

Explainable deep learning algorithm for identifying cerebral venous sinus thrombosis-related hemorrhage (CVST-ICH) from spontaneous intracerebral hemorrhage using computed tomography



Kai-Cheng Yang,^{a,k} Yunzhi Xu,^{b,k} Qing Lin,^c Li-Li Tang,^a Jia-wei Zhong,^d Hong-Na An,^e Yan-Qin Zeng,^f Ke Jia,^{g,h,i} Yujia Jin,^a Guoshen Yu,^j Feng Gao,^{a,***} Li Zhao,^{b,**} and Lu-Sha Tong^{a,*}



^aNeurology Department, The Second Affiliated Hospital of Zhejiang University, School of Medicine, Hangzhou, Zhejiang Province, China

^bKey Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering & Instrument Science of Zhejiang University, Hangzhou, Zhejiang Province, China

^cNeurology Department, The First People's Hospital of Taizhou City, Taizhou, Zhejiang Province, China

^dNeurology Department, Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang Province, China

^eNeurology Department, The Second People's Hospital of Quzhou, Quzhou, Zhejiang Province, China

^fNeurology Department, Longyan First Hospital Affiliated to Fujian Medical University, Longyan, Fujian Province, China

^gAffiliated Mental Health Center & Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, China

^hLiangzhu Laboratory, MOE Frontier Science Center for Brain Science and Brain-machine Integration, State Key Laboratory of Brain-machine Intelligence, Zhejiang University, Hangzhou, Zhejiang Province, China

ⁱNHC and CAMS Key Laboratory of Medical Neurobiology, Zhejiang University, Hangzhou, Zhejiang Province, China

^jNeurology Department, Haiyan People's Hospital, Jiaxing, Zhejiang Province, China

Summary

Background Misdiagnosis of hemorrhage secondary to cerebral venous sinus thrombosis (CVST-ICH) as arterial-origin spontaneous intracerebral hemorrhage (sICH) can lead to inappropriate treatment and the potential for severe adverse outcomes. The current practice for identifying CVST-ICH involves venography, which, despite being increasingly utilized in many centers, is not typically used as the initial imaging modality for ICH patients. The study aimed to develop an explainable deep learning model to quickly identify ICH caused by CVST based on non-contrast computed tomography (NCCT).

Methods The study population included patients diagnosed with CVST-ICH and other spontaneous ICH from January 2016 to March 2023 at the Second Affiliated Hospital of Zhejiang University, Taizhou First People's Hospital, Taizhou Hospital, Quzhou Second People's Hospital, and Longyan First People's Hospital. A transfer learning-based 3D U-Net with segmentation and classification was proposed and developed only on admission plain CT. Model performance was assessed using the area under the curve (AUC), sensitivity, and specificity metrics. For further evaluation, the average diagnostic performance of nine doctors on plain CT was compared with model assistance. Interpretability methods, including Grad-CAM++, SHAP, IG, and occlusion, were employed to understand the model's attention.

Findings An internal dataset was constructed using propensity score matching based on age, initially including 102 CVST-ICH patients (median age: 44 [29, 61] years) and 683 sICH patients (median age: 65 [52, 73] years). After matching, 102 CVST-ICH patients and 306 sICH patients (median age: 50 [40, 62] years) were selected. An external dataset consisted of 38 CVST-ICH and 119 sICH patients from four other hospitals. Validation showed AUC 0.94, sensitivity 0.96, and specificity 0.8 for the internal testing subset; AUC 0.85, sensitivity 0.87, and specificity 0.82 for the external dataset, respectively. The discrimination performance of nine doctors interpreting

eClinicalMedicine
2025;81: 103128

Published Online xxx
<https://doi.org/10.1016/j.eclinm.2025.103128>

*Corresponding author. Neurology Department, The Second Affiliated Hospital of Zhejiang University, School of Medicine, #88 Jiefang Road, Hangzhou, 310009, China.

**Corresponding author. Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering & Instrument Science of Zhejiang University, #38 Zheda Road, Hangzhou, 310007, China.

***Corresponding author. Neurology Department, The Second Affiliated Hospital of Zhejiang University, School of Medicine, #88 Jiefang Road, Hangzhou, 310009, China.

E-mail addresses: 2310040@zju.edu.cn (L.-S. Tong), lz5bf@virginia.edu (L. Zhao), 2202012@zju.edu.cn (F. Gao).

^kContributed equally to this work.

CT images significantly improved with the assistance of the proposed model (accuracy 0.79 vs 0.71, sensitivity 0.88 vs 0.81, specificity 0.75 vs 0.68, $p < 0.05$). Interpretability methods highlighted the attention of model to the features of hemorrhage edge appearance.

Interpretation The present model demonstrated high-performing and robust results on discrimination between CVST-ICH and spontaneous ICH, and aided doctors' diagnosis in clinical practice as well. Prospective validation with larger-sample size is required.

Funding The work was funded by the National Key R&D Program of China (2023YFE0118900), National Natural Science Foundation of China (No.81971155 and No.81471168), the Science and Technology Department of Zhejiang Province (LGJ22H180004), Medical and Health Science and Technology Project of Zhejiang Province (No.2022KY174), the 'Pioneer' R&D Program of Zhejiang (No. 2024C03006 and No. 2023C03026) and the MOE Frontier Science Center for Brain Science & Brain-Machine Integration, Zhejiang University.

Copyright © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Cerebral venous sinus thrombosis; Spontaneous intracerebral hemorrhage; Deep learning; Explainable AI

Research in context

Evidence before this study

Delayed diagnosis and misdiagnosis of cerebral venous sinus thrombosis-related intracerebral hemorrhage (CVST-ICH) are common, because venography is not routinely used as the initial imaging modality for ICH patients in emergency settings. A systematic literature search was conducted in PubMed using the query ("Cerebral Venous Thrombosis" [Mesh] OR "CVST" OR "venous infarction" OR "venous thrombosis") AND ("Intracerebral Hemorrhage" [Mesh] OR "ICH" OR "brain hemorrhage") AND ("Deep Learning" OR "Artificial Intelligence" OR "Machine Learning"), identifying studies that applied deep learning for ICH classification. However, no prior studies have specifically developed an AI-based model to differentiate CVST-ICH from sICH using either non-contrast CT (NCCT) or magnetic resonance image.

Added value of this study

This multi-center study developed and validated an explainable deep learning model for differentiating CVST-ICH

from spontaneous ICH using only non-contrast CT. The model achieved high diagnostic performance on independent testing and external validation sets, significantly outperforming the accuracy of unaided doctors. Furthermore, interpretability techniques such as Grad-CAM++, SHAP, and IG provided visual insights into the model's attention, particularly regarding hematoma margins and shape features.

Implications of all the available evidence

Although guidelines emphasize the importance of early identification of intracerebral hemorrhage caused by CVST, there is still a considerable risk of misdiagnosis and delayed diagnosis. The proposed deep learning model that relies solely on non-contrast CT scans has the potential to assist doctors identifying CVST-ICH from spontaneous ICH, enabling timely further confirmation and reducing misdiagnosis. However, further validation, especially larger-scale prospective trials, is necessary before its application into routine clinical practice.

Introduction

Intracerebral hemorrhage (ICH) is a common and life-threatening condition in emergency departments.¹ Unlike other etiologies such as aneurysm rupture, arteriovenous malformation and trauma, intracerebral hemorrhage secondary to cerebral venous sinus thrombosis is relatively rare and hard to recognize yet requires high caution to identify.² Previous studies have reported cases where misdiagnosis of CVST-ICH led to mismanagement and adverse outcomes, such as progression of thrombosis and edema, and decreased consciousness.³ Confirmative venography imaging, such as computed tomography venography (CTV) or contrast-enhanced magnetic resonance venography (CE-MRV), is not only challenging doctors' capacity, but also time-

consuming. Median diagnostic delays for CVST have been reported to be about seven days.⁴ Guidelines also highlight CVST presenting as hemorrhage as a major risk of misdiagnosis or delay in diagnosis. Recognizing these patients is crucial, as the hemorrhagic mechanisms in CVST-ICH differ from other ICH types, directly impacting treatment approaches.²

The spontaneous intracerebral hemorrhage with a mortality rate exceeding 30% in the first month, typically results from small vessel disease caused by hypertensive arteriopathy and cerebral amyloid angiopathy.⁵ Early blood pressure control and minimally-invasive surgical hematoma clearance are currently recommended to achieve hemostasis. In contrast, CVST-ICH is attributed to thrombosis and congestion in cerebral venous

system. Therefore, anticoagulation is required immediately while hemostatic therapy can be fatal.⁶

Nearly half of CVST patients experience intracerebral hemorrhage,⁷ indicating severe venous congestion, resulting in a mortality of approximately 40% within the first month, even under intensive care.⁸ Therefore, a quick identification between the two hemorrhagic disorders is urgently needed in clinical settings.

Non-contrast computed tomography is always the initial imaging modality as patients with intracerebral hemorrhage presenting in emergency room.¹ If secondary hemorrhage is suspected, subsequent standard care procedures involve the selection of further CT angiography (CTA)/CTV, MR angiography (MRA)/MRV, or digital subtraction angiography (DSA), based on initial CT and clinical information, with prioritization until a diagnosis is made.⁹ For CVST-ICH, while non-contrast CT is essential for initial assessment, its sensitivity for CVST is only 0.48–0.73, with a specificity of 0.50–0.98.^{10,11} CE-MRV offers improved sensitivity (0.83) and specificity (0.99) but is only applied if necessary and usually inaccessible in urgent settings.¹² Improving CT's ability to detect CVST is crucial for guiding further diagnostic steps such as CTV or CE-MRV.

Recently, deep learning models have been employed in medical emergency scenarios successfully. Previous studies employed CT scans to develop deep learning models for discriminating common etiologies of cranial hemorrhage and generated expert-level results, but CVST-ICH was omitted from the result categories.^{13,14} In this context, the current AI models have not addressed the clinically urgent demand to discriminate different etiologic categories requiring completely contrary therapies. Additionally, in real-world settings, inexperienced doctors and chaos in the emergency department can lead to a higher misdiagnosis risk compared with other routine scenarios.

Therefore, to meet the highly clinical-related demand, we aim to develop a rapid and accurate explainable deep learning model to automatically discriminate CVST-ICH from spontaneous intracerebral hemorrhage. What's more, its performance will be assessed and compared with that of doctors. Then four visualizable interpreting methods are used to improve understandability of the algorithm's judgment, and the doctors' diagnostic accuracy will be tested again with model's assistance.

Methods

Data source

We retrospectively collected initial non-contrast CT images of CVST-ICH patients on arrival to the emergency department in Second Affiliated Hospital of Zhejiang University between January 2016 and March 2023. Diagnoses were made by radiological reports from the

further confirmative imaging modalities, including CTV, CE-MRV, or DSA, and re-checked by two experienced neurologists (Yang K.C. and Tong L.S.). Since CVST-ICH mostly presented intracerebral hemorrhage located in brain lobes,¹⁵ we selected spontaneous intracerebral hemorrhage patients who also presented with lobar hemorrhage and were admitted during the corresponding period as the identification target. For sICH group, all patients had at least one vascular imaging during hospitalization, including CT/MR angiography, CT/MR venogram, or DSA. We systematically re-evaluated imaging and clinical data as follows: (1) For patients with available MRV, CTV, or DSA, CVST was excluded based on these tests. (2) For patients without MRV, CTV, or DSA, CVST was excluded by reviewing CTA venous phase, MRA, and clinical features (e.g., continuous headache, pregnancy or puerperium, or high D-dimer levels). (3) For patients with clinical features highly suggestive of CVST but with vascular imaging showing no significant abnormalities, follow-up was conducted to review subsequent imaging and clinical records to confirm the exclusion of venous origin ICH. Only patients who had supratentorial hemorrhage were analyzed.

Additionally, to evaluate the model's robustness, we compiled an external dataset of CVST-ICH and spontaneous intracerebral hemorrhage patients from January 2016 to March 2023 from four additional municipal or county hospitals: Quzhou Second People's Hospital, Taizhou First People's Hospital, Taizhou Hospital, and Longyan People's Hospital.

To address potential imaging bias, we applied propensity score matching (PSM) to balance age distributions and mitigate class imbalance between CVST-ICH and sICH groups. Given that CVST-ICH patients were significantly younger than sICH patients in both previous studies and our dataset,¹⁶ age-related imaging characteristics—such as brain atrophy, ventricular enlargement, and parenchymal low-density changes—could influence model performance. Additionally, the original dataset had a notable class imbalance, with 102 CVST-ICH cases compared to 683 sICH cases. PSM helped minimize the influence of class imbalance on model training.¹⁷ Using the MatchIt package in R (version 4.3.4), we applied a 1:3 nearest-neighbor matching without replacement, without setting a caliper, and estimated propensity scores via logistic regression, with age as the only matching variable. Matching was restricted to the internal dataset. The matched data was then randomly split into a training set (75%) and an internal testing set (25%) used for model development and internal testing, respectively. The external dataset remained unmatched to preserve its original distribution for evaluating the model's generalizability. Details are provided in the [Figure S1](#). The main process of the study was shown in [Fig. 1](#).

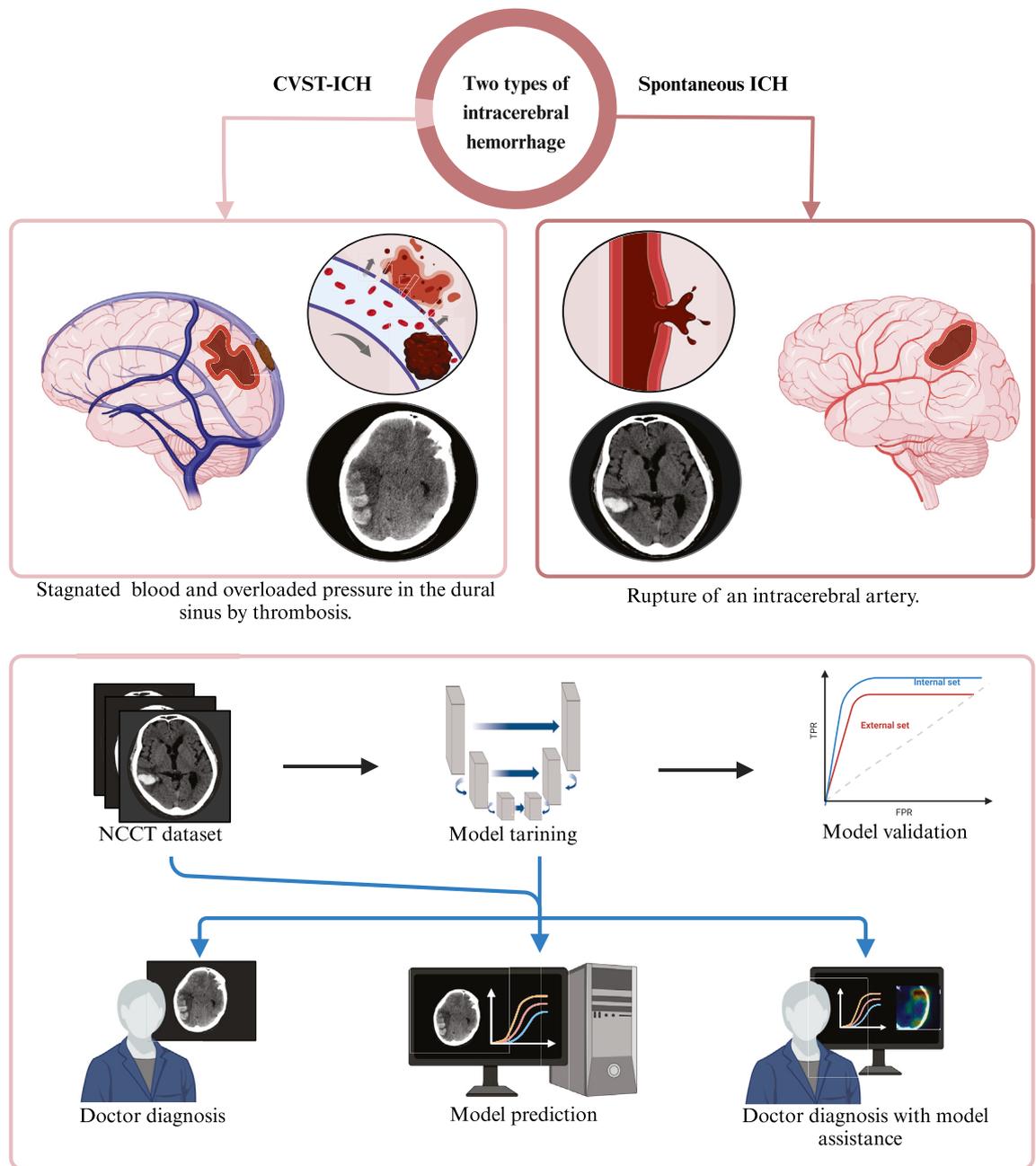


Fig. 1: Distinct pathogenic mechanisms of intracerebral hemorrhage and the main process of the study.

Ethics

The study conforms to the principles outlined in the Helsinki Declaration. Ethical approval was obtained by the Human Ethics Committee of the Second Affiliated Hospital of Zhejiang University (2023LSYD0437), the Medical Ethics Committees of the First People’s Hospital of Taizhou (2023-KY089-01), the First Hospital of Longyan (LYREC2023-k099-01), Taizhou Hospital (K20231227), and the Second People’s Hospital of

Quzhou (2023-ER-61). Informed consent was exempted due to the retrospective nature.

Non-contrast CT image acquisition, segmentation and labeling

Non-contrast CT images in Digital Imaging and Communications in Medicine (DICOM) format were extracted from Picture Archiving and Communication System (PACS) and converted into Neuroimaging

Informatics Technology Initiative (NIFTI) format using the open-source Python package SimpleITK (version 2.1.1) for further analysis. The number of slices varied from 24 to 36, with slice thickness ranging from 4.0 to 6.0 mm. Detailed image parameters are described in [Table S1](#).

Two neurologists, one with 6 years of experience (Yang K.C.) and one with 20 years of experience (Tong L.S.) manually segmented and adjusted the details of the hematoma area on the CT images using ITK-SNAP software (version 3.8.0). These manual segmentations were used as the ground truth for developing model. CVST and spontaneous intracerebral hemorrhage were labeled as “1” and “0”, respectively.

Image preprocessing

To enhance model performance and efficiency, all CT images were preprocessed. The images were resampled using bilinear interpolation and masked segmentally to a consistent voxel spacing of 1 mm, resulting in a matrix size of $140 \times 250 \times 250$. Resampling to isotropic resolution of 1 mm aids the 3D CNN model in accurately understanding spatial information, harmonizing data from different sites.¹⁸ A window of 0–200 Hounsfield Units (HU) was selected. Relevant studies demonstrate that CT images within the 0–200 window contain abundant brain information, aiding in the detection of potential subarachnoid and subdural hemorrhages, and enhancing the model generalization.^{14,19} Data augmentation techniques such as rotation, shift, and flip, were employed to expand training data and mitigate model overfitting. Differences in data across various sites and scanners arise from variations in in-plane resolution, slice thickness, and brain position. Consistency in resolution was ensured through resampling. Image matrices of consistent size, centered on the brain, were obtained to ensure accurate brain localization. These steps ensure the consistency of data across different sites.

Architecture of the deep learning model

The foundational model was an improved version of the three-dimensional U-Net architecture (3D U-Net) with the Parametric Rectified Linear Unit (PReLU), which served as the activation function to augment feature learning.²⁰ Each convolutional block comprised two convolutional layers with kernel size of $3 \times 3 \times 3$. The dilation convolution with stride of two was used as the second convolutional layer. The concatenation was used instead of residual connections in each convolutional block to increase model parameter count. The encoder integrated the max-pool layer for down-sampling, whereas the decoder employed 3D transposed convolution to restore resolution. Dropout was applied at the bottleneck of the encoding-decoding structure to mitigate model overfitting. The depth of the model was four, with the initial number of convolutional kernels of 32.

The complete model consists two networks: hematoma segmentation and hemorrhagic cause classification. The classification network was further trained based on the pre-trained part of hematoma segmentation network, regarding enhancement of the capability of encoder's feature extraction by segmentation proxy task. Specifically, based on the preprocessed non-contrast CT images, hematoma segmentation network was trained with the manually segmented hematoma mask as the ground truth. Then, the encoder part of the segmentation network was extracted as pre-trained weights for the classification network, and a fully connected layer was added with the sigmoid function as the activation for binary classification. The Convolutional Block Attention Module (CBAM) was incorporated into the segmentation and classification network, in order to further guide the model's attention to hemorrhage regions.²¹ The CBAM is composed of the Spatial Attention Module (SAM) and Channel Attention Module (CAM). In CAM, global average pool and global maximum pool were used to aggregate spatial information, and the resultant feature maps undergo computation in a shared MLP, where the channel weights are obtained using the softmax function. SAM receives the input from CAM, computing global maximum and global average in the channel dimension to yield the single-layer feature map. The outputs from both channels are merged and fed into a convolutional layer with a $7 \times 7 \times 7$ kernel. Finally, the sigmoid function is applied to determine the weight of each pixel. The CBAM module is introduced after the second convolutional layer in each convolutional block. Pre-processed non-contrast CT images and corresponding hematoma masks were used as dual-channel inputs to train the classification network. [Fig. 2](#) shows the model structure.

Interpretability of the model

Despite the high accuracy of deep learning models in diagnosing and predicting diseases, they often fail to offer comprehensible explanations and undermine the clinicians' trust.²² To address this, we employed four widely-used visualization methods to highlight the regions significantly influencing the model's outputs. Grad-CAM++ was used to generate heatmaps that portray areas where the model concentrated.²³ It is an improved version of Grad-CAM that backpropagates only through the positive gradients of the predicted class,²⁴ thereby accentuating pixels positively associated with the class.

Another gradient-based attribution method called Integrated Gradients (IG) was also used, which overcomes gradient saturation by adhering to the axioms of sensitivity and implementation invariance.²⁵ Using a black image as the baseline, a gradient linear path from the baseline image to input image was created. The gradient integral of classification results related to input

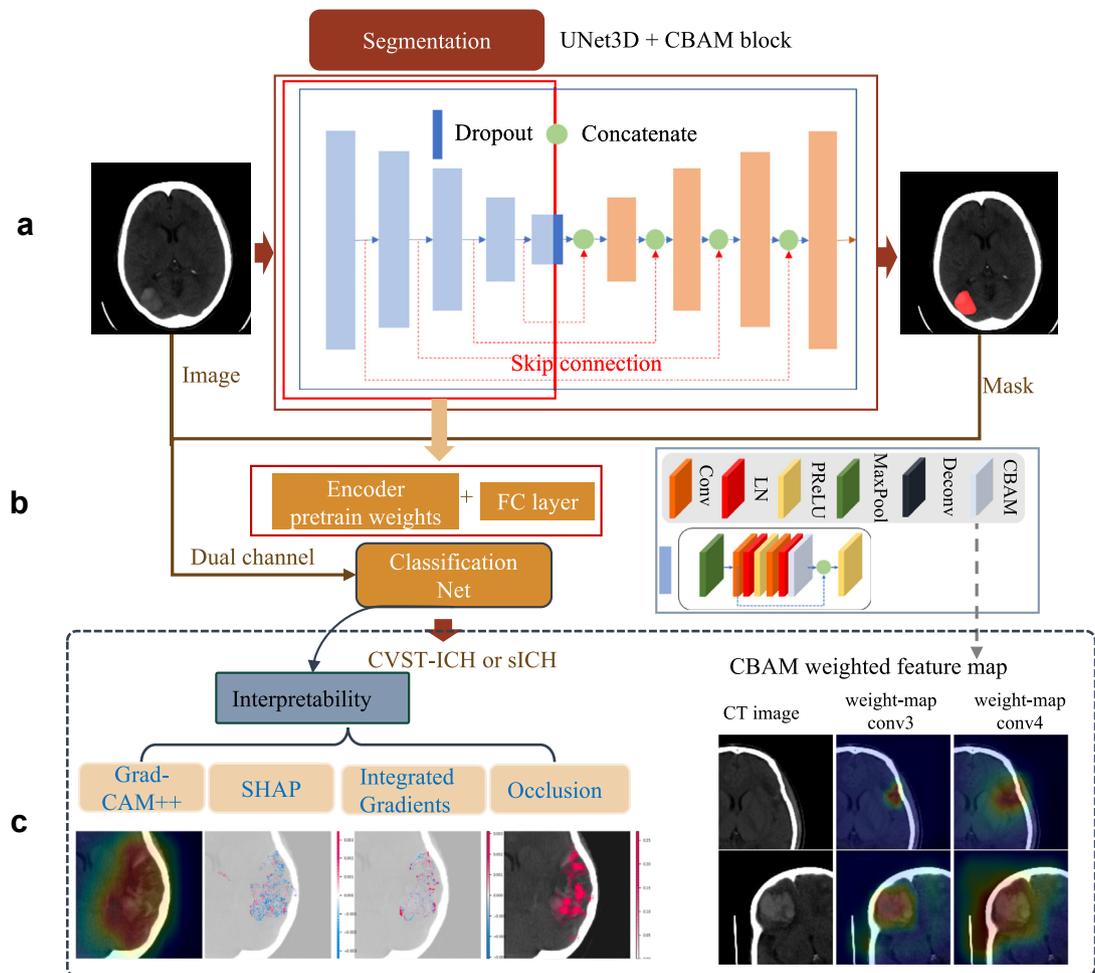


Fig. 2: The framework and process of the proposed method. (a) The hematoma segmentation task based on 3D U-Net. (b) The classification model. An encoder structure derived from the segmentation model is utilized, and pre-training weights from the segmentation task are loaded to improve the classification model's feature extraction capability. The image and segmentation mask form a dual channel as input. (c) The interpretability researches. The Grad-CAM++, SHAP, Integrated gradient, and Occlusion were employed to elucidate variations in the attention levels of classification models across different regions. Weighted attention maps were computed from the CBAM module of the third and fourth convolutional blocks. CVST-ICH, hemorrhage secondary to cerebral venous sinus thrombosis; sICH, spontaneous intracerebral hemorrhage; UNet3D, 3-dimensional U-Net; CBAM, convolutional block attention module; SHAP, shapley additive exPlanations; Grad-CAM, gradient-weighted class activation mapping; FC layer, fully connected layer; Conv, convolution; LN, layernorm; PReLU, parametric rectified linear unit; and Deconv, deconvolution.

image pixels was calculated, and each pixel's contribution was quantified.

Shapley Additive exPlanations (SHAP), grounded in coalitional game theory,²⁶ was used to further visualize the model's interested areas across different outcomes. This involves calculating the Shapley value for each pixel based on its marginal contribution to the classification. An interpretive map was then produced, where pixels were color-coded according to their contributions: positive contributions in red and negative contributions in blue.

Another perturbation-based attribution method, called occlusion method, was implemented by sliding an

occlusion slider across different image regions to assess their influence on the model's classification decisions.

These interpretability maps were overlaid onto the original non-contrast CT images to assist the diagnosis of clinicians and enhance intuitive understanding of model's decision-making process.

Training procedure of the model

The proposed model employed patch-based training. Segmentation and classification models used the $128 \times 128 \times 128$ patch as input, which completely contained the region of interest (ROI) of hematoma area. In the inference stage, the segmentation model was used to

segment hematoma areas from the complete NCCT image (with a resolution of isotropic 1 mm). The ROI region was obtained by bounding, using the segmented hematoma region as the ROI boundary. Then, the $128 \times 128 \times 128$ patch matrix was taken from the complete non-contrast CT image and additionally segmented with mask around the ROI region. Finally, the patch image and mask were input into the classification model to generate the results. Since the accuracy of hematoma segmentation can largely impact classification performance, the pre-trained segmentation model underwent an additional 30 epochs of training on the complete CT image to obtain the segmentation model for inference. To evaluate the potential impact of resampling resolution on model performance, we conducted additional experiments by resampling the Z-axis resolution to 5 mm while maintaining the in-plane resolution at 1 mm. This resulted in an adjusted matrix size of [32, 256, 256] for the model input, with a cropped region of interest at [32, 128, 128]. The input patch size for the 3D CNN was [32, 128, 128]. The same model architecture and training pipeline were used for both settings.

In the experiment, segmentation and classification models were trained using the training set (306 cases), as shown in Table 1. The entire training dataset was randomly divided into five groups (61 cases per group) for training (four groups) and validation (one group) using the five-fold cross-validation method, which was used to optimize hyperparameters, including the input ratio of positive and negative samples, learning rate, choice of optimizer, and choice of loss function. The model was then retrained from scratch using the entire training set and tested on both internal and external test sets, consisting of 102 and 157 samples, respectively. In the final model training, the learning rate was set to

$1e-4$, the dropout ratio of 0.5, and AdamW was chosen as the optimizer. The segmentation model used DICE-CE as loss function, while binary cross-entropy loss was applied to the classification model.²⁷ The models were implemented using the PyTorch 1.8.1 framework and trained on dual RTX3090 GPUs, with a batch size of two for segmentation tasks and eight for classification tasks. Segmentation models are trained for 100 epochs, while classification models are trained for 60 epochs.

Evaluation of the model

The characteristics of segmentation and classification were both evaluated.

The Dice coefficient was calculated against the manual segmentation mask on internal dataset.

The classification performance was tested both in internal and external datasets using area under the curve, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and Matthews Correlation Coefficient (MCC). The receiver operating characteristic (ROC) curve was obtained by plotting the true positive rate and the false positive rate at each discrimination threshold. The area under the ROC curve and its 95% confidence interval (CI) were computed. Two operational points on the ROC curve were selected for detailed evaluation: one for high sensitivity (specificity > 0.8) and another for high specificity (sensitivity > 0.8), and then accuracy, sensitivity, specificity, PPV, NPV, and F1 scores were calculated, respectively. The high sensitivity point was selected as the formal threshold for application, and it was employed to represent the target performance of the model. Tests at high specificity points more comprehensively reflected the model's performance range, particularly with small test sets. Under the threshold of high specificity, the model's performance was tested for

	Internal dataset			External dataset				
	CVST-ICH	sICH	Training set	Testing set	All	CVST-ICH	sICH	All
Cases	102	306	306 (76 CVST-ICH)	102 (26 CVST-ICH)	408	38	119	157
Male sex, n [%]	54 [52.9%]	203 [66.3%]	199 [65.0%]	58 [56.9%]	257 [62.9%]	19 [50.0%]	76 [63.8%]	95 [60.5%]
Age, years, median [IQR]	44 [29-61]	50 [40-62]	49 [36-57]	53 [39-66]	50 [37-62]	46 [33-59]	65 [54-74]	60 [49-72]
Number of axial slices, n	2978	9075	9001	3052	12,053	964	3371	4335
Slice thickness, mm, mean [SD]	5.01 [0.44]	5.02 [0.22]	5.04 [0.20]	5.00 [0.25]	5.02 [0.29]	5.78 [1.46]	5.27 [0.73]	5.39 [0.99]
Slices per case, mean [SD]	29.2 [3.48]	29.7 [2.69]	29.4 [2.32]	29.9 [2.78]	29.5 [2.91]	25.4 [5.73]	28.3 [4.55]	27.6 [5.02]
Time onset to imaging, days, median [IQR]	2 [1-4]	1 [1-2]	1 [1-3]	1 [1-3]	1 [1-3]	1 [1-5]	1 [1-2]	1 [1-3]
CT scanners	GE Optima CT540 SIEMENS SOMATOM Definition Flash SIEMENS SOMATOM Force SIEMENS SOMATOM Perspective			GE LightSpeed16 GE Discovery CT750 HD GE BrightSpeed SIEMENS SOMATOM Definition Edge SIEMENS SOMATOM Definition AS+ SIEMENS SOMATOM Definition Flash PHILIPS Brilliance 16CT CANON AquilionLightning TSX-035A				

IQR, interquartile range; SD, standard deviation; CVST-ICH, hemorrhage secondary to cerebral venous sinus thrombosis; sICH, spontaneous intracerebral hemorrhage.

Table 1: Dataset characteristics.

interference from specific data, revealing its lower performance limits. Additionally, the MCC values at the two operating points were assessed for the comprehensive evaluation of the classification model. The MCC is a correlation coefficient particularly adept for analyzing imbalanced classes, offering a comprehensive measure of classification effectiveness by accounting for all aspects of the confusion matrix.²⁸ The MCC value spans from -1 to $+1$, with $+1$ denoting perfect prediction accuracy, 0 signifying no better than random chance, and -1 reflecting total discordance between predictions and actual observations.

Reading test

Two neurologists (Lin Q. with 10 years, and Zhang Y.H. with two years of experience, respectively), two radiologists (Wang Q.Y. and Guan X.J. both with ten years of experience), and five emergency doctors (Lin R.X. with two years, Wang J.L. with three years, Cai Y.Y. with three years, Wang D.Q. with two years and Wu D.Q. with twelve years of experience, respectively) conducted the reading of non-contrast CT scans to classify CVST-ICH without the model assistance. Judgments were made based on their professional imaging expertise, including signs such as cord sign, high-density delta sign, or multiple flame-shaped hemorrhages indicative of CVST-ICH. Subsequently, the probability values derived from the high-sensitivity point and interpretable maps were provided as assistance to re-read the CT images. Cases were arranged in a random order, with all clinical information, radiology reports, and other imaging data concealed. Both the readers and the recorders were blinded to the classifications. The reading assessment was performed on the entire internal testing set.

Statistics

Continuous variables that adhered to a normal distribution were reported as mean \pm standard deviation (SD), while non-normally distributed data were presented as median and interquartile range (IQR). Categorical variables were summarized as percentages. The robustness of the metrics was evaluated using the bootstrapping method, using 1000 iterations with random sampling, and credit interval as defined as between 2.5th and 97.5th percentiles. During the reading test, readers and recorders were blinded to the ground truth. The average accuracy, sensitivity, and specificity of doctors, both with and without the assistance of the proposed model were calculated using Linear Mixed Model, as well as the mean differences.^{29,30} The mixed model included an indicator for model vs doctor, and for assistance vs no-assistance, as fixed effects, with image serial number as random effect. The coefficient for the fixed effect could be interpreted as the difference in accuracy, sensitivity, or specificity between the model and the doctor or between with and without assistance. Corrected t-tests were used for fixed effect estimates

based on the Satterthwaite approximation for denominator degrees of freedom. Two-tailed $p < 0.05$ was defined as statistically significant. Statistical analysis was performed using the lmerTest (version 3.1–3) and emmeans (version 1.10.0) packages within R software.

Role of the funding source

The funders of our study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. There was no commercial support.

Results

Datasets

Initial imaging and further clinical materials were collected in 102 patients with CVST-ICH (Age, 44 [29, 61] years) and 683 patients having spontaneous lobar intracerebral hemorrhage (Age, 65 [52, 73] years). Following propensity score matching, 102 patients in CVST-ICH and 306 patients in spontaneous lobar intracerebral hemorrhage (Age, 50 [40, 62] years) were selected to form the internal dataset.

The external dataset comprised 157 cases (Age, 60 [49, 72] years), including 38 patients with CVST-ICH (Age, 46 [33, 59] years) and 119 patients with spontaneous lobar intracerebral hemorrhage (Age, 65 [54, 74] years). The inclusion/exclusion flowchart of internal dataset is shown in Fig. 3, and patient demographics are presented in Table 1 and Table S2.

Model performance

Segmentation

The Dice coefficient of the segmentation network for hematoma segmentation was 0.84 ± 0.13 (95% CI [0.81, 0.87]) in the internal testing set.

Classification

Table 2 and Fig. 4 summarize the classification performance of the model.

On the internal testing set, the classification network achieved an AUC of 0.94 (95% CI [0.87, 0.98]) for hemorrhagic origin classification. At the high sensitivity operation point, the network achieved accuracy of 0.84 (95% CI [0.77, 0.91]), with sensitivity and specificity of 0.96 (95% CI [0.88, 1.00]) and 0.80 (95% CI [0.71, 0.89]), respectively. The corresponding F1 score and MCC value were 0.76 (95% CI [0.67, 0.85]) and 0.68 (95% CI [0.56, 0.81]). As to the high specificity point, the network exhibited a classification accuracy of 0.90 (95% CI [0.83, 0.95]), with specificity of 0.93 (95% CI [0.87, 0.99]) and sensitivity of 0.81 (95% CI [0.65, 0.96]). The F1 score was 0.81 (95% CI [0.68, 0.91]) and the MCC value was 0.74 (95% CI [0.59, 0.87]).

On the external dataset, the AUC of the classification network was 0.85 (95% CI [0.76, 0.92]). The accuracy, sensitivity, specificity, F1 score, and MCC of the model at the high sensitivity point were 0.83 (95% CI [0.77,

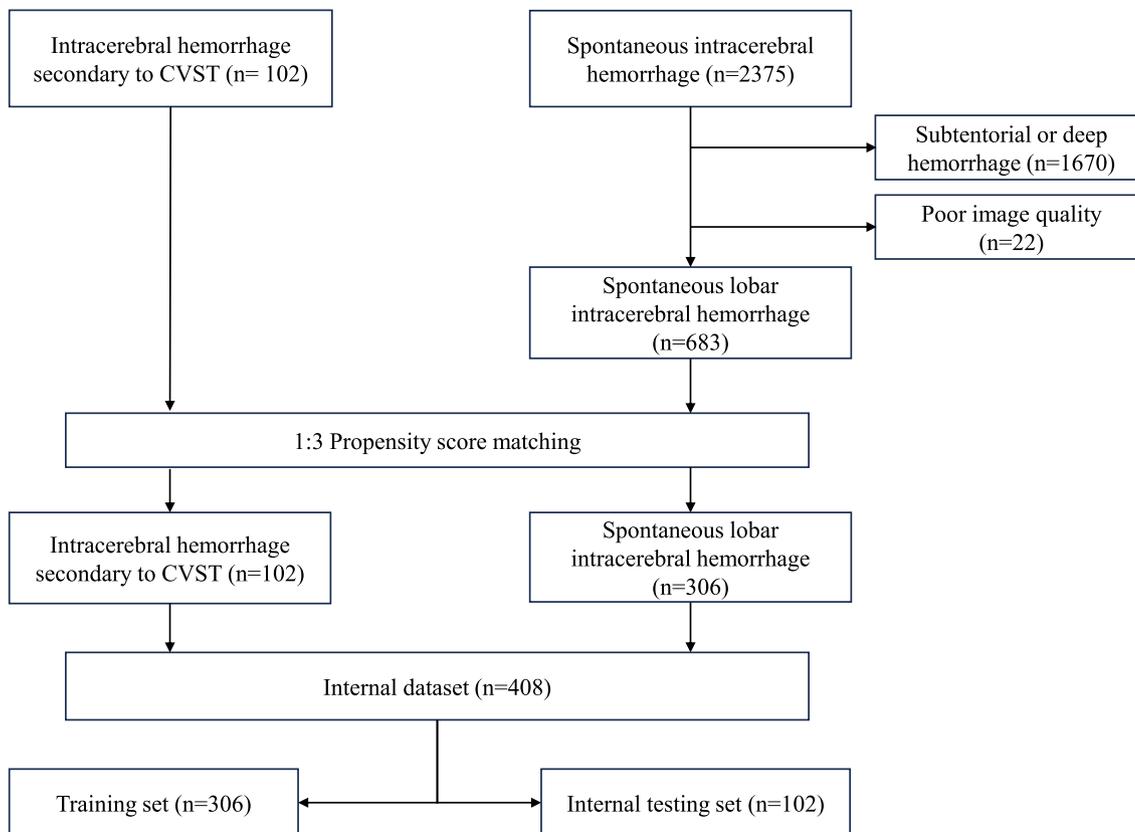


Fig. 3: Flow chart for patient extracting and matching for internal dataset.

0.89)), 0.87 (95% CI [0.76, 0.97]), 0.82 (95% CI [0.76, 0.89]), 0.72 (95% CI [0.63, 0.80]), and 0.62 (95% CI [0.50, 0.74]), respectively. At the high specificity operation point, the model accomplished 0.82 (95% CI [0.76, 0.88]) as accuracy, with sensitivity and specificity values of 0.82 (95% CI [0.68, 0.92]) and 0.82 (95% CI [0.76, 0.89]), resulting in an F1 score and MCC value of 0.69 (95% CI [0.59, 0.78]) and 0.58 (95% CI [0.45, 0.71]).

To demonstrate the stability of the classification algorithm, we summarized the cross-validation results in Table 2. The ROC curves for the five-fold cross-validation are shown in Figure S3, showing a mean AUC of 0.93 (95% CI [0.89, 0.97]). The mean sensitivity, specificity, MCC, and F1 score of the model were 0.86 (95% CI [0.82, 0.90]), 0.86 (95% CI [0.81, 0.90]), 0.75 (95% CI [0.70, 0.80]), and 0.66 (95% CI [0.59, 0.74]) at high sensitivity point, and 0.80 (95% CI [0.80, 0.80]), 0.92 (95% CI [0.86, 0.98]), 0.72 (95% CI [0.64, 0.81]), and 0.79 (95% CI [0.73, 0.85]) at high specificity point, respectively.

Reading performance

Table 3 and Table S4 summarized the diagnostic performance of doctors' manual reading. Before being aided by the current model, the mean accuracy,

sensitivity, and specificity of doctors were 0.71 (95% CI [0.67, 0.75]), 0.81 (95% CI [0.73, 0.89]), and 0.68 (95% CI [0.62, 0.73]), respectively. Compared with the model performance alone, the absolute differences in accuracy, sensitivity, and specificity were 0.13 (95% CI [0.05, 0.22], $p < 0.05$), 0.15 (95% CI [0.01, 0.29], $p < 0.05$), and 0.13 (95% CI [0.03, 0.23], $p < 0.05$).

With the model assistance, the doctors notably improved average accuracy from 0.71 to 0.79 (95% CI [0.74, 0.83]), with an absolute difference of 0.08 (95% CI [0.04, 0.11], $p < 0.05$), the sensitivity was also elevated from 0.81 to 0.87 (95% CI [0.80, 0.95]), with an absolute difference of 0.06 (95% CI [0.01, 0.12], $p < 0.05$), and the specificity was increased from 0.68 to 0.75 (95% CI [0.70, 0.81]), with a difference of 0.08 (95% CI [0.04, 0.12], $p < 0.05$).

Model interpretability

The interpretability maps generated by four visualization methods are presented in Fig. 5. The Grad-CAM++ heatmaps highlighted the model's attention on hematoma's borderline and perihematomal low-density areas, with the latter being particularly emphasized in CVST-ICH patients (Fig. 5a–d). In SHAP and IG maps, pixels with positive gradients were coded red, and those

	5-fold cross validation ^a		Internal testing data		External testing data	
	CVST-ICH	sICH	CVST-ICH	sICH	CVST-ICH	sICH
Cases	-	-	26	76	38	119
AUC [95% CI]	0.9342 [0.8949, 0.9736]		0.9352 [0.8674, 0.9818]		0.8476 [0.7629, 0.9228]	
High sensitivity operating point						
Predicted CVST-ICH	-	-	25	15	33	21
Accuracy [95% CI]	0.8567 [0.8213, 0.8920]		0.8431 [0.7745, 0.9118]		0.8344 [0.7707, 0.8917]	
Sensitivity [95% CI]	0.8533 [0.8096, 0.8971]		0.9615 [0.8846, 1.0000]		0.8684 [0.7632, 0.9737]	
Specificity [95% CI]	0.8578 [0.8158, 0.8997]		0.8026 [0.7105, 0.8947]		0.8235 [0.7563, 0.8908]	
PPV [95% CI]	0.6748 [0.6012, 0.7484]		0.6250 [0.5319, 0.7429]		0.6111 [0.5179, 0.7200]	
NPV [95% CI]	0.9461 [0.9296, 0.9626]		0.9839 [0.9500, 1.0000]		0.9515 [0.9118, 0.9896]	
F1 Score [95% CI]	0.7512 [0.6983, 0.8041]		0.7576 [0.6761, 0.8525]		0.7174 [0.6279, 0.8046]	
MCC [95% CI]	0.6641 [0.5905, 0.7378]		0.6821 [0.5633, 0.8093]		0.6239 [0.4971, 0.7427]	
High specificity operating point						
Predicted CVST-ICH	-	-	21	5	31	21
Accuracy [95% CI]	0.8900 [0.8481, 0.9319]		0.9020 [0.8333, 0.9510]		0.8217 [0.7580, 0.8790]	
Sensitivity [95% CI]	0.8000 [0.8000, 0.8000]		0.8077 [0.6538, 0.9615]		0.8158 [0.6842, 0.9211]	
Specificity [95% CI]	0.9200 [0.8641, 0.9759]		0.9342 [0.8684, 0.9868]		0.8235 [0.7563, 0.8908]	
PPV [95% CI]	0.7917 [0.6836, 0.8999]		0.8077 [0.6667, 0.9524]		0.5962 [0.5000, 0.7083]	
NPV [95% CI]	0.9321 [0.9280, 0.9363]		0.9342 [0.8846, 0.9857]		0.9333 [0.8909, 0.9709]	
F1 Score [95% CI]	0.7906 [0.7315, 0.8498]		0.8077 [0.6818, 0.9057]		0.6889 [0.5934, 0.7805]	
MCC [95% CI]	0.7211 [0.6366, 0.8056]		0.7419 [0.5870, 0.8728]		0.5818 [0.4457, 0.7089]	

AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; MCC, Matthews correlation coefficient; CI, confidence interval; CVST-ICH, intracerebral hemorrhage secondary to cerebral venous sinus thrombosis; sICH, spontaneous intracerebral hemorrhage. ^aThe mean and confidence interval of cross-validation were computed based on five outcomes of the optimal hyperparameter model. The normal approximation method is used to calculate 95% CI of cross-validation.

Table 2: The performance of the classification model on the internal and external testing datasets.

areas increased the probability of CVST-ICH were mainly distributed around the hemorrhage. Conversely, blue pixels, which enhanced the tendency of sICH, were concentrated within the hemorrhage. Masking the hematoma, particularly its margins, significantly impacted the classification results, as demonstrated by the occlusion technique.

Model performance with 5 mm Z-axis resampling resolution

Segmentation

The DICE of the segmentation model with 5 mm Z-axis resampling resolution was 0.82 (0.79, 0.85) in the internal testing set, which is similar to the performance at 1 mm resolution, 0.84 (0.81, 0.87).

Classification

Similarly, classification performance at 5 mm Z-axis resolution showed slightly lower accuracy compared to the 1 mm setting, as shown in Table 4. On the internal testing set, the classification model under 5 mm Z-axis resampling resolution achieved an AUC of 0.91 (95% CI [0.85, 0.97]). At the high sensitivity operating point, the model demonstrated an accuracy of 0.81 (95% CI [0.72, 0.89]), with a sensitivity of 0.85 (95% CI [0.69, 0.96]) and a specificity of 0.80 (95% CI [0.71, 0.89]). At the high specificity operating point, the accuracy increased to

0.89 (95% CI [0.83, 0.96]), with a specificity of 0.92 (95% CI [0.85, 0.97]) and a sensitivity of 0.81 (95% CI [0.65, 0.96]). On the external set, the model achieved an AUC of 0.83 (95% CI [0.77, 0.89]). At the high sensitivity operating point, the accuracy was 0.81 (95% CI [0.68, 0.92]), with a sensitivity of 0.82 (95% CI [0.68, 0.92]) and a specificity of 0.81 (95% CI [0.73, 0.88]). At the high specificity operating point, the model maintained an accuracy of 0.81 (95% CI [0.75, 0.87]), with a specificity of 0.81 (95% CI [0.74, 0.87]) and a sensitivity of 0.82 (95% CI [0.68, 0.95]). These findings suggest that resampling to 1 mm resolution does not degrade model performance.

Discussion

In the present model, we found that an enhanced 3D U-Net with segmentation and classification through transfer learning provides an accurate and stable approach to identify cerebral hemorrhagic lesions due to cerebral venous sinus thrombosis from other spontaneous intracerebral hemorrhage. To our knowledge, this is the first study to pioneer a deep learning algorithm using only non-contrast CT in the emergency scenario to discriminate these two etiologies with totally different treatment strategies. By guiding the model concentrating on the hematoma region through a segmentation

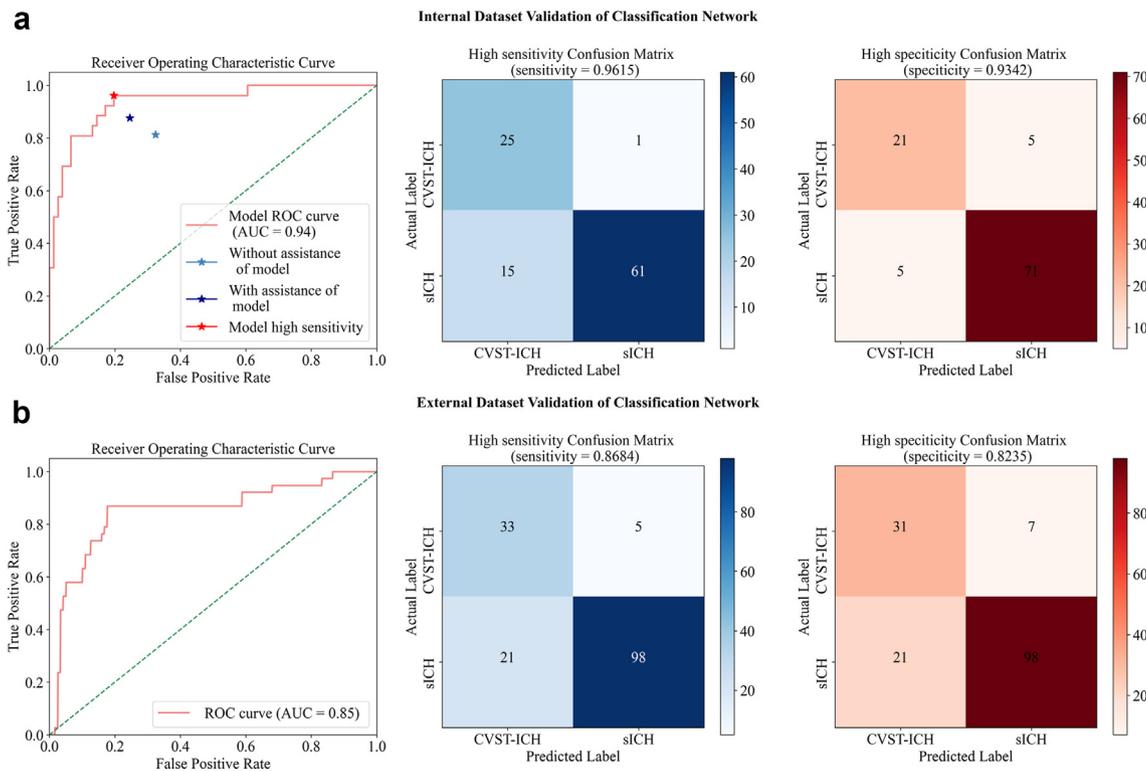


Fig. 4: ROC curves and confusion matrices for classification algorithm on internal and external testing datasets. (a) The AUC of the classification algorithm was 0.94 (95% CI 0.87–0.98) on internal testing set. At the high sensitivity operating point, the algorithm (red asterisk) outperformed the average performance of doctors (blue asterisk), and doctors’ performance improved after algorithm assistance (purple asterisk). The confusion matrix at the high sensitivity operating point was marked in blue, while the one at high specificity was in red. (b) The AUC of the classification algorithm was 0.84 (95% CI 0.76–0.92) on external testing set. The confusion matrices of high sensitivity and high specificity operating points were shown in blue and red, respectively. CVST-ICH, hemorrhage secondary to cerebral venous sinus thrombosis; siCH, spontaneous intracerebral hemorrhage; ROC, receiver operating characteristic; AUC, area under the curve.

proxy task, and enhancing this effect with a dual-channel input and integrated CBAM module, we trained the model with an AUC exceeding 0.90, as well as high sensitivity and specificity. It maintained robust performance across an external dataset originated from other four hospitals having different imaging settings. Additionally, the model outperforms doctors’ average performance and significantly improves diagnostic accuracy when used as an assistant, suggesting it could be a promising imaging aid.

Published machine learning studies have focused only on CVST. Yang X et al. developed a model using 392 patients’ MR imaging data to diagnosis CVST, with an AUC of 0.96, a sensitivity of 0.96 and specificity of 0.88 on per-patient diagnosis level among internal test data, without external validation.³¹ The model trained by 13 radiomics variables extracted from high-resolution MRI of 53 patients with CVST had a sensitivity of 0.83, a specificity of 0.94 and an AUC of 0.98 for diagnosis.³² However, a novel perspective is that our study aims to distinguish CVST from other causes among patients presenting with ICH. As CVST-ICH frequently

presents in emergency settings and requires specific anticoagulant therapy, a highly sensitive detection method is necessary.² Traditional indicators like flame-shaped and small subcortical hemorrhages were found to have limited diagnostic efficiency.³³ Other CT image signs, such as the high-density delta sign and cord sign on non-contrast CT images can implicate venous sinus thrombosis, but only appear in a small proportion of

	Average doctor performance [95% CI]	Deference [95% CI]	p value
Without model assistance (compared with classification model)			
Accuracy	0.7102 [0.6655, 0.7550]	0.1329 [0.0497, 0.2161]	0.0018
Sensitivity	0.8120 [0.7346, 0.8894]	0.1496 [0.0097, 0.2895]	0.037
Specificity	0.6754 [0.6231, 0.7278]	0.1272 [0.0262, 0.2282]	0.014
With model assistance (compared with the ones without model assistance)			
Accuracy	0.7854 [0.7392, 0.8316]	0.0752 [0.0399, 0.1104]	<0.0001
Sensitivity	0.8761 [0.8000, 0.9522]	0.0641 [0.0047, 0.1235]	0.035
Specificity	0.7544 [0.6994, 0.8093]	0.0789 [0.0362, 0.1217]	<0.0001

Table 3: The manual performance in identifying with and without the proposed model assistance.

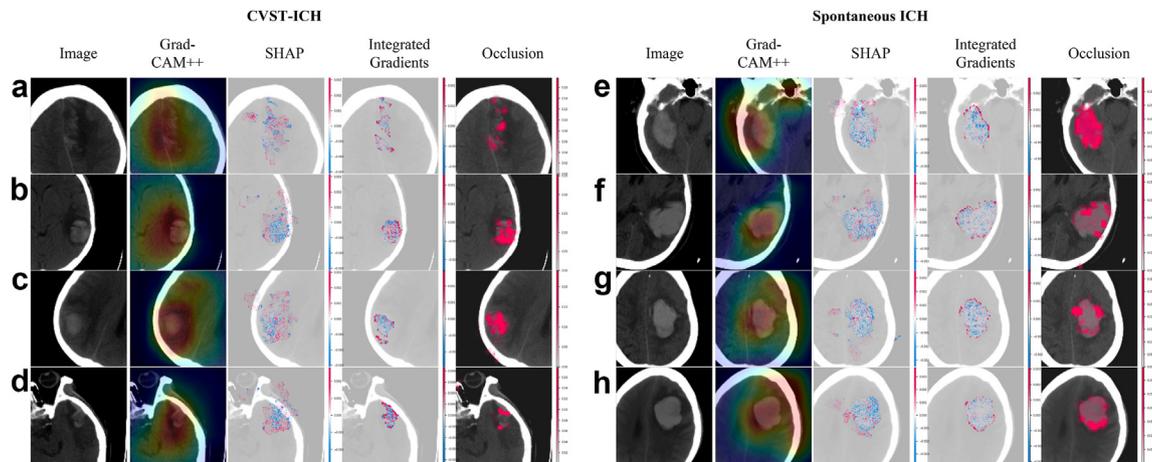


Fig. 5: Visual representation of the interpretability of the classification network. Four distinct methods are employed to elucidate the classification criteria. In the Grad-CAM++ and Occlusion maps, high-weight regions significantly affect the model prediction results. In SHAP and IG maps, red indicates that the voxel favors the model predicting CVST-ICH, while blue signifies a tendency towards sICH. For correctly predicted CVST-ICH patients, the Grad-CAM++ heatmap presentation model focuses on the hematoma and surrounding parenchyma (a, b, c, d). SHAP interpretability maps indicate that red voxels are mainly located along the edge of the hematoma and the surrounding low-density areas (a, b, c, d). IG interpretability maps demonstrate that red voxels are located at the edge of the hematoma (b, c, d). The occlusion method shows that regions masked by red blocks in the visual representation significantly impact the model's classification. For correctly predicted sICH patients, Grad-CAM++ shows the model focuses on hematoma. SHAP and IG interpretability maps indicate that blue voxels are predominantly within the hemorrhage (e, f, g, h). CVST-ICH, hemorrhage secondary to cerebral venous sinus thrombosis; ICH, intracerebral hemorrhage; SHAP, shapley additive exPlanations; Grad-CAM, gradient-weighted class activation mapping.

CVST patients.³⁴ In this context, Discriminating CVST-ICH from spontaneous intracerebral hemorrhage, using initial non-contrast CT at arrival of emergency department rather than magnetic resonance imaging, should be the priority in developing an automatic model. Therefore, we aimed to develop a model with high sensitivity to meet the diagnostic needs of doctors for this condition, and minimize the cost of specificity. On the internal testing set, the proposed model achieved a sensitivity of 0.96 while maintaining a specificity no lower than 0.80, which has potential implications for reducing missed diagnoses of CVST-ICH. Additionally, the proposed model demonstrated robustness on external data compiled from four hospitals. It is reasonable to see minor differences compared to its performance on internal data, because it may be attributed to inconsistency in imaging protocols or image quality in different hierarchy of hospitals. Overall, the model's stability and performance on external data remain satisfactory.

The transfer learning and CBAM were used to guide the model's attention towards hemorrhage regions and enhance its feature extraction capabilities. Leveraging the outstanding hematoma segmentation capability of 3D U-Net, we employed a segmentation proxy task to enhance the classification model's capacity in extracting structure and edge information of hematoma. The CBAM module incorporates two distinct attention mechanisms, channel attention and spatial attention,

allowing the feature map to obtain weights in the channel domain and spatial domain respectively, adaptively refining features, and focusing the model on learning more important features. In addition, we found that by introducing the original CT images combined with hematoma masks as dual-channel inputs in the classification network, the performance was significantly enhanced. This improvement is likely attributed to the mask serving as a guiding mechanism, reinforcing the model's focus on hemorrhagic regions, and intensifying its perception and learning of target areas.

One of the concerns in clinical practice is that standard head CT scans are typically acquired with a slice thickness of 5 mm, whereas our model was trained using isotropic 1 mm resolution. To evaluate the influence of this discrepancy, we conducted an additional experiment by resampling the Z-axis to 5 mm. The results showed that segmentation and classification performance at 5 mm in the slice direction remained comparable to that with 1 mm setting (Tables 2 and 4), indicating that 1 mm resampling did not reduce the model performance. Unlike 2D CNNs, which typically resample only in-plane resolution, isotropic 3D resampling has been widely used in deep learning-based brain CT analysis. Several studies have demonstrated its effectiveness, including Titano et al. and Tran et al., who resampled head CT images with 5 mm slice thickness to isotropic 1 mm resolution for training 3D CNNs in neurological event detection and hematoma expansion

prediction.^{18,35} Therefore, isotropic resolution of 1 mm was used in the proposed model. Regarding the influence of interpolation on medical image segmentation, our prior research investigated the effects of MRI physical resolution and interpolated resolution on segmentation, demonstrating that interpolated resolution does not introduce significant biases in segmentation tasks.³⁶

The interpretability analysis by four visualization methods, suggests that the model identifies features related to the hematoma margins, shape, and the surrounding edema, potentially reflecting the unique pathophysiology of CVST-ICH. This condition is often linked to elevated venous pressure resulting in vascular rupture and capillary wall necrosis,³⁷ leading to red blood cell extravasation.^{38,39} Unlike hypertensive hemorrhage, which originates from a single small artery and typically appears as a solitary, well-defined hematoma on CT,⁴⁰ CVST-ICH often involves multifocal subcortical veins. However, CAA-related hemorrhages can be misdiagnosed as CVST-ICH due to their irregular or multifocal nature.^{3,41} In addition, CVST-ICH is associated with significant cytotoxic or vasogenic edema,^{42,43} characterized by a mismatch in hematoma and edema volumes. This occurs due to local blood reflux and impaired blood perfusion, leading to peri-venous edema and further fluid dynamics disruption in venous sinuses.^{39,44} Beyond the visual patterns revealed by heatmaps, the proposed models may identify more nuanced and deeper level imaging characteristics, such as texture and density gradient distributions, enhancing the model's ability to distinguish between these two conditions.

Our study has several limitations. Firstly, due to the rarity of CVST-ICH, prospective data is quite challenging to obtain, thus our study utilized a sufficiently large retrospective dataset with multi-center validation. The inherent nature of retrospective data introduces selection bias and uneven baseline characteristics between CVST-ICH and other sICH, also, the notable difference in disease incidence leads to imbalance in our datasets. To resolve the risk of overfitting and underfitting due to imbalance, and to deal with the uneven baseline characters, we applied propensity score matching, alongside data augmentation and under-sampling, to minimize the impact of data imbalance, as well as bias due to retrospective nature. Additionally, we could not intervene in whether patients followed the standardized care pathway during their hospitalization, which is an inherent limitation of retrospective studies. However, we made efforts to minimize potential biases. Specifically, all CVST-ICHs were confirmed by venous imaging or DSA, and the sICH patients underwent thorough clinical, imaging, and follow-up evaluations. Secondly, the sample sizes in our training and test datasets were relatively small, compared with other diseases, although the CVST-ICH cohort is relatively large one compared to previous studies, as our

	Internal testing data		External testing data	
	CVST-ICH	sICH	CVST-ICH	sICH
Cases	26	76	38	119
AUC [95% CI]	0.9135 [0.8492, 0.9696]		0.831 [0.7747, 0.8925]	
High sensitivity operating point				
Predicted CVST-ICH	22	15	31	23
Accuracy [95% CI]	0.8137 [0.7255, 0.8922]		0.8089 [0.7389, 0.8790]	
Sensitivity [95% CI]	0.8462 [0.6923, 0.9615]		0.8158 [0.6842, 0.9211]	
Specificity [95% CI]	0.8026 [0.7105, 0.8947]		0.8067 [0.7311, 0.8824]	
PPV [95% CI]	0.5946 [0.4773, 0.7500]		0.5741 [0.4762, 0.6863]	
NPV [95% CI]	0.9385 [0.8857, 0.9848]		0.9320 [0.8857, 0.9706]	
F1 Score [95% CI]	0.6984 [0.5882, 0.8148]		0.6739 [0.5778, 0.7727]	
MCC [95% CI]	0.5881 [0.4293, 0.7496]		0.5613 [0.4238, 0.6990]	
High specificity operating point				
Predicted CVST-ICH	21	6	31	23
Accuracy [95% CI]	0.8922 [0.8333, 0.9608]		0.8089 [0.7516, 0.8726]	
Sensitivity [95% CI]	0.8077 [0.6538, 0.9615]		0.8158 [0.6842, 0.9474]	
Specificity [95% CI]	0.9211 [0.8553, 0.9737]		0.8067 [0.7395, 0.8739]	
PPV [95% CI]	0.7778 [0.6552, 0.9231]		0.5741 [0.4902, 0.6852]	
NPV [95% CI]	0.9333 [0.8861, 0.9865]		0.9320 [0.8922, 0.9780]	
F1 Score [95% CI]	0.7925 [0.6792, 0.9231]		0.6739 [0.5806, 0.7674]	
MCC [95% CI]	0.7199 [0.5669, 0.8968]		0.5613 [0.4267, 0.6968]	

AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; MCC, Matthews correlation coefficient; CI, confidence interval; CVST-ICH, intracerebral hemorrhage secondary to cerebral venous sinus thrombosis; sICH, spontaneous intracerebral hemorrhage.

Table 4: The performance of the classification model under 5 mm Z-axis resampling resolutions.

knowledge. From two large observational cohorts of ICH, comprising 1043 patients, only 10 cases (1%) were identified as CVST-ICH.⁴⁵ The ISCVT study, a prospective observational study across 21 countries and 89 centers, included 624 CVST cases, with 245 involving hemorrhage.⁴ These inherent shortcoming of obtaining larger datasets is a common limitation in CVST research, given the low incidence of CVST (with hemorrhagic cases accounting for only 30%). Despite concerns about the stability and generalizability due to the small sample size, our model showed convincing, consistent, and reliable performance across datasets, indicating its robustness. A major challenge in multi-site CT studies is the variability in acquisition protocols, scanner types, and reconstruction algorithms, which can introduce bias in models. While harmonization methods such as ComBat have been widely used to adjust radiomic features across sites, these methods are primarily designed for predefined feature-based analyses.⁴⁶ In contrast, 3D CNNs learn hierarchical features directly from images, making it technically challenging to apply ComBat to learned intermediate features. Incomplete harmonization may reduce model generalizability, leading to inconsistent performance across different clinical settings, increased misclassification, or reduced segmentation accuracy. To evaluate site-specific variations, the t-SNE was used to visualize the learned feature distribution from five sites, which

confirmed the highly overlapped features (Figure S4).⁴⁷ Finally, the non-100% specificity indicates the presence of false positive cases, thus positioning our model more as an alarm, prompting doctors to confirm the diagnosis with venography imaging.

The proposed model is provided as an open-sourced package at Github: https://github.com/CVST-Research/CVST-ICH_Classify, facilitating the generalizability testing across additional sites. Further prospective validation is necessary to ensure its reliability before it can be adopted in clinical practice. Future studies could focus on assessing the model's impact on diagnostic accuracy, workflow efficiency, and patient outcomes. Additionally, developing models that cover a broader range of hemorrhage types represents a promising direction for future research.

In conclusion, we developed an explainable and fully automated deep learning model for distinguishing intracerebral hemorrhage due to CVST from other spontaneous intracerebral hemorrhages on initial emergent CT. The model demonstrated promising generalizability and reproducibility, and the visualization tools may assist doctors in urgent recognition. Importantly, the requirement of only plain CT makes this model applicable in emergency scenarios. Future prospective validation in real-world clinical settings is needed before its integration into clinical practice.

Contributors

Conceptualization and Study Design: LST, LZ, FG, KCY, YZX; Data Curation: KCY, QL, LLT, HNA, JWZ, YQZ, GY; Investigation: KCY, YZX, JWZ, QL, LLT, YJJ; Methodology, Formal Analysis, Visualizations: KCY, YZX, LST, LZ, FG; Manuscript Writing—original draft: KCY, YZX; Manuscript Writing—review and editing: All authors; Project Administration: LST, LZ, FG, KCY, YZX; Funding: LST, LZ, FG.

KCY and YZX contributed equally to this work. KCY, YZX, LLT, QL, JWZ, YQZ, HNA, and LST had access to and verify the underlying study data. LST had final responsibility for the decision to submit the manuscript.

All authors have read and approved the final version of the manuscript.

Data sharing statement

The proposed model is provided as an open-sourced package at Github: https://github.com/CVST-Research/CVST-ICH_Classify. The data involved in this study are available from the corresponding author on reasonable request.

Declaration of interests

All authors declare no conflict of interest.

Acknowledgements

The work was funded by the National Key R&D Program of China (2023YFE0118900), National Natural Science Foundation of China (No.81971155 and No.81471168), the Science and Technology Department of Zhejiang Province (LGJ22H180004), Medical and Health Science and Technology Project of Zhejiang Province (No.2022KY174), the 'Pioneer' and 'Leading Goose' R & D Program of Zhejiang Province (No. 2024C03006 and No. 2023C03026) and the MOE Frontier Science Center for Brain Science & Brain-Machine Integration, Zhejiang University.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2025.103128>.

References

- Greenberg SM, Ziai WC, Cordonnier C, et al. 2022 Guideline for the management of patients with spontaneous intracerebral hemorrhage: a guideline from the American Heart Association/American Stroke Association. *Stroke*. 2022;53:e282–e361.
- Saposnik G, Barinagarrementeria F, Brown RD, et al. Diagnosis and management of cerebral venous thrombosis: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42:1158–1192.
- Singh T, Chakera T. Dural sinus thrombosis presenting as unilateral lobar haematomas with mass effect: an easily misdiagnosed cause of cerebral haemorrhage. *Australas Radiol*. 2002;46:351–365.
- Ferro JM, Canhão P, Stam J, Boussier M-G, Barinagarrementeria F. Prognosis of cerebral vein and dural sinus thrombosis: results of the international study on cerebral vein and dural sinus thrombosis (ISCVT). *Stroke*. 2004;35:664–670.
- Puy L, Parry-Jones AR, Sandset EC, Dowlatshahi D, Ziai W, Cordonnier C. Intracerebral haemorrhage. *Nat Rev Dis Primer*. 2023;9:14.
- Ferro JM, Boussier M-G, Canhão P, et al. European Stroke Organization guideline for the diagnosis and treatment of cerebral venous thrombosis - endorsed by the European Academy of Neurology. *Eur J Neurol*. 2017;24:1203–1213.
- Alajmi E, Zung J, Duquet-Armand M, Coutinho JM, Mandell DM. Prevalence of venous infarction in patients with cerebral venous thrombosis: baseline diffusion-weighted MRI and follow-up MRI. *Stroke*. 2023;54:1808–1814.
- Soyer B, Rusca M, Lukaszewicz A-C, et al. Outcome of a cohort of severe cerebral venous thrombosis in intensive care. *Ann Intensive Care*. 2016;6:29.
- Hilkens NA, van Asch CJJ, Werring DJ, et al. Predicting the presence of macrovascular causes in non-traumatic intracerebral haemorrhage: the DIAGRAM prediction score. *J Neurol Neurosurg Psychiatry*. 2018;89:674–679.
- Tang PH, Chai J, Chan YH, Chng SM, Lim CCT. Superior sagittal sinus thrombosis: subtle signs on neuroimaging. *Ann Acad Med Singapore*. 2008;37:397–401.
- Roland T, Jacobs J, Rappaport A, Vanheste R, Wilms G, Demaerel P. Unenhanced brain CT is useful to decide on further imaging in suspected venous sinus thrombosis. *Clin Radiol*. 2010;65:34–39.
- Sadigh G, Mullins ME, Saindane AM. Diagnostic performance of MRI sequences for evaluation of dural venous sinus thrombosis. *Am J Roentgenol*. 2016;206:1298–1306.
- Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*. 2018;392:2388–2396.
- Thanellas A, Peura H, Lavinto M, et al. Development and external validation of a deep learning algorithm to identify and localize subarachnoid hemorrhage on CT scans. *Neurology*. 2023;100:e1257–e1266.
- Affi K, Bellanger G, Buyck PJ, et al. Features of intracranial hemorrhage in cerebral venous thrombosis. *J Neurol*. 2020;267:3292–3298.
- Simaan N, Molad J, Peretz S, et al. Characteristics of cerebral sinus venous thrombosis patients presenting with intracerebral hemorrhage. *J Clin Med*. 2022;11:1040.
- He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. 2009;21:1263–1284.
- Titano JJ, Badgeley M, Schefflein J, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat Med*. 2018;24:1337–1341.
- Yeo M, Tahayori B, Kok HK, et al. Evaluation of techniques to improve a deep learning algorithm for the automatic detection of intracranial haemorrhage on CT head imaging. *Eur Radiol Exp*. 2023;7:17.
- Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham: Springer International Publishing; 2016:424–432.
- Woo S, Park J, Lee J-Y, Kwon IS. CBAM: convolutional block attention module. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, eds. *Computer Vision – ECCV 2018*. Cham: Springer International Publishing; 2018:3–19.
- Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a

- systematic review of the last decade (2011-2022). *Comput Methods Programs Biomed.* 2022;226:107161.
- 23 Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe, NV: IEEE; 2018:839–847.
 - 24 Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016:2921–2929.
 - 25 Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. In: *Proceedings of the 34th International Conference on Machine Learning*. PMLR; 2017:3319–3328.
 - 26 Lundberg S, Lee SI. A unified approach to interpreting model predictions. *arXiv*. 2017. <https://doi.org/10.48550/arXiv.1705.07874>.
 - 27 Milletari F, Navab N, Ahmadi S-A. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. Stanford, CA, USA: IEEE; 2016:565–571.
 - 28 Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS One*. 2017;12:e0177678.
 - 29 de Melo MB, Daldegan-Bueno D, Menezes Oliveira MG, de Souza AL. Beyond ANOVA and MANOVA for repeated measures: advantages of generalized estimated equations and generalized linear mixed models and its use in neuroscience research. *Eur J Neurosci*. 2022;56:6089–6098.
 - 30 Holmberg MJ, Andersen LW. Estimating risk ratios and risk differences: alternatives to odds ratios. *JAMA*. 2020;324:1098–1099.
 - 31 Yang X, Yu P, Zhang H, et al. Deep learning algorithm enables cerebral venous thrombosis detection with routine brain magnetic resonance imaging. *Stroke*. 2023;54:1357–1366.
 - 32 Chang Y-Z, Zhu H-Y, Song Y-Q, et al. High-resolution magnetic resonance imaging-based radiomic features aid in selecting endovascular candidates among patients with cerebral venous sinus thrombosis. *Thromb J*. 2023;21:116.
 - 33 Leach JL, Fortuna RB, Jones BV, Gaskill-Shiple MF. Imaging of cerebral venous thrombosis: current techniques, spectrum of findings, and diagnostic pitfalls. *Radiogr Rev Publ Radiol Soc N Am Inc*. 2006;26(Suppl 1):S19–S41. Discussion S42-43.
 - 34 Poon CS, Chang J-K, Swarnkar A, Johnson MH, Wasenko J. Radiologic diagnosis of cerebral venous thrombosis: pictorial review. *AJR Am J Roentgenol*. 2007;189:S64–S75.
 - 35 Tran AT, Zeevi T, Haider SP, et al. Uncertainty-aware deep-learning model for prediction of supratentorial hematoma expansion from admission non-contrast head computed tomography scan. *NPJ Digit Med*. 2024;7:26.
 - 36 Xu Y, Li J, Feng X, et al. Efficient segmentation of fetal brain MRI based on the physical resolution. *Med Phys*. 2024;51:7214–7225.
 - 37 Cuvinciu V, Viguier A, Calviere L, et al. Isolated acute non-traumatic cortical subarachnoid hemorrhage. *AJNR Am J Neuro-radiol*. 2010;31:1355–1362.
 - 38 Schaller B, Graf R. Cerebral venous infarction: the pathophysiological concept. *Cerebrovasc Dis*. 2004;18:179–188.
 - 39 Gotoh M, Ohmoto T, Kuyama H. Experimental study of venous circulatory disturbance by dural sinus occlusion. *Acta Neurochir*. 1993;124:120–126.
 - 40 Sheth KN. Spontaneous intracerebral hemorrhage. *N Engl J Med*. 2022;387:1589–1596.
 - 41 Samarasekera N, Rodrigues MA, Toh PS, Al-Shahi S. Imaging features of intracerebral hemorrhage with cerebral amyloid angiopathy: systematic review and meta-analysis. *PLoS One*. 2017;12:e0180923.
 - 42 Forbes KP, Pipe JG, Heiserman JE. Evidence for cytotoxic edema in the pathogenesis of cerebral venous infarction. *AJNR Am J Neuroradiol*. 2001;22:450–455.
 - 43 Nagai M, Terao S, Yilmaz G, et al. Roles of inflammation and the activated protein C pathway in the brain edema associated with cerebral venous sinus thrombosis. *Stroke*. 2010;41:147–152.
 - 44 Xiao L, Ji X, Zhao H, et al. A novel severe cerebral venous thrombosis rat model based on semi-ligation combined with ferric chloride and thrombin. *CNS Neurosci Ther*. 2022;28:2129–2140.
 - 45 Fandler-Höfler S, Ambler G, Goeldlin MB, et al. MRI-based prediction of macrovascular causes of intracerebral hemorrhage: the MACRO score. *Neurology*. 2024;103:e209950.
 - 46 Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a Method to compensate multicenter effects affecting CT radiomics. *Radiology*. 2019;291:53–59.
 - 47 Laurens VDM, Hinton G. Visualizing Data using t-SNE. *J Mach Learn Res*. 2008;9:2579–2605.