

<https://doi.org/10.1038/s41746-025-01953-z>

# Prediction of functional outcomes in aneurysmal subarachnoid hemorrhage using pre-/postoperative noncontrast CT within 3 days of admission

Check for updates

Pengzhan Yin<sup>1,6</sup>, Jiaqi Wang<sup>2,6</sup>, Chao Zhang<sup>3</sup>, Yongxiang Tang<sup>4</sup>, Xiankuo Hu<sup>5</sup>, Hongmin Shu<sup>1</sup>, Jun Wang<sup>1</sup>, Bin Liu<sup>1</sup>, Yongqiang Yu<sup>1</sup>, Yunfeng Zhou<sup>3</sup> ✉ & Xiaohu Li<sup>1</sup> ✉

Aneurysmal subarachnoid hemorrhage (aSAH) is a life-threatening condition, and accurate prediction of functional outcomes is critical for optimizing patient management within the initial 3 days of presentation. However, existing clinical scoring systems and imaging assessments do not fully capture clinical variability in predicting outcomes. We developed a deep learning model integrating pre- and postoperative noncontrast CT (NCCT) imaging with clinical data to predict 3-month modified Rankin Scale (mRS) scores in aSAH patients. Using data from 1850 patients across four hospitals, we constructed and validated five models: preoperative, postoperative, stacking imaging, clinical, and fusion models. The fusion model significantly outperformed the others (all  $p < 0.001$ ), achieving a mean absolute error of 0.79 and an area under the curve of 0.92 in the external test. These findings demonstrate that this integrated deep learning model enables accurate prediction of 3-month outcomes and may serve as a prognostic support tool early in aSAH care.

Aneurysmal subarachnoid hemorrhage (aSAH) is a life-threatening subtype of hemorrhagic stroke, with approximately one-third of survivors experiencing substantial disability<sup>1</sup>. Functional outcomes are primarily influenced by early brain injury and in-hospital complications, including delayed cerebral ischemia (DCI) and chronic hydrocephalus (CH)<sup>2,3</sup>. Given that functional outcomes after aSAH are influenced by early brain injury and complications, accurate prediction of functional outcomes within the initial 3 days of presentation is crucial in clinical settings. It supports efficient allocation of medical resources, enables timely therapeutic interventions, and aids decisions regarding ICU monitoring, timing of rehabilitation, and prognosis discussions with patients and families.

Owing to their immediate availability, previous studies have predominantly employed clinical scoring systems—such as the World Federation of Neurological Surgeons (WFNS) scale<sup>4</sup>, the Hunt-Hess score<sup>5</sup>, and the modified Fisher scale (mFS)<sup>6</sup>—to estimate 3-month outcomes using the modified Rankin Scale (mRS)<sup>7</sup>. However, these tools are limited by their inherent subjectivity and frequently fail to capture the complex interactions

between individual patient profiles and radiological findings<sup>8</sup>. Although some investigations have explored machine learning models incorporating clinical data and preoperative CT perfusion parameters, these approaches—despite showing promise<sup>9,10</sup>—are hindered by limited accessibility, labor-intensive post-processing, and inter-institutional variability, thereby constraining their applicability in routine practice.

Recent advances in deep learning (DL) have markedly improved the analysis of radiological images, enhancing diagnostic accuracy across multiple domains. DL models are capable of autonomously extracting features from medical images, thereby minimizing human bias in feature selection and uncovering latent patterns that may not be perceptible through conventional visual interpretation, such as subtle indicators of edema, hemorrhage distribution, or ventricular changes<sup>11</sup>. Noncontrast CT (NCCT), the first-line imaging modality in both the acute and follow-up phases of aSAH management, enables reliable identification of hemorrhage severity, acute hydrocephalus, cerebral edema, and infarction preoperatively, and facilitates postoperative assessment of rebleeding,

<sup>1</sup>Department of Radiology, the First Affiliated Hospital of Anhui Medical University; Research Center of Clinical Medical Imaging; Anhui Province Clinical Image Quality Control Center, Hefei, Anhui, China. <sup>2</sup>School of Information, Wannan Medical College, Wuhu, Anhui, China. <sup>3</sup>Department of Radiology, First Affiliated Hospital of Wannan Medical College, Wuhu, Anhui, China. <sup>4</sup>Department of Radiology, Tongling People's Hospital, Tongling, Anhui, China. <sup>5</sup>Department of Radiology, Fuyang People's Hospital, Fuyang, Anhui, China. <sup>6</sup>These authors contributed equally: Pengzhan Yin, Jiaqi Wang. ✉ e-mail: [zhouyunfeng808@163.com](mailto:zhouyunfeng808@163.com); [lixiaohu@ahmu.edu.cn](mailto:lixiaohu@ahmu.edu.cn)

infarction, CH, and other complications. As such, NCCT is well-suited for outcomes prediction through DL-based approaches<sup>12</sup>. Furthermore, the utility of NCCT has been validated in other neuroimaging applications, such as stroke or traumatic brain injury prediction, reinforcing its relevance in prognostic modeling<sup>13,14</sup>.

Since clinical scores and periprocedural NCCT are essential for every aSAH patient, obtaining this data is relatively convenient. Preoperative imaging reflects the initial injury burden and is pivotal for treatment risk assessment and planning, while postoperative imaging provides insight into treatment effects and emerging complications. DL algorithms can conveniently and accurately extract this information from NCCT images. However, to date, no study has developed a multimodal fusion model integrating pre- and postoperative NCCT images with clinical data for predicting mRS score in aSAH patients. Existing literature predominantly focuses on clinical variables and either preoperative or postoperative imaging features<sup>10,15-18</sup>, with limited efforts to integrate both. Therefore, this study seeks to address this gap by leveraging multicenter datasets to integrate clinical information with both preoperative and postoperative NCCT scans, thereby developing DL models aimed at predicting functional outcomes in patients with aSAH.

## Results

### Patients

A total of 3302 patients were initially enrolled. Following exclusions based on medical record review ( $n = 1236$ ) and picture archiving and communication system evaluation ( $n = 216$ ), 1850 patients remained eligible for analysis (median age, 59 years; IQR: 51–68 years; 650 men). The cohort was distributed across the following centers: First Affiliated Hospital of Wannan Medical College (WN,  $n = 1178$ ), First Affiliated Hospital of Anhui Medical University (AH,  $n = 244$ ), Fuyang People’s Hospital (FY,  $n = 225$ ), and Tongling People’s Hospital (TL,  $n = 203$ ) (Fig. 1), and then these test cohorts were also aggregated to form the Test-combined dataset ( $n = 672$ ). There were no statistically significant differences in baseline characteristics between included and excluded patients at all centers (all  $p > 0.05$ ). A summary of patient characteristics is provided in Table 1.

### Model performance

The workflow for model construction is illustrated in Fig. 2. Model performance metrics for the preoperative, postoperative, stacking imaging,

clinical, and fusion models are presented in Table 2 and Fig. 3, the confusion matrices of all models across different test sets are presented in the Supplementary Fig. 1. (The details of model construction were described in the Methods section). The fusion model demonstrated superior performance across all validation sites (AH, FY, TL, and Test-combined), consistently achieving area under the curve (AUC) values greater than 0.90. Specificity of the fusion model was notably high at all centers, exceeding 90%, with the highest observed at TL (93.5%). Sensitivity of the fusion model varied across sites, with the highest values recorded at TL (84.1%). The mean absolute error (MAE) of the fusion model remained consistently low across all centers (range, 0.73–0.88). Comparative performance of the models on the Test-combined set is illustrated in Fig. 4. Among the fusion and two alternative fusion models, the main fusion model remained superior to the two alternative models, with significantly lower MAE and higher AUC (all  $p < 0.05$ ). The importance analysis of two alternative fusion models, together with the comparisons of predictive performance among the main fusion model and the two alternative fusion models, were detailed in Supplementary Note 1.

For the regression task, the objective was to predict the continuous mRS score ranging from 0 to 6. The fusion model achieved the lowest MAE on the Test-combined set, significantly outperforming all other models ( $p < 0.001$ ). The stacking imaging model also demonstrated significantly lower MAE values compared with both the preoperative and postoperative models ( $p < 0.001$ ).

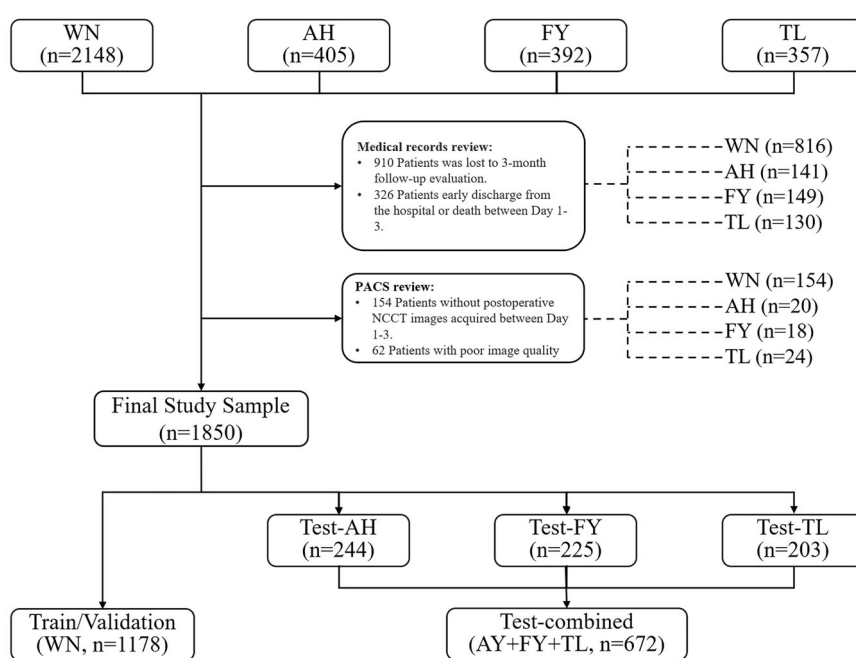
For the classification task, the objective is to predict dichotomized functional outcomes as either good ( $mRS \leq 2$ ) or poor ( $mRS > 2$ ). The fusion model attained the highest AUC on the Test-combined set, significantly outperforming all other models ( $p < 0.001$ ). The stacking imaging model also achieved a significantly higher AUC value than the preoperative and postoperative models ( $p < 0.01$ ). The performance using  $mRS > 3$  as the poor functional outcome sees in Supplementary Table 1, which yielded comparable performance to the model based on  $mRS > 2$ .

In the Test-combined set, the fusion model significantly outperformed the WFNS score, Hunt-Hess score, and mFS in terms of both regression and classification tasks. Details are provided in Supplementary Note 2.

### Subgroup analysis of DCI and CH

In the subgroup analysis, we evaluated the performance of all models across different patient populations. Details are shown in Supplementary Fig. 2.

**Fig. 1 | Flowchart of patient selection.** DCI delayed cerebral ischemia, CH chronic hydrocephalus, PACS picture archiving and communication system, NCCT noncontrast CT, WN First Affiliated Hospital of Wannan Medical College, AH First Affiliated Hospital of Anhui Medical University, FY Fuyang People’s Hospital, TL Tongling People’s Hospital.



**Table 1 | Patient characteristics, complications, and outcomes across hospital cohorts**

Variable	WN (n = 1178)	AH (n = 244)	FY (n = 225)	TL (n = 203)
Age, years	59 (51–67)	58 (49–68)	61 (52–70)	60 (52–68)
Male	404 (34.30)	91 (37.30)	82 (36.44)	73 (35.96)
GCS score	15 (13–15)	15 (11–15)	15 (12–15)	15 (12–15)
WFNS score				
1	706 (59.93)	135 (55.33)	137 (60.89)	115 (56.65)
2	159 (13.50)	22 (9.02)	24 (10.67)	26 (12.81)
3	61 (5.18)	8 (3.28)	18 (8.00)	27 (13.30)
4	170 (14.43)	61 (25.00)	35 (15.56)	18 (8.87)
5	82 (6.96)	18 (7.38)	11 (4.89)	17 (8.37)
Hunt-Hess score				
1	283 (24.02)	63 (25.82)	82 (36.44)	66 (32.51)
2	495 (42.02)	86 (35.25)	69 (30.67)	66 (32.51)
3	250 (21.22)	53 (21.72)	37 (16.44)	40 (19.70)
4	129 (10.95)	38 (15.57)	34 (15.11)	27 (13.30)
5	21 (1.78)	4 (1.64)	3 (1.33)	4 (1.97)
mFS score				
1	310 (26.32)	58 (23.77)	64 (28.44)	67 (33.00)
2	298 (25.30)	54 (22.13)	40 (17.78)	27 (13.30)
3	131 (11.12)	51 (20.90)	40 (17.78)	31 (15.27)
4	439 (37.27)	81 (33.20)	81 (36.00)	78 (38.42)
SEBES				
0	303 (25.72)	46 (18.85)	49 (21.78)	47 (23.15)
1	190 (16.13)	28 (11.48)	16 (7.11)	22 (10.84)
2	349 (29.63)	65 (26.64)	52 (23.11)	40 (19.70)
3	118 (10.02)	30 (12.30)	36 (16.00)	18 (8.87)
4	218 (18.51)	75 (30.74)	72 (32.00)	76 (37.44)
Treatment				
Clipping	51 (4.33)	18 (7.38)	11 (4.89)	8 (3.94)
Coiling	1127 (95.67)	226 (92.62)	214 (95.11)	195 (96.06)
Acute hydrocephalus	375 (31.83)	61 (25.00)	71 (31.56)	68 (33.50)
Posterior circulation	189 (16.04)	48 (19.67)	52 (23.11)	46 (22.66)
Size, mm	5 (4–6)	5 (3–6)	5 (4–5)	5 (4–7)
Hypertension	602 (51.10)	125 (51.23)	118 (52.44)	82 (40.39)
ICH	238 (20.20)	45 (18.44)	57 (25.33)	55 (27.09)
DCI	296 (25.13)	71 (29.10)	66 (29.33)	62 (30.54)
CH	199 (16.89)	55 (22.54)	50 (22.22)	47 (23.15)
mRS score				
0	638 (54.16)	121 (49.59)	98 (43.56)	91 (44.83)
1	111 (9.42)	20 (8.20)	18 (8.00)	16 (7.88)
2	102 (8.66)	21 (8.61)	26 (11.56)	15 (7.39)
3	101 (8.57)	31 (12.70)	35 (15.56)	18 (8.87)
4	89 (7.56)	21 (8.61)	18 (8.00)	25 (12.32)
5	98 (8.32)	19 (7.79)	16 (7.11)	26 (12.81)
6	39 (3.31)	11 (4.51)	14 (6.22)	12 (5.91)

Baseline clinical and imaging characteristics, complications, and outcomes from WN, AH, FY, and TL. Values are shown as median (interquartile range) for continuous variables and n (%) for categorical variables.

GCS Glasgow Coma Scale, WFNS World Federation of Neurological Surgeons, mFS Modified Fisher Scale, SEBES Subarachnoid Hemorrhage Early Brain Edema Score, ICH Intracerebral Hematoma, DCI Delayed Cerebral Ischemia, CH Chronic Hydrocephalus, mRS Modified Rankin Scale, WN First Affiliated Hospital of Wannan Medical College, AH First Affiliated Hospital of Anhui Medical University, FY Fuyang People's Hospital, TL Tongling People's Hospital.

Among DCI subgroups, both the preoperative and postoperative models demonstrated significantly superior regression performance in patients with DCI compared with those without (MAE: 1.206 vs. 1.549,  $p < 0.001$ ; 1.008 vs. 1.167,  $p = 0.021$ , respectively). In contrast, the clinical and fusion models exhibited significantly poorer regression performance in the DCI group (MAE: 1.298 vs. 0.876,  $p < 0.001$ ; 0.958 vs. 0.714,  $p < 0.001$ , respectively). The stacking imaging model showed no significant difference in regression performance between groups (MAE: 1.001 vs. 0.886,  $p = 0.134$ ). Despite these differences, the fusion model consistently achieved the lowest MAE across both DCI and non-DCI groups. With respect to classification task performance, all five models maintained comparable AUC values across DCI and non-DCI subgroups, with no statistically significant differences observed in the functional outcomes' prediction (all  $p > 0.05$ ).

Within CH subgroups, the preoperative model yielded significantly better regression performance in patients with CH relative to those without (MAE: 1.262 vs. 1.502,  $p = 0.004$ ), whereas the postoperative model demonstrated comparable regression performance across groups (MAE: 1.026 vs. 1.147,  $p = 0.108$ ). Conversely, the stacking imaging, clinical, and fusion models exhibited significantly higher MAE values in the CH subgroup (MAE: 1.074 vs. 0.875,  $p = 0.018$ ; 1.535 vs. 0.845,  $p < 0.001$ ; 1.025 vs. 0.716,  $p < 0.001$ , respectively). Nevertheless, the fusion model continued to yield the lowest MAE in both CH and non-CH cohorts. Regarding classification task performance, all models demonstrated stable AUC values across CH and non-CH subgroups (all  $p > 0.05$ ), with the fusion model exhibiting the highest overall discriminative capacity.

### Model interpretation

To enhance interpretability, Fig. 5 presents the predictive performance of the preoperative and postoperative models alongside Grad-CAM activation maps for four representative cases. The highlighted regions predominantly correspond to areas of subarachnoid hemorrhage, intraventricular blood, and low-density regions—findings that are clinically relevant to outcomes prediction in aSAH patients.

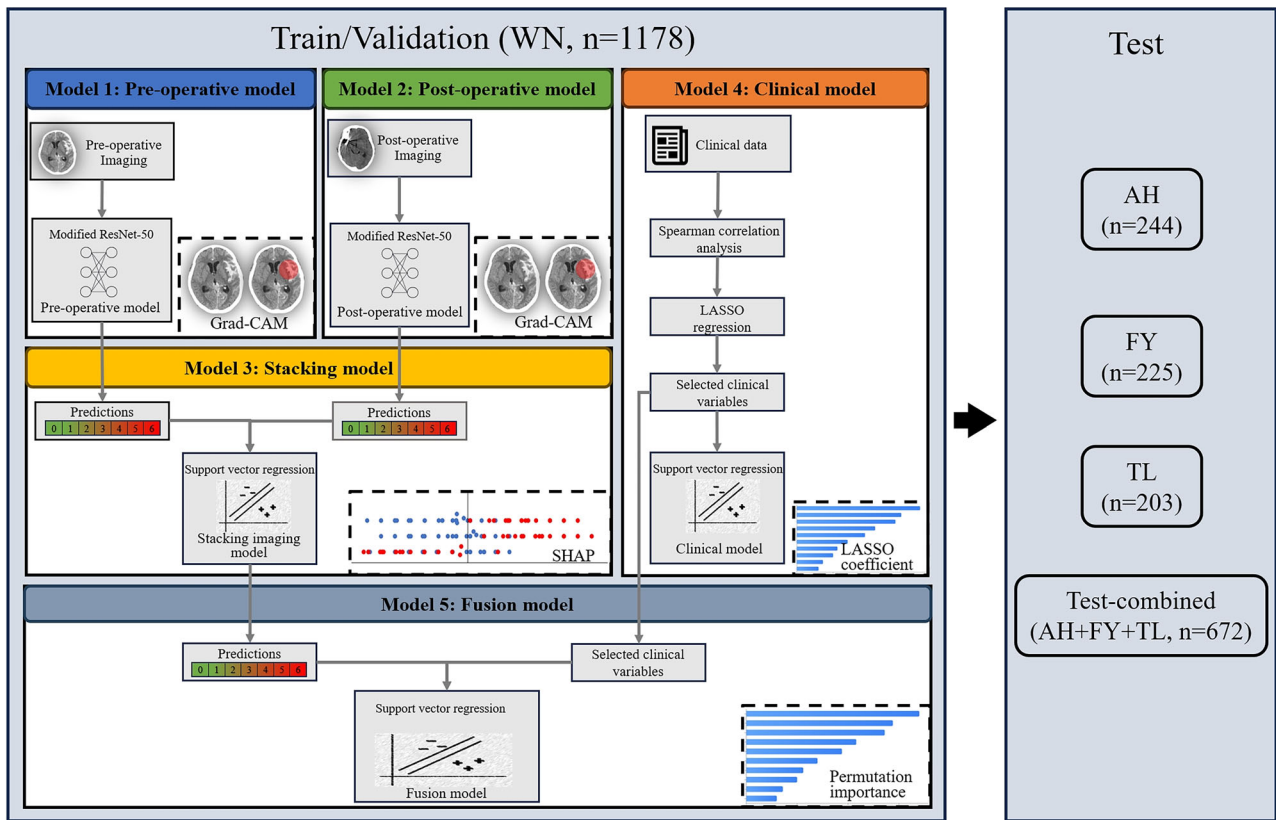
For the stacking model, SHapley Additive Explanations (SHAP) weight analysis in Fig. 6a indicated a substantially higher contribution from the postoperative model (0.879) compared with the preoperative model (0.121). Figure 6b demonstrates a complementary relationship between the preoperative and postoperative models.

In the clinical model (Fig. 6c), least absolute shrinkage and selection operator (LASSO) coefficients shows that the Hunt-Hess score emerged as the most influential variable (0.678), followed by the subarachnoid hemorrhage early brain edema score (0.379), the mFS score (0.356), and age (0.250), all of which demonstrated moderate to low importance. SHAP weight analysis shows in Supplementary Fig. 3a.

Permutation importance is shown in Fig. 6d, the fusion model included 2 imaging features—predictions from the preoperative and postoperative models (which together formed the stacking imaging model), and 4 clinical features, including Hunt-Hess score, mFS score, subarachnoid hemorrhage early brain edema score, and age. The stacking imaging model was the dominant predictor in the fusion model, with the highest importance score (0.967). Among clinical variables, the Hunt-Hess score remained the most impactful variable (0.126). SHAP weight analysis is shown in Supplementary Fig. 3b.

### Discussion

In this study, we developed and validated DL models incorporating preoperative and postoperative NCCT imaging data with clinical information to predict the 3-month functional outcomes (mRS score) in patients with aSAH. By utilizing imaging acquired at both time points, this approach captured the temporal progression of patient status, including the extent of initial brain injury (preoperative) and subsequent treatment-related changes or complications (postoperative). Furthermore, the integration of imaging data with clinical variables in a fusion model yielded superior predictive



**Fig. 2 | Workflow of model construction.** All models were trained/validated using the WN dataset ( $n = 1178$ ) and evaluated in three separate test sets and the Test-combined set. Five distinct models were constructed: (1) Pre-operative model utilizing pre-operative NCCT images with a modified ResNet-50 architecture; (2) Post-operative model employing post-operative (Day 1–3) NCCT images with the same neural network architecture; (3) Stacking model integrating 0–6 predictions from pre- and post-operative models using support vector regression; (4) Clinical variables were first filtered using Spearman correlation analysis and LASSO regression,

then the clinical model was constructed with support vector regression; (5) Fusion model combining stacking model’s predictions with selected clinical variables using support vector regression. Model interpretability was analyzed using Grad-CAM, SHAP, LASSO coefficient, and permutation importance, respectively. LASSO Least Absolute Shrinkage and Selection Operator, SHAP SHapley Additive Explanations, WN First Affiliated Hospital of Wannan Medical College, AH First Affiliated Hospital of Anhui Medical University, FY Fuyang People’s Hospital, TL Tongling People’s Hospital.

performance compared with traditional clinical models, both for specific mRS score and for the classification of poor functional outcome. Notably, within the fusion model, postoperative NCCT data contributed more substantially than preoperative data, this may be due to post-treatment findings such as early infarction, hydrocephalus, or residual hematoma, which often emerge after surgery and are strongly associated with poor outcomes. While the use of postoperative imaging may introduce the possibility of temporal bias, as it is temporally closer to the outcomes assessment, we believe this effect is minimal given the narrow time window (within 3 days) and the clinical relevance of the postoperative changes. Meanwhile, the Hunt-Hess score emerged as the most influential clinical variable. Importantly, the classification task performance of the fusion model remained robust and was not significantly affected by the presence of DCI or CH, underscoring its efficacy in predicting poor functional outcome.

For predicting poor functional outcome, the fusion model significantly outperformed the preoperative, postoperative, stacking imaging, and clinical models on the Test-combined set ( $p < 0.001$ ). When predicting continuous mRS score, the fusion model demonstrated superior performance relative to all other models in terms of MAE ( $p < 0.001$ ). These findings highlight the advantage of combining multimodal data from distinct imaging phases and clinical variables, thereby capturing complementary aspects of both initial brain injury and post-treatment alterations. This approach aligns with previous research underscoring the value of multimodal data integration for outcomes prediction<sup>19,20</sup>. While earlier studies have predominantly utilized readily available admission clinical data for functional outcomes prediction<sup>21,22</sup>, recent investigations have explored CT perfusion

imaging in this context<sup>9,23,24</sup>. In addition, Huang et al. incorporated radiomic features from CTA-based aneurysm imaging with clinical parameters to predict 3-month outcomes, with a promising result<sup>25</sup>. Although research involving MRI remains limited, Sener S et al. reported that diffusion tensor imaging parameters assessed around day 12 after injury correlated with 6-month mortality in patients with severe aSAH, suggesting a potentially valuable research direction<sup>26</sup>. However, this study leveraged NCCT imaging, which is more accessible and standardized across institutions compared with these advanced modalities, thereby enhancing the generalizability and clinical applicability of the model.

Analysis of feature importance elucidated the contribution of each data source to model performance. The postoperative imaging model demonstrated greater importance relative to the preoperative model, likely because of its capacity to capture treatment-related changes and complications that are highly predictive of outcomes. Nevertheless, the preoperative model provided essential baseline information, particularly regarding initial injury severity, which remains prognostically relevant despite the potential for subsequent recovery. In the Test-combined set, augmenting the postoperative model with preoperative data significantly reduced MAE (from 1.12 to 0.92,  $p < 0.001$ ). Further integration of clinical variables into the fusion model enhanced performance relative to the stacking imaging model. Among clinical features, the Hunt-Hess score was the most influential factor. Early neurological status is a direct indicator of injury severity, and lower levels of consciousness are strongly associated with poorer prognosis, consistent with previous studies emphasizing the predictive utility of standardized neurological assessments<sup>27–29</sup>. Although some clinical scores are

**Table 2 | Model performance across test sets**

	Regression task (mRS score)	Classification task (poor functional outcome)			
	MAE	AUC	Sensitivity (%)	Specificity (%)	PPV (%)
<b>AH</b>					
Pre-operative model	1.46 (1.35–1.58)	0.82 (0.76–0.88)	76.8 (67.5–84.9)	79.0 (72.6–85.6)	64.8 (55.7–74.1)
Post-operative model	1.17 (1.07–1.28)	0.86 (0.80–0.91)	68.3 (57.8–78.4)	86.4 (81.1–91.2)	72.0 (62.1–81.1)
Stacking imaging model	0.98 (0.86–1.10)	0.87 (0.82–0.91)	65.8 (55.1–75.6)	88.3 (82.8–93.2)	73.8 (63.3–82.7)
Clinical model	1.10 (0.95–1.25)	0.84 (0.78–0.89)	57.4 (46.9–67.1)	88.9 (83.8–93.3)	72.3 (60.0–83.6)
Fusion model	0.88 (0.76–0.99)	0.90 (0.85–0.93)	64.7 (54.2–74.6)	91.4 (87.0–95.5)	79.1 (69.2–88.7)
<b>FY</b>					
Pre-operative model	1.41 (1.30–1.53)	0.82 (0.76–0.88)	75.2 (65.2–83.8)	78.2 (71.7–84.8)	67.0 (57.1–76.5)
Post-operative model	1.07 (0.98–1.17)	0.90 (0.85–0.94)	83.2 (75.0–90.8)	79.6 (72.9–85.9)	70.2 (61.7–78.7)
Stacking imaging model	0.91 (0.79–1.02)	0.91 (0.86–0.95)	83.2 (75.0–90.5)	81.5 (75.0–87.3)	72.6 (63.2–81.2)
Clinical model	0.94 (0.81–1.08)	0.88 (0.84–0.92)	61.6 (51.8–72.4)	92.8 (88.4–96.6)	83.6 (74.2–91.9)
Fusion model	0.73 (0.63–0.84)	0.93 (0.89–0.96)	81.8 (73.1–89.5)	91.7 (86.9–95.9)	85.0 (76.9–92.3)
<b>TL</b>					
Pre-operative model	1.47 (1.35–1.58)	0.88 (0.83–0.92)	77.7 (68.4–86.7)	85.3 (79.4–91.1)	77.8 (68.3–86.2)
Post-operative model	1.12 (1.01–1.23)	0.93 (0.89–0.96)	86.3 (78.2–93.4)	87.7 (81.7–93.0)	82.2 (74.1–90.1)
Stacking imaging model	0.87 (0.76–0.99)	0.93 (0.89–0.96)	85.4 (77.5–92.2)	88.5 (82.9–93.9)	83.2 (74.7–90.7)
Clinical model	0.96 (0.81–1.12)	0.89 (0.83–0.94)	64.4 (53.9–74.7)	94.3 (89.8–97.7)	88.1 (78.9–95.5)
Fusion model	0.74 (0.63–0.85)	0.94 (0.90–0.97)	84.1 (75.6–91.5)	93.5 (88.4–97.5)	89.6 (82.2–96.1)
<b>Test-Combined</b>					
Pre-operative model	1.45 (1.38–1.52)	0.84 (0.80–0.87)	76.3 (71.2–81.7)	80.5 (76.8–84.3)	69.5 (64.2–75.3)
Post-operative model	1.12 (1.06–1.18)	0.89 (0.87–0.92)	79.2 (74.2–84.3)	84.5 (81.0–87.8)	74.7 (69.3–79.8)
Stacking imaging model	0.92 (0.86–0.99)	0.90 (0.87–0.92)	78.0 (72.8–83.2)	86.2 (82.6–89.2)	76.6 (71.1–81.7)
Clinical model	1.00 (0.92–1.09)	0.87 (0.84–0.90)	60.9 (54.8–66.9)	91.8 (89.2–94.2)	81.0 (75.0–86.3)
Fusion model	0.79 (0.72–0.85)	0.92 (0.90–0.94)	76.9 (71.5–82.0)	92.0 (89.4–94.7)	84.7 (79.9–89.1)

Model performance metrics for predicting mRS score and poor functional outcome in different test cohorts. Values shown with 95% confidence intervals in parentheses. Test-Combined includes pooled data from AH, FY, and TL cohorts. Sensitivity specificity and PPV were calculated using a threshold of 2.5 for binary classification. All performance metrics and their corresponding confidence intervals were estimated using 1000 bootstrap resamples and reported as the mean values.

MAE Mean Absolute Error, AUC Area Under the Receiver Operating Characteristic Curve, PPV Positive Predictive Value, mRS Modified Rankin Scale, WW First Affiliated Hospital of Wannan Medical College, AH First Affiliated Hospital of Anhui Medical University, FY Fuyang People’s Hospital, TL Tongling People’s Hospital.

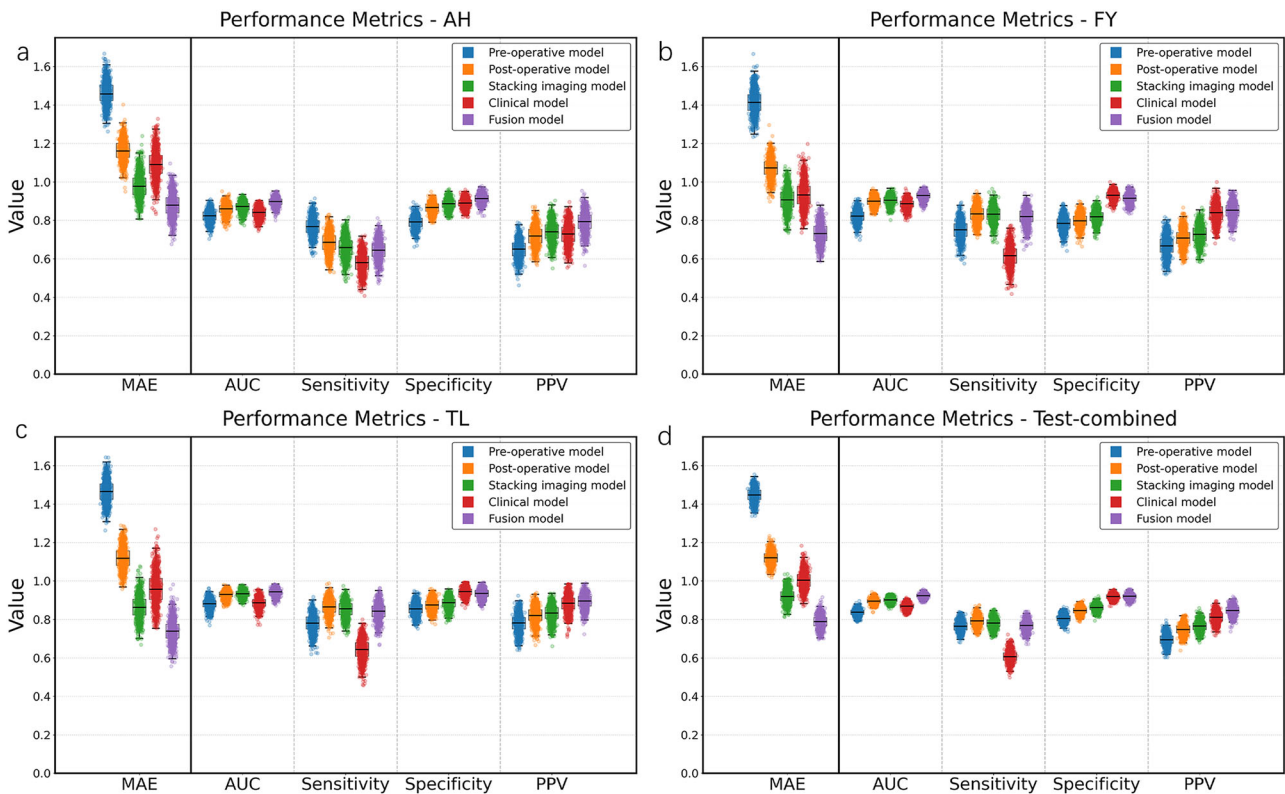
derived from NCCT image assessments and may partially overlap with the stacking imaging model, they retained high importance, suggesting they capture diagnostically salient features that may not be fully learned by the DL model. This is likely because these scores are guided by expert judgment, focusing on critical imaging findings, and incorporate clinical experience and standardized diagnostic frameworks, thereby enhancing interpretability. Grad-CAM visualizations supported this notion, revealing that the pre- and postoperative models primarily attended to regions of brain edema and subarachnoid hemorrhage—findings that correspond with clinical scoring criteria. The combination of expert-derived assessments and stacking imaging data may thus contribute to greater prediction stability and reliability. In addition, Shan et al. proposed a radiomics model based on manual segmentation of the cerebral hemorrhage area, which also provided a certain level of interpretability<sup>30</sup>. However, our DL model operates on NCCT data without manual segmentation, offering broader applicability, while also providing interpretability from a different perspective through Grad-CAM.

Prior studies have underscored the prognostic complexity introduced by complications such as DCI and CH in aSAH patients<sup>31–33</sup>. In the present study, subgroup analyses demonstrated consistent AUC values for all models among patients with and without complications, including DCI and CH. However, for MAE, the differences between the two subgroups varied greatly across different models. These findings suggested that while such complications may introduce variability in the prediction of continuous outcomes, the models retained strong discriminative capacity for predicting

functional outcomes. This discrepancy may be attributable to the limited clinical variables incorporated into the fusion model, which may be insufficient to accurately predict mRS score in patients with complications. The inclusion of additional imaging biomarkers or complication-specific molecular features could potentially improve MAE in predicting continuous mRS score.

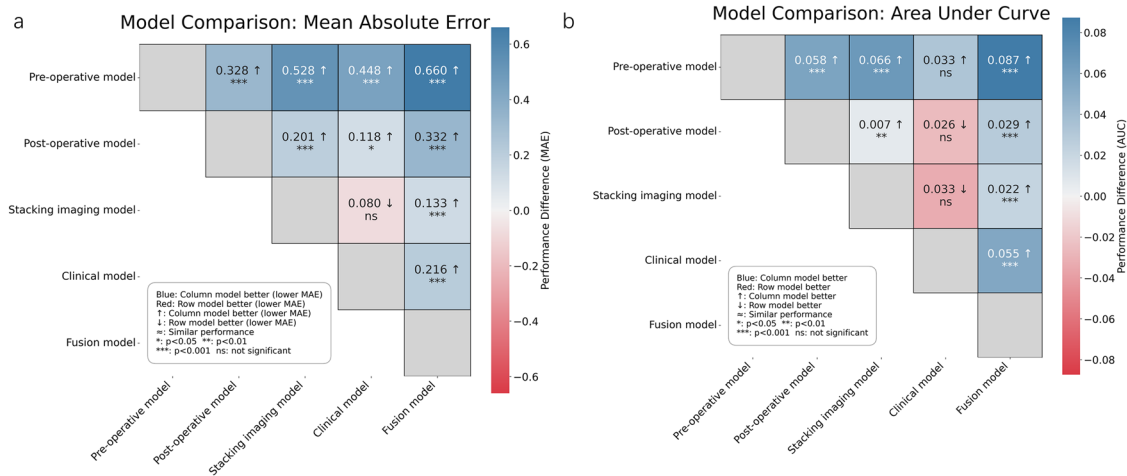
In this study, a modified ResNet-50 was selected as the backbone for image-based DL framework due to its proven effectiveness and stability in medical imaging tasks, enabling it to meet clinical expectations with lower resource consumption<sup>14,34,35</sup>. Compared to Transformer-based architectures, which typically require larger datasets and higher-dimensional inputs to achieve optimal performance, ResNet-based framework offers a more practical and robust solution for this medical dataset<sup>36–39</sup>. For multimodal integration, we employed support vector regression, which demonstrated superior predictive performance over other methods such as random forest and gradient boosting. This advantage may be attributed to the continuous and low-dimensional nature of our integrated features, which align well with our framework. Although SHAP are more naturally aligned with tree-based models, we placed greater emphasis on predictive performance. As such, support vector regression was selected as an optimal trade-off between accuracy and interpretability.

This study offers several notable strengths. First, the use of a multi-center dataset enhanced the generalizability and robustness of the models across different populations and imaging protocols. Unlike many DL studies that limit distribution shifts and thereby risk overestimating performance



**Fig. 3 | Trend of model performance in 4 test sets.** Boxes represent interquartile ranges with median lines; whiskers extend to 1.5 times the IQR. Individual points show bootstrap resampling results (1000 repetitions). **a–c** Represent model performance from AH, FY, and TL, respectively, **d** represents model performance from the overall 3 test sets. For MAE, lower values indicate better performance, while for all other metrics (AUC, Sensitivity, Specificity, and PPV), higher values indicate

better performance. Sensitivity and specificity are presented in decimal format and correspond to percentage values. IQR Interquartile Range, MAE Mean Absolute Error, AUC Area Under the Curve, PPV Positive Predictive Value, WN First Affiliated Hospital of Wannan Medical College, AH First Affiliated Hospital of Anhui Medical University, FY Fuyang People’s Hospital, TL Tongling People’s Hospital.

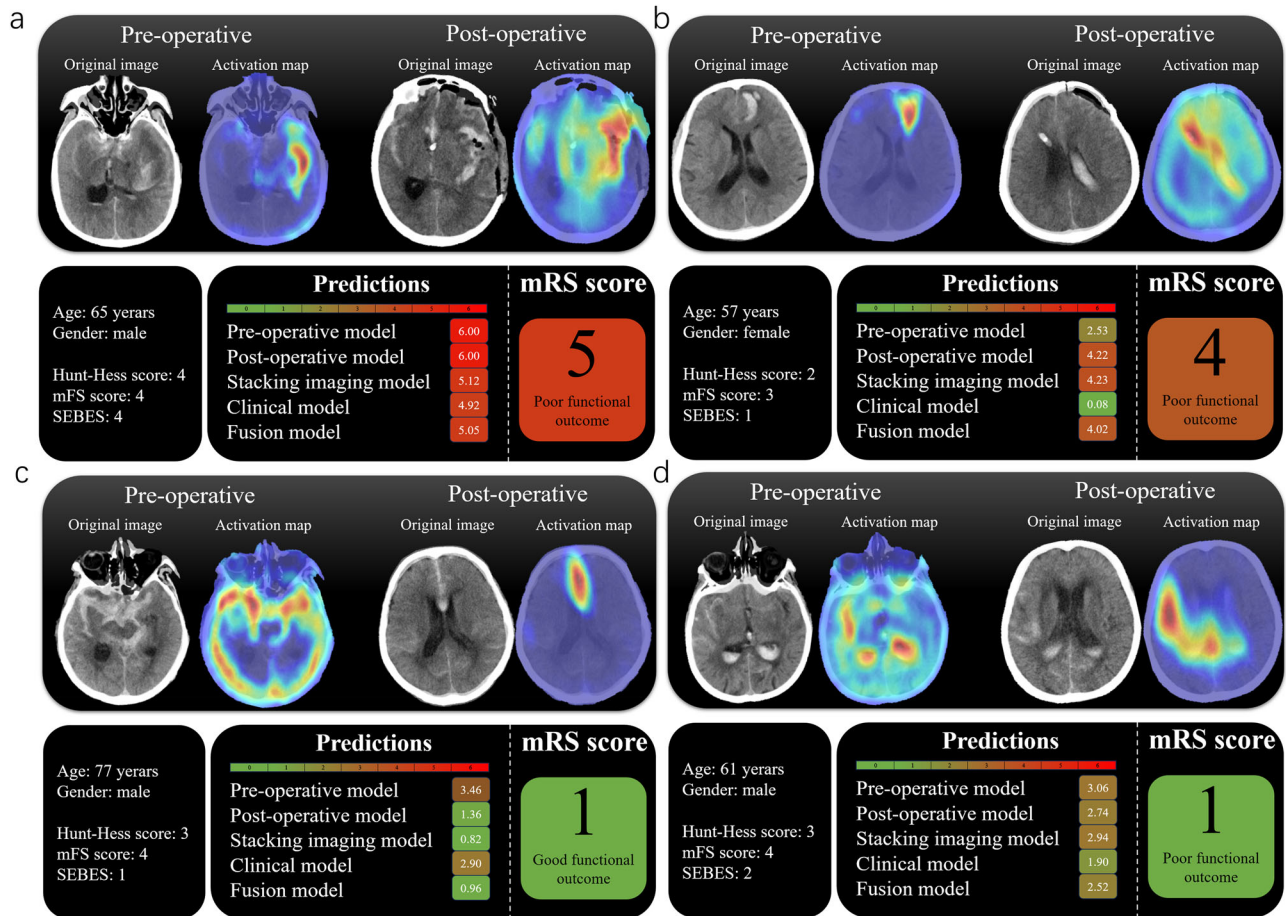


**Fig. 4 | Model performance comparison in Test-combined set.** In the heatmaps, MAE comparisons were conducted using paired t-tests (a), while the significance of AUC differences between two models was evaluated using the DeLong test (b). Each cell shows the absolute difference between the column model and row model, with arrows indicating better (†) or worse (‡) performance of the column model

compared to the row model. Statistical significance is denoted by asterisks (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , ns: not significant). Blue cells indicate the column model performs better than the row model, while red cells indicate the opposite. MAE Mean Absolute Error, AUC Area Under the Curve.

and reducing clinical relevance, our models were evaluated under realistic test conditions incorporating heterogeneous data. The consistent performance under these conditions suggests the models identify meaningful imaging biomarkers rather than dataset-specific features. Second, the

application of interpretability techniques such as Grad-CAM, SHAP, LASSO coefficients, and permutation importance strengthened the clinical utility of the models by elucidating key predictive contributors. The post-operative model exhibited greater importance for outcomes prediction,



**Fig. 5 | Examples of five models predicting different functional outcomes and their corresponding activation maps of pre- and postoperative NCCT images.** The activation maps are shown at selected representative slices to highlight typical features. The maps use color gradients, with warmer colors (red/yellow) indicating regions of higher model attention, reflecting areas critical for prediction. Background colors in the prediction table correspond to risk levels: green (low risk), yellow (moderate risk), and red (high risk). Actual mRS score are shown on the right, categorized as “Good functional outcome” (mRS  $\leq 2.5$ ) or “Poor functional outcome” (mRS  $> 2.5$ ). **a** Images of a 65-year-old male patient. The fusion model provided the most accurate prediction for the actual mRS score of 5, with activation maps highlighting the lesion area in both pre-operative and post-operative phases. **b** Images of a 57-year-old female patient. The fusion model predicted mRS score of 4.02, which closely approximated the actual mRS score of 4. Activation maps

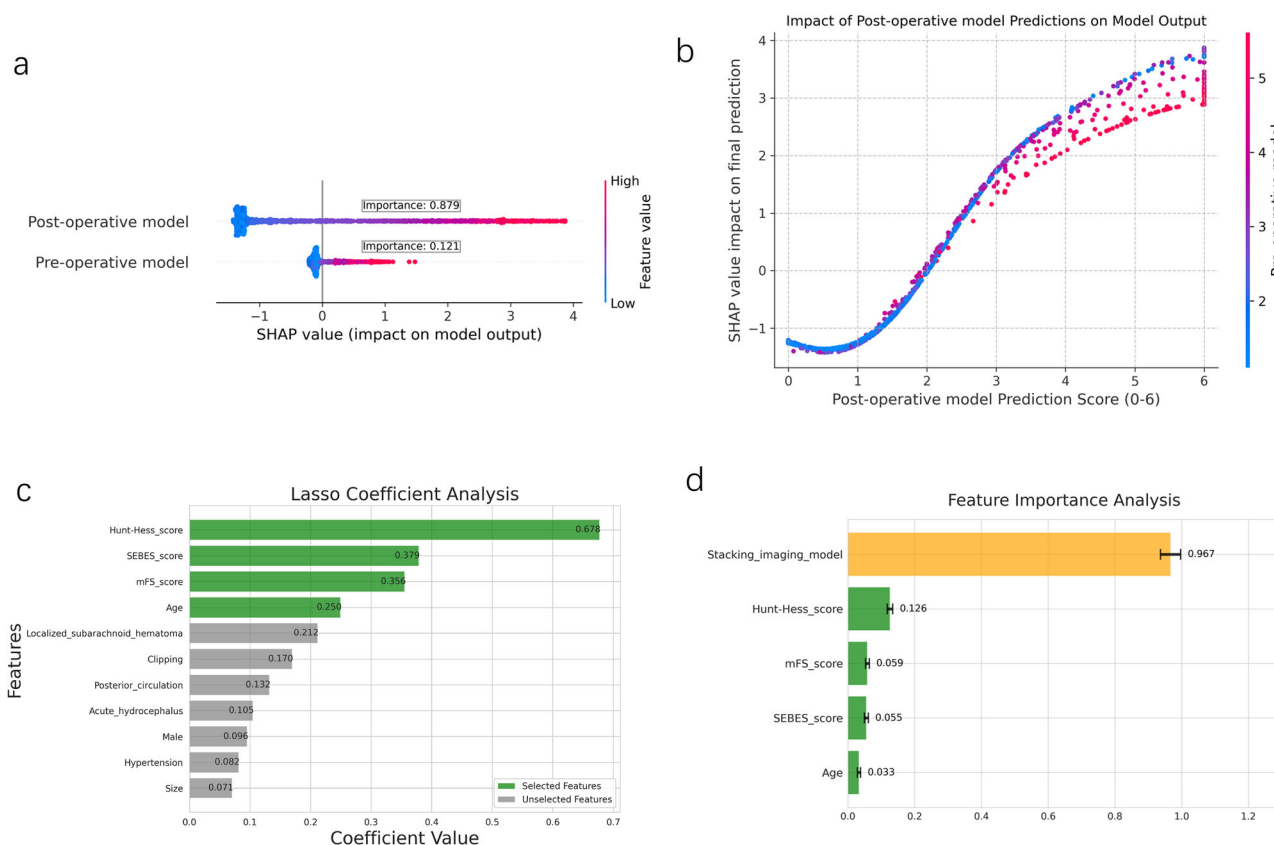
consistently highlighted relevant regions across both phases. **c** Images of a 77-year-old male patient. At admission, the pre-operative and clinical models produced relatively high-risk scores of 3.46 and 2.90, respectively. However, when post-operative information was incorporated, the stacking imaging model and fusion model made accurate predictions, with activation maps focusing on the critical areas. **d** A misclassified image of a 61-year-old male patient. All deep learning models overestimated the mRS score (above 2.5), while the clinical model provided a prediction of 1.9, relatively closer to the actual score of 1. This rare case of misprediction highlights a potential mismatch between imaging severity and clinical presentation, warranting further investigation into the underlying mechanisms. WFNS World Federation of Neurological Surgeons scale, SEBES Subarachnoid Hemorrhage Early Brain Edema Score, mFS Modified Fisher Scale, mRS Modified Rankin Scale.

aligning with clinical understanding that post-treatment imaging changes are pivotal for prognosis. Importantly, the models were designed for flexible application at various stages of clinical care. The preoperative model can be used when only admission NCCT is available, while the fusion model yields enhanced predictive performance when additional postoperative imaging or clinical data are accessible, which is particularly beneficial in settings lacking comprehensive imaging resources. Third, rather than limiting analyses to binary classifications, our approach directly predicted continuous mRS score, thereby offering a more nuanced assessment of functional outcomes.

Despite these strengths, the study is subject to several limitations. First, although the models were deliberately validated across multiple independent institutions with diverse imaging protocols and patient populations, additional validation in larger and more geographically heterogeneous cohorts is needed. Domain adaptation techniques could also be leveraged to mitigate distribution shifts and improve model robustness in varied clinical environments. Second, the retrospective design may introduce biases in follow-up assessments and treatment consistency. Moreover, CH was only

tracked for  $\leq 14$  days, and the 3-month follow-up may not be sufficient to assess long-term prognostic outcomes. While our fusion model outperformed conventional clinical scores, we acknowledge the potential tradeoff between its increased complexity and clinical applicability, especially in early decision-making settings. Future work should consider incorporating longitudinal data, such as serial imaging or neurological assessments, to better capture patients’ evolving clinical trajectories and mitigating potential temporal confounding issues. Additionally, more complex architectures such as transformers may benefit from the integration of molecular features and multi-modal fusion strategies, especially when applied to larger-scale datasets that can support their data demands.

In conclusion, this study demonstrates the feasibility and predictive utility of integrating preoperative and postoperative NCCT imaging with clinical variables for functional outcomes prediction in aSAH patients. The DL-generated fusion model represents a promising tool for individualized prognostication during the initial 3 days of presentation, with the potential to optimize patient management and resource allocation in aSAH care.



**Fig. 6 | Importance analysis of stacking, clinical and fusion models.** **a** SHAP summary plot shows the impact of pre- and post-operative model predictions on the stacking model output. Each point represents a single prediction, with color indicating the feature value (blue for low values, red for high values). The position on the x-axis indicates the SHAP value, representing how much each feature contributed to pushing the model output away from the baseline prediction. The post-operative model demonstrates substantially higher importance (0.879) compared to the pre-operative model (0.121), indicating its stronger influence on the predictions. **b** SHAP dependence plot of the post-operative model's predicted probability. This pattern suggests a complementary relationship, in which the post-operative model compensates when the pre-operative model is less confident. Moreover, the nonlinear

distribution of colors and SHAP values indicates that this complementarity is not simply additive but exhibits a nonlinear interaction between the two models. **c** LASSO coefficient analysis identified 11 clinical variables with non-zero coefficients. Features were further refined using recursive feature elimination to determine the optimal subset for modeling. Green bars indicate selected features, while gray bars represent unselected features. **d** Permutation importance ranking from support vector regression-based analysis, showing relative contributions of imaging and clinical features in fusion model. SHAP SHapley Additive exPlanations, LASSO Least Absolute Shrinkage and Selection Operator Regression, WFNS World Federation of Neurological Surgeons scale, SEBES Subarachnoid Hemorrhage Early Brain Edema Score, mFS Modified Fisher Scale, GCS Glasgow Coma Scale.

## Methods

### Ethics statement

The study was approved by the local ethics committee and institutional review board of each hospital (First Affiliated Hospital of Anhui Medical University: PJ2024-12-59; First Affiliated Hospital of Wannan Medical College:2024-185; Fuyang People's Hospital: [2025]-6; Tongling People's Hospital: 2025ky004D), and complied with the Declaration of Helsinki. The ethics committee waived the requirement for informed consent as this retrospective study analyzed only existing, fully anonymized clinical data with no additional patient interventions. All data were deidentified prior to analysis to protect participant privacy. Informed consent for the use of anonymized clinical data was obtained from patients or their legal representatives upon hospital admission. All methods were carried out following institutional guidelines and regulations.

### Patients

This multicenter retrospective study analyzed data from four hospitals, each with distinct patient demographics and imaging protocols. All datasets were independent of one another: WN ( $n = 2148$ ), AH ( $n = 405$ ), FY ( $n = 392$ ), and TL ( $n = 357$ ). The inclusion criteria comprised adult patients (age  $\geq 18$  years) who were initially suspected of having subarachnoid hemorrhage between June 2012 and September 2024 at the participating institutions. Eligible patients presented at symptom onset, underwent preoperative

NCCT imaging within 24 hours, followed by digital subtraction angiography. Only those confirmed with aSAH who subsequently received treatment were included in the study. Clinical data were extracted from electronic medical records, and NCCT images were retrieved from the picture archiving and communication system. Patients were excluded if postoperative NCCT images (acquired between Days 1 and 3) were unavailable or showed motion artifacts, or if clinical data or the 3-month mRS score were missing. To assess potential selection bias, baseline characteristics were compared between included and excluded patients across all centers, including age, sex, GCS score, WFNS score, Hunt-Hess score, mFS score, and subarachnoid early brain edema score. A flowchart of patient inclusion and exclusion is presented in Fig. 1.

### Clinical outcome and in-hospital complications

The mRS, ranging from 0 (no symptoms) to 6 (death), was obtained from medical records at the 3-month follow-up<sup>40</sup>. DCI was defined as the presence of a new infarct on hospitalization, NCCT or MRI not attributable to treatment, a new hypodense region, or unexplained neurologic deterioration accompanied by a decline in Glasgow Coma Scale (GCS) score<sup>41</sup>; DCI status was extracted from medical records. Hydrocephalus was evaluated by radiologists at each center (P.Y., 5 years post-training; X.H., 8 years post-training; C.Z., 8 years post-training; Y.T., 18 years post-training) using hospitalization NCCT. It was categorized as acute (0–3 days post-aSAH),

subacute (4–13 days post-aSAH), or CH ( $\geq 14$  days post-aSAH)<sup>42</sup>. Diagnostic criteria included marked enlargement of the temporal horns or a frontal horn-to-biparietal diameter ratio  $>30\%$ . Patients discharged within 14 days without radiological evidence of hydrocephalus on preoperative or hospitalization NCCT were classified as not having CH.

### Clinical data and NCCT imaging

Clinical data obtained within 24 hours of presentation included age, sex, history of hypertension, GCS score, WFNS score (1–5), Hunt-Hess score (1–5), mFS score (1–4), subarachnoid hemorrhage early brain edema score (0–4), presence of acute hydrocephalus, presence of localized subarachnoid hematoma, aneurysm location and size based on digital subtraction angiography, and treatment modality (coiling or clipping). Details are summarized in Table 1.

Preoperative NCCT images acquired within 24 hours of presentation and postoperative NCCT images obtained between Days 1 and 3 were collected. For patients with multiple post-treatment scans, only the earliest available image was analyzed. NCCT imaging was performed using various scanners; specific acquisition parameters are provided in Supplementary Note 3. All images were resampled to a standardized voxel size of  $320 \times 320 \times 64$  and normalized to canonical orientation, without skull stripping. An image display window level of 40 and a window width of 80 were applied. Subsequently, Z-score normalization was performed.

### Model construction

The workflow for model construction is illustrated in Fig. 2. The WN dataset ( $n = 1178$ ) was utilized for model development, while the remaining three datasets (AH, FY, TL) served as independent testing sets to assess the model's robustness in the presence of real-world heterogeneity. These were also aggregated to form the Test-combined dataset ( $n = 672$ ). Center selection and rationale for Test-combined sees in Supplementary Note 4. Five models were developed: preoperative, postoperative, stacking imaging, clinical, and fusion models.

For the preoperative and postoperative models, a modified ResNet-50 architecture incorporating three-dimensional convolutional neural networks was implemented to directly generate continuous outputs for the regression task (see Supplementary Fig. 4)<sup>14</sup>. A stratified 5-fold cross-validation procedure was applied within the WN dataset for manual hyperparameter optimization. During training, CT volumes were resampled to  $320 \times 320 \times 64$  voxels and normalized to a canonical orientation. Z-score normalization was applied across all volumes. Random data augmentation techniques—including random flipping (random horizontal or vertical flips with 50% probability), affine transformations (scale: 0.9–1.1, rotation:  $\pm 15^\circ$ , translation:  $\pm 10$  voxels), gamma correction ( $\gamma \in [-0.3, 0.3]$ ), contrast adjustments (range: 0.75–1.25), and additive Gaussian noise ( $\mu = 0$ ,  $\sigma \in [0, 0.1]$ )—were applied to enhance generalizability. Model training was performed using the AdamW optimizer with an initial learning rate of  $8 \times 10^{-4}$ , weight decay of  $1 \times 10^{-5}$ , and a batch size of 10. Training was conducted over a maximum of 80 epochs, with an 8-epoch warm-up phase. If validation performance plateaued for 12 epochs, the learning rate was reduced by a factor of 0.3. Early stopping was employed with a patience value of 10 epochs. A weighted mean squared error loss function was used<sup>43</sup>, with class weights assigned based on sample distribution, assigning a weight of 2 to a positive outcome ( $mRS > 2.5$ ). Output values were constrained within the range of 0 to 6 to align with the mRS. The final model was retrained on the entire WN dataset using the optimal hyperparameters derived from cross-validation.

For the construction of the stacking imaging model, support vector regression was employed to integrate predictions from the preoperative and postoperative models. Each base model was trained independently using stratified 5-fold cross-validation with a radial basis function kernel. Final predictions were derived by aggregating the validation outputs from each fold and restricting them to the continuous interval between 0 and 6.

For the clinical model, multicollinearity of clinical variables was addressed through stepwise variance inflation factor selection. Features with

variance inflation factor values greater than 5 were sequentially removed until all remaining features had values below the threshold. The remaining features were standardized and subsequently processed using LASSO regression ( $\alpha = 0.01$ , determined through cross-validation)<sup>44</sup>. Features with non-zero coefficients were retained and further refined using recursive feature elimination to identify the optimal subset for modeling. The final model generated a continuous prediction of mRS score within the range of 0 to 6. Further details of the feature selection procedure are described in Supplementary Note 5.

For the fusion model, predictions from the stacking imaging model were integrated with the selected clinical features to develop a support vector regression model. This model also produced a continuous mRS prediction value ranging from 0 to 6. Besides, we also evaluated two alternative fusion models, fusion-Alt1 and fusion-Alt2, which combine selected clinical features with predictions from either preoperative or postoperative models, respectively. For the stacking imaging, clinical, and fusion models, support vector regression was compared with random forest and gradient boosting regression for multimodal integration. Details are provided in the Supplementary Note 6.

### Model evaluation for regression and classification tasks

For the regression task, the objective was to predict the continuous mRS score ranging from 0 to 6. Model performance was assessed using mean absolute error (MAE).

For the classification task, which was derived from the regression outputs, the objective was to predict functional outcomes by dichotomizing the mRS score as good ( $mRS \leq 2$ ) or poor ( $mRS > 2$ ) using a threshold of 2.5. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC). Sensitivity and specificity were calculated based on the continuous predictions using the same threshold of 2.5. To assess the robustness of our approach, additional analyses were performed using  $mRS > 3$  as the poor functional outcome.

To assess the predictive performance of the fusion model relative to conventional clinical tools, comparisons were made with widely used clinical scoring systems, including the WFNS score, Hunt-Hess score, and mFS score, in the Test-combined set. Sensitivity and specificity were calculated using the optimal cut-off points determined by the Youden Index.

### Subgroup analysis of in-hospital complications

Patients were stratified into DCI versus non-DCI groups and CH versus non-CH groups based on the presence of complications. Model performance within each subgroup was evaluated by comparing MAE and AUC metrics within each complication group (DCI vs. non-DCI and CH vs. non-CH) in the Test-combined set. These metrics were derived from predictions previously generated by the five models on the Test-combined dataset.

### Model interpretation

Multiple interpretability techniques were employed to elucidate model behavior. Grad-CAM<sup>45</sup> was utilized for both preoperative and postoperative models to visualize salient regions in the NCCT images. The stacking imaging model was interpreted using SHAP<sup>46</sup> to quantify the contribution of individual model predictions. The clinical model was interpreted via LASSO coefficients, whereas permutation importance was applied to the fusion model to assess the influence of clinical features and stacking model outputs.

### Statistical analysis

Model development was based on data from WN, and evaluation was performed using an external cohort comprising AH, FY, and TL. Performance metrics included MAE, AUC, sensitivity, and specificity. Model comparative statistical analyses were conducted using paired t-tests and the DeLong test, where appropriate. Sensitivity and specificity were computed using a threshold of 2.5. MAE and AUC were designated as the primary performance metrics for regression and classification tasks, respectively. Baseline characteristics comparison was conducted using the Mann-Whitney U test or chi-square test as appropriate. A two-sided p-value of  $<0.05$  was

considered statistically significant. All statistical analyses were performed using Python version 3.12.

### Data availability

The imaging data used in this study are protected and not publicly available due to personal information protection, patient privacy regulations, and institutional data-sharing policies. However, the data can be made available to qualified researchers for academic purposes upon reasonable request. Interested parties may contact the corresponding author, who will review each request for scientific merit and compliance with data governance requirements. Approved requests will be processed within 14 working days.

### Code availability

The source code for model development, trained model weights, and full reproduction scripts have been made publicly available at <https://github.com/Seaburg97/aSAH>.

Received: 19 May 2025; Accepted: 13 August 2025;

Published online: 24 August 2025

### References

- Robba, C. et al. Contemporary management of aneurysmal subarachnoid haemorrhage. *Intensive Care Med.* **50**, 646–664 (2024).
- Veldeman, M. et al. Delayed cerebral infarction after aneurysmal subarachnoid hemorrhage. *Neurology* **103**, e209607 (2024).
- Fang, Y. et al. Cerebrospinal fluid markers of neuroinflammation and coagulation in severe cerebral edema and chronic hydrocephalus after subarachnoid hemorrhage. *J. Neuroinflamm.* **21**, 237 (2024).
- Connolly, E. S. Jr. et al. Guidelines for the management of aneurysmal subarachnoid hemorrhage. *Stroke* **43**, 1711–1737 (2012).
- Kasner, S. E. et al. Clinical interpretation and use of stroke scales. *Lancet Neurol.* **5**, 603–612 (2006).
- Sagues, E. et al. Outcomes measures in subarachnoid hemorrhage research. *Transl. Stroke Res.* **10**, 1284–1293 (2024).
- van Donkelaar, C. E. et al. Prediction of outcome after aneurysmal subarachnoid hemorrhage. *Stroke* **50**, 837–844 (2019).
- Degen, L. A. et al. Interobserver variability of grading scales for aneurysmal subarachnoid hemorrhage. *Stroke* **42**, 1546–1549 (2011).
- Yin, P. et al. Machine learning using presentation CT perfusion imaging for predicting clinical outcomes in patients with aneurysmal subarachnoid hemorrhage. *AJR Am. J. Roentgenol.* **221**, 817–835 (2023).
- Rubbert, C. et al. Prediction of outcome after aneurysmal subarachnoid haemorrhage using data from patient admission. *Eur. Radiol.* **28**, 4949–4958 (2018).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Neifert, S. N. et al. Aneurysmal subarachnoid hemorrhage: the last decade. *Transl. Stroke Res.* **12**, 428–446 (2021).
- Pease, M. et al. Outcome prediction in patients with severe traumatic brain injury using deep learning from head CT scans. *Radiology* **304**, 385–394 (2022).
- Liu, Y. et al. Prediction of ischemic stroke functional outcomes from acute-phase noncontrast CT and clinical information. *Radiology* **313**, e240137 (2024).
- Maldaner, N. et al. Development of a complication and treatment-aware prediction model for favorable functional outcome in aneurysmal subarachnoid hemorrhage based on machine learning. *Neurosurgery* **88**, E150–E157 (2021).
- Jaja, B. N. R. et al. Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: the SAHIT multinational cohort study. *BMJ* **360**, j5745 (2018).
- García-García, S. et al. Mortality prediction of patients with subarachnoid hemorrhage using a deep learning model based on an initial brain CT scan. *Brain Sci.* **14**, 10 (2023).
- Moriya, M. et al. Interpretable machine learning model for outcome prediction in patients with aneurysmal subarachnoid hemorrhage. *Crit. Care* **29**, 36 (2025).
- Zhou, Z. et al. Pre and post-operative online prediction of outcome in patients undergoing endovascular coiling after aneurysmal subarachnoid hemorrhage: visual and dynamic nomograms. *Brain Sci.* **13**, 1185 (2023).
- Huenges Wajer, I. M. et al. CT perfusion on admission and cognitive functioning 3 months after aneurysmal subarachnoid haemorrhage. *J. Neurol.* **262**, 623–628 (2015).
- Gaastra, B. et al. CRP in outcome prediction after subarachnoid hemorrhage and the role of machine learning. *Stroke* **52**, 3276–3285 (2021).
- Feghali, J. et al. External validation of a neural network model in aneurysmal subarachnoid hemorrhage: a comparison with conventional logistic regression models. *Neurosurgery* **90**, 552–561 (2022).
- Russin, J. J. et al. Permeability imaging as a predictor of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. *J. Cereb. Blood Flow. Metab.* **38**, 973–979 (2018).
- Zhang, C. et al. Early and delayed blood-brain barrier permeability predicts delayed cerebral ischemia and outcomes following aneurysmal subarachnoid hemorrhage. *Eur. Radiol.* **34**, 5287–5296 (2024).
- Huang, T. et al. Can the radiomics features of intracranial aneurysms predict the prognosis of aneurysmal subarachnoid hemorrhage?. *Front. Neurosci.* **18**, 1446784 (2024).
- Sener, S. et al. Diffusion tensor imaging: A possible biomarker in severe traumatic brain injury and aneurysmal subarachnoid hemorrhage?. *Neurosurgery* **79**, 786–793 (2016).
- Thilak, S. et al. Diagnosis and management of subarachnoid haemorrhage. *Nat. Commun.* **15**, 1850 (2024).
- Hofmann, B. B. et al. Revisiting the WFNS score: native computed tomography imaging improves identification of patients with “false poor grade” aneurysmal subarachnoid hemorrhage. *Neurosurgery* **94**, 515–523 (2024).
- Goldberg, J. et al. Survival and outcome after poor-grade aneurysmal subarachnoid hemorrhage in elderly patients. *Stroke* **49**, 2883–2889 (2018).
- Shan, D. et al. Non-contrast CT radiomics for SAH prognosis prediction. *Bioengineering* **10**, 967 (2023).
- Labib, H. et al. Sodium and its impact on outcome after aneurysmal subarachnoid hemorrhage in patients with and without delayed cerebral ischemia. *Crit. Care Med.* **52**, 752–763 (2024).
- Macdonald, R. L. Delayed neurological deterioration after subarachnoid haemorrhage. *Nat. Rev. Neurol.* **10**, 44–58 (2014).
- Adams, H. et al. Risk of shunting after aneurysmal subarachnoid hemorrhage: a collaborative study and initiation of a consortium. *Stroke* **47**, 2488–2496 (2016).
- Gibson, E. et al. Artificial intelligence with statistical confidence scores for detection of acute or subacute hemorrhage on noncontrast CT head scans. *Radio. Artif. Intell.* **4**, e210115 (2022).
- Liu, Y. et al. Functional outcome prediction in acute ischemic stroke using a fused imaging and clinical deep learning model. *Stroke* **54**, 2316–2327 (2023).
- Takahashi, S. et al. Comparison of vision transformers and convolutional neural networks in medical image analysis: a systematic review. *J. Med. Syst.* **48**, 84 (2024).
- Yadav, S. S. & Jadhav, S. M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **6**, 113 (2019).

38. Jeevan, P. P. & Sethi, A. *Which backbone to use: a resource-efficient domain specific comparison for computer vision*. *arXiv* 2406.05612v1 (2024).
  39. Jain, A. et al. A comparative study of CNN, ResNet, and vision transformers for multi-classification of chest diseases. *arXiv* 2406.00237v1 (2024).
  40. Saver, J. L. et al. Standardized nomenclature for modified Rankin Scale global disability outcomes: consensus recommendations from stroke therapy academic industry roundtable XI. *Stroke* **52**, 3054–3062 (2021).
  41. Vergouwen, M. D. et al. Definition of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage as an outcome event in clinical trials and observational studies: proposal of a multidisciplinary research group. *Stroke* **41**, 2391–2395 (2010).
  42. Kuo, L. T. & Huang, A. P. The pathogenesis of hydrocephalus following aneurysmal subarachnoid hemorrhage. *Int. J. Mol. Sci.* **22**, 5050 (2021).
  43. Zhou, P. et al. Towards understanding convergence and generalization of AdamW. *IEEE Trans. Pattern Anal. Mach. Intell.* **46**, 6486–6493 (2024).
  44. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **58**, 267–288 (1996).
  45. Selvaraju, R. R. et al. Visual explanations from deep networks via gradient-based localization: Grad-CAM. *Int. J. Comput. Vis.* **128**, 618–626 (2017).
  46. Nohara, Y. et al. N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput. Methods Prog. Biomed.* **214**, 106584 (2022).
- all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, Y.Z, X.L, P.Y, J.W.; clinical studies, C.Z, Y.T, X.H, H.S, J.W, B.L.; experimental studies, P.Y, J.W, C.Z, Y.T, X.H, H.S; statistical analysis, Y.Z, X.L, P.Y, J.W; and manuscript editing, Y.Z, X.L, P.Y, J.W, J.L.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41746-025-01953-z>.

**Correspondence** and requests for materials should be addressed to Yunfeng Zhou or Xiaohu Li.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

### Acknowledgements

This work was partly supported by the Anhui Provincial Scientific Research Program (2308085Y48; 2023AH040251; 202304295107020028) and National Science Foundation for Distinguished Young Scholars of the Higher Education Institutions of Anhui Province (2022AH020071). We would like to thank Dr. Jianying Li for reviewing the manuscript and providing valuable suggestions. We also thank Dr. Phoebe Chi from Liwen Bianji (Edanz) for editing an early draft of the manuscript.

### Author contributions

Guarantors of integrity of entire study, Y.Z, X.L, P.Y, J.W; study concepts/ study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content,