Technology in Cancer Research & Treatment
Volume 23: 1-12

DOI: 10.1177/15330338241242654
journals.sagepub.com/home/tct

A Comparative Study of Deep Learning Dose Prediction Models for Cervical Cancer Volumetric Modulated Arc Therapy

Zhe Wu, PhD[1,2,*], Mujun Liu, PhD[1,*], Ya Pang, BS[2], Lihua Deng, PhD[3], Yi Yang, MS[1], and Yi Wu, PhD[1]

Abstract

Purpose: Deep learning (DL) is widely used in dose prediction for radiation oncology, multiple DL techniques comparison is often lacking in the literature. To compare the performance of 4 state-of-the-art DL models in predicting the voxel-level dose distribution for cervical cancer volumetric modulated arc therapy (VMAT). **Methods and Materials:** A total of 261 patients' plans for cervical cancer were retrieved in this retrospective study. A three-channel feature map, consisting of a planning target volume (PTV) mask, organs at risk (OARs) mask, and CT image was fed into the three-dimensional (3D) U-Net and its 3 variants models. The data set was randomly divided into 80% as training-validation and 20% as testing set, respectively. The model performance was evaluated on the 52 testing patients by comparing the generated dose distributions against the clinical approved ground truth (GT) using mean absolute error (MAE), dose map difference (GT-predicted), clinical dosimetric indices, and dice similarity coefficients (DSC). **Results:** The 3D U-Net and its 3 variants DL models exhibited promising performance with a maximum MAE within the PTV $0.83\% \pm 0.67\%$ in the UNETR model. The maximum MAE among the OARs is the left femoral head, which reached $6.95\% \pm 6.55\%$. For the body, the maximum MAE was observed in UNETR, which is $1.19 \pm 0.86\%$, and the minimum MAE was $0.94 \pm 0.85\%$ for 3D U-Net. The average error of the Dmean difference for different OARs is within 2.5 Gy. The average error of V40 difference for the bladder and rectum is about 5%. The mean DSC under different isodose volumes was above 90%. **Conclusions:** DL models can predict the voxel-level dose distribution accurately for cervical cancer VMAT treatment plans. All models demonstrated almost analogous performance for voxel-wise dose prediction maps. Considering all voxels within the body, 3D U-Net showed the best performance. The state-of-the-art DL models are of great significance for further clinical applications of cervical cancer VMAT.

Keywords

cervical cancer, volumetric modulated arc therapy (VMAT), 3D dose prediction, deep learning, 3D U-Net variants

[1] Department of Digital Medicine, School of Biomedical Engineering and Medical Imaging, Army Medical University (Third Military Medical University), Chongqing, China
[2] Department of Radiation Oncology, Zigong Disease Prevention and Control Center Mental Health Center, Zigong First People's Hospital, Zigong, Sichuan, China
[3] Department of Radiology, The First Affiliated Hospital of the Army Medical University, Chongqing, China

*These authors contributed equally.

Corresponding Authors:
Yi Wu, Department of Digital Medicine, School of Biomedical Engineering and Medical Imaging, Army Medical University, Gaotanyan Main Street, Shapingba District, Chongqing 400038, China.
Email: wuy1979@tmmu.edu.cn

Mujun Liu, Department of Digital Medicine, School of Biomedical Engineering and Medical Imaging, Army Medical University, Gaotanyan Main Street, Shapingba District, Chongqing 400038, China.
Email: 2249717501@qq.com

Abbreviations

3D, three-dimensional; 3DAtten U-Net, 3D attention U-Net; AXB, acuros external beam algorithm; CNN, convolutional neural network; DL, deep learning; DSC, dice similarity coefficients; DVH, dose-volume histogram; GT, ground truth; H&N, head and neck; HD U-Net, hierarchically densely connected U-Net; HI, homogeneity index; IN, instance normalization; KBP, knowledge-based planning; MAE, mean absolute error; OARs, organs at risk; PTV, planning tumor volume; ReLU, rectified linear unit; ROIs, regions of interest; TPS, treatment planning system; UNETR, unet transformers; VMAT, volumetric modulated arc therapy.

Received: September 20, 2023; Revised: December 19, 2023; Accepted: February 19, 2024.

Introduction

Cervical cancer is a highly malignant tumor of the female genital system. It is the fourth most common malignancy in females globally.[1] The global incidence of cervical cancer is ~500 000 cases annually.[2] Radiotherapy is a cornerstone of treatment in most stages of cervical cancer, and concurrent chemotherapy bestows an additional survival benefit.[3] In the process of designing a radiation treatment plan, a time-consuming, labor-intensive inverse optimization executed by an experienced dosimetrist in a trial-and-error manner is needed. The quality of the plan determines the efficacy of radiotherapy and even affects the safety of radiotherapy.

Currently, the research of artificial intelligence and deep learning (DL) has made astonishing progress, particularly in the field of computer vision and decision making.[4] In 2015, Ronneberger et al[5] proposed a DL convolutional neural network (CNN) architecture for semantic segmentation, known as U-Net, which was widely used in medical image segmentation and radiation dose prediction. Osman et al[6] proposed an attention-aware three-dimensional (3D) U-Net CNN for knowledge-based planning (KBP) 3D dose distribution prediction of head-and-neck (H&N) cancer based on the OpenKBP-Grand Challenge.[7] The proposed attention-gated 3D U-Net model showed high capability in accurately predicting 3D dose distributions that closely replicated the ground-truth dose distributions of 68 plans in the test set. Nguyen et al[8] researched the 3D radiotherapy dose prediction for 120 H&N patients based on the proposed U-Net variant (hierarchically densely connected U-Net, HD U-Net). As a U-Net variant, the new state-of-the-art model, U-Net transformers (UNETR), has acquired high dice coefficients in segmentation.[9,10] UNETR utilizes a U-Net architecture and transformer as the encoder to extract important global information and has never been used in radiation dose prediction. The performance of these DL models for dose prediction strongly depends on the data set used for training. So it is difficult to compare the testing results of different models in different data pools. It is also necessary to make meaningful comparisons of these DL models using the same evaluation metrics in the same data sets.

Attention-gating mechanism highlights important anatomy features and suppresses redundant information propagation, and the advantage of the transformer-based networks is the self-attention mechanism and the capability of learning long-range dependencies, which could be compensated for the shortcomings of CNN. To our knowledge, there are no studies about the comparison of these diverse novel networks used for cervical cancer dose prediction. Which CNN is the optimal model, that can be used in cervical cancer radiotherapy is still unknown.

In this study, we aimed to use the same patients' data sets and quantitative assessment metrics to evaluate the performance of various novel KBP 3D U-Net and its variants models for cervical cancer volumetric modulated arc therapy (VMAT) 3D dose predictions. The comprehensive comparative study facilitates the exploration of the performance of different models and provides a base for further innovations in algorithms.

Materials and Methods

Patient Materials

Data from 261 patients with cervical cancer in our center were collected as a retrospective study. Manual planning and treatment were performed for all patients with 2 coplanar VMAT full arcs with a photon energy of 6 MV beam. All the dose volumes were scaled to 95% of the PTV receiving 100% of the prescription dose with 50 Gy for 25 fractions. Data of 261 patients were randomly divided into 80%, and 20%, namely 209 patients for 5-fold cross-training-validation and 52 patients for testing.

All treatments were planned using the Varian Eclipse treatment planning system (TPS, version 13.6, Varian Oncology Systems, Palo Alto, CA, USA) with the Acuros External Beam Algorithm (AXB) and Trilogy linear accelerator (Varian Oncology Systems, Palo Alto, CA, USA). All these plans were designed by 3-year and 13-year experienced dosimetrists. The study protocol was approved by the Ethics Committee of the authors' hospital (No. Ethics [M] 2024-015).

Dose Prediction Models

The network used was based on Pytorch 1.10.0 and Python 3.8. Figure 1 shows the dose prediction DL models for dose prediction, which consists of a down-sampling (encoding) path and an up-sampling (decoding) path. Four 3D U-Net-based variants architectures for KBP dose distribution predictions of cervical cancer were evaluated in our study. These architectures are, namely (A) 3D U-Net, (B) 3DAtten U-Net, (C) HD U-Net, and (D) UNETR.

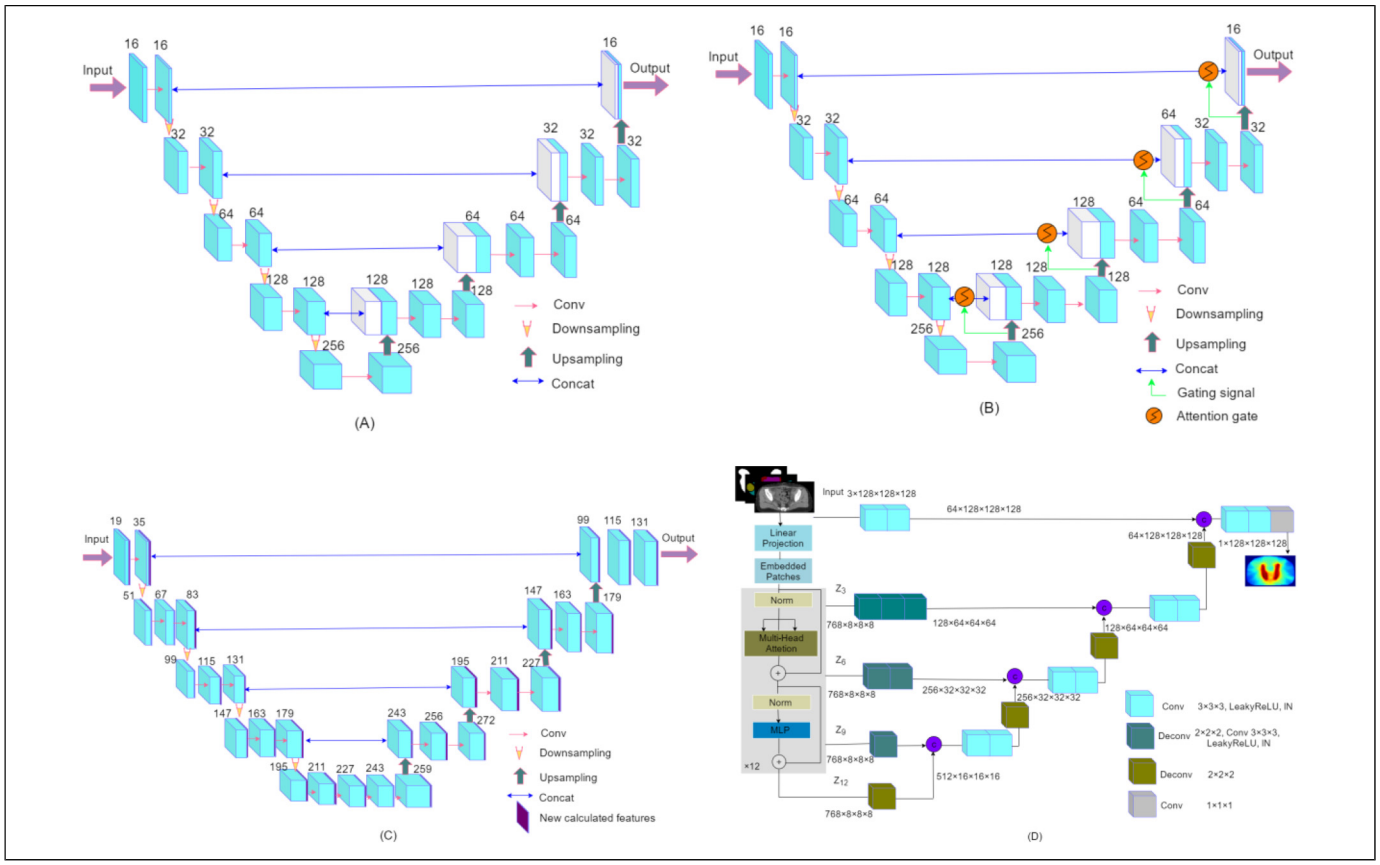

Figure 1. The 4 model architectures used in this study are (A) 3D U-Net, (B) 3DAtten U-Net, (C) HD U-Net, and (D) UNETR.
Abbreviations: 3D, three-dimensional; UNETR, U-Net transformers; HD, hierarchically densely.

As shown in Figure 1(A), the 3D U-Net is made up of 5 multiscale-resolution hierarchical levels. The encoding part typically contains a convolutional block at each hierarchy level followed by a downsampling layer ($3 \times 3 \times 3$ kernel size; voxel stride = 2, and padding =1). The convolutional block is composed of one 3D convolutional layer ($3 \times 3 \times 3$ kernel size; voxel stride = 1, and padding = 1), and each convolutional operation is followed by an instance normalization (IN) and a rectified linear unit (ReLU) layer. In the decoding part, on the right side, each step consists of an upsampling of the feature map (transposed convolution), and two 3D convolutional layers, each convolutional operation is followed by an IN and a ReLU. The encoder output convolutional block is concatenated with the input in the decoder. The upsampling layer was operated in a reverse way.

The 3DAtten U-Net was designed based on the 3D U-Net mentioned above. The output at each hierarchy level in the encoder is concatenated to the corresponding one in the decoder through attention-gated connections. The attention-gating mechanism was utilized to enhance network concern to important regions in images, which is shown in Figure 1(B).

The HD U-Net utilizes 3 operations, dense convolve, dense encoder, and U-Net decoder. For each dense operation, a growth rate can be defined as the number of new features calculated during the convolution step. Specifically, we utilized a growth rate of 16 (16 new features added after each "dense" operation), 4 dense decoder operations, and 64 features returned during the encoder operation. The architecture is described in Figure 1(C).

The architecture of the UNETR is illustrated in Figure 1(D). The model utilizes a transformer as the encoder to learn sequence representations of the input and effectively capture the global multiscale information. The network design resembles U-Net for the encoder and decoder. The transformer encoder is directly connected to a decoder via skip connections at different resolutions to compute the final dose output.

In all these architectures, CT images, contoured PTV, and OAR masks were first fed into the DL models for training and verification. At last, a single-channel dose distribution tensor with a size of $1 \times 128 \times 128 \times 128$ dose file was produced.

Model 5-Fold Cross-Training-Validation

For all patients, 209 patients were trained with 5-fold cross-validation, and the remaining 52 patients were randomly used as testing sets. The loss function used in the optimization was a mean absolute error (MAE). Adam optimizer and cosine annealing scheduler were used (learning rate: 3×10^{-4} and weight decay: 10^{-4}). The stopping criterion was 100 epochs. The batch size was set as 2. The networks were run on a

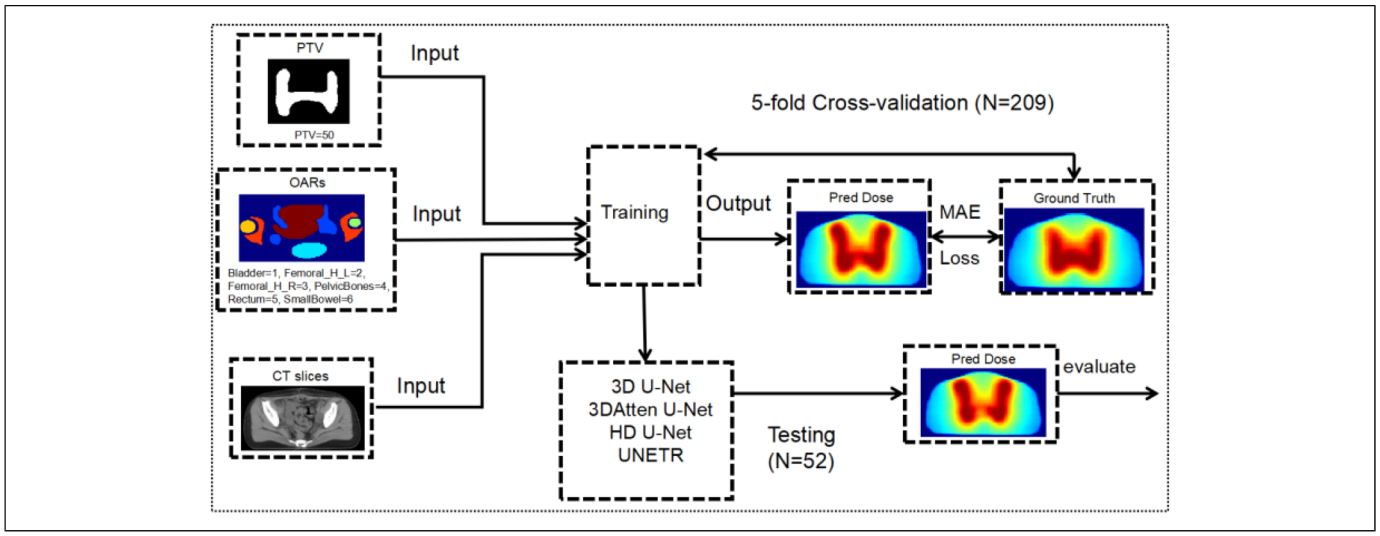

Figure 2. The flowchart of the model training, validation, and testing.

Table 1. The MAE of Different Structures in 5-Fold Cross-Validation (Mean ± Std).

Structures	3D U-Net	3DAtten U-Net	HD U-Net	UNETR
PTV (%)	0.77 ± 0.54	0.67 ± 0.53	**0.62 ± 0.46**	0.83 ± 0.67
Bladder (%)	4.28 ± 3.28	**4.13 ± 3.16**	4.29 ± 3.09	4.92 ± 3.47
Rectum (%)	4.25 ± 3.44	**4.24 ± 3.65**	4.41 ± 3.64	4.69 ± 3.74
Femoralhead_L (%)	6.95 ± 6.55	6.27 ± 6.33	**6.15 ± 6.00**	6.28 ± 5.33
Femoralhead_R (%)	6.72 ± 6.24	6.25 ± 5.80	6.17 ± 5.57	**6.09 ± 5.16**
Pelvic bones (%)	4.55 ± 4.51	**3.81 ± 3.05**	4.24 ± 3.18	4.33 ± 3.09
Small bowel (%)	**2.28 ± 1.76**	2.45 ± 2.46	2.58 ± 1.98	2.55 ± 1.99
Body (%)	**0.94 ± 0.85**	1.05 ± 1.07	1.17 ± 0.86	1.19 ± 0.86

Abbreviations: 3D, three-dimensional; UNETR, U-Net transformers; HD, hierarchically densely.

workstation with an NVIDIA TITAN graphics card and 32 GB RAM. Once the model was well trained, the time for a dose-predicted map was produced in 1 s. The flowchart of the model's training, validation, and testing is shown in Figure 2.

Quantitative Dose Prediction Evaluation

The performance of the 4 proposed models was evaluated based on dosimetric parameters, 3D dose distributions, and DVH parameters of OARs and PTV between the prediction and clinical truth (ground truth, GT). The predicted 3D dose distribution was evaluated by calculating the voxel-level MAE, which is defined as follows:

$$MAE = \frac{1}{N}\sum_{i=0}^{N}\frac{|D_{pred}(i) - D_{GT}(i)|}{D_p} \times 100 \quad (1)$$

where N is the total number of voxels, $D_{pred}(i)$ is the predicted dose value of the ith voxel, and $D_{GT}(i)$ is the GT of the ith voxel. D_p is the prescription dose value. Dosimetric parameters were evaluated for PTV and OARs according to the clinical interest as follows: PTV: D2, D98, Dmean, and homogeneity index (HI).[11] HI was defined in (2). OARs: bladder or rectum V40, V50, Dmean, and D2cc. Left and right femoral head V30, V40, and V50. Pelvic bones V20, V30, V40, and Dmean. Small bowel V40, Dmean, and D2cc are used to evaluate the accuracy of the model. Vn means the percentage of the volume received radiation n Gy. The dosimetric parameter differences are defined in (3) and (4). All comparisons were done with paired t tests. Statistical significance was set as $P < .05$.

$$HI = \frac{D2 - D98}{D50} \quad (2)$$

$$|\delta D| = |D_{x,GT} - D_{x,pred}| \quad (3)$$

$$|\delta V| = |V_{n,GT} - V_{n,pred}| \quad (4)$$

The dice similarity coefficients (DSCs) between the 3D isodose volumes of the predicted and clinical dose distribution images were defined in (5)

$$DSC = \frac{2 * |V_{x,GT} \cap V_{x,pred}|}{|V_{x,GT}| + |V_{x,pred}|} \quad (5)$$

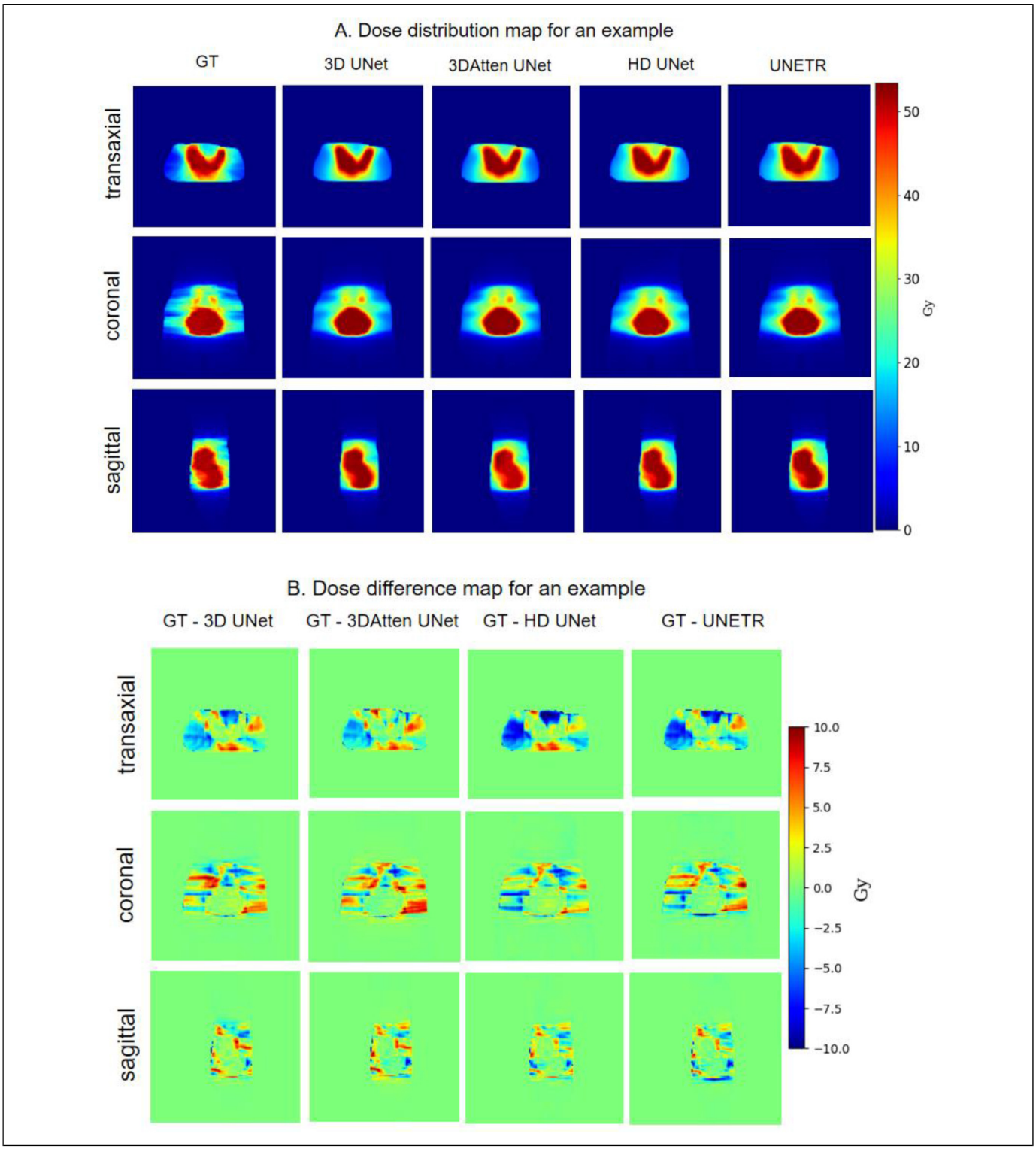

Figure 3. Two examples of the GT and predicted dose maps (transaxial, coronal, and sagittal) with 4 models in the test set: (A) dose distribution map for an example; (B) dose difference for an example; (C) dose distribution map for additional examples, and (D) dose difference for additional examples.

(continued)

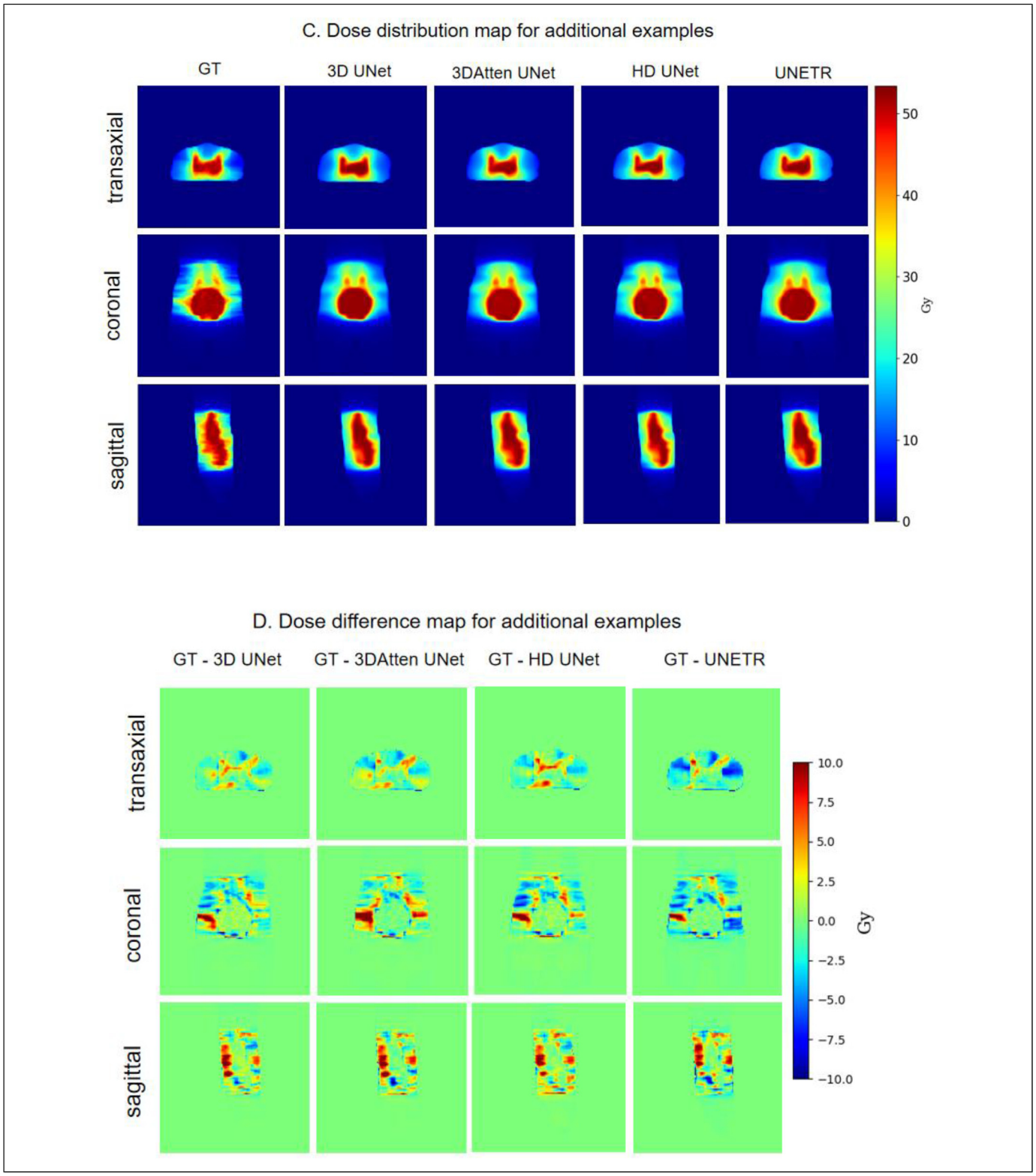

Figure 3. Continued.

Table 2. Dosimetric Parameters of GT and Predicted Dose for 52 Test Patients (Mean ± Std).

		Predicted dose			
PTV	GT	3D U-Net	3DAtten U-Net	HD U-Net	UNETR
D2 (Gy)	53.06 ± 0.51	51.83 ± 0.48 ($P < .001$)	51.65 ± 0.47 ($P < .001$)	52.24 ± 0.48 ($P < .001$)	53.00 ± 0.48 ($P = .055$)
D98 (Gy)	49.31 ± 0.67	48.41 ± 0.57 ($P < .001$)	49.31 ± 0.57 ($P = .246$)	49.63 ± 0.57 ($P = .013$)	49.77 ± 0.57 ($P < .001$)
Dmean (Gy)	51.47 ± 0.29	50.78 ± 0.38 ($P < .001$)	50.97 ± 0.38 ($P < .001$)	51.32 ± 0.38 ($P = .036$)	51.86 ± 0.38 ($P < .001$)
HI	0.07 ± 0.02	0.06 ± 0.01 ($P = .718$)	0.05 ± 0.01 ($P < .001$)	0.05 ± 0.01 ($P < .001$)	0.06 ± 0.01 ($P = .042$)
Bladder					
V40 (%)	51.07 ± 16.48	49.71 ± 17.41 ($P = .182$)	47.59 ± 17.60 ($P = .241$)	48.96 ± 17.15 ($P = .631$)	50.64 ± 18.55 ($P = .042$)
V50 (%)	26.92 ± 14.65	27.28 ± 15.42 ($P < .001$)	24.88 ± 16.14 ($P < .001$)	27.04 ± 16.20 ($P = .891$)	28.36 ± 17.50 ($P = .021$)
Dmean (Gy)	37.41 ± 4.88	38.07 ± 4.55 ($P = .150$)	38.14 ± 4.77 ($P = .666$)	37.82 ± 4.63 ($P = .375$)	38.76 ± 5.05 ($P = .006$)
D2cc (Gy)	52.09 ± 1.02	51.27 ± 0.53 ($P < .001$)	51.21 ± 0.35 ($P < .001$)	51.58 ± 0.37 ($P < .001$)	52.03 ± 0.50 ($P = .258$)
Rectum					
V40 (%)	62.19 ± 20.37	60.91 ± 21.52 ($P = .046$)	59.69 ± 21.23 ($P = .309$)	58.18 ± 21.46 ($P = .799$)	63.96 ± 21.28 ($P < .001$)
V50 (%)	24.83 ± 20.71	16.57 ± 17.39 ($P < .001$)	21.79 ± 19.13 ($P = .002$)	24.31 ± 19.71 ($P = .433$)	25.81 ± 20.38 ($P = .412$)
Dmean (Gy)	39.33 ± 5.97	40.22 ± 5.22 ($P = .0358$)	40.53 ± 5.06 ($P = .004$)	40.30 ± 5.23 ($P = .024$)	41.51 ± 5.23 ($P < .001$)
D2cc (Gy)	50.33 ± 2.32	50.01 ± 1.36 ($P < .001$)	50.27 ± 1.02 ($P < .001$)	50.51 ± 1.24 ($P < .001$)	51.24 ± 1.35 ($P = .325$)
FemoralH_L					
V30 (%)	12.09 ± 19.51	12.29 ± 8.66 ($P = .513$)	8.83 ± 6.94 ($P = .009$)	9.48 ± 7.432 ($P = .024$)	12.66 ± 9.20 ($P = .675$)
V40 (%)	2.52 ± 5.66	1.62 ± 2.12 ($P = .472$)	1.01 ± 1.47 ($P = .015$)	1.29 ± 1.85 ($P = .085$)	1.73 ± 2.19 ($P = .693$)
V50 (%)	0.00 ± 0.02	0.00 ± 0.00 ($P = .176$)	0.00 ± 0.00 ($P = .176$)	0.00 ± 0.00 ($P = .187$)	0.00 ± 0.00 ($P = .201$)
Dmean (Gy)	15.48 ± 6.61	18.90 ± 5.31 ($P = .266$)	17.75 ± 4.98 ($P = .173$)	17.86 ± 4.98 ($P = .244$)	19.08 ± 5.24 ($P = .149$)
FemoralH_R					
V30 (%)	10.36 ± 17.50	11.09 ± 8.37 ($P = .660$)	8.44 ± 7.05 ($P = .192$)	9.31 ± 8.00 ($P = .466$)	11.76 ± 9.36 ($P = .369$)
V40 (%)	2.48 ± 6.83	1.40 ± 2.12 ($P = .922$)	1.07 ± 1.65 ($P = .337$)	1.27 ± 2.14 ($P = .741$)	1.53 ± 2.33 ($P = .657$)
V50 (%)	0.03 ± 0.15	0.00 ± 0.00 ($P = .232$)	0.00 ± 0.00 ($P = .232$)	0.00 ± 0.00 ($P = .232$)	0.01 ± 0.08 ($P = .232$)
Dmean (Gy)	14.96 ± 6.60	14.99 ± 5.25 ($P = .067$)	15.74 ± 6.09 ($P = .993$)	14.97 ± 5.06 ($P = .979$)	16.50 ± 5.72 ($P = .021$)
Pelvic bones					
V20 (%)	75.11 ± 10.14	75.21 ± 7.47 ($P = .008$)	74.20 ± 7.15 ($P = .044$)	75.32 ± 7.41 ($P = .008$)	77.78 ± 8.28 ($P = .002$)
V30 (%)	48.95 ± 12.55	47.12 ± 8.52 ($P = .270$)	45.09 ± 8.57 ($P = .761$)	47.04 ± 8.28 ($P = .297$)	49.67 ± 9.46 ($P = .011$)
V40 (%)	25.94 ± 11.41	22.58 ± 6.65 ($P = .109$)	21.83 ± 6.93 ($P = .009$)	23.18 ± 6.66 ($P = .392$)	23.96 ± 7.11 ($P = .850$)
Dmean (Gy)	29.99 ± 4.01	29.29 ± 2.54 ($P = .239$)	28.98 ± 2.60 ($P = .685$)	29.42 ± 2.57 ($P = .141$)	30.05 ± 2.79 ($P = .005$)
Small bowel					
V40 (%)	14.25 ± 7.11	15.53 ± 7.79 ($P = .853$)	15.27 ± 7.71 ($P = .492$)	15.71 ± 7.84 ($P = .846$)	15.87 ± 7.83 ($P = .591$)
Dmean (Gy)	18.83 ± 5.21	19.28 ± 5.61 ($P = .511$)	19.35 ± 5.58 ($P = .351$)	19.61 ± 5.64 ($P = .081$)	19.76 ± 5.67 ($P = .015$)
D2cc (Gy)	52.07 ± 0.85	50.85 ± 1.38 ($P < .001$)	50.88 ± 1.28 ($P < .001$)	51.36 ± 1.29 ($P < .001$)	52.52 ± 1.69 ($P < .001$)

Abbreviations: 3D, three-dimensional; UNETR, U-Net transformers; HD, hyperdense; GT, ground truth; PTV, planning target volume.

Results

Mean Absolute Error

We first quantified the MAE results of different structures in 4 models used in our study. The predicted 3D dose distribution was evaluated by calculating the voxel-level MAE, which is the dose error, averaged across all voxels of a structure (PTV, OARs, or entire body contour), and normalized to the prescription dose. The results of MAE on the testing set are listed in Table 1. The mean MAE of PTV and body contour are within 1.19%. The femoral head shows a relatively larger MAE than other OARs. The table shown in bold indicates the lowest predicted error.

Comparison of Dose Map

Figure 3 shows the dose map of the predicted dose distribution and clinical one, and their 3D spatial dose difference (GT–prediction) maps between the clinical and predicted dose distributions. It can be visually seen that 4 models generate results that are similar to the clinical dose. The prediction accuracy of the 40 to 50 Gy region is of high overlap, which may be due to the good consistency of the target and dose distribution in this region for different clinical plans. It is a remarkable fact that the dose difference between the clinical and predicted dose in the 4 models is minor.

Dosimetric Parameters and DVHs

The dosimetric parameters in the test set are shown in Table 2. Dmean is defined as the mean dose, D98 is defined as the minimum dose, and D2 is defined as the maximum dose. The dosimetric analysis for OARs included Vn, D2cc (2 cm^3 receiving radiation n Gy). Details of other OAR results are listed in Table 2. The boxplot of dosimetric parameters differences $|\delta D|$ (|GT−prediction|) of all structures within the body for the 52 testing patients is shown in Figure 4. The average

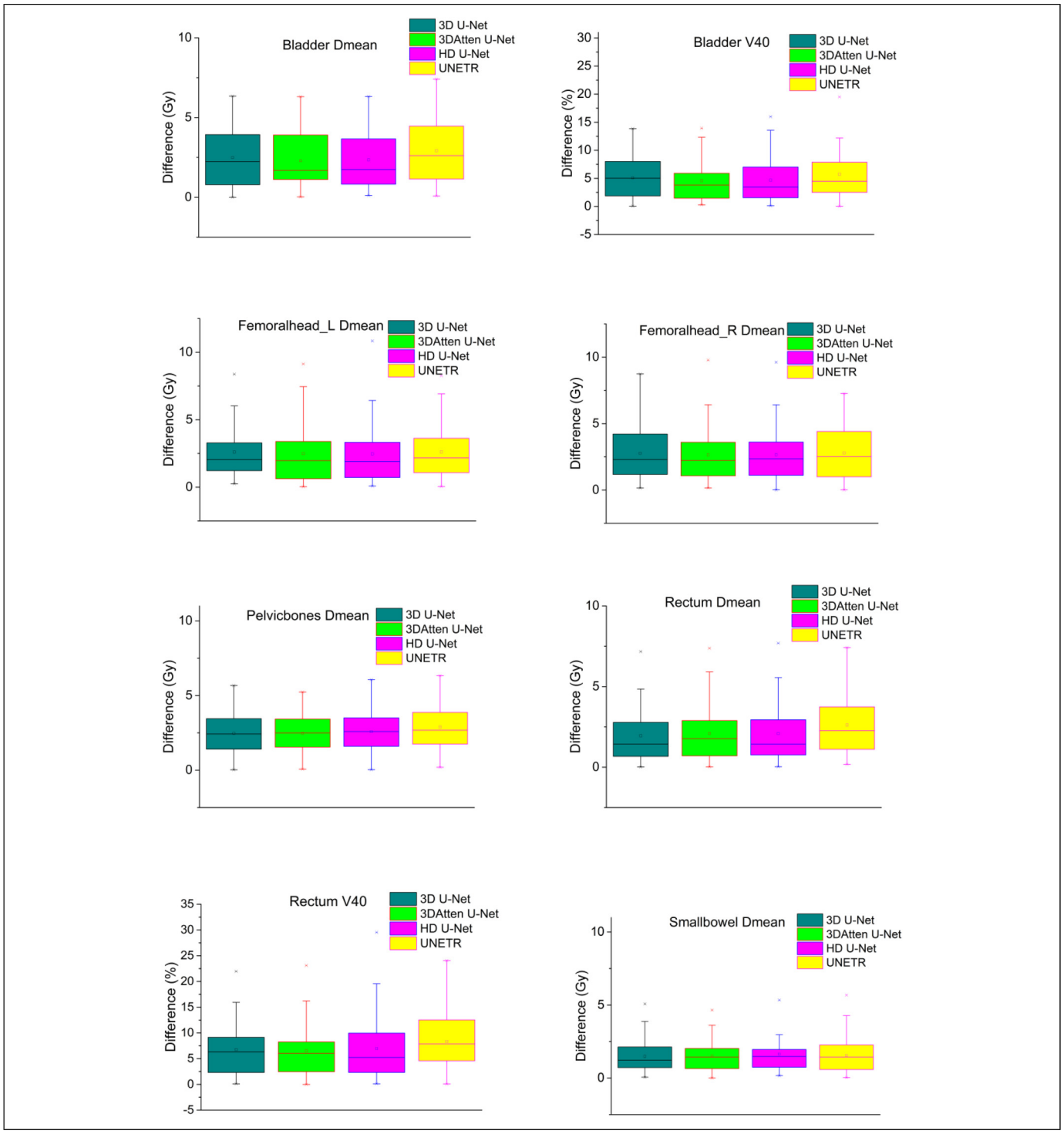

Figure 4. Dose difference boxplot for bladder, rectum, left and right femoral head, pelvic bone, and small bowel.

error of the Dmean difference for different OARs is within 2.5 Gy. The average error of V40 difference for the bladder and rectum is about 5%.

Figure 5 shows the DVH parameters of OARs and PTV between the prediction and GT in a test example. Four models predicted an accurate dose curve to GT. Four DL models were able to predict the dose curves in this example.

We also compared the DSC results of the 52 testing sets for different models, which are illustrated in Figure 6. The 4 models exhibit similar DSC values, and the average value of DSC is above 90%.

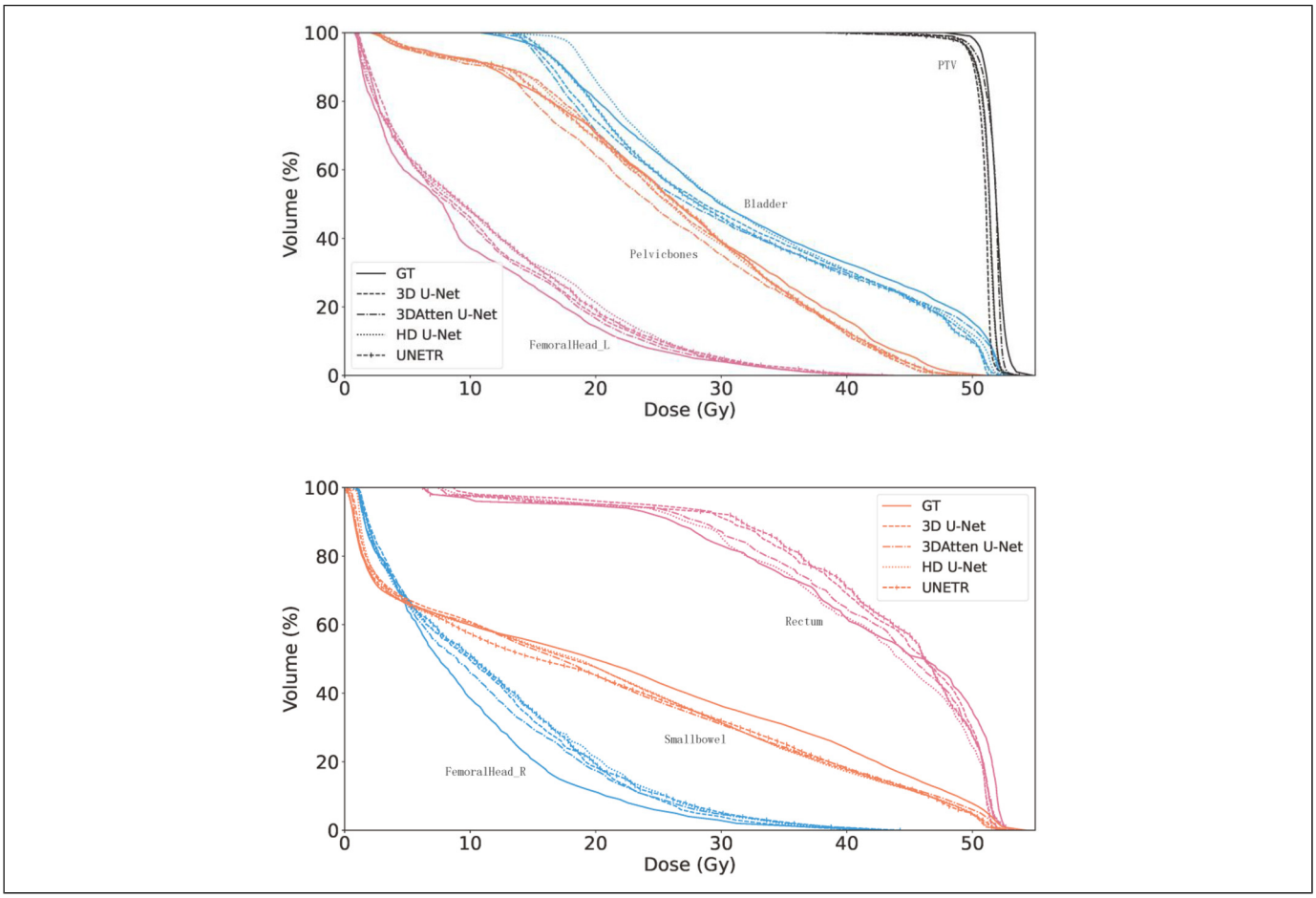

Figure 5. DVH comparison between GT and predicted dose distribution.
Abbreviations: DVH, dose–volume histogram; GT, ground truth.

Discussion

This study aimed to compare the performance of 4 DL dose prediction models for cervical cancer on the same test set using the same evaluation criteria. It is challenging to make meaningful comparisons of the state-of-the-art DL models since previous studies were carried out using different patient databases and limited evaluation criteria.

In this study, the inputs of the model were binary PTV, OARs mask, and CT images, which are consistent with the majority of studies.[12–14] The output was a spatial dose objective specifying 50 Gy per voxel, which eliminates the need to tune the complexity of inverse optimization. Although there were statistically significant differences in some dosimetric parameters, the predicted results are clinically acceptable. In our studies, the dosimetrists spend 40 min re-optimizing the dose distributions, but DL models take about one second to predict it. The dose distribution information-aid replan reduced the total planning time. The results were similar to Song et al's study.[15] Song et al predicted dose distribution for rectal cancer based on DeepLabv3+ and demonstrated that the DL model produced a clinically acceptable dose distribution with time savings. We expected no statistical differences in all predicted metrics. However, compared with clinical ground truth, not all dosimetric parameters were statistically different. This is consistent with the findings of many reports.[16–20]

The predicted dose map showed almost identical performance of the 4 models as they are almost similar to the clinical dose. The error in the PTV region is lower than that in the OARs, as seen in the dose difference map (Figure 3). In the OAR regions, due to the various dose constraint methods and trade-off balance principles by different dosimetrists, the dose distribution in the clinical plan was less consistent, which resulted in relatively lower OAR prediction accuracy. This phenomenon is common in most other reports[6,16–19] that include dose prediction for cervical cancer.[16–18] Due to the good consistency of the target and the relatively small span of dose values in this region. The various shapes and locations of OARs for different patients resulted in poor accuracy. Furthermore, the MAE results of PTV also confirmed the conclusion. The maximum MAE of PTV is only $0.83\% \pm 0.67\%$, but the maximum MAE among the OARs is the femoral head, which reached $6.95\% \pm 6.55\%$. More recently, the attention U-Net model based on the self-attention mechanism was widely used in semantic segmentation,[21–23] as well as radiation dose prediction[6] and achieved high accuracy. Moreover, transformers, which have the advantage of a self-attention mechanism

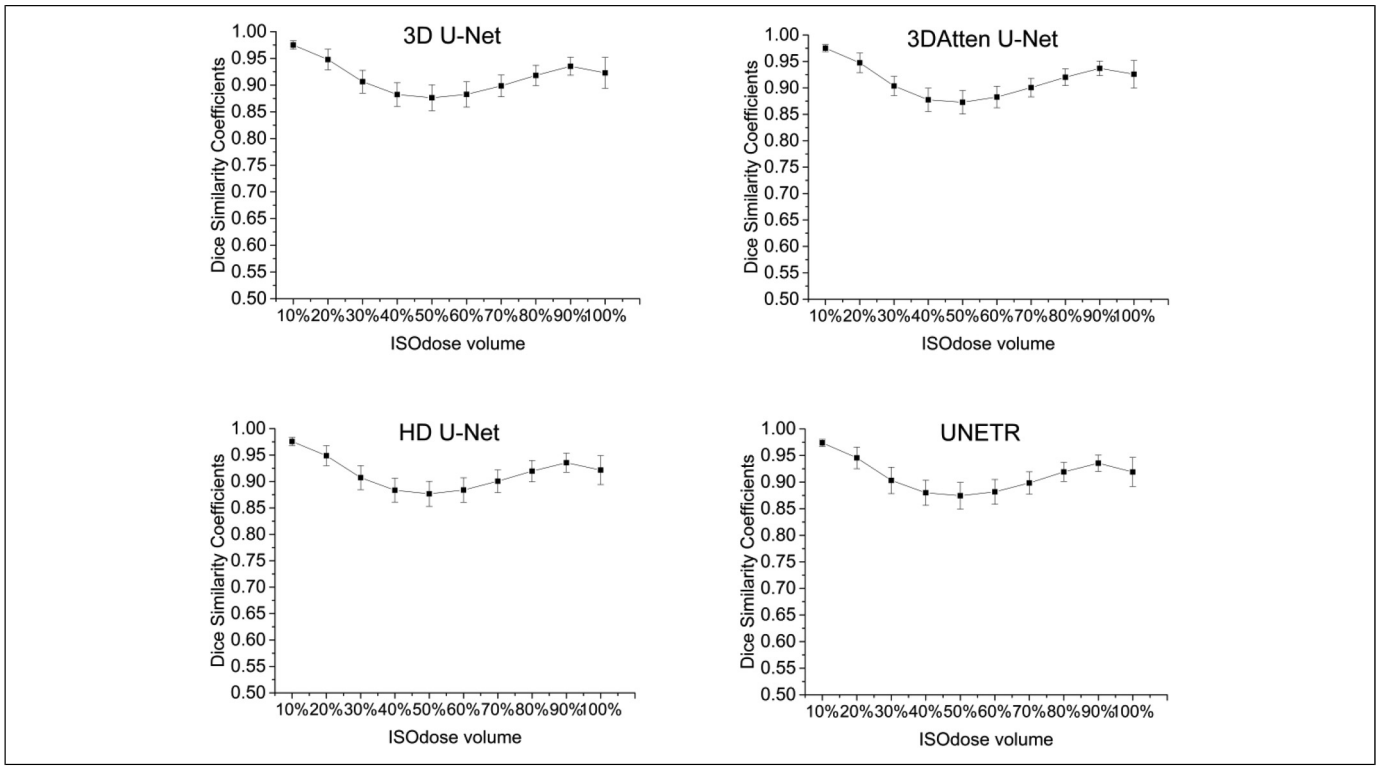

Figure 6. Dice similarity coefficients between ground truth (GT) and predicted isodose volumes for 52 test patients.

and the capability of learning long-range dependencies,[24] have made great progress in natural language processing. Transformers-based dose prediction models have been proposed in recent years.[12,25,26] As a kind of typical transformers-based DL model, UNETR has never been used in dose prediction.[27] Research from Osman et al[28] showed that U-Net, attention U-Net, residual U-Net, and attention Res U-Net for H&N plans from OpenKBP-Grand Challenge data[7] had an almost comparable performance for voxel-wise dose prediction. Inspired by Osman et al's research,[28] we compared the 3D U-Net and its typical variants. Our results show that the 4 DL models have high capability in accurately predicting 3D dose distributions for cervical cancer VMAT plans.

The difference map between the predicted and clinical dose from a recent study on cervical cancer shows that the error range is from -10 to 10 Gy,[18] which is consistent with our results (Figure 3B to D). In Yu et al's[17] results, the maximum dose error of PTV was less than $1.64\% \pm 1.51\%$ and $1.27\% \pm 1.67\%$ for 3D U-Net and 3DRes U-Net, respectively, whereas our results showed the maximum mean error is 0.83%. Zhang et al[29] proposed a densely connected network to predict esophageal radiotherapy dose distribution. The MAE within the PTV, OARs, and body is 2.1%, 4.2% to 7%, and 3.4%, respectively. In Nguyen et al's[8] results for H&N dose prediction, the mean error for OARs is 5% to 10%. Although their esophageal or H&N results are not comparable to our results for cervical cancer, the values in our results still demonstrated the validity of the state-of-the-art model we studied. The mean absolute error results of the body for different groups in 7-fold cross-validation were from 0.9% to 2.3% in previous results of cervical cancer.[18] Our results are similar to the prediction accuracy of the above studies. The average error of the Dmean difference for different OARs is within 2.5 Gy. The average error of V40 difference for bladder and rectum is about 5%, which is similar to or better than previous studies.[18] Considering all voxels within the body, 3D U-Net showed the best performance or is at least one of the best. More complex models seem to have not yielded better results. This is mainly because the complex model is more likely to exhibit overfitting, which hinders the performance of DL. Furthermore, the average DSC of our results was above 90%, which indicates that the predicted values were close to the clinical actual values and the results (Figure 6) were similar to previous studies.[16,18]

There are several limitations in our study. The number of enrolled patients was not large enough, and it is single-center research, which lacks external validation. There is still room to further improve our dose prediction accuracy. The depth of the models cannot be increased due to the limited GPU capacity.

Conclusions

Four DL models that use 3D U-Net and its state-of-the-art variants were studied in this retrospective study on our institute's cervical cancer VMAT plans using various assessment metrics. The DL models proposed in this research all exhibited

robustness and feasibility in the 3D VMAT dose prediction for cervical cancer. Considering all voxels within the body, 3D U-Net showed the best performance. Intelligent dose map distribution prediction of VMAT is of paramount importance for further clinical applications for cervical cancer plans.

Acknowledgments

The authors thank Dr Shengxian Peng for his illuminating discussion.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the General program of the National Natural Science Foundation of China (31971113); Chongqing Science and Technology Talent Project (CQYC201905037); and Chongqing key research and development program (CSTB2022TIAD-KPX0181).

Ethical Approval

The study protocol was approved by the Ethics Committee of Zigong First People's Hospital (No. Ethics [M] 2024-015). The requirement for patient consent was waived due to the retrospective nature of the study.

ORCID iD

Zhe Wu PhD https://orcid.org/0000-0001-5191-0248

References

1. Sung H, Ferlay J, Siegel RL, et al.. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209-249.
2. Gopalani SV, Janitz AE, Campbell JE. Trends in cervical cancer incidence and mortality in Oklahoma and the United States, 1999–2013. *Cancer Epidemiol*. 2018;56:140-145.
3. Rahimy E, von Eyben R, Lewis J, et al. Evaluating dosimetric parameters predictive of hematologic toxicity in cervical cancer patients undergoing definitive pelvic chemoradiotherapy. *Strahlenther Onkol*. 2022;198(9):773-782.
4. Strauß S. From big data to deep learning: a leap towards strong AI or 'intelligentia obscura'? *Big Data Cogn Comput*. 2018;2(3):16.
5. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, 2015, Vol. 9351, pp. 234–241.
6. Osman AFI, Tamam NM. Attention-aware 3D U-Net convolutional neural network for knowledge-based planning 3D dose distribution prediction of head-and-neck cancer. *J Appl Clin Med Phys*. 2022;23(7):e13630.
7. Babier A, Zhang B, Mahmood R, et al. OpenKBP: the open-access knowledge-based planning grand challenge and dataset. *Med Phys*. 2021;48(9):5549-5561.
8. Nguyen D, Jia X, Sher D, et al. 3D Radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. *Phys Med Biol*. 2019;64(6):065020.
9. Li Z, Zhou L, Tan S, Tang A. Application of UNETR for automatic cochlear segmentation in temporal bone CTs. *Auris Nasus Larynx*. 2023;50(2):212-217.
10. Gillot M, Baquero B, Le C, et al. Automatic multi-anatomical skull structure segmentation of cone-beam computed tomography scans using 3D UNETR. *PLoS One*. 2022;17(10):e0275033.
11. Liu Z, Fan J, Li M, et al. A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy. *Med Phys*. 2019;46(5):1972-1983.
12. Xiao F, Cai J, Zhou X, et al. TransDose: a transformer-based UNet model for fast and accurate dose calculation for MR-LINACs. *Phys Med Biol*. 2022;67(12):125013.
13. Hedden N, Xu H. Radiation therapy dose prediction for left-sided breast cancers using two-dimensional and three-dimensional deep learning models. *Phys Med*. 2021;83:101-107.
14. Gronberg MP, Gay SS, Netherton TJ, et al. Technical note: dose prediction for head and neck radiotherapy using a three-dimensional dense dilated U-net architecture. *Med Phys*. 2021;48(9):5567-5573.
15. Song Y, Hu J, Liu Y, et al. Dose prediction using a deep neural network for accelerated planning of rectal cancer radiotherapy. *Radiother Oncol*. 2020;149:111-116.
16. Zhang QL, Bao P, Qu A, et al. The feasibility study on the generalization of deep learning dose prediction model for volumetric modulated arc therapy of cervical cancer. *J Appl Clin Med Phys*. 2022;23(6):e13583. doi:10.1002/acm2.13583
17. Yu W, Xiao C, Xu J, et al. Direct dose prediction with deep learning for postoperative cervical cancer underwent volumetric modulated arc therapy. *Technol Cancer Res Treat*. 2023;22:15330338231167039. doi:10.1177/15330338231167039
18. Zhang G, Jiang Z, Zhu J, et al. Dose prediction for cervical cancer VMAT patients with a full-scale 3D-cGAN-based model and the comparison of different input data on the prediction results. *Radiat Oncol*. 2022;17(1):179.
19. Ahn SH, Kim E, Kim C, et al. Deep learning method for prediction of patient-specific dose distribution in breast cancer. *Radiat Oncol*. 2021;16(1):154. doi:10.1186/s13014-021-01864-9
20. Ma M, Buyyounouski M K, Vasudevan V, et al. Dose distribution prediction in isodose feature-preserving voxelization domain using deep convolutional neural network. *Med Phys*. 2019;46(7):2978-2987.
21. Rajamani KT, Rani P, Siebert H, ElagiriRamalingam R, Heinrich MP. Attention-augmented U-Net (AA-U-Net) for semantic segmentation. *Signal Image Video Process*. 2023;17(4):981-989.
22. Zhang H, Lian Q, Zhao J, Wang Y, Yang Y, Feng S. RatUNet: residual U-Net based on attention mechanism for image denoising. *PeerJ Comput Sci*. 2022;10(8):e970. doi:10.7717/peerj-cs.970
23. Lin H, Li Z, Yang Z, Wang Y. Variance-aware attention U-Net for multi-organ segmentation. *Med Phys*. 2021;48(12):7864-7876.
24. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Proceedings of the 31st international conference on neural information processing systems*. NIPS'17, Red Hook, NY, USA: Curran Associates Inc. 2017, pp. 6000-6010.

25. Pastor-Serrano O, Dong P, Huang C, et al. Sub-second photon dose prediction via transformer neural networks. *Med Phys*. 2023;50(5):3159-3171.
26. Yang J, Zhao Y, Zhang F, et al. Deep learning architecture with transformer and semantic field alignment for voxel-level dose prediction on brain tumors. *Med Phys*. 2023;50(2):1149-1161.
27. Hatamizadeh A, Yang D, Roth H, et al. UNETR: Transformers for 3D Medical Image Segmentation. In: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2021, pp. 1748–1758..
28. Osman AFI, Tamam NM, Yousif YAM. A comparative study of deep learning-based knowledge-based planning methods for 3D dose distribution prediction of head and neck. *J Appl Clin Med Phys*. 2023;3(9):e14015. doi:10.1002/acm2.14015
29. Zhang J, Liu S, Yan H, et al. Predicting voxel-level dose distributions for esophageal radiotherapy using densely connected network with dilated convolutions. *Phys Med Biol*. 2020;65(20):205013.