Radiation Oncology

RESEARCH

Open Access

Smart contours: deep learning-driven internal gross tumor volume delineation in non-small cell lung cancer using 4D CT maximum and average intensity projections

Yuling Huang[1,2,3], Mingming Luo[1,2,3], Zan Luo[1,2,3], Mingzhi Liu[1,2,3], Junyu Li[1,2,3], Junming Jian[1,2,3*] and Yun Zhang[1,2,3*]

Abstract

Background Delineating the internal gross tumor volume (IGTV) is crucial for the treatment of non-small cell lung cancer (NSCLC). Deep learning (DL) enables the automation of this process; however, current studies focus mainly on multiple phases of four-dimensional (4D) computed tomography (CT), which leads to indirect results. This study proposed a DL-based method for automatic IGTV delineation using maximum and average intensity projections (MIP and AIP, respectively) from 4D CT.

Methods We retrospectively enrolled 124 patients with NSCLC and divided them into training (70%, $n=87$) and validation (30%, $n=37$) cohorts. Four-dimensional CT images were acquired, and the corresponding MIP and AIP images were generated. The IGTVs were contoured on 4D CT and used as the ground truth (GT). The MIP or AIP images, along with the corresponding IGTVs ($IGTV_{MIP-manu}$ and $IGTV_{AIP-manu}$, respectively), were fed into the DL models for training and validation. We assessed the performance of three segmentation models—U-net, attention U-net, and V-net—using the Dice similarity coefficient (DSC) and the 95th percentile of the Hausdorff distance (HD95) as the primary metrics.

Results The attention U-net model trained on AIP images presented a mean DSC of 0.871 ± 0.048 and mean HD95 of 2.958 ± 2.266 mm, whereas the model trained on MIP images achieved a mean DSC of 0.852 ± 0.053 and mean HD95 of 3.209 ± 2.136 mm. Among the models, attention U-net and U-net achieved similar results, considerably surpassing V-net.

Conclusions DL models can automate IGTV delineation using MIP and AIP images, streamline contouring, and enhance the accuracy and consistency of lung cancer radiotherapy planning to improve patient outcomes.

Keywords Internal gross tumor volume, Automatic delineation, Deep learning, Non-small cell lung cancer, Radiotherapy, Four-dimensional computed tomography

*Correspondence:
Junming Jian
jianjunming94@163.com
Yun Zhang
zhangyun_1983@sohu.com

[1]Department of Radiation Oncology, Jiangxi Cancer Hospital & Institute (The Second Affiliated Hospital of Nanchang Medical College), Nanchang 330029, Jiangxi, PR China
[2]Jiangxi Key Laboratory of Oncology (2024SSY06041), Nanchang 330029, Jiangxi, PR China
[3]Jiangxi Clinical Research Centre for Cancer, Nanchang 330029, Jiangxi, PR China

Background

Lung cancer remains the leading cause of cancer-related mortality worldwide [1, 2], with non-small cell lung cancer (NSCLC) accounting for approximately 85% of all cancer cases [3]. Radiation therapy is a cornerstone in lung cancer management [4, 5] and involves delivering high-dose radiation to malignant cells while sparing the surrounding healthy tissue. Respiratory-induced intrafractional movement significantly affects the precise radiation dose delivered to a tumor, particularly in lung cancer cases. This variation can lead to insufficient tumor control or damage to organs at risk. To counter the effects of intrafractional motion, various adaptive techniques, such as respiratory gating, tumor tracking, breath holding, and motion-encompassing strategies, have been developed.

The internal target volume (ITV) concept is widely accepted and utilized. Derived from four-dimensional (4D) computed tomography (CT), the ITV method expands the target volume to include the full range of tumor motion throughout the respiratory cycle. Four-dimensional CT scans produce a series of three-dimensional (3D) images, with each image capturing a distinct phase of the normal breathing cycle. Moreover, the internal gross tumor volume (IGTV) concept has been adopted to define the ITV more efficiently [6] because the ITV is calculated as the IGTV plus a margin that is often a fixed value, thus encompassing microscopic diseases. Thus, precise delineation of the IGTV is crucial for effective radiation therapy planning. The most precise approach for delineating the IGTV on 4D CT datasets involves manually contouring the GTVs in all 10 respiratory phases and merging them to create a comprehensive IGTV. However, this dependency on manual delineation poses challenges in terms of efficiency, which highlights the critical need for innovative solutions in radiation oncology.

Maximum intensity projection (MIP) [7–9] and average intensity projection (AIP) [8, 9] have been extensively discussed as alternatives. MIP captures the highest density of each pixel during the respiratory cycle of 4D CT. In practice, IGTV delineation is typically simplified by direct definition on the MIP image and is followed by adjustment according to the motion information from the breathing phases [6, 10]. Several studies [11, 12] have validated its efficacy in target delineation; however, certain concerns [13–15] exist regarding its potential to underestimate the size of the ITV. AIP is another postprocessing method that creates a single 3D image by averaging the voxel intensities over all phases of the respiratory cycle, thus leveraging all the information from 4D CT. Similarly, concerns have been raised regarding the potential issue of underestimating tumor volumes. Borm et al. [15] demonstrated that the IGTV contoured in AIP can differ significantly from the ITV contoured in 10 phases of 4D CT, with average deviations reaching 18.7%. In clinical practice, the aforementioned methods, which use AIP or MIP for contouring, typically employ uniform expansion for adjustment rather than individualized expansion, thus failing to reflect the unique breathing patterns of each patient. This implies that they are not universally applicable to all types of respiratory motion [16], particularly in cases of irregular motion patterns.

AI offers promising solutions by automating and standardizing the tumor delineation process [17, 18], thereby improving the accuracy and consistency of radiation treatment planning. Given the potential of AI in medical imaging, several studies have explored its application for automatic lung cancer delineation. However, most previous studies [19–21] have focused on GTV segmentation, whereas relatively few studies have addressed the prediction of IGTV. Li et al. [16] proposed a novel deep machine-learning algorithm with a linear exhaustive optimal combination framework for IGTV delineation on full 4D CT in patients with lung cancer treated via stereotactic body radiation therapy. As an indirect solution, the proposed algorithm obtains the GTVs for all phases and then merges them to generate the IGTV, which is tedious and time-consuming. Furthermore, the errors generated from the automatic delineation of the GTV of each phase are not controllable, and they may gradually accumulate and result in significant errors in the merged IGTV. To address these issues, Ma et al. [6] designed deep-learning (DL) models for IGTV delineation directly on 4D CT. However, owing to the limitations of GPU memory, only three phases (100% In, 0% Ex, and an intermediate phase of 60% Ex) were utilized for model training. Although the pipeline is drastically simplified, the information in 4D CT images may not be fully exploited, leading to suboptimal results.

In this study, considering the significant potential of MIP and AIP in IGTV delineation and the powerful modeling capabilities of AI, we established a new paradigm for automatic IGTV delineation. Compared with 4D CT, AIP and MIP images enable more efficient IGTV delineation by reducing resource usage during model training and inference while preserving key information and improving delineation speed. Furthermore, AIP images are widely utilized for delineating normal tissues and organs at risk. Using AIP for IGTV delineation helps streamline the radiotherapy process, improving both workflow efficiency and accuracy. In this study, IGTVs were first manually contoured on MIP or AIP images. Subsequently, the contoured IGTVs and the corresponding images were sent to the DL network to generate the final IGTV. Three segmentation models, U-net [22], attention U-net [23], and V-net [24], were assessed for their performance on this task. To the best of our

knowledge, this study is the first to generate an IGTV directly from AIP or MIP images.

Methods

Materials

This study was approved by the Institutional Review Board (IRB) of Jiangxi Cancer Hospital, and the requirement for informed consent was waived owing to the retrospective nature of the study. The inclusion criteria for this study were as follows: (1) a histologically confirmed diagnosis of NSCLC, (2) indication for radiotherapy and successful completion of the prescribed treatment, and (3) availability of high-quality 4D CT scans. The exclusion criteria included (1) incomplete imaging data or poor image quality, (2) the requirement for radiotherapy at additional anatomical sites simultaneously, and (3) the presence of contraindications to radiotherapy. Figure 1 shows the pipeline of data collection, processing, and model development. One hundred and twenty-four patients with lung cancer treated at Jiangxi Cancer Hospital were enrolled in this study. None of the patients underwent post-induction chemotherapy. The 4D CT images were acquired before treatment as part of routine clinical practice using a Siemens SOMATOM Definition AS CT scanner (Siemens, Erlangen, Germany). The scanner operated with a voxel spacing of either $0.976 \times 0.976 \times 2$ or $1.269 \times 1.269 \times 2$ mm, ensuring high-resolution imaging. The imaging parameters were set to 80 mA and 120 kV. The 4D CT data were sorted into 10 respiratory bins, covering 0–90% of the respiratory cycle. In this scheme, the 0% phase represents peak inhalation, whereas the other bins correspond to varying phases of the respiratory cycle, capturing tumor motion across different breathing states. Among these, 70% ($n = 87$) were randomly assigned to the training cohort for model training, 10% ($n = 12$) to the validation cohort for hyperparameter tuning, and 20% ($n = 25$) to the testing cohort for model evaluation.

AIP and MIP images were generated from 4D CT, and the corresponding IGTVs were delineated by a thoracic radiation oncologist with 10 years of experience using the Eclipse software (version 13.5). The GTV included only the primary tumor. A preset lung window setting (W = 1600 and L = 600) was used to delineate tumors surrounded by lung tissue, whereas a mediastinum preset window setting (W = 400 and L = 20) was used to delineate primary tumors invading the mediastinum or chest wall [25]. Similarly, the GTV was contoured phase by phase on 4D CT by the same oncologist and then fused to generate the ground truth IGTV ($IGTV_{gt}$). A senior thoracic radiation oncologist reviewed all the target volumes.

Data preprocessing

The AIP (or MIP) images, along with the corresponding $IGTV_{AIP-manu}$ (or $IGTV_{MIP-manu}$) and $IGTV_{gt}$, were resampled to a resolution of $1 \times 1 \times 2$ mm. To reduce both training and inference times while simplifying model coverage, a $112 \times 112 \times 80$-pixel region centered around the $IGTV_{AIP-manu}$ (or $IGTV_{MIP-manu}$) was automatically cropped. The intensities of the images were normalized to a range of 0–1. During the model-training stage, data augmentation techniques were applied, including random flips across the three axes ($p = 0.1$), random 90-degree rotations ($p = 0.1$), and random intensity shifts with an offset of 0.1 ($p = 0.5$).

Networks

Convolutional neural networks have been extensively used for image segmentation tasks in medical image analysis. U-net [22], attention U-net [23], and V-net [24] are three popular architectures, each with unique strengths. In this study, these models were adopted for IGTV delineation.

A two-channel structure was used for the input to the models: one channel comprised the AIP or MIP image, and the other channel contained the corresponding manually contoured IGTV (either the $IGTV_{AIP-manu}$ or $IGTV_{MIP-manu}$). The number of channels in U-net and attention U-net were set to 32, 64, 128, and 256, whereas in V-net, the channels were configured to 16, 32, 64, 128, and 256.

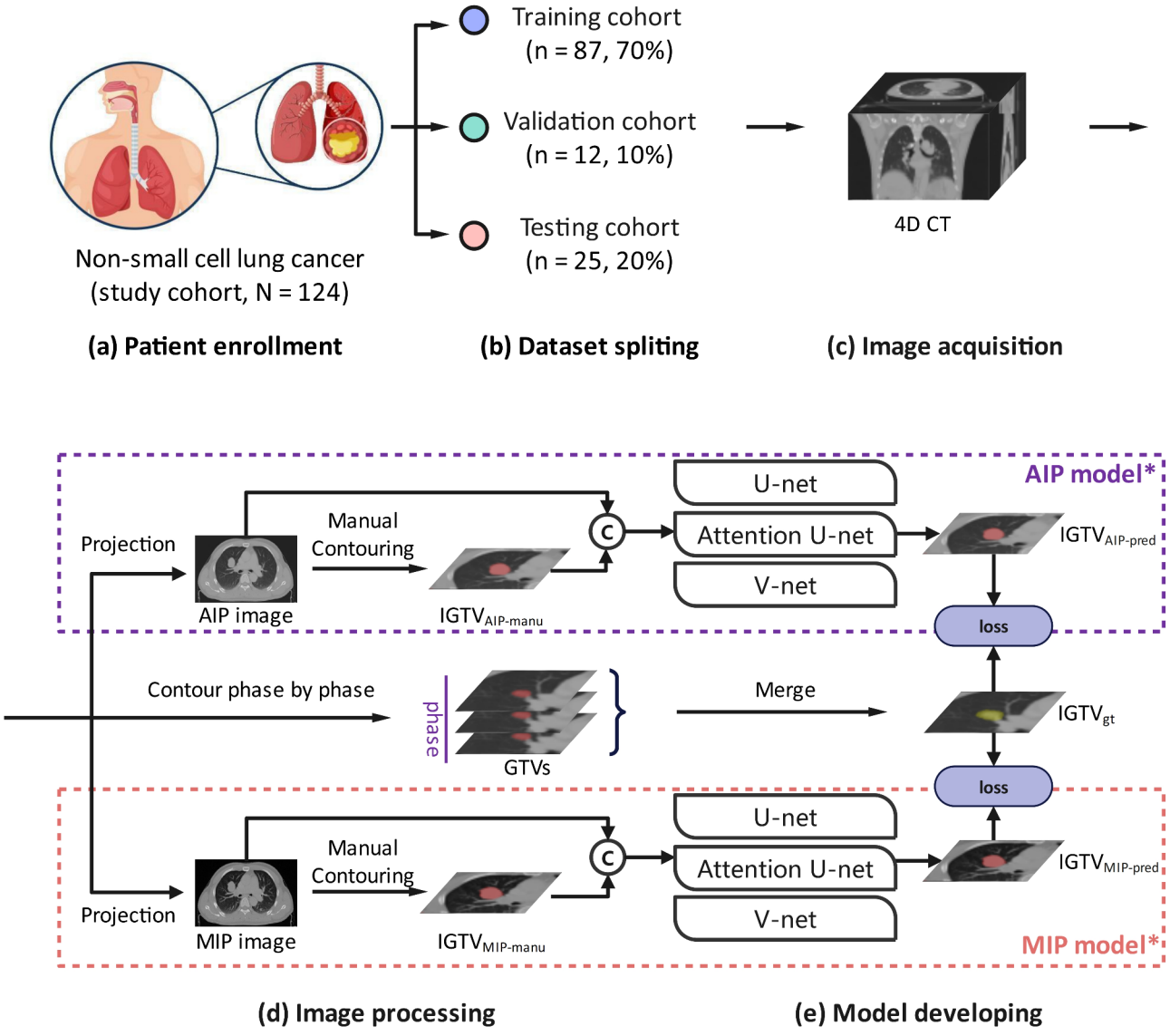
Implementation details

The models were implemented in Python (version 3.10.9) using PyTorch [26] (version 2.0.1; https://github.com/pytorch/pytorch) on a server equipped with eight NVIDIA GeForce RTX 4090 GPUs. The adaptive gradient algorithm (AdamW) [27] with a weight decay of 1e – 5 served as the optimizer. All the networks were trained using the optimal learning rate with a maximum of 100 iterations. The batch size was set to either 32 or 64. The final hyperparameters were chosen by evaluating the model performance on the validation set.

Evaluation metrics

The Dice similarity coefficient (DSC), surface DSC with a 1 mm tolerance value, 95th percentile of the Hausdorff distance (HD95), sensitivity, and specificity were used to evaluate the performance of IGTV delineation. Differences between the models were assessed using the Mann–Whitney U test. The statistical significance was set at a p-value < 0.05.

The performance of the proposed model was assessed by comparing the predicted IGTV with the $IGTV_{gt}$. Additionally, we compared the manually contoured IGTV ($IGTV_{AIP-manu}$ or $IGTV_{MIP-manu}$) to the $IGTV_{gt}$, which



Ⓒ AIP or MIP images, along with their corresponding manually contoured IGTV, are combined into a two-channel input and fed into the network

* The AIP and MIP models are trained using AIP and MIP images, respectively, along with the corresponding manually contoured IGTV.

Fig. 1 Pipeline of data collection, preprocessing, and model training. MIP and AIP: maximum and average intensity projections, respectively; IGTV: internal gross tumor volume. IGTV$_{AIP-manu}$ and IGTV$_{MIP-manu}$: manually contoured tumor volumes on AIP and MIP images, respectively, used as inputs for the model. IGTV$_{gt}$: ground truth IGTV, which is derived from the fusion of GTVs contoured across different breathing phases

enabled us to evaluate the error introduced by manual delineation directly on AIP or MIP images and assess the additional value of our model in improving IGTV delineation accuracy.

Results

The characteristics of the enrolled patients are presented in Table 1. No significant differences were observed between the training cohort ($N = 87$) and test cohort ($N = 25$) in terms of age (p-value = 0.804), gender (p-value = 0.127), or overall cancer stage (p-value = 0.721).

Table 1 Characteristics of patients in the training and test cohorts

Characteristics	Training cohort	Test cohort	*P*-value
Number of patients	***N*** **= 87**	***N*** **= 25**	
Age (y), median [Q1, Q3]	66.0 [59.0,74.0]	66.0 [60.0,75.0]	0.804
Gender, n (%)			0.127
Male	57 (65.5)	21 (84.0)	
Female	30 (34.5)	4 (16.0)	
Overall stage, n (%)			0.721
I	15 (17.2)	4 (16.0)	
II	4 (4.6)	0 (0)	
III	8 (9.2)	2 (8.0)	
IV	60 (69.0)	19 (76.0)	

Note: P-values are provided to assess the statistical significance between the training and test cohorts for each characteristic. A p-value < 0.05 indicates a statistically significant difference between the groups

The performance of the models for MIP-based IGTV delineation on the test cohort is summarized in Table 2. Attention U-net showed the highest DSC (0.852 ± 0.053), followed by U-net (0.829 ± 0.055) and V-net (0.819 ± 0.073). The surface DSC showed similar trends, with attention U-net achieving the highest score (0.760 ± 0.130) compared with U-net (0.688 ± 0.150) and V-net (0.673 ± 0.123). For HD95, the attention U-net performed best (3.209 ± 2.136 mm), whereas U-net and V-net achieved higher HD95 values (3.264 ± 2.157 and 3.903 ± 2.157 mm, respectively). In terms of sensitivity, U-net (0.849 ± 0.095) outperformed the other models, followed by attention U-net (0.818 ± 0.108) and V-net (0.759 ± 0.142). Finally, all models demonstrated excellent specificity (0.999 ± 0.001). All models showed improved delineated IGTVs compared with the manually contoured IGTVs on MIP images ($IGTV_{MIP-manu}$).

Table 3 summarizes the performance of the models for AIP-based IGTV delineation for the test cohort. The attention U-net achieved the highest DSC (0.871 ± 0.048) and surface DSC (0.801 ± 0.133), followed closely by U-net (0.870 ± 0.048 for the DSC and 0.797 ± 0.139 for the surface DSC). V-net achieved a slightly lower performance, with a DSC of 0.853 ± 0.051 and surface DSC of 0.742 ± 0.129. For HD95, the U-net model performed best (2.932 ± 2.262 mm), whereas attention U-net and V-net yielded slightly higher values (2.958 ± 2.266 and 3.263 ± 2.319 mm, respectively). In terms of sensitivity, V-net was the best (0.859 ± 0.091), closely followed by attention U-net (0.855 ± 0.093) and U-net (0.852 ± 0.090). All the models showed excellent specificity, achieving perfect scores (0.999 ± 0.001). All models demonstrated improvements in IGTV delineation accuracy compared with that of AIP-based manual contouring of IGTVs ($IGTV_{AIP-manu}$).

Compared with MIP, AIP generally showed slightly better performance across nearly all the metrics, with a DSC of 0.871 vs. 0.852, surface DSC of 0.801 vs. 0.760, HD95 of 2.932 mm vs. 3.209 mm, and sensitivity of 0.859 vs. 0.849.

The visualization results for a typical case from the test cohort, along with the AIP images, are shown in Fig. 2. The DSC and HD95 values for this case were 0.929 and 1.414 mm, respectively. The red contours represent the $IGTV_{gt}$, whereas the green and orange contours represent the $IGTV_{AIP-pred}$ (predictions from the attention U-net) and the manually contoured IGTV ($IGTV_{AIP-manu}$), respectively. Twenty-four slices are presented, each labeled with sensitivity and specificity, which were calculated by comparing the $IGTV_{AIP-pred}$ with the $IGTV_{gt}$. In most slices, our model correctly predicted the tumor, showing a significant improvement over the manually

Table 2 Performance of deep-learning models for IGTV delineation using MIP images for the test cohort

	$IGTV_{MIP-manu}$	Attention U-net	U-net	V-net
DSC	0.713 ± 0.117	**0.852 ± 0.053**	0.829 ± 0.055	0.819 ± 0.073
Surface DSC	0.466 ± 0.231	**0.760 ± 0.130**	0.688 ± 0.150	0.673 ± 0.123
HD95 (mm)	4.146 ± 2.229	**3.209 ± 2.136**	3.264 ± 2.015	3.903 ± 2.157
Sensitivity	0.571 ± 0.151	0.818 ± 0.108	**0.849 ± 0.095**	0.759 ± 0.142
Specificity	1.000 ± 0.000	**0.999 ± 0.001**	**0.999 ± 0.001**	**0.999 ± 0.001**

Note: The $IGTV_{MIP-manu}$ column presents the comparison between the manually contoured IGTV on MIP images and the gold standard IGTV ($IGTV_{gt}$), whereas the model columns present the comparison between the model predictions and the $IGTV_{gt}$. The model with the best performance in each metric is highlighted in bold

Table 3 Performance of deep-learning models for IGTV delineation using AIP images for the test cohort

	$IGTV_{AIP-manu}$	Attention U-net	U-net	V-net
DSC	0.689 ± 0.088	**0.871 ± 0.048**	0.870 ± 0.048	0.853 ± 0.051
Surface DSC	0.390 ± 0.183	**0.801 ± 0.133**	0.797 ± 0.139	0.742 ± 0.129
HD95 (mm)	4.222 ± 2.394	2.958 ± 2.266	**2.932 ± 2.262**	3.263 ± 2.319
Sensitivity	0.533 ± 0.102	0.855 ± 0.093	0.852 ± 0.090	**0.859 ± 0.091**
Specificity	1.000 ± 0.000	**0.999 ± 0.001**	**0.999 ± 0.001**	**0.999 ± 0.001**

Note: The $IGTV_{AIP-manu}$ column presents the comparison between the manually contoured IGTV on AIP images and the gold standard IGTV ($IGTV_{gt}$), whereas the model columns present the comparison between the model predictions and the $IGTV_{gt}$. The model with the best performance in each metric is highlighted in bold

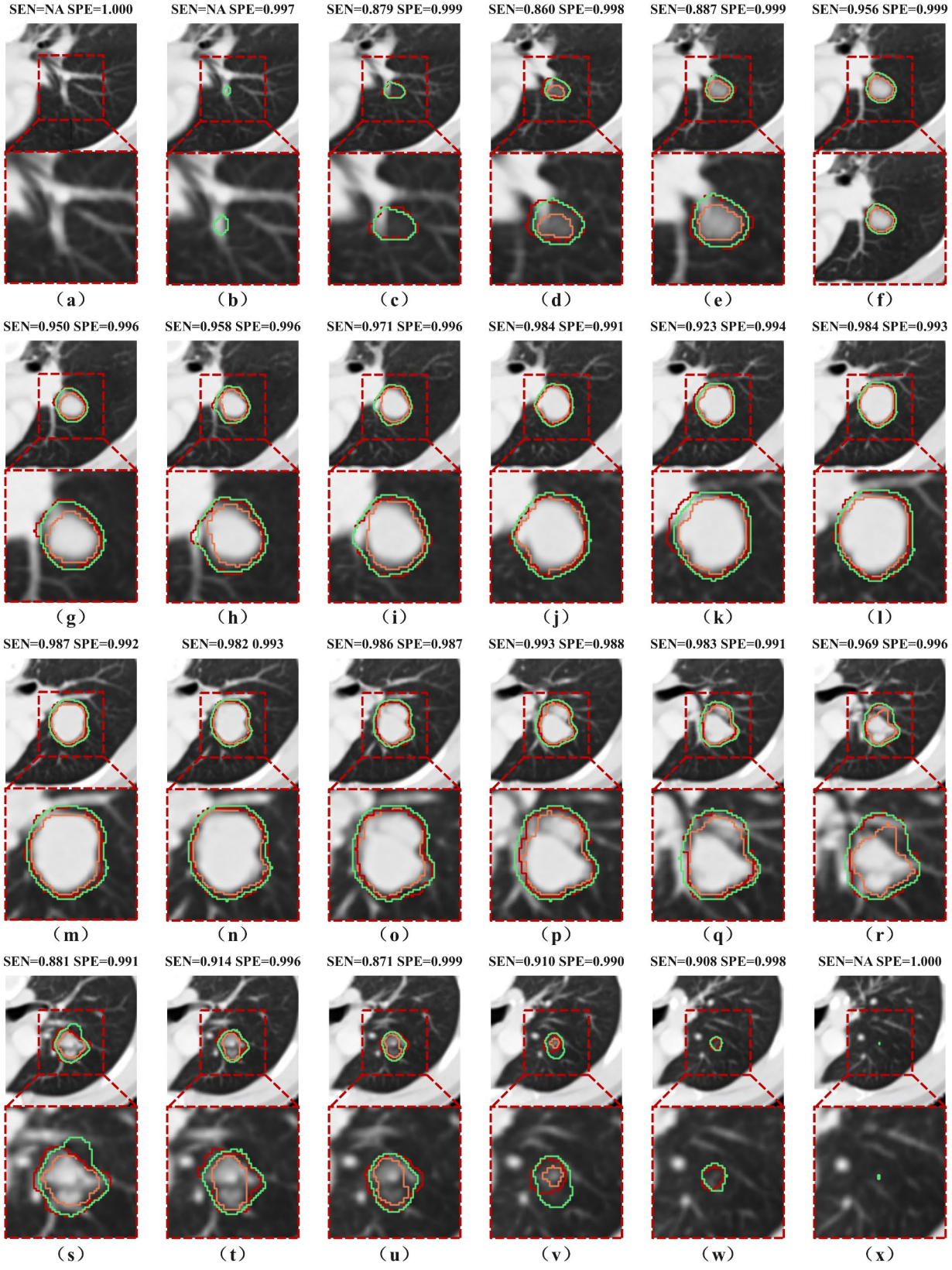


Fig. 2 Visualization results for a case with average intensity projection images. The red contours represent the ground truth internal gross tumor volume (IGTV), the green contours represent the IGTV predicted by the attention U-net model, and the orange contours represent the manually contoured IGTV

contoured IGTV. Note that in Figs. 2(c) and 2(w), the tumor target was almost invisible, making manual contouring of the IGTV on the AIP images impractical; however, our model performed well and successfully identified the tumor targets. Moreover, in the top and bottom slices, as shown in Figs. 2(b) and 2(x), our model failed and tended to make false positive predictions, mistakenly identifying the presence of tumors where none existed.

Discussion

In this study, we introduced a novel automated IGTV contouring method specifically designed for lung cancer radiotherapy using DL models. Contouring the IGTV is a crucial aspect of radiotherapy treatment planning for mitigating the impact of intra-fractional target motion. This process is often time consuming and subject to variability among observers. Therefore, DL-assisted automatic segmentation is gaining popularity.

Regarding the automated delineation of the IGTV in patients with lung cancer, Li et al. [16] proposed a DL framework for automated IGTV delineation using 4D CT, where GTVs from each respiratory phase are generated and combined to form the final IGTV. Although innovative, this method is labor intensive and prone to errors because inaccuracies in phase-specific GTVs accumulate when the GTVs are merged into the final IGTV. Ma et al. [6] proposed an improved method using 4D CT; however, owing to GPU memory limitations, only three specific phases (0% Ex, 60% Ex, and 100% In) were used, yielding a DSC of 0.7405 and HD95 of 15.61 mm, which are lower than those obtained by our model. We believe that using only these phases leads to inaccuracies in IGTV delineation because they do not capture full tumor motion. To overcome this, our study utilized AIP and MIP images, maximizing the available information in 4D CT scans and providing a more comprehensive and accurate delineation of the IGTV by capturing a broader range of tumor motion and density variations [28]. However, the methodological differences between our study and previous works may influence segmentation performance. Factors such as variations in patient demographics, CT acquisition settings, breathing motion management strategies, and tumor localization criteria could contribute to differences in reported metrics. Therefore, future studies would benefit from more rigorous comparisons that account for these variables to ensure a fair and reliable evaluation. Additionally, in current clinical practice, AIP is used to delineate normal tissues and organs at risk, whereas GTVs are contoured across multiple breathing phases and fused to create the final IGTV. Our method improves upon this by directly using AIP for IGTV delineation, thus eliminating the need for multiple phase delineation and fusion. We streamlined the process by relying solely on AIP, improving efficiency without compromising accuracy, and reducing workflow complexity.

MIP provides a visualization method in which the highest intensity voxel is projected along the viewing direction, helping identify the full extent of tumors that exhibit higher radiodensity than the surrounding tissues [13]. This is particularly useful in thoracic imaging, where tumors must be distinguished from the surrounding normal lung tissue, which can vary significantly in density owing to respiratory motion. In contrast, AIP averages the intensity of all voxels along each projection path, offering a more holistic view of tumor density variations. While AIP mitigates the potential overemphasis on high-density artifacts commonly observed in MIP, it introduces the risk of underestimating tumor volumes. Specifically, due to averaging intensities across all respiratory phases, AIP may obscure the extremes of tumor motion, potentially leading to an incomplete delineation of the IGTV. Previous studies have reported underestimation risks associated with AIP-based delineation methods [14]. For example, Borm et al. [15] indicated that the ITV contoured in MIP and AIP differs significantly from the ITV contoured in 10 phases of a 4D CT scan, with average deviations potentially as large as 25%. Furthermore, Tibdewal et al. [7] concluded that compared with ITV delineation in 10 phases of 4D CT, contouring the ITV using MIP is significantly smaller and may miss tumors. Given that IGTV incorporates tumor motion similarly to ITV, these findings suggest a likely risk of underestimation when using AIP images for IGTV delineation. Our findings are consistent with these observations because the consistency between $IGTV_{MIP-manu}$ and $IGTV_{gt}$ was low, as shown in Tables 2 and 3. When DL models were introduced, the generated IGTV showed a significant improvement. Specifically, the IGTV generated using AIP images ($IGTV_{AIP-pred}$) demonstrated superior performance compared with those generated using MIP images ($IGTV_{MIP-pred}$). This is because AIP images contain intensity information from all phases, whereas MIP images reflect only the phase with the highest intensity.

Three models—U-net, attention U-net, and V-net—were compared in terms of their performance in generating IGTVs. The results indicate that both U-net and attention U-net outperformed V-Net, although the statistical analysis revealed no significant differences between the models in either the DSC or HD95. The performance of attention U-net can be attributed to its enhanced focus mechanism, which aids in capturing fine details and variations in the tumor region. However, the lack of a significant difference suggests that the basic U-net architecture is sufficiently robust for IGTV segmentation in this dataset. The slightly lower performance of V-net could be due to its complexity and the difficulty of model convergence

with the AIP and MIP images. This might be attributed to the design of V-net, which is optimized for more heterogeneous structures, rendering it less suited to the simpler intensity variations in these images.

While numerical metrics such as the DSC, Surface DSC, HD95, sensitivity, and specificity provide quantitative insights into segmentation performance, their clinical implications deserve explicit interpretation. In clinical radiation therapy planning, higher DSC and lower HD95 values translate directly into more precise delineation of tumor boundaries, potentially enabling clinicians to reduce PTV margins without compromising treatment effectiveness [29, 30]. Smaller margins can subsequently reduce the radiation dose to adjacent healthy tissues, thus decreasing treatment-related toxicity and improving patients' quality of life. Furthermore, higher sensitivity reduces the risk of missing tumor regions, whereas higher specificity decreases false positives, collectively contributing to more precise and reliable treatment planning [31, 32]. Therefore, the observed improvements in these metrics using our proposed approach could significantly impact clinical practice by improving targeting accuracy, reducing inter-observer variability, streamlining workflows, and ultimately enhancing patient outcomes [33].

This study has several limitations. First, this was a preliminary feasibility study conducted at a single center with a small number of patients to explore IGTV delineation using AIP or MIP images. The findings have not been validated via bootstrapping, via cross-validation, or across multiple centers, which may affect the generalizability of the results. To address this limitation, future studies will focus on multi-institutional validation, which will involve a more diverse patient population, variations in imaging protocols, and different clinical workflows. This will help assess model robustness across heterogeneous datasets and improve its applicability in real-world clinical settings. Second, the models used—U-net, attention U-net, and V-net—are standard models and are not specifically tailored for our dataset or task. Future studies should focus on developing customized models for better performance. Third, the study relied on manual IGTV contouring, which is time-consuming and may be subject to interobserver variability. Although we tested multiple models, including U-net, attention U-net, and V-net, fully automatic contouring of AIP or MIP images without any manual delineation was unsuccessful owing to model convergence issues. Consequently, we adopted a clinical practice approach by manually contouring the IGTV on AIP or MIP images and using these contours to train the model. To mitigate the variability introduced by manual contouring, future studies should explore standardized contouring protocols, structured observer training programs, and inter-observer agreement analyses to improve consistency. Additionally, incorporating consensus-based annotations and leveraging AI-assisted refinement techniques could enhance model robustness and generalizability. Fourth, the tumor motion amplitude and tumor location (e.g., upper-lobe tumors) may impact segmentation performance. The automated delineation for cases with larger tumor motion is more challenging, potentially reducing segmentation accuracy. Future studies should investigate the influence of these factors and explore motion-aware network architectures or adaptive training strategies to improve performance in such scenarios. Finally, we did not explore the potential benefits of fusing MIP and AIP images, which could enhance IGTV segmentation by leveraging the strengths of both methods. Future research should investigate this integration to determine whether it can provide better results.

Conclusions

This study aimed to improve the accuracy of IGTV segmentation using DL models. The results indicate that these models have the potential for automatic IGTV delineation using both MIP and AIP images, along with their corresponding manually contoured IGTVs. Among these, AIP images demonstrated superior performance across multiple quantitative metrics. The proposed method streamlines the contouring process, leading to more accurate and consistent radiation therapy planning for lung cancer, thereby improving patient outcomes. In future, validating these findings in multi-institutional settings will be crucial to assess their generalizability and robustness across diverse patient populations and clinical practices. Further research should focus on refining the model's performance, optimizing the workflow for clinical implementation, and exploring its impact on treatment efficiency and patient outcomes in broader clinical contexts.

Abbreviations

3D	Three-dimensional
4D	Four-dimensional
AdamW	Adaptive gradient algorithm
AIP	Average intensity projection
CT	Computed tomography
DL	Deep learning
DSC	Dice similarity coefficient
GT	Ground truth
HD95	The 95th percentile of the Hausdorff distance
IGTV	Internal gross tumor volume
$IGTV_{AIP-manu}$	Manually contoured IGTV on AIP images
$IGTV_{gt}$	Ground truth IGTV
$IGTV_{MIP-manu}$	Manually contoured IGTV on MIP images
$IGTV_{pred}$	Predicted IGTV
ITV	Internal target volume
MIP	Maximum intensity projection
NSCLC	Non-small cell lung cancer
SD	Standard deviation

Author contributions

Y. H.: Conceptualization, Data curation, Funding acquisition, Resources, Validation, and Writing– original draft.M. L.: Formal Analysis, Software,

Visualization, and Writing– review & editing.Z. L.: Data curation, Investigation, and Writing– review & editing.M. L.: Data curation, Resources, and Writing– review & editing.J. L.: Data curation, Resources, Supervision, and Writing– review & editing.J. J.: Conceptualization, Formal Analysis, Funding acquisition, Methodology, Supervision, and Writing– review & editing.Y. Z.: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, and Writing– review & editing.All authors read and approved the final manuscript.

Funding
This work was funded by the "Five-level Progressive" talent cultivation project of Jiangxi Cancer Hospital (WCDJ2024JQ01), Jiangxi Cancer Hospital scientific research open fund project (KFJJ2023YB19), Jiangxi Provincial Natural Science Foundation (20242BAB21040), and Doctoral Startup Fund of Jiangxi Cancer Hospital (BSQDJ202308).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
The studies involving human participants were reviewed and approved by Ethics Committee of Jiangxi Cancer Hospital (Approval ID:2023ky192). Owing to the retrospective nature of the study, the requirement for informed consent was waived by the ethics committee. The study was conducted in accordance with the principles of the Declaration of Helsinki and in accordance with local statutory requirements.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Declaration of Generative AI and AI-assisted technologies in the writing process
We declare that the assistance of generative AI, specifically ChatGPT by OpenAI, was employed in the preparation of this manuscript. The AI was utilized to aid in the drafting of text, formulation of ideas, and refinement of language. All scientific content, including data analysis, interpretation of results, and conclusions, were conducted and reviewed by the authors. The authors take full responsibility for the integrity and accuracy of the content presented in this paper.

Received: 8 September 2024 / Accepted: 11 April 2025
Published online: 18 April 2025

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023;73:17–48.
2. Lundberg FE, Ekman S, Johansson ALV, Engholm G, Birgisson H, Ólafsdóttir EJ, et al. Trends in lung cancer survival in the nordic countries 1990–2016: the NORDCAN survival studies. Lung Cancer. 2024;192:107826.
3. Zhang S-L, Tian Y, Yu J, Zhang J-H, Sun L, Huang L-T, et al. Is neoadjuvant immunotherapy necessary in patients with programmed death ligand 1 expression-negative resectable non-small cell lung cancer? A systematic review and meta-analysis. Lung Cancer. 2024;191:107799.
4. Wang T-W, Hong J-S, Huang J-W, Liao C-Y, Lu C-F, Wu Y-T. Systematic review and meta-analysis of deep learning applications in computed tomography lung cancer segmentation. Radiother Oncol. 2024;197:110344.
5. Haasbeek CJA, Slotman BJ, Senan S. Radiotherapy for lung cancer: clinical impact of recent technical advances. Lung Cancer. 2009;64:1–8.
6. Ma Y, Mao J, Liu X, Dai Z, Zhang H, Zhang X, et al. Deep learning-based internal gross target volume definition in 4D CT images of lung cancer patients. Med Phys. 2023;50:2303–16.
7. Tibdewal A, Bushra S, Mummudi N, Kinhikar R, Ghadi Y, Agrawal JP. Is maximum intensity projection an optimal approach for internal target volume delineation in lung cancer?? J Med Phys. 2021;46:59–65.
8. Qi H, Li JB, Zhang YJ. Comparison of internal gross tumor volume for peripheral lung Cancer based on 4-dimensional CT and 3-dimensional CT assisted with active breathing control. Int J Radiat Oncol Biol Phys. 2012;84:S606–7.
9. Zhang S, Lin S, Yu H, Zhang H, Zhang G, Han P, Application of AIP and MIP CT on Individual GTV Delineation for Tumor Moving with Respiration. 2012 International Conference on Biomedical Engineering and Biotechnology [Internet]. 2012 [cited 2024 May 9]. pp. 736–9. Available from: https://ieeexplore.ieee.org/abstract/document/6245224/authors#authors
10. Beasley M, Brown S, McNair H, Faivre-Finn C, Franks K, Murray L et al. The advanced radiotherapy network (ART-NET) UK lung stereotactic ablative radiotherapy survey: National provision and a focus on image guidance. Br J Radiol. 2019;20180988.
11. Wang L, Chen X, Lin M, Xue J, Lin T, Fan J, et al. Evaluation of the cone beam CT for internal target volume localization in lung stereotactic radiotherapy in comparison with 4D MIP images. Med Phys. 2013;40:111709.
12. Underberg RWM, Lagerwaard FJ, Slotman BJ, Cuijpers JP, Senan S. Use of maximum intensity projections (MIP) for target volume generation in 4DCT scans for lung cancer. Int J Radiat Oncol*Biol*Phys. 2005;63:253–60.
13. Muirhead R, McNee SG, Featherstone C, Moore K, Muscat S. Use of maximum intensity projections (MIPs) for target outlining in 4DCT radiotherapy planning. J Thorac Oncol. 2008;3:1433–8.
14. Ezhil M, Vedam S, Balter P, Choi B, Mirkovic D, Starkschall G, et al. Determination of patient-specific internal gross tumor volumes for lung cancer using four-dimensional computed tomography. Radiat Oncol. 2009;4:4.
15. Borm KJ, Oechsner M, Wiegandt M, Hofmeister A, Combs SE, Duma MN. Moving targets in 4D-CTs versus MIP and AIP: comparison of patients data to Phantom data. BMC Cancer. 2018;18:760.
16. Li X, Deng Z, Deng Q, Zhang L, Niu T, Kuang Y. A novel deep learning framework for internal gross target volume definition from 4D computed tomography of lung Cancer patients. IEEE Access. 2018;6:37775–83.
17. Bollen H, Willems S, Wegge M, Maes F, Nuyts S. Benefits of automated gross tumor volume segmentation in head and neck cancer using multi-modality information. Radiother Oncol. 2023;182:109574.
18. Li L, Xu B, Zhuang Z, Li J, Hu Y, Yang H, et al. Accurate tumor segmentation and treatment outcome prediction with deeptop. Radiother Oncol. 2023;183:109550.
19. Cui Y, Arimura H, Nakano R, Yoshitake T, Shioyama Y, Yabuuchi H. Automated approach for segmenting gross tumor volumes for lung cancer stereotactic body radiation therapy using CT-based dense V-networks. J Radiat Res. 2021;62:346–55.
20. Momin S, Lei Y, Tian Z, Wang T, Roper J, Kesarwala AH, et al. Lung tumor segmentation in 4D CT images using motion convolutional neural networks. Med Phys. 2021;48:7141–53.
21. Yu X, Jin F, Luo H, Lei Q, Wu Y. Gross tumor volume segmentation for stage III NSCLC radiotherapy using 3D ResSE-Unet. Technol Cancer Res Treat. 2022;21:15330338221090847.
22. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015 [Internet]. Springer, Cham; 2015. pp. 234–41. Available from: https://link.springer.com/chapter/https://doi.org/10.1007/978-3-319-24574-4_28
23. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K et al. Attention U-Net: Learning Where to Look for the Pancreas [Internet]. arXiv; 2018 [cited 2024 May 21]. Available from: http://arxiv.org/abs/1804.03999
24. Milletari F, Navab N, Ahmadi S-AV. -Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. IEEE; 2016 [cited 2017 Oct 18]. pp. 565–71. Available from: http://ieeexplore.ieee.org/document/7785132/
25. De Ruysscher D, Faivre-Finn C, Moeller D, Nestle U, Hurkmans CW, Le Péchoux C, et al. European organization for research and treatment of Cancer (EORTC) recommendations for planning and delivery of high-dose, high precision radiotherapy for lung cancer. Radiother Oncol. 2017;124:1–10.
26. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc F dextquotesingle, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32 [Internet]. Curran Associates, Inc.; 2019. pp. 8024–35. Available from: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
27. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization [Internet]. arXiv.org. 2017 [cited 2024 May 28]. Available from: https://arxiv.org/abs/1711.05101v3

28. Rietzel E, Liu AK, Chen GTY, Choi NC. Maximum-Intensity volumes for fast contouring of lung tumors including respiratory motion in 4DCT planning. Int J Radiat Oncol*Biol*Phys. 2008;71:1245–52.
29. Savjani RR, Lauria M, Bose S, Deng J, Yuan Y, Andrearczyk V. Automated tumor segmentation in radiotherapy. Semin Radiat Oncol. 2022;32:319–29.
30. Ferreira Silvério N, van den Wollenberg W, Betgen A, Wiersema L, Marijnen C, Peters F, et al. Evaluation of deep learning clinical target volumes Auto-Contouring for magnetic resonance Imaging-Guided online adaptive treatment of rectal Cancer. Adv Radiation Oncol. 2024;9:101483.
31. Cho W, Yoo GS, Kim WD, Kim Y, Kim JS, Min BJ. Institution-Specific autosegmentation for personalized radiotherapy protocols. Progress Med Phys. 2024;35:205–13.
32. Johnston N, De Rycke J, Lievens Y, van Eijkeren M, Aelterman J, Vandersmissen E, et al. Dose-volume-based evaluation of convolutional neural network-based auto-segmentation of thoracic organs at risk. Phys Imaging Radiation Oncol. 2022;23:109–17.
33. Mikalsen SG, Skjøtskift T, Flote VG, Hämäläinen NP, Heydari M, Rydén-Eilertsen K. Extensive clinical testing of deep learning segmentation models for thorax and breast cancer radiotherapy planning. Acta Oncol. 2023;62:1184–93.

Publisher's note