Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Multicenter Study on a Multimodal Transformer Model for Predicting Early Hematoma Expansion

Xueqing Lv
Yiwu Traditional Chinese Medicine Hospital
Yue Wang
the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University
Jin Wang
Yiwu Traditional Chinese Medicine Hospital
Haobin Fu
Yiwu Traditional Chinese Medicine Hospital
Xuan Cheng
Yiwu Traditional Chinese Medicine Hospital
Yonghong Tao
Yiwu Chouzhou Hospital
Shanshan Jin
Yiwu Chouzhou Hospital
Fanxuan Meng
Yiwu Traditional Chinese Medicine Hospital
Yongren Wang
Yiwu Traditional Chinese Medicine Hospital
Qingdian Cong
the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University
Jiankuai Zhou
the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University
Jun Mo
the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University
Sheng Hu
8015135@zju.edu.cn

the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University

---

Research Article

Keywords: Intracerebral hemorrhage, hematoma expansion, radiomics, Deep learning, multimodal Transformer model

Posted Date: June 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9660726/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

Objective

This study aimed to develop and externally validate a multimodal transformer-based model that integrates deep learning (DL)-derived features, hematoma radiomics, perihematomal edema (PHE) radiomics, and clinical variables for early prediction of hematoma expansion (HE) in patients with spontaneous intracerebral hemorrhage (ICH).

Methods

This retrospective multicenter study included patients with ICH from three institutions (training set n = 232, internal set n = 100, external set n = 338). Three-dimensional DL features were extracted from baseline non-contrast computed tomography, alongside radiomics features derived from hematoma and PHE regions. After feature selection, single-modality K-nearest neighbor (KNN) models were developed and subsequently integrated using a multi-head self-attention transformer architecture. Model performance was evaluated utilizing the area under the receiver operating characteristic curve (AUC), DeLong test, decision curve analysis (DCA), and calibration curves.

Results

The multimodal transformer model demonstrated robust predictive performance, achieving an AUC of 0.856 (95% confidence interval: 0.802–0.901) in the external testing set. This integrated model significantly outperformed the multimodal KNN model (AUC = 0.715) and all single-modal models, including the DL model (AUC = 0.630), hematoma radiomics model (AUC = 0.616), PHE radiomics model (AUC = 0.540), and clinical model (AUC = 0.632). DeLong tests confirmed its statistical superiority over KNN-based models (*all P* < 0.05). Additionally, DCA and calibration curves demonstrated that the transformer model provided the greatest net clinical benefit and optimal predictive calibration.

Conclusion

The proposed multimodal transformer model, notably enhanced by the novel inclusion of PHE radiomics, demonstrates strong performance in predicting early HE expansion. This comprehensive framework highlights the robust feasibility of multimodal data fusion, offering a highly reliable tool for early clinical decision-making.

1. Introduction

Intracerebral hemorrhage (ICH) is a severe cerebrovascular event associated with high disability and mortality rates, comprising approximately 10%–15% of all stroke cases[1]. Clinical evidence demonstrates that approximately 20% of patients with acute ICH develop hematoma expansion (HE) within 24 h of onset, as confirmed by follow-up non-contrast computed tomography (NCCT). Notably, patients with HE exhibit a two-to-threefold increased risk of poor prognosis compared to those with stable hematomas[2]. Consequently, the early and precise identification of HE risk is essential for timely clinical intervention and the enhancement of patient outcomes.

Currently, various imaging markers on NCCT, combined with clinical features, have been substantiated as HE predictors[3, 4]. However, conventional radiological indicators frequently rely on qualitative or semi-quantitative evaluations, which are intrinsically susceptible to inter-observer variability and possess restricted sensitivity[5]. To address these limitations, artificial intelligence (AI) has been increasingly integrated into ICH research. Radiomics facilitates the extensive high-throughput extraction of quantitative features from hematoma regions that are undetectable by the human eye, functioning as a robust tool for HE prediction[6, 7]. Despite these advancements, most existing models primarily focus on the

hematoma, often neglecting the perihematomal edema (PHE). As a critical biomarker of secondary brain injury, PHE demonstrates a relatively prolonged growth window and diverse patterns, offering a crucial opportunity for early clinical assessment and intervention[8, 9, 10]. Recent studies have begun to identify correlations between PHE-based radiomic signatures and clinical outcomes, indicating their unexploited potential in improving predictive models[11, 12]. Furthermore, deep learning has demonstrated remarkable proficiency in autonomous feature learning, bypassing the subjectivity inherent in manual feature extraction. However, single-modality approaches inadequately represent the complex, non-linear correlations between clinical, radiomic, and deep learning data[13, 14].

The transformer architecture, characterized by its multi-head self-attention mechanism, has recently emerged as a robust framework for multimodal information integration. Unlike conventional models, Transformers can dynamically assign weights to various data sources and capture complex feature interdependencies, providing a novel avenue for high-precision clinical prediction[15, 16].

This study aimed to develop and validate a comprehensive multimodal transformer model that synergistically integrates conventional clinical-radiological features, hematoma and PHE radiomics, and deep learning signatures. Utilizing a multicenter dataset, we assess the model's performance across multiple dimensions to establish a robust and generalizable framework for the early prediction of hematoma expansion among patients with ICH.

2. Materials and methods

2.1. Patient selection and study design

This retrospective multicenter study was approved by the Ethics Committee of Yiwu Hospital of Traditional Chinese Medicine (Approval No. 2025-009). Due to the retrospective nature of the study, the need for informed patient consent was waived. We systematically collected clinical and imaging data from three institutions: Hospital 1 (patients with spontaneous intracerebral hemorrhage diagnosed and treated at Yiwu Hospital of Traditional Chinese Medicine between January 2021 and December 2024), Hospital 2 (patients with spontaneous intracerebral hemorrhage who were diagnosed and treated by the Yiwu Chouzhou Hospital between December 2019 and June 2024), and Hospital 3 (patients with spontaneous intracerebral hemorrhage who were diagnosed and treated by the Fourth Affiliated Hospital of Zhejiang University School of Medicine between December 2015 and December 2024).

The inclusion criteria were as follows: (1) A confirmed diagnosis of primary ICH; (2) an initial NCCT scan performed within 24 h of symptom onset; (3) a follow-up NCCT scan completed within 72 h after the initial scan.

The exclusion criteria included the following: (1) Baseline computed tomography (CT) scan acquired > 24 h after onset; (2) follow-up CT acquired > 72 h after the initial scan; (3) surgical intervention performed before the follow-up NCCT; (4) ICH secondary to tumors, aneurysms, or arteriovenous malformations; (5) severe image artifacts, distortion, or blurring; (6) inability to precisely delineate the PHE region of interest (ROI); (7) incomplete clinical records.

The combined data from Hospitals 1 and 2 were randomly divided into an internal training set and an internal testing set at a 7:3 ratio. The data from Hospital 3 were utilized as an independent external testing set. Figure 1 depicts the detailed patient screening and enrollment workflow.

2.2. Clinical data collection and image feature assessment

Figure 1 displays the overall workflow of our study. Baseline clinical parameters were extracted from the electronic medical record systems. These included age, sex, initial CT hematoma volume, Glasgow Coma Scale (GCS) score, Graeb score, admission systolic and diastolic blood pressure, and a comprehensive medical history (hypertension, antihypertensive treatment, heart disease, diabetes, renal insufficiency, smoking, and the use of antiplatelet or anticoagulation therapies).

Conventional imaging signs were assessed independently by two neuroradiologists and encompassed brain herniation, intraventricular hemorrhage (IVH), black hole sign, swirl sign, island sign, satellite sign, blend sign, fluid level sign, and midline shift. Any diagnostic discrepancies were resolved through consensus with a senior neuroimaging physician who has 20 years of experience. Missing clinical data were addressed through multiple imputation for variables with < 5% missingness, whereas variables with ≥ 5% missing data were excluded from the analysis[17]. HE was explicitly characterized by an absolute volume increase of ≥ 6 mL or a relative volume increase of ≥ 33% on the follow-up NCCT compared to the baseline scan[18]. Table 1 presents comprehensive clinical and imaging features.

2.3. Image acquisition and ROI segmentation

Baseline and follow-up NCCT scans were obtained using standard protocols (120 kV, 5 mm slice thickness; Section 1 of the Appendix). Figure 2 illustrates the overall analytical pipeline. Regions of interest(ROI)segmentation was conducted using the Deepwise Multimodal Research Platform (version 2.5.1; https://keyan.deepwise.com, Hangzhou Deepwise and League of PHD Technology Co., Ltd, Hangzhou, Zhejiang, China). To minimize inter-scanner variability, all images were resampled to 1.0 ×1.0 × 1.0 mm^3 isotropic voxels, discretized into 256 gray-level bins. Baseline hematoma volumes (V1) and follow-up volumes within 72 h (V2) were calculated automatically[19]. When multiple follow-up head CT scans were available within the 72-h window, the final scan was selected for volumetric analysis to effectively assess HE. In patients with concurrent IVH, hematoma volumetry was exclusively limited to the parenchymal components of the deep ICH, thereby omitting intraventricular blood.

Figure. 2 Workflow of the multimodal Transformer model for prediction

of early hematoma expansion

Two radiologists employed the software to semi-automatically delineate the ICH-ROI and manually trace the PHE-ROI. PHE was recognized as the hypoattenuating region surrounding the hematoma. The ICH boundary was initially traced layer-by-layer through automatic level tracing. Manual adjustments, including contour expansion or deletion, were implemented as required. When the hematoma infiltrated the ventricles, the ventricular contours constituted the definitive boundary, ensuring that only the parenchymal hematoma was encompassed within the ROI, whereas surrounding brain tissue and IVH were excluded. To thoroughly evaluate the reliability and reproducibility of the extracted radiomic features, 30 randomly selected lesions were re-evaluated to calculate the intraclass correlation coefficient (ICC) for inter-reader and intra-reader consistency (Section 2 of the Appendix).

2.4. Radiomics feature extraction and model construction

Before feature extraction, all NCCT images underwent systematic preprocessing, which involved spatial resampling to an isotropic resolution of 1 × 1 × 1 mm^3 and grayscale discretization into 25 bins using nearest-neighbor interpolation. We extracted 1,834 radiomics features from the respective ROIs of the hematoma and PHE. The hematoma and PHE radiomics feature set comprised 14 shape features, 252 first-order features, and various higher-order textures: 308 gray-level co-occurrence matrix (GLCM) features, 224 gray-level run-length matrix (GLRLM) features, 224 gray-level size zone matrix (GLSZM) features, 196 gray-level dependence matrix (GLDM), and 616 additional texture features, including the neighborhood gray-tone difference matrix (NGTDM). The extracted parameters denote geometric forms, shape-related characteristics, and complex texture features. Various image transformation techniques, including filtering, logarithmic, and Gaussian transformations, were utilized to enhance feature extraction.

To prevent data bias, the extracted radiomics features underwent Z-score standardization. Redundant features with a Pearson correlation coefficient > 0.9 were systematically removed. Subsequently, the least absolute shrinkage and selection operator (LASSO) regression was employed to identify the optimal predictive features for hematoma and PHE. These selected features were subsequently independently input into K-nearest neighbor (KNN) machine learning classifiers to develop the respective Hematoma and PHE radiomics models.

2.5. Deep learning feature extraction

A three-dimensional volume of interest (VOI) was reconstructed for each patient and analyzed using an ImageNet-pretrained ResNet18 network. To meet the architectural specifications, images were cropped and scaled to a standard resolution of 224 × 224 pixels employing linear interpolation. Deep learning is defined using a standardized fixed brain window (width: 90 HU, level: 40 HU). To enhance the model's generalizability, data augmentation strategies, including random horizontal and vertical flipping, were employed during the training phase.

The network was trained over 50 epochs using the Adam optimizer, with a batch size of 32, an initial learning rate of 1×10^{-5}, L2 regularization, and an early stopping criterion. Post-training, high-dimensional deep learning feature vectors were obtained from the penultimate average pooling layer. Principal component analysis (PCA) was initially employed for dimensionality reduction. After Z-score normalization, LASSO regression was utilized to identify the most highly predictive deep learning features. The core features were employed to train a KNN algorithm, forming the DL model.

2.6. Multimodal Transformer model construction

To develop the unified multimodal prediction framework, LASSO regression was initially utilized to screen the clinical and conventional imaging variables. The clinical features were integrated with the selected radiomics and deep learning features, creating a comprehensive, unified feature vector that served as the input for the KNN or transformer architecture.

The Transformer model inherently utilizes a multi-head self-attention mechanism to identify complex, non-linear correlations across the diverse modalities, facilitating adaptive weight distribution[20]. The multimodal model was refined using stochastic gradient descent (SGD) over 200 epochs, with an initial learning rate of 0.001 and a batch size of 64. Regularization techniques and early stopping were diligently maintained.

2.7. Statistical analysis

All statistical analyses were performed using the Statistical Package for the Social Sciences software (version 27.0). Categorical variables are presented as frequencies (percentages) and analyzed using the chi-square test or Fisher’s exact test. Continuous variables were assessed using the independent t-test or Mann–Whitney U test, contingent upon data distribution, and presented as means ± standard deviations or medians (interquartile ranges).

The model’s predictive efficacy was assessed using receiver operating characteristic (ROC) curve analysis, which reported the AUC alongside 95% confidence intervals (CI), accuracy, sensitivity, and specificity. Moreover, the consistency between predicted and actual HE outcomes was assessed through calibration curves, whereas the clinical utility and net patient benefit of the models were evaluated using decision curve analysis (DCA).

3. Results

3.1. Patient characteristics

We allocated 232 patients to the internal training set. The internal testing set comprised 100 patients. The external testing set comprised 338 patients. Figure 1 depicts the detailed patient screening workflow. Table 1 presents baseline clinical characteristics and conventional imaging signs.

Table 1
Patient baseline characteristics of the training set and the testing set

HE, n(%)	training set(N = 232)	Internal testing set(N = 100)	External testing set(N = 338)	P
	43(18.5)	**19(19)**	**56(16.6)**	**0.77**
Gender(male),n(%)	166(71.6)	73(73)	241(71.3)	0.95
Age	58 ± 15.3	58 ± 15.9	55 ± 14.6	0.56
GCS Scale	14 ± 2.1	14 ± 1.6	14 ± 3.1	< 0.001
Initial CT Hematoma Volume	12.5(6.3–25.6)	14.6(7.5–27.9)	15.2(7.7–32)	0.19
Graeb Score	0 ± 1.9	0 ± 2	0 ± 2	0.56
Hypertension History,n(%)	170(73.3)	66(66)	239(70.7)	0.41
Antihypertensive Treatment,n(%)	99(42.7)	31(31)	132(39.1)	0.14
Heart Disease History,n(%)	17(7.3)	5(5)	10(3)	0.06
Diabetes History,n(%)	27(11.6)	10(10)	32(9.5)	0.70
Renal Insufficiency,n(%)	5(2.2)	0(0)	8(2.4)	0.31
Smoking History,n(%)	61(26.3)	24(24)	101(29.9)	0.43
Antiplatelet History,n(%)	25(10.8)	7(7)	28(8.3)	0.45
Anticoagulation History,n(%)	5(2.2)	0(0)	6(1.8)	0.35
Systolic Blood Pressure on Admission	161 ± 25	158 ± 25.4	162 ± 29.8	0.4
Diastolic Blood Pressure on Admission	92 ± 15.7	95.5 ± 15	95 ± 20.2	0.99
Brain Herniation,n(%)	7(3)	4(4)	11(3.3)	0.9
Ventricular System Hemorrhage,n(%)	61(26.3)	24(24)	88(26)	0.9
Black Hole Sign,n(%)	40(17.2)	18(18)	69(20.4)	0.62
Swirl Sign,n(%)	171(73.7)	67(67)	239(70.7)	0.45
Island Sign,n(%)	40(17.2)	16(16)	49(14.5)	0.67
Satellite Sign,n(%)	83(35.8)	50(50)	132(39.1)	0.05
Blend Sign,n(%)	55(23.7)	15(15)	81(24)	0.15
Fluid Level Sign,n(%)	1(0.4)	1(1)	6(1.8)	0.34
Midline Displacement,n(%)	53(22.8)	21(21)	136(40.2)	< 0.001

HE: hematoma expansion;GCS：glasgow coma scale；Graeb：graeb intraventricular hemorrhage scale.

3.2. Clinical model

Among the 24 initial clinical and conventional imaging variables, LASSO regression identified six optimal predictors: baseline hematoma volume on CT, swirl sign, island sign, anticoagulation history, brain herniation, and GCS score. These

variables were incorporated into a KNN classifier to construct the clinical model. The clinical model achieved an AUC of 0.839 (95% CI: 0.786–0.889) in the training set. In the validation cohort, the clinical model yielded an AUC of 0.628 (95% CI: 0.488–0.769) in the internal testing set and 0.632 (95% CI: 0.553–0.707) in the external testing set.

3.3. Radiomics feature selection and model construction

Initially, 1,834 radiomics features were extracted from the hematoma and PHE ROIs. After Z-score normalization and removal of highly correlated features via Pearson correlation, LASSO regression identified 13 optimal hematoma features and 6 optimal PHE features for model development.

The hematoma radiomics model achieved an AUC of 0.876 (95% CI: 0.832–0.922) in the internal training set, 0.744 (95% CI: 0.630–0.849) in the internal testing set, and 0.616 (95% CI: 0.535–0.697) in the external testing set. Conversely, the PHE radiomics model yielded an AUC of 0.835 (95% CI: 0.778–0.883) in the internal training set, with validation AUCs of 0.530 (95% CI: 0.396–0.663) and 0.540 (95% CI: 0.464–0.614) in internal and external testing sets, respectively.

3.4. Deep learning feature selection and model construction

Imaging data from all 670 enrolled patients were processed using a pre-trained ResNet18 network. After extracting 512-dimensional feature vectors from the penultimate average pooling layer, PCA was used to compress these vectors to 256 dimensions. Following Z-score standardization, LASSO regression further isolated six highly predictive deep learning features.

The DL model constructed from these features achieved an AUC of 0.888 (95% CI: 0.845–0.925) in the training cohort. In the validation phases, the DL model demonstrated reduced performance, with an AUC of 0.621 (95% CI: 0.476–0.765) in the internal testing set and 0.630 (95% CI: 0.549–0.710) in the external testing set.

3.5. Multimodal KNN model performance

Six optimized clinical features, 13 hematoma radiomics features, six PHE radiomics features, and six deep learning features were integrated to construct a multimodal KNN model. The model achieved an AUC of 0.912 (95% CI: 0.878–0.943) in the training set, 0.723 (95% CI: 0.584–0.846) in the internal testing set, and 0.715 (95% CI: 0.633–0.790) in the external testing set.

3.6. Multimodal Transformer model performance

A unified fusion feature set was constructed by integrating 6 optimized clinical features, 13 hematoma radiomics features, 6 PHE radiomics features, and 6 deep learning features were integrated to construct a unified fusion feature set. The resulting multimodal transformer model exhibited exceptional predictive capability, achieving the highest AUC of 0.955 (95% CI: 0.922–0.982) in the training set. This exceptional performance was robustly maintained during validation, with an AUC of 0.860 (95% CI: 0.746–0.955) in the internal testing set and 0.856 (95% CI: 0.802–0.901) in the external testing set. Table 2 presents detailed performance metrics.

Table 2
Effectiveness of different models in different subgroups

Model name	Accuracy	AUC	95%CI	Sensitivity	Specificity	PPV	NPV	Set
Clinic_Model	0.595	0.839	0.786–0.889	1	0.503	0.314	1	training
Rad_ICH_Model	0.647	0.876	0.832–0.922	1	0.566	0.344	1	training
Rad_PHE_Model	0.591	0.835	0.778–0.883	1	0.497	0.312	1	training
DL_Model	0.853	0.888	0.845–0.925	0.698	0.889	0.588	0.928	training
KNN_Combine_Model	0.741	0.912	0.878–0.943	1	0.683	0.417	1	training
Transformer_Combine_Model	0.888	0.955	0.922–0.982	0.884	0.889	0.644	0.971	training
Clinic_Model	0.75	0.628	0.488–0.769	0.474	0.815	0.375	0.868	internal testing
Rad_ICH_Model	0.59	0.744	0.630–0.849	0.895	0.519	0.304	0.955	internal testing
Rad_PHE_Model	0.67	0.530	0.396–0.663	0.368	0.741	0.250	0.833	internal testing
DL_Model	0.75	0.621	0.476–0.765	0.368	0.840	0.35	0.850	internal testing
KNN_Combine_Model	0.64	0.723	0.584–0.846	0.737	0.617	0.311	0.909	internal testing
Transformer_Combine_Model	0.86	0.860	0.746–0.955	0.789	0.877	0.6	0.947	internal testing
Clinic_Model	0.731	0.632	0.553–0.707	0.429	0.791	0.289	0.875	external testing
Rad_ICH_Model	0.716	0.616	0.535–0.697	0.429	0.773	0.273	0.872	external testing
Rad_PHE_Model	0.438	0.540	0.464–0.614	0.679	0.390	0.181	0.859	external testing
DL_Model	0.541	0.630	0.549–0.710	0.679	0.514	0.217	0.890	external testing
KNN_Combine_Model	0.642	0.715	0.633–0.790	0.714	0.628	0.276	0.917	external testing
Transformer_Combine_Model	0.805	0.856	0.802–0.901	0.821	0.801	0.451	0.958	external testing

3.7. Comparison of model performance

The multimodal Transformer model consistently outperformed all individual modality models across the training, internal testing, and external testing cohorts. Notably, the PHE radiomics model demonstrated the least predictive efficacy among the evaluated models (Figs. 3A–C). Table 2 presents detailed performance comparisons.

The DeLong test analysis demonstrated that the predictive performance of the transformer model was markedly superior to all KNN-based models across all three datasets (Figs. 3D–F). DCA revealed that the multimodal Transformer model yielded

the highest net clinical benefit across all cohorts when compared to the independent clinical, hematoma radiomics, PHE radiomics, and DL models (Figs. 3G–I). Furthermore, calibration curves demonstrated that the probabilities of HE predicted by the Transformer model exhibited the highest concordance with actual observed HE occurrences compared to the other baseline models in the testing set (Figs. 3J–L).

Figure. 3. Comparative analysis in the ROC curves (A-C) of the clinic model, hematoma radiomics, perihematoma edema radiomics,deep learning, KNN combine and transformer models. DeLong test for comparing the statistical significance of the AUC differences between models (D-F) in the DeLong test for six model among different subgroups. (G-I) DCA decision curve analysis to assess the comparative clinical utility of the six models across subgroups. (J-L) Calibration curve analysis to analyze observations in different cohorts to visualize predictive accuracy.

4. Discussion

Herein, we developed and externally validated a multimodal Transformer model that comprehensively integrates DL signatures, hematoma radiomics, PHE radiomics, and conventional clinical-radiological features. Our findings indicate that the multimodal transformer architecture markedly surpasses single-modality and conventional machine learning models, attaining a benchmark AUC of 0.856 (95% CI: 0.802–0.901) in the external testing set. This remarkable predictive capability offers a reliable, non-invasive tool for early identification of high-risk ICH patients, thereby facilitating prompt therapeutic interventions and potentially improving clinical outcomes.

Recent advancements in artificial intelligence have significantly improved the accuracy of HE prediction in medical image analysis[21, 22]. Previous studies have demonstrated that integrating multi-source data produces superior prognostic models compared to unimodal methods. For instance, Li et al.[23] introduced a DL-based framework that achieved an AUC of 0.806 in external validation, whereas Lee et al.[24] developed a multi-task DL model that integrated clinical factors, achieving an AUC of 0.83. However, the study by Lee et al. was limited to a single-center study. Based on these foundations, our multimodal transformer model exhibits superior performance and has been validated in external testing sets. This performance leap validates the hypothesis that highly dimensional, cross-domain feature fusion is crucial for elucidating the complex pathophysiology of HE.

Our research findings indicate that the AUC of the multimodal transformer model on the external testing set is 0.856 (95% CI: 0.802–0.901), which is statistically significantly superior to the multimodal KNN model (0.715 (95% CI: 0.633–0.790); $P < 0.05$, DeLong test). Consequently, the enhanced efficacy of our proposed model is primarily due to the transformer architecture, specifically its fundamental multi-head self-attention mechanism. In contrast to conventional machine learning algorithms, Transformer proficiently interprets complex, non-linear correlations across highly heterogeneous multimodal inputs. It dynamically assigns weights, allowing the network to prioritize the most prognostically salient features and effectively integrating diverse data streams within a unified framework[25, 26]. Compared to a recent study by Feng et al.[27], which employed EfficientNet-B0 to integrate DL and radiomics features (external AUC = 0.82), our transformer-based strategy demonstrated a significant performance increase. This highlights the unique benefit of self-attention mechanisms in managing high-dimensional biomedical data compared to conventional convolutional classifiers. Moreover, our findings align with those of Yang et al.[28], who highlighted the prognostic significance of PHE. however, their machine learning model—which depends solely on hematoma and PHE radiomics without DL or clinical features—experienced a significant performance drop in external validation (AUC = 0.751). Conversely, our comprehensive integration maximized predictive accuracy and ensured remarkable model generalizability. Notably, although the performance of all single-modality models in our study degraded significantly during external validation (the clinical model decreased from 0.839 to 0.632, and the DL model decreased from 0.888 to 0.630), the multimodal transformer model maintained robust stability (external AUC = 0.856). This resilience highlights the robustness of our multimodal fusion strategy, validating its appropriateness for broad, cross-center clinical deployment.

A key innovation of this study is the prospective integration of PHE radiomics into the multimodal fusion framework. Although conventional HE prediction models primarily focus on the hematoma core, PHE serves as an essential, dynamic biomarker of secondary brain injury. The heterogeneous texture and morphological evolution of PHE reflect the underlying biological behavior of the hematoma, offering an independent and highly valuable dimension of predictive information[29, 30]. For instance, although Zhang et al.[31] employed a transformer architecture for multimodal fusion (AUC of 0.822), our inclusion of PHE radiomics, combined with an optimized feature integration strategy, yielded a significant performance enhancement.

Despite these promising results, several limitations warrant consideration. First, the retrospective, multicenter design inherently introduced selection bias, and variations in CT scanners, acquisition parameters may result in data heterogeneity. Consequently, large-scale, prospective, multinational cohorts are crucial for rigorously validating the model's generalizability. Second, the manual and semi-automated delineation of hematoma and PHE ROIs is susceptible to intra- and inter-observer variability, potentially impacting the stability of extracted radiomic features. The future integration of fully automated, DL-based segmentation algorithms will be essential for improving analytical efficiency and reproducibility. The comparatively low natural incidence of HE led to a class imbalance in our dataset. Although LASSO regularization and data augmentation techniques were utilized to mitigate overfitting, training highly parameterized models on imbalanced data remains challenging. Expanding the dataset with a larger, more balanced cohort will be essential to enhance the model's robustness and clinical reliability.

5. Conclusion

This study effectively developed and validated a multimodal transformer-based model for the early prediction of HE in patients with spontaneous intracerebral hemorrhage. By synergistically combining clinical imaging signs, deep learning signatures, and radiomics features from hematoma and PHE, our model demonstrated enhanced predictive efficacy and robustness compared to unimodal and conventional machine learning methods. This multimodal fusion strategy may function as an essential clinical tool for identifying patients with ICH at high risk of HE.

Declarations

Ethics approval and consent to participate

This retrospective study was approved by by the Ethics Committee of Yiwu Hospital of Traditional Chinese Medicine (Approval No. 2025-009),and informed consent was waived,given the retrospective study and anonymised data of patients. This study was conducted in accordance with the World Medical Association Declaration of Helsinki-Ethical Principles for Medical Research Involving Human Subjects.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding.

Author Contribution

All authors contributed to the conception and design of this study.The initial draft was written by X. L.Y. W,J.W,H.Fu,X. C,F. M,Y. W and Q.C completed the data extraction and evidence synthesis.Y. W,Y.T and S. J performed the quality assessment of the literature.Y. W,J.Z and J. M conducted the database literature search.S.H proof reading and polished the final manuscript.This study was conducted under the guidance of S. H.All authors reviewed and approved the final version of the manuscript.

Acknowledgements

Not Applicable.

Data Availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, Boehme AK, Buxton AE, Carson AP, Commodore-Mensah Y, Elkind MSV, Evenson KR, Eze-Nliam C, Ferguson JF, Generoso G, Ho JE, Kalani R, Khan SS, Kissela BM, Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. American Heart Association. (2022). Heart Disease and stroke statistics—2022 Update: A report from the American Heart Association. In Circulation (Vols. 145–e639) [Report]. https://doi.org/10.1161/CIR.001052
2. Li Z, You M, Long C, Bi R, Xu H, He Q, Hu B. Hematoma expansion in intracerebral hemorrhage: An update on prediction and treatment. Front Neurol. 2020;11:702. https://doi.org/10.3389/fneur.2020.00702.
3. Chen S, Zhao 1, Wang B, 1 W, Shi L, 1, Reis C, 2 J. 1, Department of Neurosurgery, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, PR China, Department of Physiology and Pharmacology, Loma Linda University, Loma Linda, California, USA, & Department of Preventive Medicine, Loma Linda University, Loma Linda, California, USA. (2017). Predictors of hematoma expansion predictors after intracerebral hemorrhage. In Oncotarget (Vol. 8, Issue 51, pp. 89348–89363) [Journal-article]. https://www.impactjournals.com/oncotarget
4. Li H, Xie Y, Liu H, Wang X. Non-Contrast CT-Based radiomics score for predicting hematoma enlargement in spontaneous intracerebral hemorrhage. Clin Neuroradiol. 2021;32(2):517–28. https://doi.org/10.1007/s00062-021-01062-w.
5. Yu F, Yang M, He C, Yang Y, Peng Y, Yang H, Lu H, Liu H. CT radiomics combined with clinical and radiological factors predict hematoma expansion in hypertensive intracerebral hemorrhage. Eur Radiol. 2024;35(1):6–19. https://doi.org/10.1007/s00330-024-10921-2.
6. Xu W, Ding Z, Shan Y, Chen W, Feng Z, Pang P, Shen Q. A nomogram model of radiomics and satellite sign number as imaging predictor for intracranial hematoma expansion. Front NeuroSci. 2020;14:491. https://doi.org/10.3389/fnins.2020.00491.
7. Jiang Y, Xu X, Wang R, Chen C. Efficacy of non-enhanced computer tomography-based radiomics for predicting hematoma expansion: A meta-analysis. Front Oncol. 2023;12:973104. https://doi.org/10.3389/fonc.2022.973104.
8. Dowlatshahi D, Demchuk A, Flaherty M, Ali M, Lyden P, Smith E. Defining hematoma expansion in intracerebral hemorrhage. Neurology. 2011;76(14):1238–44. https://doi.org/10.1212/wnl.0b013e3182143317.

9. Peng W, Li Q, Tang J, Reis C, Araujo C, Feng R, Yuan M, Jin L, Cheng Y, Jia Y, Luo Y, Zhang J, Yang J. The risk factors and prognosis of delayed perihematomal edema in patients with spontaneous intracerebral hemorrhage. CNS Neurosci Ther. 2019;25(10):1189–94. https://doi.org/10.1111/cns.13219.
10. Loan JJ, Gane AB, Middleton L, Sargent B, Moullaali TJ, Rodrigues MA, Cunningham L, Wardlaw J, Salman RA, Samarasekera N, Addison A, Ahmad K, Alhadad S, Andrews P, Bisset E, Bodkin P, Bouhaidar R, Brennan P, Campbell B, Broadbent S. Association of baseline hematoma and edema volumes with one-year outcome and long-term survival after spontaneous intracerebral hemorrhage: A community-based inception cohort study. Int J Stroke. 2020;16(7):828–39. https://doi.org/10.1177/1747493020974282.
11. Huang X, Wang D, Ma Y, Zhang Q, Ren J, Zhao H, Li S, Deng J, Yang J, Zhao Z, Xu M, Zhou Q, Zhou J. Perihematomal edema-based CT-radiomics model to predict functional outcome in patients with intracerebral hemorrhage. Diagn Interv Imaging. 2023;104(9):391–400. https://doi.org/10.1016/j.diii.2023.04.008.
12. Wang J, Xiong X, Zou J, Fu J, Yin Y, Ye J. Combination of hematoma volume and perihematoma radiomics analysis on baseline CT scan predicts the growth of perihematomal edema. Clin Neuroradiol. 2022;33(1):199–209. https://doi.org/10.1007/s00062-022-01201-x.
13. Feng C, Ding Z, Lao Q, Zhen T, Ruan M, Han J, He L, Shen Q. Prediction of early hematoma expansion of spontaneous intracerebral hemorrhage based on deep learning radiomics features of noncontrast computed tomography. Eur Radiol. 2023;34(5):2908–20. https://doi.org/10.1007/s00330-023-10410-y.
14. Zhao X, Zhou B, Luo Y, Chen L, Zhu L, Chang S, Fang X, Yao Z. CT-based deep learning model for predicting hospital discharge outcome in spontaneous intracerebral hemorrhage. Eur Radiol. 2023;34(7):4417–26. https://doi.org/10.1007/s00330-023-10505-6.
15. Zhi K, Wang Y, Yan L, Hou F, Wu J, Zhang S, Zhu H, Zhao L, Wang N, Zhao X, Li X, Wang Y, Chen C, Wang N, Xu Y, Yang G, Nie P. The interpretable CT-based vision transformer model for preoperative prediction of clear cell renal cell carcinoma SSIGN score and outcome. Insights Into Imaging. 2025;16(1):98. https://doi.org/10.1186/s13244-025-01972-0.
16. Li Y, Habib Daho E, Conze M, Zeghlache P-H, Le Boité R, Tadayoni H, Cochener R, Lamard B, Quellec M. G., LaTIM UMR 1101, Inserm, University of Western Brittany, IMT Atlantique, Sorbonne University, Ophthalmology Department, Lariboisière Hospital, AP-HP, Paris Cité University, & Ophthalmology Department, CHRU Brest. (2024). A review of deep learning-based information fusion techniques for multimodal medical image classification. In *Computers in Biology and Medicine* (Vol. 177, p. 108635). https://doi.org/10.1016/j.compbiomed.2024.108635
17. Farrar D, Fairley L, Santorelli G, Tuffnell D, Sheldon T, Wright J, Van Overveld L, Lawlor D. Bradford Institute for Health Research, Bradford Teaching Hospitals, Department of Health Sciences, Hull York Medical School, University of York, Clinical Trials Research Unit, University of Leeds, Bradford Women's and Newborn Unit, Radboud University, MRC Integrative Epidemiology Unit, & School of Social and Community Medicine. (2015). Association between hyperglycaemia and adverse perinatal outcomes in south Asian and white British women: analysis of data from the Born in Bradford cohort. Lancet Diabetes Endocrinol, *3*(October), 795–804. http://dx.doi.org/10.1016/S2213-8587(15)00255-7
18. Song Z, Guo D, Tang Z, Liu H, Li X, Luo S, Yao X, Song W, Song J, Zhou Z. Noncontrast Computed Tomography-Based Radiomics Analysis in Discriminating Early Hematoma Expansion after Spontaneous Intracerebral Hemorrhage. Korean J Radiol. 2020;22(3):415. https://doi.org/10.3348/kjr.2020.0254.
19. Xia X, Zhang X, Cui J, Jiang Q, Guan S, Liang K, Wang H, Wang C, Huang C, Dong H, Han K, Meng X. Difference of mean Hounsfield units (dHU) between follow-up and initial noncontrast CT scan predicts 90-day poor outcome in spontaneous supratentorial acute intracerebral hemorrhage with deep convolutional neural networks. NeuroImage Clin. 2023;38:103378. https://doi.org/10.1016/j.nicl.2023.103378.
20. Li Y, Habib Daho E, Conze M, Zeghlache P-H, Le Boité R, Tadayoni H, Cochener R, Lamard B, Quellec M. G., LaTIM UMR 1101, Inserm, University of Western Brittany, IMT Atlantique, Sorbonne University, Ophthalmology Department, Lariboisière Hospital, AP-HP, Paris Cité University, & Ophthalmology Department, CHRU Brest. (2024). A review of deep

learning-based information fusion techniques for multimodal medical image classification. In *Computers in Biology and Medicine* (Vol. 177, p. 108635). https://doi.org/10.1016/j.compbiomed.2024.108635

21. Jiang Y, Xu X, Wang R, Chen C. Efficacy of non-enhanced computer tomography-based radiomics for predicting hematoma expansion: A meta-analysis. Front Oncol. 2023;12:973104. https://doi.org/10.3389/fonc.2022.973104.
22. Fan W, Wu Z, Zhao W, Jia L, Li S, Wei W, Chen X. The performance of artificial intelligence in image-based prediction of hematoma enlargement: a systematic review and meta-analysis. Ann Med. 2025;57(1):2515473. https://doi.org/10.1080/07853890.2025.2515473.
23. Li N, Ding S, Liu Z, Ye W, Liu P, Jing J, Jiang Y, Zhao X, Liu T. A Deep Learning-Based framework for predicting intracerebral hematoma expansion using head non-contrast CT scan. Acad Radiol. 2024;32(1):347–58. https://doi.org/10.1016/j.acra.2024.07.039.
24. Lee H, Lee J, Jang J, Hwang I, Choi KS, Park JH, Chung JW, Choi SH. Predicting hematoma expansion in acute spontaneous intracerebral hemorrhage: integrating clinical factors with a multitask deep learning model for non-contrast head CT. Neuroradiology. 2024;66(4):577–87. https://doi.org/10.1007/s00234-024-03298-y.
25. Song L, Li M, Lin Y, Li X, An P. Advancing multimodal predictive models for targeted therapy response in EGFR-mutant lung adenocarcinoma: opportunities and challenges. J Translational Med. 2025;23(1):589. https://doi.org/10.1186/s12967-025-06601-4.
26. Chen Y, Shao Z, Bian H, Fang Z, Wang Y, Cai Y, Wang H, Liu G, Li X, Zhang Y. DMIL-Transformer: multiple instance learning via integrating morphological and spatial information for lymph node metastasis classification. IEEE J Biomedical Health Inf. 2023;27(9):4433–43. https://doi.org/10.1109/jbhi.2023.3285275.
27. Feng C, Ding Z, Lao Q, Zhen T, Ruan M, Han J, He L, Shen Q. Prediction of early hematoma expansion of spontaneous intracerebral hemorrhage based on deep learning radiomics features of noncontrast computed tomography. Eur Radiol. 2023;34(5):2908–20. https://doi.org/10.1007/s00330-023-10410-y.
28. Yang T, Zhao Z, Gu Y, Yang S, Zhang Y, Li L, Wang T, Miao Z. Evaluating the value of machine learning models for predicting hematoma expansion in acute spontaneous intracerebral hemorrhage based on CT imaging features of hematomas and surrounding oedema. Front Neurol. 2025;16:1567525. https://doi.org/10.3389/fneur.2025.1567525.
29. Chen Y, Chen S, Chang J, Wei J, Feng M, Wang R. Perihematomal edema after Intracerebral hemorrhage: An update on pathogenesis, risk factors, and therapeutic advances. Front Immunol. 2021;12:740632. https://doi.org/10.3389/fimmu.2021.740632.
30. Dierksen F, Sommer JK, Tran AT, Lin H, Haider SP, Maier IL, Aneja S, Sanelli PC, Malhotra A, Qureshi AI, Claassen J, Park S, Murthy SB, Falcone GJ, Sheth KN, Payabvash S. 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(24):2827. https://doi.org/10.3390/diagnostics14242827.
31. Zhang Y, Gao X, Zhou M, Hu T, Ta N, Li D, Wang D, Qu X, Wang H. Transformer-based multimodal fusion model predicts early hematoma expansion in spontaneous cerebral hemorrhage: A multicenter study. Eur J Radiol. 2025;194:112465. https://doi.org/10.1016/j.ejrad.2025.112465.

Figures

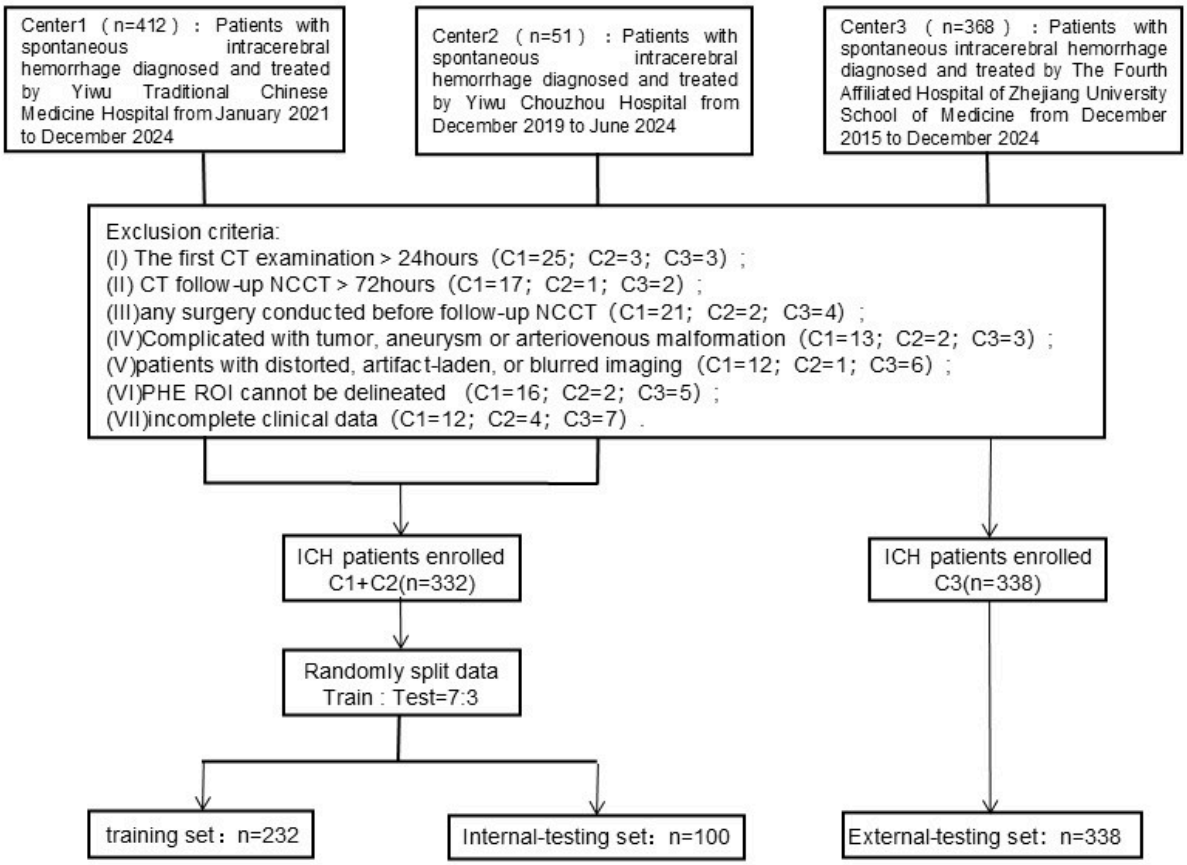


Figure. 1 Flow chart for enrolling the study population

Figure 1

See image above for figure legend.

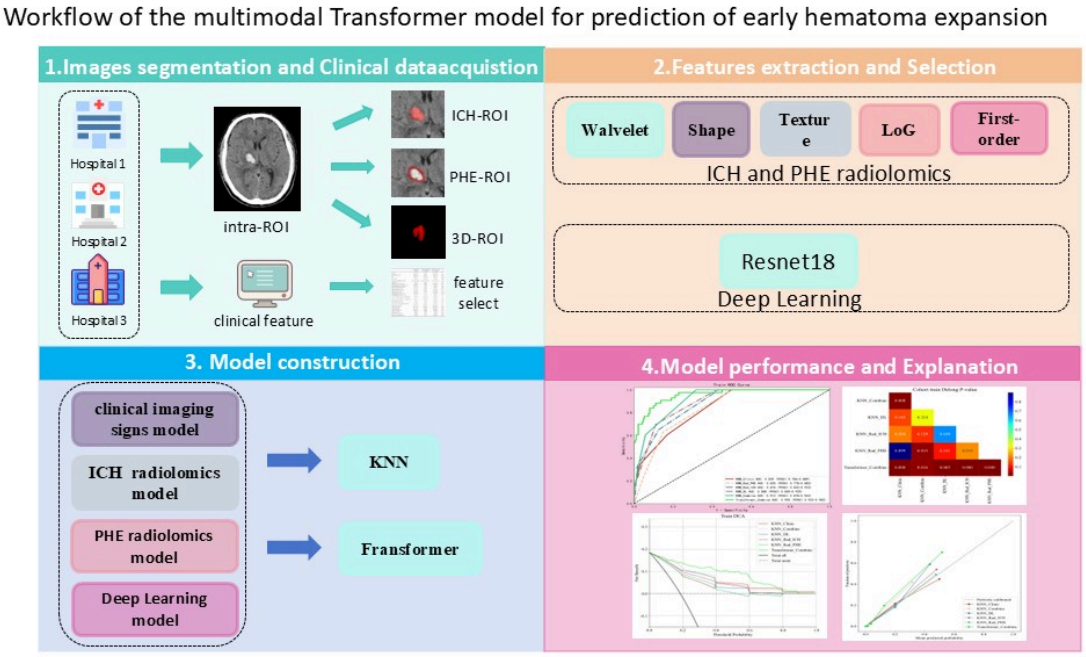


Figure. 2 Workflow of the multimodal Transformer model for prediction of early hematoma expansion. Diagram shows the four main processing steps:1. Images segmentation and Clinical dataacquistion; 2.Features extraction and Selection;3.Model construction;4.Model performance and Explanation.

Figure 2

See image above for figure legend.

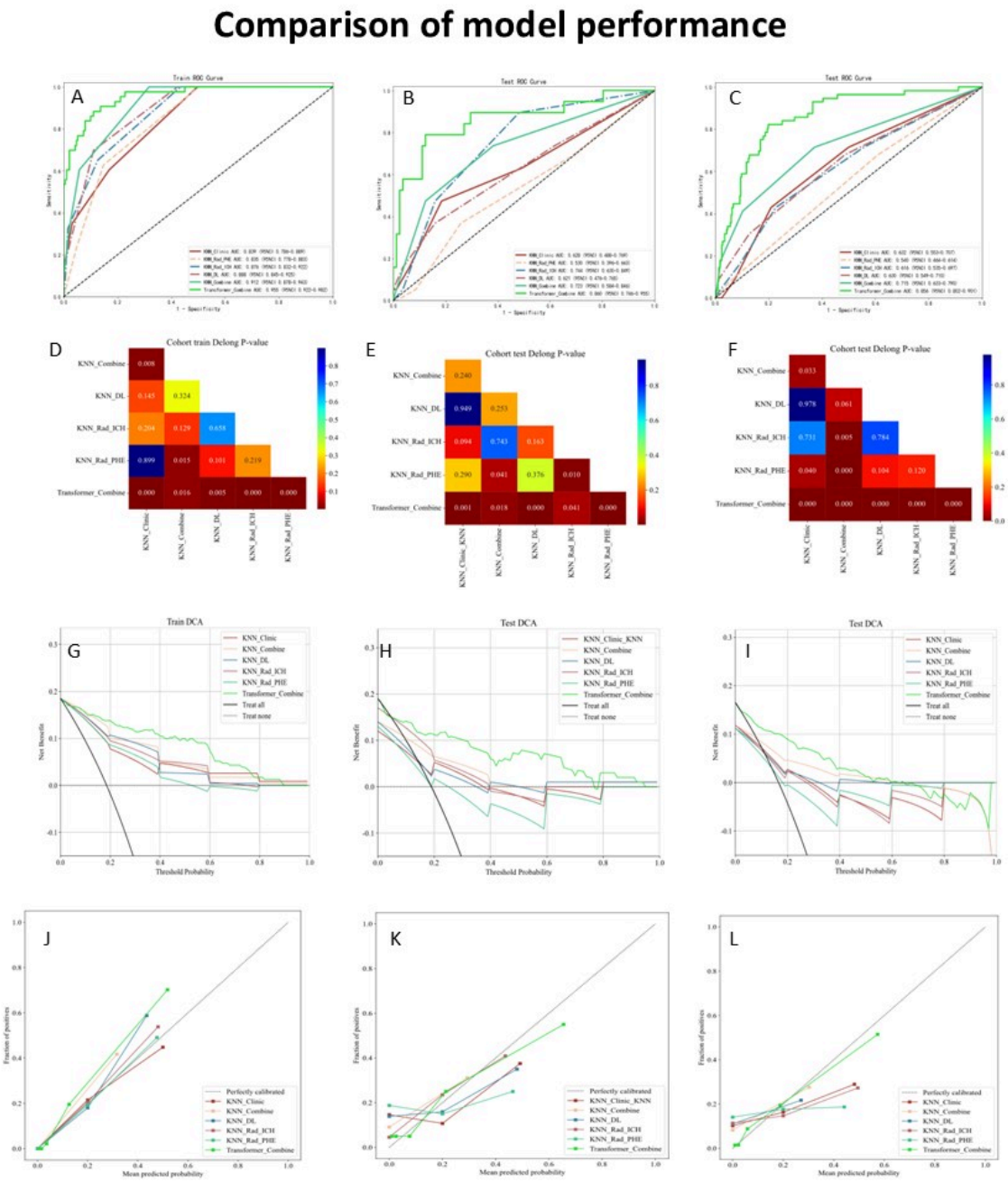


Figure. 3. Comparative analysis in the ROC curves (A-C) of the clinic model, hematoma radiomics, perihematoma edema radiomics,deep learning, KNN combine and transformer models. DeLong test for comparing the statistical significance of the AUC differences between models (D-F) in the DeLong test for six model among different subgroups. (G-I) DCA decision curve analysis to assess the comparative clinical utility of the six models across subgroups. (J-L) Calibration curve analysis to analyze observations in different cohorts to visualize predictive accuracy.

Figure 3

See image above for figure legend.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.docx