Received: 14 May 2024 | Revised: 16 October 2024 | Accepted: 17 October 2024

DOI: 10.1002/mp.17493

RESEARCH ARTICLE

MEDICAL PHYSICS

Radiomics-guided generative adversarial network for automatic primary target volume segmentation for nasopharyngeal carcinoma using computed tomography images

Juebin Jin[1] | **Jicheng Zhang**[1] | **Xianwen Yu**[1] | **Ziqing Xiang**[2] | **Xuanxuan Zhu**[1] | **Mingrou Guo**[1] | **Zeshuo Zhao**[1] | **WenLong Li**[1] | **Heng Li**[1] | **Jiayi Xu**[1] | **Xiance Jin**[1,3]

[1]Department of Radiotherapy Center, 1st Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[2]Department of Medical Engineering, 2nd Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[3]School of Basic Medical Science, Wenzhou Medical University, Wenzhou, China

Correspondence
Xiance Jin, Department of Radiotherapy Center, 1st Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325000, China.
Email: jinxc1979@hotmail.com

Funding information
National Natural Science Foundation, Grant/Award Number: 12475352; Key project of Zhejiang Natural Science Foundation, Grant/Award Number: Z24A050009; Key project of Zhejiang Provincial Health Science and Technology Program, Grant/Award Number: WKJ-ZJ-2437; Major project of Wenzhou Science and Technology Bureau, Grant/Award Numbers: ZY2022016, ZY2020011; Wenzhou Science and Technology Bureau, Grant/Award Number: Y2023798

Abstract

Background: Automatic primary gross tumor volume (GTVp) segmentation for nasopharyngeal carcinoma (NPC) is a quite challenging task because of the existence of similar visual characteristics between tumors and their surroundings, especially on computed tomography (CT) images with severe low contrast resolution. Therefore, most recently proposed methods based on radiomics or deep learning (DL) is difficult to achieve good results on CT datasets.

Purpose: A peritumoral radiomics-guided generative adversarial network (PRG-GAN) was proposed to address this challenge.

Methods: A total of 157 NPC patients with CT images was collected and divided into training, validation, and testing cohorts of 108, 9, and 30 patients, respectively. The proposed model was based on a standard GAN consisting of a generator network and a discriminator network. Morphological dilation on the initial segmentation results from GAN was first conducted to delineate annular peritumoral region, in which radiomics features were extracted as priori guide knowledge. Then, radiomics features were fused with semantic features by the discriminator's fully connected layer to achieve the voxel-level classification and segmentation. The dice similarity coefficient (DSC), 95% Hausdorff distance (HD95), and average symmetric surface distance (ASSD) were used to evaluate the segmentation performance using a paired samples *t*-test with Bonferroni correction and Cohen's d (*d*) effect sizes. A two-sided *p*-value of less than 0.05 was considered statistically significant.

Results: The model-generated predictions had a high overlap ratio with the ground truth. The average DSC, HD95, and ASSD were significantly improved from 0.80 ± 0.12, 4.65 ± 4.71 mm, and 1.35 ± 1.15 mm of GAN to 0.85 ± 0.18 ($p = 0.001$, $d = 0.71$), 4.15 ± 7.56 mm ($p = 0.002$, $d = 0.67$), and 1.11 ± 1.65 mm ($p < 0.001$, $d = 0.46$) of PRG-GAN, respectively.

Conclusion: Integrating radiomics features into GAN is promising to solve unclear border limitations and increase the delineation accuracy of GTVp for patients with NPC.

Juebin Jin and Jicheng Zhang contributed equally to this study.

KEYWORDS
adversarial learning, computed tomography (CT), nasopharyngeal carcinoma segmentation, radiomics

1 | INTRODUCTION

Nasopharyngeal carcinoma (NPC) is an epithelial carcinoma arising from the nasopharyngeal mucosal lining, which is prevalent in Southeast Asia and South China.[1] With the advance of radiotherapy, intensity-modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT) have become the standard precise radiotherapy techniques for NPC.[2–4] The proximity to critical neural and other organs demands accurate delineation of primary gross tumor volume (GTVp) to avoid unnecessary toxicities from radiation.[5] Accurate delineation of target volumes is critical in the radiotherapy treatment planning of NPC, which is usually conducted manually by radiation oncologists in a treatment planning system (TPS).[6] However, manual delineation of NPC is time-consuming and labor-intensive. It takes around 2.7 h on average to manually draw targets for a single head and neck case,[7] and the process is prone to variations among physicians.[8] Therefore, methods of automatic GTVp segmentation are urgently needed to relieve oncologists from heavy work and improve the consistency and accuracy.

Traditionally, automatic segmentation methods based on atlas, grayscale, and shape have been investigated for NPC.[9–11] The atlas-based method with unsupervised mode is one of the most widely used techniques in the automatic segmentation of targets or organs at risk (OARs) in head-and-neck radiotherapy.[12] However, atlas-based methods rely on image registration for new contours, which results in an inherent weakness of sensitive to the accuracy of deformable registration algorithm.[13] Grayscale based methods rely on grayscale difference, which limits their application in NPC due to a low contrast pair. Shape based methods require high professional knowledge of users to manually adjust many parameters, which hinders its widespread clinical application.[14]

In the last decade, with the development and achievement of deep learning (DL) in computer vision processing and medical image analysis,[15] convolutional neural networks (CNNs) have achieved wide applications in automatic segmentation for many cancers, such as breast cancer,[16] esophageal cancer,[17] and NPC.[18] DL-based segmentation methods are mainly divided into 2D and 3D models. Individual slice analysis with 2D convolutions cannot retrieve all useful information and may affect the segmentation performance.[19] On the other side, 3D CNNs are able to achieve better results using convolution kernels in three directions to learn inter-slice information [20]. However, the unclear boundary problems are still not fully addressed in NPC segmentation.[21] Suggested methods of specialized network structures or modules or loss functions, and so forth, are lack of universality and flexibility.[22–24] Increasing data and data quality need appropriate enhancement techniques and strategies, and may increase computational costs or overfitting.[25]

In recent years, because of the advantages of generative adversarial network (GAN),[26] such as less demand for training data, good generation effect, and easy combination with other neural networks, the segmentation methods based on GAN have been increasingly applied in medical image research.[27] Huang et al.[28] proposed a GAN with a modified UNet[29] integrating with a Transformer[30] as the generator to achieve automatic segmentation for NPC. The introduction of Transformer improved the discriminant ability on tumor boundaries. However, the images were treated as 2D slices instead of 3D volumes in the experiments. Liu et al.[31] proposed an automatic segmentation network for NPC based on adversarial learning and UNet to solve the problem of sample imbalance. Overall, these studies provided some solutions for NPC segmentation, but the issue of unclear boundaries has not been well resolved.

With the emergence of radiomics, studies have demonstrated that radiomics features can characterize tumor microstructures and heterogeneity for various cancers, including NPC.[32,33] Specifically, radiomics features from peritumoral regions are helpful in describing tumor heterogeneity and prognostic prediction for NPC,[34] which may provide potential information to solve the unclear boundary problem. Additionally, studies demonstrated that introducing additional information as priori knowledge and solving it pertinently may provide a deeper understanding of tasks and data to guide model design and training during automatic segmentation.[35] High accuracy of voxel segmentation was achieved with comprehensive spatiotemporal features and radiomics features in the simultaneous segmentation of myocardial infarction on MRI images and liver tumor segmentation on CT images without contrast agents.[36,37]

In summary, although several studies have explored the effectiveness of combining radiomics and DL features in disease classification and segmentation tasks,[36–38] none have directly and systematically investigated the integration of peritumoral radiomics features with DL features specifically for segmentation tasks. This study is the first attempt to validate the feasibility of integrating radiomics features from peritumoral regions into GAN to improve the automatic segmentation

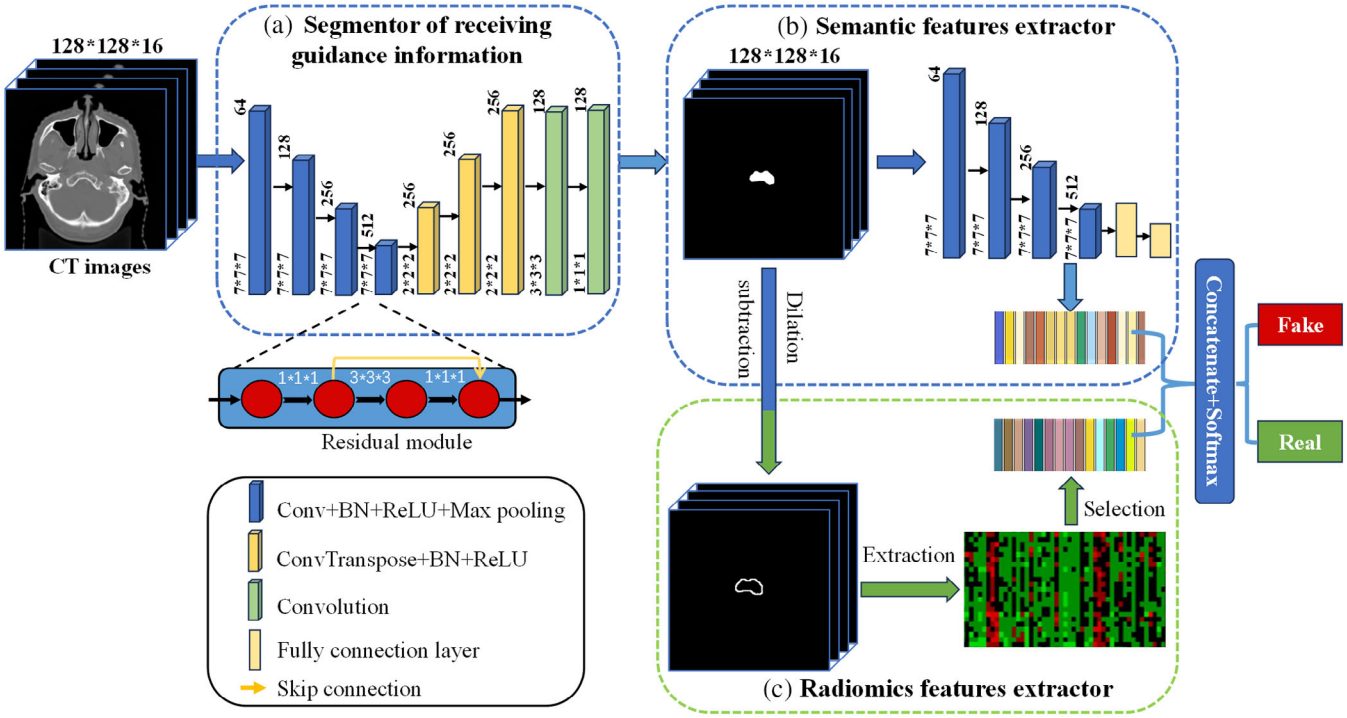

FIGURE 1 Overview of the proposed PRG-GAN. It segments NPC tumors contains three parts: (a) 3D segmentor of receiving guidance information. (b) Deep learning networks for extracting semantic features. (c) Radiomics features extraction and selection. NPC : nasopharyngeal carcinoma; PRG-GAN: peritumoral radiomics-guided GAN.

accuracy of GTVp for NPC on CT images. This study is the first attempt to validate the feasibility of integrating radiomics features from peritumoral regions into GAN to improve the automatic segmentation accuracy of GTVp for NPC on CT images. First, morphological dilation on the initial segmentation results from GAN was conducted to delineate the annular peritumoral region, in which radiomics features were extracted as priori guide knowledge. Then, radiomics features were fused with semantic features through the discriminator's fully connected layer to enhance the model's discriminatory ability to identify tumor boundaries, ultimately to achieve voxel-level classification and segmentation.

2 | MATERIALS AND METHODS

2.1 | Study design

The workflow and network architecture of the proposed peritumoral radiomics guided GAN (PRG-GAN) were shown in Figure 1, which includes three steps: (1) obtaining the optimal architecture of generator and discriminator for GAN among six different combinations; (2) selecting the most related and stable radiomics features to assist the discriminator in classification and segmentation; (3) developing a radiomics-guided adversarial learning architecture for automatic segmentation of GTVp for NPC. Basically, relying on the preliminary segmentation results with GAN, radiomics features were extracted and screened and then added as priori knowledge into the fully connected layer of the discriminator to enhance the discriminant ability during automatic segmentation.

2.2 | Patients and imaging data

A total of 552 patients diagnosed as NPC from Authors' Hospital from January 1 2017 to January 31 2021 were enrolled and analyzed in this study through searching electronic medical records. The inclusion criteria are as follows: (1) histologically confirmed as NPC; (2) without lymph node metastasis and bone metastasis; (3) with planning CT including nasopharynx and neck. Patients with low resolution CT images that will affect the accurate delineation of GTVp were excluded. The flowchart of patient enrollment was shown in Figure 2. Finally, a total of 157 patients were enrolled and randomly split into three cohorts: a training cohort (118 patients), a validation cohort (9 patients), and a testing cohort (30 patients), respectively. A number of 97 to 176 images slices were acquired for NPC patients during CT scan at 3 mm slice spice with an average resolution of 0.68 mm × 0.68 mm. CT images were acquired with a standard setting of 120 kV and 100–400 mA with a

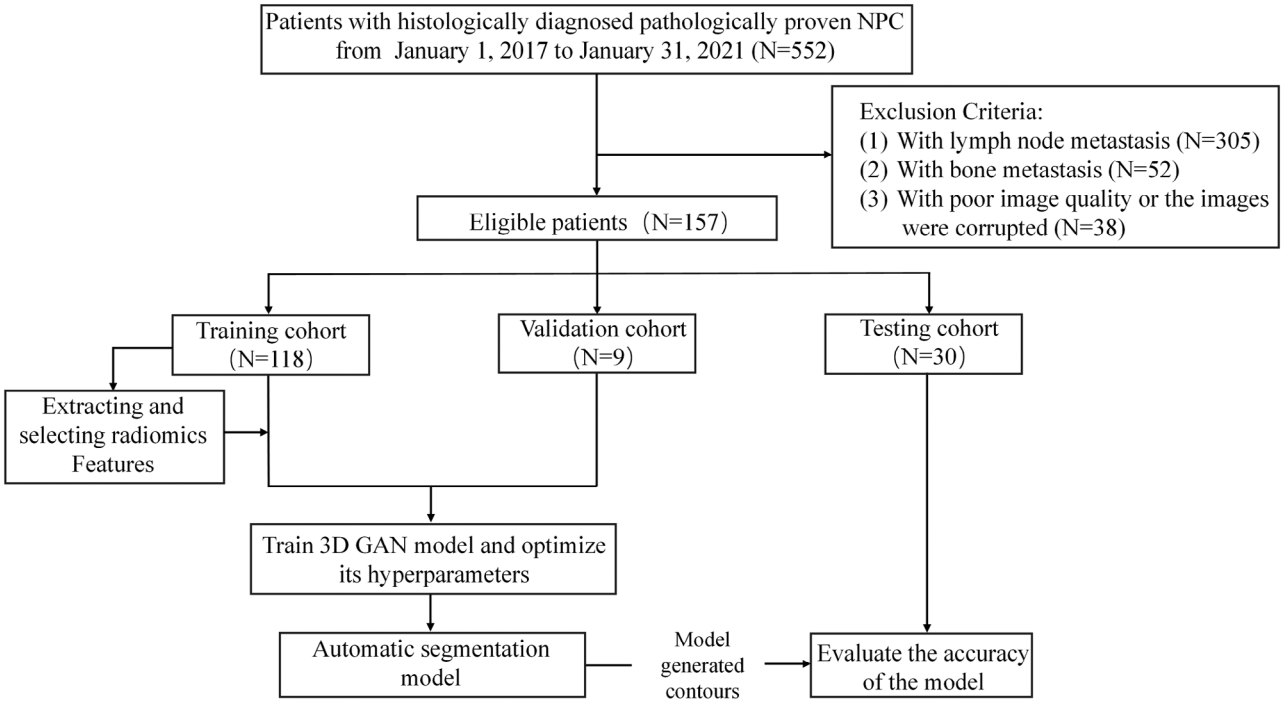

FIGURE 2 Study flow diagram. GAN: Generative Adversarial Network, NPC: nasopharyngeal carcinoma, 3D: three dimensional.

simulator CT of Philips Brilliance (Phillips, USA). GTVp were delineated by a radiation oncologist with more than 5 years clinical experience and confirmed by another oncologist with over 20 years of experience.

2.3 | Preprocessing and network architecture selection

CT images were firstly resized to 512 × 512 via bilinear interpolation. The voxel value of each image was normalized to the range of [0, 1] using min-max normalization. Generator and discriminator are the two competing modules in the GAN, which interact with each other and consist of two seamlessly connected networks. The input of the generator network (as shown in Figure 1a) is the ground truth images, which are randomly cropped into a patch of 128 × 128 × 16. The downsampling gradually reduces the spatial resolution and increases the semantic dimension to restore the abstract image representation, and then the upsampling process restores the detailed information and gradually restores the precision of images. Residual connections in a generator can effectively improve the performance of backpropagation with updated weights to learn better and adapt to data.[39] In this study, the radiomics features were combined with high-dimensional semantic features by discriminator to guide the generator using radiomics-guided adversarial mechanism.[36] In order to obtain the optimal combination of generator and discriminator for GAN, generators of 3D UNet,[40] 3D DenseNet,[41] 3D ResNet,[42] and discriminators of ResNet[39] and VGG[43] were investigated to determine the basic architecture of GAN.

These networks were optimized using optimizer Adam[44] with a batch size of 8 and trained in pytorch v1.11.0 and CUDA v11.3 library on the Ubuntu 16.04.7 LTS 64 bit platform with one Intel Core i7-8700K CPU @ 3.7 GHz × 12 and one NVIDIA GeForce RTX 2080Ti/PCIe/SSE2 with 11GB GPU memory. All the models were trained for 600 epochs with a learning rate of 2 × 10-4 and 2 × 10-6 for the generator and discriminator, respectively.

2.4 | Evaluation metrics

The Dice similarity coefficient (DSC), 95% Hausdorff distance (HD95), and average symmetric surface distance (ASSD) were applied to evaluate the performance of GAN in the segmentation of GTVp. These metrics are defined as follows.

$$\mathrm{DSC} = \frac{2\,|A \cap B|}{|A| + |B|} \tag{1}$$

DSC is a spatial overlap-based metric that can be used to assess the spatial overlap between the ground truth B and the automated segmentation A. The value of DSC lies in the range of [0, 1], where 0 indicates no

spatial overlap between the two segmentation volumes, and 1 indicates a complete overlap.

$$\mathrm{ASSD} = \frac{1}{2}\left\{\underset{b\in B_s}{\text{mean}}\ \underset{a\in A_s}{\min}\, d(a,b) + \underset{a\in A_s}{\text{mean}}\ \underset{b\in B_s}{\min}\, d(a,b)\right\} \quad (2)$$

$$\mathrm{HD95}\ (B_s, A_s) = \max\left(d_{95}(B_s, A_s), d_{95}(A_s, B_s)\right) \quad (3)$$

where

$$d_{95}\ (B_s, A_s) = K^{95}_{b\in B_s}\left(\min_{a\in A_s}\|a-b\|\right) \quad (4)$$

HD95 and ASSD describe the mean surface distance between the ground truth and automated delineation. A lower value indicates a higher delineation accuracy. For distance metrics, where B_s and A_s are the corresponding surface voxel sets of B and A, $d(a, b)$ and $\|a - b\|$ are the Euclidean distances of the voxel between a and b; $d_{B_sA_s}$ describes the point $x \in X_s$ that is farthest from any point of Y_s and calculates the distance from x to its nearest neighbor in Y_s.

2.5 | Radiomics features extraction and screening

Patients in the training cohort were used for radiomics feature screening. A morphologic dilation of 1, 3, 5, and 7 mm was conducted outside of the segmented GTVp from GAN to generated peritumoral regions. Therefore, a total of nine ROIs were generated from each sample, namely, intratumoral ROI (I), annular peritumoral ROIs of 1, 3, 5, and 7 mm (P1, P3, P5, and P7), intratumoral + peritumoral ROIs of 1, 3, 5 and 7 mm (IP1, IP3, IP5, and IP7), for radiomics features extraction. The diagram and examples of the image segmentation scheme are shown in Figure 3a. A total of 1274 features, which includes 18 first-order features, 22 Gray-level co-occurrence matrix (GLCM) features, 16 Gray-level size zone matrix (GLSZM) features, 16 Gray-level run length matrix (GLRLM) features, 5 Neighbouring Gray Tone Difference Matrix (NGTDM) features, 14 Gray Level Dependence Matrix (GLDM) Features, 455 log-sigma features and 728 wavelet features, were extracted from each ROIs using PyRadiomics.[45] The recap of the feature extraction parameters was presented in Table S1.

The process of radiomics features selection was mainly divided into the following three steps. First, a pairwise *t*-test was used to obtain the *p*-value of each feature. Then, to eliminate the collinearity between features, the Pearson correlation coefficient was calculated between each pair of features. If the absolute value of the correlation coefficient was greater than 0.80 and the *p*-value was less than 0.05, the feature with the smaller *p*-value in the paired *t*-test was retained. Finally, the least absolute shrinkage and selection operator (LASSO) regression with 5-fold cross-validation were conducted to select optimal features to differentiate the segmentation results generated by GAN and the ground truth. Specifically, an additional model combining the model with the features from the intratumoral ROI (Table 1: model I) and the model with the features from the annular peritumoral ROI that showed the best performance (Table 1: one of the models P1-7 and IP1-7) was constructed to explore whether the classification performance could be improved. Two machine learning algorithms, namely support vector machine (SVM)[46] and logistic regression (LR),[47] were applied to establish classification models with the data randomly divided into training and testing sets at a ratio of 7:3 for 100 times. The workflow of feature screening was shown in Figure S1.

2.6 | Radiomics guided adversarial learning

The semantic features and radiomics features were integrated into the discriminator of GAN. As shown in Figure 1b,c, the intratumoral radiomics features were extracted and screened from the preliminary segmentation results with GAN, peritumoral radiomics features were obtained from the perform morphological dilation regions around the preliminary segmentation results. The semantic features were mapped to the same number as the screened radiomics features through a fully connected layer. The output of discriminator was obtained by feeding the concatenation of the semantic features and radiomics features to another fully connected layer. Then these results continuously optimized hyperparameters through the back propagation of the discriminator loss function, and finally extracted more representative depth features under the guidance of radiomics. With learning through this adversarial strategy, a more excellent model can be trained.

2.7 | Ablation and comparative experiments

In this paper, two experiments were designed to illustrate the advantages of the proposed method. First, an ablation experiment was conducted to demonstrate the effectiveness of peritumoral radiomics features in DL segmentation. Specifically, an adversarial learning strategy was first removed from GAN to obtain a GAN without a discriminator (GAN w/o D). Subsequently, the intratumoral and annular peritumoral radiomics features were added to the GAN, resulting in the intratumoral radiomics-guided GAN (IRG-GAN) and the PRG-GAN, respectively. Finally, the performance of the proposed model was compared with that of five currently popular 3D segmentation models.

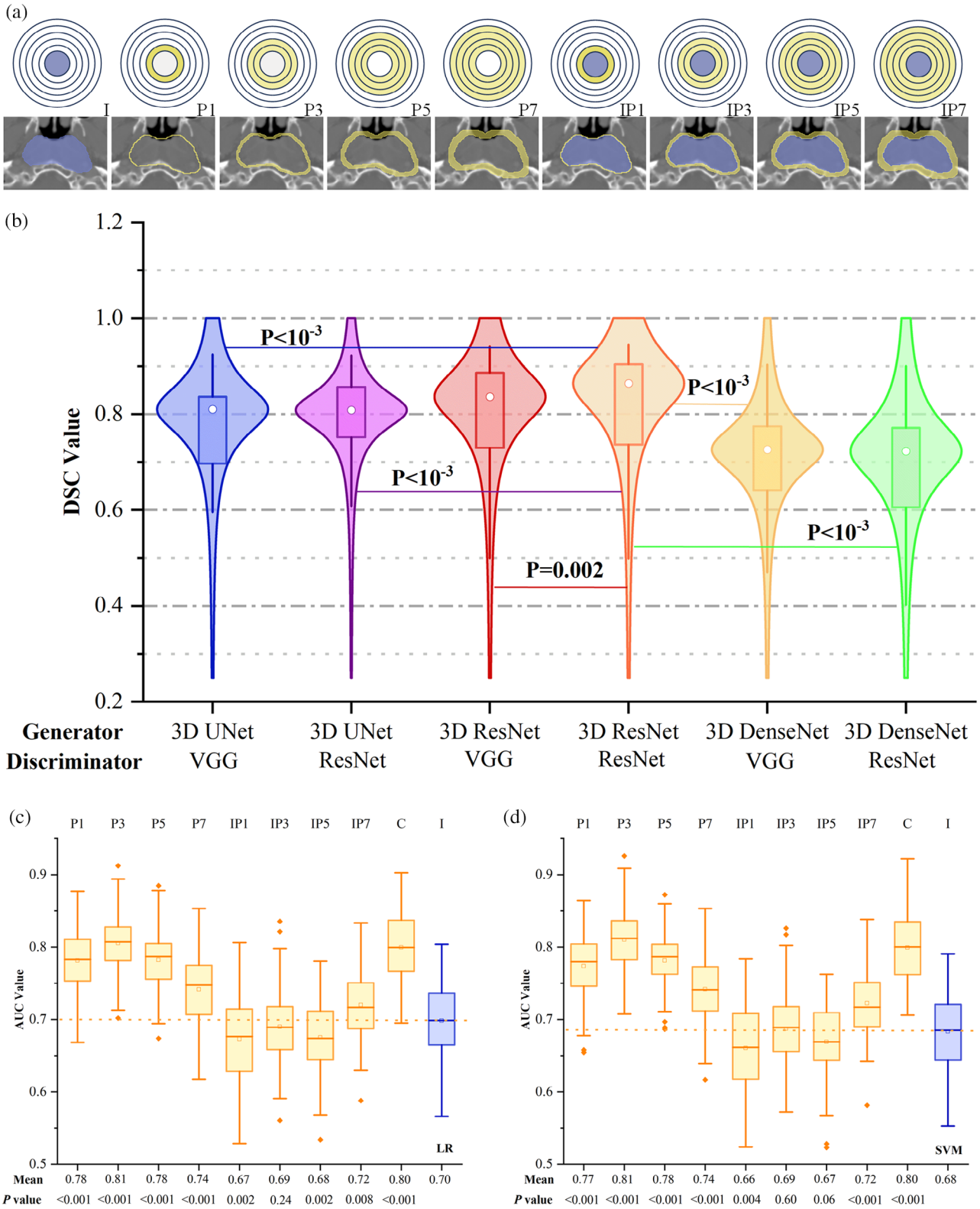

FIGURE 3 (a) Schematic illustration of the intratumoral and peritumoral ROIs and examples of the image segmentation scheme. (b) Results of violin plots for different combinations of input generator and discriminator. (c, d) Areas under the receiver operating characteristic curve (AUCs) of the ten radiomics models in the testing set. C, Combined model; DSC, Dice similarity coefficient; I, Intratumor; IP1, Intratumoral + peritumoral ROI of 1 mm; IP3, Intratumoral + peritumoral ROI of 3 mm; IP5, Intratumoral + peritumoral ROI of 5 mm; IP7, Intratumoral + peritumoral ROI of 7 mm; LR, logistic regression; P1, Annular peritumoral ROI of 1 mm; P3, Annular peritumoral ROI of 3 mm; P5, Annular peritumoral ROI of 5 mm; P7, Annular peritumoral ROI of 7 mm; ROI, regions of inSterest; SVM, support vector machine.

TABLE 1 Mean values of area under curve (AUC) of the receiver operating characteristic curve of the test set (71 samples) for radiomics models.

	LR				SVM			
Radiomics models	**Mean ± SD**	**95% CI**	***p*-value**	**Cohen's d**	**Mean ± SD**	**95% CI**	***p*-value**	**Cohen's d**
Intratumor (I)	0.70 ± 0.06	(0.69–0.71)	–	–	0.68 ± 0.05	(0.67–0.69)	–	–
Annular peritumoral ROI of 1 mm (P1)	0.78 ± 0.04	(0.77–0.79)	< 0.001	1.65	0.77 ± 0.05	(0.76–0.78)	< 0.001	1.80
Annular peritumoral ROI of 3 mm (P3)	**0.81 ± 0.04**	(0.79–0.81)	< 0.001	2.18	**0.81 ± 0.04**	(0.80–0.82)	< 0.001	2.65
Annular peritumoral ROI of 5 mm (P5)	0.78 ± 0.04	(0.77–0.79)	< 0.001	1.94	0.78 ± 0.04	(0.77–0.79)	< 0.001	2.21
Annular peritumoral ROI of 7 mm (P7)	0.74 ± 0.05	(0.73–0.75)	< 0.001	0.89	0.74 ± 0.05	(0.73–0.75)	< 0.001	1.10
Intratumor + peritumoral ROI of 1 mm (IP1)	0.67 ± 0.06	(0.66–0.68)	0.025	0.45	0.66 ± 0.06	(0.65–0.67)	0.070	0.41
Intratumor+peritumoral ROI of 3 mm (IP3)	0.69 ± 0.05	(0.68–0.70)	0.284	0.16	0.69 ± 0.05	(0.68–0.70)	0.867	0.07
Intratumor+peritumoral ROI of 5 mm (IP5)	0.68 ± 0.05	(0.67–0.68)	0.016	0.45	0.67 ± 0.05	(0.66–0.68)	0.154	0.28
Intratumor+peritumoral ROI of 7 mm (IP7)	0.72 ± 0.05	(0.71–0.73)	0.050	0.42	0.72 ± 0.05	(0.71–0.73)	< 0.001	0.78
Combined model: I + P3	0.80 ± 0.05	(0.79–0.81)	< 0.001	1.97	0.80 ± 0.05	(0.79–0.81)	< 0.001	2.31

Note: The bold statements represent the best results numerically. The reference base is Intratumor (I), the number of comparisons is 9, *differences were significant at $p < 0.006$ (0.05/9) corrected with Bonferroni's method.
Abbreviations: CI, confidence interval; LR, logistic regression; ROI, region of interest; SD, standard deviation; SVM, support vector machine.

2.8 | Constraint strategy of PRG-GAN

The basic adversarial learning strategy of the proposed PRG-GAN is coming from the minimax game between generator and discriminator in the primary GAN, in which the minimax optimization is formulated as:

$$\min_G \max_D \mathcal{L}_{\mathrm{GAN}}(G, D) = E_{y\sim p_{\mathrm{data}}(y)}\ [\log D(y)] + E_{x\sim p_x(x)}[\log(1 - D(G(x))] \quad (5)$$

where, G and D means the generator and discriminator, $\mathcal{L}_{\mathrm{GAN}}(G, D)$ is the loss function of GAN. In other words, G and D play the minimax game with the loss function of $\mathcal{L}_{\mathrm{GAN}}(G, D)$. The $p_{\mathrm{data}}(y)$ corresponds to the label of segmentation y and $p_x(x)$ is the distribution of input noise for the G. $\log D(y)$ and $\log(1 - D(G(x))$ are the logarithmic forms of the D outputs for real and generated samples, respectively. $G(x)$ corresponds to the prediction $\hat{y}$. PRG-GAN was trained to minimize the probability of $\hat{y}$ to be recognized while maximizing the probability of making mistakes of the discriminator when discriminating the $\hat{y}$. Studies showed that combining GAN losses with more traditional losses can effectively reduce ambiguity.[48,49] An additional Dice loss function $\mathcal{L}_{\mathrm{Dice}}$ and a cross entropy loss function $\mathcal{L}_{\mathrm{CE}}$ were added in the generator to obtain tumor segmentation loss function $\mathcal{L}_{\mathrm{SEG}}$, which is defined as:

$$\mathcal{L}_{\mathrm{Seg}} = \alpha\mathcal{L}_{\mathrm{Dice}} + (1 - \alpha)\,\mathcal{L}_{\mathrm{CE}} \quad (6)$$

where:

$$\mathcal{L}_{\mathrm{Dice}}\ (y, \hat{y}) = 1 - \frac{2\sum_{i=1}^{N} y_i\hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon} \quad (7)$$

$$\mathcal{L}_{\mathrm{CE}}\ (y, \hat{y}) = -(y\log(\hat{y}) + (1 - y)\log(1 - \hat{y})) \quad (8)$$

Here y_i is the real voxel value and $\hat{y}_i$ is the predicted voxel value. N is the total number of voxels, and ϵ is a small constant to avoid division by zero.

2.9 | Statistical analysis

All statistical analysis was performed with the SPSS software package (Version 26.0, IBM SPSS Inc). Categorical variables for the combined training-validation and testing cohorts were compared using the χ^2 test; numeric variables were compared using the Mann-Whitney U test. If the data were normally distributed, the paired t-test was performed; otherwise, the Wilcoxon Signed-Rank Test for Paired Samples non-parametric test was performed, and Cohen's d (d) effect sizes were calculated to quantify the observed difference between two sets of data. The criteria used were: 0.2 for small, 0.5 for medium, and 0.8 for large effects.[50] Numeric variables were denoted as mean ± standard deviation. The Wilcoxon rank-sum test with Bonferroni correction and the Kruskal-Wallis rank-sum test was used to

TABLE 2 Clinical and tumor characteristics.

Characteristic	Entire cohort (n = 157)	Training-validation cohort(n = 127)	Test cohort (n = 30)	p-value
Sex	…	…	…	0.638
Male	126 (80.25%)	101 (79.53%)	25 (83.33%)	…
Female	31 (19.75%)	26 (20.47%)	5 (16.67%)	…
Age	…	…	…	0.712
Overall age (years)	55.20 ± 11.84 (24–81)	55.15 ± 11.63 (28–81)	55.43 ± 12.90 (24–79)	…
< 30 y	3 (1.91%)	1 (0.79%)	2 (6.66%)	…
30–60 y	98 (62.42%)	82 (64.57%)	16 (53.33%)	…
≥ 60 y	56 (35.67%)	44 (34.64%)	12 (40.01%)	…
Tumor characteristic	…	…	…	
Number of tumor slices per case	12 (3–34)	12 (4–24)	13 (3–34)	0.564
Primary GTV (mL)	36.83 ± 25.67 (4.24–120.74)	36.72 ± 24.09 (4.24–120.74)	37.31 ± 31.98 (6.33–116.52)	0.492

Abbreviation: GTV, gross tumor volume.

compare the mean areas under the receiver operating characteristic (ROC) curve (AUCs) of each kind of ROI. A two-sided p-value of less than 0.05 was considered statistically significant. The multiple comparisons were corrected using the Bonferroni method to decrease the risk of a type I error.[51]

$$\alpha' = \frac{\alpha}{k} \tag{9}$$

where α is the level of type I error of each single test, most often takes 0.05. k denotes the number of comparisons, and α' stands for the corrected significance level.

3 | RESULTS

3.1 | The demographic and clinical characteristics of the participants

A total of 157 NPC patients were enrolled with a mean age of 55.20 ± 11.84 years (ranging from 24 to 81 years), of which 80.25% (126/157) were male, as shown in Table 2. The volume of GTVp ranged from 4.24 to 120.74 cm^3 (mean 36.83 ± 25.67 cm^3). No significant differences in the clinical variables were observed among the training, validation cohorts and testing cohort. Detailed characteristics of enrolled patients were shown in Table 2.

3.2 | The segmentation performance of different GAN combinations

The quantitative performance of the proposed DL models in test cohorts was summarized in Table 3 with consistent training epochs among these models. The combination of 3D ResNet and ResNet achieved an average DSC, HD95, and ASSD of 0.80 ± 0.12, 4.65 ± 4.71 mm, and 1.35 ± 1.15 mm, respectively. Except for the HD95 metric of 3D ResNet + VGG combination (p = 0.438, d = 0.21), other combination models demonstrated a significant difference in comparison with the baseline model of 3D ResNet + ResNet (all p-values < 0.01, all d > 0.5). The violin plots corresponding to the DSC values of segmentation results of different combinations were shown in Figure 3b, which visually reflect the segmentation performance of the combinations (all p-values < 0.01, represented by five lines in Figure 3b).

3.3 | The performance of radiomics models in different regions

Radiomics features screened after paired t-test, Pearson correlation analysis, and LASSO for different combinations along with their frequency of occurrence were summarized in Table S3–11. After re-splitting the whole dataset 100 times, the mean AUC values for 10 radiomics models in distinguishing segmentation results generated by GAN and ground truth were shown in Table 1. Model P3 constructed using the features from the annular peritumoral ROI of 3 mm showed good overall classification performance in the testing sets with an AUC of 0.81 ± 0.04 (95% confidence interval (CI) 0.79–0.81) and 0.81 ± 0.04 (95% CI 0.80–0.82) for LR and SVM, respectively. The performance of the models constructed by the features from the annular peritumoral ROI (Table 1: models P1–7) was good, with all AUCs greater than 0.74 in the testing set.

However, Except for model IP7 which achieved an AUC value of 0.72, all other models constructed by

TABLE 3 Accuracy comparison for model-generated GTVp in the testing cohorts by different combinations.

Model	DSC ↑			HD95 ↓			ASSD ↓		
	Values	*p*-value	Cohen's d	values	*p*-value	Cohen's d	Values	*p*-value	Cohen's d
3D ResNet + ResNet	**0.80 ± 0.12**	–	–	**4.65 ± 4.71**	–	–	**1.35 ± 1.15**	–	–
3D ResNet + VGG	0.78 ± 0.14	0.002*	0.55	4.94 ± 5.95	0.438	0.21	1.41 ± 1.12	0.005*	0.57
3D UNet + ResNet	0.76 ± 0.14	< 0.001*	1.06	6.44 ± 8.13	< 0.001*	0.83	1.82 ± 1.66	< 0.001*	1.16
3D UNet + VGG	0.75 ± 0.14	< 0.001*	1.05	6.54 ± 6.73	< 0.001*	0.55	1.74 ± 1.34	< 0.001*	0.75
3D DenseNet + ResNet	0.69 ± 0.13	< 0.001*	1.10	8.94 ± 7.08	< 0.001*	0.98	2.37 ± 1.36	< 0.001*	1.26
3D DenseNet + VGG	0.70 ± 0.12	< 0.001*	0.84	7.77 ± 6.01	< 0.001*	0.83	2.18 ± 1.15	< 0.001*	0.97

Note: The bold statements represent the best results numerically. The reference base is 3D ResNet + ResNet, and the number of comparisons is 5, *differences were significant at $p < 0.01$ (0.05/5) corrected with Bonferroni's method.
Abbreviations: ASSD, average symmetric surface distance; DSC, Dice similarity coefficient; GAN, generative adversarial network; GTVp, primary gross tumor volume; HD95, 95% Hausdorff distance; PRG-GAN: peritumoral radiomics-guided GAN.

TABLE 4 Quantitative results of the ablation experiment on the test set (30 patients).

Method	DSC ↑			HD95 ↓			ASSD ↓		
	Values	*p*-value	Cohen's d	Values	*p*-value	Cohen's d	Values	*p*-value	Cohen's d
GAN	0.80 ± 0.12	–	–	4.65 ± 4.71	–	–	1.35 ± 1.15	–	–
GAN w/o D	0.80 ± 0.17	0.504	0.14	4.82 ± 7.86	0.374	0.09	1.28 ± 1.67	0.229	0.17
IRG-GAN	0.82 ± 0.19	0.008*	0.27	4.68 ± 8.00	0.059	0.20	1.34 ± 1.66	0.002*	0.29
PRG-GAN	**0.85 ± 0.18**	0.001*	0.71	**4.15 ± 7.56**	0.002*	0.67	**1.11 ± 1.65**	<0.001*	0.46

Note: The bold statements represent the best results numerically. The reference base is GAN, and the number of comparisons is 3, *differences were significant at $p < 0.017$ (0.05/3) corrected with Bonferroni's method.
Abbreviations: ASSD, average symmetric surface distance; DSC, Dice similarity coefficient; GAN, generative adversarial network; GAN w/o D: GAN without discriminator; GTVp, primary gross tumor volume; HD95, 95% Hausdorff distance; IRG-GAN: intratumoral radiomics-guided GAN; PRG-GAN: peritumoral radiomics-guided GAN.

the features from the intratumoral + peritumoral ROI (Table 1: models IP1–7) did not show satisfactory performance, with all AUCs lower than 0.70 in the testing set. As shown in Figure 3c,d the visualized results using LR and SVM indicate that all annular peritumoral ROI models (Table 1: models P1-7) and the combined model have achieved significantly higher AUCs than model I (all *p*-values < 0.006, as shown at the bottom of Figure 3c,d, all $d > 0.8$).

3.4 | The segmentation performance of the ablation study

To verify the contributions of radiomics, an adversarial learning strategy was first removed from GAN to obtain GAN w/o D. The segmentation performance did not change significantly (DSC: 0.80 ± 0.12 vs. 0.80 ± 0.17, $p = 0.504$, $d = 0.14$; HD95: 4.65 ± 4.71 vs. 4.82 ± 7.86, $p = 0.374$, $d = 0.09$; ASSD: 1.35 ± 1.15 vs. 1.28 ± 1.67, $p = 0.229$, $d = 0.17$). Then the intratumoral and the annular peritumoral ROI of 3 mm radiomics features were added to GAN to obtain the IRG-GAN and the PRG-GAN, respectively. The quantitative results (6th row in Table 4) showed that the segmentation performance of PRG-GAN increased more significantly and achieved an average DSC, HD95, and ASSD of 0.85 ± 0.18 ($d = 0.71$), 4.15 ± 7.56 mm ($d = 0.67$), and 1.11 ± 1.65 mm ($d = 0.46$), respectively (all *p*-values < 0.017).

For visual comparison, the segmentation results of each method were showed in Figure 4. In the first and second examples, the tumor had a well-defined shape with a distinguishable border for easy segmentation. The third and fourth examples were more challenging than the first two. In the third example, the tumor was unshapely. In the fourth example, the tumor was surrounded by low contrast tissues. Additionally, both examples had a small volume. The four models were found to accurately segment the tumors in the first and second examples. Thus, the obtained contours fit well with the ground truth. In the third example, GAN mistakenly segmented the tumor. The error was partially corrected by combining intratumoral radiomics to obtain IRG-GAN. However, there were mistakes remained. By using peritumoral radiomics, PRG-GAN perfectly segmented the suspicious tumor. In the fourth example, GAN was able to determine the overall contours. Although IRG-GAN further refined the contours, the segmentation results of the proposed PRG-GAN were more consistent with the labeled results. The experiments demonstrated that the proposed PRG-GAN achieved the best visualization results compared with other methods.

3.5 | Comparison with current popular methods

As shown in Table 5, PRG-GAN combined with radiomics features from P3 achieved a best average

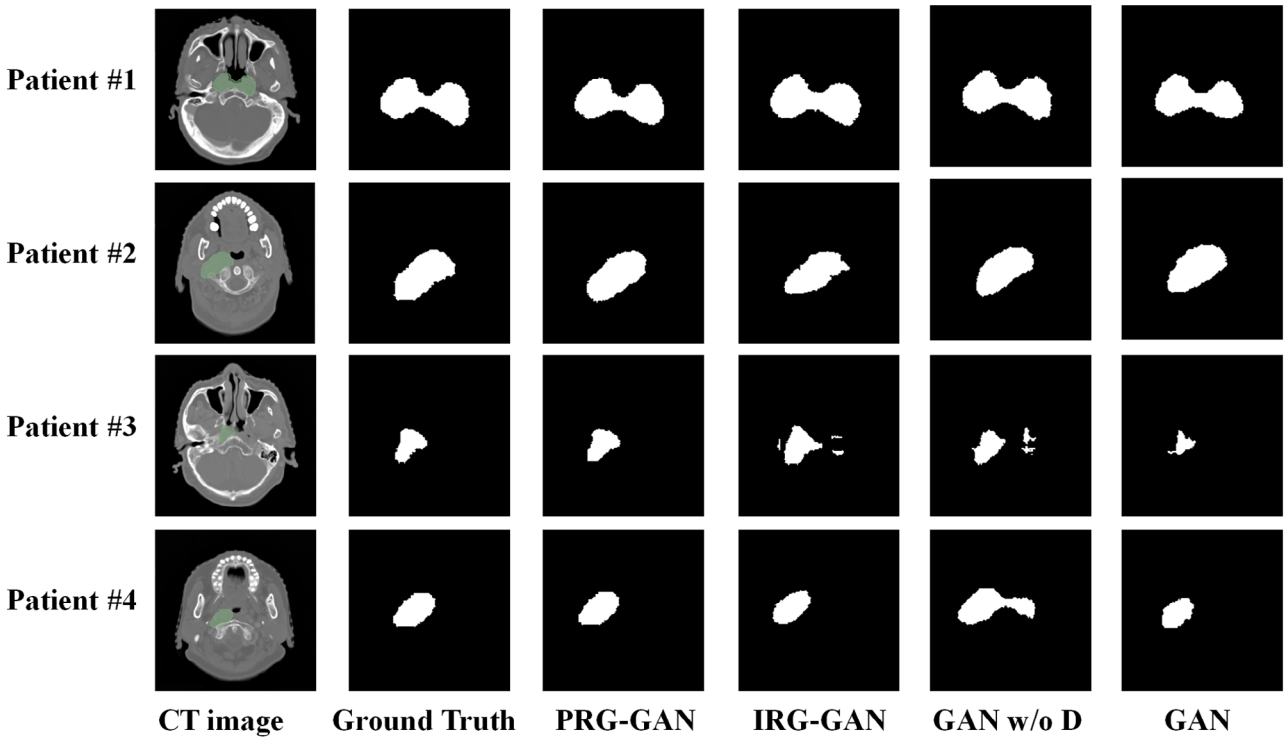

FIGURE 4 Four examples of the comparison of segmentation results from the proposed PRG-GAN, IRG-GAN, GAN w/o D, and GAN. GAN, generative adversarial network; GAN w/o D, GAN without a discriminator; IRG-GAN, intratumoral radiomics-guided GAN; PRG-GAN, peritumoral radiomics-guided GAN.

TABLE 5 Quantitative measurements on six segmentation methods.

Method	DSC ↑			HD95 ↓			ASSD ↓		
	Values	*p*-value	Cohen's d	Values	*p*-value	Cohen's d	Values	*p*-value	Cohen's d
3D-UNet[40]	0.76 ± 0.13	< 0.001*	1.27	6.39 ± 7.06	< 0.001*	0.44	1.70 ± 1.30	< 0.001*	0.81
3D-Vnet[52]	0.76 ± 0.13	0.001*	0.87	7.30 ± 8.68	< 0.001*	0.70	1.69 ± 1.39	< 0.001*	0.90
3D-ResUNet	0.76 ± 0.14	< 0.001*	0.83	5.82 ± 6.49	< 0.001*	0.43	1.56 ± 1.34	< 0.001*	0.71
3D-DenseNet[41]	0.67 ± 0.14	< 0.001*	0.92	10.80 ± 11.89	< 0.001*	0.77	2.56 ± 2.08	< 0.001*	0.89
3D-Dense-VoxNet[53]	0.70 ± 0.12	< 0.001*	1.16	11.85 ± 14.46	< 0.001*	0.54	2.74 ± 2.29	< 0.001*	0.71
PRG-GAN	**0.85 ± 0.18**	–	–	**4.15 ± 7.56**	–	–	**1.11 ± 1.65**	–	–

Note: The bold statements represent the best results numerically. The reference base is PRG-GAN, and the number of comparisons is 5, *Differences were significant at $p < 0.01(0.05/5)$ corrected with Bonferroni's method.
Abbreviation: ASSD, average symmetric surface distance; DSC, Dice similarity coefficient; HD95, 95% Hausdorff distance; PRG-GAN: peritumoral radiomics-guided GAN.

DSC, HD95, and ASSD of 0.85 ± 0.18, 4.15 ± 7.56 mm, and 1.11 ± 1.65 mm, respectively, which was significantly higher than those of other DL methods (all *p*-values < 0.01, all $d > 0.4$). The segmentation results on the four difficult cases after different methods were visualized in Figure 5. In the first and second examples, the tumor borders were unclear, and the volumes were small. In the third and fourth images, the tumor contrast was weak and the shapes were irregular. PRG-GAN also achieved the best visualization results, and its predicted contours fitted well with the ground truth. In comparison, other methods segmented the tumor mistakenly, and the obtained segmentation failed to overlap with the labeled truth significantly.

4 | DISCUSSION

In this work, a novel PRG-GAN model integrating peritumoral radiomics features with GAN was proposed for the robust delineation of GTVp for NPC on planning CT images. The peritumoral radiomics features and deep semantic features were concatenated in the fully connected layer of the discriminator of GAN to improve

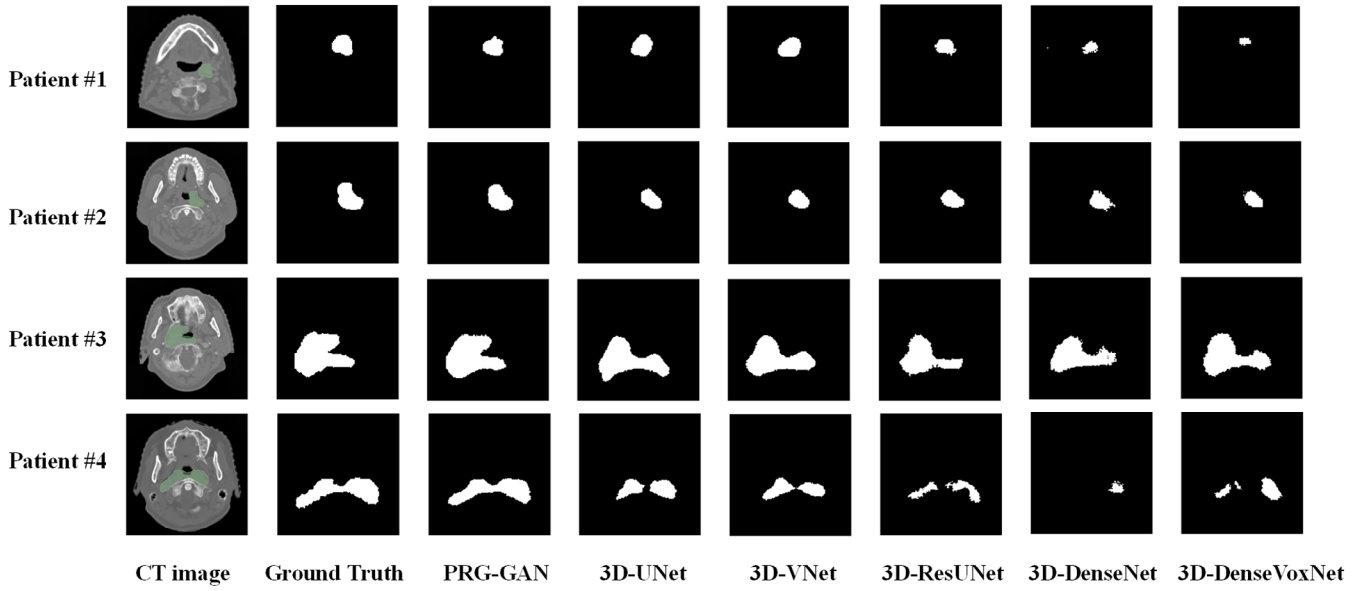

FIGURE 5 Comparative results on NPC segmentation after different 3D segmentation models. NPC, nasopharyngeal carcinoma

the model's ability of learning high discriminative features from tumors from CT images. In the design of PRG-GAN architecture, 3D convolution kernels instead of 2D were adopted to learn deep semantic features to capture slide to slide tumor connectivity. The average DSC, HD95, and ASSD were significantly improved from 0.80 ± 0.12, 4.65 ± 4.71 mm, and 1.35 ± 1.15 mm of GAN to 0.85 ± 0.18 ($p = 0.001$, $d = 0.71$), 4.15 ± 7.56 mm ($p = 0.002$, $d = 0.67$), and 1.11 ± 1.65 mm ($p < 0.001$, $d = 0.46$) of PRG-GAN, respectively.

DL models have been widely applied for the delineation of GTVp for NPC and other solid tumors.[18,54] In this study, different combinations of generators and discriminators in GAN were investigated for the GTVp delineation for NPC. As shown in Table 3, 3D DenseNet and 3D UNet combined with VGG and ResNet achieved a DSC ranging from 0.69 to 0.76, which was close to the reported DSC from 0.72 to 0.74 of 2D, 2.5D, and 3D UNet in the delineation of GTVp for NPC.[55–57] Using 3D ResNet as the generator and ResNet as the discriminator achieved the best DSC, HD95, and ASSD of 0.80 ± 0.12, 4.65 ± 4.71 mm, and 1.35 ± 1.15 mm, respectively. Similarly, Men et al.[58] achieved a mean DSC of 0.81 and 0.62 for GTVp and metastatic lymph nodes (GTVnd) with a deep deconvolutional neural network, respectively. As shown in Table 3, when the generator 3D ResNet was fixed, not all metrics showed significant improvement when the discriminator was replaced from VGG to ResNet (DSC: 0.80 ± 0.12 vs. 0.78 ± 0.14, $p = 0.002$, $d = 0.55$; HD95: 4.65 ± 4.71 vs. 4.94 ± 5.95, $p = 0.438$, $d = 0.21$; ASSD: 1.35 ± 1.15 vs. 1.41 ± 1.12, $p = 0.005$, $d = 0.57$). Among them, the improvement in the HD95 metric was not significant ($p = 0.438$) with only a small effect size ($d = 0.21$). Therefore, when the segmentation performance of the generator tends to saturate, hard to improve the performance with different discriminators.

Radiomics has been demonstrated to be able to quantitatively evaluate the quality of automatic delineation.[59,60] Specifically, Liu et al.[61] developed a radiomics-based quality control system to predict DSC scores for automatic cardiac segmentation at both the 2D and 3D models. In this study, ten radiomics models were generated with radiomics features extracted from intratumoral and peritumoral regions to evaluate the delineation results from GAN. As shown in Table 1, peritumoral radiomics models (Table 1, models P1-7) significantly outperform the intratumoral radiomics model (Table 1, models I) in the evaluation of GAN segmentation (all p-values < 0.006), and these improvements showed large effect sizes (all $d > 0.8$). This indicated that radiomics features from the peritumoral region were more valuable for automatic segmentation evaluation. On the other hand, the combined model did not improve performance significantly than the model P3 did (LR: 0.80 ± 0.05 [95% CI 0.79–0.81] vs. 0.81 ± 0.04 [95% CI 0.80–0.82]; SVM: 0.80 ± 0.05 [95% CI 0.79–0.81] vs. 0.81 ± 0.04 [95% CI 0.80–0.82]). Therefore, introducing the features of the annular peritumoral region of 3 mm into GAN may maximize the segmentation performance.

The value of radiomics in delineation was further evaluated in this study by combining radiomics features from different regions with GAN for GTVp delineation. As shown in Table 4, the adversarial learning strategy had no impact on segmentation performance because there was no statistical difference between the results of GAN and GAN w/o D (all p-values > 0.017) with small effect

sizes (all $d < 0.2$). Then the intratumoral radiomics features were added to GAN to obtain IRG-GAN. Although all metrics showed statistical improvement except for the HD95 metric (4.65 ± 4.71 vs. 4.68 ± 8.00, $p = 0.059$), all comparisons only had small to medium effect sizes (DSC: $d = 0.27$; HD95: $d = 0.20$; ASSD: $d = 0.29$). In contrast, the results demonstrated that introducing peritumoral radiomics features into GAN can further improve their delineation performance as shown in Table 4 that the performance of PRG-GAN was significantly better than GAN (DSC: 0.85 ± 0.18 vs. 0.80 ± 0.12, $p = 0.001$, $d = 0.71$; HD95: 4.15 ± 7.56 vs. 4.65 ± 4.71, $p = 0.002$, $d = 0.67$; ASSD: 1.11 ± 1.65 vs. 1.35 ± 1.15, $p < 0.001$, $d = 0.46$), the practical effect sizes were from medium to large. Cohen's d effect sizes between the model PRG-GAN and GAN were higher than those between the IRG-GAN and GAN (DSC: 0.71 vs. 0.27; HD95: 0.67 vs. 0.20; ASSD: 0.46 vs. 0.29). A possible reason was that there may be significant information overlap between intratumoral radiomics features and the DL features within the tumor, whereas peritumoral radiomics features can provide additional complementary information.

Table 5 shows the results of comparing our method with the current popular methods for 3D medical image segmentation. Among Cohen's d effect sizes across three metrics, the improvement in the DSC metric represents large effect sizes (all $d > 0.8$), while the ASSD metric improvement demonstrates medium or large effect sizes (all $d > 0.5$). The enhancement in the HD95 metric, though relatively modest, also represents close to or greater than medium effect sizes ($0.4 <$ all $d < 0.8$). Furthermore, the performance of PRG-GAN significantly surpasses that of other DL methods across all metrics listed in Table 5 (all p-values < 0.01). Figure 5 also shows that our method significantly outperforms the other compared methods.

This study had several limitations. First, the proposed PRG-GAN was conducted on CT images and only GTVp was investigated for the PRG-GAN delineation. Other target volumes, such as GTVnd, PTV should be further investigated with PRG-GAN in the future. Studies demonstrated that MRI images have superior soft-tissue contrast than CT images,[62] which is usually combined with CT images in NPC targets delineation in clinical practice. The combined use of FDG PET and MRI can comprehensively depict the pattern of nodal metastasis in NPC patients.[63] It is of great clinical significance to extract radiomics features from CT, MRI, and FDG PET to generate a radiomics features guided GAN for NPC delineation in the coming future. Another limitation of this study is that the number of samples is relatively small and from a single center. In the future, collecting more data from multiple institutions to improve the robustness and accuracy of the model is necessary with attempt to develop strong algorithms, such as transformer-based models.[64]

5 | CONCLUSIONS

In summary, peritumoral radiomics features were integrated into GAN to improve the accuracy of GTVp segmentation for NPC on planning CT images. Our findings demonstrated that peritumoral radiomics features can effectively improve contouring performance in adversarial learning strategy. It is promising to add radiomics features into GAN to solve unclear border limitation during GTVp delineation for patients with NPC.

ACKNOWLEDGMENTS

This research was supported partially by a National Natural Science Foundation (12475352), a Key project of Zhejiang Natural Science Foundation (Z24A050009), a Key project of Zhejiang Provincial Health Science and Technology Program (WKJ-ZJ-2437), a Major project of Wenzhou Science and Technology Bureau (ZY2022016, ZY2020011), a project of Wenzhou Science and Technology Bureau (Y2023798), Zhejiang Engineering Research Center for innovation and application of Intelligent Radiotherapy Technology, Zhejiang-Hong Kong Precision Theranostics of Thoracic Tumors Joint Laboratory, and Wenzhou key Laboratory of basic science and translational research of radiation oncology, Zhejiang Key Laboratory of Intelligent Cancer Biomarker Discovery and Translation, Discipline Cluster of Oncology, Wenzhou Medical University.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet North Am Ed*. 2019;394(10192):64–80. doi:10.1016/S0140-6736(19)30956-0
2. Verbakel WFAR, Cuijpers JP, Hoffmans D, Bieker M, Slotman BJ, Senan S. Volumetric intensity-modulated arc therapy Vs. conventional IMRT in head-and-neck cancer: a comparative planning and dosimetric study. *Int J Radiat Oncol Biol Phys*. 2009;74(1):252–259. doi:10.1016/j.ijrobp.2008.12.033
3. Yu W, Shang H, Xie C, et al. Feasibility of constant dose rate VMAT in the treatment of nasopharyngeal cancer patients. *Radiat Oncol*. 2014;9(1):235. doi:10.1186/s13014-014-0235-1
4. Jin X, Han C, Zhou Y, Yi J, Yan H, Xie C. A modified VMAT adaptive radiotherapy for nasopharyngeal cancer patients based on CT-CT image fusion. *Radiat Oncol*. 2013;8(1):277. doi:10.1186/1748-717X-8-277
5. Xia P, Fu KK, Wong GW, Akazawa C, Verhey LJ. Comparison of treatment plans involving intensity-modulated radiotherapy for nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. 2000;48(2):329–337. doi:10.1016/S0360-3016(00)00585-X
6. Zhang Q, Ou L, Peng Y, Yu H, Wang L, Zhang S. Evaluation of automatic VMAT plans in locally advanced nasopharyngeal

carcinoma. *Strahlenther Onkol*. 2021;197(3):177–187. doi:10.1007/s00066-020-01631-x
7. Harari PM, Song S, Tomé WA. Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2010;77(3):950–958. doi:10.1016/j.ijrobp.2009.09.062
8. Teguh DN, Levendag PC, Voet PWJ, et al. Clinical Validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncol Biol Phys*. 2011;81(4):950–957. doi:10.1016/j.ijrobp.2010.07.009
9. Tao CJ, Yi JL, Chen NY, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: a multi-institution clinical study. *Radiother Oncol*. 2015;115(3):407–411. doi:10.1016/j.radonc.2015.05.012
10. Lee FKH, Yeung DKW, King AD, Leung SF, Ahuja A. Segmentation of nasopharyngeal carcinoma (NPC) lesions in MR images. *Int J Radiat Oncol Biol Phys*. 2005;61(2):608–620. doi:10.1016/j.ijrobp.2004.09.024
11. Han X, Hoogeman MS, Levendag PC. Atlas-based auto-segmentation of head and neck CT images. In: Metaxas D, Axel L, Fichtinger G, Székely G, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008*. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2008:434–441. doi:10.1007/978-3-540-85990-1_52
12. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: a survey. *Med Image Anal*. 2015;24(1):205–219. doi:10.1016/j.media.2015.06.012
13. Cabezas M, Oliver A, Lladó X, Freixenet J, Cuadra MB. A review of atlas-based segmentation for magnetic resonance brain images. *Comput Methods Programs Biomed*. 2011;104(3):e158–e177. doi:10.1016/j.cmpb.2011.07.015
14. Huang KW, Zhao ZY, Gong Q, Zha J, Chen L, Yang R. Nasopharyngeal carcinoma segmentation via HMRF-EM with maximum entropy. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2015:2968–2972. doi:10.1109/EMBC.2015.7319015
15. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans Neural Netw Learning Syst*. 2022;33(12):6999–7019. doi:10.1109/TNNLS.2021.3084827
16. Li H, Chen D, Nailon WH, Davies ME, Laurenson D. Dual Convolutional Neural Networks for Breast Mass Segmentation and Diagnosis in Mammography. *IEEE Transactions on Medical Imaging*. 2021;41(1):3–13. doi:10.1109/TMI.2021.3102622
17. Wu Z, Ge R, Wen M, et al. ELNet:automatic classification and segmentation for esophageal lesions using convolutional neural network. *Med Image Anal*. 2021;67:101838. doi:10.1016/j.media.2020.101838
18. Lin L, Dou Q, Jin YM, et al. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. *Radiology*. 2019;291(3):677–686. doi:10.1148/radiol.2019182012
19. Niyas S, Pawan SJ, Kumar MA, Rajan J. Medical Image Segmentation with 3D Convolutional Neural Networks: A Survey. *Neurocomputing*. 2022;493:397–413. doi:10.48550/arXiv.2108.08467
20. El-Assal M, Tirilly P, Bilasco IM. 2D versus 3D Convolutional Spiking Neural Networks Trained with Unsupervised STDP for Human Action Recognition. *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022:1–8. doi:10.1109/IJCNN55064.2022.9892063
21. Wang Z, Fang M, Zhang J, et al. Radiomics and deep learning in nasopharyngeal carcinoma: a review. *IEEE Rev Biomed Eng*. 2024;17:118–135. doi:10.1109/RBME.2023.3269776
22. Hao Y, Jiang H, Diao Z, et al. MSU-Net: multi-scale sensitive U-Net based on pixel-edge-region level collaborative loss for nasopharyngeal MRI segmentation. *Comput Biol Med*. 2023;159:106956. doi:10.1016/j.compbiomed.2023.106956
23. Li Y, Dan T, Li H, et al. NPCNet: jointly segment primary nasopharyngeal carcinoma tumors and metastatic lymph nodes in MR images. *IEEE Trans Med Imaging*. 2022;41(7):1639–1650. doi:10.1109/TMI.2022.3144274
24. Liao W, He J, Luo X, et al. Automatic delineation of gross tumor volume based on magnetic resonance imaging by performing a novel semisupervised learning framework in nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. 2022;113(4):893–902. doi:10.1016/j.ijrobp.2022.03.031
25. Garcea F, Serra A, Lamberti F, Morra L. Data augmentation for medical imaging: a systematic literature review. *Comput Biol Med*. 2023;152:106391. doi:10.1016/j.compbiomed.2022.106391
26. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. *Communications of the ACM*. 2020;63(11):139–144. doi:10.48550/arXiv.1406.2661
27. Jeong JJ, Tariq A, Adejumo T, Trivedi H, Gichoya JW, Banerjee I. Systematic review of Generative Adversarial Networks (GANs) for medical image classification and segmentation. *J Digit Imaging*. 2022;35(2):137–152. doi:10.1007/s10278-021-00556-w
28. Huang Z, Tang S, Chen Z, et al. TG-Net: combining transformer and GAN for nasopharyngeal carcinoma tumor segmentation based on total-body uEXPLORER PET/CT scanner. *Comput Biol Med*. 2022;148:105869. doi:10.1016/j.compbiomed.2022.105869
29. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015:234–241. doi:10.48550/arXiv.1505.04597
30. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. Advances in Neural Information Processing Systems, 2017. doi:10.48550/arXiv.1706.03762
31. Liu Y, Yuan X, Jiang X, et al. Dilated adversarial U-Net network for automatic gross tumor volume segmentation of nasopharyngeal carcinoma. *Appl Soft Comput*. 2021;111:107722. doi:10.1016/j.asoc.2021.107722
32. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–577. doi:10.1148/radiol.2015151169
33. Mei ZY, Zhong GG, Tao QQ, Wei HY, Ming LH, Yin Y. Radiomics for diagnosis and radiotherapy of nasopharyngeal carcinoma. *Front Oncol*. 2022;11:767134. doi:10.3389/fonc.2021.767134
34. Li S, Wan X, Deng YQ, et al. Predicting prognosis of nasopharyngeal carcinoma based on deep learning: peritumoral region should be valued. *Cancer Imaging*. 2023;23(1):14. doi:10.1186/s40644-023-00530-5
35. Tao G, Li H, Huang J, et al. SeqSeg: a sequential method to achieve nasopharyngeal carcinoma segmentation free from background dominance. *Med Image Anal*. 2022;78:102381. doi:10.1016/j.media.2022.102381
36. Xiao X, Zhao J, Qiang Y. Radiomics-guided GAN for segmentation of liver tumor without contrast agents. In: Shen D, Liu T, Peters TM, eds. *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*. Lecture Notes in Computer Science. Springer International Publishing; 2019:237–245. doi:10.1007/978-3-030-32245-8_27
37. Xu C, Xu L, Brahm G, Zhang H, Li S. MuTGAN: simultaneous segmentation and quantification of myocardial infarction without contrast agents via joint adversarial learning. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, eds. *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*. Lecture Notes in Computer Science. Springer International Publishing; 2018:525–534. doi:10.1007/978-3-030-00934-2_59

38. Choi YS, Bae S, Chang JH, et al. Fully automated hybrid approach to predict the *IDH* mutation status of gliomas via deep learning and radiomics. *Neuro-oncol*. 2021;23(2):304–313. doi:10.1093/neuonc/noaa177
39. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:770–778. doi:10.1109/CVPR.2016.90
40. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer International Publishing, 2016:424–432. doi:10.48550/arXiv.1606.06650
41. Bui TD, Shin J, Moon T. 3D Densely Convolutional Networks for Volumetric Segmentation. arXiv preprint arXiv:1709.03199, 2017. doi:10.48550/arXiv.1709.03199
42. Hara K, Kataoka H, Satoh Y. Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition. In: *Proceedings of the IEEE international conference on computer vision workshops*. 2017:3154–3160. doi:10.48550/arXiv.1708.07632
43. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556, 2014. doi:10.48550/arXiv.1409.1556
44. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980, 2014. doi:10.48550/arXiv.1412.6980
45. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–e107. doi:10.1158/0008-5472.CAN-17-0339
46. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–297. doi:10.1007/BF00994018
47. Cramer JS. The origins of logistic regression. *SSRN Journal*. Tinbergen Institute Working Paper No. 2002-119/4. doi:10.2139/ssrn.360300
48. Saqlain AS, Fang F, Ahmad T, Wang L, Abidin Zu. Evolution and effectiveness of loss functions in generative adversarial networks. *China Commun*. 2021;18(10):45–76. doi:10.23919/JCC.2021.10.004
49. Pan Z, Yu W, Wang B, et al. Loss functions of Generative Adversarial Networks (GANs): opportunities and challenges. *IEEE Trans Emerg Top Comput Intell*. 2020;4(4):500–522. doi:10.1109/TETCI.2020.2991774
50. Chen H, Cohen P, Chen S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Commun Stat Simul Comput*. 2010;39(4):860–864. doi:10.1080/03610911003650383
51. Curtin F, Schulz P. Multiple correlations and bonferroni's correction. *Biol Psychiatry*. 1998;44(8):775–777. doi:10.1016/S0006-3223(98)00043-2
52. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 fourth international conference on 3D vision (3DV). Ieee, 2016:565–571. doi:10.48550/arXiv.1606.04797
53. Yu L, Cheng JZ, Dou Q, et al. Automatic 3D Cardiovascular MR Segmentation with Densely-Connected Volumetric ConvNets. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20. Springer International Publishing, 2017:287–295. doi:10.48550/arXiv.2103.10504
54. Boers TGW, Hu Y, Gibson E, et al. Interactive 3D U-net for the segmentation of the pancreas in computed tomography scans. *Phys Med Biol*. 2020;65(6):065002. doi:10.1088/1361-6560/ab6f99
55. Li S, Xiao J, He L, Peng X, Yuan X. The tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning methods. *Technol Cancer Res Treat*. 2019;18:153303381988456. doi:10.1177/1533033819884561
56. Domoguen JKL, Manuel JJA, Cañal JPA, Naval PC. Automatic segmentation of nasopharyngeal carcinoma on CT images using efficient UNet-2.5D ensemble with semi-supervised pretext task pretraining. *Front Oncol*. 2022;12:980312. doi:10.3389/fonc.2022.980312
57. Wang X, Yang G, Zhang Y, et al. Automated delineation of nasopharynx gross tumor volume for nasopharyngeal carcinoma by plain CT combining contrast-enhanced CT using deep learning. *J Radiation Res Appl Sci*. 2020;13(1):568–577. doi:10.1080/16878507.2020.1795565
58. Men K, Chen X, Zhang Y, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Front Oncol*. 2017;7:315. doi:10.3389/fonc.2017.00315
59. Sunoqrot MRS, Selnæs KM, Sandsmark E, et al. A quality control system for automated prostate segmentation on T2-weighted MRI. *Diagnostics*. 2020;10(9):714. doi:10.3390/diagnostics10090714
60. Maffei N, Manco L, Aluisio G, et al. Radiomics classifier to quantify automatic segmentation quality of cardiac sub-structures for radiotherapy treatment planning. *Physica Med*. 2021;83:278–286. doi:10.1016/j.ejmp.2021.05.009
61. Liu Q, Lu Q, Chai Y, et al. Radiomics-based quality control system for automatic cardiac segmentation: a feasibility study. *Bioeng*. 2023;10(7):791. doi:10.3390/bioengineering10070791
62. Olmi P, Fallai C, Colagrande S, Giannardi G. Staging and follow-up of nasopharyngeal carcinoma: magnetic resonance imaging versus computerized tomography. *Int J Radiat Oncol Biol Phys*. 1995;32(3):795–800. doi:10.1016/0360-3016(94)00535-S
63. Ng SH, Chang JC, Chan SC, et al. Nodal metastases of nasopharyngeal carcinoma: patterns of disease on MRI and FDG PET. *Eur J Nucl Med Mol Imaging*. 2004;31(8):1073–1080. doi:10.1007/s00259-004-1498-9
64. Hatamizadeh A, Tang Y, Nath V, et al. Unetr: Transformers for 3d medical image segmentation. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2022:574–584. doi:10.48550/arXiv.2103.10504

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jin J, Zhang J, Yu X, et al. Radiomics-guided generative adversarial network for automatic primary target volume segmentation for nasopharyngeal carcinoma using computed tomography images. *Med Phys*. 2024;1-14. https://doi.org/10.1002/mp.17493