

Residual–Dense–Attention U-Net with Convolutional Block Attention Module for Automatic Intracerebral Hemorrhage Segmentation from Sensor-derived Computed Tomography Data

Cheng-Tang Pan,^{1,2,3,4,5†} Nai-Yu Chen,^{1†} Ming-Chan Lee,³ and Meng-Chi Lin^{6*}

¹International PhD Program for Science, National Sun Yat-sen University,
70 Lienhai Rd., Kaohsiung 804, Taiwan

²Institute of Medical Science and Technology, National Sun Yat-sen University,
70 Lienhai Rd., Kaohsiung 804, Taiwan

³Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-sen University,
70 Lienhai Rd., Kaohsiung 804, Taiwan

⁴National Center for Instrumentation Research, National Institutes of Applied Research,
20 R&D Rd. VI, Hsinchu Science Park, Hsinchu 300, Taiwan

⁵Institute of Advanced Semiconductor Packaging and Testing, College of Semiconductor and Advanced Technology
Research, National Sun Yat-sen University, 70 Lienhai Rd., Kaohsiung 804, Taiwan

⁶Department of Surgery, Zuoying Armed Forces General Hospital,
553 Junxiao Rd., Zuoying Dist., Kaohsiung 813, Taiwan

(Received August 5, 2026; accepted September 30, 2026)

Keywords: X-ray computed tomography detector, sensor-derived imaging data, intracerebral hemorrhage segmentation, attention mechanism, uniform design

Computed tomography (CT) measures X-ray attenuation with a detector array, typically of scintillator–photodiode elements, and reconstructs the signals into Hounsfield-unit images, making it the most widely used imaging sensor for assessing acute intracerebral hemorrhage. However, sensor-derived CT images contain quantum noise, partial-volume blurring, and weak density contrast at lesion boundaries, which limit the consistency of manual interpretation. We developed the Residual–Dense–Attention U-Net with a convolutional block attention module (RDAC U-Net), a signal-processing and feature-fusion back end that converts sensor-derived CT data into pixel-level hemorrhage masks and volumetric models. Residual blocks, dense blocks, channel–spatial attention, and attention gates were combined to enhance lesion signals, suppress background noise, and sharpen boundaries. Hyperparameters were optimized by uniform design with a Kriging surrogate model, and fivefold cross-validation was used to evaluate generalizability. For 1340 slices from 100 patients, the optimized RDAC U-Net achieved a mean Dice coefficient of 0.9213 with a standard deviation of 0.0063, outperforming four attention-based U-Net variants while requiring the shortest training time per epoch. Stacking the slice-wise masks yielded three-dimensional hemorrhage models, extending the CT sensing chain from attenuation measurement to quantitative hematoma volumetry. The study is limited by single-center data and slice-based two-dimensional processing; multicenter, multivendor validation and three-dimensional extension will be conducted in future work.

*Corresponding author: e-mail: a0978060004@mail.ngh.com.tw

†These authors contributed equally to this work.

<https://doi.org/10.18494/SAM6566>

1. Introduction

The timeliness and accuracy of intracerebral hemorrhage (ICH) diagnosis directly affect clinical decision-making and patient prognosis. In current practice, ICH is identified on noncontrast computed tomography (CT) images by visual interpretation, and hematoma volume is estimated manually, commonly with the ABC/2 approximation,⁽¹⁾ which is time-consuming and observer-dependent. An automatic and quantitative method of extracting hemorrhage information from sensor-derived CT data is therefore required.

From the sensing perspective, a CT scanner is an X-ray transmission sensing system. An X-ray tube rotates around the patient, and a curved detector array, typically consisting of scintillator elements (e.g., gadolinium oxysulfide or garnet ceramics) coupled to silicon photodiodes, converts the transmitted X-ray photons into visible light and then into electrical signals.^(2,3) After analog-to-digital conversion, the projection data are reconstructed into maps of linear attenuation coefficients and calibrated to Hounsfield units (HU) relative to water.⁽⁴⁾ Because an acute hematoma (approximately 50–90 HU) is denser than normal brain parenchyma (approximately 20–45 HU), hemorrhage is sensed as a density contrast. This contrast, however, is degraded by detector quantum noise, beam hardening near the skull, and partial-volume effects, so that lesion boundaries in sensor-derived CT images are often ill-defined.⁽²⁾

CT is the imaging modality of choice for ICH because it acquires whole-brain cross-sectional images within minutes and reveals hemorrhage location, volume, and associated findings such as intraventricular extension and edema.⁽⁵⁾ A recent meta-analysis showed that deep learning algorithms delineate ICH on noncontrast CT accurately and overcome the shape assumption of the ABC/2 method,⁽⁶⁾ whereas magnetic resonance imaging has a low diagnostic yield in spontaneous deep ICH.⁽⁷⁾

Deep learning has become the core technology of medical image analysis, including organ segmentation, lesion detection, and three-dimensional (3D) reconstruction.^(8,9) For ICH, deep networks have achieved expert-level detection on head CT.^(10,11) U-Net,⁽¹²⁾ whose encoder–decoder structure with skip connections captures both global context and fine local features from limited data, has become the standard segmentation backbone.⁽¹³⁾ Nevertheless, the segmentation of small or low-contrast hematomas remains difficult; for example, a plain U-Net achieved a Dice coefficient of only 0.31 on a public head CT dataset.⁽¹⁴⁾

In this study, we developed the Residual–Dense–Attention U-Net with a convolutional block attention module (RDAC U-Net) as a signal-processing and feature-fusion back end for sensor-derived CT data. Residual blocks (RBs),⁽¹⁵⁾ dense blocks (DBs),⁽¹⁶⁾ a convolutional block attention module (CBAM),⁽¹⁷⁾ and attention gates (AGs)⁽¹⁸⁾ were integrated to enhance hemorrhage signals, suppress background noise, and delineate lesion boundaries. The novelty of this study lies in three aspects: (1) HU windowing and the network form a processing chain that converts detector-calibrated attenuation values into hemorrhage masks; (2) hyperparameters were optimized systematically by uniform design experiment (UDE)⁽¹⁹⁾ combined with a Kriging surrogate model,⁽²⁰⁾ instead of by trial and error; and (3) slice-wise masks were reconstructed into 3D hematoma models, turning the CT scanner into a quantitative hematoma-volume sensing system. Fivefold cross-validation was used to evaluate stability across heterogeneous data subsets.

2. Data, Materials, and Methods

2.1 Computational environment

All model training and image analysis procedures were performed on a local workstation. Python 3.7 was selected as the primary programming language, supported by deep learning frameworks including TensorFlow-GPU 2.8 and Keras 2.3, as well as scientific computing libraries including NumPy, pandas, and SciPy, enabling the efficient processing of large-scale sensor-derived imaging data. Graphics processing unit (GPU) acceleration was implemented via CUDA 11.2 to exploit parallel computation capabilities. Full hardware and software specifications are listed in Table 1.

2.2 Image acquisition and preprocessing

Head CT images were acquired with the approval of the institutional review board (IRB No. KAFGHIRB 114-066). Scans were performed with a 64-row multidetector CT scanner (Aquilion 64, model TSX-101A; Toshiba Medical Systems, Otawara, Japan) routinely used in the radiology department for first-line noncontrast brain CT in outpatient, inpatient, and emergency settings. The scanner is equipped with a solid-state scintillator–photodiode detector array of 64 rows × 0.5 mm (32 mm z-axis coverage) operating in energy-integrating mode. The acquisition parameters were a tube voltage of 120 kVp, a tube current–time product of 250–350 mAs with automatic tube current modulation, and a standard (soft-tissue) reconstruction kernel; the reconstructed slice thickness was 5 mm. The sensor-derived data used in this study were 512 × 512-pixel axial images of HU values stored in Digital Imaging and Communications in Medicine (DICOM) format. The images were converted into Neuroimaging Informatics Technology Initiative (Nifti) format using Medical Imaging Interaction Toolkit (MITK) Workbench, with spatial resolution, slice thickness, and positional metadata preserved. The processing workflow is illustrated in Fig. 1.

Table 1
Hardware configuration and software environment specifications.

Component	Specification
Central processing unit (CPU)	Intel Core i5-13600K
Graphics processing unit (GPU)	NVIDIA GeForce RTX 3060 Ti (8 GB VRAM)
Solid-state drive (SSD)	M.2 PCIe 1 TB × 1; M.2 PCIe 2 TB × 1
Random access memory (RAM)	DDR4-3600 16 GB × 2; DDR4-3600 32 GB × 2

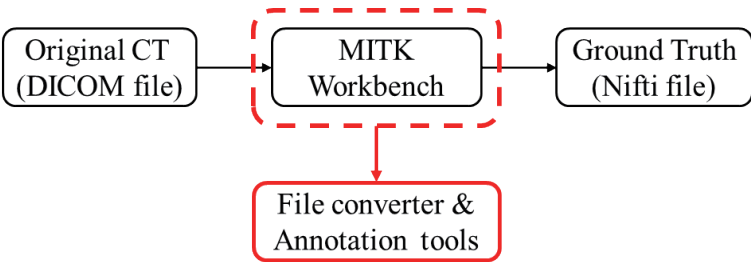


Fig. 1. (Color online) Image format conversion and rendering workflow.

Prior to ground truth (GT) annotation, sensor-derived CT images were subjected to HU windowing to enhance lesion signal contrast and suppress irrelevant tissue signals. HU quantifies the linear X-ray attenuation coefficient of tissue relative to water, as defined by Eq. (1).

$$HU = \frac{\mu_x - \mu_{water}}{\mu_{water}} \times 1000 \quad (1)$$

Here, μ_x denotes the linear attenuation coefficient of the target material and μ_{water} that of water. Following consultation with clinical collaborators, an HU window of -15 to 75 was selected to optimize hemorrhagic lesion contrast, as shown in Figs. 2(a) and 2(b).

GT annotation was subsequently performed using the built-in segmentation tools in MITK Workbench via two complementary methods: threshold-based automatic region selection, in which all pixels within the specified HU range are automatically delineated as shown in Fig. 3(a), and edge-fitting annotation, in which boundary pixels at a selected location are detected and fitted automatically as shown in Fig. 3(b). Both methods may be applied in combination according to lesion morphology to ensure annotation accuracy and cross-sample consistency.

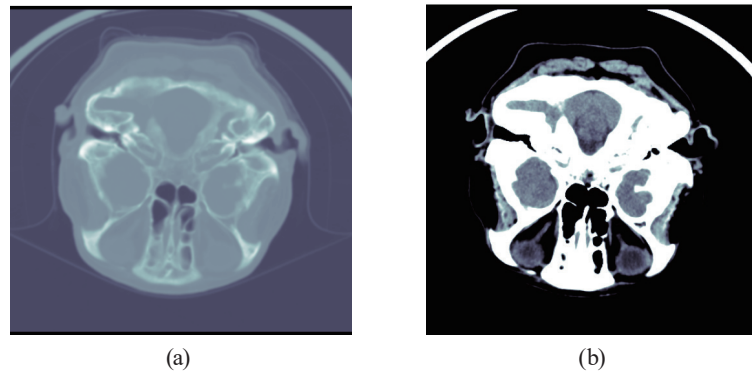


Fig. 2. (Color online) (a) Original CT image. (b) CT image after HU adjustment.

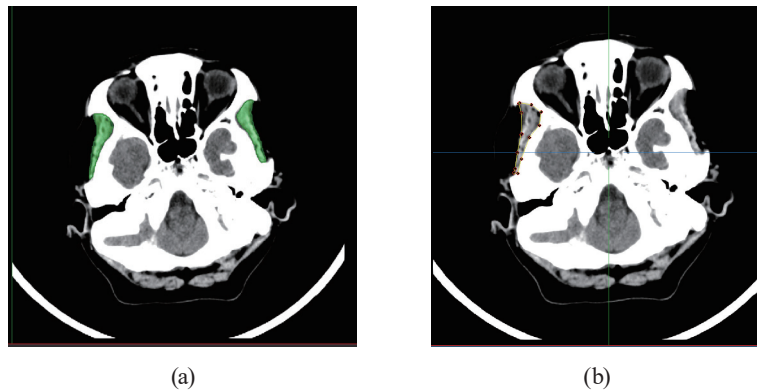


Fig. 3. (Color online) (a) Automatic thresholding of the selected image range. (b) Mask delineation by curve fitting to lesion boundaries.

2.3 Data augmentation and cross-validation

The performance of deep learning models is highly dependent on the scale and diversity of training data; insufficient samples directly degrade evaluation metrics and increase the risk of overfitting. Because the acquisition of sensor-derived imaging data is constrained by clinical enrollment procedures and GT annotation requires manual delineation by qualified clinicians, data availability is inherently limited. To address this, data augmentation, including rotation, scaling, shearing, and translation, was applied to expand the effective training dataset, and K-fold cross-validation was adopted for robust model performance assessment.

For comprehensive and unbiased model evaluation, K-fold cross-validation was employed. The complete dataset was partitioned into K mutually exclusive subsets; in each of K independent training-validation iterations, one subset served as the validation set, while the remaining K–1 subsets were combined as the training set. This procedure ensures that every sample is used exactly once for validation, yielding more robust performance estimates. The validation scheme is illustrated in Fig. 4.

2.4 RDAC U-Net architecture

The proposed RDAC U-Net adopts Attention U-Net as its backbone framework and integrates RBs, DBs, CBAM, and AGs within both the encoder and decoder pathways to enhance lesion boundary segmentation accuracy and deep feature learning capability. The complete architecture is illustrated in Fig. 5.

In the proposed RDAC U-Net architecture, four key modules are integrated to enhance feature extraction and segmentation accuracy. RBs incorporate skip connections within convolutional layers, allowing gradients to propagate more effectively during backpropagation, thereby mitigating the vanishing gradient problem and enabling the network to learn deeper and

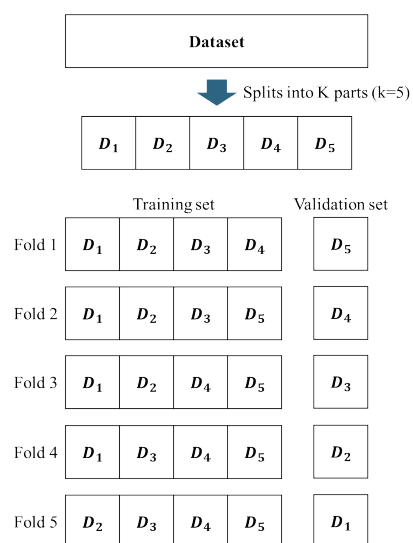


Fig. 4. K-fold cross-validation scheme.

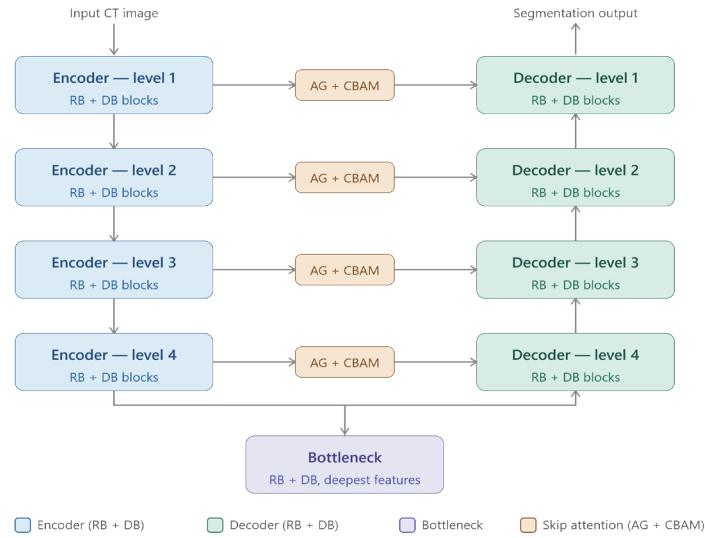


Fig. 5. (Color online) Architecture diagram of the RDAC U-Net model.

more discriminative representations of hemorrhage regions. DBs connect each layer to all preceding layers within the block, promoting feature reuse and strengthening feature propagation, which improves the network's ability to capture fine-grained lesion characteristics while reducing the number of parameters required. CBAM sequentially applies channel and spatial attention mechanisms to the feature maps, enabling the model to emphasize informative regions and suppress irrelevant background noise, thereby improving the discrimination between hemorrhagic lesions and the surrounding brain tissue. Finally, AGs are embedded within the skip connections between the encoder and the decoder, allowing the network to selectively highlight salient lesion-related features while suppressing irrelevant information from background regions before feature fusion, which is particularly beneficial for delineating the ill-defined and irregular boundaries characteristic of intracerebral hemorrhage. Collectively, these four modules complement one another to strengthen feature representation, suppress background interference, and improve boundary delineation accuracy in the segmentation task.

2.5 Hyperparameter optimization

Hyperparameters including batch size, epoch count, the number of iterations, and learning rate exert direct effects on training performance, as expressed by the relationship in Eq. (2).

$$Epoch \times Batch\ size = Dataset\ size \times Iteration \quad (2)$$

A larger batch size improves gradient estimation stability at the cost of increased computational demand, whereas a smaller batch size introduces greater stochasticity, facilitating escape from local optima. An insufficient number of epochs risk underfitting, while an excessive

number of epochs promote overfitting. A higher learning rate accelerates early convergence but may cause loss function oscillation, whereas a lower learning rate permits finer parameter adjustment at the expense of convergence speed.

To systematically identify the optimal hyperparameter configuration, UDE⁽¹⁹⁾ was applied using the uniform design table U_{13}^* (13⁸). Following its usage table for three factors ($s = 3$), the columns with generating vector (1, 9, 11) were assigned to batch size, learning rate, and the number of epochs, respectively, yielding 13 runs. The uniformity of the resulting design, evaluated by the centered L_2 discrepancy,⁽¹⁹⁾ was $CD_2 = 0.0790$, equal to the minimum among all three-column combinations of the table. Factor ranges were established on the basis of published studies: excessively large batch sizes reduce model generalizability,⁽²¹⁾ so the batch size was set to 2–14; deep networks can memorize training data when trained excessively,⁽²²⁾ so the epoch range was limited to 130–250; and following the comparative analysis of learning rate policies by Wu *et al.*,⁽²³⁾ the learning rate was set to 1×10^{-5} – 1×10^{-3} . The factor ranges and levels are summarized in Table 2.

2.6 Evaluation metrics

Model segmentation performance was assessed using the confusion matrix, F_1 -score, and Dice coefficient, all of which are derived from four fundamental classification statistics: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The confusion matrix structure is shown in Fig. 6, and schematic definitions of TP, TN, FP, and FN are illustrated in Fig. 7.

Table 2
Factor ranges and levels.

	Batch size	Epoch	Learning rate
Minimum	2	130	1×10^{-5}
Maximum	14	250	1×10^{-5}

Confusion matrix		Predict	
		False	True
Ground truth	False	True Negative	False Positive
	True	False Negative	True Positive

Fig. 6. Confusion matrix.

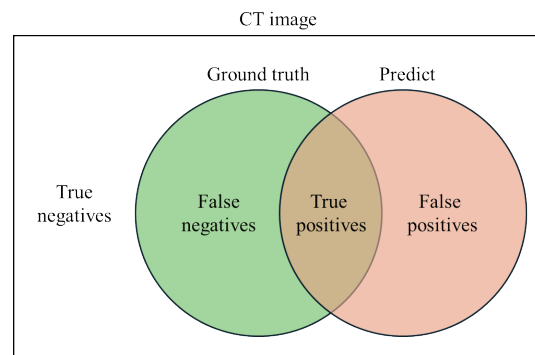


Fig. 7. (Color online) Schematic definitions of TP, TN, FP, and FN.

Precision quantifies the proportion of correctly identified positive predictions among all predicted positives and recall quantifies the proportion of correctly identified positive samples among all actual positives, as defined in Eqs. (3) and (4), respectively.

$$Precision = \frac{TP}{TP + FP}, \quad (3)$$

$$Recall = \frac{TP}{TP + FN}. \quad (4)$$

The F_1 -score is the harmonic mean of precision and recall, providing a balanced composite metric suitable for evaluating performance on imbalanced datasets, as expressed in Eq. (5).

$$F_1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \quad (5)$$

When the differential weighting of precision and recall is required, a weight parameter β may be introduced to obtain the generalized F_β -score, as defined in Eq. (6), where $\beta = 1$ assigns equal weight to both metrics.

$$F_\beta\text{-score} = \frac{(1 + \beta^2) \times Precision \times Recall}{(\beta^2 \times Precision) + Recall} \quad (6)$$

As precision and recall are considered equally important in this study, the F_1 -score ($\beta = 1$) is adopted as the classification performance indicator. Overall classification accuracy is computed as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (7)$$

However, because hemorrhagic lesions occupy a relatively small proportion of the total CT image volume, the abundance of TN samples may artificially inflate Accuracy. The Dice coefficient is therefore adopted as the primary segmentation performance metric, as it directly quantifies the spatial overlap between model predictions and GT masks over a range of 0 to 1, where 1 indicates perfect agreement, as defined in Eq. (8).

$$Dice\ coefficient = \frac{2 \times |Prediction \cap GT|}{|Prediction| + |GT|} \quad (8)$$

Although the F_1 -score and Dice coefficient are mathematically equivalent, they serve distinct analytical purposes: the F_1 -score is primarily employed in classification tasks to evaluate the detection of the positive class, whereas the Dice coefficient is employed in segmentation tasks to assess pixel-level spatial overlap. The concurrent use of both metrics provides a more comprehensive evaluation framework, mitigating the bias inherent in reliance on any single indicator and strengthening confidence in model reliability for clinical sensing applications.

3. Results and Discussion

3.1 Training dataset and data augmentation

The sensor-derived CT dataset comprised brain CT scans of 100 patients with ICH, with approximately 10 to 15 axial slices per patient, yielding a total of 1340 images. To prevent data leakage, the dataset was split at the patient level before augmentation: 20 patients (20%) were held out as an independent test set and the remaining 80 patients were used for fivefold cross-validation. Data augmentation was applied only to the training data.

To comply with model input specifications and reduce GPU memory demand, all images were downsampled from 512×512 to 224×224 pixels, which also improved computational efficiency. Sequential data augmentation, comprising affine transformation, horizontal flip, and elastic transformation, was applied to the training images, expanding the training dataset approximately threefold.

The augmentation pipeline comprised three sequential stages: affine transformation, horizontal flip, and elastic transformation. During the affine transformation stage, four geometric operations were applied simultaneously: rotation over a range of $\pm 15^\circ$ [Figs. 8(a) and 8(b)]; scaling over a range of 100% to 120%; shearing over a range of $\pm 5^\circ$; and translation along both the x- and y-axes over a range of ± 10 pixels.

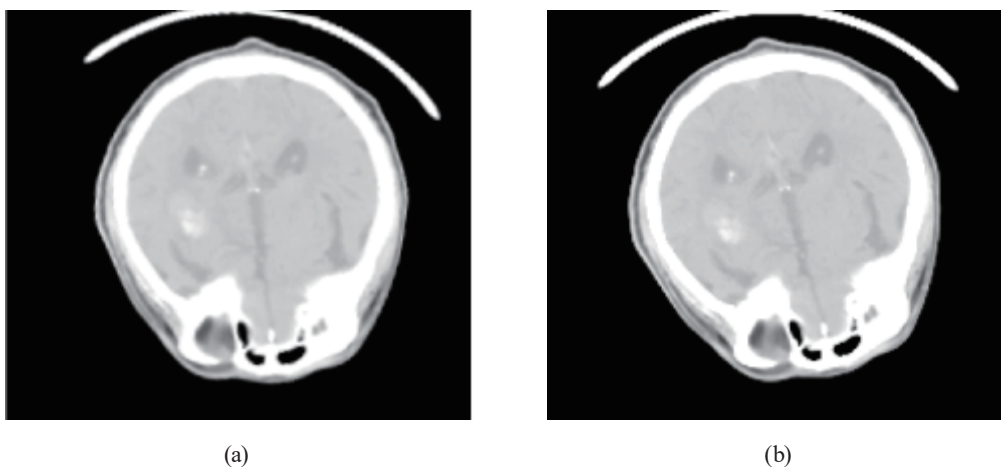


Fig. 8. (Color online) (a) Original and (b) augmented CT images (rotation).

3.2 Comparative model performance

RDAC U-Net was compared with four reference architectures, namely, Attention U-Net⁽¹⁸⁾ and its variants incorporating residual blocks (Attention Res U-Net),⁽¹⁵⁾ dense blocks (Attention Dense U-Net),⁽¹⁶⁾ and CBAM (CBA U-Net),⁽¹⁷⁾ in terms of parameter count, training time per epoch, and segmentation performance. Parameter counts and training times are summarized in Table 3.

To minimize the evaluation bias introduced by dataset partitioning, all models were trained using fivefold cross-validation with the same hyperparameters (batch size = 7, epoch = 200, and learning rate = 1×10^{-4}). The means and standard deviations (SDs) of the fivefolds are summarized in Table 4.

RDAC U-Net required the shortest training time per epoch (127 s), whereas Attention Res U-Net, Attention Dense U-Net, CBA U-Net, and Attention U-Net required 3.93, 36.22, 40.15, and 33.07% longer, respectively (Fig. 9). This demonstrates that RDAC U-Net achieves an effective balance between model complexity and computational efficiency.

Table 3
Parameter count and training time per epoch of compared models.

Model	Parameter count	Training time per epoch (s)
RDAC U-Net (proposed)	13119854	127
Attention Res U-Net	11456201	132
Attention Dense U-Net	14374021	173
CBA U-Net	35311449	178
Attention U-Net	35238293	169

Table 4
Fivefold cross-validation performance of compared models (mean $\pm$ SD).

Model	F_1 -score	Precision	Recall	Dice coefficient
RDAC U-Net	0.9324 ± 0.0021	0.9635 ± 0.0019	0.9198 ± 0.0121	0.9042 ± 0.0061
Attention Res U-Net	0.9147 ± 0.0062	0.9382 ± 0.0040	0.8925 ± 0.0109	0.8504 ± 0.0075
Attention Dense U-Net	0.9156 ± 0.0052	0.9386 ± 0.0101	0.8940 ± 0.0116	0.8570 ± 0.0067
CBA U-Net	0.9153 ± 0.0054	0.9477 ± 0.0049	0.8850 ± 0.0114	0.8577 ± 0.0083
Attention U-Net	0.9217 ± 0.0060	0.9518 ± 0.0038	0.8935 ± 0.0087	0.8656 ± 0.0078

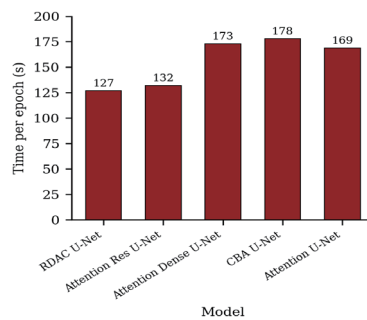


Fig. 9. (Color online) Comparison of training time.

With respect to segmentation performance, higher values of F_1 -score, precision, recall, and Dice coefficient indicate superior lesion classification and boundary fitting. RDAC U-Net outperformed all four reference architectures in every metric (Table 4). Relative to Attention Res U-Net, Attention Dense U-Net, CBA U-Net, and Attention U-Net, the F_1 -score improved by 1.89, 1.80, 1.83, and 1.14% and the Dice coefficient by 5.95, 5.22, 5.14, and 4.26%, respectively (Figs. 10–13). These results indicate that the multimodule design of RDAC U-Net enhances lesion signal feature learning and boundary precision.

3.3 Hyperparameter optimization

UDE was applied to systematically optimize the training hyperparameters of RDAC U-Net. Because the learning rate spans two orders of magnitude, its 13 levels were spaced uniformly on a logarithmic scale (1.0, 1.5, 2.2, 3.2, 4.6, and 6.8×10^n), whereas batch size and epoch were spaced linearly with steps of 1 and 10, respectively (Table 2). All factor values were normalized to [0, 1] before surrogate modeling.

Thirteen experimental runs were constructed from the uniform design table. The Dice coefficient was selected as the target index (larger-the-better), and the configurations and results of all runs are listed in Table 5.

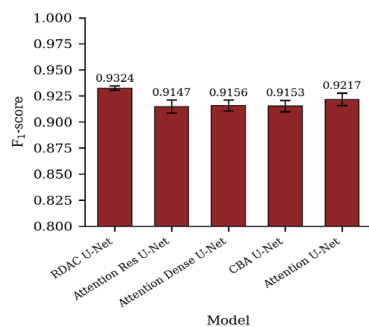


Fig. 10. (Color online) Comparison of F_1 -score.

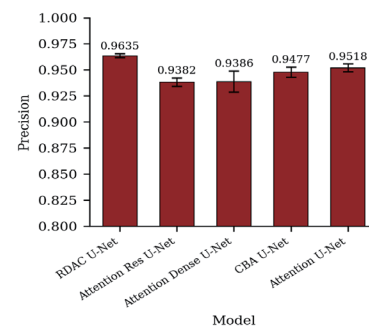


Fig. 11. (Color online) Comparison of precision.

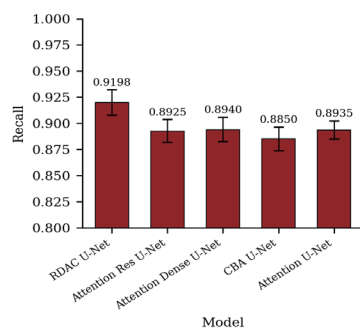


Fig. 12. (Color online) Comparison of recall.

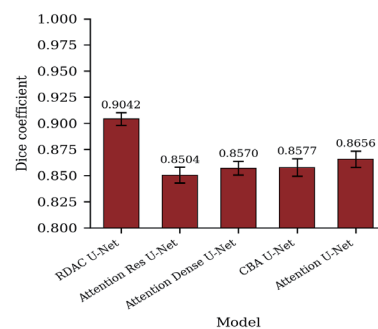


Fig. 13. (Color online) Comparison of Dice coefficient.

Table 5
Uniform design runs and resulting Dice coefficients.

No.	Factors			Target indicators (Dice coefficient)
	Batch size	Learning rate	Epoch	
1	2	2.2×10^{-4}	230	0.9114
2	3	3.2×10^{-5}	200	0.9058
3	4	1×10^{-3}	170	0.9005
4	5	1.5×10^{-4}	140	0.9151
5	6	2.2×10^{-5}	250	0.9135
6	7	6.8×10^{-4}	220	0.9176
7	8	1×10^{-4}	190	0.9186
8	9	1.5×10^{-5}	160	0.9101
9	10	4.6×10^{-4}	130	0.9142
10	11	6.8×10^{-5}	240	0.9202
11	12	1×10^{-5}	210	0.9179
12	13	3.2×10^{-4}	180	0.9192
13	14	4.6×10^{-5}	150	0.9147

A Kriging surrogate model⁽²⁰⁾ was subsequently constructed using MATLAB Optimization Toolbox, with the maximization of the target index as the optimization criterion. The optimal hyperparameter configuration identified was batch size = 10, epoch = 210, and learning rate = 1.55×10^{-4} , yielding an optimized target index of 0.9212. Response surfaces of the Kriging surrogate model are presented in Figs. 14(a)–14(c), showing the interactions among hyperparameters and the target index.

Linear regression analysis yielded the following regression equation, as expressed in Eq. (9).

$$Y = 0.89826 + 0.000095A_1 - 3.32A_2 + 0.000046A_3 \quad (9)$$

Here, Y is the target index, A_1 is the batch size, A_2 is the learning rate, and A_3 is epoch.

The statistical significance of each factor is summarized in Table 6. Factors with probability values (P) below 0.05 are considered statistically significant, with smaller P values indicating greater effects. Batch size was identified as the dominant factor ($P = 0.026$), followed by epoch ($P = 0.225$) and learning rate ($P = 0.481$). The ranking of factor effects is batch size > epoch > learning rate, with explained variance contributions of 76, 18.2, and 5.8%, respectively.

3.4 Optimized training results

RDAC U-Net was retrained using the optimized hyperparameters (batch size = 10, epoch = 210, and learning rate = 1.55×10^{-4}), and fivefold cross-validation was applied to examine training performance across all data partitions. Quantitative results are presented in Table 7.

The mean Dice coefficient of 0.9213 differed from the UDE-predicted optimized target index of 0.9212 by only 0.01%, confirming the predictive accuracy of the Kriging surrogate model. Fold 5 achieved the highest individual Dice coefficient of 0.9296; its training convergence curve is shown in Fig. 15.

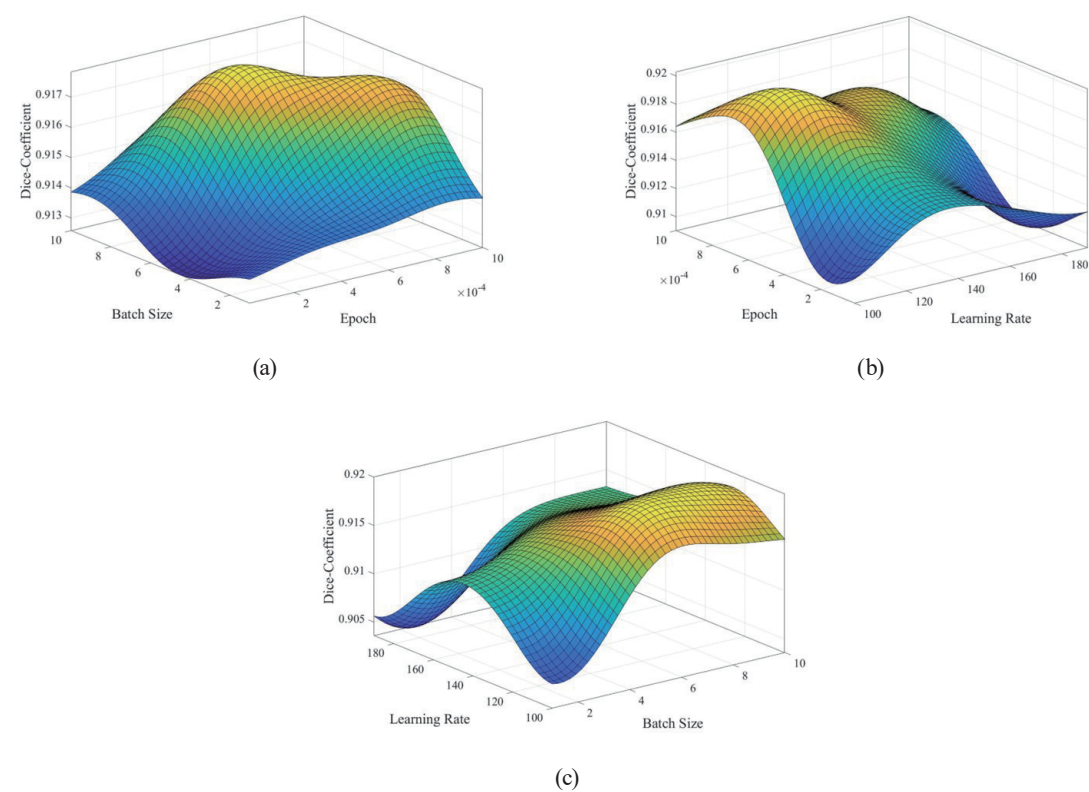


Fig. 14. (Color online) Kriging surrogate model: (a) effects of batch size and epoch on target index, (b) effects of epoch and learning rate on target index, and (c) effects of learning rate and batch size on target index.

Table 6
Statistical significance of each factor.

	Constant	Batch size	Epoch	Learning rate
Regression coefficient	0.8982	0.000095	0.000046	−3.32
<i>P</i> value	—	0.026	0.225	0.481
Percentage of variation explained (%)	—	76	18.2	5.8

Table 7
Fivefold cross-validation training results of RDAC U-Net under optimized hyperparameters.

	<i>F</i> ₁ -score	Precision	Recall	Accuracy	Dice coefficient
Fold 1	0.9275	0.9305	0.9244	0.9989	0.9240
Fold 2	0.9193	0.9361	0.9031	0.9988	0.9162
Fold 3	0.9250	0.9456	0.9052	0.9988	0.9231
Fold 4	0.9185	0.9340	0.9034	0.9988	0.9139
Fold 5	0.9313	0.9412	0.9216	0.9989	0.9296
Avg.	0.9243	0.9375	0.9115	0.9988	0.9213
SD	0.0054	0.0059	0.0105	0.00005	0.0063

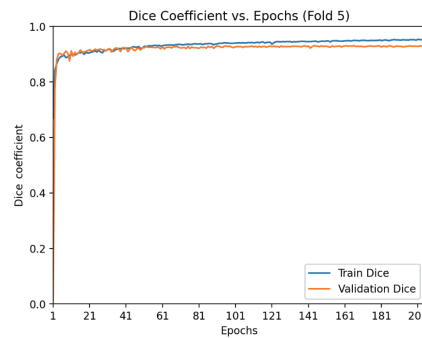


Fig. 15. (Color online) Training convergence curve.

3.5 ICH lesion segmentation results

The weight parameters derived from Fold 5 training were adopted as the final model configuration and evaluated on the independent held-out test set of 20 patients, none of whose images were used for training or augmentation, to provide an unbiased estimate of generalization performance. Segmentation predictions on the test set are shown in Fig. 16, presenting in sequence the original sensor-derived CT images, manually annotated GT masks, RDAC U-Net prediction outputs, and overlay comparisons between GT and predictions.

RDAC U-Net demonstrated consistent segmentation capability across diverse hemorrhage morphologies and anatomical locations. In cases with larger hemorrhage volumes and pronounced grayscale contrast (rows one and two), the model achieved complete lesion coverage with predicted boundaries closely matching manual annotations. In cases involving smaller hemorrhage volumes and reduced signal contrast relative to the surrounding brain parenchyma (rows three and four), the model identified the primary lesion regions despite boundary ambiguity and near-parenchymal signal intensity, with only minor boundary underestimation. These findings reflect the model's sensitivity and robustness in detecting small-volume lesions from sensor-derived CT data, and demonstrate its potential as a clinically deployable auxiliary diagnostic sensing system.

3.6 3D hemorrhage model reconstruction

The 3D visualization of ICH lesions provides significant value for both clinical evaluation and research applications. Volumetric reconstruction enables the accurate representation of hemorrhage morphology, volume, and spatial distribution within the cranium, supporting precise lesion localization, volumetric quantification, and clinical severity assessment. Compared with single-plane sensor-derived CT slices, 3D models intuitively display the spatial relationships between hemorrhagic regions and surrounding structures including brain parenchyma, ventricles, and cranial vault, thereby informing treatment decisions such as whether hematoma expansion threatens eloquent cortical areas or whether surgical evacuation is indicated.

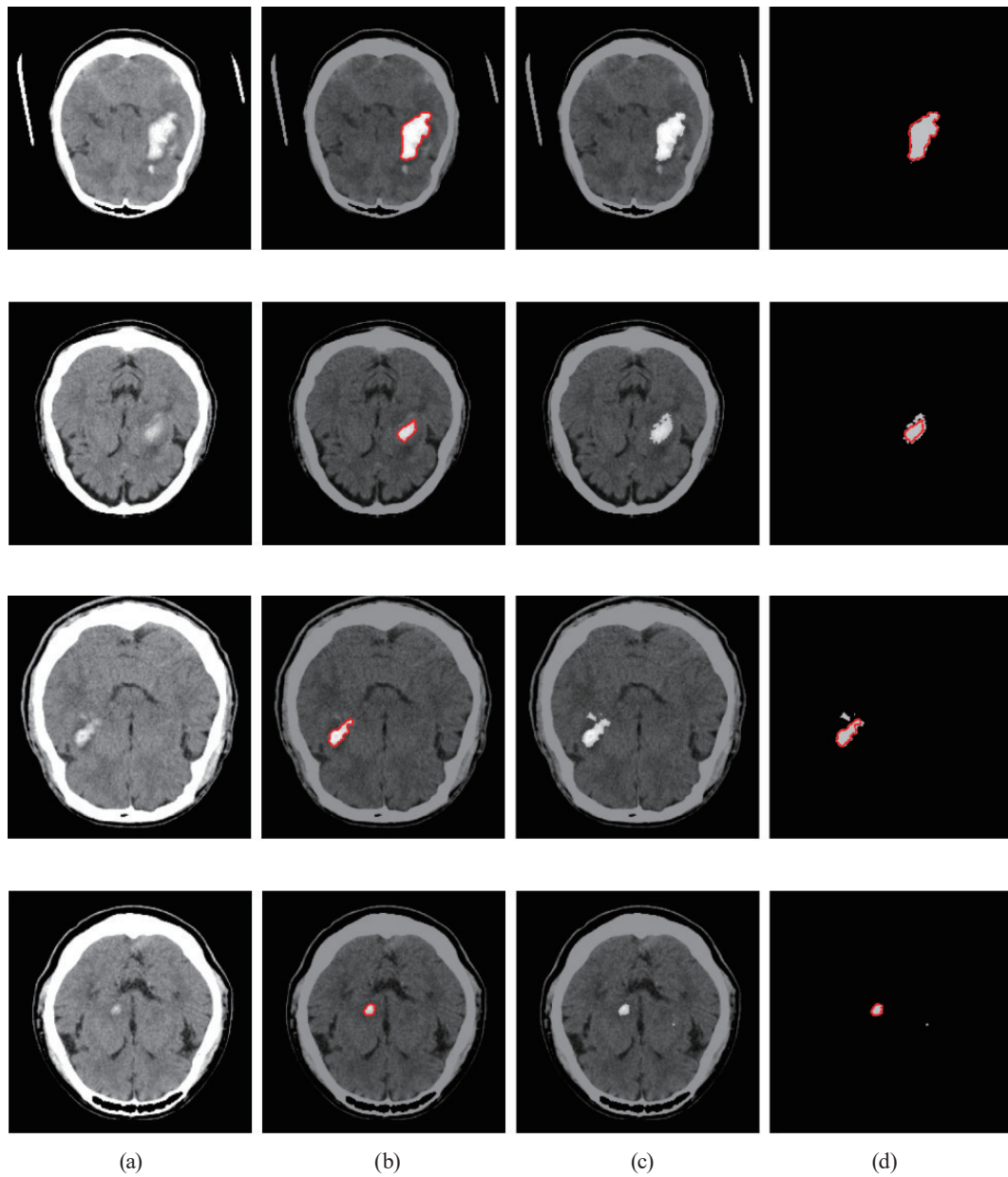


Fig. 16. (Color online) Segmentation results of RDAC U-Net: (a) original CT, (b) ground truth, (c) RDAC U-Net prediction, and (d) overlay comparison of ground truth and prediction.

In this study, the layer-wise two-dimensional segmentation masks generated by RDAC U-Net were integrated through volumetric reconstruction to produce complete 3D hemorrhage volume models. Reconstruction results are presented in Fig. 17, including the original GT-annotated images and corresponding 3D visualizations, which may serve as auxiliary references for real-time CT sensing interpretation and surgical planning.

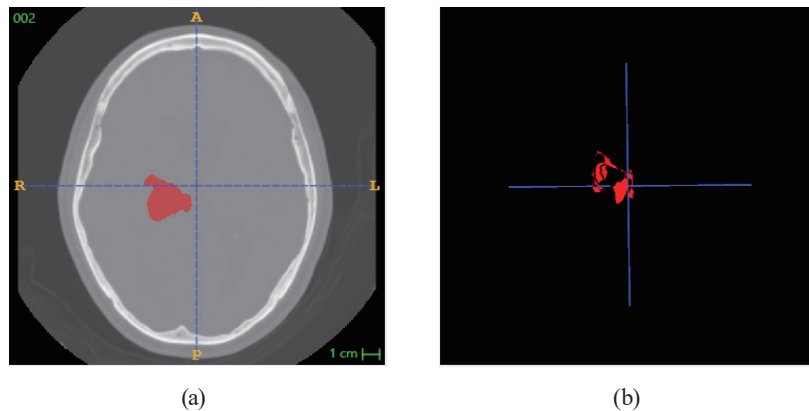


Fig. 17. (Color online) 3D modeling of brain images: (a) original ground-truth image and (b) 3D visualization model.

3.7 Comparison with previous studies and limitations

Previous deep learning studies on ICH in head CT have mainly focused on slice- or scan-level detection. Kuo *et al.*⁽¹⁰⁾ and Chilamkurthy *et al.*⁽¹¹⁾ achieved radiologist-level detection performance, but their outputs are labels or coarse localization rather than volumetric masks. Among segmentation studies, Hssayeni *et al.*⁽¹⁴⁾ applied a plain U-Net to a public dataset of 82 head CT scans and obtained a Dice coefficient of 0.31, reflecting the difficulty of segmenting small, low-contrast hematomas. nnU-Net⁽¹³⁾ provides a strong self-configuring baseline but determines hyperparameters by heuristic rules rather than by systematic experimentation. In comparison, RDAC U-Net achieved a mean Dice coefficient of 0.9213 and outperformed the four attention-based U-Net variants trained on the same data (Table 4) while requiring the shortest training time per epoch. Because datasets, slice selection, and annotation protocols differ among studies, these values are not directly comparable; nevertheless, the results indicate that combining gradient-preserving residual learning (RB), feature reuse (DB), and two-stage attention (CBAM and AGs) effectively compensates for the noise and partial-volume blurring of sensor-derived CT data.

From the sensing standpoint, the proposed system acts as the back-end processing layer of the CT sensing chain: X-ray source, scintillator–photodiode detector, reconstruction and HU calibration, RDAC U-Net, and hematoma mask and volume. It converts detector-calibrated attenuation values into a quantitative, clinically actionable output, namely, hematoma volume, which the scanner alone does not provide. The same processing chain can be applied to emerging photon-counting CT detectors,⁽³⁾ whose higher spatial resolution and lower electronic noise are expected to further improve boundary delineation. Moreover, because the data were acquired with a routine 64-row scanner widely deployed for emergency brain CT, the system can be added to existing CT sensing chains without hardware modification.

This study has limitations. First, all data were acquired at a single center with one scanner and acquisition protocol; differences in detector type, tube voltage, reconstruction kernel, and

slice thickness among vendors may change HU values and noise texture, and thus segmentation performance. Second, images were processed slice by slice and downsampled to 224×224 pixels, which discards inter-slice context and may cause the minor boundary underestimation of small lesions observed in Fig. 16. Third, GT masks were annotated semi-automatically without inter-observer agreement analysis. Future work will include multicenter, multivendor validation, 2.5D and 3D network extensions, and the comparison of automatically derived hematoma volumes with ABC/2 estimates.⁽¹⁾

4. Conclusion

Deep learning has become an essential tool for analyzing sensor-derived medical images. For ICH, the automatic segmentation of sensor-derived CT data provides objective and quantitative hematoma measurements for assessment, disease monitoring, and surgical planning, reducing the burden and variability of manual interpretation.

This study introduced RDAC U-Net, a signal-processing and feature-fusion back end for sensor-derived CT data that integrates RB, DB, AGs, and CBAM to enhance hemorrhage signals, suppress noise, and delineate lesion boundaries at a low computational cost.

The principal contributions of this work are summarized as follows:

- (i) In terms of training efficiency, RDAC U-Net required the shortest training time per epoch (127 s), whereas Attention Res U-Net, Attention Dense U-Net, CBA U-Net, and Attention U-Net required 3.93, 36.22, 40.15, and 33.07% longer, respectively.
- (ii) In terms of segmentation performance, RDAC U-Net surpassed all reference architectures in F_1 -score, precision, recall, and Dice coefficient. Relative to Attention Res U-Net, Attention Dense U-Net, CBA U-Net, and Attention U-Net, the F_1 -score improved by 1.89, 1.80, 1.83, and 1.14% and the Dice coefficient by 5.95, 5.22, 5.14, and 4.26%, respectively.
- (iii) In terms of hyperparameter optimization, UDE combined with a Kriging surrogate model identified the optimal configuration (batch size = 10, epoch = 210, and learning rate = 1.55×10^{-4}) with a target index of 0.9212, only 0.01% different from the experimentally validated mean Dice coefficient of 0.9213. Fivefold cross-validation confirmed robust generalizability, with an SD of 0.0063.
- (iv) In terms of clinical sensing application, the 3D reconstruction of RDAC U-Net segmentation outputs produced volumetric hemorrhage models capable of intuitively representing the spatial distribution of hemorrhagic structures of varying sizes and morphological characteristics, supporting hemorrhage range quantification, lesion localization, and treatment decision-making.

In summary, RDAC U-Net extends the CT sensing chain from X-ray attenuation measurement to quantitative hematoma volumetry and shows potential as a real-time auxiliary diagnostic tool for emergency imaging. Its main limitations are the single-center dataset, slice-based two-dimensional processing, and the lack of external validation. Future work will address these issues through multicenter, multivendor validation including photon-counting CT data, 3D network extensions, and deployment on scanner-side or edge computing hardware for real-time use.

Acknowledgments

The authors gratefully acknowledge the support from the Zuoying Armed Forces General Hospital (Project No. ZYAFGH_A114019) and the National Science and Technology Council (NSTC), Taiwan (Project Nos. 115-2221-E-110-088, 115-2622-E-110-009, 115-2640-E-006-004, 115-2622-E-002-020, 115-2622-E-110-005, and 115-2813-C-110-051-E).

References

- 1 R. U. Kothari, T. Brott, J. P. Broderick, W. G. Barsan, L. R. Sauerbeck, M. Zuccarello, and J. Khoury: *Stroke* **27** (1996) 1304. <https://doi.org/10.1161/01.STR.27.8.1304>
- 2 L. W. Goldman: *J. Nucl. Med. Technol.* **35** (2007) 115. <https://doi.org/10.2967/jnmt.107.042978>
- 3 M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann: *Radiology* **289** (2018) 293. <https://doi.org/10.1148/radiol.2018172656>
- 4 G. N. Hounsfield: *Br. J. Radiol.* **46** (1973) 1016. <https://doi.org/10.1259/0007-1285-46-552-1016>
- 5 S. M. Greenberg, W. C. Ziai, C. Cordonnier, D. Dowlatshahi, B. Francis, J. N. Goldstein, J. C. Hemphill, R. Johnson, K. M. Keigher, and W. J. Mack: *Stroke* **53** (2022) e282. <https://doi.org/10.1161/STR.0000000000000407>
- 6 D. Zarei, M. Issaiy, S. Kolahi, and D. S. Liebeskind: *Stroke Vasc. Interv. Neurol.* **4** (2024) e001314. <https://doi.org/10.1161/SVIN.123.001314>
- 7 H. McKinney, B. A. Kirk, A. J. Jailwala, A. McFarlane, J. L. Sullivan, R. Agarwal, and K. D. Hiatt: *Emerg. Radiol.* **32** (2025) 545. <https://doi.org/10.1007/s10140-025-02348-z>
- 8 G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez: *Med. Image Anal.* **42** (2017) 60. <https://doi.org/10.1016/j.media.2017.07.005>
- 9 X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng: *IEEE Trans. Med. Imaging* **37** (2018) 2663. <https://doi.org/10.1109/TMI.2018.2845918>
- 10 W. Kuo, C. Häne, P. Mukherjee, J. Malik, and E. L. Yuh: *Proc. Natl. Acad. Sci. U.S.A.* **116** (2019) 22737. <https://doi.org/10.1073/pnas.1908021116>
- 11 S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier: *Lancet* **392** (2018) 2388. [https://doi.org/10.1016/S0140-6736\(18\)31645-3](https://doi.org/10.1016/S0140-6736(18)31645-3)
- 12 O. Ronneberger, P. Fischer, and T. Brox: *Proc. Medical Image Computing and Computer-Assisted Intervention* (Springer, Cham, 2015) 234. https://doi.org/10.1007/978-3-319-24574-4_28
- 13 F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein: *Nat. Methods* **18** (2021) 203. <https://doi.org/10.1038/s41592-020-01008-z>
- 14 M. D. Hssayeni, M. S. Croock, A. D. Salman, H. F. Al-khafaji, Z. A. Yahya, and B. Ghoraani: *Data* **5** (2020) 14. <https://doi.org/10.3390/data5010014>
- 15 K. He, X. Zhang, S. Ren, and J. Sun: *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (IEEE, Las Vegas, 2016) 770. <https://doi.org/10.1109/CVPR.2016.90>
- 16 G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger: *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (IEEE, Honolulu, 2017) 2261. <https://doi.org/10.1109/CVPR.2017.243>
- 17 S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon: *Proc. European Conf. Computer Vision* (Springer, Cham, 2018) 3. https://doi.org/10.1007/978-3-030-01234-2_1
- 18 O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert: *arXiv:1804.03999* (2018). <https://doi.org/10.48550/arXiv.1804.03999>
- 19 K.-T. Fang, D. K. J. Lin, P. Winker, and Y. Zhang: *Technometrics* **42** (2000) 237. <https://doi.org/10.1080/00401706.2000.10486045>
- 20 D. R. Jones, M. Schonlau, and W. J. Welch: *J. Global Optim.* **13** (1998) 455. <https://doi.org/10.1023/A:1008306431147>
- 21 T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li: *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition* (IEEE, Long Beach, 2019) 558. <https://doi.org/10.1109/CVPR.2019.00065>
- 22 C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals: *Commun. ACM* **64** (2021) 107. <https://doi.org/10.1145/3446776>
- 23 Y. Wu, L. Liu, J. Bae, K. H. Chow, A. Iyengar, C. Pu, W. Wei, L. Yu, and Q. Zhang: *Proc. 2019 IEEE Int. Conf. Big Data* (IEEE, Los Angeles, 2019) 1971. <https://doi.org/10.1109/BIGDATA47090.2019.9006104>