

Comparison of two- and three-dimensional computed tomography-based deep-learning models with neurosurgeon assessments for predicting functional outcome at discharge in patients with spontaneous intracerebral hemorrhage: a single-center retrospective study

Jichao Zhu

Longgang Central Hospital of Shenzhen

Yue Zhang

Longgang Central Hospital of Shenzhen

Yuqian Wang

Longgang Central Hospital of Shenzhen

Mifang Li

Longgang Central Hospital of Shenzhen

Gan Cao

Longgang Central Hospital of Shenzhen

Shanghuang Xie

Longgang Central Hospital of Shenzhen

Xiang Wan

Shenzhen Research Institute of Big Data

Changmiao Wang

Shenzhen Research Institute of Big Data

Lingyan Zhang

18819818005@163.com

Longgang Central Hospital of Shenzhen

Research Article

Keywords: Intracerebral hemorrhage, Computed tomography, Glasgow Coma Scale, Artificial intelligence, Prognosis

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11030363/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Early prediction of short-term functional outcome after spontaneous intracerebral hemorrhage (sICH) remains challenging. Non-contrast computed tomography (NCCT) provides essential imaging information, whereas the Glasgow Coma Scale (GCS) reflects early neurological severity. This study evaluated whether two-dimensional (2D) slice-based and three-dimensional (3D) volume-based deep-learning models integrating NCCT and clinical information could predict short-term functional outcome in patients with sICH and compared their performance with neurosurgeon assessments.

Methods

This single-center retrospective study included 286 patients with sICH who underwent NCCT within 24 h after hemorrhage onset. Functional outcome at discharge was dichotomized according to the Glasgow Outcome Scale. Patients were divided into training and testing cohorts. A 2D model was trained using hemorrhage-containing NCCT slices (without manual voxel-wise segmentation), with patient-level prediction obtained through majority voting. A 3D model was trained using whole-brain CT volumes combined with clinical information. Three neurosurgeons independently predicted outcomes. Model performance was evaluated using accuracy, sensitivity, specificity, and area under the curve (AUC).

Results

The 2D image-only baseline model achieved 81.0% accuracy (AUC = 0.812). Incorporating admission GCS improved the 2D model performance to 90.3% accuracy (AUC = 0.903). The 3D model achieved 90.6% accuracy (AUC = 0.895). Both deep-learning models showed higher accuracy in short-term outcome prediction than did neurosurgeon assessments. No significant difference was observed between the best-performing 2D and 3D models.

Conclusion

NCCT-based deep-learning models incorporating admission GCS may improve short-term outcome prediction after sICH. The 2D model achieved performance comparable to that of the 3D model, required less clinical information, and showed higher predictive efficiency. Prospective multicenter validation is required.

Background

Spontaneous intracerebral hemorrhage (sICH), characterized by rupture of blood vessels within the cerebral parenchyma, remarkably affects global health. The annual incidence of sICH is approximately

24.6 per 100,000 individuals worldwide, with only 12%–39% achieving long-term functional independence [1]. Furthermore, the 30-day mortality rate ranges from 35% to 52% [2]. In China, the incidence of sICH has increased to approximately 69.6 per 100,000 people annually, imposing a considerable burden on society and families [3]. Given the high incidence of disability and mortality associated with sICH, prompt diagnosis and management are important to address the situation.

In clinical practice, patient prognosis often depends on the initial volume of the hemorrhage and the location of the hematoma, which inform the treatment plan [4]. However, such plans are inherently subjective and rely on the expertise of neurosurgeons. Notable prognostic determinants of sICH include the volume of cerebral hemorrhage, the Glasgow Coma Scale (GCS) score, and the presence of intraventricular hemorrhage [5–9]. Non-contrast computed tomography (NCCT) is vital for diagnosing and managing sICH; however, computed tomography (CT) signs require interpretation by experienced clinicians and are subject to inter-observer variability, thereby reducing diagnostic consistency [10].

Machine-learning (ML) and deep-learning (DL) models have been increasingly applied to sICH research, encompassing hemorrhage detection, hematoma expansion prediction, and automated segmentation [11–14]. Recent advances have extended these applications to prognostic prediction. A MultiStep Transformer model incorporating dynamic postoperative features achieved area under the receiver operating characteristic curve values of 0.87 for 30-day survival and 0.75 for 180-day favorable functional outcome in patients undergoing minimally invasive surgery for ICH [15]. ML models combining ICH and perihematoma edema radiomics with clinical variables have demonstrated prognostic accuracy comparable to or exceeding the ICH score for 3-month outcome prediction [16]. Furthermore, latent deep features extracted from U-Net segmentation models and generative autoencoders have shown potential to complement handcrafted radiomics for hemorrhagic stroke prognostication [17]. However, existing models often lack the integration of dynamic clinical indicators like GCS with high-dimensional three-dimensional (3D) imaging features. Although two-dimensional (2D) and 3D radiomics features have shown comparable performance in predicting hematoma expansion, and 2D-CNN models have outperformed 3D-CNN models in external validation due to the higher complexity and susceptibility of the latter to overfitting with limited sample sizes [18], direct comparisons of 2D slice-based versus 3D volume-based architectures for short-term functional outcome prediction at discharge remain limited. Furthermore, cross-modal fusion of imaging and clinical data, though promising in other neurological diseases [19], has been underexplored in sICH prognostication.

Therefore, this study aimed to systematically evaluate the performance of DL models in predicting short-term functional outcomes at hospital discharge for patients with sICH. Specifically, we investigated the incremental prognostic value of integrating admission clinical domain knowledge, particularly the GCS, with imaging features from NCCT. Furthermore, this study provided a comprehensive comparison between 2D slice-based and 3D volume-based architectures against neurosurgeons of varying experience levels under a standardized patient-level evaluation framework. By developing GCS-ICHNet and ICHPro, we aimed to establish a robust and objective computational approach for acute sICH prognosis.

Methods

Study population

This retrospective study included 286 patients diagnosed with sICH who were admitted to Longgang Central Hospital of Shenzhen and completed the treatment protocol between August 2013 and May 2021. The study was approved by the Institutional Review Board of Longgang Central Hospital of Shenzhen. The requirement for written informed consent was waived because of the retrospective nature of the study and the use of anonymized data analysis.

The inclusion criteria were as follows: (1) confirmed diagnosis of sICH; (2) availability of NCCT images acquired within 24 h post-hemorrhage; (3) documented GCS scores at admission; (4) prognostic evaluation based on the Glasgow Outcome Scale (GOS) at the time of discharge; and (5) availability of clinical data. Patients with secondary cerebral hemorrhage due to arteriovenous malformations, cerebral aneurysms, traumatic brain injury, brain tumors, or cerebral infarction were excluded.

The demographic information and clinical characteristics of the patients were extracted from electronic medical records of our hospital. These included age, sex, length of hospital stay, treatment approach, hematoma volume and location, and presence of intraventricular hemorrhage. Patients were stratified into two prognostic groups based on the GOS: those with GOS scores ≥ 4 constituted the good prognosis group, whereas those with GOS scores ≤ 3 comprised the poor prognosis group.

Although no universally accepted GCS threshold is established to predict patient outcomes following cerebral hemorrhage, clinicians generally consider a GCS score ≥ 9 to be associated with more favorable prognosis. The case collection process is shown in Fig. 1.

Data preprocessing

CT was performed using Philips Brilliance 16 CT (Philips Healthcare, Eindhoven, Netherlands) and Toshiba Aquilion One 320 CT (Toshiba Medical, Tokyo, Japan) scanners. The scanning parameters were set to a slice thickness of 5–6 mm and a matrix size of 512×512. CT images were acquired in Digital Imaging and Communications in Medicine format. To construct the 2D model, images were automatically selected under radiologist supervision to ensure that the training data focused primarily on the hemorrhagic region of the patient, thereby excluding extraneous information. This ensured that each CT image layer incorporated into the study encompassed hemorrhagic areas, accounting for variations in hemorrhage area, size, and shape across patients. The data selected for this study were standardized by calculating the mean and standard deviation. Subsequently, radiologists conducted a validation check to confirm the appropriateness of the dataset.

In the 3D model, the CT image data encompassed the entire brain. In addition, the construction of the model entailed transforming basic clinical information into a textual medical report for each patient. To minimize interference from non-cerebral regions, multi-threshold segmentation and connected component filtering techniques were employed. These methods facilitate the separation of the skull

from other tissues while preserving the hemorrhagic area and intact brain tissue. The connected components were then batch-processed to remove extraneous noise and enhance the accuracy of the results. The final images retained for analysis included only brain-related information, specifically hemorrhagic regions, gray matter, and normal brain tissue. The overall study workflow is shown in Fig. 2. The pipeline comprised four main stages: (1) data collection, involving 286 patients with sICH and NCCT within 24 h and clinical data; (2) data preprocessing, including image normalization, automatic hemorrhage extraction, and construction of 2D slice-level and 3D volume-level inputs; (3) model prediction, involving 2D CNN with majority voting and 3D CNN integrating volumetric and clinical features via cross-modal attention fusion; and (4) result analysis, including performance evaluation against neurosurgeons using confusion matrices, activation maps, and metrics (area under the curve [AUC], accuracy, sensitivity, specificity).

2D and 3D models

This study introduced two deep neural network (DNN) architectures: the 2D and 3D models. These models were designed to synergistically integrate clinical parameters, including GCS scores, to enhance the detection of characteristics associated with the hemorrhagic zone in patients with sICH. By fusing clinical data with imaging inputs, our models aimed to improve the precision of feature identification, which is crucial for the assessment and management of sICH.

The 2D model utilized a ResNet-50 backbone, initialized with ImageNet pre-trained weights, to extract visual features from individual NCCT slices. All input images were resized to 224×224. Following the backbone and subsequent linear and Dropout layers, a 128-dimensional CT feature vector is extracted. Simultaneously, the GCS score of the patient is processed through a multilayer perceptron consisting of three hidden layers (32, 128, and 256 neurons), yielding a 64-dimensional feature vector. The image and GCS vectors are concatenated and fed into a transformer encoder equipped with 16 self-attention heads to capture intra-modal correlations. Subsequently, the features were forwarded to the classification head (using a default probability threshold of 0.5) to generate predictive outcomes. The architecture of the 2D model is visually detailed in Fig. 3.

The 3D model extracted volumetric features utilizing a pre-trained 3D ResNet-50 visual encoder, combined with a pre-trained BioClinicalBERT text encoder. The input CT volumes were standardized to a dimension of 96×96×96. Prior to feature extraction, demographic and clinical variables were structured into a textual medical report template including: age, sex, onset-to-CT time, hospital stay, GCS score, treatment method, hematoma position, and volume. (Note: Since this study specifically models retrospective known clinical outcome parameters, variables such as hospital stay and treatment method were retained in the template without introducing forward data leakage). Subsequently, the joint attention fusion module merged these multimodal representations. This integration leveraged a cross-modal attention fusion block, using 1×1 convolutions to map features into query, key, and value spaces with a SoftPool layer to reduce overhead, and a multi-head self-attention fusion block to capture internal dependencies across subspaces. Following the fusion process, the enriched text-enhanced visual

feature representation was introduced into the final classification head, which delivered the classification results. The network architecture of the 3D model is depicted in Fig. 3.

Both models were optimized using the Adam optimizer with a learning rate of 0.0001. The training process was conducted for 300 epochs with a batch size of 32 for the 2D model and 128 for the 3D model. The loss function utilized for classification was the cross-entropy loss. Early stopping with a patience of 20 epochs was implemented to prevent overfitting. All models were trained and evaluated on two NVIDIA HGX A100 Tensor Core GPUs.

The study dataset included 286 patients, focusing on hematoma as imaged by CT and admission GCS scores. To assess model performance, three sub-datasets were used: (1) a 2D baseline model dataset comprising solely CT image information of cerebral hemorrhage, without accompanying GCS data; (2) a 2D model dataset containing both CT image information and corresponding GCS information for cerebral hemorrhage; and (3) a 3D model dataset including whole-brain CT image information, corresponding GCS data, and additional clinical information for patients with cerebral hemorrhage. Patients were randomly divided into two groups (Fig. 1), 80% for training and 20% for testing, to guarantee the assignment of CT images from a single patient of the same set. The CT image of each patient was allocated to its respective dataset. Segmentation was performed three times for validation. The 2D model classified individual CT slices and determined patient prediction accuracy based on majority voting, where a prediction was considered correct if more than half of the images of a patient were accurately predicted.

Model evaluation and comparison to neurosurgeons

Two groups of CT images containing hemorrhage regions were randomly selected from the dataset for observer evaluation. Multiple images could originate from the same patient. Neurosurgeons evaluated selected hemorrhage-containing CT slices rather than the complete NCCT series. Therefore, the physician assessment represented a study-specific image-review setting rather than routine clinical practice. The images were anonymized, and all patient-related identifiers were stripped. The first group represented the CT-only evaluation setting, consisting of isolated hemorrhage-containing CT slices. The second group represented the CT + GCS evaluation setting, in which the same CT slices were presented together with corresponding admission GCS information. Three neurosurgeons with varying levels of expertise, a junior neurosurgeon (3 years of experience), an intermediate neurosurgeon (8 years of experience), and a senior neurosurgeon (14 years of experience), were invited to perform a standardized benchmarking task to evaluate their ability to identify prognostic markers from synchronized and constrained data inputs, including key-slice sequences and a standardized clinical briefing. These predictions were then compared with clinical outcomes and the results from the 2D baseline model and 3D model. Because multiple CT slices could originate from the same patient, image-level observations were not fully statistically independent. Sensitivity, specificity, accuracy (ACC), AUC, positive predictive value (PPV), and negative predictive value (NPV) were calculated to evaluate the predictive performance of the model.

The 3D model dataset consisted of 223 patients from the 2D model dataset and was completed using whole-brain CT images and clinical information. Of the initial 286 patients, the 3D model utilized a subset of 223 cases that possessed both complete whole-brain NCCT volumes and comprehensive clinical baseline data. This reduction in sample size was primarily attributed to the lack of high-quality, continuous whole-brain scans or missing specific clinical features required for the 3D joint-attention fusion architecture in some earlier cases. A 3D model was employed to predict prognosis using only whole-brain CT images and a combination of whole-brain CT images, GCS data, and other clinical information. The results from the 3D model were compared with the clinical prognosis and predictions from the 2D model and neurosurgeons. Sensitivity, specificity, ACC, AUC, PPV, and NPV were measured to assess the predictive performance of the model.

Statistical analysis

Statistical analyses were conducted using SPSS version 23.0. The Kolmogorov–Smirnov test was employed to assess the normality of distribution of continuous variables associated with clinical features, which were described as means $\pm$ standard deviations. However, those that did not meet the criteria of normality were characterized by medians and interquartile ranges.

An independent sample t-test was used to analyze normally distributed continuous data. Conversely, the Mann–Whitney U test was used to evaluate continuous data with skewed distributions. Categorical data were presented as counts and percentages (n [%]), and comparisons between groups were performed using the chi-square test. A receiver operating characteristic (ROC) curve was generated to ascertain the predictive performance of the proposed model. To account for potential intra-patient correlation arising from multiple slice samples per individual, a generalized estimating equation approach was implemented to ensure the statistical independence of p -value calculations in slice-level comparisons.

Results

Clinical characteristics

Significant differences were observed in hospitalization duration, treatment regimens, pneumonia, and admission systolic blood pressure between the two groups (all $p < 0.001$). In contrast, no significant differences were observed in age, sex, smoking, diabetes mellitus, hyperlipidemia, hypertension, or admission diastolic blood pressure between the two groups (all $p > 0.05$). Analysis of the bleeding sites revealed statistically significant differences in the incidence of hemorrhage within the basal ganglia, cerebellum, and secondary intraventricular regions ($p = 0.001$, $p = 0.003$, and $p = 0.026$, respectively) between the two groups. However, no such differences were noted in other hemorrhagic locations. Furthermore, the groups demonstrated significant differences in hematoma volume and GCS scores ($p < 0.001$). Detailed comparative results are presented in Table 1.

Table 1

Characteristics of patients with spontaneous intracerebral hemorrhage with available CT scans, stratified by study cohort

Age, years, median (IQR)	GOS $\geq$ 4 (n = 144)	GOS $\leq$ 3 (n = 142)	p-value
	53 (47,61)	53 (44,61)	0.757
Sex, male, n (%)	106 (73.61%)	108 (76.06%)	0.684
Length of stay, days, median (IQR)	17 (12.25, 23)	27.5 (18.75, 45.25)	< 0.001
Treatment approach, surgical treatment, n (%)	25 (17.36%)	75 (52.82%)	< 0.001
Smoking, n (%)	23 (15.97%)	25 (17.61%)	0.753
Pneumonia, n (%)	42 (29.17%)	93 (65.49%)	< 0.001
Diabetes mellitus, n (%)	10 (6.94%)	19 (13.38%)	0.080
Hyperlipidemia, n (%)	24 (16.67%)	19 (13.38%)	0.509
Hypertension, n (%)	135 (93.75%)	133 (93.66%)	1.000
Admission systolic blood pressure, mmHg (mean $\pm$ SD)	174.26 $\pm$ 24.18	183.43 $\pm$ 27.28	0.003
Admission diastolic blood pressure, mmHg (mean $\pm$ SD)	104.83 $\pm$ 16.79	107.92 $\pm$ 17.37	0.127
GCS, median (IQR)	15 (14,15)	11 (8,14)	< 0.001
Hematoma volume, mL (IQR)	9.87 (2.35, 12.43)	27.85 (8.60, 35.85)	< 0.001
Location of hematoma, n (%)			
Basal ganglia	48 (33.33%)	76 (53.52%)	0.001
Thalamus	29 (20.14%)	18 (12.68%)	0.110
External capsule	5 (3.47%)	6 (4.23%)	0.769
Brain lobe	20 (13.89%)	14 (9.86%)	0.362
Brain stem	16 (11.11%)	13 (9.15%)	0.696
Cerebellum	14 (9.72%)	2 (1.41%)	0.003
Primary intraventricular hemorrhage	6 (4.17%)	1 (0.70%)	0.120
GOS, Glasgow Outcome Scale; IQR, interquartile range; GCS, Glasgow Coma Scale; SD, standard deviation			

Age, years, median (IQR)	GOS $\geq$ 4 (n = 144)	GOS $\leq$ 3 (n = 142)	p-value
	53 (47,61)	53 (44,61)	0.757
Secondary intraventricular hemorrhage	32 (22.22%)	49 (34.51%)	
Multiple locations	5 (3.47%)	10 (7.04%)	0.196
GOS, Glasgow Outcome Scale; IQR, interquartile range; GCS, Glasgow Coma Scale; SD, standard deviation			

Comparative analysis of prediction performance

When relying solely on key-slice sequences containing hematoma to predict outcomes, a significant disparity in prediction accuracy was noted between the three clinicians and the 2D baseline model ($p < 0.001$). The 2D baseline model exhibited higher accuracy (ACC = 81.0%), and the ROC curve analysis indicated superior predictive performance (AUC = 0.812). Detailed predictive performance metrics are listed in Table 2, and the corresponding confusion matrix is shown in Fig. 4.

Table 2
Performance comparisons using key-slice sequences for prognostic prediction in intracerebral hemorrhage

Test	AUC (95% CI)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Doctor 1 (CT)	0.609 (0.555, 0.664)	57.8 (231/400)	94.5 (171/181)	27.4 (60/219)	51.8 (171/330)	85.7 (60/70)
Doctor 2 (CT)	0.615 (0.561, 0.669)	58.0 (232/400)	97.8 (177/181)	25.1 (55/219)	51.9 (177/341)	93.2 (55/59)
Doctor 3 (CT)	0.737 (0.687, 0.786)	73.3 (293/400)	77.9 (141/181)	69.4 (152/219)	67.8 (141/208)	79.2 (152/192)
2D baseline model (CT)	0.812 (0.767, 0.856)	81.0 (324/400)	82.9 (150/181)	79.5 (174/219)	76.9 (150/195)	84.9 (174/205)
ResNet50 (2D)	0.679 (0.595, 0.762)	63.8 (255/400)	65.2 (118/181)	62.6 (137/219)	59.0 (118/200)	68.5 (137/200)
“Doctor 1,” “Doctor 2,” and “Doctor 3” are professional neurosurgeons with 3, 8, and 14 years of working experience, respectively.						
2D, two-dimensional; AUC, area under the curve; CT, computed tomography; NPV, negative predictive value; PPV, positive predictive value						

When prognostication was based on key-slice sequences containing hematoma supplemented with GCS data, the 2D model demonstrated significantly greater accuracy than did the other three models ($p < 0.001$), attaining an ACC of 90.3%. The ROC curve corroborated the enhanced predictive capability of the 2D model (AUC = 0.903).

We used the publicly available convolutional neural network 2D model, ResNet50, to analyze our data and found an ACC of 63.8% and an AUC of 0.679, which were lower than those of our 2D model [20]. Specific predictive performance indicators are compiled in Table 3, and the confusion matrix is shown in Fig. 5.

Table 3
Performance comparisons using CT images plus GCS information for prognostic prediction in intracerebral hemorrhage

Test	AUC (95% CI)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Doctor 1 (CT + GCS)	0.573 (0.518, 0.629)	54.8 (219/400)	91.9 (170/185)	22.8 (49/215)	50.6 (170/336)	71.9 (49/64)
Doctor 2 (CT + GCS)	0.693 (0.642, 0.745)	67.3 (269/400)	96.8 (179/185)	41.9 (90/215)	58.9 (179/304)	93.8 (90/96)
Doctor 3 (CT + GCS)	0.779 (0.732, 0.826)	77.5 (310/400)	82.7 (153/185)	73.0 (157/215)	72.5 (153/211)	83.1 (157/189)
2D model (CT + GCS)	0.903 (0.869, 0.936)	90.3 (361/400)	90.8 (168/185)	89.8 (193/215)	88.4 (168/190)	91.9 (193/210)
“Doctor 1,” “Doctor 2,” and “Doctor 3” are professional neurosurgeons with 3, 8, and 14 years of working experience, respectively.						
2D, two-dimensional; AUC, area under the curve; CT, computed tomography; NPV, negative predictive value; PPV, positive predictive value; GCS, Glasgow Coma Scale						

Prognostic predictions were evaluated across several different input scenarios. The three physicians and the 2D model utilized CT images combined with GCS data, whereas the 3D model was assessed using three distinct inputs: CT images alone, CT images combined with GCS data, and CT images combined with clinical baseline information. Across all these scenarios, a noteworthy difference in predictive accuracy was observed, both the 2D and 3D models demonstrated higher predictive performance than did physician assessment under the study-specific image-review setting (McNemar test, $p < 0.001$).

The 3D model achieved slightly higher accuracy (ACC = 90.6%), whereas the 2D model achieved a slightly higher AUC (AUC = 0.902); however, the difference was not statistically significant. Moreover, no significant difference was observed in the prediction accuracy between the 2D model, which used CT images and GCS data, and the 3D model, which incorporated CT images with GCS scores and additional clinical information. We used the publicly available convolutional neural network 3D DL-based model to analyze our data and found an ACC of 80.7% and an AUC of 0.812, both of which were lower than those of our 3D model [21]. Furthermore, the inter-observer agreement among the three neurosurgeons was evaluated using Fleiss’ kappa, yielding a value of 0.72, indicating substantial consistency. The precise predictive performance metrics are listed in Table 4, and the confusion matrix is shown in Fig. 6.

Table 4

Performance comparisons for using CT images plus baseline information for prognostic prediction in intracerebral hemorrhage

Test	AUC (95% CI)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Doctor 1 (CT + GCS)	0.552 (0.477,0.628)	53.8 (120/223)	90.7 (97/107)	19.8 (23/116)	51.1 (97/190)	69.7 (23/33)
Doctor 2 (CT + GCS)	0.645 (0.573,0.717)	63.2 (141/223)	96.3 (103/107)	32.8 (38/116)	56.9 (103/181)	90.5 (38/42)
Doctor 3 (CT + GCS)	0.729 (0.662,0.797)	72.7 (162/223)	80.4 (86/107)	65.5 (76/116)	68.3 (86/126)	78.4 (76/97)
2D model (CT + GCS)	0.902 (0.856,0.947)	90.1 (201/223)	90.7 (97/107)	89.7 (104/116)	89.0 (97/109)	91.2 (104/114)
3D model (CT)	0.797 (0.736,0.859)	79.8 (178/223)	77.6 (83/107)	81.9 (95/116)	79.8 (83/104)	79.8 (95/119)
3D model (only baseline characteristics)	0.643 (0.618,0.668)	63.2 (141/223)	64.5 (69/107)	62.1 (72/116)	61.1 (69/113)	65.5 (72/110)
3D model (CT + GCS)	0.881 (0.858,0.904)	89.7 (200/223)	88.8 (95/107)	90.5 (105/116)	89.6 (95/106)	89.7 (105/117)
3D model (CT + baseline characteristics)	0.895 (0.878,0.917)	90.6 (202/223)	91.6 (98/107)	89.7 (104/116)	89.1 (98/110)	92.0 (104/113)
3D deep learning-based model (CT + baseline characteristics)	0.812 (0.765,0.859)	80.7 (180/223)	86.0 (92/107)	75.9 (88/116)	76.7 (92/120)	85.4 (88/103)
2D, two-dimensional; 3D, three-dimensional; AUC, area under the curve; CT, computed tomography; NPV, negative predictive value; PPV, positive predictive value; GCS, Glasgow Coma Scale						

Discussion

In this study, larger hematoma volumes, infratentorial locations, and lower admission GCS scores were found to be associated with poorer short-term outcomes in patients with sICH. These clinical indicators,

although informative individually, are often integrated subjectively in clinical practice, highlighting the need for objective prognostic tools that combine imaging biomarkers with dynamic clinical parameters.

The integration of deep neural networks with medical imaging has shown significant progress in extracting high-dimensional prognostic information from CT that is challenging to discern with the naked eye. The GCS, which reflects the severity of neurological damage upon admission, is an important index for predicting short-term cerebral hemorrhage outcomes. Prior studies have demonstrated that additional clinical information, such as hematoma location and volume, enhances the accuracy of prognostic predictions [22, 23]. Admission GCS has been consistently validated as a prognostic factor for ICH across diverse populations [24]. However, compared with composite scores such as the ICH score or FUNC score, GCS alone may have limited discriminative ability, highlighting the need for integrated models that combine GCS with imaging biomarkers [25].

In this context, we proposed two novel DL models for short-term functional outcome prediction. The 2D model integrated GCS information with CT image features of cerebral hemorrhage through a self-attention mechanism, whereas the 3D model employed a cross-modal network based on joint attention fusion, mimicking the multimodal reasoning process used in clinical prognostic assessment. A key advantage of these models is that they do not require manual voxel-wise hematoma segmentation, enabling direct input of CT images and clinical information for automated prognosis prediction. This approach is more efficient than methods reported in previous studies [26–28].

Our findings align with prior research establishing ML -based prognostic models for cerebral hemorrhage. Wang et al. [29] reported accuracies of 83.1% for 1-month and 83.9% for 6-month prognoses using NCCT images and clinical information. Nawabi et al. [30] developed a model with 77% accuracy at discharge. In our analysis, the publicly available 2D ResNet50 model achieved an ACC of 63.8% and an AUC of 0.679, whereas the 3D DL-based model achieved an ACC of 80.7% and an AUC of 0.812, both lower than those of our proposed models [20, 21]. These results indicate that self-attention and joint-attention fusion techniques can capture valuable prognostic information through learning and that DNN technology can extract meaningful relationships between clinical data and CT images, outperforming traditional architectures in capturing complex patterns.

Visual analysis of the performance of the 2D model indicated higher accuracy than did the baseline model in identifying sICH regions within CT images, highlighting the critical role of GCS information in guiding model predictions (Fig. 7). The 2D baseline model exceeded the accuracy of physicians under the constrained review setting, revealing that AI models can identify high-dimensional prognostic information imperceptible to humans. Notably, the addition of GCS as auxiliary information improved the accuracy of the 2D model by 9.3% over the baseline, whereas clinician assessments showed inconsistent improvement with GCS data. These findings highlight the challenges of prognostic assessment under information-restricted conditions and emphasize the importance of integrating GCS into prediction models.

The 3D model achieved slightly higher accuracy (ACC = 90.6%), whereas the 2D model achieved a slightly higher AUC (AUC = 0.902); however, the difference was not statistically significant ($p > 0.05$). The prediction accuracy of the 2D model using CT images combined with GCS data was comparable to that of the 3D model incorporating additional clinical information. These findings affirm the significance of the GCS in predicting the short-term prognosis of sICH [24]. The prognostic value of the GCS has been highlighted in recent studies, although its specific contribution within DL architectures requires further clarification. The comparable performance of the 2D model, which requires fewer parameters and less computational resources, indicates that simpler architectures may be preferable for sICH prognostic prediction when sample sizes are limited. This observation aligns with recent evidence that 3D-CNN models exhibit higher complexity and are more susceptible to overfitting with modest datasets owing to their larger parameter counts and lack of pretrained weights [31].

Several methodological limitations regarding the comparison between artificial intelligence (AI) models and neurosurgeons should be acknowledged. First, neurosurgeons evaluated selected hemorrhage-containing CT slices rather than the complete NCCT series routinely reviewed in clinical practice. Therefore, physician assessment in this study represented a constrained experimental image-review setting rather than real-world clinical decision-making. Second, the AI models generated patient-level predictions using multiple CT slices and majority voting, whereas physicians assessed isolated images, resulting in unequal information input between human readers and AI systems. Third, because multiple CT slices may have originated from the same patient, image-level analyses may have introduced statistical dependency among observations. Future studies should adopt patient-level evaluation frameworks in which physicians review complete CT series together with comprehensive clinical information and provide probability-based predictions for more rigorous comparison with AI systems.

This study has several additional limitations. Its single-center retrospective nature and modest sample size ($n = 286$ for 2D, $n = 223$ for 3D) necessitate larger prospective multicenter studies to validate generalizability. We only recorded functional outcomes at discharge without long-term follow-up; consequently, the predictive capabilities of the model for sustained recovery (e.g., 90-day modified Rankin scale) require further exploration. The GOS scores at discharge should be interpreted as short-term functional outcomes inherently influenced by confounding factors such as length of hospital stay, treatment strategies, and rehabilitation availability. Future studies should incorporate standardized long-term assessments and explore automated bleeding area delineation using weakly supervised learning to enhance localization accuracy.

Notably, although these AI models demonstrate promising prognostic performance, their ultimate goal is to aid and support clinical decision-making, not to replace the judgment of human neurosurgeons. Such models must serve as collaborative tools to assist professionals in complex clinical environments, with transparent interpretation of model outputs and equitable validation across diverse populations before clinical deployment.

Conclusions

In conclusion, NCCT-based DL models incorporating admission GCS showed promising predictive performance for short-term functional outcome at discharge in sICH. The 2D model achieved comparable accuracy to the 3D model while requiring less clinical information and computational resources, indicating that simpler architectures may be preferable for modest-sized datasets. These models are intended to augment, not replace, clinical decision-making. Prospective multicenter validation is required before clinical implementation.

Abbreviations

spontaneous intracerebral hemorrhage (sICH)

Non-contrast computed tomography (NCCT)

Glasgow Coma Scale (GCS)

two-dimensional (2D)

three-dimensional (3D)

area under the curve (AUC)

non-contrast computed tomography (NCCT)

computed tomography (CT)

machine-learning (ML)

deep-learning (DL)

Glasgow Outcome Scale (GOS)

deep neural network (DNN)

multilayer perceptron (MLP)

positive predictive value (PPV)

negative predictive value (NPV)

receiver operating characteristic (ROC)

artificial intelligence (AI)

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of Longgang Central Hospital of Shenzhen. The requirement for written informed consent was waived because of the retrospective nature of the study and the use of anonymized data analysis.

Consent for publication

Not applicable.

Availability of data and materials

Data will be made available on request.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Authors' contributions

Jichao Zhu: Methodology, Writing–original draft. Yue Zhang: Data collection, Writing–original draft. Yuqian Wang: Data collection, Writing–original draft. Mifang Li: Data collection. Gan Cao: Data collection,. Shanghuang Xie: Data collection. Xiang Wan: Reviewing editing. Changmiao Wang: Conceptualization, Reviewing and editing, Supervision. Lingyan Zhang: Writing, Conceptualization, Reviewing and editing, Supervision.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of this paper. We also appreciate the support from our institution.

References

1. Van Asch CJ, Luitse MJ, Rinkel GJ, van der Tweel I, Algra A, Klijn CJ. Incidence, case fatality, and functional outcome of intracerebral hemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. *Lancet Neurol*. 2010;9:167–76.
2. Broderick J, Connolly S, Feldmann E, Hanley H, Kase C, Krieger D, et al. Guidelines for the management of spontaneous intracerebral hemorrhage in adults: 2007 update: a guideline from the American Heart Association/American Stroke Association stroke council, high blood pressure research council, and the quality of care and outcomes in research interdisciplinary working group. *Stroke*. 2007;38:2001–23.

3. 2016 GBD, Stroke Collaborators, Nguyen M, Roth GA. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurol.* 2019;18:439 – 58.
4. Morotti A, Li Q, Nawabi J, Mazzacane F, Schlunk F, Shoamanesh A, et al. Volume tolerance and prognostic impact of hematoma expansion in deep and lobar intracerebral hemorrhage. *Stroke.* 2025;56:1224–31.
5. Ho RW, Cai JF, Hu YJ, So CH, Leung KL, Wu J, et al. Location-specific analysis of hematoma expansion in intracerebral hemorrhage. *J Am Heart Assoc.* 2026;15:e042399.
6. Gunda B, Böjti P, Takács T, Zhubi E, Bereczki D, Varga A, et al. Spontaneous intracerebral hemorrhage during computed tomography scanning-assessment of hyperacute hematoma growth. *GeroScience.* 2026;48:871–7.
7. Yu Q, Fan X, Li J, Hao Q, Ning Y, Long S, et al. An end-to-end deep learning pipeline for hematoma expansion prediction in spontaneous intracerebral hemorrhage based on non-contrast computed tomography. *NPJ Digit Med.* 2025;9:39.
8. Pensato U, Dowlathshahi D, Rodriguez-Luna D, Ospel JM, Morotti A, Tanaka K, et al. Spot sign in intracerebral hemorrhage: critical reappraisal and future clinical implications. *Stroke.* 2025;56:1612–24.
9. Emblem KE, Pinho MC, Zöllner FG, Due-Tonnessen P, Hald JK, Schad LR, et al. A generic support vector machine model for preoperative glioma survival associations. *Radiology.* 2015;275:228–34.
10. Wei L, Wu B, Guo T, Ru D, Gao C, Wu J, et al. Development and validation of a machine learning-based model for 90-day prognosis outcome in spontaneous intracerebral hemorrhage patients based on non-contrast computed tomography: a multicenter retrospective observational study. *EClinicalMedicine.* 2025;88:103507.
11. Abdel-Salam M, Houssein EH, Emam MM, Samee NA, Gharehchopogh FS, Bacanin N. EATHOA: elite-evolved hiking algorithm for global optimization and precise multi-thresholding image segmentation in intracerebral hemorrhage images. *Comput Biol Med.* 2025;196:110835.
12. Wang K, Liu Q, Mo S, Zheng K, Li X, Li J, et al. A decision tree model to help treatment decision-making for severe spontaneous intracerebral hemorrhage. *Int J Surg.* 2024;110:788–98.
13. Ma D, Li C, Du T, Qiao L, Tang D, Ma Z, et al. PHE-SICH-CT-IDS: a benchmark CT image dataset for evaluation semantic segmentation, object detection and radiomic feature extraction of perihematomal edema in spontaneous intracerebral hemorrhage. *Comput Biol Med.* 2024;173:108342.
14. Chavoshi M, Mansuri A, Bala W, Brown-Mulry B, Dapamede T, Isaac R, et al. Real-world performance evaluation of a commercial deep learning model for intracranial hemorrhage detection. *NPJ Digit Med.* 2025;9:66.
15. Wang J, Shi J, Ye Q, Chen F, Sun Y, Pan C, et al. Deep learning for dynamic prognostic prediction in minimally invasive surgery for intracerebral hemorrhage: model development and validation study. *JMIR Med Inf.* 2026;14:e86327.

16. Dierksen F, Sommer JK, Tran AT, Lin H, Haider SP, Maier IL, et al. Machine learning models for 3-month outcome prediction using radiomics of intracerebral hemorrhage and perihematomal edema from admission head computed tomography (CT). *Diagnostics*. 2024;14:2827.
17. Tran AT, Wen J, Abou Karam G, Zeevi D, Qureshi AI, Malhotra A, et al. Comparing handcrafted radiomics versus latent deep learning features of admission head CT for hemorrhagic stroke outcome prediction. *Biotech*. 2025;14:87.
18. Ning Y, Yu Q, Fan X, Jiang W, Chen X, Jiang H, et al. Noncontrast CT-based deep learning for predicting intracerebral hemorrhage expansion incorporating growth of intraventricular hemorrhage. *Sci Rep*. 2025;15:32021.
19. Ma W, Chen C, Gong Y, Chan NY, Jiang M, Mak CH, et al. Causal effect estimation on imaging and clinical data for treatment decision support of aneurysmal subarachnoid hemorrhage. *IEEE Trans Med Imaging*. 2024;43:2778–89.
20. McAllister P, Zheng H, Bond R, Moorhead A. Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Comput Biol Med*. 2018;95:217–33.
21. Del Pérez A, Esteve Domínguez AS, Menéndez Fernández-Miranda P, Sanz Bellón P, Rodríguez González D, Lloret Iglesias LL, et al. A deep learning model for prognosis prediction after intracranial hemorrhage. *J Neuroimaging*. 2023;33:218–26.
22. Gao DQ, Feng WL, Qiao YY, Jiang X, Zhang Y. Development and validation of a random forest model to predict functional outcome in patients with intracerebral hemorrhage. *Neurol Sci*. 2023;44:3615–27.
23. Zhang KW, Zhou X, Xi Q, Wang X, Yang B, Meng J, et al. Outcome prediction of spontaneous supratentorial intracerebral hemorrhage after surgical treatment based on non-contrast computed tomography: a multicenter study. *J Clin Med*. 2023;12:1580.
24. Abou Karam G, Chen MC, Zeevi D, Harms BC, Berson E, Torres-Lopez BM, et al. Voxel-wise map of intracerebral hemorrhage locations associated with worse outcomes. *Stroke*. 2025;56:868–77.
25. Yu X, Elazab A, Ge R, Zhu J, Zhang L, Jia G, et al. ICH-PRNet: a cross-modal intracerebral haemorrhage prognostic prediction method using joint-attention interaction mechanism. *Neural Netw*. 2025;184:107096.
26. Qi X, Hu GR, Sun HY, Chen Z, Yang C. Machine learning-based perihematomal tissue features to predict clinical outcome after spontaneous intracerebral hemorrhage. *J Stroke Cerebrovasc Dis*. 2022;31:106475.
27. Wang S, Chen F, Zhang MY, Zhao X, Wen L, Wu W, et al. Predicting prognosis of primary pontine hemorrhage using CT image and deep learning. *NeuroImage Clin*. 2022;36:103257.
28. Yindeeddej V, Setprapha C, Komarapaj C, Osirichaivait K, Nimmannitya P, Noiphithak R. Applications of machine learning model for prediction of outcomes in primary pontine hemorrhage. *World Neurosurg*. 2023;175:e1348–59.

29. Wang HL, Hsu WY, Lee MH, Weng HH, Chang SW, Yang JT, et al. Automatic machine-learning-based outcome prediction in patients with primary intracerebral hemorrhage. *Front Neurol.* 2019;10:910.
30. Nawabi J, Kniep H, Elsayed S, Friedrich C, Sporns P, Rusche T, et al. Imaging-based outcome prediction of acute intracerebral hemorrhage. *Transl Stroke Res.* 2021;12:958–67.
31. Huang F, Chen N, Qiu A. 3D-CNN enhanced multiscale progressive vision transformer for AD diagnosis. *IEEE J Biomed Health Inf.* 2026;30:402–12.

Figures

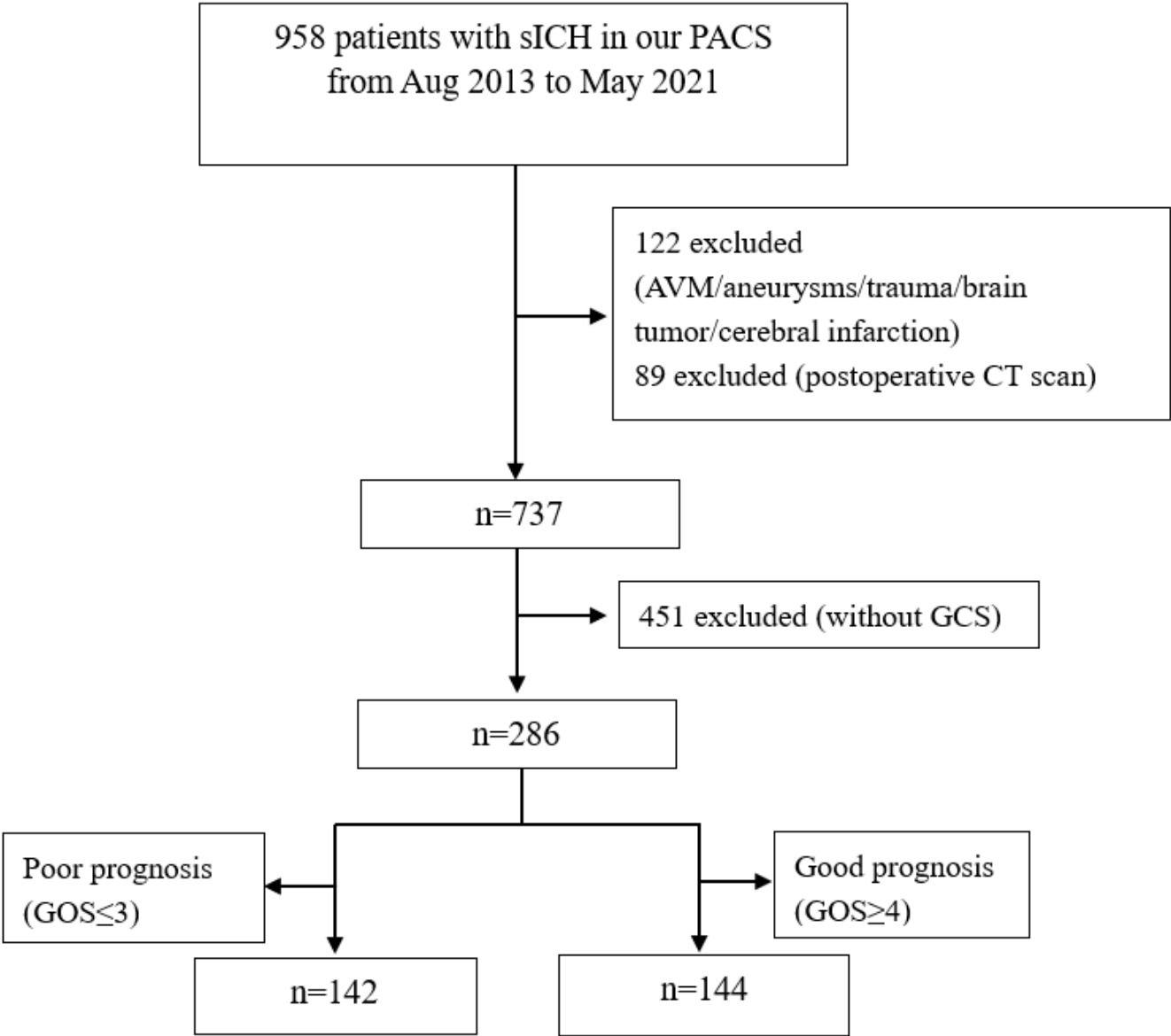


Figure 1

Flowchart of the study population

slCH, spontaneous intracerebral hemorrhage; PACS, picture archiving and communication system; AVM, arteriovenous malformation; CT, computed tomography; GCS, Glasgow Coma Scale; GOS, Glasgow Outcome Scale

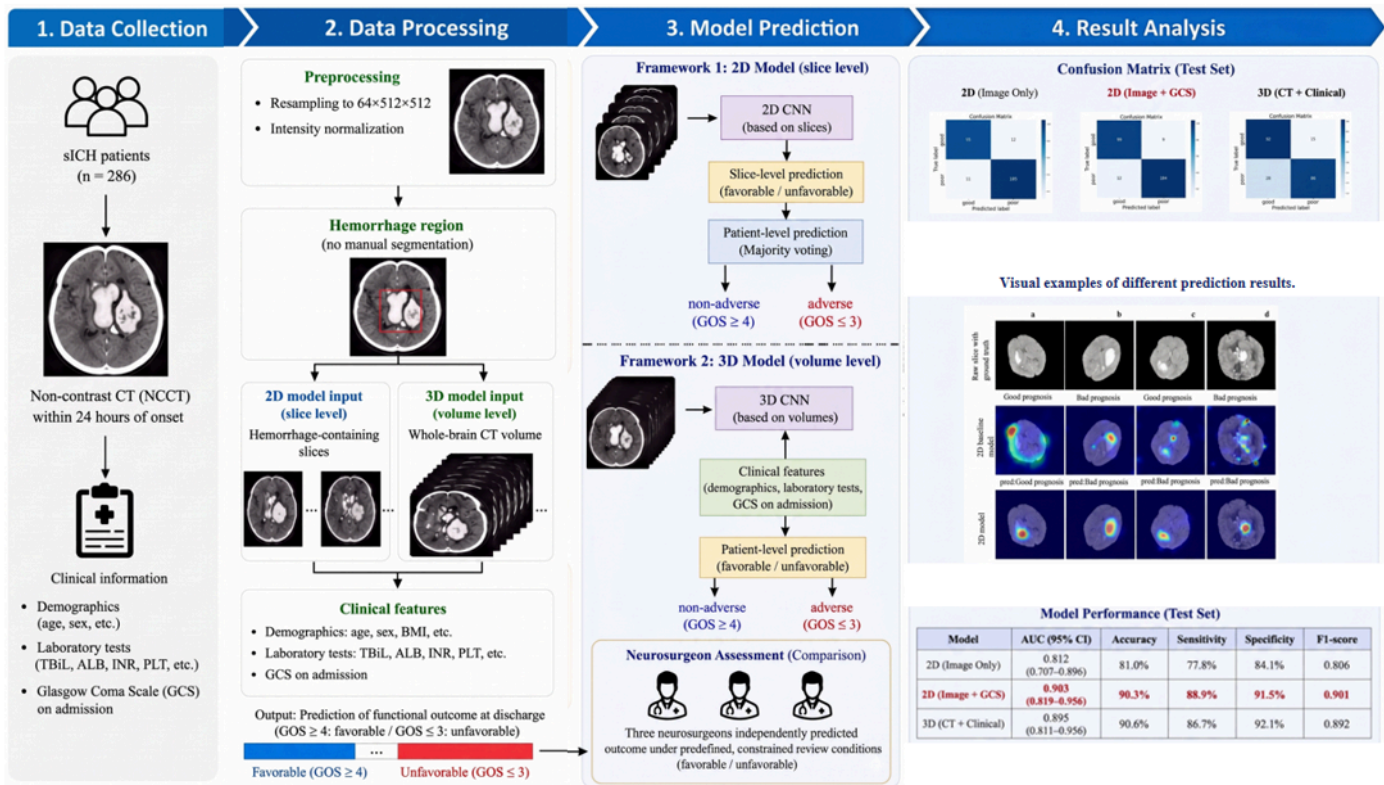


Figure 2

Study workflow overview

slCH, spontaneous intracerebral hemorrhage; NCCT, non-contrast computed tomography; TBil, total bilirubin; ALB, albumin; INR, international normalized ratio; PLT, platelet; GCS, Glasgow Coma Scale; 2D, two-dimensional; 3D, three-dimensional; CT, computed tomography; BMI, body mass index; GOS, Glasgow Outcome Scale; AUC, area under the curve; CI, confidence interval

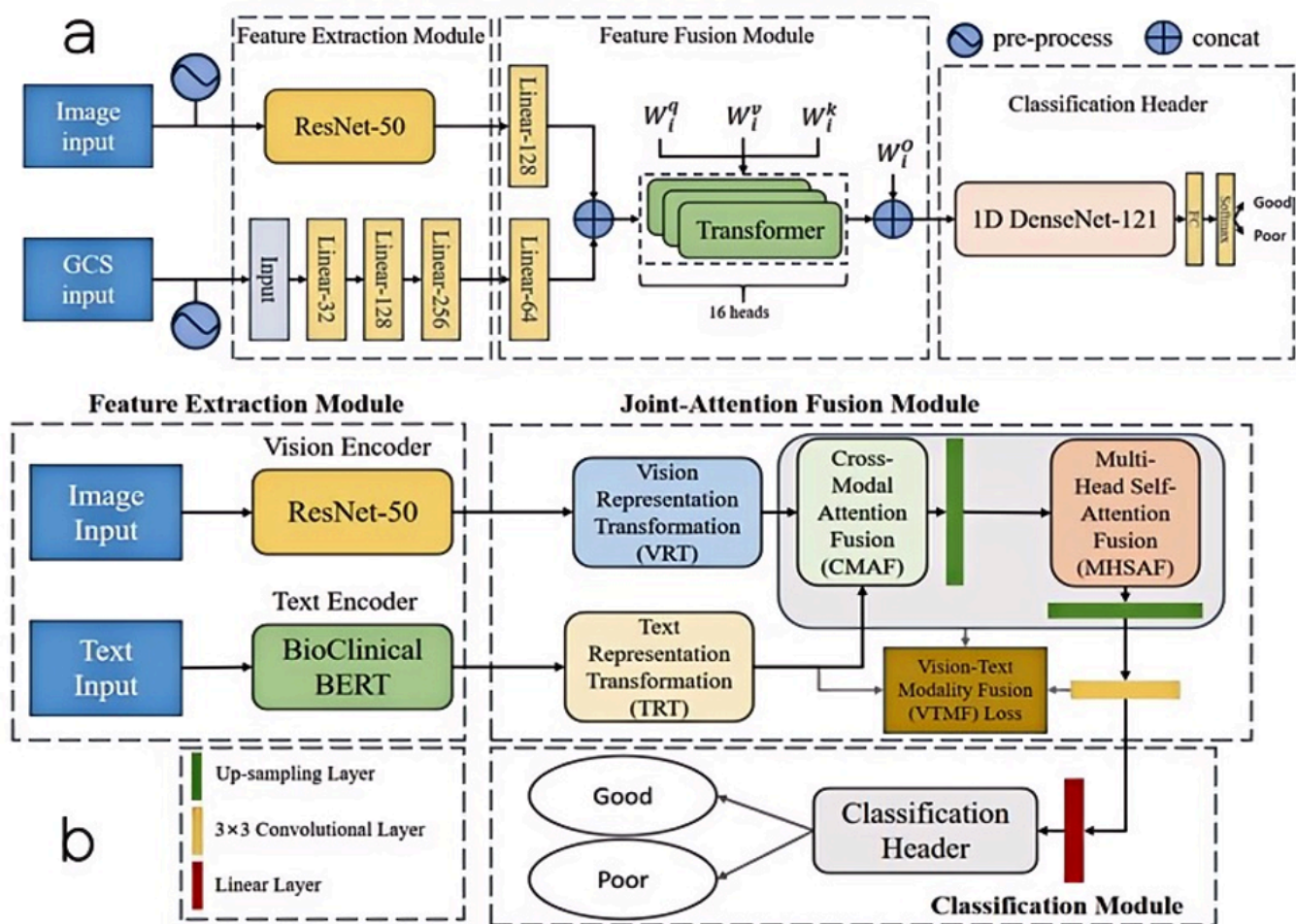


Figure 3

2D and 3D models

a. A 2D model network structure diagram. It is comprised of three main modules: feature extraction module, feature fusion module, and final classification head.

b. A 3D model network structure diagram. It contains three main modules: feature extraction module, joint attention fusion module, and classification module.

GCS, Glasgow Coma Scale; 2D, two-dimensional; 3D, three-dimensional

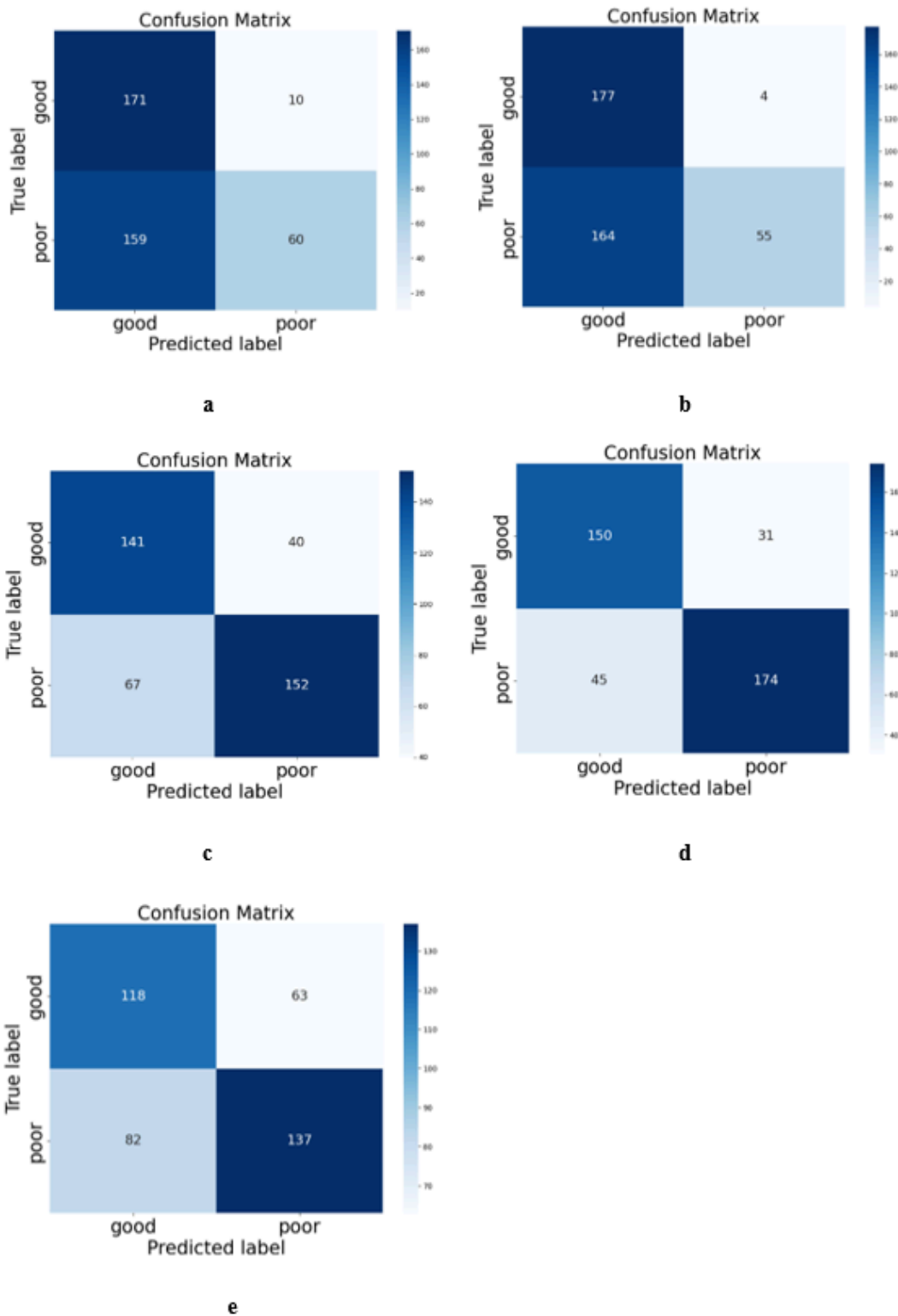
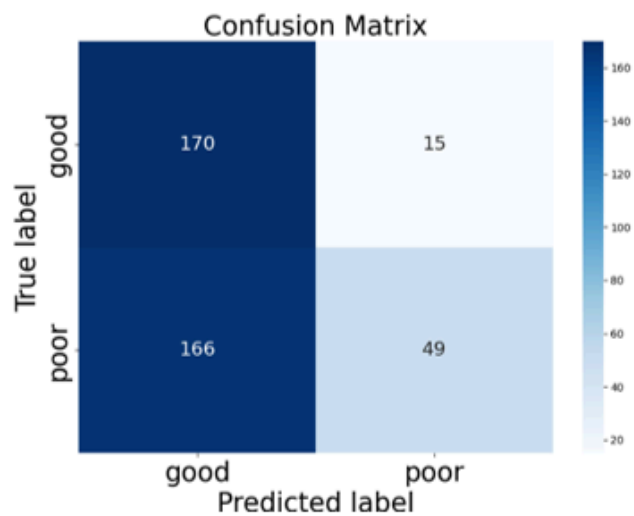


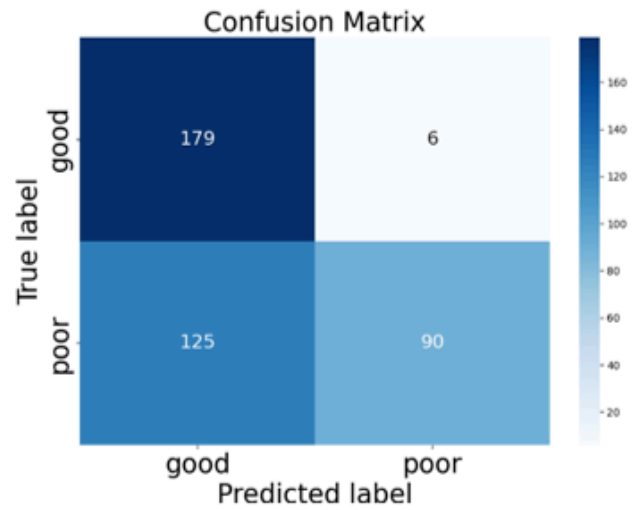
Figure 4

Confusion matrix for doctor 1 (a), doctor 2 (b), doctor 3 (c), 2D baseline model (d), and ResNet50 (2D) (e) using CT images

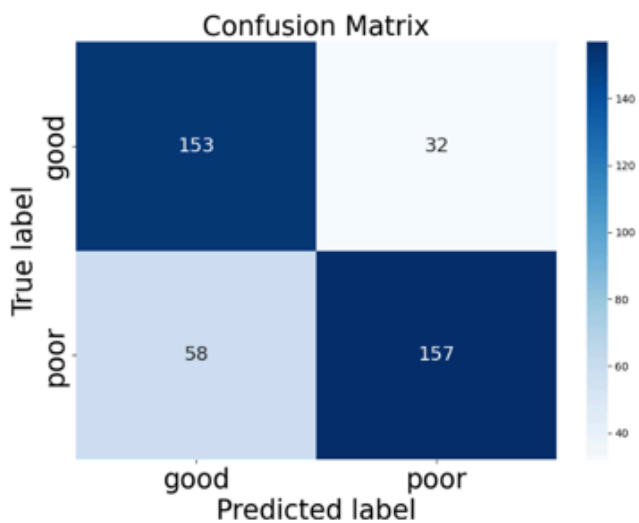
CT, computed tomography; 2D, two-dimensional



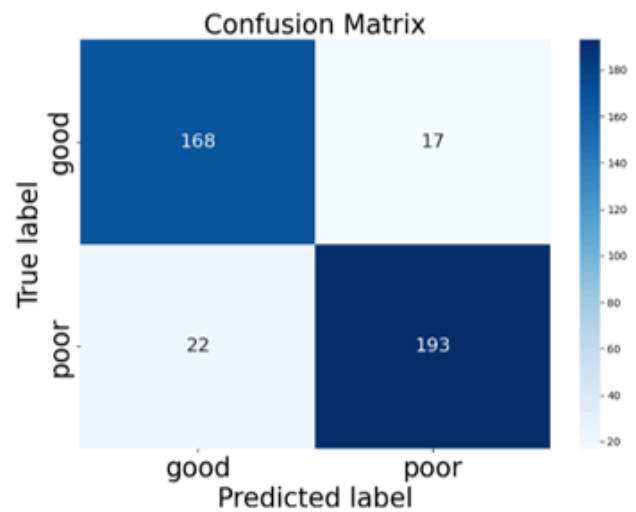
a



b



c



d

Figure 5

Confusion matrix for doctor 1 (a), doctor 2 (b), doctor 3 (c), and the 2D model (d) with computed tomography image and the GCS

GCS, Glasgow coma scale; 2D, two-dimensional

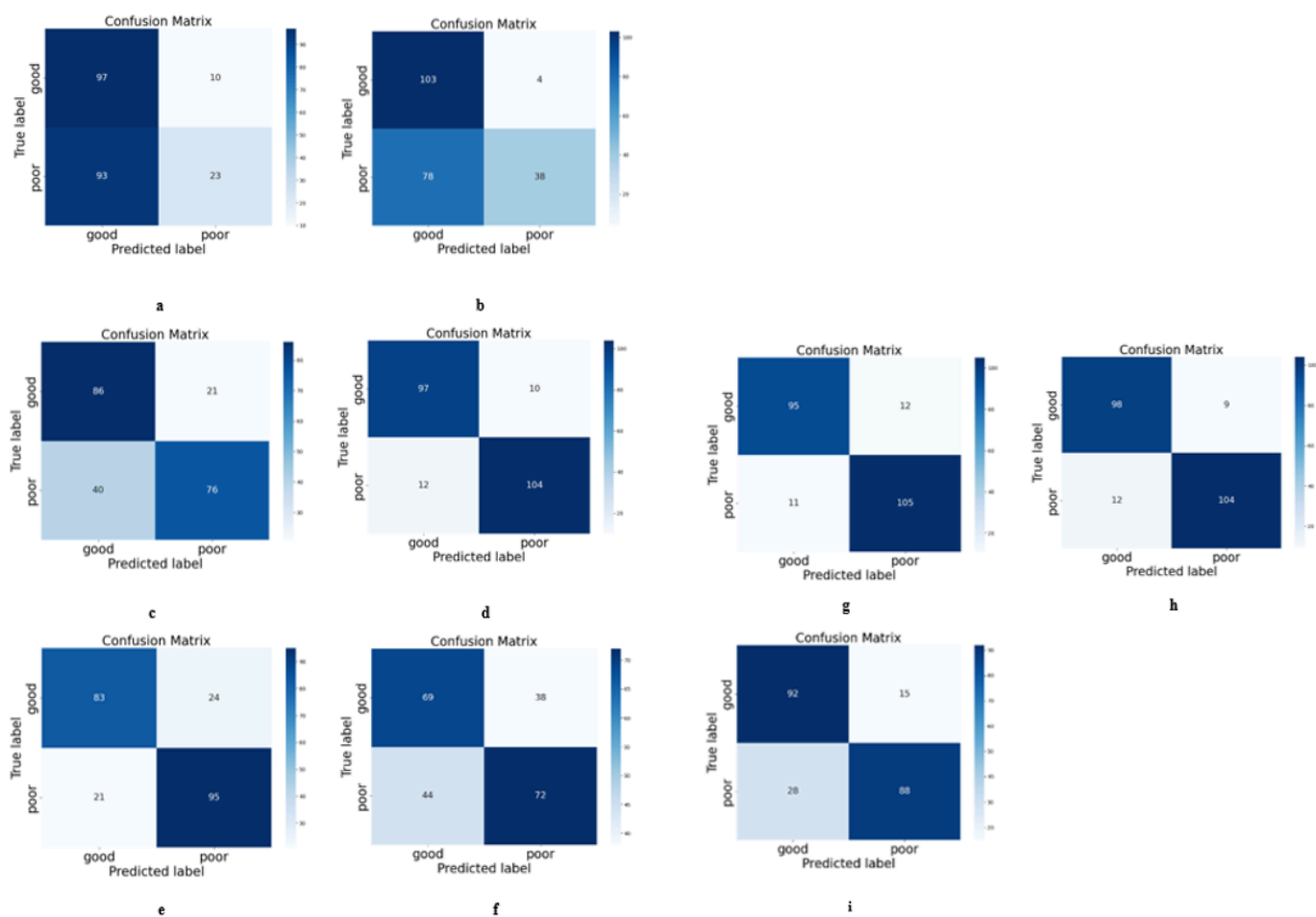


Figure 6

Confusion matrix for doctor 1 (a), doctor 2 (b), doctor 3 (c), 2D model (d) with CT and GCS, 3D model (e) with only CT, 3D model (f) with only baseline characteristics, 3D model (g) with CT and GCS, 3D model (h) with CT and baseline characteristics, and 3D deep learning-based model (i) with CT and baseline characteristics

CT, computed tomography; GCS, Glasgow Coma Scale; 2D, two-dimensional; 3D, three dimensional

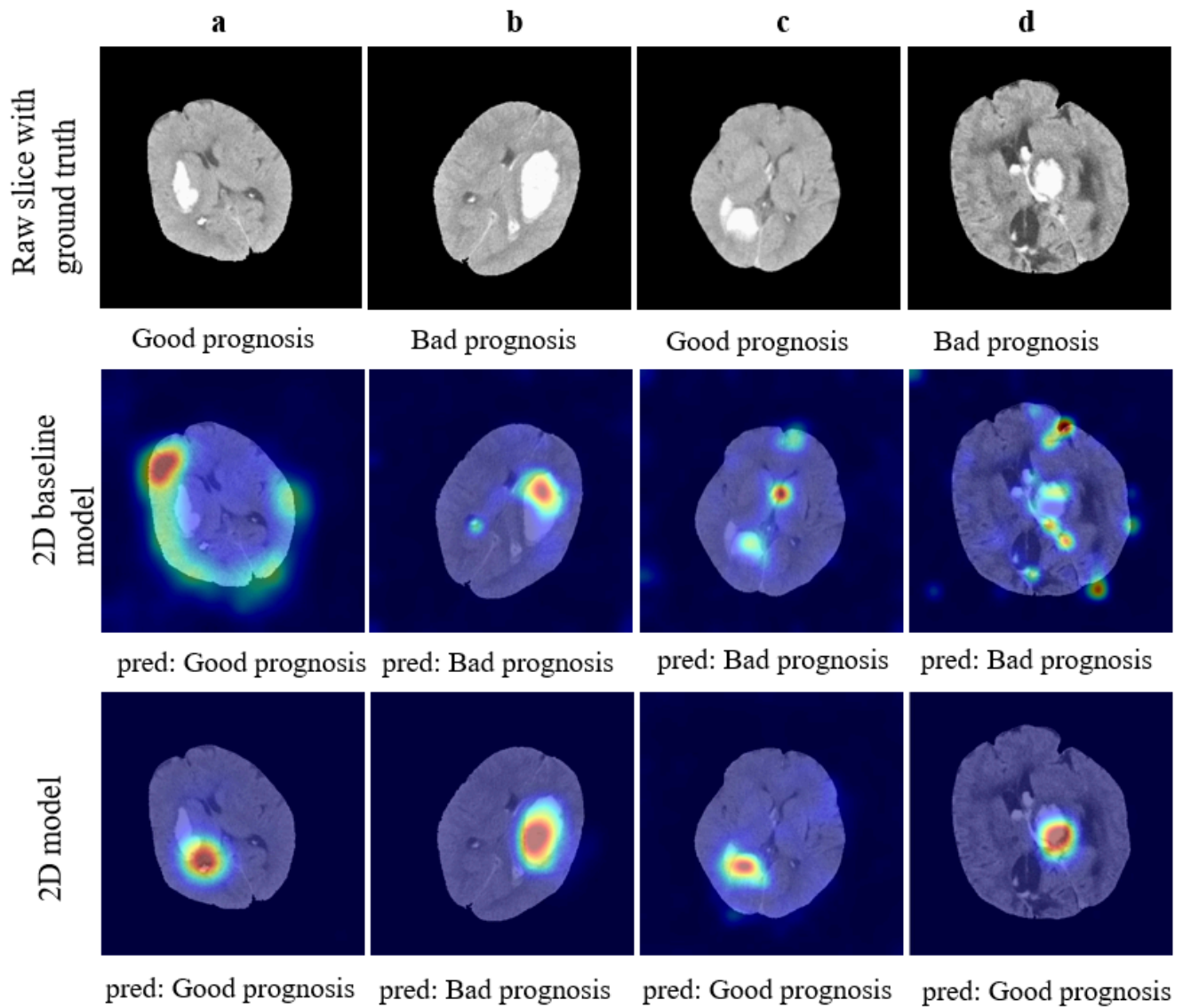


Figure 7

Visual examples of different prediction results

The red highlighted section indicates the class association activation area of the network for the prediction results.