



Carotid artery segmentation in computed tomography angiography (CTA) using multi-scale deep supervision with Swin-UNet and advanced data augmentation

Haodong Xie^{1#}, Hongmei Gu^{2#}, Minda Li², Li Zhu³, Tianle Wang³, Zhaotong Li¹, Huiqun Wu¹

¹Department of Medical Informatics, Medical School of Nantong University, Nantong, China; ²Department of Medical Imaging, Affiliated Hospital of Nantong University, Nantong, China; ³Department of Medical Imaging, Nantong First People's Hospital, Nantong, China

Contributions: (I) Conception and design: H Xie, H Wu; (II) Administrative support: H Gu; (III) Provision of study materials or patients: M Li; (IV) Collection and assembly of data: L Zhu; (V) Data analysis and interpretation: T Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Tianle Wang, MD. Department of Medical Imaging, Nantong First People's Hospital, 6 North Road, Nantong 226001, China. Email: wangtianle9192@163.com; Zhaotong Li, PhD; Huiqun Wu, MD, PhD. Department of Medical Informatics, Medical School of Nantong University, 19 Qixiu Road, Nantong 226001, China. Email: zhaotong_li@ntu.edu.cn; wuhuiqun@ntu.edu.cn.

Background: Carotid artery disease (CAD) is a serious disease caused by atherosclerosis, resulting in reduced cerebral blood flow and an increased risk of stroke. Traditionally, CAD diagnosis involves manual segmentation of computed tomography angiography (CTA) images, a time-consuming and complex process. This study aimed to address the need for an automated and accurate method for three-dimensional (3D) carotid artery segmentation using deep learning (DL) techniques.

Methods: A total of 214 CTA images from patients at the Affiliated Hospital of Nantong University and Nantong First People's Hospital were collected. The data were annotated using 3Dslicer software and calibrated by experienced radiologists. Preprocessing and augmentation of the CTA images were conducted using a novel window/level (W/L) adjustment method to enhance vascular imaging. The segmentation is performed using the Multi-Flux-Swin-Deepsup-UNet (MFSD-UNet) model, which incorporates multi-scale deep supervision and multi-flux fusion architecture. Performance was evaluated based on accuracy, dice coefficient, sensitivity, and specificity, and compared with state-of-the-art models. Ablation studies were conducted, removing the Swin transformer and deep supervision components to demonstrate the superiority of our method.

Results: The proposed model showed excellent performance, achieving an average dice coefficient of 0.9119 and an accuracy of 0.9819, outperforming the average dice coefficients of 0.8770 and 0.8910 for the two state-of-the-art models. Furthermore, it demonstrated high stability across various segmentation categories. Ablation studies revealed that removing the Swin transformer and deep supervision components resulted in a decrease in the dice coefficient to 0.8630 and 0.8371. Significant differences were observed when comparing these four models with MFSD-UNet ($P < 0.05$), and seven-fold cross-validations were performed on MFSD-UNet to demonstrate its robustness.

Conclusions: This study introduced a novel DL-based method for automatic 3D carotid artery segmentation from CTA images. The integration of Swin transformers, deep supervision mechanisms, and innovative data augmentation techniques significantly enhanced the accuracy and robustness of segmentation.

[^] ORCID: 0009-0005-0956-558X.

This method offers valuable support for the clinical diagnosis and treatment of CAD and exhibits great potential for future medical image segmentation.

Keywords: Carotid artery segmentation; deep learning (DL); carotid artery disease (CAD); computed tomography angiography (CTA)

Submitted Sep 28, 2024. Accepted for publication Feb 28, 2025. Published online Mar 28, 2025.

doi: 10.21037/qims-24-2087

View this article at: <https://dx.doi.org/10.21037/qims-24-2087>

Introduction

Carotid artery disease (CAD) is common in the elderly and can occur as carotid artery stenosis or occlusion, primarily affecting the internal carotid artery (ICA). Atherosclerosis is the main pathological cause of CAD, and this condition leads to the formation of plaques of varying degrees within the arterial wall (1-3). As these plaques progressively enlarge, they narrow the lumen of the carotid artery, potentially resulting in severe stenosis or complete occlusion (4,5). CAD can significantly reduce blood flow to the brain, substantially increasing the risk of stroke (6). Additionally, complications such as visual impairment and transient ischemic attacks may occur (7). One key issue that remains unresolved is the development of an automated, high-precision three-dimensional carotid artery segmentation model based on computed tomography angiography (CTA). Imaging examinations for CAD patients usually include ultrasound, CTA, magnetic resonance imaging (MRI), etc. In most cases, non-invasive, real-time, and fast ultrasound is the preferred examination method for the initial diagnosis of CAD, while CTA and MRI can provide higher accuracy and more comprehensive evaluation in more complex cases. Although “percent stenosis” measurement methods using different guideline-approved imaging modalities are required and decision thresholds need to be readjusted (8), the European Society of Cardiology (ESC) guidelines (latest version 2024) recommend that CTA is also important for patients with CAD (9). First, CTA’s high-resolution imaging can evaluate the surface morphology of plaques, including smooth, rough or ulcerated features (10), especially for severely calcified plaques (11), which are clearer, and the total volume of plaques and their subcomponents (such as fat, mixed and calcified areas) can be accurately calculated (12). Then, for severe stenosis (exceeding 70%), CTA has a sensitivity and specificity of 95% and 98% (13), respectively,

which exceeds ultrasound (90% and 94%) (14), with a high detection rate for mild stenosis too (15). Next, CTA is a sequential imaging that can perform three-dimensional (3D) reconstruction of the vascular lumen to better observe the anatomical location of the stenosis. Moreover, the combination of ultrasound and CTA can provide a more comprehensive diagnostic perspective and optimize the management of carotid stenosis. 3D carotid vascular lumen reconstruction based on CTA is also very meaningful (16). The reconstructed blood vessels can effectively identify the location and degree of stenosis, facilitate quantitative analysis, enable more accurate lesion positioning, and optimize surgical plans (17-19). Additionally, when combined with the reconstruction results, hemodynamic analysis can be performed to better assess the patient’s stroke risk, aid in formulating intervention strategies, and reduce the likelihood of stroke (20). Traditionally, these data are manually segmented using specialized software to create an accurate 3D model of the carotid artery (21). However, conventional 3D segmentation methods in medical imaging, particularly those based on CTA—such as the level set method with dual adaptive thresholds—face challenges due to complex anatomical structures, inconsistent image quality, and significant individual variations (22,23).

Consequently, there is an urgent need to develop a more effective method for creating personalized and accurate 3D models of the carotid artery from CTA images.

In recent years, deep learning (DL) has made significant advancements in the identification and segmentation of carotid arteries in medical imaging (24). Compared to traditional methods, DL-based medical image segmentation offers several advantages, including automatic feature extraction, end-to-end learning, and multi-scale information fusion (25,26). For instance, the Nested attention-guided Net (NAG-Net) model incorporates clinical knowledge by generating a visual attention map that highlights

areas of interest to physicians, enhancing the extraction of key features and achieving precise carotid artery segmentation (27). Additionally, end-to-end DL frameworks can automatically extract segmentation masks containing the carotid intima-media layer directly from the original image, while simultaneously predicting intima-media thickness and detecting arterial plaques, thus streamlining workflow efficiency (28). Moreover, transformer-based models such as Cross-Shaped Window (CSWin) outperform convolutional neural network (CNN) and hybrid models in carotid artery 3D ultrasound data, demonstrating the strength of transformers in multi-scale information fusion (29).

Despite the impressive performance of current medical image segmentation networks, there remains significant potential for improvement. For instance, integrating transformers (29) and employing data augmentation techniques (30,31) in carotid artery segmentation tasks could enhance the generalization capabilities of these models, particularly for challenging applications such as stuttering detection and chromosome histopathology. The U-Net network, without the use of transformers, achieved a dice score of only 0.55 in the automatic segmentation of carotid plaques (32). In contrast, an advanced model achieved a dice score of $95.8\% \pm 1.9\%$ at the media-adventitia boundary, though it relied on relatively simple data augmentation techniques (33). Additionally, while the dual-channel U-Net accurately segmented the common carotid artery (CCA) and ICA, it lacked emphasis on interpretability, validating its results solely through comparison with other methods (34).

This study aimed to develop a fully automatic 3D carotid artery segmentation model based on CTA images. First, an innovative data augmentation technique was applied to improve the model's robustness to complex geometric deformations. Next, a novel Swin-UNet network was developed to segment both local detail and global semantic information of the carotid artery. Finally, a multi-scale fusion deep supervision mechanism was integrated to enhance the model's ability to capture the intricate geometric structure of the artery. The resulting model achieves personalized, accurate, and efficient automatic segmentation of the carotid artery, with an average dice score of 0.9119. This high level of accuracy demonstrated the model's capacity to capture subtle variations in vascular structure, providing reliable support for clinical applications. We presented this article in accordance with the TRIPOD + AI reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-2087/rc>).

Methods

Data description

This study was conducted in accordance with the Declaration of Helsinki (revised in 2013) and received approval from the Ethics Committee of Nantong First People's Hospital (approval No. 2022-K105-01) and the Affiliated Hospital of Nantong University (approval No. 2023KT274). Since this was a retrospective analysis, written informed consent was waived. Data were collected retrospectively from 270 patients, including 150 treated at the Affiliated Hospital of Nantong University from July 11, 2022, to June 1, 2023, and 120 treated at Nantong First People's Hospital from December 19, 2023, to June 1, 2024. A total of 56 cases were excluded due to phase delay, artifacts, blurred images, and missing important structures, including 32 cases at the Affiliated Hospital of Nantong University and 24 cases at Nantong First People's Hospital.

In this study, the remaining 214 samples were randomly divided into training, validation, and test sets in an 8:1:1 ratio. The training set comprised 172 samples, with 94 from the Affiliated Hospital of Nantong University and 78 from Nantong First People's Hospital. Both the validation and test sets contained 21 samples each, with 12 from the Affiliated Hospital of Nantong University and 9 from Nantong First People's Hospital.

All CTA images were acquired using a dual-source CTA system (Siemens Somatom Definition Flash 128; Siemens, Forchheim, Germany) under the following scanning parameters: 120 kV, 100–400 mA, exposure time of 0.3 to 0.6 seconds, slice thickness of 0.7 mm, field of view of 350 mm × 350 mm, and matrix of 512 mm × 512 mm. The images were reconstructed using Sinogram Affirmed Iterative Reconstruction (SAFIRE) algorithms.

Data processing

CTA data were processed using 3D Slicer software (version 5.4.0). Four undergraduate medical students, under the guidance of two radiologists with 5 and 7 years of experience, annotated the vascular structures and exported the data in Neuroimaging Informatics Technology Initiative (NIFTI) format. Two physicians with 10 and 15 years of experience subsequently reviewed and corrected the annotations independently, without prior knowledge of the initial annotations. *Figure 1* presented the cross-sectional, coronal, and sagittal views of the CTA data, as well as the

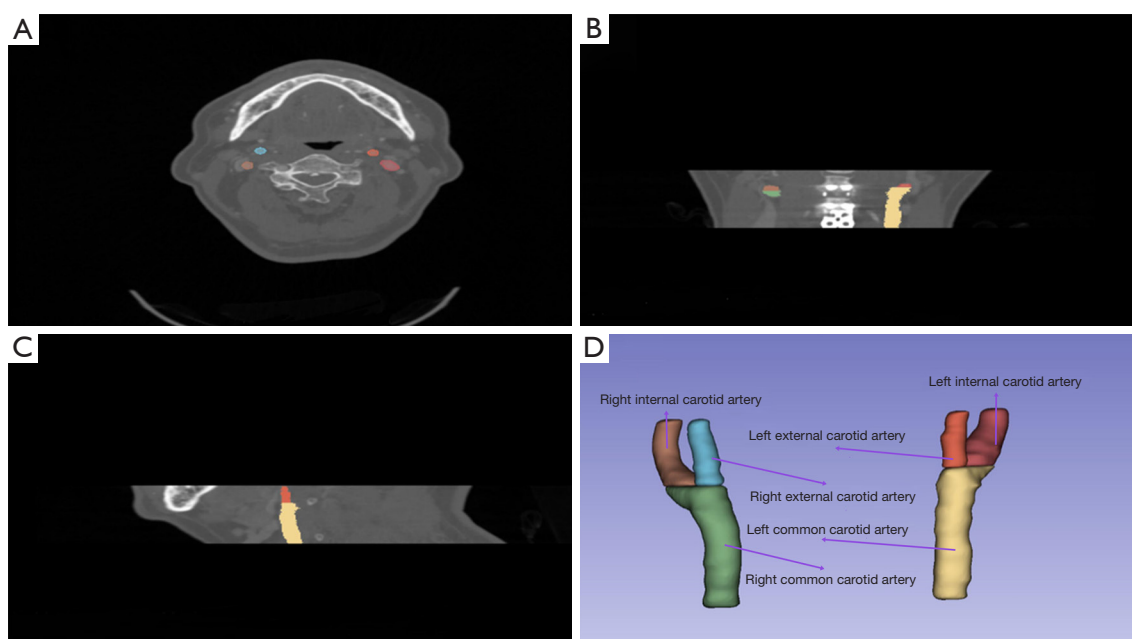


Figure 1 Visualization of CTA images from different perspectives. (A) Transverse section. (B) Coronal plane. (C) Sagittal plane. (D) 3D visualization of annotations. CTA, computed tomography angiography; 3D, three-dimensional.

3D visualization of the annotated vessels. The segmentation labels were categorized as follows: right common carotid artery (RCCA), left common carotid artery (LCCA), right external carotid artery (RECA), right internal carotid artery (RICA), left external carotid artery (LECA), and left internal carotid artery (LICA).

Data augmentation

To improve the clarity of vascular structures in the CTA images, a new data augmentation method was developed. This method was used to adjust the window/level (W/L) parameters of the CTA data to better highlight the carotid artery. As shown in *Figure 1A*, the carotid artery occupied a small proportion of the CTA image, and the contrast with surrounding fat tissue was minimal, which led to challenges in segmentation and irregular vascular contours.

The data augmentation process involved adjusting the initial W/L parameter (W0) in three steps. First, both the window and level parameters were increased by a factor of 1, resulting in a high-contrast CTA image (W1). Next, both parameters were decreased by a factor of 1, producing a low-contrast image (W2). Lastly, the window parameter remained unchanged, while the level was reduced by a factor of 1, yielding the adjusted image (W3). These W/L

adjustments may require slight modifications based on the specific dataset. The comparison of these four imaging types after W/L adjustments was shown in *Figure 2*.

Model architecture

This study presented a novel approach to CTA carotid vessel segmentation by leveraging a DL model that combines the Swin transformer, deep supervision, the channel attention mechanism (CAM), and the U-Net model, as illustrated in *Figure 3A*. Our model was designed to process four input channels corresponding to the four W/L-varied images shown in *Figure 2*, and utilized the Swin transformer for feature extraction instead of traditional convolutional layers. Class Activation Mapping was applied at each scale, and multiple auxiliary outputs were employed for deep supervision and multi-scale fusion. During down-sampling, MaxPooling was applied, while in the decoding stage, a recurrent module (RM) was used for feature connection and up-sampling, ultimately restoring matrix dimensions to achieve precise segmentation.

The Multi-Flux-Swin-Deepsup-UNet (MFSD-UNet) framework involves the first four MaxPooling layers and the last four convolutional layers, resembling the U-Net structure. It further incorporates four 3D convolutional

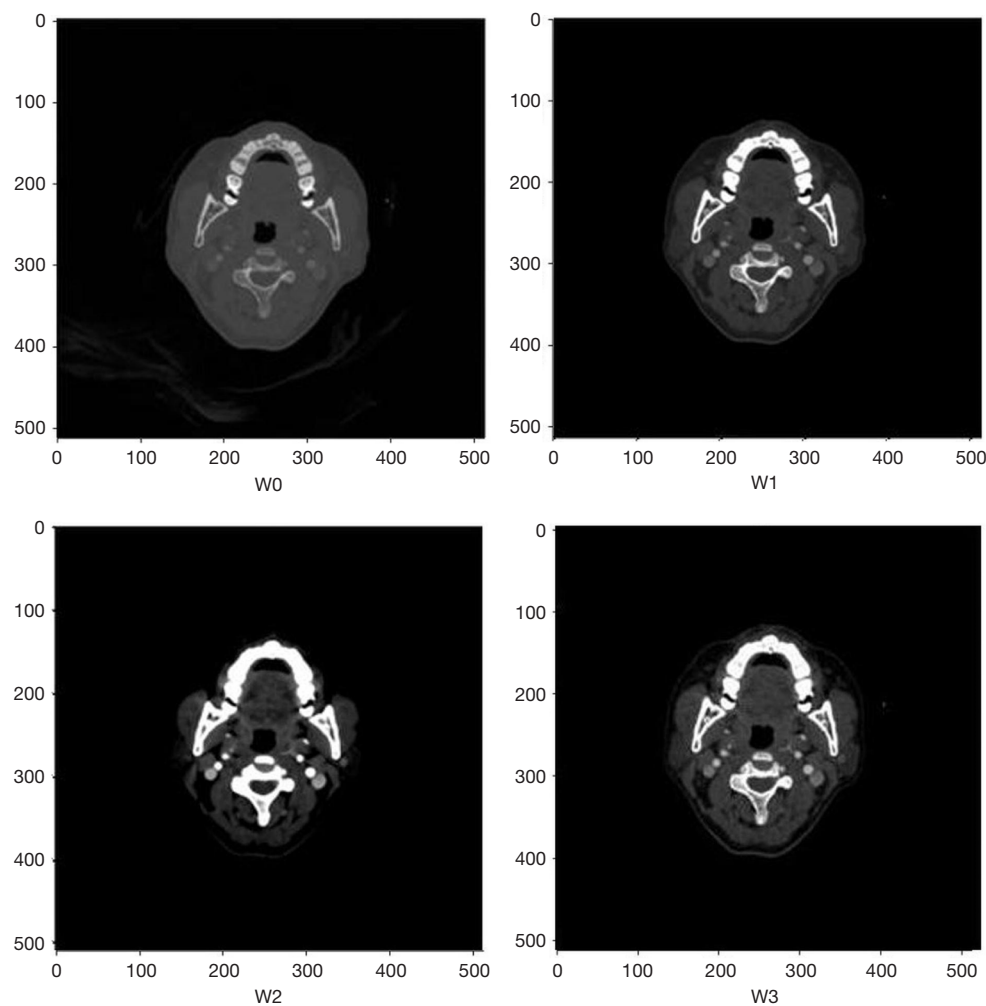


Figure 2 Four images and masks after W/L data augmentation. W0 represents the original CTA image. W1 is created by enlarging the W/L parameter by a factor of 1. W2 is formed by reducing the W/L parameter by a factor of 1 based on W0. W3 is achieved by keeping the window parameters unchanged while reducing the levels parameter by a factor of 1. X axis represents the length of the cross section (512 mm), Y axis represents the width (512 mm). CTA, computed tomography angiography; W/L, window/level.

layers for auxiliary outputs and four transposed convolutional layers for upsampling. The final 3D convolutional layer generates the output, which produces four auxiliary masks and a final mask for accurate carotid artery segmentation.

As shown in *Figure 3B*, the Swin transformer was designed as a hierarchical visual transformer architecture for image feature extraction. Its structure included patch partitioning, which divided the input image into non-overlapping patches of fixed size; linear embedding, which mapped each patch to a feature space of fixed dimension; and Swin transformer blocks. These blocks contained layer

normalization (LN), the window multi-head self-attention mechanism (W-MSA), sliding window multi-head self-attention (SW-MSA), and a multi-layer perceptron (MLP). Through patch merging, which reduced spatial dimensions, the Swin transformer captured global and local image information, significantly improving feature representation and the effectiveness of subsequent segmentation tasks.

In *Figure 3C*, CAM was used to enhance the model's ability to focus on task-relevant features while suppressing irrelevant or redundant information by adaptively adjusting channel weights. Its structure included global average pooling, which was used to compute global features for

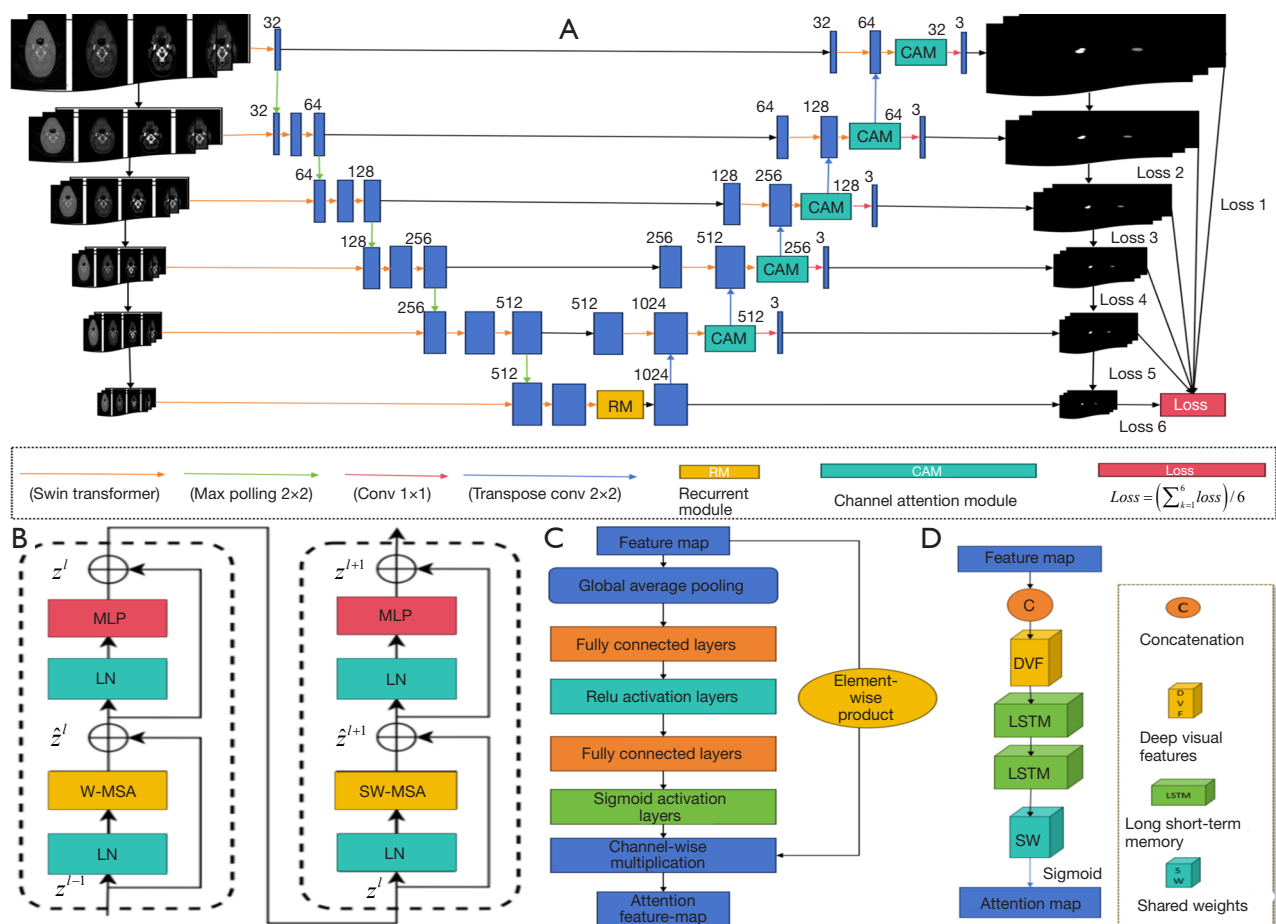


Figure 3 Model and its component architecture diagram. (A) is the overall model architecture diagram; (B) is the structure diagram of the Swin transformer block; (C) is the CAM structure diagram; (D) is the RM structure diagram. CAM, channel attention mechanism; LN, layer normalization; MLP, multi-layer perceptron; RM, recurrent module; SW-MSA, sliding window multi-head self-attention; W-MSA, window multi-head self-attention mechanism.

each feature map; fully connected layers, which generated weights for each channel through nonlinear transformations; sigmoid activation, which produced channel weights; and channel weighting, where weights were multiplied by the original feature map on a channel-by-channel basis. CAM increased attention to critical features, thereby improving the model's segmentation performance.

As depicted in Figure 3D, RM was used for feature connection and up-sampling during the decoding stage. Its structure includes feature connection (which links encoder features with those from the previous decoder layer) and recurrent operations. Recurrent neural networks (such as Long Short-Term Memory, LSTM) are employed for feature processing to capture temporal dependencies. By integrating multi-level feature information and maintaining

consistency during up-sampling, RM achieved highly accurate image segmentation.

Training parameters and environment

The experiments were conducted on a platform equipped with a Ray Tracing Texel eXtreme (RTX) 3090 Graphics Processing Unit (GPU) (24 GB of video memory) and an Intel Xeon Platinum 8255C Central Processing Unit (CPU) with a clock speed of 2.50 GHz and 43 GB of memory. The model was implemented using the PyTorch 1.12.1 framework and based on Python 3.8.6 (Python Software Foundation, Wilmington, Delaware, USA). The Adam optimizer was utilized, with a batch size of 32, an initial learning rate of 0.0001, and training over 500 epochs.

Table 1 Comparison of four evaluation indicators between our model and two advanced models

Models	Accuracy	Dice coefficient	Sensitivity	Specificity
C2F-DS-UNet*	0.9774±0.0020	0.8770±0.0050	0.9468±0.0025	0.9890±0.0025
Swin-UNet*	0.9889±0.0025	0.8910±0.0045	0.9361±0.0030	0.9931±0.0020
MFSD-UNet	0.9891±0.0028	0.9119±0.0060	0.9924±0.0028	0.9895±0.0030

Data are presented as mean ± SD. The C2F-Deepsup-UNet (C2F-DS-UNet) model is a UNet with “Channel-to-Feature” units and deep supervision, while the Swin-UNet combines the Swin transformer with UNet. *, denotes a significant ($P < 0.05$) difference between these two models and MFSD-UNet. MFSD-UNet, Multi-Flux-Swin-Deepsup-UNet; SD, standard deviation.

The loss value (L) is evaluated by the mean of the cross-entropy loss (L_{Cro}) and the dice loss (L_{Dice}), which can be written as:

$$L_{Cro} = -[Y \times \log(p) + (1 - Y) \times \log(1 - p)] \quad [1]$$

$$L_{Dice} = 1 - \left(2 \times |X \cap Y| / (|X| + |Y|) \right) \quad [2]$$

$$L = (L_{Cro} + L_{Dice}) / 2 \quad [3]$$

where p is the positive class probability predicted by the model, X represents the segmentation result, Y represents the ground truth.

The area stenosis rate calculation formula can be described as:

$$S_r = \frac{S_n - S_s}{S_n} \times 100\% \quad [4]$$

Where S_r represented the percentage of area reduction, S_n and S_s represented the luminal area when unaffected and the luminal area when stenotic, respectively.

Statistical analysis

To assess the performance of the carotid artery segmentation model, accuracy, Dice coefficient, sensitivity, and specificity, along with their means and standard deviations (SD), were calculated. Seven-fold cross-validation was implemented to enhance model reliability and mitigate potential bias.

Statistical analyses were performed using Statistical Package for the Social Sciences (SPSS) (version 21.0) and MedCalc (version 20.0.4). For comparison experiments with state-of-the-art models, analysis of variance (ANOVA) was used to test for statistical differences across all evaluation metrics. In ablation studies, paired t -tests were conducted. Two-tailed P values < 0.05 were considered statistically significant.

Results

Performance compared to state-of-the-art models

We selected two state-of-the-art models for comparison: C2F-Deepsup-UNet (C2F-DS-UNet) (35) and Swin-UNet (36). These models only applied basic data augmentation techniques, such as rotation and cropping, allowing us to highlight the superiority of our novel data augmentation approach combined with the MFSD-UNet model.

Our model showed a statistically significant improvement in all evaluation metrics compared to the other models ($P < 0.05$, Table 1). Specifically, MFSD-UNet outperformed in accuracy (mean: 0.9891), Dice coefficient (mean: 0.9119), and sensitivity (mean: 0.9924). While the specificity of our model (mean: 0.9895) was slightly lower than that of Swin-UNet (mean: 0.9931), this may reflect the inherent trade-off between sensitivity and specificity in the model design. Overall, our data augmentation method and model demonstrated clear advantages, with notable improvements in key performance metrics.

Visualize contrast between normal and narrow vessels

The segmentation results of different models are shown in Figure 4. Specifically, Recurrent Attention-UNet (RA-UNet) had spatial misalignment and under-segmentation in samples 2 and 3, as shown in column (I). Multi-Flux-Deepsup-UNet (MFD-UNet) showed under-segmentation in LCCA and RICA of samples 2 and 3, as shown in column (II). C2F-Deepsup-UNet (C2F-DS-UNet) showed under-segmentation in sample 1 and over-segmentation of moderately stenotic LICA and RICA in samples 3 and 4, as shown in column (III). Swin-UNet (IV) showed that only LICA was segmented in sample 1, and LICA and RICA in samples 3 and 4 were slightly over-segmented. Finally, the Multi-Flux-Swin-Deepsup-UNet (MFSD-UNet) model performed best and was basically consistent with the ground



Figure 4 Inference results of five models on four samples. “Image” refers to cross-sectional slices arranged from top to bottom, with the first two images showing normal segmentation of the vascular lumen, while the third and fourth images depict moderate stenosis of the left and right internal carotid arteries, respectively, due to the presence of mixed plaques; the yellow dotted line indicates the normal vascular area, and the green dotted line marks the area affected by stenosis. The term ‘Ground Truth’ refers to the annotated segmentation. C2F-DS-UNet, C2F-Deepsup-UNet; MFD-UNet, Multi-Flux-Deepsup-UNet; MFSD-UNet, Multi-Flux-Swin-Deepsup-UNet; RA-UNet, Recurrent Attention-UNet; Swin-UNet, Swin transformer-UNet.

results. The LICA of sample 1 had only a slight spatial misalignment, showing better performance than other models in normal and narrow conditions.

Performance of the DL model in each category

Figure 5 illustrates the performance of the MFSD-UNet model across different carotid artery categories. The RCCA category achieved the highest sensitivity (median near 1.00), specificity, and accuracy (both around 0.99), with a median Dice coefficient of approximately 0.92. However, the RICA category underperforms, with a median Dice coefficient around 0.88 and sensitivity of 0.98. It also contains low-value outliers in specificity and accuracy (both around 0.97), indicating some instability. Other categories, such as LCCA and RECA, perform well, achieving Dice coefficients of approximately 0.92 and 0.90, respectively, though they

exhibited some outliers in specificity and accuracy. LECA and LICA were generally stable but perform slightly below RCCA.

Ablation experiments

To validate the contributions of the Swin transformer, deep supervision mechanism, and data augmentation components in the proposed model, three ablation experiments were conducted. First, the Swin transformer was removed to generate the MFD-UNet model. Next, the deep supervision mechanism was eliminated from MFSD-UNet, producing the RA-UNet model. Finally, the proposed data augmentation method was compared against a common augmentation approach. The three models were evaluated using the same metrics to assess the significance of each component.

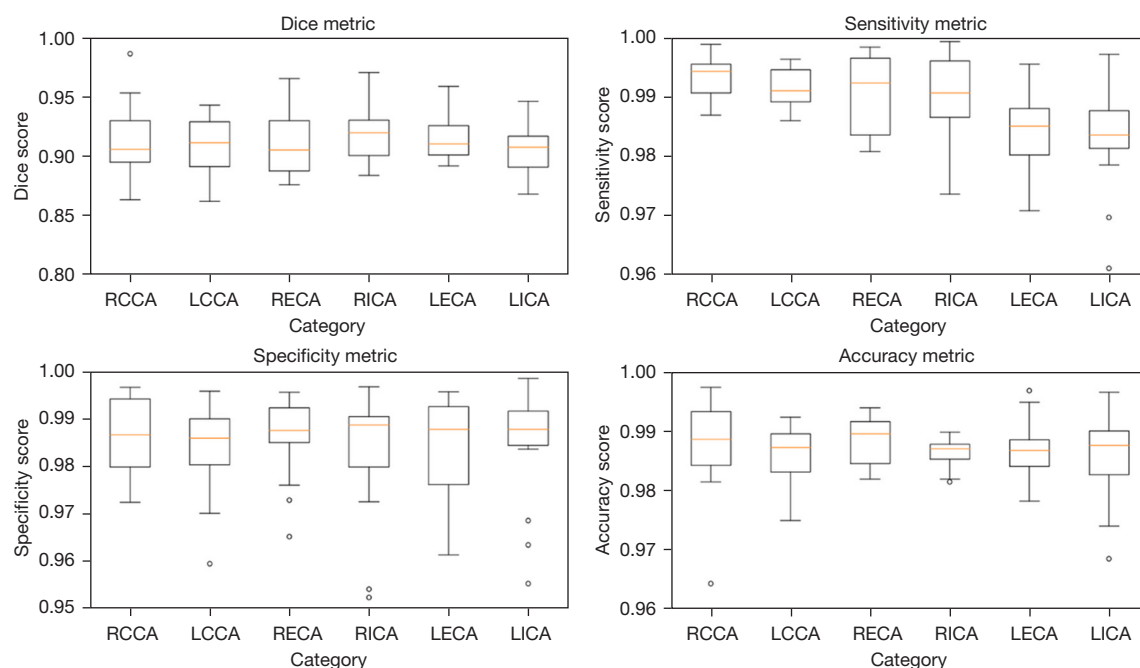


Figure 5 Model performance across different categories. LCCA, left common carotid artery; LECA, left external carotid artery; LICA, left internal carotid artery; RCCA, right common carotid artery; RECA, right external carotid artery; RICA, right internal carotid artery.

Table 2 Average of accuracy, dice, sensitivity and specificity of different models

Models	Accuracy	Dice coefficient	Sensitivity	Specificity
RA-UNet*	0.9818±0.0020	0.8371±0.0050	0.9966±0.0015	0.9883±0.0038
MFD-UNet*	0.9850±0.0025	0.8630±0.0055	0.9910±0.0037	0.9871±0.0042
MFSD-UNet	0.9891±0.0028	0.9119±0.0060	0.9924±0.0028	0.9895±0.0030

Data are presented as mean ± SD. *, denotes a significant ($P < 0.05$) difference between these two models and MFSD-UNet. MFD-UNet, Multi-Flux-Deepsup-UNet; MFSD-UNet, Multi-Flux-Swin-Deepsup-UNet; RA-UNet, Recurrent Attention-UNet; SD, standard deviation.

A statistically significant difference ($P < 0.05$) was observed in all evaluation metrics between the ablated models and the full MFSD-UNet model (Table 2). As shown in Table 2, the Dice coefficient and accuracy exhibited an increasing trend across the three models. MFSD-UNet performed the best on the test set, achieving a mean Dice coefficient of 0.9119 and a mean accuracy of 0.9891, while RA-UNet performed the worst, with a mean Dice coefficient of 0.8371. All three models demonstrated high sensitivity and specificity, both reaching 0.99 and 0.98, respectively, with a maximum SD not exceeding 0.01. These findings indicated that the Swin transformer and deep supervision mechanisms significantly enhanced model performance.

To validate the effectiveness of the new data augmentation method and MFSD-UNet model, the Brain Tumor

Challenge 2023 (BraTC 2023) and the Brain Tumor Challenge T1N Enhanced (BraTC TE) datasets were selected, totaling 214 samples. Each sample in the BraTC 2023 dataset includes four MRI sequences: T1-weighted contrast-enhanced (T1C), T1-weighted non-contrast (T1N), T2-weighted fat-suppressed (T2F), and T2-weighted (T2W). Each sample in the BraTC TE dataset was generated from the T1N sequence in BraTC 2023, with W/L data augmentation applied to create derived data similar to the other three sequences in BraTC 2023.

Table 3 showed that the sensitivity of the BraTC TE dataset (mean: 0.8754) was slightly higher than that of the BraTC 2023 dataset (mean: 0.8579), indicating improved detection of positive samples. The similar Dice coefficients of BraTC TE (mean: 0.7939) and BraTC 2023 (mean:

Table 3 Comparison of MFSD-UNet model on BraTC 2023 and BraTC T1N datasets

Datasets	Accuracy	Dice coefficient	Sensitivity	Specificity
BraTC 2023	0.9853±0.0009	0.8097±0.0069	0.8579±0.0066	0.9956±0.0004
BraTC TE	0.9642±0.0027	0.7939±0.0119	0.8754±0.0040	0.9888±0.0010

Data are presented as mean ± SD. BraTC 2023, Brain Tumor Challenge 2023; BraTC TE, Brain Tumor Challenge T1N Enhanced; MFSD-UNet, Multi-Flux-Swin-Deepsup-UNet; SD, standard deviation.

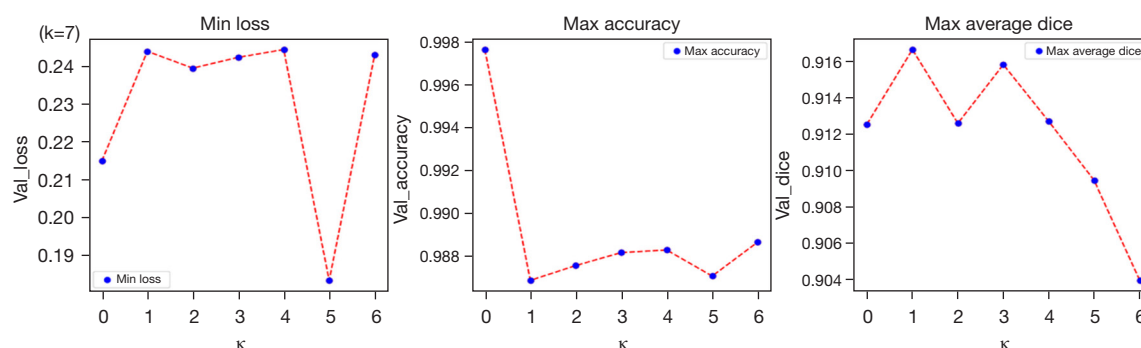


Figure 6 Evaluation metrics of the MFSD-UNet model under 7-fold cross-validation. The comprehensive loss, maximum accuracy, and maximum dice coefficient of the MFSD-UNet. MFSD-UNet, Multi-Flux-Swin-Deepsup-UNet.

0.8097) suggested that the W/L method effectively enhanced data variability, contributing to the model's robustness in handling complex samples.

The model reliability by cross-validation

The proposed model demonstrated strong stability in Dice and accuracy during seven -fold cross-validation, with the MFSD-UNet achieving a maximum Dice of 0.9167 and accuracy of 0.9976, as shown in *Figure 6*. These results confirmed the model's reliability and effective generalization across different datasets, highlighting its excellent segmentation performance and minimal misclassification, thereby reinforcing its potential for practical applications.

Performance of manual and our model measured in percentage of area reduction

For 35 stenosis sites in 21 test sets, the percentage reduction of their area relative to the unaffected normal lumen was measured by manual annotation and model (as shown in *Figure 7A*). Based on these data, we plotted a Bland-Altman plot in *Figure 7B* to assess the consistency between the two measurement methods. The results showed that the 95% consistency limit between the two was between 4.3% and

–5.8%, with an average difference of –0.7%. It is worth noting that the difference in case 10 was 4.7%, which might be due to the different morphology or composition of the plaque here, making it more difficult for the model to identify its boundaries, thus exceeding the 95% consistency limit. While the results indicated good consistency between the two methods, the irregular shape of the plaque and the subjectivity in the manual annotation process may lead to blurred boundary recognition when stenosis is present, causing the model to slightly underestimate the effective vascular area.

Discussion

This study proposed a W/L data augmentation method and developed a fully automatic 3D segmentation model for carotid CTA images, using a Swin transformer integrated with a deep supervision approach. The MFSD-UNet model achieved excellent performance, with a statistically significant improvement over four other models ($P < 0.05$, *Tables 1,2*), achieving a mean Dice coefficient of 0.9119 and an accuracy of 0.9819. This model showed promise in providing valuable technical support for future clinical diagnosis and treatment. In the carotid segmentation study, the DL model demonstrated strong consistency with

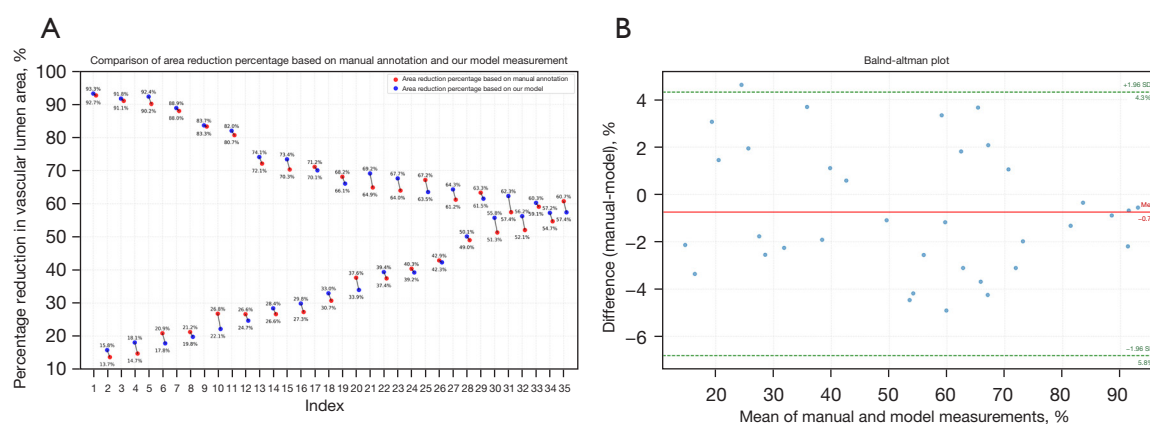


Figure 7 Comparison of area reduction percentage measured by manual annotation and MFSD-UNet model results. (A) A comparison chart of area reduction percentage. (B) A Bland-altman analysis chart. MFSD-UNet, Multi-Flux-Swin-Deepsup-UNet; SD, standard deviation.

experienced physicians in evaluating carotid stenosis and atherosclerotic plaques. The average evaluation time was significantly reduced to 27.3 ± 4.4 seconds, compared to the physicians' 296.8 ± 81.1 seconds (37). Additionally, the time for diagnosis and report writing decreased from 28.8 ± 5.6 to 12.4 ± 2.0 minutes (38). Notably, since an ICA angle $\leq 90^\circ$ predicts longer endovascular thrombectomy duration, DL was utilized to assess whether the ICA angle exceeded 90° , achieving a higher accuracy rate (39).

Previous studies on 3D segmentation of carotid arteries in medical images have largely focused on ultrasound and MRI images, with little attention to novel data augmentation methods or the use of transformers for efficient feature extraction. For instance, Jiang *et al.* (33) designed a dual-channel Unet for automatic segmentation of the common carotid artery and its branches but primarily focused on ultrasound imaging. When applied to CTA imaging, the simple Unet structure requires improvement in both segmentation accuracy and robustness. Other approaches, such as modifications to attention mechanisms or increasing the number of channels in ultrasound-based methods, may similarly struggle with CTA data processing (40,41).

Recent studies have introduced transformers, such as the CSWin transformer, for 3D segmentation of carotid arteries in ultrasound images, achieving Dice scores of $94.6\% \pm 3.0\%$, indicating that transformer architectures can effectively capture long-range dependencies in medical images (29). However, the three-dimensional nature of CTA images may limit the feature extraction capabilities of these architectures. Moreover, no research has yet proposed novel data augmentation methods for CTA images. Given that

data augmentation has been shown to significantly improve model performance in medical image segmentation (42,43), this study introduced a new data augmentation method combined with a Swin transformer to achieve superior segmentation of carotid arteries in CTA images.

Compared to state-of-the-art models, the MFSD-UNet model demonstrated an average Dice coefficient of 0.9119, which was significantly higher than the 0.8910 and 0.8770 achieved by two other models, likely due to advanced feature extraction and data augmentation techniques.

The model achieved an average sensitivity of 0.9924, demonstrating its exceptional ability to detect positive samples, which is crucial for clinical applications. As illustrated in Figure 4, other models had issues such as over-segmentation, under-segmentation, and spatial misalignment due to background noise, indicating a lack of robustness. In contrast, MFSD-UNet demonstrated high consistency with the ground truth in both narrow and normal cases, exhibiting only slight spatial misalignment in sample 1. This highlighted its excellent context-aware feature integration.

Figure 5 further demonstrated the model's performance, indicating that the RCCA, LCCA, and RECA categories all had median Dice coefficients above 0.92, with sensitivity and accuracy above 0.97. However, LECA and LICA exhibited slightly lower median Dice coefficients of 0.89 and 0.90, respectively, with sensitivities between 0.96–0.97. This drop may be attributed to the complexity of these categories or their indistinct features, making segmentation more challenging. The seven-fold cross-validation in Figure 6 underscored the robustness of our model across different

categories.

In the ablation experiments, we observed a consistent upward trend in both Dice coefficient and accuracy across the three models. Notably, MFSD-UNet outperformed all models, with a Dice coefficient of 0.9119, suggesting that the Swin transformer and deep supervision mechanisms are highly effective for capturing both global and local features, mitigating gradient vanishing, and improving segmentation accuracy. The W/L data augmentation method further enhanced the diversity of the dataset and the model's ability to learn distinctive features, increasing the sensitivity to positive samples (*Table 3*).

Despite these promising results, the study had some limitations. First, the relatively small dataset suggests that expanding to larger, multi-center datasets could enhance the model's generalization capabilities. Second, the network architecture was not compared with other mainstream segmentation networks, such as residual networks or convolutional neural networks, limiting a comprehensive evaluation of its advantages. Lastly, this study was based solely on carotid CTA imaging; future work could explore multimodal data, integrating MRI or other modalities to further improve segmentation accuracy.

Conclusions

This study presents a novel Swin-UNet model for efficient and accurate carotid artery segmentation, leveraging advanced DL techniques and data augmentation to improve feature extraction and segmentation precision. The model's results closely align with manual annotations, highlighting its potential for reliable carotid artery 3D modeling in clinical practice.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD + AI reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-24-2087/rc>

Funding: This work was supported by the Science and Technology Project of Nantong City (No. MS2023050) and University-Industry Collaborative Education Program of the Ministry of Education (No. 230700562265543),

Engineering Research Center of Integration and Application of Digital Learning Technology, Ministry of Education (No. 1311016), and Jiangsu Students' Platform for Innovation and Entrepreneurship Training Program (Nos. 202410304163E and 202410304169T).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-2087/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study has been approved by the Ethics Committee of the Affiliated Hospital of Nantong University (No. 2022-K105-01) and the Ethics Committee of Nantong First People's Hospital (No. 2023KT274), and the requirement for individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Bonati LH, Jansen O, de Borst GJ, Brown MM. Management of atherosclerotic extracranial carotid artery stenosis. *Lancet Neurol* 2022;21:273-83.
2. Tan J, Liang Y, Yang Z, He Q, Tong J, Deng Y, Guo W, Liang K, Tang J, Shi W, Yu B. Single-Cell Transcriptomics Reveals Crucial Cell Subsets and Functional Heterogeneity Associated With Carotid Atherosclerosis and Cerebrovascular Events. *Arterioscler Thromb Vasc Biol* 2023;43:2312-32.
3. Saba L, Nardi V, Cau R, Gupta A, Kamel H, Suri JS, Balestrieri A, Congiu T, Butler APH, Gieseg S, Fanni D, Cerrone G, Sanfilippo R, Puig J, Yang Q, Mannelli L, Faa G, Lanzino G. Carotid Artery Plaque Calcifications:

- Lessons From Histopathology to Diagnostic Imaging. *Stroke* 2022;53:290-7.
4. Bos D, Arshi B, van den Bouwhuijsen QJA, Ikram MK, Selwaness M, Vernooij MW, Kavousi M, van der Lugt A. Atherosclerotic Carotid Plaque Composition and Incident Stroke and Coronary Events. *J Am Coll Cardiol* 2021;77:1426-35.
 5. van Dam-Nolen DHK, van Egmond NCM, Dilba K, Nies K, van der Kolk AG, Liem MI, Kooi ME, Hendrikse J, Nederkoorn PJ, Koudstaal PJ, van der Lugt A, Bos D. Sex Differences in Plaque Composition and Morphology Among Symptomatic Patients With Mild-to-Moderate Carotid Artery Stenosis. *Stroke* 2022;53:370-8.
 6. Kopczak A, Schindler A, Sepp D, Bayer-Karpinska A, Malik R, Koch ML, Zeller J, Strecker C, Janowitz D, Wollenweber FA, Hempel JM, Boeckh-Behrens T, Cyran CC, Helck A, Harloff A, Ziemann U, Poli S, Poppert H, Saam T, Dichgans M. Complicated Carotid Artery Plaques and Risk of Recurrent Ischemic Stroke or TIA. *J Am Coll Cardiol* 2022;79:2189-99.
 7. Yaghi S, de Havenon A, Rostanski S, Kvernland A, Mac Grory B, Furie KL, Kim AS, Easton JD, Johnston SC, Henninger N. Carotid Stenosis and Recurrent Ischemic Stroke: A Post-Hoc Analysis of the POINT Trial. *Stroke* 2021;52:2414-7.
 8. Tekieli L, Mazurek A, Dzierwa K, Stefaniak J, Kablak-Ziembicka A, Knapik M, Moczulski Z, Banyś RP, Urbanczyk-Zawadzka M, Dabrowski W, Krupinski M, Paluszek P, Weglarz E, Wiewiórka Ł, Trystula M, Przewlocki T, Pieniazek P, Musialek P. Misclassification of carotid stenosis severity with area stenosis-based evaluation by computed tomography angiography: impact on erroneous indication to revascularization or patient (lesion) migration to a higher guideline recommendation class as per ESC/ESVS/ESO/SVS and CMS-FDA thresholds. *Postępy Kardiol Interwencyjnej* 2022;18:500-13.
 9. Mazzolai L, Teixido-Tura G, Lanzi S, Boc V, Bossone E, Brodmann M, et al. 2024 ESC Guidelines for the management of peripheral arterial and aortic diseases. *Eur Heart J* 2024;45:3538-700.
 10. Saba L, Agarwal N, Cau R, Gerosa C, Sanfilippo R, Porcu M, Montisci R, Cerrone G, Qi Y, Balestrieri A, Lucatelli P, Politi C, Faa G, Suri JS. Review of imaging biomarkers for the vulnerable carotid plaque. *JVS Vasc Sci* 2021;2:149-58.
 11. Zhao X, Hippe DS, Li R, Canton GM, Sui B, Song Y, Li F, Xue Y, Sun J, Yamada K, Hatsukami TS, Xu D, Wang M, Yuan C; CARE-II Study Collaborators. Prevalence and Characteristics of Carotid Artery High-Risk Atherosclerotic Plaques in Chinese Patients With Cerebrovascular Symptoms: A Chinese Atherosclerosis Risk Evaluation II Study. *J Am Heart Assoc* 2017;6:e005831.
 12. Saba L, Saam T, Jäger HR, Yuan C, Hatsukami TS, Saloner D, Wasserman BA, Bonati LH, Wintermark M. Imaging biomarkers of vulnerable carotid plaques for stroke risk prediction and their potential clinical implications. *Lancet Neurol* 2019;18:559-72.
 13. Adla T, Adlova R. Multimodality Imaging of Carotid Stenosis. *Int J Angiol* 2015;24:179-84.
 14. Jahromi AS, Cinà CS, Liu Y, Clase CM. Sensitivity and specificity of color duplex ultrasound measurement in the estimation of internal carotid artery stenosis: a systematic review and meta-analysis. *J Vasc Surg* 2005;41:962-72.
 15. Xiaofang H, Yajuan H, Pengfeng Y, Qiongjuan Z, Hongyuan S, Siyi S, Chujun S, Weixiang L, Tao L. Analysis of the reasons for the inconsistency between carotid vascular ultrasound and computed tomography angiography in the diagnosis of carotid artery stenosis. *International Medicine and Health Guidance News* 2024;30:1334-8.
 16. Mayer-Suess L, Peball T, Pereverzyev S Jr, Steiger R, Galijasevic M, Kiechl S, Knoflach M, Gizewski ER, Mangesius S. Cervical artery tortuosity-a reliable semi-automated magnetic resonance-based method. *Quant Imaging Med Surg* 2024;14:1383-91.
 17. van Dam-Nolen DHK, Truijman MTB, van der Kolk AG, Liem MI, Schreuder FHBM, Boersma E, Daemen MJAP, Mess WH, van Oostenbrugge RJ, van der Steen AFW, Bos D, Koudstaal PJ, Nederkoorn PJ, Hendrikse J, van der Lugt A, Kooi ME; PARISK Study Group. Carotid Plaque Characteristics Predict Recurrent Ischemic Stroke and TIA: The PARISK (Plaque At RISK) Study. *JACC Cardiovasc Imaging* 2022;15:1715-26.
 18. Yu S, Huo R, Qiao H, Ning Z, Xu H, Yang D, Shen R, Xu N, Han H, Chen S, Liu Y, Zhao X. Carotid artery perivascular adipose tissue on magnetic resonance imaging: a potential indicator for carotid vulnerable atherosclerotic plaque. *Quant Imaging Med Surg* 2023;13:7695-705.
 19. Poloni S, Bozzetto M, Du Y, Aiani L, Goddi A, Fiorina I, Remuzzi A. Velocity vector comparison between vector flow imaging and computational fluid dynamics in the carotid bifurcation. *Ultrasonics* 2023;128:106860.
 20. Leng X, Lan L, Ip HL, Abrigo J, Scalzo F, Liu H, et al. Hemodynamics and stroke risk in intracranial atherosclerotic disease. *Ann Neurol* 2019;85:752-64.
 21. Müller MD, Lyrer P, Brown MM, Bonati LH. Carotid

- artery stenting versus endarterectomy for treatment of carotid artery stenosis. *Cochrane Database Syst Rev* 2020;2:CD000515.
22. Zhou R, Guo F, Azarpazhooh MR, Spence JD, Ukwatta E, Ding M, Fenster A. A Voxel-Based Fully Convolution Network and Continuous Max-Flow for Carotid Vessel-Wall-Volume Segmentation From 3D Ultrasound Images. *IEEE Trans Med Imaging* 2020;39:2844-55.
 23. Luo L, Liu S, Tong X, Jiang P, Yuan C, Zhao X, Shang F. Carotid artery segmentation using level set method with double adaptive threshold (DATLS) on TOF-MRA images. *Magn Reson Imaging* 2019;63:123-30.
 24. Zhou R, Fenster A, Xia Y, Spence JD, Ding M. Deep learning-based carotid media-adventitia and lumen-intima boundary segmentation from three-dimensional ultrasound images. *Med Phys* 2019;46:3180-93.
 25. Cao L, Wang Q, Hong J, Han Y, Zhang W, Zhong X, Che Y, Ma Y, Du K, Wu D, Pang T, Wu J, Liang K. MVI-TR: A Transformer-Based Deep Learning Model with Contrast-Enhanced CT for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *Cancers (Basel)* 2023.
 26. Han Y, Holste G, Ding Y, Tewfik A, Peng Y, Wang Z. Radiomics-Guided Global-Local Transformer for Weakly Supervised Pathology Localization in Chest X-Rays. *IEEE Trans Med Imaging* 2023;42:750-61.
 27. Huang Q, Zhao L, Ren G, Wang X, Liu C, Wang W. NAG-Net: Nested attention-guided learning for segmentation of carotid lumen-intima interface and media-adventitia interface. *Comput Biol Med* 2023;156:106718.
 28. Gago L, Vila MDM, Grau M, Remeseiro B, Igual L. An end-to-end framework for intima media measurement and atherosclerotic plaque detection in the carotid artery. *Comput Methods Programs Biomed* 2022;223:106954.
 29. Lin Y, Huang J, Xu W, Cui C, Xu W, Li Z. Method for Carotid Artery 3-D Ultrasound Image Segmentation Based on CSWin Transformer. *Ultrasound Med Biol* 2023;49:645-56.
 30. Sheikh SA, Sahidullah M, Hirsch F, Ouni S. Advancing Stuttering Detection via Data Augmentation, Class-Balanced Loss and Multi-Contextual Deep Learning. *IEEE J Biomed Health Inform* 2023;27:2553-64.
 31. Faryna K, van der Laak J, Litjens G. Automatic data augmentation to improve generalization of deep learning in H&E stained histopathology. *Comput Biol Med* 2024;170:108018.
 32. Meshram NH, Mitchell CC, Wilbrand S, Dempsey RJ, Varghese T. Deep Learning for Carotid Plaque Segmentation using a Dilated U-Net Architecture. *Ultrason Imaging* 2020;42:221-30.
 33. Jiang M, Chiu B. A Dual-Stream Centerline-Guided Network for Segmentation of the Common and Internal Carotid Arteries From 3D Ultrasound Images. *IEEE Trans Med Imaging* 2023;42:2690-705.
 34. Jiang M, Zhao Y, Chiu B. Segmentation of common and internal carotid arteries from 3D ultrasound images based on adaptive triple loss. *Med Phys* 2021;48:5096-114.
 35. Wang J, Yu Y, Yan R, Liu J, Wu H, Geng D, Yu Z. Coarse-to-fine multiplanar D-SEA UNet for automatic 3D carotid segmentation in CTA images. *Int J Comput Assist Radiol Surg* 2021;16:1727-36.
 36. Pecco N, Della Rosa PA, Canini M, Nocera G, Scifo P, Cavoretto PI, Candiani M, Falini A, Castellano A, Baldoli C. Optimizing Performance of Transformer-based Models for Fetal Brain MR Image Segmentation. *Radiol Artif Intell* 2024;6:e230229.
 37. Zhu Y, Chen L, Lu W, Gong Y, Wang X. The application of the nnU-Net-based automatic segmentation model in assisting carotid artery stenosis and carotid atherosclerotic plaque evaluation. *Front Physiol* 2022;13:1057800.
 38. Fu F, Shan Y, Yang G, Zheng C, Zhang M, Rong D, Wang X, Lu J. Deep Learning for Head and Neck CT Angiography: Stenosis and Plaque Classification. *Radiology* 2023;307:e220996.
 39. Nageler G, Gergel I, Fangerau M, Breckwoldt M, Seker F, Bendszus M, Möhlenbruch M, Neuberger U. Deep Learning-based Assessment of Internal Carotid Artery Anatomy to Predict Difficult Intracranial Access in Endovascular Recanalization of Acute Ischemic Stroke. *Clin Neuroradiol* 2023;33:783-92.
 40. Naik V, Gamad RS, Bansod PP. Carotid artery segmentation in ultrasound images and measurement of intima-media thickness. *Biomed Res Int* 2013;2013:801962.
 41. Groves LA, VanBerlo B, Veinberg N, Alboog A, Peters TM, Chen ECS. Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction. *Int J Comput Assist Radiol Surg* 2020;15:1835-46.
 42. Sanford TH, Zhang L, Harmon SA, Sackett J, Yang D, Roth H, Xu Z, Kesani D, Mehralivand S, Baroni RH, Barrett T, Girometti R, Oto A, Purysko AS, Xu S, Pinto PA, Xu D, Wood BJ, Choyke PL, Turkbey B.

- Data Augmentation and Transfer Learning to Improve Generalizability of an Automated Prostate Segmentation Model. *AJR Am J Roentgenol* 2020;215:1403-10.
43. Liu W, Zhuo Z, Liu Y, Ye C. One-shot segmentation of

novel white matter tracts via extensive data augmentation and adaptive knowledge transfer. *Med Image Anal* 2023;90:102968.

Cite this article as: Xie H, Gu H, Li M, Zhu L, Wang T, Li Z, Wu H. Carotid artery segmentation in computed tomography angiography (CTA) using multi-scale deep supervision with Swin-UNet and advanced data augmentation. *Quant Imaging Med Surg* 2025;15(4):3161-3175. doi: 10.21037/qims-24-2087