Radiation Oncology

RESEARCH

Open Access

Deep learning-powered radiotherapy dose prediction: clinical insights from 622 patients across multiple sites tumor at a single institution

Zhen Hou[1†], Lang Qin[2,3†], Jiabing Gu[4†], Zidong Liu[5], Juan Liu[1], Yuan Zhang[1], Shanbao Gao[1*], Jian Zhu[6,7,8*] and Shuangshuang Li[1*]

Abstract

Purpose Accurate pre-treatment dose prediction is essential for efficient radiotherapy planning. Although deep learning models have advanced automated dose distribution, comprehensive multi-tumor analyses remain scarce. This study assesses deep learning models for dose prediction across diverse tumor types, combining objective and subjective evaluation methods.

Methods and materials We included 622 patients with planning data across various tumor sites: nasopharyngeal carcinoma ($n=29$), esophageal carcinoma ($n=82$), left-sided breast carcinoma ($n=107$), right-sided breast carcinoma ($n=95$), cervical carcinoma treated with radical radiotherapy ($n=84$), postoperative cervical carcinoma ($n=122$), and rectal carcinoma ($n=103$). Dose predictions were generated using U-Net, Flex-Net, and Highres-Net models, with data split into training (60%), validation (20%), and testing (20%) sets. Quantitative comparisons used normalized dose difference (NDD) and dose-volume histogram (DVH) metrics, and qualitative assessments by radiation oncologists were performed on the testing set.

Results Predicted and clinical doses correlated well, with NDD values under 3% for tumor targets in nasopharyngeal, breast, and postoperative cervical cancer. Qualitative assessments revealed that U-Net, Flex-Net, and Highres-Net achieved the highest accuracy in cervical radical, breast/rectal/postoperative cervical, and nasopharyngeal/esophageal cancers, respectively. Among the test cases ($n=123$), 53.7% were deemed clinically acceptable and 32.5%

[†]Zhen Hou, Lang Qin and Jiabing Gu contributed equally to this work.

*Correspondence:
Shanbao Gao
gbsource@sina.com
Jian Zhu
zhujian@sdfmu.edu.cn
Shuangshuang Li
lishuangshuang@njglyy.com

Full list of author information is available at the end of the article

required minor adjustments. The "Best Selection" approach, combining strengths of all three models, raised clinical acceptance to 62.6%.

Conclusion This study demonstrates that automated dose prediction can provide a robust starting point for rapid plan generation. Leveraging model-specific strengths through the "Best Selection" approach enhances prediction accuracy and shows potential to improve clinical efficiency across multiple tumor types.

Keywords Deep learning, Dose prediction, Radiotherapy, Multi-tumor

Background

Radiotherapy (RT) is a critical treatment modality for tumors [1]. Among available techniques, Intensity-Modulated Radiation Therapy (IMRT) has become the preferred method for targeting solid tumors with precision, minimizing exposure to vital organs and surrounding tissues. Volumetric Modulated Arc Therapy (VMAT) [2], an advanced form of IMRT, employs rotating arcs for superior modulation capabilities, allowing for greater conformity to tumor shape and location. This precision enhances dose delivery while better sparing healthy tissues. However, planning VMAT treatments is time-intensive, often requiring iterative adjustments by experienced dosimetrists.

Key time-consuming steps in the planning process include contouring plan-assisted structures and adjusting parameters [3], both of which demand significant time and expertise to optimize. In this context, having predictive information about dose distribution in advance could streamline the process by allowing optimization based on limited plan-assisted contours, potentially saving time. Additionally, predictive dose distribution could offer valuable visual input and dose-volume metrics, supporting physicians' decision-making before initiating the optimization process [4].

Recent advances in deep learning, especially convolutional neural networks (CNNs), have shown promising potential in radiation dose prediction [2–22]. These methods exceed the accuracy and efficiency of conventional knowledge-based planning (KBP) techniques [23, 24]. Studies have highlighted the predictive capabilities of deep learning models across diverse cancer types, including head and neck [7–12], esophageal [5, 6], breast [13, 14], lung [15], pancreatic [16], cervical [17–21], prostate [2–4, 15, 16], and rectal cancer [22]. Nevertheless, despite these advances, existing literature lacks a comprehensive evaluation of deep learning models' effectiveness across multiple tumor types.

This study addresses this gap by assessing the effectiveness of deep learning models in predicting radiation doses across multiple tumor sites. Specifically, we conduct a detailed quantitative and qualitative evaluation of three distinct model architectures.

Materials and methods

Patient characteristics and imaging details

The study was approved by the Ethics Committee of Nanjing Drum Tower Hospital, and informed consent was waived due to the retrospective nature of the study, which posed no additional health risks. Planning data from patients with nasopharyngeal, esophageal, breast, cervical, and rectal carcinomas who received radiotherapy at Nanjing Drum Tower Hospital were evaluated in reverse chronological order, starting from October 8, 2021. The inclusion criteria were as follows: (a) patients with a single radiation target treated at our institution; (b) no metastases; (c) availability of planning CT scans, RT structure, and RT dose in DICOM format; and (d) all plans and doses reviewed by expert dosimetrists and subsequently approved by oncologists with over eight years of experience.

A total of 622 patients met the inclusion criteria were identified across seven datasets, including nasopharyngeal (NPC, $n = 29$), esophageal (EC, $n = 82$), left-sided breast cancer cases ($n = 107$) and right-sided cases ($n = 95$), cervical radical carcinoma ($n = 84$), postoperative cervical carcinoma ($n = 122$), and rectal carcinoma ($n = 103$).

The planning CT images were acquired using a Siemens SOMATOM Definition CT scanner (Siemens, Erlangen, Germany) following a standardized clinical protocol: 120 kVp, 120–300 mAs, pixel spacing of 0.8×0.8, image resolution of 512×512 pixels, and slice thickness of 5 mm.

Planning details

The planning target volume (PTV) and organs at risk (OARs) were meticulously contoured and thoroughly reviewed by experienced radiation oncologists. Volumetric Modulated Arc Therapy (VMAT) plans were designed using the Monaco planning system (Elekta, Stockholm, Sweden) using a 6MV 360-degree arc. All plans were subsequently clinically approved by the responsible senior oncologists.

Prescription doses and fractions were administered based on the specific tumor site. For patients with nasopharyngeal carcinoma (NPC), the prescribed doses were 70 Gy/64Gy/58Gy for PGTV/PTV-LN (PTV1)/PTV (PTV2), respectively, delivered over 33 fractions. Patients with esophageal cancer (EC) received 50.4 Gy over 28

fractions to the PTV. Breast carcinoma patients were treated with 50 Gy over 25 fractions to the PTV. Cervical carcinoma patients received a prescription dose of 50.4 Gy over 28 fractions to the PTV, while rectal carcinoma patients received 50.4 Gy over 28 fractions to their respective PTVs.

Deep learning models for dose prediction

As depicted in Fig. 1, three CNN-based deep learning models namely U-Net, Flex-Net, and Highres-Net, were assessed for each dataset. These models are commercially available on the Acculearning AI Model Training Platform, provided by Manteia Medical Technologies. The platform enables no-coding workflows with features such as automatic data cleansing, adjustable multi-layer CNN architectures, and model performance evaluation, allowing users to customize training parameters for dose prediction.

Training and validation

The dataset included multiple tumor types from a total of 622 patients. Data were divided into training, validation, and testing sets at a 6:2:2 ratio. The models were trained using the mean absolute error (MAE) loss function and the Adam optimizer, which was chosen for its ability to adapt the learning rate during training. An initial learning rate of 0.001 was used, with exponential decay incorporated into the optimizer. To improve model generalization without significantly affecting performance, a higher dropout rate of 0.9 was selected. This dropout rate effectively reduced overfitting on our smaller dataset, contributing to better model robustness. To enhance model robustness, data augmentation techniques were applied, including zoom ratios from 0.85 to 1.2, rotation angles of ± 15°, and gamma variations from 0.5 to 2.0. The training was conducted over 200 epochs.

Quantitative evaluation of dose distribution by dose metrics

Model performance was evaluated by comparing predicted and actual dose distributions using dose-volume-histogram (DVH)-based metrics for OARs, as well as D_{98} and D_2 (the minimum doses covering 98% and 2% of the PTV, respectively). Additionally, the normalized dose difference (NDD) for the PTV was calculated, defined as the difference between predicted and actual doses normalized by the prescribed dose. This evaluation was conducted separately for each dataset, with emphasis on the testing set.

Qualitative evaluation of dose distribution by experienced radiation oncologists

For qualitative evaluation, predicted results from the testing sets were visually assessed by a committee of senior radiation oncologists specializing in different tumor types. All committee members have extensive experience in radiotherapy, exceeding eight years. In this study, we performed qualitative assessments of each model's dose prediction results and used the "Best Selection" method to identify the top predictions among U-Net, Flex-Net, and Highres-Net based on qualitative scores. The "Best Selection" method does not involve creating a new model from scratch or combining multiple models. Instead, it is a process applied on each test data to identify the best possible prediction among the three models.

The oncologists graded each predicted dose distribution on a 4-point scale: grade A-acceptance (clinical acceptance), grade B-minor revision (slight optimization required), grade C-major revision (significant optimization required), and grade D-rejection (requiring re-planning).

Statistical analysis

A repeated measures analysis of variance (ANOVA) was used to compare model differences. When the assumptions of sphericity or normality for ANOVA were violated, the Friedman test, a non-parametric alternative, was employed. Paired t-tests were used to compare the predicted and actual outcomes for each model, and when the assumptions of normality or homoscedasticity were violated, the Wilcoxon signed-rank test was applied as a non-parametric alternative.

Results

Dose-volume histogram (DVH) analysis for predicted PTV doses across different tumor sites

Table 1 presents the predicted PTV results across various tumor types, highlighting performance differences among the models. For nasopharyngeal carcinoma, there was no significant difference in the D_{98} predictions for $PGTV_{69.96\ Gy}$, $PTV1_{60.06\ Gy}$, and $PTV2_{53.46\ Gy}$ across the three models ($P > 0.05$). Highres-Net provided the most accurate predictions to the approved dose for the D_2 parameters of all PTVs, indicating its superior accuracy for this tumor type. In esophageal carcinoma, U-Net performed best in predicting $PTV_{45Gy}\_D_2$, with minimal deviation from the approved dose (48.97 ± 1.80 Gy vs. 48.22 ± 0.47 Gy, NDD = 2.3 ± 9.87%). For left-sided breast cancer, Flex-Net had the highest accuracy (NDD = -0.29 ± 4.08%), while for right-sided breast cancer, Highres-Net's prediction for $PTV_{50Gy}\_D_2$ (52.42 ± 0.10 Gy) was closest to the approved dose (53.84 ± 0.26 Gy, NDD = 1.62 ± 4.39%). In cervical carcinoma, U-Net consistently produced predictions closely matching the approved dose for both radical and post-operative cases ($PTV_{50Gy}\_D_{98}$ and $\_D_2$). For instance, $PTV_{50Gy}\_D_{98}$ predictions differed by only 0.31 Gy (49.67 Gy vs. 49.98 Gy). In rectal carcinoma, despite no

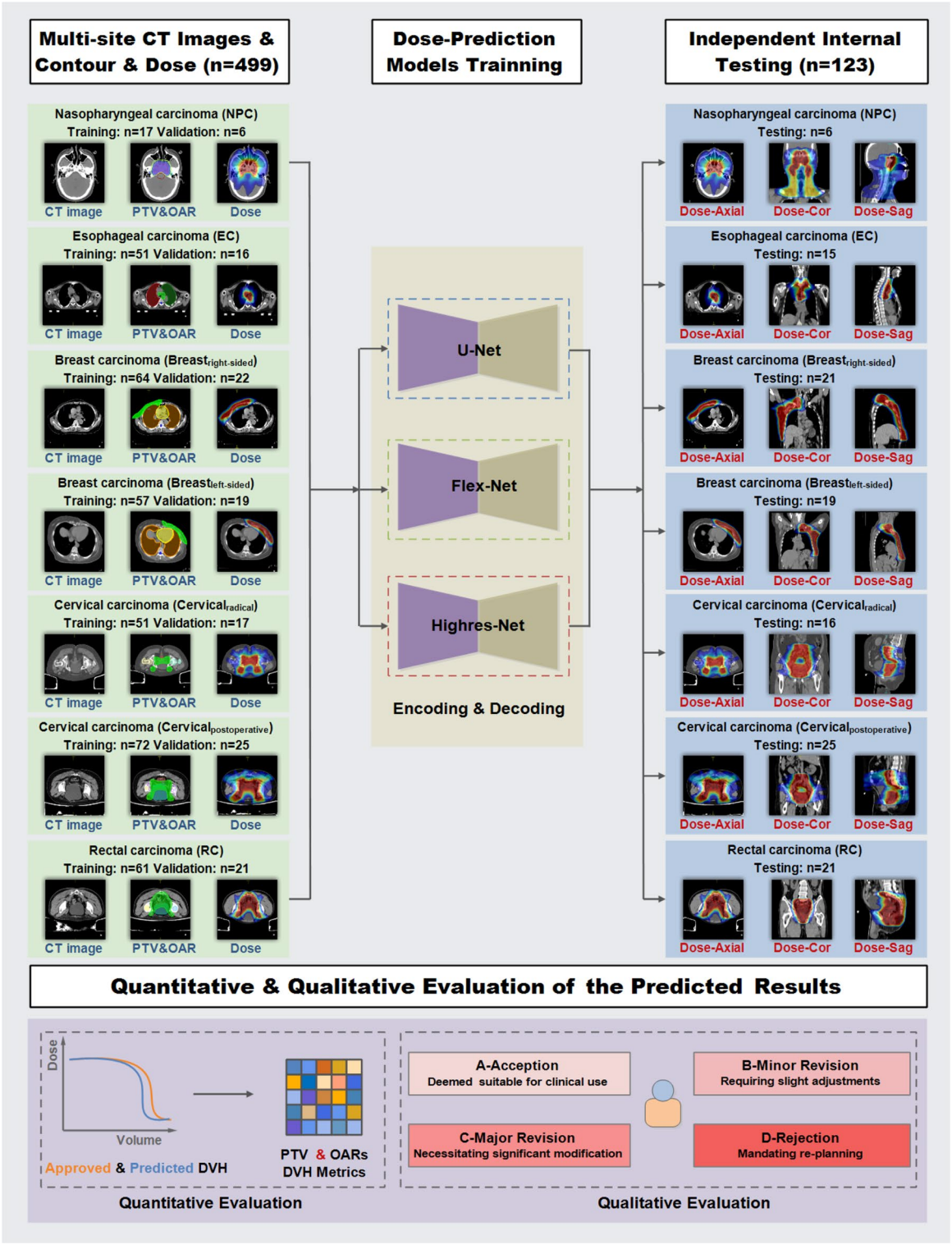

Fig. 1 Workflow for automated dose prediction across multiple tumor sites. (**A**) Input multi-site CT images with RT contours and dose data; (**B**) Deep learning model pipeline; (**C**) Axial, sagittal, and coronal views of the predicted dose distribution; (**D**) Quantitative and qualitative evaluation of prediction accuracy

Table 1 Predicted dose-volume histogram (DVH) parameters of planning target volumes (PTVs) across multiple tumor sites for each mode

Metrics	Dose predicted by DL models				*P*-value
	$Dose_{Approved}$	$Dose_{U-Net}$	$Dose_{Flex-Net}$	$Dose_{Highres-Net}$	
Nasopharyngeal carcinoma					
$PGTV_{69.96Gy}\_D_{98}$ (Gy)	67.28 ± 2.21	66.76 ± 2.86	66.86 ± 2.13	67.28 ± 2.21	0.568
$PGTV_{69.96Gy}\_D_{2}$ (Gy)	75.22 ± 0.48	76.22 ± 0.70	74.83 ± 0.50	74.84 ± 0.43	0.012†
NDD_{PGTV} (%)	N/A	-2.01 ± 3.72	-0.45 ± 3.51	0.01 ± 4.05	0.503
$PTV1_{60.06Gy}\_D_{98}$ (Gy)	61.77 ± 0.66	61.70 ± 1.06	61.01 ± 1.06	61.23 ± 0.69	0.062
$PTV1_{60.06Gy}\_D_{2}$ (Gy)	74.88 ± 0.51	75.74 ± 0.79	74.54 ± 0.51	74.59 ± 0.33	0.045†
NDD_{PTV1} (%)	N/A	-1.94 ± 6.03	-0.02 ± 4.35	2.57 ± 6.09	0.046†
$PTV2_{53.46Gy}\_D_{98}$ (Gy)	53.46 ± 0.56	53.09 ± 0.49	52.62 ± 0.86	52.87 ± 0.66	0.080
$PTV2_{53.46Gy}\_D_{2}$ (Gy)	74.12 ± 0.51	75.10 ± 0.75	73.84 ± 0.46	74.04 ± 0.47	0.049†
NDD_{PTV2} (%)	N/A	-1.20 ± 6.27	0.88 ± 5.72	-0.78 ± 5.95	0.053
Esophageal carcinoma					
$PTV_{45Gy}\_D_{98}$ (Gy)	45.10 ± 0.79	44.81 ± 1.02	44.99 ± 0.70	42.72 ± 0.72**	< 0.001†
$PTV_{45Gy}\_D_{2}$ (Gy)	48.22 ± 0.47	48.97 ± 1.80	50.03 ± 1.81**	47.19 ± 0.25**	< 0.001†
NDD (%)	N/A	2.30 ± 9.87	3.41 ± 8.85	7.4 ± 7.71	0.002†
$Breast_{left-sided}$					
$PTV_{50Gy}\_D_{98}$ (Gy)	49.61 ± 0.63	48.64 ± 0.38**	48.89 ± 0.22**	48.51 ± 0.28**	< 0.001†
$PTV_{50Gy}\_D_{2}$ (Gy)	53.77 ± 0.19	52.16 ± 0.038**	52.45 ± 0.23**	52.46 ± 0.10**	< 0.001†
NDD (%)	N/A	1.79 ± 4.52	-0.29 ± 4.08	1.96 ± 4.33	< 0.001†
$Breast_{right-sided}$					
$PTV_{50Gy}\_D_{98}$ (Gy)	49.75 ± 0.63	48.81 ± 0.24**	48.93 ± 0.21**	48.36 ± 0.21**	< 0.01†
$PTV_{50Gy}\_D_{2}$ (Gy)	53.84 ± 0.26	52.39 ± 0.20**	52.36 ± 0.23**	52.42 ± 0.10**	0.422
NDD (%)	N/A	-1.72 ± 4.98	-2.13 ± 4.37	1.62 ± 4.39	< 0.001†
$Cervical_{radical}$					
$PTV_{50Gy}\_D_{98}$ (Gy)	49.98 ± 0.48	49.67 ± 0.28**	49.39 ± 0.35**	48.49 ± 0.13**	< 0.001†
$PTV_{50Gy}\_D_{2}$ (Gy)	53.74 ± 0.25	53.09 ± 0.40**	52.69 ± 0.45**	52.42 ± 0.07**	< 0.001†
NDD (%)	N/A	8.09 ± 22.60	9.45 ± 22.10	8.80 ± 22.28	0.223
$Cervical_{postoperative}$					
$PTV_{50Gy}\_D_{98}$ (Gy)	49.76 ± 0.48	49.42 ± 0.36**	49.29 ± 0.33**	48.61 ± 0.18**	< 0.001†
$PTV_{50Gy}\_D_{2}$ (Gy)	53.77 ± 0.18	53.05 ± 0.17**	52.72 ± 0.32**	52.51 ± 0.06**	< 0.001†
NDD (%)	N/A	-1.24 ± 6.03	0.56 ± 6.81	0.05 ± 6.41	0.136
Rectal carcinoma					
$PTV_{45Gy}\_D_{98}$ (Gy)	45.10 ± 0.37	44.67 ± 0.29**	45.11 ± 0.21	44.38 ± 0.19**	< 0.001†
$PTV_{45Gy}\_D_{2}$ (Gy)	48.34 ± 0.18	47.41 ± 0.31**	47.71 ± 0.18**	47.25 ± 0.15**	< 0.001†
NDD (%)	N/A	5.35 ± 20.42	5.29 ± 21.56	5.97 ± 20.30	0.595

*indicate a statistically significant difference ($P < 0.05$) compared to the approved dose; **indicate a highly significant difference ($P < 0.01$) compared to the approved dose; †indicate a significant difference among the three models; D_{98}, dose received by 98% of the volume; D_2, dose received by 2% of the volume; NDD, normalized dose difference; $Breast_{left-sided}$, breast-conserving surgery for left-sided; $Breast_{right-sided}$, breast-conserving surgery for right-sided; $Cervical_{radical}$, cervical carcinoma before surgery; $Cervical_{postoperative}$, cervical carcinoma after surgery

significant differences among models in NDD ($P = 0.595$), Flex-Net showed better performance in predicting $PTV_{45Gy}\_D_{98}$ and D_2, closely aligning with the approved dose (D_{98}: 45.11 Gy vs. 45.10 Gy).

Figures 2, 3 and 4 illustrates the dose agreement between model predictions and the approved dose across different carcinoma types, showcasing typical examples from the test sets.

Predicted DVH parameters of OARs by different deep learning models

Table 2 compares the predicted DVH parameters for organs at risk (OARs) in nasopharyngeal, esophageal, and breast carcinoma across the three models. For nasopharyngeal carcinoma, all models provided similar predictions for critical OARs (brain stem, spinal cord, optic nerves) ($P > 0.05$), but Flex-Net overestimated the D_{max} of the left lens (6.18 ± 1.36 Gy vs. 4.56 ± 1.35 Gy, $P < 0.05$). For the parotid glands, all models overestimated the D_{mean} and V_{30}, showing highly significant differences ($P < 0.05$ or < 0.01). In esophageal cancer, no significant differences were found for spinalcord_D_{max} ($P = 0.188$) and lung V_5 ($P = 0.087$), V_{20} ($P = 0.466$), and D_{mean} ($P = 0.318$) predictions. However, V_{25} predictions for the heart differed significantly among models ($P = 0.003$), with Flex-Net and Highres-Net overestimating this parameter. For left-sided

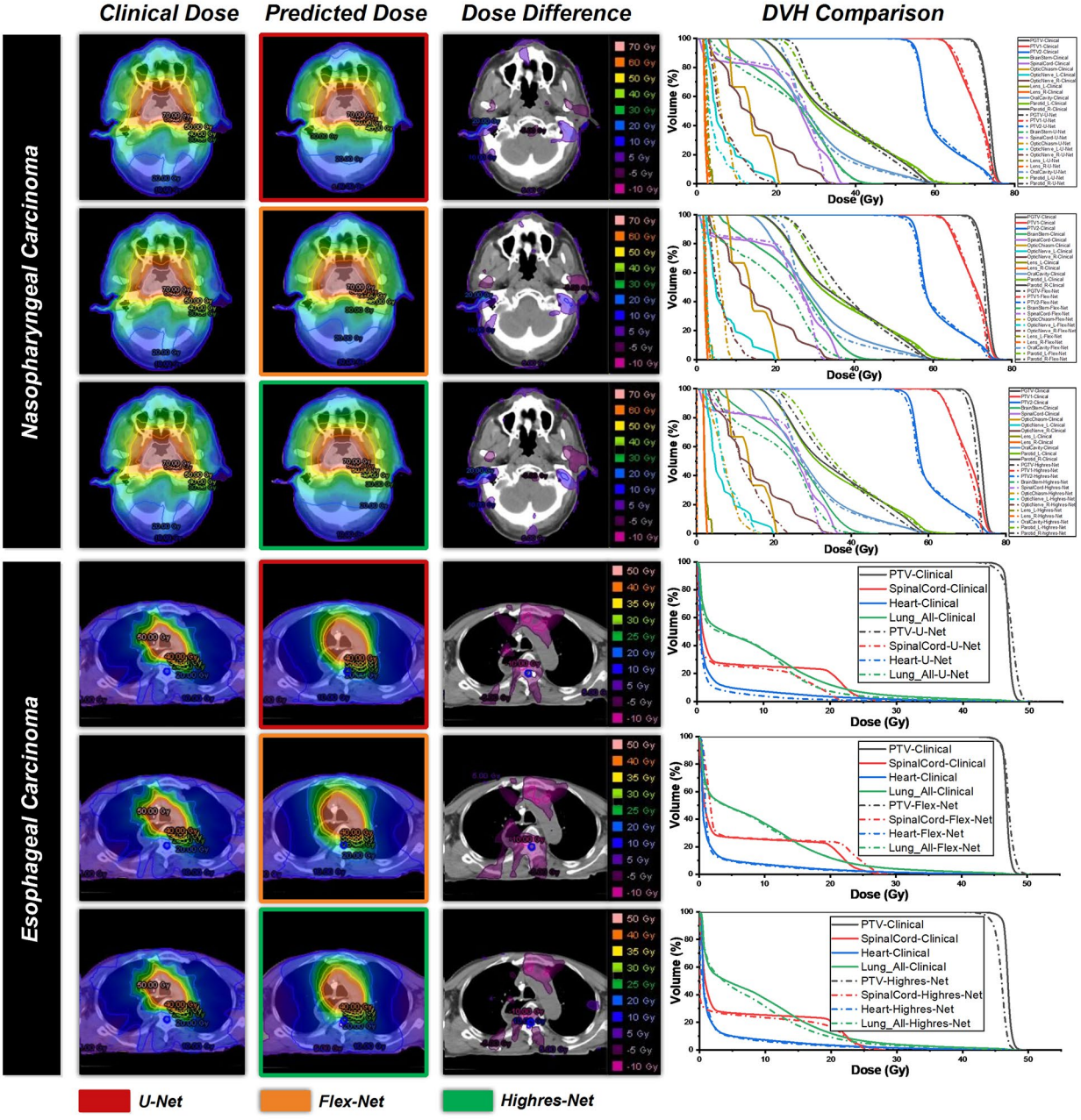

Fig. 2 Comparison of clinical versus predicted dose distributions for representative cases of nasopharyngeal and esophageal carcinoma. The columns show: (1) clinical dose, (2) predicted dose, (3) dose difference map between clinical and predicted doses, and (4) dose-volume histogram (DVH) of the cases

breast cancer, U-Net and Flex-Net significantly overestimated spinalcord_D_{max} (both $P<0.01$). In right-sided breast cancer, Highres-Net provided the most accurate lung_V_{20} prediction (26.17%) compared to the approved dose (26.77%).

Table 3 presents DVH parameters for cervical and rectal carcinoma. In cervical cancer, all models underestimated bladder D_{2cc} ($P<0.001$), while femoral head V_{30} and D_{mean} were generally overestimated in radical cases ($P<0.01$). In postoperative cervical cancer, all models underestimated intestine (D_{2cc}, V_{30}, V_{40}) and colon (V_{30},V_{40}) dose parameters. Similarly, in rectal cancer, all models predicted lower values for intestine and colon D_{2cc}, V_{30}, and V_{40}, with Highres-Net providing the closest prediction of D_{2cc} to the approved dose. U-Net and

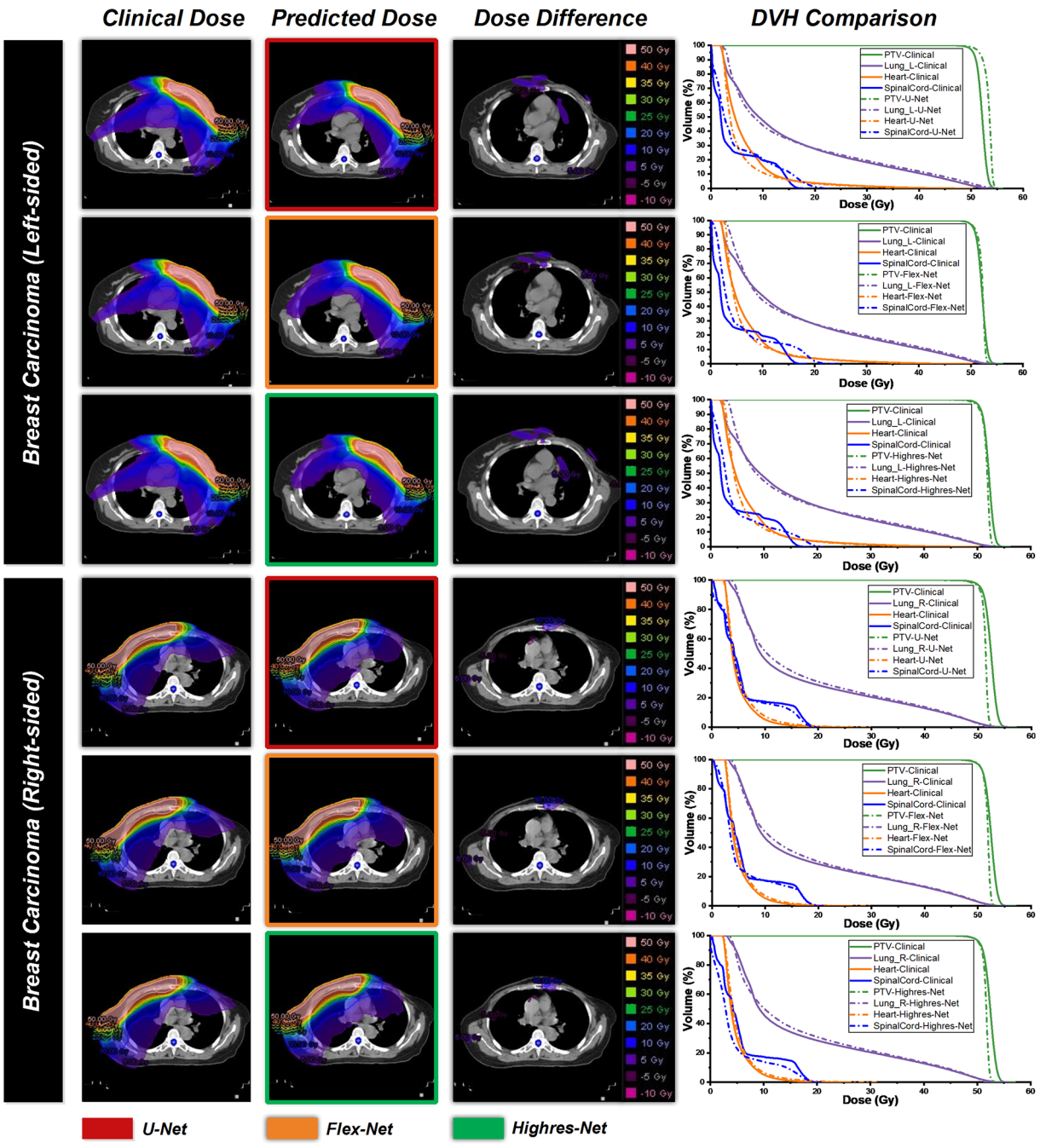

Fig. 3 Comparison of clinical versus predicted dose distributions for representative cases of breast carcinoma. The columns show: (1) clinical dose, (2) predicted dose, (3) dose difference map between clinical and predicted doses, and (4) dose-volume histogram (DVH) of the cases

Flex-Net performed better in predicting femoral head V30, with Flex-Net being the most consistent.

Qualitative evaluation

Table 4 and Fig. 5 provide qualitative scores for the three models across different tumor types. For nasopharyngeal carcinoma, Highres-Net achieved the highest percentage of grade-A (33%) and grade-B (50%) scores, while U-Net had higher proportions of lower grades, with 33% grade-D scores. Highres-Net produced no grade-D predictions, indicating its superior reliability for this tumor type. For esophageal carcinoma, Flex-Net and Highres-Net performed similarly, with Flex-Net achieving more grade-B predictions (33%) and Highres-Net performing slightly better in grade-A (20%). U-Net showed moderate performance, with 47% of predictions rated as grade-B but

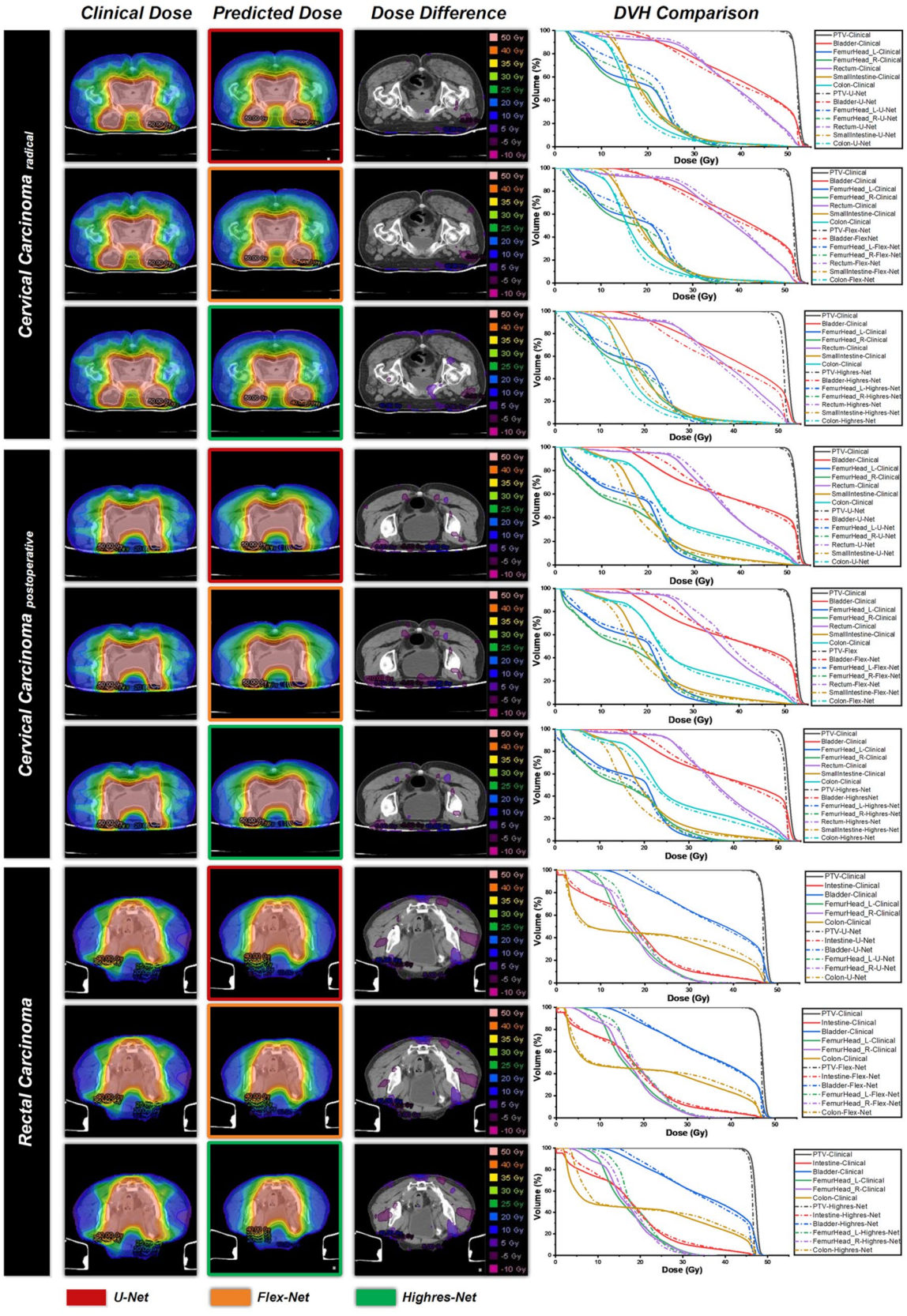

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Comparison of clinical versus predicted dose distributions for representative cases of cervical and rectal carcinoma. The columns show: (1) clinical dose, (2) predicted dose, (3) dose difference map between clinical and predicted doses, and (4) dose-volume histogram (DVH) of the cases

only 13% as grade-A. For both left- and right-sided breast cancer, Flex-Net consistently outperformed the other models in grade-A predictions (38%), while U-Net had lower scores (19% for left-sided and 16% for right-sided). Highres-Net performed better in the grade-B category (for left-sided) but showed fewer grade-A predictions. In cervical cancer (both radical and postoperative), U-Net and Flex-Net clearly outperformed Highres-Net. U-Net achieved 69% grade-A predictions in radical cases, followed by Flex-Net at 44%, while Highres-Net had only 6% grade-A predictions. In postoperative cases, U-Net and Flex-Net performed exceptionally well, with 72% and 76% grade-A scores, respectively, compared to Highres-Net's 20%. For rectal carcinoma, Flex-Net showed the best performance, with 76% of predictions rated as grade-A, outperforming U-Net (57%) and Highres-Net (19%). Flex-Net had no grade-C or -D predictions, demonstrating its reliability for this tumor type.

The "Best Selection" method, which selects the top prediction from U-Net, Flex-Net, and Highres-Net based on qualitative scores, consistently outperformed individual models across all tumor types. This method combines the best aspects of each model, offering more reliable and optimal prediction results compared to using any single model alone.

Discussion

This study systematically compared the performance of three deep learning models, namely U-Net, Flex-Net, and Highres-Net, in predicting radiotherapy dose distributions across various tumor types, including nasopharyngeal, esophageal, breast, cervical, and rectal carcinomas. For nasopharyngeal, breast, and postoperative cervical cancer, the average normalized dose difference (NDD) within the tumor targets was below 3%. Qualitative evaluation indicated that Flex-Net achieved superior accuracy in breast cancer dose predictions, while U-Net performed best for cervical carcinoma (radical). Highres-Net was more effective for nasopharyngeal carcinoma. Additionally, the "Best Selection" method (Table 4; Fig. 5) proved to be a highly reliable approach, consistently yielding the best qualitative outcomes by selecting the top-performing prediction from the three models.

In recent years, deep learning techniques have gained significant traction in the field of radiation oncology [25], presenting innovative approaches for automating dose distribution predictions. These models have already demonstrated successful applications across various tumor sites [2–22], showcasing their potential to enhance treatment precision. However, to date, no studies have conducted multi-model dose prediction research across different tumor types. This is primarily due to the time-consuming nature of data preprocessing, model training, and code development, which has presented challenges for radiation oncologists and medical physicists in advancing this area of research. With the increasing availability of commercial software platforms such as the AccuLearning platform used in this study, it is now possible to conduct various studies with higher efficiency and minimal coding effort [26]. These advancements enable tasks such as automatic segmentation of target volumes and organs at risk [26–29], as well as the automated dose prediction utilized in this study. In previous studies on nasopharyngeal carcinoma [10], esophageal cancer [5, 6], breast cancer [14, 30], cervical cancer [17–19], and rectal cancer [3, 22], model prediction results were typically evaluated only through dosimetric parameters, without incorporating subjective qualitative analysis. However, considering that the dose distributions in radiotherapy plans used for training lack a definitive gold standard, subjective evaluation of the prediction results is essential [31]. Such qualitative analysis provides a more comprehensive assessment of model performance, especially when dosimetric parameters may appear similar, but the clinical applicability of the predictions could vary. By integrating both quantitative and qualitative analyses, this approach offers a more accurate basis for model optimization and strengthens the foundation for their application in clinical practice.

In this study, we trained and evaluated three deep learning models using data from 622 patients across five distinct carcinoma types, representing a diverse range of target shapes, sizes, and anatomical locations. Our quantitative and qualitative analyses underscore the tumor-specific strengths of different dose prediction models. Quantitatively, as shown in Tables 1, 2 and 3, Highres-Net demonstrated superior performance in anatomically complex regions such as nasopharyngeal carcinoma, accurately predicting dose-volume histogram (DVH) parameters with minimal deviation from the approved dose. In contrast, Flex-Net consistently performed excelled in breast and rectal cancers, likely due to the simpler anatomical structures in these regions. This trend was corroborated by the qualitative analysis in Table 4; Fig. 5, where Flex-Net achieved higher qualitative scores for rectal carcinoma, while Highres-Net excelled in nasopharyngeal carcinoma. These findings suggest that each model leverages specific strengths aligned with the tumor type and its associated anatomical complexity. This tumor specificity may result from the variations in target volume and organ-at-risk (OAR) geometry across cancer types, with certain model architectures better

suited to these characteristics. Highres-Net's strong performance in complex regions like the head and neck may be due to its enhanced feature extraction capabilities, while Flex-Net's reliable predictions in breast and rectal cancers likely reflect its generalized training on simpler structures. Prior studies, such as those by Wen et al. [21] and Li et al. [13], have highlighted the utility of different models for dose prediction in cervical and breast cancers, respectively. Our proposed "Best Selection" scheme, which consolidates the strengths of all three models, consistently outperformed individual models across all tumor types, supporting the potential of a hybrid approach to achieve reliable prediction outcomes. Furthermore, patient-specific dose predictions generated by this scheme can guide further refinement of treatment plans developed from initial dose predictions. When initial fluence proves suboptimal, optimization can proceed by using these patient-specific dose predictions as target objectives, with the initial fluence serving as a baseline [12]. As a result, this approach enhances the precision and effectiveness of radiotherapy, ensuring that dose distributions are finely tuned to both the tumor and surrounding organs-at-risk.

Despite the promising performance of these deep learning models across multiple tumor types, several limitations remain. First, the models were trained on data from a single institution, suggesting a need for further training on multi-institutional datasets to enhance generalizability. Additionally, external validations are essential to confirm model robustness and accuracy. The "black-box" nature of deep learning models, such as U-Net, Flex-Net, and Highres-Net, limits their interpretability, which could hinder clinical adoption. Moreover, this study only includes convolution-based models. To offer a more comprehensive comparison, we plan to explore Transformer-based models in future research. This will provide valuable insights into the performance differences between convolutional and attention-based architectures in predicting radiotherapy dose distributions, further enriching our analysis. Finally, although the clinical doses used for model training were physician-approved, they may not represent optimal values, highlighting the need for future studies to incorporate case-specific quality assurance (QA) to enhance precision. Addressing these limitations will enhance model applicability and support the development of reliable predictive tools for clinical radiotherapy.

Conclusion

In conclusion, this study evaluated the predictive performance of three deep learning models for radiotherapy dose distribution across five carcinoma types, revealing tumor-specific strengths among the models. The "Best Selection" approach effectively combined these strengths, achieving enhanced prediction accuracy across all tumor sites. These findings underscore the potential of deep learning models to advance radiotherapy planning, minimize clinician workload, and ultimately improve treatment outcomes.

Table 2 Predicted DVH parameters of organs-at-risk (OARs) for nasopharyngeal, esophageal, and breast carcinomas across the three models

Metrics	Dose predicted by DL models				*P*-value
	$Dose_{Approved}$	$Dose_{U-Net}$	$Dose_{Flex-Net}$	$Dose_{Highres-Net}$	
Nasopharyngeal carcinoma					
Brain Stem					
D_{max} (Gy)	49.17 ± 5.68	52.68 ± 7.18	50.63 ± 6.56	50.66 ± 6.71	0.232
Spinal cord					
D_{max} (Gy)	38.46 ± 4.20	41.59 ± 6.34	41.65 ± 7.08	40.71 ± 6.26	0.714
Len_L					
D_{max} (Gy)	4.56 ± 1.35	4.54 ± 0.51	6.18 ± 1.36*	3.87 ± 1.09	0.041†
Len_R					
D_{max} (Gy)	4.26 ± 1.48	4.19 ± 0.48	5.09 ± 0.67	5.30 ± 0.89	0.088
OpticNerve_L					
D_{max} (Gy)	34.69 ± 13.57	23.92 ± 12.46	23.38 ± 13.57	21.96 ± 12.69	0.429
OpticNerve_R					
D_{max} (Gy)	36.48 ± 12.37	22.91 ± 7.63*	24.36 ± 8.67	24.96 ± 10.57	0.516
OpticChiasm					
D_{max} (Gy)	36.07 ± 17.42	25.29 ± 12.62	25.74 ± 15.37	23.80 ± 15.96	0.452
Parotid_L					
D_{mean} (Gy)	37.59 ± 1.77	39.99 ± 3.02*	39.94 ± 2.07**	40.66 ± 2.57**	0.182
V_{30} (%)	65.12 ± 9.15	77.57 ± 9.58*	77.48 ± 5.01**	80.63 ± 7.45**	0.282
Parotid_R					
D_{mean} (Gy)	36.95 ± 1.93	39.15 ± 2.72*	39.60 ± 2.66*	39.53 ± 2.57*	0.350
V_{30} (%)	60.60 ± 6.73	71.06 ± 7.19**	76.90 ± 7.71**	75.18 ± 6.55**	0.045†
OralCavity					
D_{mean} (Gy)	30.20 ± 2.68	29.92 ± 1.41	30.20 ± 2.68	29.90 ± 1.20	0.824
Esophageal carcinoma					
Spinal cord					
D_{max} (Gy)	26.71 ± 4.51	28.09 ± 1.45	28.38 ± 2.07	27.63 ± 1.29	0.188
Heart					
D_{mean} (Gy)	3.03 ± 3.05	3.32 ± 3.22*	3.48 ± 3.16*	3.56 ± 3.24	0.443
V_{30} (%)	1.93 ± 2.65	1.86 ± 2.21	2.06 ± 2.50	1.98 ± 2.38	0.069
V_{25} (%)	3.06 ± 4.20	2.90 ± 3.53	3.20 ± 3.84	3.20 ± 3.88	0.003†
Lung_All					
D_{mean} (Gy)	7.38 ± 1.80	7.53 ± 2.04	7.58 ± 2.01	7.61 ± 2.02	0.318
V_5 (%)	41.72 ± 10.07	43.24 ± 1.96	44.33 ± 12.40	43.25 ± 12.75	0.087
V_{20} (%)	9.42 ± 2.89	9.92 ± 3.20	10.36 ± 4.20	10.77 ± 4.12	0.466
$Breast_{left-sided}$					
Spinal cord					
D_{max} (Gy)	20.44 ± 2.82	22.25 ± 1.13**	22.58 ± 2.42**	20.84 ± 1.19	< 0.001†
Heart					
D_{mean} (Gy)	6.38 ± 0.90	6.42 ± 0.72	6.44 ± 0.75	6.27 ± 0.67	0.008†
V_{30} (%)	2.08 ± 0.97	2.10 ± 0.90	1.89 ± 0.92	1.72 ± 0.87*	< 0.001†
Lung_L					
D_{mean} (Gy)	15.43 ± 1.13	15.71 ± 1.45	15.56 ± 1.49	15.37 ± 1.56	< 0.001†
V_{20} (%)	26.87 ± 1.98	28.36 ± 3.95*	27.83 ± 3.94	27.03 ± 4.25	< 0.001†
$Breast_{right-sided}$					
Spinal cord					
D_{max} (Gy)	20.82 ± 3.09	20.46 ± 1.34	21.38 ± 1.48	22.12 ± 1.14	< 0.001†
Heart					
D_{mean} (Gy)	4.50 ± 0.51	4.44 ± 0.42	4.43 ± 0.36	4.39 ± 0.27	< 0.001†
Lung_R					

Table 2 (continued)

Metrics	Dose predicted by DL models				*P*-value
	Dose $_{Approved}$	**Dose $_{U-Net}$**	**Dose $_{Flex-Net}$**	**Dose $_{Highres-Net}$**	
D_{mean} (Gy)	15.54 ± 1.14	15.78 ± 1.27	15.67 ± 1.25	15.38 ± 1.18	< 0.001†
V_{20} (%)	26.77 ± 2.52	27.26 ± 3.12	27.29 ± 3.04	26.17 ± 3.14	< 0.001†

*indicate a statistically significant difference ($P < 0.05$) compared to the approved dose; **indicate a highly significant difference ($P < 0.01$) compared to the approved dose; †indicate a significant difference among the three models; D_{max}, maximum dose received by the volume; D_{mean}, average dose received by the volume; V_x, percentage of the volume receiving a dose greater than or equal to x Gy; $Breast_{left-sided}$, breast-radical mastectomy for left-sided; $Breast_{right-sided}$, breast-radical mastectomy for right-sided

Table 3 Predicted DVH parameters of OARs for cervical and rectal carcinomas across the three models

Metrics	Dose predicted by DL models				*P*-value
	$Dose_{Approved}$	$Dose_{U-Net}$	$Dose_{Flex-Net}$	$Dose_{Highres-Net}$	
$Cervical_{radical}$					
Bladder					
D_{2cc} (Gy)	53.91 ± 0.34	53.14 ± 0.40**	52.5 ± 0.43**	52.27 ± 0.30**	< 0.001†
V_{30} (%)	69.96 ± 11.55	71.73 ± 10.02	71.55 ± 10.16	69.26 ± 10.79	< 0.001†
V_{40} (%)	56.23 ± 12.25	56.37 ± 10.01	56.00 ± 9.96	53.09 ± 10.14**	< 0.001†
Femurhead_L					
V_{30} (%)	4.99 ± 2.29	7.85 ± 1.65**	9.15 ± 2.24**	5.55 ± 1.59	< 0.001†
D_{mean} (Gy)	15.50 ± 1.90	17.23 ± 1.48**	17.03 ± 1.66**	16.59 ± 0.88*	0.012†
Femurhead_R					
V_{30} (%)	4.60 ± 2.41	8.67 ± 1.77**	7.64 ± 2.72**	7.39 ± 1.28**	0.057
D_{mean} (Gy)	14.33 ± 1.90	16.02 ± 1.57**	15.93 ± 1.08**	16.25 ± 0.88**	0.443
Intestine					
D_{2cc} (Gy)	52.61 ± 0.72	52.16 ± 0.63**	52.32 ± 0.81	52.33 ± 0.60	0.514
V_{30} (%)	13.93 ± 8.34	14.47 ± 9.46	14.20 ± 9.27	13.35 ± 9.27	< 0.001†
V_{40} (%)	6.36 ± 5.64	6.56 ± 6.31	6.47 ± 6.41	6.04 ± 6.03	< 0.001†
Colon					
D_{2cc} (Gy)	49.09 ± 7.70	49.95 ± 6.32*	49.35 ± 6.21	49.48 ± 6.44	< 0.001†
V_{30} (%)	20.19 ± 11.51	20.38 ± 11.51	20.28 ± 11.71	18.95 ± 11.40	< 0.001†
V_{40} (%)	10.93 ± 6.85	11.46 ± 7.95	11.41 ± 7.86	10.55 ± 7.54	< 0.001†
Rectum					
D_{2cc} (Gy)	52.39 ± 0.70	52.15 ± 0.58	52.02 ± 0.31	52.06 ± 0.45*	0.555
V_{40} (%)	57.61 ± 13.31	62.27 ± 13.17*	63.48 ± 13.45**	62.21 ± 14.08*	0.177
V_{50} (%)	23.04 ± 9.70	22.52 ± 6.73	25.56 ± 7.23	21.88 ± 6.85	< 0.001†
$Cervical_{postoperative}$					
Bladder					
D_{2cc} (Gy)	53.95 ± 0.27	52.75 ± 0.52**	52.70 ± 0.41**	52.69 ± 0.11**	0.478
V_{30} (%)	65.60 ± 9.64	65.64 ± 8.48	65.16 ± 8.93	64.46 ± 9.33	0.002†
V_{40} (%)	49.79 ± 9.54	48.75 ± 9.20	48.31 ± 9.51*	48.45 ± 9.78	0.087
Femurhead_L					
V_{30} (%)	7.66 ± 5.17	6.58 ± 2.03	7.30 ± 2.77	5.77 ± 1.13	0.002†
D_{mean} (Gy)	15.99 ± 2.59	16.46 ± 1.62	16.44 ± 1.70	15.15 ± 1.14	< 0.001†
Femurhead_R					
V_{30} (%)	7.63 ± 3.87	7.31 ± 2.34	7.24 ± 2.39	6.25 ± 1.69	0.001†
D_{mean} (Gy)	15.44 ± 2.63	16.08 ± 1.91**	15.99 ± 2.12*	15.80 ± 1.31	0.168
Intestine					
D_{2cc} (Gy)	53.05 ± 0.54	52.55 ± 0.42**	52.68 ± 0.66*	52.34 ± 0.41**	0.072
V_{30} (%)	12.50 ± 4.35	11.70 ± 4.71	11.55 ± 4.60	11.08 ± 4.66**	< 0.001†
V_{40} (%)	5.38 ± 2.46	5.33 ± 2.90	5.15 ± 2.82	5.13 ± 2.89	0.006†
Colon					
D_{2cc} (Gy)	44.43 ± 8.93	45.43 ± 8.08	45.01 ± 8.40	44.75 ± 7.51	0.001†
V_{30} (%)	16.12 ± 12.38	15.43 ± 11.05	15.20 ± 11.23	14.60 ± 10.57	0.001†
V_{40} (%)	8.66 ± 9.43	7.52 ± 7.15	7.30 ± 6.97	7.18 ± 6.70	0.038†
Rectum					
D_{2cc} (Gy)	51.82 ± 0.83	51.74 ± 0.85	52.29 ± 0.53*	52.40 ± 0.50**	< 0.001†
V_{40} (%)	47.39 ± 11.79	49.79 ± 11.69*	49.76 ± 10.67*	48.54 ± 10.42	0.072
V_{50} (%)	14.44 ± 6.09	16.42 ± 7.12*	16.19 ± 5.95*	13.02 ± 4.19*	< 0.001†
Rectal carcinoma					
Bladder					
D_{2cc} (Gy)	48.28 ± 0.23	48.32 ± 0.34**	47.55 ± 0.18**	47.22 ± 0.46**	< 0.001†
V_{30} (%)	67.87 ± 9.30	66.59 ± 9.13	66.55 ± 9.23	66.96 ± 9.45	0.435
V_{40} (%)	49.45 ± 8.02	47.27 ± 8.42*	47.84 ± 8.90*	45.84 ± 8.80**	< 0.001†
Femurhead_L					

Table 3 (continued)

Metrics	Dose predicted by DL models				*P*-value
	$Dose_{Approved}$	$Dose_{U-Net}$	$Dose_{Flex-Net}$	$Dose_{Highres-Net}$	
D_{mean}	19.01 ± 1.88	18.94 ± 1.11	19.18 ± 1.12	18.80 ± 0.71	0.073
V_{30} (%)	2.84 ± 2.89	2.45 ± 1.37	3.01 ± 1.69	1.37 ± 0.74	< 0.001†
Femurhead_R					
D_{mean}	18.51 ± 1.82	18.08 ± 1.21	18.82 ± 1.46	17.76 ± 0.98	< 0.001†
V_{30} (%)	2.86 ± 3.19	2.11 ± 1.44	2.91 ± 2.11	1.49 ± 1.01	< 0.001†
Intestine					
D_{2cc} (Gy)	44.81 ± 5.44	44.39 ± 5.92*	44.35 ± 6.00	44.53 ± 6.06	0.041
V_{30} (%)	12.72 ± 7.68	11.67 ± 7.27	11.78 ± 7.34	11.82 ± 7.57	0.442
V_{40} (%)	6.32 ± 5.61	5.55 ± 5.06*	5.61 ± 5.15*	5.41 ± 5.08**	0.052
Colon					
D_{2cc} (Gy)	44.06 ± 6.51	43.80 ± 6.81	43.77 ± 6.47**	43.90 ± 6.77	0.404
V_{30} (%)	21.59 ± 14.49	20.56 ± 13.68	21.36 ± 14.32	20.35 ± 14.30	0.039†
V_{40} (%)	13.43 ± 11.72	12.70 ± 11.77	12.71 ± 11.56	12.49 ± 11.60	0.195

*indicate a statistically significant difference ($P < 0.05$) compared to the approved dose; **indicate a highly significant difference ($P < 0.01$) compared to the approved dose; †indicate a significant difference among the three models; D_{mean}, average dose received by the volume; V_x, percentage of the volume receiving a dose greater than or equal to x Gy; D_{2cc}, minimum dose received by the 2 cc hottest volume within the region; $Cervical_{radical}$, cervical carcinoma before surgery; $Cervical_{postoperative}$, cervical carcinoma after surgery

Table 4 Qualitative scores assigned to predicted dose distributions from each model

Tumor sites (testing set, $n = 123$)	Qualitative Scores (A/B/C/D)			
	Grade A	Grade B	Grade C	Grade D
Nasopharyngeal carcinoma ($n = 6$)				
U-Net	1 (17%)	1 (17%)	2 (33%)	2 (33%)
Flex-Net	1 (17%)	1 (17%)	3 (49%)	1 (17%)
Highres-Net	2 (33%)	3 (50%)	1 (17%)	0 (0%)
Best Selection	2 (33%)	3 (50%)	1 (17%)	0 (0%)
Esophageal carcinoma ($n = 15$)				
U-Net	2 (13%)	7 (47%)	5 (33%)	1 (7%)
Flex-Net	2 (13%)	5 (33%)	3 (21%)	5 (33%)
Highres-Net	3 (20%)	3 (20%)	4 (27%)	5 (33%)
Best Selection	5 (33%)	4 (27%)	5 (33%)	1 (7%)
Breast$_{\text{left–sided}}$ ($n = 21$)				
U-Net	4 (19%)	12 (57%)	5 (24%)	0 (0%)
Flex-Net	8 (38%)	12 (57%)	1 (5%)	0 (0%)
Highres-Net	1 (5%)	13 (62%)	6 (28%)	1 (5%)
Best Selection	8 (38%)	12 (57%)	1 (5%)	0 (0%)
Breast$_{\text{right–sided}}$ ($n = 19$)				
U-Net	3 (16%)	11 (58%)	5 (26%)	0 (0%)
Flex-Net	7 (37%)	8 (42%)	4 (21%)	0 (0%)
Highres-Net	1 (5%)	7 (37%)	11 (58%)	0 (0%)
Best Selection	7 (37%)	8 (42%)	4 (21%)	0 (0%)
Cervical$_{\text{radical}}$ ($n = 16$)				
U-Net	11 (69%)	3 (19%)	2 (12%)	0 (0%)
Flex-Net	7 (44%)	8 (50%)	1 (6%)	0 (0%)
Highres-Net	1 (6%)	14 (88%)	1 (6%)	0 (0%)
Best Selection	13 (81%)	3 (19%)	0 (0%)	0 (0%)
Cervical$_{\text{postoperative}}$ ($n = 25$)				
U-Net	18 (72%)	7 (28%)	0 (0%)	0 (%)
Flex-Net	19 (76%)	6 (24%)	0 (0%)	0 (%)
Highres-Net	5 (20%)	20 (80%)	0 (0%)	0 (%)
Best Selection	23 (92%)	2 (8%)	0 (0%)	0 (%)
Rectal carcinoma ($n = 21$)				
U-Net	12 (57%)	8 (38%)	1 (5%)	0 (%)
Flex-Net	16 (76%)	5 (24%)	0 (%)	0 (%)
Highres-Net	4 (19%)	16 (76%)	1 (5%)	0 (%)
Best Selection	19 (90%)	2 (10%)	0 (%)	0 (%)

Best Selection, chosen as the best prediction from U-Net, Flex-Net, and Highres-Net based on qualitative scores (e.g., ABB→A); Breast$_{\text{left–sided}}$, breast-radical mastectomy for left-sided; Breast$_{\text{right–sided}}$, breast-radical mastectomy for right-sided; Cervical$_{\text{preoperative}}$, cervical carcinoma before surgery; Cervical$_{\text{postoperative}}$, cervical carcinoma after surgery

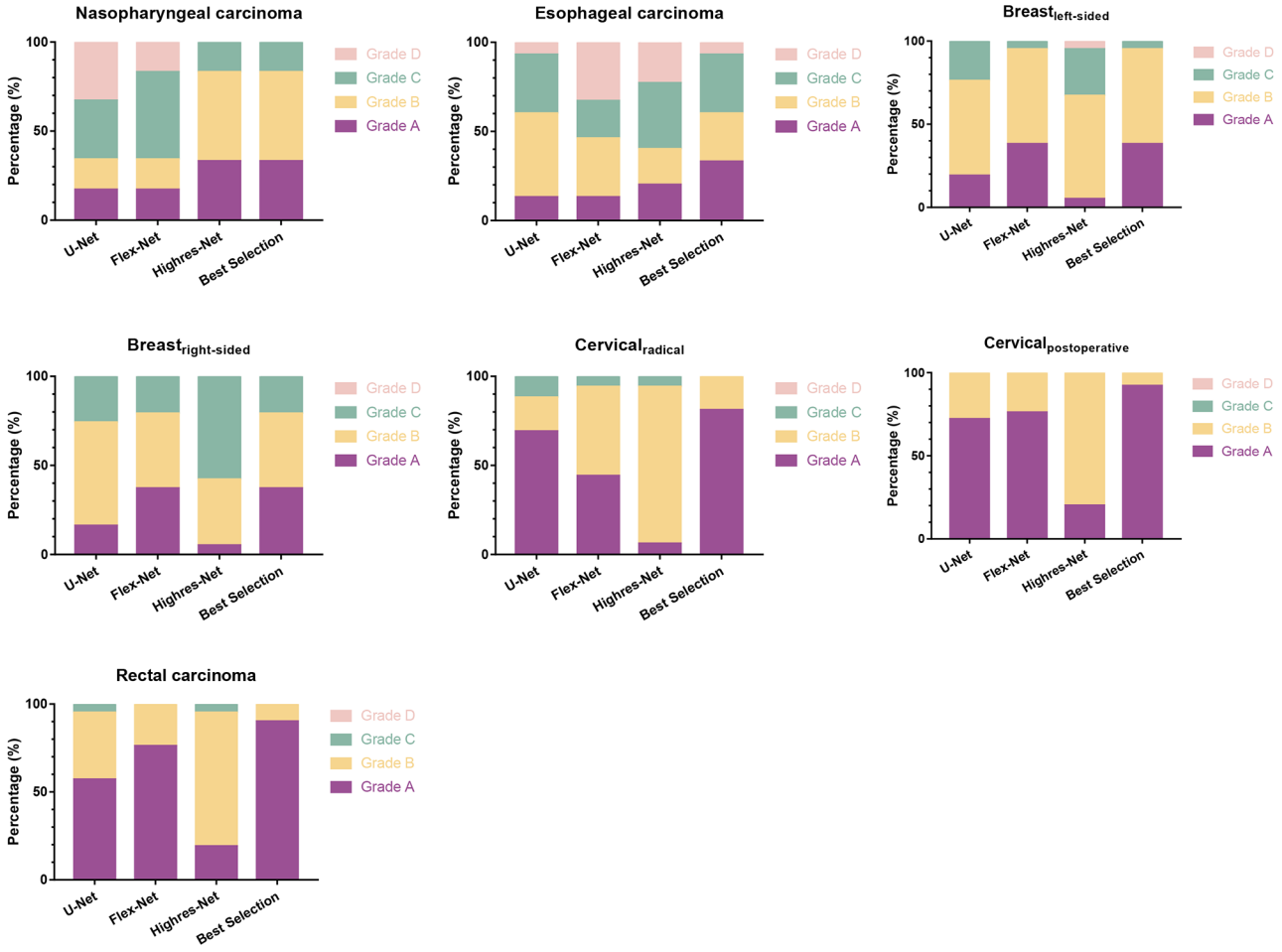

Fig. 5 Stacked histogram showing the distribution of qualitative scores for each carcinoma type across different models

Abbreviations

NPC	Nasopharyngeal Carcinoma
EC	Esophageal Carcinoma
$Breast_{left-sided}$	Left-Sided Breast Carcinoma
$Breast_{right-sided}$	Right-Sided Breast Carcinoma
$Cervical_{radical}$	Cervical Carcinoma Treated with Radical Radiotherapy
$Cervical_{postoperative}$	Cervical Carcinoma Receiving Postoperative Radiotherapy
RC	Rectal Carcinoma
DVH	Dose-Volume Histograms
RT	Radiotherapy
IMRT	Intensity-Modulated Radiation Therapy
VMAT	Volumetric Modulated Arc Therapy
CNN	Convolutional Neural Networks
KBP	Knowledge-Based Planning
MAE	Mean Absolute Error
PTV	Planning Target Volume
OAR	Organ at Risk
NDD	Normalized Dose Difference

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13014-025-02634-7.

Supplementary Material 1

Acknowledgements

We would like to thank Qichao Zhou, Yunfeng Gu, Wei Zhang, Jiayao Gao, and Xingwen Yuan (Manteia Technologies Co.,Ltd) for their technical advice on deep learning-based dose prediction.

Author contributions

Original idea and study design (SSL, JZ, ZH); Data collection and analysis (SBG, ZH, YZ, LQ, JBG, JL); Manuscript writing (ZH, SSL, JZ); Important contributions to model Training (ZH, ZDL, JBG). All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (No. 82202300, 82172072), the Foundation of Shandong Provincial Key Research and Development (No. 2024CXPT084), and the Noncommunicable Chronic Diseases-National Science and Technology Major Project (No. 2024ZD0519901).

Data availability

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

Declarations

Ethics approval and consent to participate

The ethics committee approved this retrospective research at Nanjing Drum Tower Hospital, and informed consent was waived. The radiotherapy data were anonymized for the scientific purpose of this work. This retrospective study was performed in line with the principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]The Comprehensive Cancer Centre of Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, Jiangsu 210000, China
[2]The School of Electronic Science and Engineering, Nanjing University, Nanjing, Jiangsu 210000, China
[3]Department of Radiology, The First Affiliated Hospital with Nanjing Medical University, Nanjing, Jiangsu 210000, China
[4]Department of Radiotherapy Technology Center, The Affiliated Hospital of Qingdao University, Qingdao, Shandong 266001, China
[5]The Laboratory of Image Science and Technology, The School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China
[6]Department of Radiation Oncology Physics and Technology, Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences, Jinan, Shandong 250000, China
[7]Shandong Provincial Key Medical and Health Laboratory of Pediatric Cancer Precision Radiotherapy, (Shandong Cancer Hospital), Jinan, Shandong 250000, China
[8]Centre de Recherche en Information BioMedicale Sino-Français, Nanjing, China

Received: 26 November 2024 / Accepted: 6 April 2025
Published online: 19 May 2025

References

1. Chandra RA, Keane FK, Voncken FE, Thomas CR. Contemporary radiotherapy: present and future[J]. Lancet. 2021;398(10295):171–84.
2. Hunte SO, Clark CH, Zyuzikov N, Nisbet A. Volumetric modulated Arc therapy (VMAT): a review of clinical outcomes—what is the clinical evidence for the most effective implementation?[J]. Br J Radiol. 2022;95(1136):20201289.
3. Song Y, Hu J, Liu Y, Hu H, Huang Y, Bai S, et al. Dose prediction using a deep neural network for accelerated planning of rectal cancer radiotherapy[J]. Radiother Oncol. 2020;149:111–6.
4. Kandalan RN, Nguyen D, Rezaeian NH, Barragán-Montero AM, Breedveld S, Namuduri K, et al. Dose prediction with deep learning for prostate cancer radiation therapy: model adaptation to different treatment planning practices[J]. Radiother Oncol. 2020;153:228–35.
5. Barragán-Montero AM, Thomas M, Defraene G, Michiels S, Haustermans K, Lee JA, et al. Deep learning dose prediction for IMRT of esophageal cancer: the effect of data quality and quantity on model performance[J]. Phys Med. 2021;83:52–63.
6. Zhang J, Liu S, Yan H, Li T, Mao R, Liu J. Predicting voxel-level dose distributions for esophageal radiotherapy using densely connected network with dilated convolutions[J]. Phys Med Biol. 2020;65(20):205013.
7. Nguyen D, Jia X, Sher D, Lin M-H, Iqbal Z, Liu H, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture[J]. Phys Med Biol. 2019;64(6):065020.
8. Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique[J]. Med Phys. 2019;46(1):370–81.
9. Liu S, Zhang J, Li T, Yan H, Liu J. A cascade 3D U-Net for dose prediction in radiotherapy[J]. Med Phys. 2021;48(9):5574–82.
10. Chen X, Men K, Li Y, Yi J, Dai J. A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning[J]. Med Phys. 2019;46(1):56–64.
11. Chandran LP, KA AN, Puzhakkal N, Makuny D, MemU-Net. A new volumetric dose prediction model using deep learning techniques in radiation treatment planning[J]. Biomed Signal Process Control. 2023;85:104940.
12. Li Y, Cai W, Xiao F, Zhou X, Cai J, Zhou L, et al. Simultaneous dose distribution and fluence prediction for nasopharyngeal carcinoma IMRT[J]. Radiat Oncol. 2023;18(1):110.
13. Bakx N, Bluemink H, Hagelaar E, van der Sangen M, Theuws J, Hurkmans C. Development and evaluation of radiotherapy deep learning dose prediction models for breast cancer[J]. Phys Imaging Radiation Oncol. 2021;17:65–70.
14. Bai X, Zhang J, Wang B, Wang S, Xiang Y, Hou Q. Sharp loss: a new loss function for radiotherapy dose prediction based on fully convolutional networks[J]. Biomed Eng Online. 2021;20:1–15.
15. Shao Y, Zhang X, Wu G, Gu Q, Wang J, Ying Y, et al. Prediction of three-dimensional radiotherapy optimal dose distributions for lung cancer patients with asymmetric network[J]. IEEE J Biomedical Health Inf. 2020;25(4):1120–7.
16. Momin S, Lei Y, Wang T, Zhang J, Roper J, Bradley JD, et al. Learning-based dose prediction for pancreatic stereotactic body radiation therapy using dual pyramid adversarial network[J]. Phys Med Biol. 2021;66(12):125019.

17. Zhang G, Jiang Z, Zhu J, Wang L. 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[J]. Radiat Oncol. 2022;17(1):179.
18. Li H, Peng X, Zeng J, Xiao J, Nie D, Zu C, et al. Explainable attention guided adversarial deep network for 3D radiotherapy dose distribution prediction[J]. Knowl Based Syst. 2022;241:108324.
19. Qilin Z, Peng B, Ang Q, Weijuan J, Ping J, Hongqing Z, et al. The feasibility study on the generalization of deep learning dose prediction model for volumetric modulated Arc therapy of cervical cancer[J]. J Appl Clin Med Phys. 2022;23(6):e13583.
20. Zhan B, Xiao J, Cao C, Peng X, Zu C, Zhou J, et al. Multi-constraint generative adversarial network for dose prediction in radiotherapy[J]. Med Image Anal. 2022;77:102339.
21. Wen L, Xiao J, Zeng J, Zu C, Wu X, Zhou J, et al. Multi-level progressive transfer learning for cervical cancer dose prediction[J]. Pattern Recogn. 2023;141:109606.
22. Zhou J, Peng Z, Song Y, Chang Y, Pei X, Sheng L, et al. A method of using deep learning to predict three-dimensional dose distributions for intensity-modulated radiotherapy of rectal cancer[J]. J Appl Clin Med Phys. 2020;21(5):26–37.
23. Chatterjee A, Serban M, Abdulkarim B, Panet-Raymond V, Souhami L, Shenouda G, et al. Performance of knowledge-based radiation therapy planning for the glioblastoma disease site[J]. Int J Radiation Oncology* Biology* Phys. 2017;99(4):1021–8.
24. Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy[J]. Radiother Oncol. 2016;120(3):473–9.
25. Huynh E, Hosny A, Guthier C, Bitterman DS, Petit SF, Haas-Kogan DA, et al. Artificial intelligence in radiation oncology[J]. Nat Reviews Clin Oncol. 2020;17(12):771–81.
26. Zhang W, Chen Z, Liang Z, Hu Y, Zhou Q. AccuLearning: a user-friendly deep learning auto-segmentation platform for radiotherapy[J]. Int J Radiat Oncol Biol Phys. 2021;111(3):e122.
27. Qiu W, Zhang W, Ma X, Kong Y, Shi P, Fu M, et al. Auto-segmentation of important centers of growth in the pediatric skeleton to consider during radiation therapy based on deep learning[J]. Med Phys. 2023;50(1):284–96.
28. Hou Z, Gao S, Liu J, Yin Y, Zhang L, Han Y, et al. Clinical evaluation of deep learning-based automatic clinical target volume segmentation: a single-institution multi-site tumor experience[J]. Radiol Med. 2023;128(10):1250–61.
29. Johnson CL, Press RH, Simone B 2, Tsai P, Hu L et al. Clinical validation of commercial deep-learning based auto-segmentation models for organs at risk in the head and neck region: a single institution study[J]. Front Oncol. 2024;14.
30. Hedden N, Xu H. Radiation therapy dose prediction for left-sided breast cancers using two-dimensional and three-dimensional deep learning models[J]. Phys Med. 2021;83:101–7.
31. Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D, et al. Overview of artificial intelligence-based applications in radiotherapy: recommendations for implementation and quality assurance[J]. Radiother Oncol. 2020;153:55–66.

Publisher's note