Original Article

Comparative analysis of U-Mamba and no new U-Net for the detection and segmentation of esophageal cancer in contrast-enhanced computed tomography images

Yifan Hu[1,2#^], Yi Zhang[3#^], Zeyu Tang[1^], Xin Han[4^], Huimin Hong[5^], Lin Kong[1,2^], Zhihan Xu[6^], Shanshan Jiang[7^], Xiaojin Yu[1^], Lei Zhang[3^]

[1]Department of Radiology, Dongtai People's Hospital, Yancheng, China; [2]Department of Radiology, Nantong University Affiliated Hospital, Nantong, China; [3]Department of Radiology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; [4]Department of Thoracic Surgery, Dongtai People's Hospital, Yancheng, China; [5]Department of Pathology, Dongtai People's Hospital, Yancheng, China; [6]Department of CT Collaboration, Siemens Healthineers, Shanghai, China; [7]Department of Clinical and Technical Support, Philips Healthcare, Xi'an, China

Contributions: (I) Conception and design: Y Hu, L Zhang; (II) Administrative support: X Yu; (III) Provision of study materials or patients: Y Zhang, X Han, H Hong; (IV) Collection and assembly of data: Z Tang, Y Zhang, L Kong; (V) Data analysis and interpretation: Y Hu, Z Xu, S Jiang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Lei Zhang, MD. Department of Radiology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, #650 Songjiang Rd, Shanghai 201600, China. Email: lei.Zhang2@shgh.cn; Xiaojin Yu, MD. Department of Radiology, Dongtai People's Hospital, Nantong University Faculty of Health Science, 2# Kangfuxi Rd, Dongtai 224200, China. Email: yxjfhl@163.com.

Background: Radiomics research in esophageal cancer (EC) has made considerable advancements. However, manual segmentation, which is relied upon in clinical and scientific workflows, remains time-consuming and inconsistent. This study aimed to develop and validate a deep learning (DL) model for the automatic detection and segmentation of EC lesions in contrast-enhanced computed tomography (CT) images.

Methods: We retrospectively collected the CT data of patients with EC confirmed by pathology from January 2017 to September 2021 at three hospitals and from individuals with a healthy esophagus. Manual labeling of EC lesions was conducted, and DL networks [no new U-Net (nnU-Net) and U-Mamba] were trained for automatic segmentation. An optimal threshold volume for EC lesion detection was determined and integrated into the postprocessing module. The performance of DL models was evaluated in internal, external, and thin-slice image test cohorts and compared with diagnoses by radiologists. The sensitivity, specificity, accuracy, Dice similarity coefficient (DSC), and Hausdorff distance (HD) were calculated.

Results: A total of 871 patients (564 males) were included, with a median age of 67 years. DL models exhibited no significant difference from radiologists' diagnoses (P>0.05). Median DSC values for the internal, external, and thin-slice cohorts were 0.795, 0.811, and 0.797, respectively, with a corresponding HD of 9.733 mm, 7.860 mm, and 8.168 mm. An intraclass correlation coefficient greater than 0.7 was observed for 97.2% of the radiomic features extracted from thin-slice images.

^ ORCID: Yifan Hu, 0000-0002-0770-1067; Yi Zhang, 0000-0002-6540-0051; Zeyu Tang, 0009-0009-6674-7573; Xin Han, 0009-0009-3371-9074; Huimin Hong, 0009-0002-1447-7486; Lin Kong, 0007-0005-6749-3377; Zhihan Xu, 0000-0003-4057-8299; Shanshan Jiang, 0000-0003-1721-573X; Xiaojin Yu, 0000-0001-8554-6739; Lei Zhang, 0000-0002-0952-3057.

Conclusions: The DL methods demonstrated exceptional sensitivity and robustness in EC detection and segmentation on contrast-enhanced CT images, not only reducing missed EC diagnoses but also providing radiologists with consistent lesion annotations.

Keywords: Esophageal cancer (EC); object detection; automatic segmentation; no new U-Net (nnU-Net); U-Mamba

Submitted Jun 04, 2024. Accepted for publication Jan 13, 2025. Published online Feb 26, 2025.
doi: 10.21037/qims-24-1116
View this article at: https://dx.doi.org/10.21037/qims-24-1116

Introduction

Esophageal cancer (EC) is the seventh most prevalent and the sixth most deadly form of cancer worldwide (1). To meet clinical needs and improve patient survival rates, researchers have been applying radiomics technology to analyze quantitative features of tumor images, making notable progress. However, most radiomics studies rely on manual segmentation, which is inefficient and lacks stability (2,3).

In clinical practice, automatic lesion segmentation is fundamental to artificial intelligence (AI)-driven target detection tasks, including extending lesion classification. The widespread use of commercial AI software for breast tumor and lung nodule detection highlights the critical role of automatic lesion segmentation in AI-driven target detection tasks (4,5). However, the automatic segmentation of EC has not yet reached a mature stage of application.

EC often grows longitudinally along the esophageal wall, and when the tumor is small, thick-slice (reconstructed slice thickness of 5 mm) images provide very limited information. In contrast, thin-slice (reconstructed slice thickness of 1 or 0.625 mm) computed tomography (CT) images can offer a wealth of information for research (6). However, the manual annotation of thin-slice images by professional physicians is impractical due to the nearly 10-fold increase in workload compared to thick-slice images.

The rapid development of convolutional neural networks (CNNs) in medical image segmentation in recent years may provide a means to solving this issue, driving the transition from traditional manual radiomics to radiomics based on deep learning (DL) models (7,8). Sui *et al.* and Takeuchi *et al.* used the V-Net and visually-aware biomimetic network (VB-Net) and visual geometry group 16 network architecture for esophageal segmentation, respectively, and achieved EC target detection by measuring its thickness (9,10). However, their models were only internally validated and did not segment the tumor lesions. Both the mature U-Net-based no new U-Net (nnU-Net) and the novel U-Mamba adaptive segmentation model have demonstrated excellent performance in the field of medical segmentation. We hypothesized that nnU-Net and U-Mamba could achieve automated detection and segmentation of EC with higher stability and consistency than those of manual segmentation, addressing clinical needs and improving radiomics feature reproducibility for research applications. We present this article in accordance with the TRIPOD+AI reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-24-1116/rc).

Methods

Dataset

The multicenter study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of Dongtai People's Hospital (No. 2020-dtry-K-16). Ethical approval of this study was filed for record-keeping at Nantong University Affiliated Hospital and Shanghai General Hospital, who were informed and agreed with the study. Informed consent was waived given the retrospective design and the use of anonymized data.

The positive cohort included patients admitted for the treatment of EC between January 2017 and January 2021 at Nantong University Affiliated Hospital (Hospital 1) and Dongtai People's Hospital (Hospital 2) from March 2021 to September 2021 at Shanghai General Hospital (Hospital 3). The negative cohort comprised patients who underwent enhanced chest CT scans at these hospitals for indications other than esophageal tumors. These patients underwent follow-up chest CT scans more than 6 months after the first CT scan for unrelated reasons and with no esophageal abnormalities detected in the subsequent imaging.

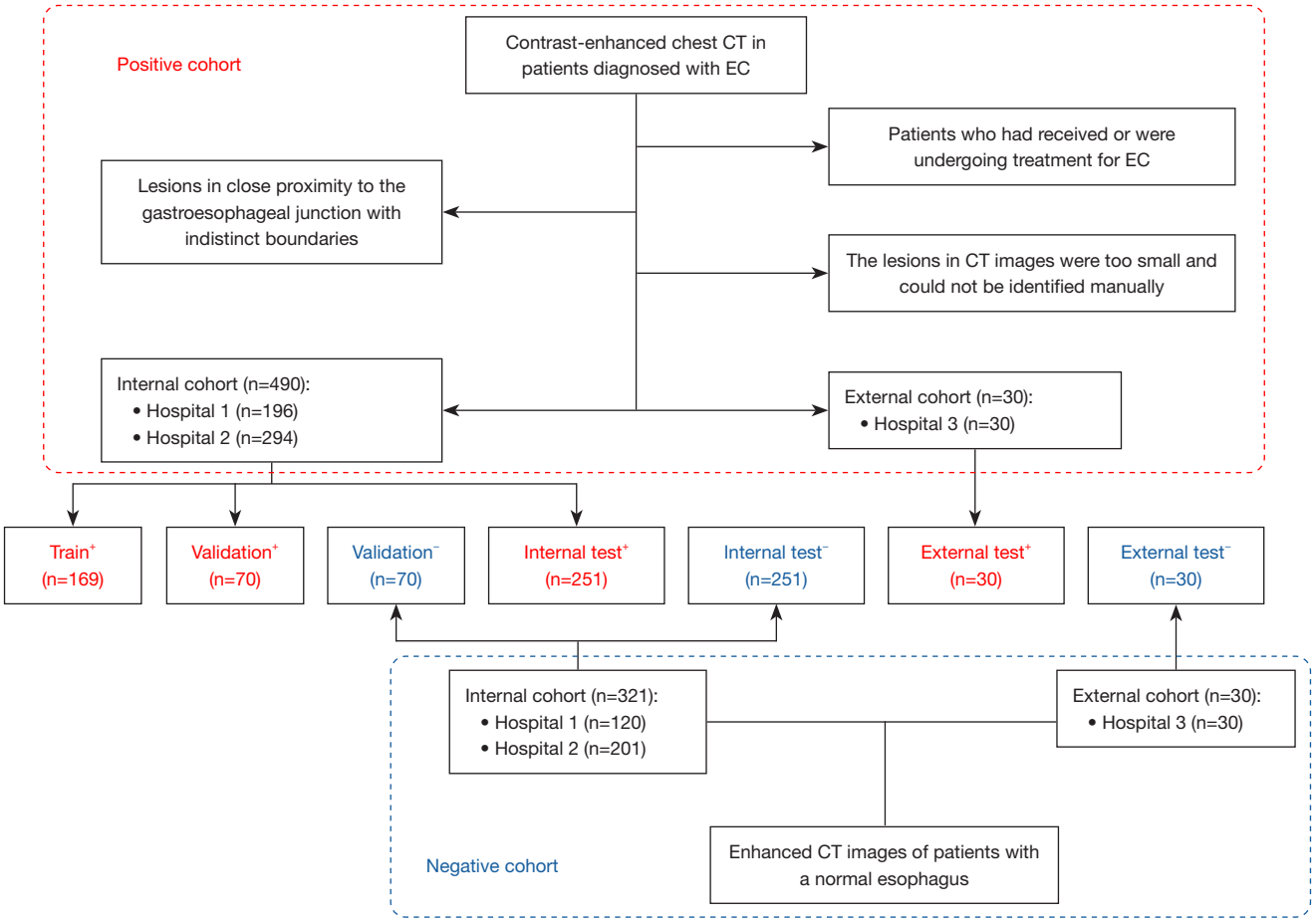

Figure 1 Schematic diagram of the enrollment process and data set allocation. +, the positive cohort; –, the negative cohort. EC, esophageal cancer; CT, computed tomography; Hospital 1, Nantong University Affiliated Hospital; Hospital 2, Dongtai People's Hospital; Hospital 3, Shanghai General Hospital.

The inclusion criteria for the positive cohort were as follows: (I) confirmed diagnosis of EC through surgical pathology and (II) availability of complete thoracic contrast-enhanced CT imaging data. Meanwhile, the exclusion criteria were as follows: (I) patients who had received or were undergoing treatment for EC prior to the enhanced CT scan; (II) lesions in CT images too small to be identified manually; and (III) lesions in close proximity to the gastroesophageal junction with indistinct boundaries.

A total of 871 participants were included, comprising 316 participants from Hospital 1, 495 from Hospital 2, and 60 from Hospital 3. Cases recruited prior to June 2019 were included in the training and validation cohorts, while those recruited after June 2019 were included in the test cohort. The specific distribution of patients is illustrated in *Figure 1*.

Image acquisition and labeling

The scanning range extended from the thoracic inlet to the lower edge of the bilateral adrenal glands. Arterial phase images were acquired upon triggering of the CT monitoring threshold, with the region of interest (ROI) designated as the aorta, and venous phase images were collected 35 seconds after the initiation of the arterial phase acquisition. The scanning parameters were as follows: collimation of 256×0.625 mm, tube voltage of 120 kV, slice thickness and spacing of 5 mm, reconstruction slice thickness of 5 and 1 mm, and a matrix of 512×512 pixels.

To assess the DL model's robustness, this study analyzed data from three centers with different imaging equipment and contrast agent protocols. The scanning equipment used

at Hospital 1 included the Brilliance iCT and Brilliance 64 scanners (Philips Healthcare, Best, the Netherlands), the SOMATOM Force scanner (Siemens Healthineers, Erlangen, Germany), and the Revolution 1.5 M3C Global scanner (GE Healthcare, Chicago, IL, USA). A bolus of 50–60 mL of contrast agent (iopromide, 370 mg iodine/mL) was injected at a rate of 2–3 mL/s into the antecubital vein and was followed by a 40-mL saline flush. Oral ingestion of the contrast agent diluted at a ratio of 1:10 was administered before scanning.

At Hospital 2, the equipment used was the SOMATOM Definition AS (Siemens Healthineers). The contrast agent used was iodixanol (80 mL; 300 mg iodine/mL), with the same injection rate and flush technique as those described above. No oral contrast agent was taken before scanning.

Hospital 3 used the SOMATOM Flash and Force (Siemens Healthineers). The contrast agent used was iodixanol (80 mL; 300 mg iodine/mL), injected at a rate of 3 mL/s, with the same injection and flush technique used as those mentioned above. No oral contrast agent was taken before scanning.

The patients' venous phase Digital Imaging and Communications in Medicine (DICOM) images were imported into 3D Slicer software version 4.13 (https://download.slicer.org/) via the picture archiving and communication system for delineation of the ROIs. This process was performed by an attending physician with over 6 years of relevant work experience. During delineation, blood vessels and lymph nodes were excluded, and care was taken to avoid areas containing air and contrast agent. The tumor's inner edge was manually traced layer by layer, and upon completion, the ROI was confirmed by another attending physician.

Automatic EC detection and segmentation framework

In the Python 3.9.12 environment (Python Software Foundation, Wilmington, DE, USA), we employed two advanced medical image segmentation frameworks: nnU-Net version v. 2 (11) and U-Mamba (12) for model training. Both frameworks are designed to automatically tune all hyperparameters according to the specific characteristics of the dataset in use. For our investigation, we chose two architectures from nnU-Net [two-dimensional (2D) U-Net and 3D full resolution (FullRes) U-Net] and four from U-Mamba (2D U-Mamba_Bot, 3D U-Mamba_Bot, 2D U-Mamba_Enc, and 3D U-Mamba_Enc) to conduct training and validation.

The nnU-Net framework comprises two architectures: 2D U-Net and 3D FullRes U-Net. Two-dimensional U-Net is specifically designed for 2D image segmentation, employing a traditional encoder-decoder structure with skip connections that fuse high-resolution and low-resolution features, thereby enhancing segmentation accuracy. Three-dimensional FullRes U-Net expands this architecture into the 3D domain, making it particularly effective for volumetric data analysis via the ability to capture spatial information across multiple slices.

The U-Mamba framework offers four architectures: 2D and 3D U-Mamba_Bot and 2D and 3D U-Mamba_Enc. U-Mamba_Bot integrates U-Mamba blocks at the bottleneck for high-level feature extraction. The U-Mamba block introduces a novel mechanism for simultaneously capturing both short- and long-range dependencies within the data. By integrating CNNs and state-space models (SSMs), this block combines the strengths of local feature extraction and dynamic system modeling. SSMs, originally designed for sequential data analysis, are employed in this case to encode long-range spatial dependencies efficiently. Their recursive structure enables the modeling of complex spatial relationships without requiring the resource-intensive operations of attention mechanisms. Meanwhile, the CNNs in the block extract high-resolution local features, complementing the SSMs' global spatial modeling capabilities. This synergistic design allows the U-Mamba block to maintain computational efficiency while effectively capturing detailed spatial context. The detailed network architecture is depicted in *Figure 2*.

Training parameters were standardized across all networks, and the number of training epochs was fixed at 1,000. We used both cross-entropy and Dice loss functions and employed the Adam optimizer, incorporating a strategy for dynamic adjustment of the learning rate (initial learning rate 0.01). This consistent approach facilitated a fair comparison of the network architectures under identical training conditions. To monitor the training process, we plotted the loss and pseudo-Dice curves (Figure S1).

Postprocessing modules

Physiological peristalsis of the esophagus can lead to segmental wall thickening, which may cause DL models to misinterpret these changes as tumors on CT images. To mitigate the risk of misclassifying physiological esophageal changes as pathological findings, we developed a lesion volume-based postprocessing module. In our validation

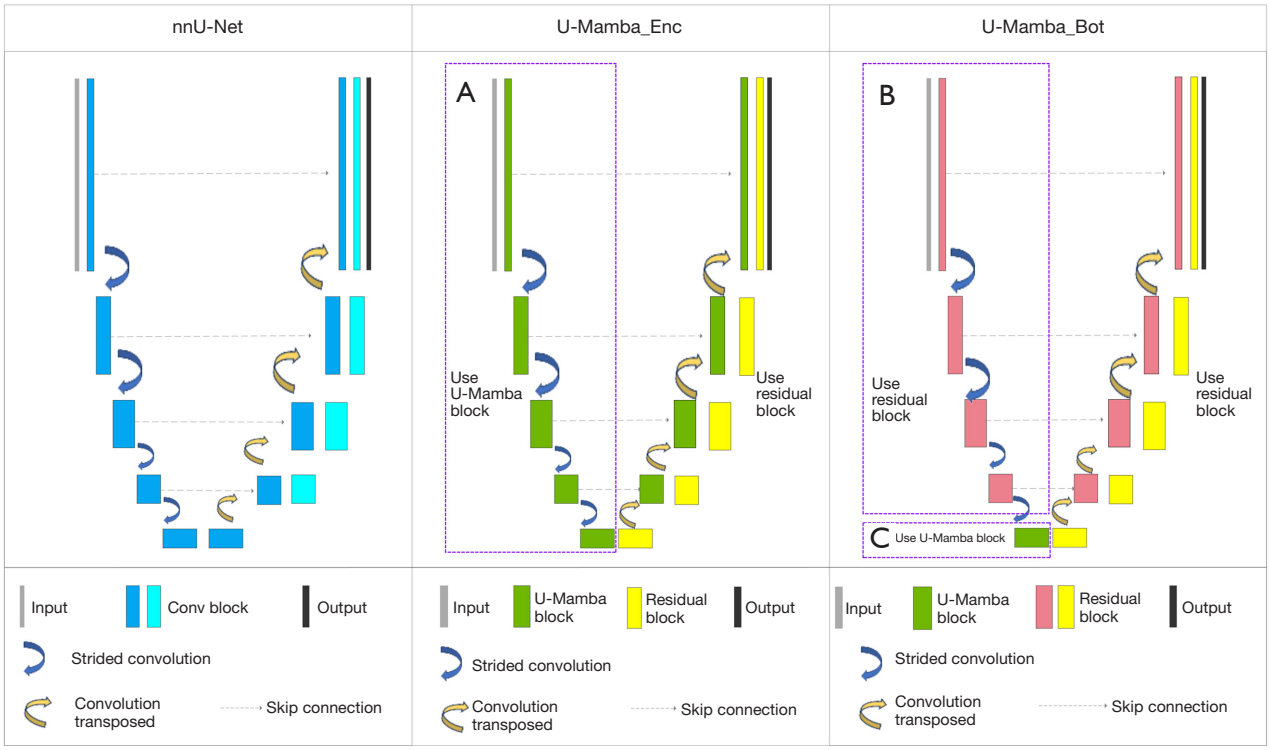

Figure 2 nnU-Net, U-Mamba-Enc, and U-Mamba-Bot architecture diagram. (A) U-Mamba_Enc employs U-Mamba blocks in the encoder. (B) U-Mamba_Bot employs residual block in the encoder. (C) U-Mamba_Bot places U-Mamba blocks at the bottom layer, concentrating feature learning at the model's bottleneck. The distinction between 2D and 3D versions of the models lies in the data dimension (2D slice-based versus 3D volumetric segmentation), while the network architecture remains the same. 2D, two-dimensional; 3D, three-dimensional; Conv, convolution; nnU-Net, no new U-Net.

cohort, receiver operating characteristic (ROC) curves and the Youden index were used to establish the optimal cutoff volume for identifying EC, with segments with volumes below this threshold being classified as nontumorous (*Figure 3*).

Evaluation metrics and statistical analysis

Objective detection performance evaluation

The DL model's object detection performance was evaluated using sensitivity, specificity, and accuracy metrics. The McNemar test was employed to compare the DL model's detections to radiologists' interpretations. Two radiologists independently interpreted patient images in internal and external cohorts, with the patient grouping being concealed.

Automatic segmentation performance evaluation

Using the validation cohort's outcomes, we selected the best-performing nnU-Net and U-Mamba models for testing on the internal and external cohorts. True-positive EC cases identified by the DL model were included in further analysis. Segmentation performance was evaluated using the Dice similarity coefficient (DSC) and Hausdorff distance (HD), with radiologists' manual segmentations being used as the reference. We conducted a quantitative analysis of tumor dimensions—including length, height, width, and volume—and of radiomics texture features to evaluate the consistency between the U-Mamba and nnU-Net segmentation outputs. To extract features from CT images, we used the PyRadiomics software package in Python 3.9.12. This process included first-order features, shape, gray-level co-occurrence matrix (GLCM), gray-level size zone matrix (GLSZM), gray-level run-length matrix (GLRLM), neighborhood gray-tone difference matrix (NGTDM), and gray-level dependence matrix (GLDM) features. For detailed information, visit the PyRadiomics documentation online (https://pyradiomics.readthedocs.io/

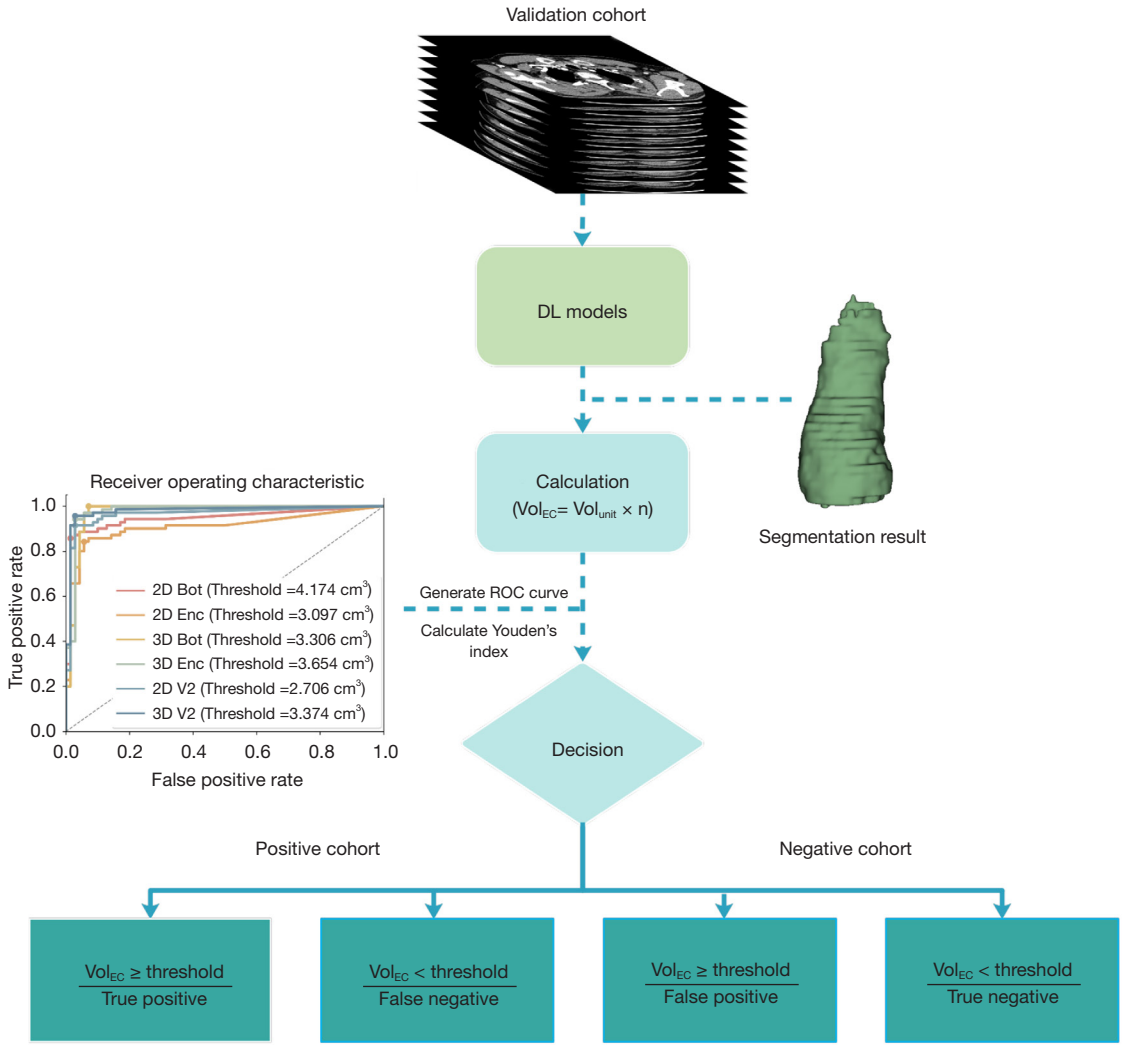

Figure 3 Postprocessing pipeline. Vol_{EC} represents the EC lesion volume, while Vol_{unit} is the volume of each voxel, calculated as the product of voxel spacing and slice spacing (obtained from the DICOM file). The variable n represents the number of times a "1" appears in the binary image output by the model. The Youden index is calculated as the sum of sensitivity and specificity, minus one. Finally, the Vol_{EC} corresponding to the maximum Youden index on the ROC curve is identified, serving as the model's threshold volume. EC, esophageal cancer; DL, deep learning; 2D, two-dimensional; 3D, three-dimensional; ROC, receiver operating characteristic; DICOM, Digital Imaging and Communications in Medicine.

en/latest/) or the Table S1.

Statistical analysis

Statistical analyses of the research data were conducted using MedCalc version 20.019 (MedCalc Software, Ostend, Belgium). Continuous variables are presented as the mean ± standard deviation ($\bar{x}$±s) if they followed a normal distribution (as determined by the Shapiro-Wilk test) or as median and interquartile range if they did not.

Results

Study population

A total of 871 people (564 males) were enrolled in this study, with a median age of 67 (IQR 59–73) years. There

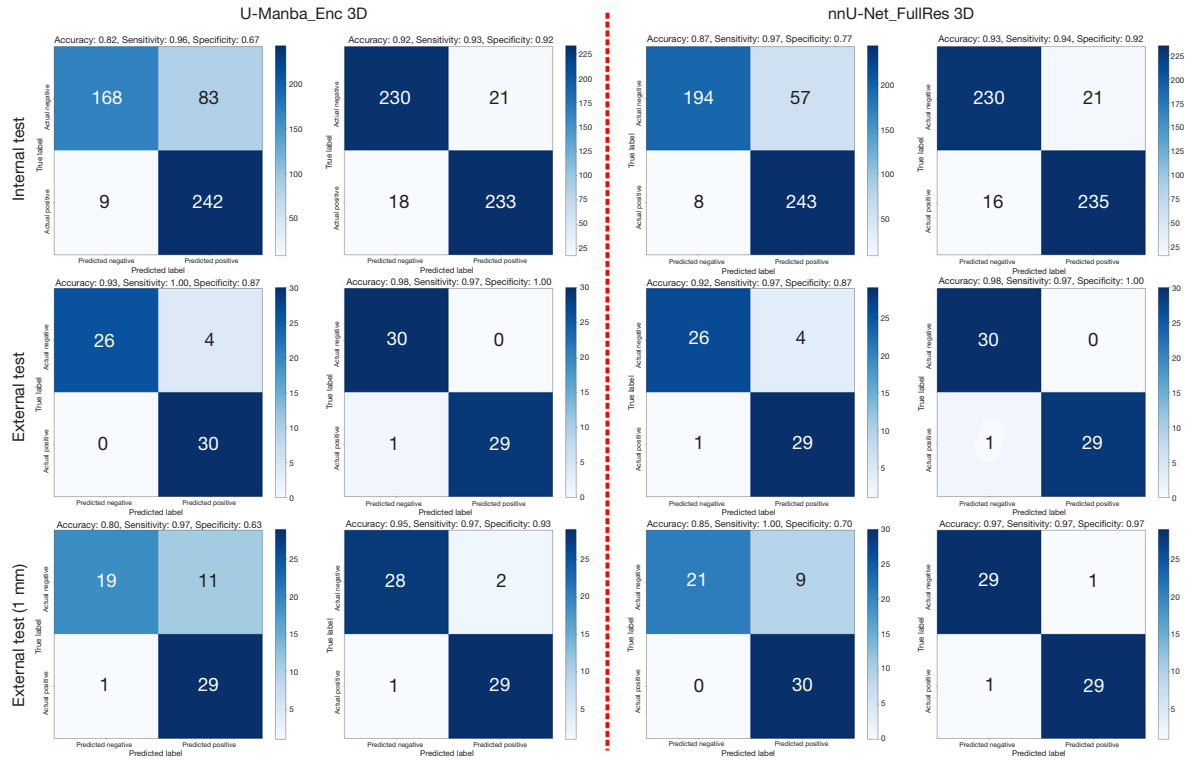

Figure 4 Comparison of the target detection performance of the two models in different cohorts with and without the threshold volume postprocessing module. The column on the left represents the results before application, and the column on the right represents the results after application. It can be seen that after the addition of the threshold volume postprocessing module, the number of cases of incorrect identification of normal esophagus as tumors was significantly reduced, indicating significant improvement in the specificity of the model. 3D, three-dimensional; nnU-Net, no new U-Net.

are 520 cases in the positive cohort and 351 cases in the negative cohort.

Model thresholds and performance

The cutoff values (threshold volumes) for each model in the tuning cohort are illustrated in *Figure 3*. By setting a threshold volume, we substantially improved the specificity of the model (*Figure 4*). In the validation cohort, the optimal models for both object detection and segmentation tasks were the 3D models within the nnU-Net architecture (*Table 1*). Across the U-Mamba and nnU-Net frameworks, 3D models notably outperformed their 2D counterparts.

Object detection task

The 3D FullRes model from the nnU-Net and the 3D U-Mamba_Bot model achieved the highest accuracy in the validation cohort, with a score of 0.964. Therefore, for subsequent analysis, these two models were chosen for validation in the internal and external test cohorts, respectively. The performance of the DL models compared to manual interpretation on the object detection task in the internal and external test cohorts is shown in the *Table 2*. The McNemar test indicated no statistically significant differences between the two DL models and manual interpretation results.

Segmentation task

For the segmentation task, we included cases for further analysis in which esophageal tumors were correctly identified as true positives by both models. These cases consisted of 227 samples from the internal test set, 29 from the external test set, and 28 from the external test set with a 1-mm slice thickness. In terms of segmentation performance, the nnU-Net 3D FullRes model exhibited the best results across all cohorts [DSC 0.795–0.811;

Table 1 The results of each model in the validation cohort

Task	Evaluation metrics	nnU-Net		U-Mamba			
		2D	3D FullRes	2D Bot	3D Bot	2D Enc	3D Enc
Target detection	Threshold volume (cm^3)	2.706	3.374	4.174	3.306	3.097	3.654
	Sensitivity	0.914	0.958	0.857	1*↑	0.843	0.943
	Specificity	0.971	0.971	0.986*↑	0.929	0.929	0.971
	Accuracy	0.943	0.964*↑	0.921	0.964*↑	0.886	0.957
Segmentation	DSC	0.778	0.813*↑	0.772	0.796	0.772	0.800
	HD95 (mm)	11.072	7.248*↓	10.986	10.202	11.698	9.809

*, the best performance, with arrows indicating whether higher (↑) or lower (↓) values are better. DSC, Dice similarity coefficient; HD95, 95% Hausdorff distance; 2D, two-dimensional; 3D, three-dimensional; FullRes, full resolution.

Table 2 Comparison of results for models in the test cohorts

Dataset	Observer	Target detection			McNemar test	Segmentation	
		Sensitivity	Specificity	Accuracy	P	DSC	HD95 (mm)
Internal test	Attending physician	0.980	0.928	0.954	–	–	–
	nnU-Net	0.936*↑	0.916*↑	0.926*↑	0.215	0.795*↑	9.733*↓
	U-Mamba	0.928	0.916*↑	0.922	0.121	0.774	11.177
External test	Attending physician	1	0.967	0.983	–	–	–
	nnU-Net	0.967*↑	1*↑	0.983*↑	0.5	0.811*↑	7.860*↓
	U-Mamba	0.967*↑	1*↑	0.983*↑	0.5	0.794	8.072
External test (1 mm)	Attending physician	1	0.967	0.983	–	–	–
	nnU-Net	0.967*↑	0.967*↑	0.967*↑	1	0.797*↑	8.168*↓
	U-Mamba	0.967*↑	0.933	0.950	1	0.792	11.387

The McNemar test compared the target detection results of the DL model with the manual interpretation results. *, the best performance, with arrows indicating whether higher (↑) or lower (↓) values are better. DSC, Dice similarity coefficient; HD95, 95% Hausdorff distance; DL, deep learning; nnU-Net, no news U-Net.

95% Hausdorff distance (HD95) 7.860–9.733]. *Figure 5* illustrates the segmentation results of these two DL models on thick (5 mm) and thin (1 mm) slice images in the external testing cohort. A detailed comparison between all six models is presented in Figure S2.

Quantitative analysis and radiomic features

Quantitative measurements of the segmentation outcomes from both models demonstrated good stability, particularly in terms of volumetric measurements [intraclass correlation coefficient (ICC) 0.954–0.997]. Furthermore, radiomic features extracted from thin-slice images exhibited stronger consistency than did those derived from thick-slice images, with 97.2% of features showing an ICC greater than 0.700 for thin-slice images as compared to 91.1% for thick-slice images (*Table 3*). To intuitively display the consistency of radiomic features, we plotted the measurement results of all features into ICC scatter plots and included them in the Figure S3.

Discussion

This study developed and tested a DL model for the automated detection and segmentation of esophageal tumors. The 3D FullRes model, based on the nnU-

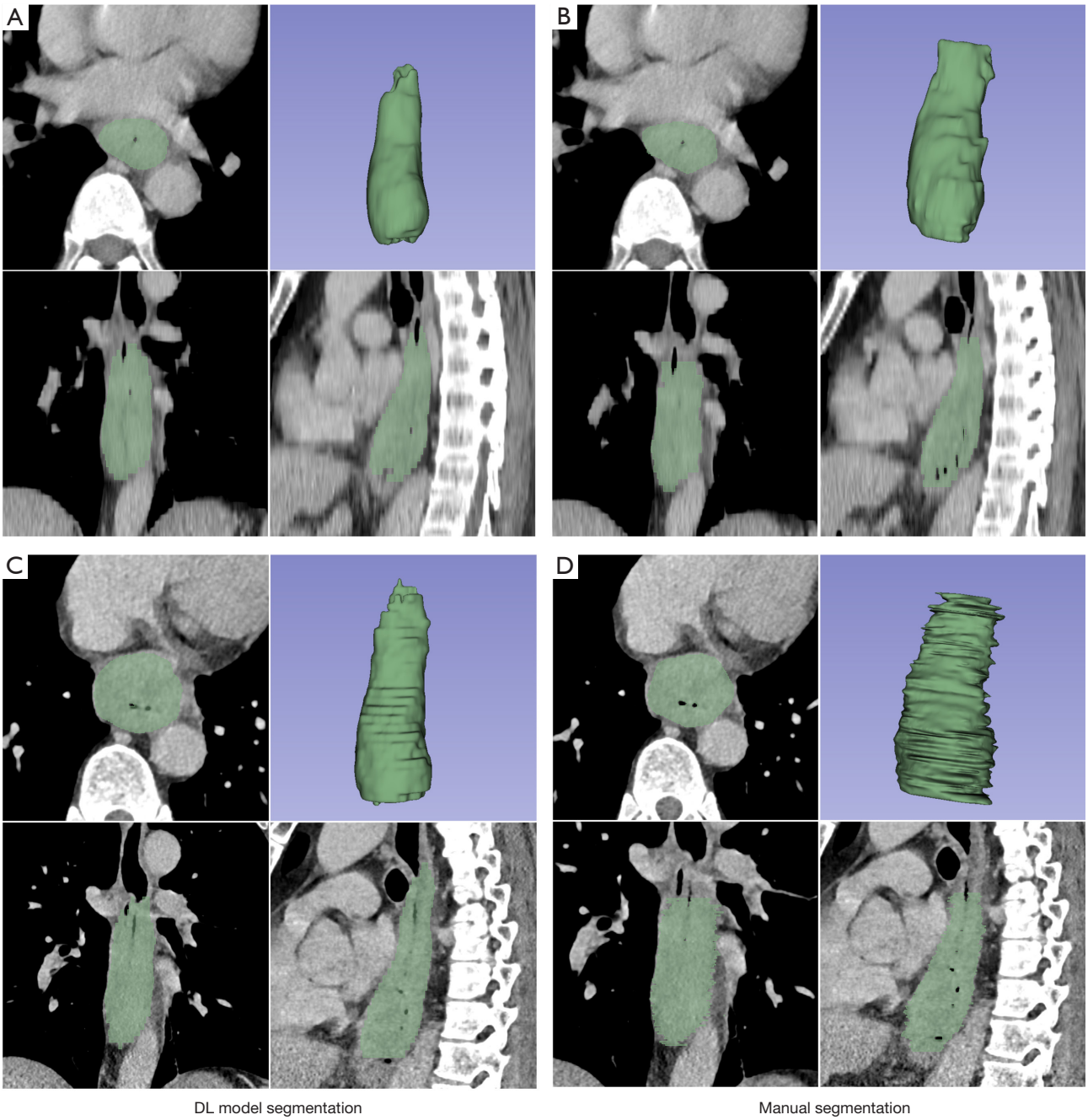

Figure 5 Comparison of DL (nnU-Net) model segmentation and manual segmentation in thick (5 mm) and thin (1 mm) layer images. (A,B) Thick-slice CT images. (C,D) Thin-slice images. (D) Manual segmentation in thin-slice images not only increases the workload several times but also renders the boundaries more jagged, while (C) with DL model segmentation, this defect is greatly improved, and the three-dimensional volume of interest shows smoother boundaries for DL model segmentation. DL, deep learning; nnU-Net, no new U-Net; CT, computed tomography.

Table 3 Quantitative analysis of DL model segmentation of EC lesions

Cohort	Models	Anteroposterior diameter (cm)	Width (cm)	Length (cm)	Volume (cm^3)	Radiomics features (ICC >0.700)
Internal test	nnU-Net	3.315 (2.712, 3.841)	2.627 (2.260, 3.260)	5.000 (3.500, 7.000)	17.480 (9.757, 27.123)	1,092/1,158 (94.3%)
	U-Mamba	3.262 (2.730, 3.822)	2.617 (2.242, 3.188)	4.500 (3.500, 6.500)	16.863 (10.512, 25.770)	
	ICC	0.818	0.909	0.792	0.981	
External test	nnU-Net	3.217±0.813	2.761 (2.300, 3.052)	5.000 (3.500, 6.000)	16.108 (9.655, 28.612)	1,055/1,158 (91.1%)
	U-Mamba	3.241±0.801	2.781 (2.492, 3.094)	5.000 (3.500, 5.500)	16.190 (10.411, 28.077)	
	ICC	0.938	0.943	0.818	0.997	
External test (1 mm)	nnU-Net	3.385±0.783	2.829 (2.425, 3.214)	5.500 (4.500, 7.150)	21.366 (12.046, 30.887)	1,126/1,158 (97.2%)
	U-Mamba	3.346±0.825	2.773 (2.284, 3.162)	5.250 (4.000, 6.625)	16.782 (11.088, 31.151)	
	ICC	0.895	0.803	0.745	0.954	

Normally distributed continuous variables are presented as the mean ± standard deviation, nonnormally distributed continuous variables as the median and interquartile range (Q1–Q3), and categorical variables as the frequency and percentage (%). DL, deep learning; EC, esophageal cancer; nnU-Net, no new U-Net; ICC, intraclass correlation coefficient.

Net framework, exhibited the highest performance, accurately detecting EC and achieving precise tumor segmentation. The DL model demonstrated exceptional robustness, and variations in image slice thickness, types of imaging equipment, and the use of oral contrast agents showed minimal impact on the model's performance. The performance differences between the U-Mamba and nnU-Net models were marginal, and the models demonstrated considerable stability in the quantitative measurement of tumors.

This study compared two leading adaptive segmentation networks, nnU-Net and U-Mamba. nnU-Net, based on the U-Net architecture, has consistently ranked highly in segmentation challenges such as the Medical Segmentation Decathlon (13) and has been validated across numerous organ segmentation tasks (14,15). U-Mamba, incorporating U-net and mamba modules, combines convolutional layers with SSMs to capture both local features and long-range dependencies (16). Sui *et al.* and Lin *et al.* employed an improved VB-Net (a modified version of V-Net) and nnU-Net after segmentation, respectively, to detect esophageal tumors on CT images by measuring the average diameter and wall thickness of the esophagus (10,17). However, these methods fail to distinguish the physiological peristalsis-induced thickening of the esophageal wall. We found that changes in the esophageal wall due to normal peristalsis affected fewer layers, and by setting a threshold volume, physiological peristalsis-induced errors could be largely excluded.

Our target detection model outperformed previous approaches. In the test cohort, the sensitivity, specificity, and accuracy were 0.936–0.967, 0.916–1, and 0.926–0.983, respectively, surpassing the results of Lin *et al.*'s model (0.900, 0.880, and 0.882, respectively) (17). Furthermore, our DL models achieved better performance in segmenting EC lesions compared to Amyar *et al.*'s multitask model, which reported a best DSC of 0.79 for EC segmentation (18).

Radiomics has been applied in the diagnosis and treatment of EC for years (19-22), and various mathematical models have become key tools for processing data and have strong robustness. However, human involvement in tumor segmentation introduces uncertainties, particularly due to the indistinct boundaries of esophageal tumors in CT images. Manual delineation is not only time-consuming but also prone to inconsistency. For instance, in Li *et al.*'s study, only approximately 80% of radiomic features assessed by ICC exhibited satisfactory consistency (23), while features demonstrating both high interobserver reproducibility and test-retest reliability accounted for just 66% (24). Our research confirms the excellent stability of DL models in EC segmentation, with more than 90% of features showing good ICC values (>0.70) in thick-slice images and even higher, over 95%, in thin-slice images. Thus, developing tools for CT image-based detection and segmentation of EC tumors is of practical significance for advancing the application of radiomics in EC, particularly as DL models can easily perform tumor segmentation in thin-slice images, thereby enhancing the reproducibility of radiomic features and, consequently, the predictive performance of radiomic

models (25,26).

Our study validated the robustness and generalizability of DL models in both the internal and external test cohorts, confirming their stability in segmenting EC. Whether measuring tumor diameters, volumes, or extracting radiomic features, DL models provide fully automated and highly reliable measurements. These models not only support stable outcome assessments in therapy for EC but also offer consistent lesion segmentation for radiomics research, contributing to a unified standard and high-quality annotations for radiologic image databases. This not only facilitates the transition of radiomics from research to clinical practice but also provides precise volumes of interest for future interactive reports which can connect directly to hypertext descriptions in reports (6).

It should be noted, however, that due to the significant lack of noncontrast data in this retrospective study, we were only able to validate the DL models using contrast-enhanced CT images, without assessing their applicability to non-contrast images. This limitation restricts conclusions regarding its applicability.

Conclusions

DL models reliably identified and segmented EC lesions in contrast-enhanced CT images. The segmentation results from the nnU-Net and U-Mamba models demonstrated excellent stability, and thus these models are capable of providing radiologists with high-quality annotations of EC lesions.

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-24-1116/rc

Funding: This study was supported by the Nantong University Clinical Medicine Special Project (No. 2022LQ001) and the Jiangsu Vocational College of Medicine School-Regional Collaborative Innovation Research Project (No. 20239515).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-24-1116/coif). Z.X. was an employee of Siemens Healthineers throughout her involvement in the study. S.J. was an employee of Philips Healthcare China (2021–2024) throughout her involvement in the study and then became an employee of Bayer Healthcare China. The other 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. This multicenter study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of Dongtai People's Hospital (No. 2020-dtry-K-16). Ethical approval of this study was filed for record-keeping at Nantong University Affiliated Hospital and Shanghai General Hospital, who were informed of and agreed with the study. Informed consent was waived given the retrospective design and the use of anonymized data.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Xie CY, Pang CL, Chan B, Wong EY, Dou Q, Vardhanabhuti V. Machine Learning and Radiomics Applications in Esophageal Cancers Using Non-Invasive Imaging Methods-A Critical Review of Literature. Cancers (Basel) 2021;13:2469.
3. Kocak B, Yardimci AH, Nazli MA, Yuzkan S, Mutlu S, Guzelbey T, Sam Ozdemir M, Akin M, Yucel S, Bulut E, Bayrak ON, Okumus AA. REliability of consensus-based segMentatIoN in raDiomic feature reproducibility

(REMIND): A word of caution. Eur J Radiol 2023;165:110893.
4. Vachon CM, Scott CG, Norman AD, Khanani SA, Jensen MR, Hruska CB, Brandt KR, Winham SJ, Kerlikowske K. Impact of Artificial Intelligence System and Volumetric Density on Risk Prediction of Interval, Screen-Detected, and Advanced Breast Cancer. J Clin Oncol 2023;41:3172-83.
5. Tang TW, Lin WY, Liang JD, Li KM. Artificial intelligence aided diagnosis of pulmonary nodules segmentation and feature extraction. Clin Radiol 2023;78:437-43.
6. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, Folio LR, Summers RM, Rubin DL, Lungren MP. Preparing Medical Imaging Data for Machine Learning. Radiology 2020;295:4-15.
7. Wong PK, Chan IN, Yan HM, Gao S, Wong CH, Yan T, Yao L, Hu Y, Wang ZR, Yu HH. Deep learning based radiomics for gastrointestinal cancer diagnosis and treatment: A minireview. World J Gastroenterol 2022;28:6363-79.
8. Cao L, Zhang Q, Fan C, Cao Y. Not Another Dual Attention UNet Transformer (NNDA-UNETR): a plug-and-play parallel dual attention block in U-Net with enhanced residual blocks for medical image segmentation. Quant Imaging Med Surg 2024;14:9169-92.
9. Takeuchi M, Seto T, Hashimoto M, Ichihara N, Morimoto Y, Kawakubo H, Suzuki T, Jinzaki M, Kitagawa Y, Miyata H, Sakakibara Y. Performance of a deep learning-based identification system for esophageal cancer from CT images. Esophagus 2021;18:612-20.
10. Sui H, Ma R, Liu L, Gao Y, Zhang W, Mo Z. Detection of Incidental Esophageal Cancers on Chest CT by Deep Learning. Front Oncol 2021;11:700210.
11. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203-11.
12. Ma J, Li F, Wang B. U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation. ArXiv 2024;abs/2401.04722.
13. Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, et al. The Medical Segmentation Decathlon. Nat Commun 2022;13:4128.
14. Park HJ, Shin K, You MW, Kyung SG, Kim SY, Park SH, Byun JH, Kim N, Kim HJ. Deep Learning-based Detection of Solid and Cystic Pancreatic Neoplasms at Contrast-enhanced CT. Radiology 2023;306:140-9.
15. Hu Y, Jiang S, Yu X, Huang S, Lan Z, Yu Y, Zhang X, Chen J, Zhang J. Automatic epicardial adipose tissue segmentation in pulmonary computed tomography venography using nnU-Net. Quant Imaging Med Surg 2023;13:6482-92.
16. Tsai TY, Lin L, Hu S, Chang MC, Zhu H, Wang X. UU-Mamba: Uncertainty-aware U-Mamba for Cardiac Image Segmentation. 2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 2024, pp. 267-73
17. Lin C, Guo Y, Huang X, Rao S, Zhou J. Esophageal cancer detection via non-contrast CT and deep learning. Front Med (Lausanne) 2024;11:1356752.
18. Amyar A, Modzelewski R, Vera P, Morard V, Ruan S. Multi-task multi-scale learning for outcome prediction in 3D PET images. Comput Biol Med 2022;151:106208.
19. Jayaprakasam V, Gibbs P, Gangai N, Bajwa R, Sosa R, Yeh R, Greally M, Ku G, Gollub M, Paroder V. Can F-FDG PET/CT Radiomics Features Predict Clinical Outcomes in Patients with Locally Advanced Esophageal Squamous Cell Carcinoma? Cancers 2022;14:3035.
20. Xie CY, Hu YH, Ho JW, Han LJ, Yang H, Wen J, Lam KO, Wong IY, Law SY, Chiu KW, Fu JH, Vardhanabhuti V. Using Genomics Feature Selection Method in Radiomics Pipeline Improves Prognostication Performance in Locally Advanced Esophageal Squamous Cell Carcinoma-A Pilot Study. Cancers (Basel) 2021;13:2145.
21. Wu YP, Wu L, Ou J, Cao JM, Fu MY, Chen TW, Ouchi E, Hu J. Preoperative CT radiomics of esophageal squamous cell carcinoma and lymph node to predict nodal disease with a high diagnostic capability. Eur J Radiol 2024;170:111197.
22. Huang YL, Yan C, Lin X, Chen ZP, Lin F, Feng ZP, Ke SK. The development of a nomogram model for predicting left recurrent laryngeal nerve lymph node metastasis in esophageal cancer based on radiomics and clinical factors. Ann Transl Med 2022;10:1282.
23. Li Y, Yu M, Wang G, Yang L, Ma C, Wang M, Yue M, Cong M, Ren J, Shi G. Contrast-Enhanced CT-Based Radiomics Analysis in Predicting Lymphovascular Invasion in Esophageal Squamous Cell Carcinoma. Front Oncol 2021;11:644165.
24. Li Y, Liu J, Li HX, Cai XW, Li ZG, Ye XD, Teng HH, Fu XL, Yu W. Radiomics Signature Facilitates Organ-Saving Strategy in Patients With Esophageal Squamous Cell Cancer Receiving Neoadjuvant Chemoradiotherapy. Front Oncol 2020;10:615167.

25. Hu P, Chen L, Zhong Y, Lin Y, Yu X, Hu X, Tao X, Lin S, Niu T, Chen R, Wu X, Sun J. Effects of slice thickness on CT radiomics features and models for staging liver fibrosis caused by chronic liver disease. Jpn J Radiol 2022;40:1061-8.
26. Barragán-Montero AM, Thomas M, Defraene G, Michiels S, Haustermans K, Lee JA, Sterpin E. Deep learning dose prediction for IMRT of esophageal cancer: The effect of data quality and quantity on model performance. Phys Med 2021;83:52-63.

Cite this article as: Hu Y, Zhang Y, Tang Z, Han X, Hong H, Kong L, Xu Z, Jiang S, Yu X, Zhang L. Comparative analysis of U-Mamba and no new U-Net for the detection and segmentation of esophageal cancer in contrast-enhanced computed tomography images. Quant Imaging Med Surg 2025;15(3):2119-2131. doi: 10.21037/qims-24-1116

