion in patients with spontaneous intracerebral hemorrhage. *Neurology* 2015;84:989-94.
23. Li Z, You M, Long C, Bi R, Xu H, He Q, Hu B. Hematoma Expansion in Intracerebral Hemorrhage: An Update on Prediction and Treatment. *Front Neurol* 2020;11:702.
24. Chen S, Zhao B, Wang W, Shi L, Reis C, Zhang J. Predictors of hematoma expansion predictors after intracerebral hemorrhage. *Oncotarget* 2017;8:89348-63.
25. Sporns PB, Schwake M, Kemmling A, Minnerup J, Schwindt W, Niederstadt T, Schmidt R, Hanning U. Comparison of Spot Sign, Blend Sign and Black Hole Sign for Outcome Prediction in Patients with Intracerebral Hemorrhage. *J Stroke* 2017;19:333-9.
26. Fu F, Sun S, Liu L, Gu H, Su Y, Li Y. Iodine Sign as a Novel Predictor of Hematoma Expansion and Poor Outcomes in Primary Intracerebral Hemorrhage Patients. *Stroke* 2018;49:2074-80.
27. Song Z, Guo D, Tang Z, Liu H, Li X, Luo S, Yao X, Song W, Song J, Zhou Z. Noncontrast Computed Tomography-Based Radiomics Analysis in Discriminating Early Hematoma Expansion after Spontaneous Intracerebral Hemorrhage. *Korean J Radiol* 2021;22:415-24.
28. Chen Q, Zhu D, Liu J, Zhang M, Xu H, Xiang Y, Zhan C, Zhang Y, Huang S, Yang Y. Clinical-radiomics Nomogram for Risk Estimation of Early Hematoma Expansion after Acute Intracerebral Hemorrhage. *Acad Radiol* 2021;28:307-17.
29. Li H, Xie Y, Liu H, Wang X. Non-Contrast CT-Based Radiomics Score for Predicting Hematoma Enlargement in Spontaneous Intracerebral Hemorrhage. *Clin Neuroradiol* 2022;32:517-28.
30. Yogendrakumar V, Moores M, Sikora L, Shamy M, Ramsay T, Fergusson D, Dowlatshahi D. Evaluating Hematoma Expansion Scores in Acute Spontaneous Intracerebral Hemorrhage: A Systematic Scoping Review. *Stroke* 2020;51:1305-8.
31. Huang K, Ji Z, Sun L, Gao X, Lin S, Liu T, Xie S, Zhang Q, Xian W, Zhou S, Gu Y, Wu Y, Wang S, Lin Z, Pan S. Development and Validation of a Grading Scale for Primary Pontine Hemorrhage. *Stroke* 2017;48:63-9.
32. Volbers B, Giede-Jeppe A, Gerner ST, Sembill JA, Kuramatsu JB, Lang S, Lücking H, Staykov D, Huttner HB. Peak perihemorrhagic edema correlates with functional outcome in intracerebral hemorrhage. *Neurology* 2018;90:e1005-12.
33. Wang JL, Jin GL, Yuan ZG. Artificial neural network predicts hemorrhagic contusions following decompressive craniotomy in traumatic brain injury. *J Neurosurg Sci* 2021;65:69-74.

**Cite this article as:** Xie Y, Chen F, Li H, Wu Y, Fu H, Zhong Q, Chen J, Wang X. Development and validation of a clinical-radiomics nomogram for predicting a poor outcome and 30-day mortality after a spontaneous intracerebral hemorrhage. *Quant Imaging Med Surg* 2022;12(10):4900-4913. doi: 10.21037/qims-22-128

## Methods

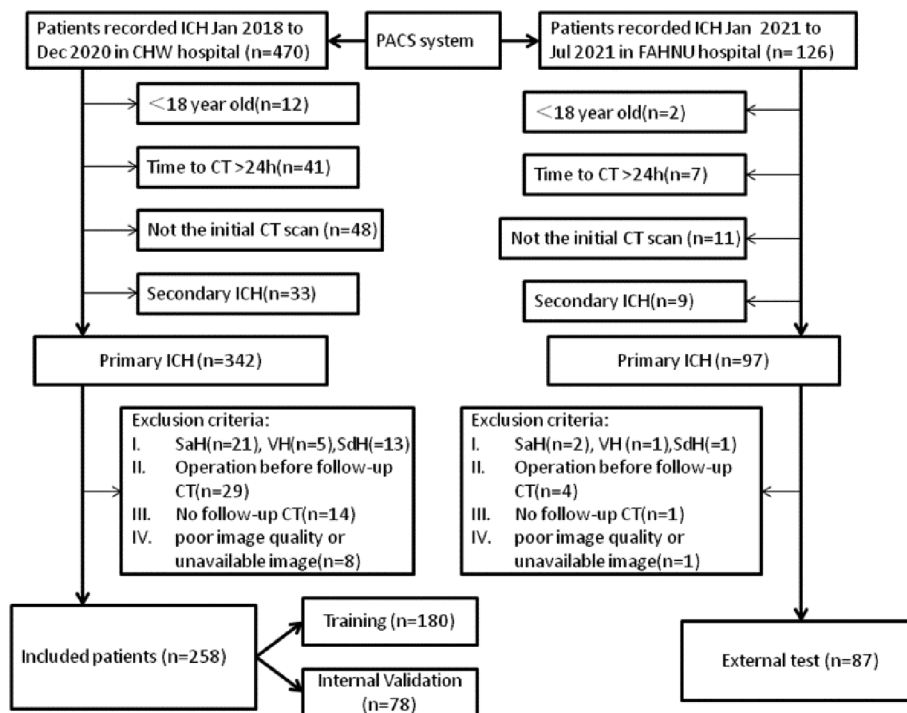
### Patients

We retrospectively reviewed data for 258 patients hospitalized for a spontaneous intracerebral hemorrhage (ICH) at the Central Hospital of Wuhan (CHW) as either training or internal validation cohort, 87 ICH patients at the Fifth Affiliated Hospital of Nanchang University (FAHNU) were prospectively enrolled as an external test cohort. Patient selection process is shown in *Figure S1*.

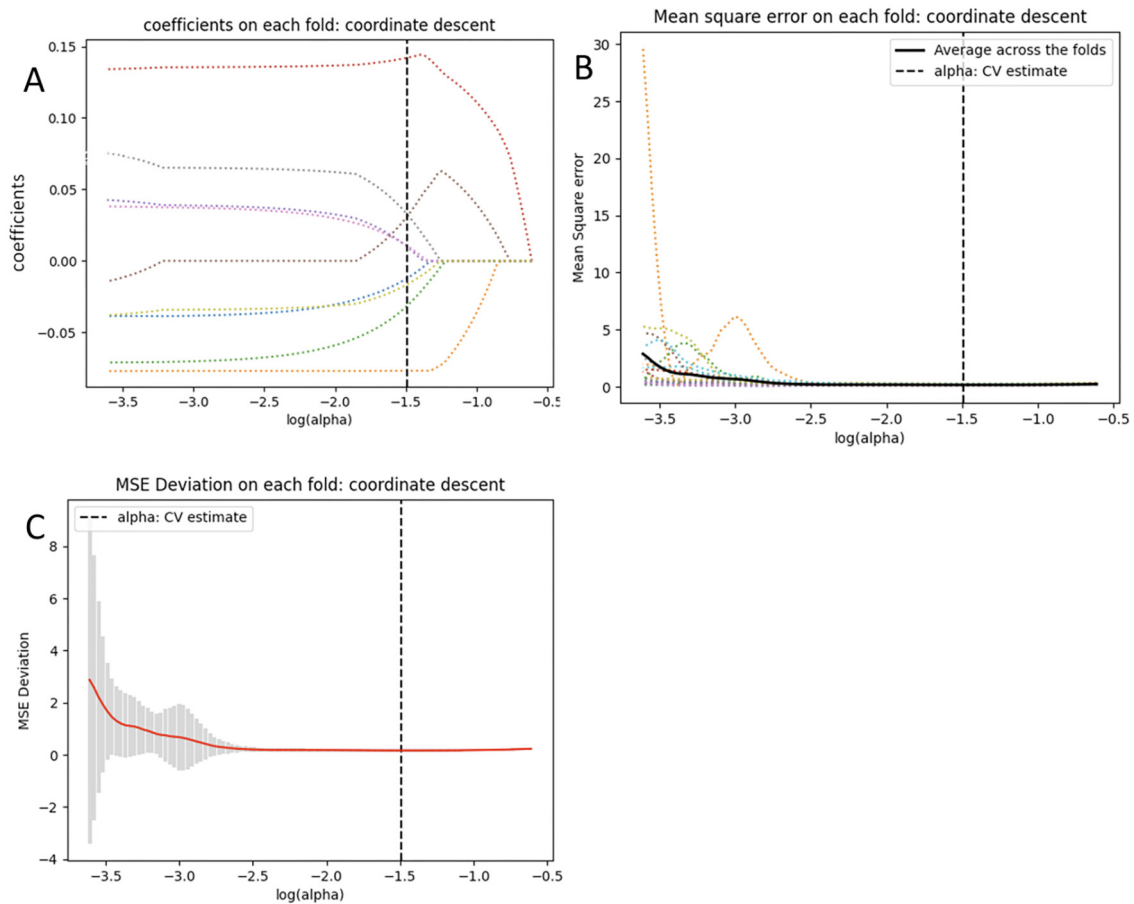
## Results

### Establishment of the rad-score

The least absolute shrinkage and selection operator method (LASSO) is a popular method for the regression of high-dimensional predictors (34,35)\_ENREF\_6. The method uses an L1 penalty to shrink some regression coefficients to exactly zero (*Figure S2*). Rad-score =  $(\sum \beta_i * X_i) + \text{Intercept}$  ( $i=0, 1, 2, 3, \dots$ ), where  $X_i$  represents the  $i^{\text{th}}$  selected feature and  $\beta_i$  is its



**Figure S1** Flowchart of the patient selection process. ICH, intracerebral hemorrhage; CT, computed tomography; CHW, Central Hospital of Wuhan; FAHNU, Fifth Affiliated Hospital of Nanchang University.



**Figure S2** Texture feature selection using the least absolute shrinkage and selection operator (LASSO) logistic regression models. (A) LASSO coefficient profiles of the selected 9 features among 1072 features. (B) Mean square error on each fold. (C) Deviation of mean square error on each fold.

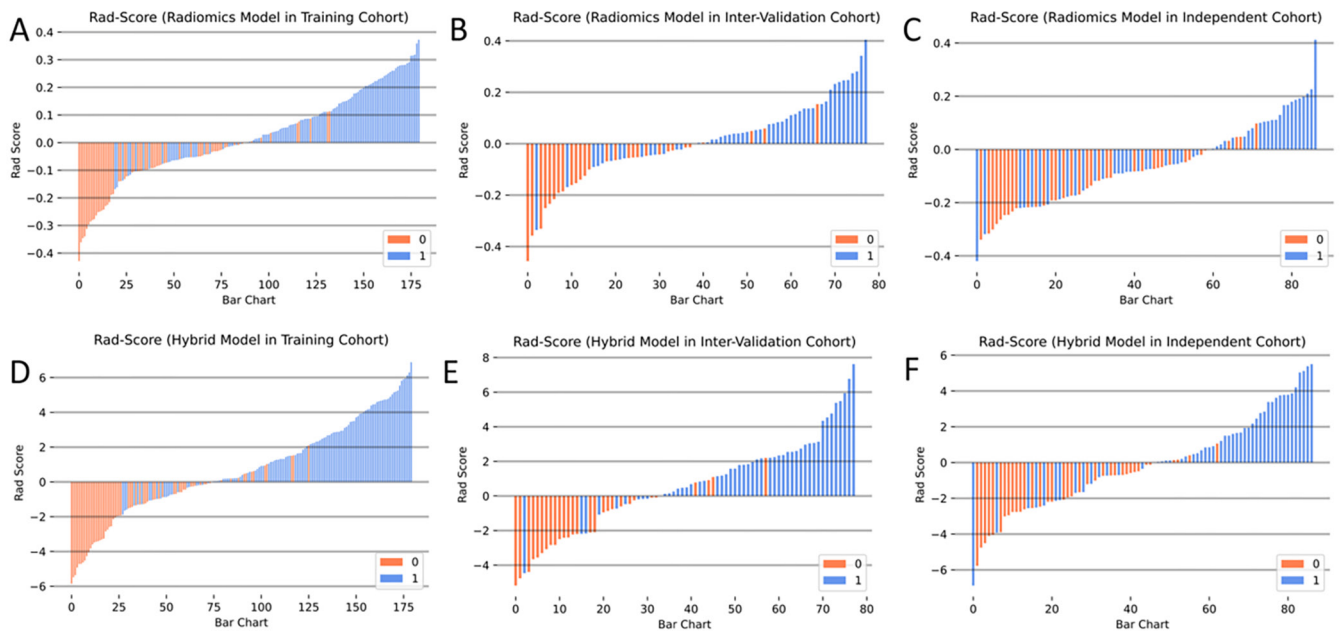
coefficient. The radiomic score formula is as follows:

$$\begin{aligned}
 \text{Radscore} = & -0.27896054 * \log\text{-sigma-2-0-mm-3D\_firstorder\_RootMeanSquared} \\
 & + -0.445686359 * \log\text{-sigma-2-0-mm-3D\_gldm\_DependenceVariance} \\
 & + -0.421542144 * \log\text{-sigma-3-0-mm-3D\_gldm\_ClusterTendency} \\
 & + 0.819735954 * \text{original\_shape\_Maximum2DDiameterColumn} \\
 & + 0.528278381 * \text{original\_shape\_MinorAxisLength} \\
 & + 0.120289332 * \text{wavelet-HHH\_gldm\_LargeDependenceEmphasis} \\
 & + 0.248174582 * \text{wavelet-HHH\_gldm\_LongRunHighGrayLevelEmphasis} \\
 & + 0.236439734 * \text{wavelet-LHH\_gldm\_LargeDependenceEmphasis} \\
 & + -0.251098979 * \text{wavelet-LLH\_gldm\_DependenceVariance} + 0.39407328
 \end{aligned}$$

### **Model construction, calibration and validation**

All 258 SICH patients from CHW were randomly divided into a training dataset (n=180) and an internal validation dataset (n=78) according to a 7:3 ratio for models establishment and validation, finally 87 SICH patients from FAHNU were enrolled for models test. Three models, including radiomics (rad-score based), clinical (clinical factors based), hybrid (rad-score combined with clinical factors) models were established respectively. Nomogram of hybrid model was constructed.





**Figure S3** Bar chart of Rad-Score in the training (left), internal validation (middle) and external testing (right) cohorts. (A-C) Rad-Score model; (D-F) hybrid model. The y-axis refers to the Rad-Score minus the optimal cutoff value. Upper and lower bars refer to the predicted positive and negative poor outcomes, respectively. Blue and orange bars refer to actual positive and negative poor outcomes, respectively.

### Models construction

*Rad-score model:* According to the dichotomy criterion of mRS 0-3 or mRS 4-6, the Rad-score model was established in the training cohort by the following equations:  $\text{logit } \pi_{\text{Rad-Score}} = 1.002 + 0.884 \times \text{Rad-Score}$  (Figure S3).

*Clinical model:* We firstly used univariate logistic regression analysis for screening clinical independent risk factors. Then, multivariate logistic regression model was applied for clinical-based model construction according to those independent risk factors in training cohort.

*Hybrid nomogram:* Based on Rad-Score and independent clinical risk factors, a hybrid model nomogram was established for poor outcome prediction using multivariate logistic regression (Figure S3).

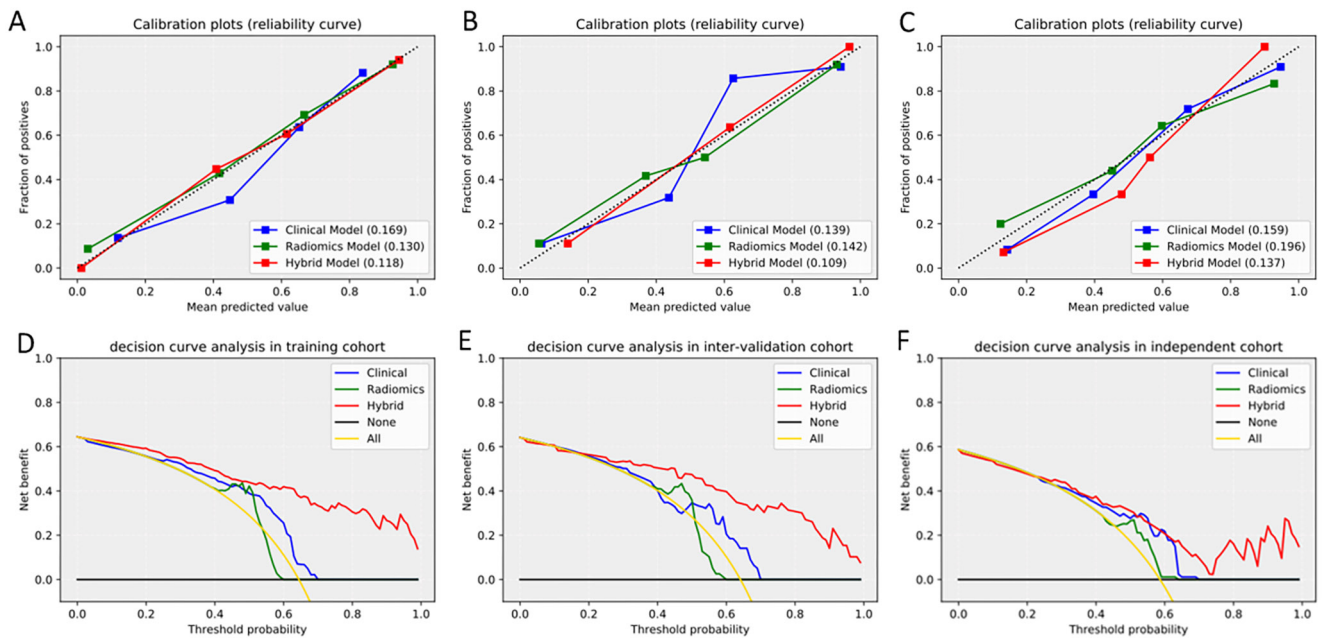
### Models calibration

*Discrimination:* The AUCs under ROCs were used to assess the predictive performances of Rad-Score based, clinical-based and hybrid models in discriminating SICH patients with poor outcome (mRS 4-6) from those with mRS 0-3.

*Calibration:* A calibration curve was plotted in the training, internal validation and the independent test cohorts for the purpose of examining the agreement between the observed outcomes and predicted probabilities. Hosmer-lemeshow test was performed to test the calibration, and decision curve analysis (DCA) to evaluate the clinical net benefit of the models (Figure S4).

### Models validation

Rad-Score model, clinical model and hybrid model nomogram constructed in training cohort (CHW cohort1) were introduced into internal validation cohort (CHW cohort2), external validation cohort (FAHNU cohort) to validate respectively.



**Figure S4** Calibration curves and decision curve analysis (DCA) of three models in the training (left), internal validation (middle) and external testing (right) cohorts. Calibration curves (A-C) depict agreement between the predicted risks of a poor outcome and actual observed poor outcomes following ICH. The solid blue, green and red lines represent good predictive ability of a poor outcome using the clinical, Rad-Score and hybrid models, respectively. DCA (D-F) for clinical, Rad-Score and hybrid model in the training (left), internal validation (middle) and external testing (right) cohorts. The yellow curve represents the assumption that all will have a poor outcome. The black line represents the assumption that none will have a poor outcome.

## References

34. Jiang Y, He Y, Zhang H. Variable Selection with Prior Information for Generalized Linear Models via the Prior LASSO Method. *J Am Stat Assoc* 2016;111:355-76.
35. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997;16:385-95.