Automatic segmentation of high-risk clinical target volume for tandem-and-ovoids brachytherapy patients using an asymmetric dual-path convolutional neural network

Yufeng Cao[1,#], **April Vassantachart**[2,#], **Omar Ragab**[1], **Shelly Bian**[1], **Priya Mitra**[1], **Zhengzheng Xu**[1], **Audrey Zhuang Gallogly**[1], **Jing Cui**[1], **Zhilei Liu Shen**[1], **Salim Balik**[1], **Michael Gribble**[3], **Eric L. Chang**[1], **Zhaoyang Fan**[1,4], **Wensha Yang**[1]

[1]Department of Radiation Oncology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

[2]Department of Radiation Oncology, LAC+USC Medical Center, Los Angeles, California, USA

[3]Keck School of Medicine, University of Southern California, Los Angeles, California, USA

[4]Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

Abstract

Purposes: Preimplant diagnostic magnetic resonance imaging is the gold standard for image-guided tandem-and-ovoids (T&O) brachytherapy for cervical cancer. However, high dose rate brachytherapy planning is typically done on postimplant CT-based high-risk clinical target volume (HR-CTV$_{CT}$) because the transfer of preimplant Magnetic resonance (MR)-based HR-CTV (HR-CTV$_{MR}$) to the postimplant planning CT is difficult due to anatomical changes caused by applicator insertion, vaginal packing, and the filling status of the bladder and rectum. This study aims to train a dual-path convolutional neural network (CNN) for automatic segmentation of HR-CTV$_{CT}$ on postimplant planning CT with guidance from preimplant diagnostic MR.

Methods: Preimplant T2-weighted MR and postimplant CT images for 65 (48 for training, eight for validation, and nine for testing) patients were retrospectively solicited from our institutional database. MR was aligned to the corresponding CT using rigid registration. HR-CTV$_{CT}$ and HR-CTV$_{MR}$ were manually contoured on CT and MR by an experienced radiation oncologist. All images were then resampled to a spatial resolution of $0.5 \times 0.5 \times 1.25$ mm. A dual-path 3D asymmetric CNN architecture with two encoding paths was built to extract CT and MR image features. The MR was masked by HR-CTV$_{MR}$ contour while the entire CT volume was included. The network put an asymmetric weighting of 18:6 for CT: MR. Voxel-based dice similarity coefficient (DSC$_V$), sensitivity, precision, and 95% Hausdorff distance (95-HD) were used to evaluate model performance. Cross-validation was performed to assess model stability. The study

Correspondence: Wensha Yang, Department of Radiation Oncology, Keck School of Medicine, University of Southern California, 1441 Eastlake Ave., Los Angeles, CA 90033, USA. Wensha.Yang@med.usc.edu.
[#]Yufeng Cao and April Vassantachart contributed equally as the first author.

cohort was divided into a small tumor group (<20 cc), medium tumor group (20–40 cc), and large tumor group (>40 cc) based on the HR-CTV$_{CT}$ for model evaluation. Single-path CNN models were trained with the same parameters as those in dual-path models.

Results: For this patient cohort, the dual-path CNN model improved each of our objective findings, including DSC$_V$, sensitivity, and precision, with an average improvement of 8%, 7%, and 12%, respectively. The 95-HD was improved by an average of 1.65 mm compared to the single-path model with only CT images as input. In addition, the area under the curve for different networks was 0.86 (dual-path with CT and MR) and 0.80 (single-path with CT), respectively. The dual-path CNN model with asymmetric weighting achieved the best performance with DSC$_V$ of 0.65 ± 0.03 (0.61–0.70), 0.79 ± 0.02 (0.74–0.85), and 0.75 ± 0.04 (0.68–0.79) for small, medium, and large group. 95-HD were 7.34 (5.35–10.45) mm, 5.48 (3.21–8.43) mm, and 6.21 (5.34–9.32) mm for the three size groups, respectively.

Conclusions: An asymmetric CNN model with two encoding paths from preimplant MR (masked by HR-CTV$_{MR}$) and postimplant CT images was successfully developed for automatic segmentation of HR-CTV$_{CT}$ for T&O brachytherapy patients.

Keywords

CNN; Deep learning; HDR; high-risk CTV; segmentation; Tandem and Ovoids

1 | INTRODUCTION

The American Brachytherapy Society and Groupe European Curietherapie-European Society of Therapeutic Radiation Oncology (GEC-ESTRO) Gynecology (GYN) working group supports brachytherapy after external beam radiotherapy for patients with locally advanced cervical cancers (FIGO stages IB-IVA stage number) as it improves local control and survival rates.[1–3] There are various applicators available for brachytherapy treatment of cervical cancer. Tandem-and-ovoids (T&O) is one commonly used applicator for patients with barrel-shaped cervixes.[3,4] The appropriate selection of the applicator depends on the patient's clinical history (e.g., prior hysterectomy) and anatomy.[3] The European studies on magnetic resonance imaging (MRI)-guided brachytherapy in locally advanced cervical cancer (EMBRACE) demonstrated that 98% of local failures were located within the high-risk clinical target volume (HR-CTV) and the intermediate-risk CTV (IR-CTV).[3] A correlation between local control and target volume was demonstrated in the retroEMBRACE data that supports adaptive brachytherapy planning with target dose escalation. With the advent of three-dimensional imaging modalities, such as MRI or computed tomography (CT), the concept of image-guided adaptive brachytherapy has been developed and implemented by the GEC-ESTRO gynecology working group in the EMBRACE studies. Compared to CT, the MR scan can detect tumor regression during radiation therapy, which allows for adaptive brachytherapy plans and ensures sufficient target dose coverage without overdosing the organs at risk (OAR).[5–7] The results of the EMBRACE studies confirmed the safety, feasibility, advantages of MRI-based treatment planning, clinical outcomes, and late toxicities supporting the implementation of MRI-based brachytherapy.[3]

Currently, the application of MRI for each applicator implantation is still limited. Major drawbacks include limited availability of MRI scanners, time needed for MRI scanning, and additional patient transportation and setup time leading to increased uncertainty in the applicator position, and older MRI noncompatible applicators. In cases where MRI is not available in the subsequent fractions of brachytherapy, the MRI-based target contours can be reused through image fusions in the process of contouring on CT. Uncertainties in image and registration quality result in inconsistency in contouring at different fractions of brachytherapy or from different clinicians. Moreover, the fusion of pretreatment MRI to the CT for planning is challenging due to anatomical changes caused by applicator insertion, vaginal packing, and physiologic changes of the bowels and bladder. Therefore, we seek to train a dual-path convolutional neural network (CNN) utilizing the preimplant diagnostic MR and postimplant CT for automatic segmentation of HR-CTV.

The development of deep learning has pushed the limits of what is possible in the domain of medical image processing, particularly in image registration, detection, segmentation, regression, and classification.[8–11] Meanwhile, improved performance has been reported on solving a large variety of tasks in radiation oncology, such as treatment planning, contouring, organ segmentation, quality improvement, and treatment response.[12–14] Specifically, CNN has achieved remarkable success in 2D and 3D medical image segmentation,[15–17] most of which are for normal organs. A few studies reported automatic segmentation of CTV. For the head and neck, Cardenas et al. built a deep auto-encoders model to auto-delineate HR-CTV.[18] Zhang et al. used a UNet to auto-segment HR-CTV on CT images alone for T&O patients.[19] However, a potential limitation of this study is that MR was not used. Without MR, reliable manual segmentation of the HR-CTV is less than ideal in our experience due to the poor soft-tissue contrast, tumor visibility, and applicator-caused artifacts in the CT images. Dyer et al. used a deformable image registration algorithm to aid HR-CTV contours by involving preimplant MR and postimplant CT, with the need to manually segment both target and OAR, including cervix, uterus, bladder, and rectum. The contour-guided deformable image registration achieved a relatively low voxel-based DSC of 0.61.[20] In the current study, we aim to first understand the inter- and intraoperator HR-CTV contour consistencies for T&O brachytherapy patients, which serve as a reasonable upper bound to evaluate automated segmentation algorithms. We then developed a novel dual-path CNN to incorporate both preimplant diagnostic MR and postimplant planning CT for the automated segmentation of the HR-CTV, with a layer-level fusion of symmetric weightings to improve performance. Moreover, an asymmetric learning architecture from multi-modal MR images was built by using a different number of filters for different paths to improve the results. Model performance is compared with contouring uncertainties existing in the real clinical setting.

2 | METHODS AND MATERIALS

2.1 | Patient data

Under an institutional review board approved protocol, we retrospectively solicited 65 T&O patients from our institutional database from 2017 to 2021. Preimplant T2-weighted MR and postimplant CT images (typically done one week after the MR scan) were included

for each patient. In addition, a second postimplant CT two weeks after the MR scan was also included to quantify intra- and interoperator uncertainties. The CT-based HR-CTV (HR-CTV$_{CT}$) was manually delineated on the CT after reviewing the tumor measurements on the diagnostic MR images as part of the standard clinical workflow. For the study, the HR-CTV$_{CT}$ was verified by a single primary radiation oncologist (PRO), who was the brachytherapy fellow trained in our institution with 5 years of experience. The MR-based HR-CTV (HR-CTV$_{MR}$) were manually contoured by a medical resident and then verified by the same PRO. We divided the patient cohort into small (<20 cc), medium (20–40 cc), and large volume group (>40 cc) based on the HR-CTV$_{CT}$ of the first postimplant CT.

2.2 | Image acquisition and preprocessing

Postimplant CT images were performed on GE medical systems, with original volumetric dimensions of $512 \times 512 \times 220$, voxel spacing of $0.5 \times 0.5 \times 1.25$ mm^3, kVp of 120, and mAs of 300. Preimplant T2-weighted MR images were acquired on GE medical systems, with original volumetric dimensions of $256 \times 256 \times 42$ and a voxel spacing of $1.25 \times 1.25 \times 5$ mm^3, TP of (2316–5422) ms, echo time of (109–114) ms, percent phase field of view of 100, and flip angle of 180°.

Digital Imaging and Communications in Medicine (DICOM) files containing CT and MR images were exported to VelocityAI (Varian, Palo Alto, CA). The MR scans were then rigidly registered to the CT coordinates using the femur heads, sacrum, and coccyx as landmarks. Figure 1 compares the location and volume of preimplant MR-based HR-CTV$_{MR}$ and postimplant CT-based HR-CTV$_{CT}$ for an example patient. The voxel intensities of CT and MR images were normalized to be between 0 and 1. To manage the data size in training, we resampled the registered CT and MR images to $128 \times 128 \times 80$ voxels by nearest-neighbor interpolation. All images have a final spatial resolution of $0.5 \times 0.5 \times 1.25$ mm^3. The model training and testing were performed using a graphics processing unit (GPU) workstation equipped with 4x RTX 2080 Ti and a total of 44 Gigabyte (GB) graphic memory.

2.3 | Cross-validation procedure

Final model performance was assessed using eight-fold cross-validation, in which each fold consisted of randomly selected 48 subjects for training, eight for validation, and nine for testing.

2.4 | Network architecture

As shown in Figure 2, a dual-path 3D asymmetric CNN architecture with two encoding paths was built for image features of CT and MR. The MR was masked by HR-CTV$_{MR}$ contour while the entire CT volume was included. The network employed N$_{CT}$ filters on the CT path and N$_{MR}$ filters on the MR path, where N$_{CT}$ and N$_{MR}$ are two tunable parameters to control the relative weighting of the imaging modalities. In this study, we fixed N$_{CT}$ = 18 and increased N$_{MR}$ from 2 to 18 with an increment of 2. Inside each encoding path, the corresponding kernel convolution was applied twice with a rectified linear unit, a dropout layer between layers with a dropout rate of 0.4, a shortcut with kernel $1 \times 1 \times 1$, and a $2 \times 2 \times 2$ max-pooling operation in each layer.[21,22] The number of feature channels was

doubled after the max-pooling operation. In the upsampling path, each layer consisted of an upconvolution kernel of $2 \times 2 \times 2$, followed by two $3 \times 3 \times 3$ convolution kernels to halve the feature channels. We then concatenated the corresponding channels in the two paths. In the final step, $1 \times 1 \times 1$ convolution and soft-max were used to map the feature vectors to binary classes. The loss function used cross-entropy, which was defined as

$$E = - \sum_x y_k(x) \log(p_k(x)) + (1 - y_k(x)) \log(1 - p_k(x))$$ (1)

$y_k(x)$ was the true label at pixel position x, which was 0 or 1 for this binary case. $p_k(x)$ was the pixel-wise soft-max, which was given by

$$p_k(x) = \frac{\exp(f_k(x))}{\sum_{k'}^C \exp(f_{k'}(x))}$$ (2)

$f_k(x)$ was the activation function in feature channel k at the pixel position x. C was the number of classes. Glorot (Xavier) normal initializer was used for this dual-path CNN,[23] which drew samples from a truncated normal distribution centered on zero with $\sigma = \sqrt{\frac{2}{fan_{in} + fan_{out}}}$. fan_{in} and fan_{out} were the numbers of input and output units, in the weighting tensor, respectively. Adam optimizer was used to train this model.[24] We tested learning rates ranging from 1×10^{-4} to 1×10^{-7}. As a result, a learning rate of 6×10^{-5} with 1000 epochs was chosen to train this model.

The dual-path CNN model was compared with two scenarios using a single path in CNN, respectively. In the first scenario, the single path only used the CT as input; in the second scenario, the CT and MR were concatenated into a single matrix for input. The same network parameters, loss function, and training methods were used for comparisons. We define the following acronyms for the four models studied: SP_{CT}, SP_{CTMR}, DP_{SW}, and DP_{AW} stand for single-path with input of CT, single-path with combined inputs of CT and MR, dual-path with symmetric weighting, and dual-path with asymmetric weighting, respectively.

2.5 | Intra- and interoperator uncertainty estimation

To benchmark the performance of our dual-path CNN model, we studied intra- and interoperator uncertainties. For the intraoperator uncertainty, 10 patients were randomly selected from each volume group for a total of 30 patients. Postimplant CTs at the first and second fractions were contoured by the PRO as part of the standard clinical workflow. For the interoperator uncertainty, four patients were randomly selected from each volume group for a total of 12 patients. The HR-CTV$_{CT}$ from these 12 patients was independently contoured by two additional radiation oncologists (RO1 and RO2), who occasionally treat brachytherapy patients. HR-CTV$_{CT}$ from PRO, RO1 and RO2 were used to calculate intra- and interoperator dice similarity coefficients (DSC$_{intra}$ and DSC$_{inter}$) as follows:

$$DSC_{intra} = \frac{2 * (HR-CTV_{week1} \cap HR-CTV_{week2})}{HR-CTV_{week1} + HR-CTV_{week2}} \tag{3}$$

$$DSC_{inter} = \frac{2 * (HR-CTV_{PRO} \cap HR-CTV_{RO1or2})}{HR-CTV_{PRO} + HR-CTV_{RO1or2}} \tag{4}$$

Additionally, to evaluate the dosimetric uncertainties of the auto-segmented volumes, we compared the dose to the 90% (D90%) of the HR-CTV to the PRO plan of the same 12 patients. To compare these findings with intraoperator variability, rigid registration of PRO contours from the second fraction (PRO2) was used on the first fraction with alignment of the tandem. Interoperator D90% was also determined on the first fraction using RO1 and RO2 contours.

2.6 | Evaluation

The performance of the developed automated segmentation model was similarly determined using the DSC_V, sensitivity, precision, and Hausdorff distance (HD).

$$DSC_V = \frac{2 \times TP_V}{2 \times TP_V + FP_V + FN_V} \tag{5}$$

where TP_V is the number of voxels correctly detected, FN_V is the number of voxels not detected, and FP_V is the number of voxels falsely detected.

Voxel-based sensitivity (S_V) and precision (P_V) are the similarity measure often used in medical image processing to evaluate the performance of the segmentation algorithm that has a predefined ground truth.[25] S_V and P_V are defined in Equations (6) and (7), respectively.

$$S_V = \frac{TP_V}{TP_V + FN_V} \tag{6}$$

$$P_V = \frac{TP_V}{TP_V + FP_V} \tag{7}$$

HD is the maximum distance between a boundary point of auto-segmentation A and the nearest boundary point of manual contour B:

$$HD(A, B) = \max(h(A, B), h(B, A)) \tag{8}$$

where

$$h(A, B) = max_{a \in A} min_{b \in B} |a - b| \tag{9}$$

|·| denotes the Euclidean distance. To mitigate the sensitivity of HD to point outliers, we used 95-HD as the 95%-percentile HD between set A and B.

2.7 | Statistical testing

Due to the relatively small sample size, statistical analyses were performed by taking $p < 0.05$ as statistically significant, calculated by an independent two-sample t-test. In addition, mean ± standard deviation (SD), 95% confidence interval, and the median (minimum, maximum) were used to summarize the results. Following a significant one-way Analysis of Variance (ANOVA) result, Bonferroni correction was performed on DSC_V results, comparing four models, and including six comparisons between any two models. The original alpha level (0.05) was adjusted to 0.0083 (0.05/6) with Bonferroni correction, which means the test is significant if the p-value is <0.0083. Additionally, the relative SD (RSD) and two-sided Wilcoxon signed rank test were calculated for dosimetric evaluation.

3 | RESULTS

Figure 3a shows the volume distribution of HR-CTV$_{CT}$ in CT and HR-CTV$_{MR}$ in MR. The 65 patients are divided based on HR-CTV$_{CT}$ into small (<20 cc, $n = 13$), medium (20–40 cc, $n = 39$), and large (>40cc, $n = 13$) groups, respectively. As shown in Figure 3b, the mean, median, and SD of volumes for HR-CTV$_{MR}$ (HR-CTV$_{CT}$) are 38.0 (31.7) cc, 32.8 (28.2) cc, and 19.2 (16.7) cc. The volume difference between HR-CTV$_{MR}$ and HR-CTV$_{CT}$ is not statistically significant ($p = 0.06$).

3.1 | Intra- and interoperator uncertainty analysis

From the 30 patients' week one and week two CTs (total of 60 CTs), the intraoperator DSC_{intra} was calculated for the PRO. The DSC_{intra} were 0.67 ± 0.03 (0.62–0.75), 0.75 ± 0.06 (0.66–0.88), and 0.73 ± 0.08 (0.57–0.88), for the small, medium, and large volume groups, respectively. From the 12 patients' week one and week two CTs (total of 24 CTs), the interoperator DSC_{inter} was calculated for RO1 versus PRO and RO2 versus PRO. The DSC_{inter} were 0.64 ± 0.10 (0.49–0.83), 0.69 ± 0.08 (0.55–0.79), and 0.68 ± 0.09 (0.54–0.84) for the three volume groups.

3.2 | Model performance and comparison among different models

Figure 4a shows the four trained models' receiver operating characteristic curves. Figure 4b shows the loss function convergence for the training data. In addition, Figure 4c shows loss function convergence for training and validation of the dual-path model. The dual-path model with asymmetric weighting clearly showed the best diagnostic ability among all four models from Figure 4a. The average results from the eight-fold cross-validation were reported for each model. The mean area under the curve values are 0.86 (0.85–0.87), 0.83 (0.81–0.84), 0.82 (0.81–0.83), 0.80 (0.78–0.81) for dual-path with asymmetric weighting, dual-path with symmetric weighting, single-path with input CT+MR, and single-path with input CT, respectively. Compared to single-path with CT-only input, all other three models performed significantly better, $p < 0.01$.

Resulted DSC_v values from the four models were compared. After Bonferroni correction, the DP_{AW} resulted in significantly greater DSC than DP_{SW}, SP_{CTMR}, and SP_{CT}, $p <$ 0.0083. DP_{SW} also resulted in significantly greater DSC than SP_{CT}, although it was not significantly different from SP_{CTMR}, emphasizing the importance of asymmetric weightings. For single path models, the addition of MRs did not significantly improve the DSC. The Sv is significantly higher in the DP_{AW} model than DP_{SW} and SP_{CT} ($p < 0.0083$), although it is not significantly different from SP_{CTMR} ($p = 0.04$). DP_{AW} also resulted in significantly greater Pv than DP_{SW}, SP_{CTMR}, and SP_{CT}, $p < 0.0083$. DP_{SW} and SP_{CTMR} both resulted in significantly higher Pv than SP_{CT}, emphasizing the importance of MR data. The P_V from SP_{CTMR} is not significantly different from DP_{SW}, with $p = 0.023$, again emphasizing the importance of asymmetric weightings. 95-HD from DP_{AW} is significantly smaller than SP_{CT} and SP_{CTMR}, yet not significantly different from DP_{SW} with $p = 0.021$.

Table 1 shows the detailed model validation results by different volume groups. Single-path CNN models with CT-only input and CT + MR (masked by HR-CTV$_{MR}$) input were trained with the same learning parameters as those used in dual-path models. For the single-path CNN with only CT input, the results of DSCv, Sv, Pv, and 95-HD were 0.68 (0.52–0.81), 0.74 (0.65–0.82), 0.72 (0.63–0.83), and 7.65 (4.57–14.62) mm, respectively. The DSCv, Sv, Pv, and 95-HD improved to 0.71 (0.54–81), 0.79 (0.70–0.84), 0.76 (0.67–0.84), and 7.16 (4.26–13.45) mm when training the single-path CNN with both CT and MR inputs. Moreover, to explore the role of HR-CTV$_{MR}$, dual-path CNN models were constructed with two encoding paths based on CT and MR with varying filter weightings. Compared to the single-path model, the dual-path model with the symmetric weighting of 18:18 for CT: MR improved DSCv, Pv, and 95-HD to 0.72 (0.54–0.82), 0.78 (0.65–0.83), and 6.73 (4.61–10.55) mm, respectively. However, there was a minor reduction of Sv to 0.76 (0.68–0.83). Among all dual-path models with varying CT: MR weightings, the best performing model was the dual-path with an optimized asymmetric weighting of 18:6 (CT: MR), which achieved the highest DSCv, Sv, and Pv of 0.76 (0.61–0.85), 0.81 (0.72–0.86), and 0.83 (0.71–0.90). The 95-HD was also decreased to 5.99 (3.67–10.45) mm.

3.3 | Model comparison with the intra- and interoperator uncertainties

Figure 5 shows the model comparison with the intra- and interoperator uncertainties for the 12 randomly selected patients. The dual-path asymmetric model yielded a favorable contouring result with a mean DSC_V (auto-segmentation vs. PRO) of 0.74 ± 0.06. This result is superior to interoperator uncertainties for RO1 versus PRO ($DSC_{inter} = 0.68 \pm 0.11$) and RO2 versus PRO ($DSC_{inter} = 0.69 \pm 0.07$), as well as the mean intraoperator agreement of PRO ($DSC_{intra} = 0.74 \pm 0.07$). However, the differences were statistically insignificant with p-values of 0.14, 0.08, and 0.72, respectively.

Figure 6 shows a representative example comparing manual contours from RO1, RO2, PRO (ground truth), and the auto-contour from the dual-path model with the optimized asymmetric weighting. The volume of HR-CTV$_{CT}$ is 43 cc. For this representative case, our model achieved $DSC_V = 0.77$, between the radiation oncologist manual interoperator variations ($DSC_{inter} = 0.84$ for RO1 and $DSC_{inter} = 0.63$ for RO2).

Comparison of intra- and interoperator D90% to our model of the 12 patients is shown in Table 2. The median RSD% of our model compared to PRO was 7.0. This was between the interoperator RO1 and RO2 RSD% with medians of 7.6 and 5.7, respectively. However, the RSD% evaluating intraoperator D90% was lower with a median of 2.8. Wilcoxon signed rank test showed no statistically significant difference between the D90% for the intraoperator comparison and the auto segmented versus PRO dosimetric evaluation. There were statistically significant differences in both the RO1 and RO2 D90% compared to PRO.

4 | DISCUSSION

Deep learning networks have become a state-of-the-art method for automated multiorgan segmentation on CT images.[26–28] Nevertheless, for pelvis anatomy, the blurry boundaries and low soft-tissue contrast could reduce the segmentation accuracy.[26] To address this issue, Wang et al. developed a multistage segmentation framework consisting of an organ localization model to extract the segmentation region of each organ.[26] Tong et al. proposed a multi-task edge-recalibrated network to adaptively enhance its segmentation performance by extracting the edge-related features during training.[16] Meanwhile, very little work has been done to auto-segment the HR-CTV for high dose rate brachytherapy, which was considerably more challenging due to the lack of visible anatomical edges in CT.[18,19] GEC-ESTRO recommends performing the "pre-exam" MR scan for tumor size and anatomical evaluation, and applicator selection. With the applicator in situ, the "pretreatment" MR scan is recommended for contouring and treatment planning at each implantation of the applicator.[29,30] The multiplanar T2-weighted MR scan of less than 5-mm slice thickness acquired with the pelvic surface coils is considered as the gold standard for visualization and contouring of the tumor and OAR.[30] The use of complementary MRI sequences (e.g., contrast-enhanced T1-weighted MR or 3D isotropic MRI sequences) is optional.[30] Without the applicator in situ, MRI acquired before brachytherapy treatment is used to improve contouring of the HR-CTV on subsequent postimplant CT images.[31,32] Therefore, HR-CTV delineation should benefit from features of preimplant MR as well as the postimplant CT. Nevertheless, it is challenging to incorporate preimplant MR imaging information into the HR-CTV delineation workflow due to large variabilities between HR-CTV$_{CT}$ and HR-CTV$_{MR}$ in the location, shape, and size. Dyer et al. used the preimplant MR to aid target contours through deformable image registration, which achieved a relatively low DSC$_V$ of 0.61.[20] We avoided the unreliable MR-CT deformable registration by training a dual-path deep learning network to synthesize both the CT and MR information. To the best of our knowledge, this is the first report of auto HR-CTV segmentation for T&O patients using deep learning imaging features extracted from both postimplant CT and preimplant MR.

Motivated by the success of asymmetric learning from two different kernels from one input source,[33] in this study, we employed dual-path for CT and MR inputs, respectively, to allow separate control of the filters, channels, depths, and kernel sizes. Specifically, we studied the asymmetric features learned from CT and MR controlled by the relative number of filters and determined the optimal ratio of 18:6 for CT versus MR. The higher contribution from the CT can be explained by the fact that the CT is in the treatment geometry and contains more directly relevant information than the preimplant MR. Yet, the nonzero weighting of MR indicates a non-negligible contribution of MR to the segmentation performance. The

statistical testing results on DSC, S_V, P_V, and 95-HD comparing different models also help understand the importance of MR contribution and asymmetric weighting. The comparisons between SP_{CT} and SP_{CTMR} showed insignificant differences in DSC and HD-95 using Bonferroni corrected p-value. Yet, the sensitivity and specificity values are significantly different, indicating that the MR dataset's contribution is less on the location or volume of the CTVs and more on the tumor identification and detection. Using standard $p < 0.05$ as a significance threshold, DP_{AW} achieved significantly better results on all metrics than DP_{SW}. After Bonferroni correction, only HD-95 became statistically insignificant. This particular result supported the importance of asymmetric weighting.

It is worth noting that the HR-CTV volume influenced our model's segmentation performance. The medium group presented the best DSC_V and 95-HD values with both single and dual path models. This could be attributed to the nonuniform dataset distribution. As shown in Figure 3, the medium group comprised more samples than the small and large CTV groups. The results indicate space for further improvement with a larger dataset enriched in CTVs toward both tail ends of the size distribution.

A unique contribution of the study is that we compared the automated segmentation performance with both inter- and intraoperator variations, as shown in Figure 5. The interoperator variation indicates the potential variation between radiation oncologists who are specialized in T&O brachytherapy and who are not. This variation reflects the real-world uncertainties in treating T&O patients at a low-volume clinic versus a high-volume clinic. As shown in the results, after model training, our automated segmentation method performed better than the nonspecialists and closer to the specialist. Therefore, the automated segmentation method, after further validation, will be valuable to help low volume clinics to improve their consistency and quality of defining the HR-CTV for T&O treatments. Dosimetric evaluation also showed RSD% was not statistically different between the intraoperator and auto-segmentation versus PRO D90%. This further supports that auto-segmentation may allow for improved consistency compared to nonspecialist practitioners. While the DSC for auto-segmentation and RO1 was higher than RO2, the RSD% was lower for RO2. This can be explained by the fact that the mean volume of the evaluated contours was smaller for RO2 (21.6 cc, range 11–31.2) compared to PRO, PRO1, RO1, and the model – 27.5 cc (12.8–57.3), 25.8 cc (14.5–59.2), 33 cc (18.6–65.4), and 27.5 cc (19.6–60.9), respectively. Therefore, although the contours may not be as accurate when compared to PRO, D90% can be higher due to a smaller volume needing coverage by the plan. The RSD% seen in our study is also comparable to a prior study evaluating dosimetric variability with MRI-based contours.[34]

The study is limited by the patient sample size and the number of radiation oncologists who have repeated the HR-CTV segmentation. Larger patient size and broader sampling of manual segmentation from a variety of clinics will improve the representation of HR-CTV segmentation.

5 | CONCLUSION

A 3D asymmetric CNN model with two encoding paths from preimplant MR and postimplant CT was successfully developed for automatic segmentation of HR-CTV for T&O brachytherapy patients.

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

1. Pötter R, Tanderup K, Kirisits C, et al. The EMBRACE II study: the outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the EMBRACE studies. Clin Transl Radiat Oncol. 2018;9:48–60. [PubMed: 29594251]

2. Pötter R, Georg P, Dimopoulos JC, et al. Clinical outcome of protocol based image (MRI) guided adaptive brachytherapy combined with 3D conformal radiotherapy with or without chemotherapy in patients with locally advanced cervical cancer. Radiother Oncol. 2011;100:116–123. [PubMed: 21821305]

3. Viswanathan AN, Thomadsen B. Committee ABSCCR. American Brachytherapy Society consensus guidelines for locally advanced carcinoma of the cervix. Part I: general principles. Brachytherapy. 2012;11:33–46. [PubMed: 22265436]

4. Nomden CN, de Leeuw AA, Roesink JM, et al. Clinical outcome and dosimetric parameters of chemo-radiation including MRI guided adaptive brachytherapy with tandem-ovoid applicators for cervical cancer patients:a single institution experience. Radiother Oncol. 2013;107:69–74. [PubMed: 23639580]

5. Haie-Meder C, Pötter R, Van Limbergen E, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group☆(I): concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV. Radiother Oncol.2005;74:235–245. [PubMed: 15763303]

6. Hricak H, Gatsonis C, Coakley FV, et al. Early invasive cervical cancer: CT and MR imaging in preoperative evaluation—ACRIN/GOG comparative study of diagnostic performance and interobserver variability. Radiology. 2007;245:491–498. [PubMed: 17940305]

7. Mitchell DG, Snyder B, Coakley F, et al. Early invasive cervical cancer: tumor delineation by magnetic resonance imaging, computed tomography, and clinical examination, verified by pathologic results, in the ACRIN 6651/GOG 183 intergroup study. J Clin Oncol. 2006;24:5687–5694. [PubMed: 17179104]

8. Balakrishnan G, Zhao A, Sabuncu MR, et al. An unsupervised learning model for deformable medical image registration. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; June 18–23, 2018; Salt Lake City, UT.

9. Milletari F, Navab N, Ahmadi S-A. V-net: fully convolutional neural networks for volumetric medical image segmentation. Paper presented at:2016 Fourth International Conference on 3D Vision (3DV); October 25–28, 2016; Stanford, CA.

10. Zhu J-Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. Paper presented at: Proceedings of the IEEE International Conference on Computer Vision; October 22–29, 2017; Venice, Italy.

11. Kumar A, Kim J, Lyndon D, et al. An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE J Biomed Health Inform. 2016;21:31–40. [PubMed: 28114041]

12. Fan J, Wang J, Chen Z, et al. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Med Phys. 2019;46:370–381. [PubMed: 30383300]

13. Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. Clin Cancer Res. 2019;25:3266–3275. [PubMed: 31010833]

14. Gibson E, Giganti F, Hu Y, et al. Automatic multi-organ segmentation on abdominal CT with dense v-networks. IEEE Trans Med Imaging. 2018;37:1822–1834. [PubMed: 29994628]

15. Moeskops P, Wolterink JM, van der Velden BH, et al. Deep learning for multi-task medical image segmentation in multiple modalities. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2016:478–486.

16. Tong N, Gou S, Chen S, et al. Multi-task edge-recalibrated network for male pelvic multi-organ segmentation on CT images. Phys Med Biol. 2021;66:035001. [PubMed: 33197901]

17. Chen Y, Ruan D, Xiao J, et al. Fully automated multi-organ segmentation in abdominal magnetic resonance imaging with deep neural networks. Med Phys. 2020;47:4971. [PubMed: 32748401]

18. Cardenas CE, McCarroll RE, Court LE, et al. Deep learning algorithm for auto-delineation of high-risk oropharyngeal clinical target volumes with built-in dice similarity coefficient parameter optimization function. Int J Radiat Oncol Biol Phys. 2018;101: 468–478. [PubMed: 29559291]

19. Zhang D, Yang Z, Jiang S, et al. Automatic segmentation and applicator reconstruction for CT-based brachytherapy of cervical cancer using 3D convolutional neural networks. J Appl Clin Med Phys. 2020;21:158–169.

20. Dyer BA, Yuan Z, Qiu J, et al. Clinical feasibility of MR-assisted CT-based cervical brachytherapy using MR-to-CT deformable image registration. Brachytherapy. 2020;19:447–456. [PubMed: 32327343]

21. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout:a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1929–1958.

22. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition;June 27–30, 2016; Las Vegas, NV.

23. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Paper presented at:Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; PMLR 9:249–256, 2010.

24. Kingma DP, Adam BaJ. A method for stochastic optimization. ArXiv14126980 Cs. 2017. https://arxiv.org/pdf/1412.6980.pdf

25. Fenster A, Chiu B. Evaluation of segmentation algorithms for medical imaging. Conf Proc IEEE Eng Med Biol Soc. 2005;2005:7186–7189. [PubMed: 17281935]

26. Wang S, He K, Nie D, et al. CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation. Med Image Anal. 2019;54:168–178. [PubMed: 30928830]

27. Hatamizadeh A, Terzopoulos D, Myronenko A. Edge-gated CNNs for volumetric semantic segmentation of medical images. ArXiv Prepr ArXiv200204207. 2020. https://arxiv.org/pdf/2002.04207.pdf

28. Tong N, Gou S, Yang S, et al. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Med Phys. 2018;45:4558–4567. [PubMed: 30136285]

29. Hellebust TP, Kirisits C, Berger D, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group: considerations and pitfalls in commissioning and applicator reconstruction in 3D image-based treatment planning of cervix cancer brachytherapy. Radiother Oncol. 2010;96:153–160. [PubMed: 20663578]

30. Dimopoulos JC, Petrow P, Tanderup K, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (IV): basic principles and parameters for MR imaging within the frame of image based adaptive cervix cancer brachytherapy. Radiother Oncol. 2012;103:113–122. [PubMed: 22296748]

31. Tanderup K, Nesvacil N, Pötter R, et al. Uncertainties in image guided adaptive cervix cancer brachytherapy:impact on planning and prescription. Radiother Oncol. 2013;107:1–5. [PubMed: 23541642]

32. RU I. Prescribing, recording, and reporting brachytherapy for cancer of the cervix. J ICRU. 2013;13:2.

33. Cao Y, Vassantachart A, Jason CY, et al. Automatic detection and segmentation of multiple brain metastases on magnetic resonance image using asymmetric UNet architecture. Phys Med Biol. 2021;66:015003. [PubMed: 33186927]

34. Hellebust TP, Tanderup K, Lervåg C, et al. Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2013;107:13–19.
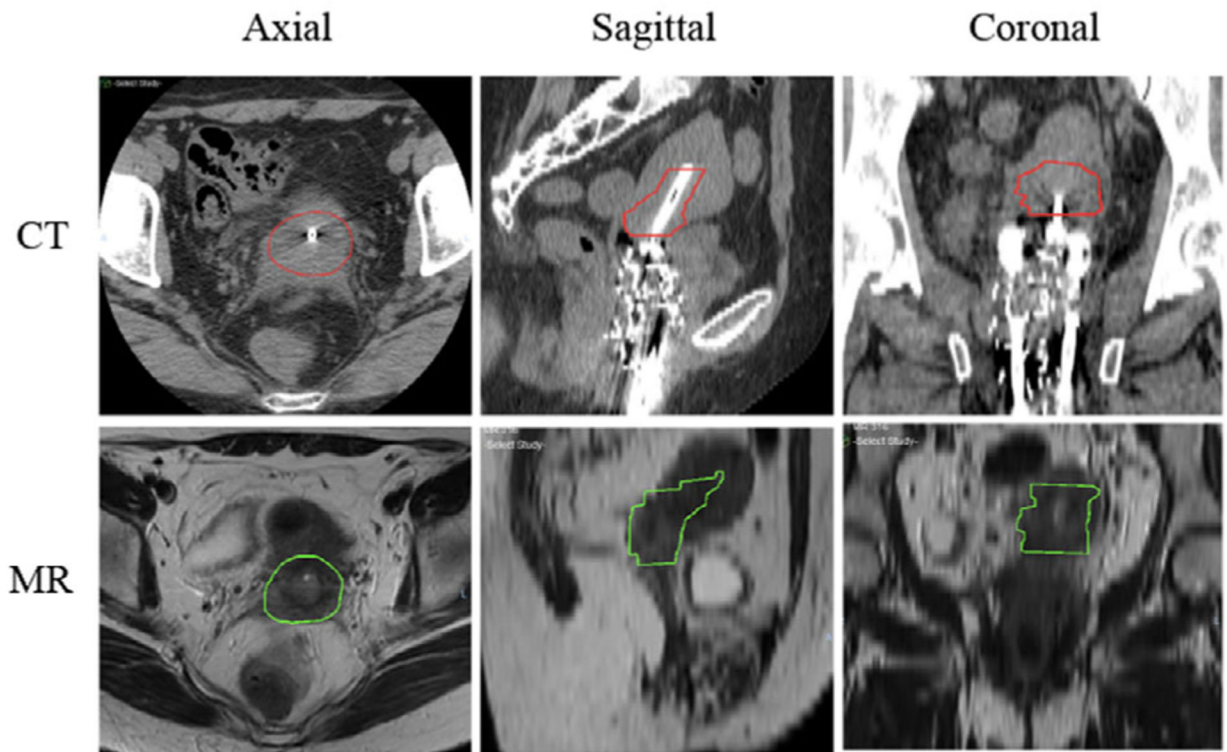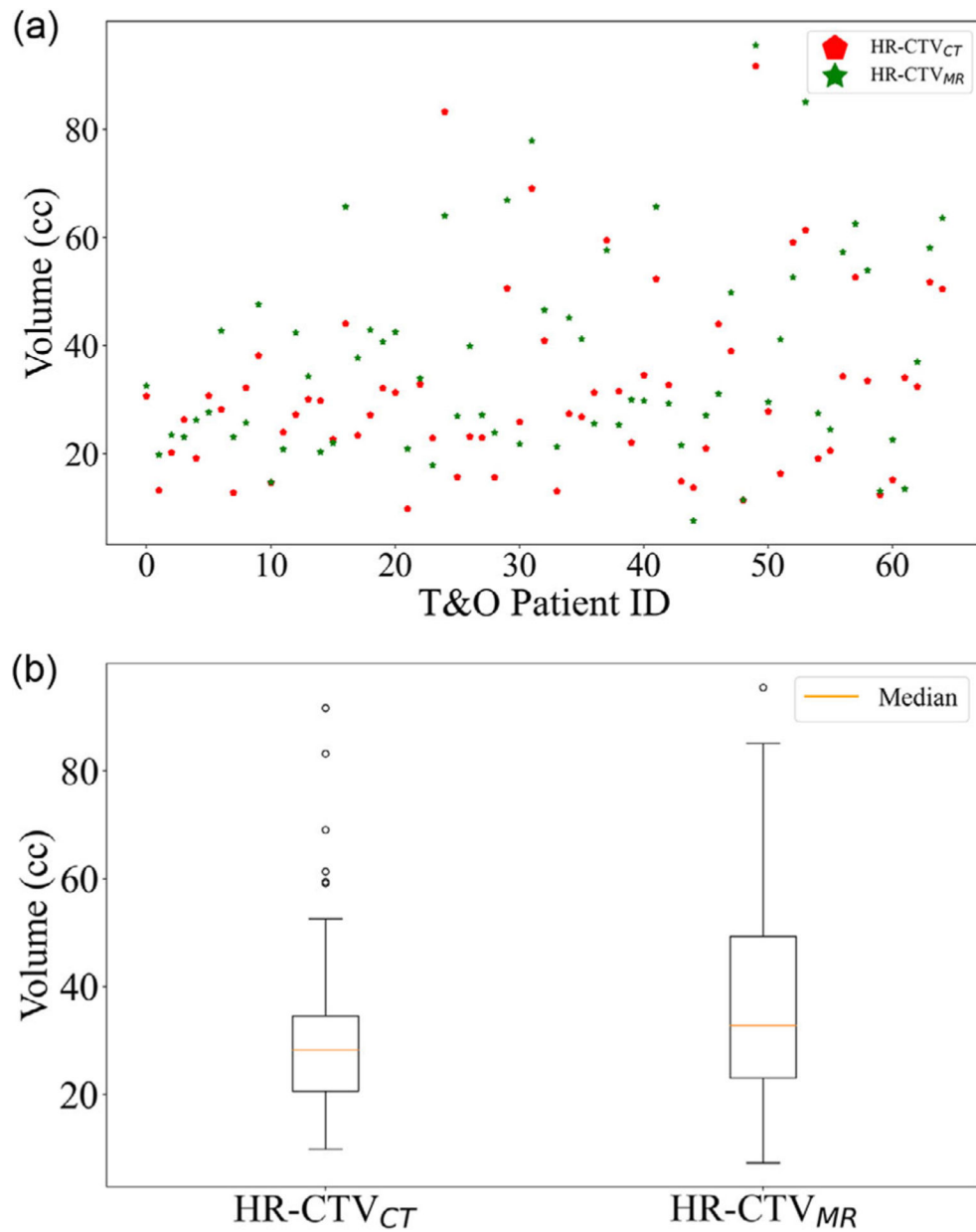
FIGURE 1.
Axial, sagittal, and coronal views of computed tomography (CT) and MR for a representative patient, which show the comparison of location and size of preimplant MR-based high-risk clinical target volume (HR-CTV$_{MR}$) and postimplant planning-CT-based HR-CTV$_{CT}$
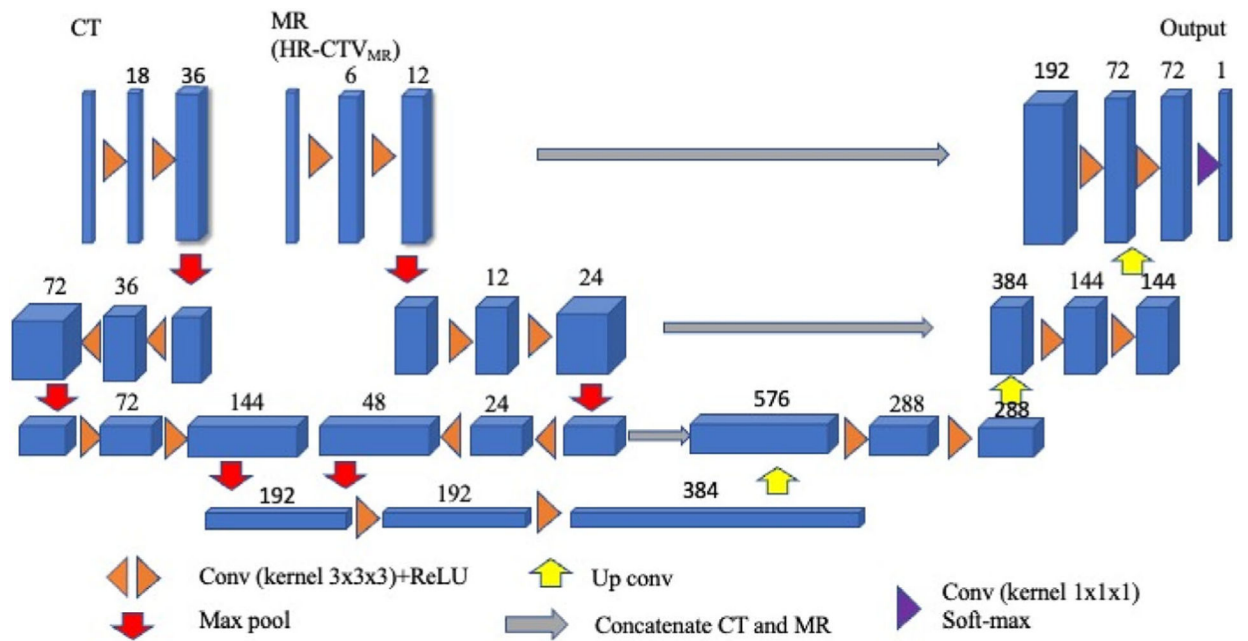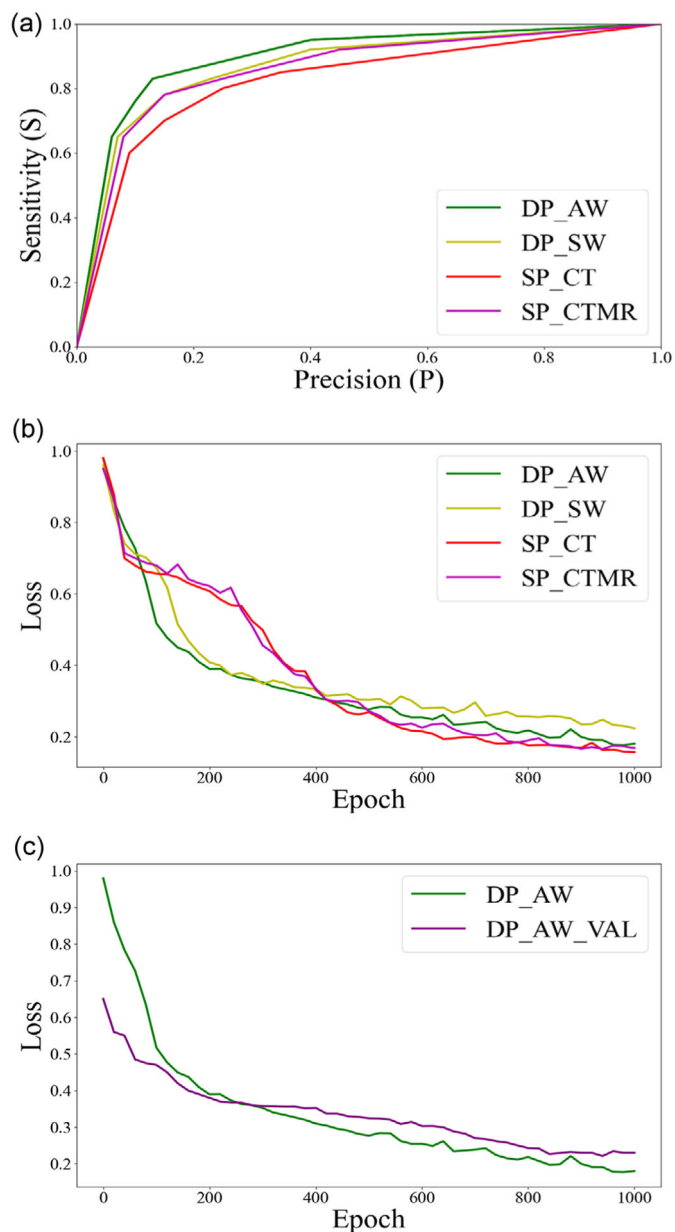
FIGURE 2.
The architecture of the asymmetric convolutional neural network (CNN) model. Each blue cuboid corresponds to a feature map. The number of channels is denoted on the top of the cuboid
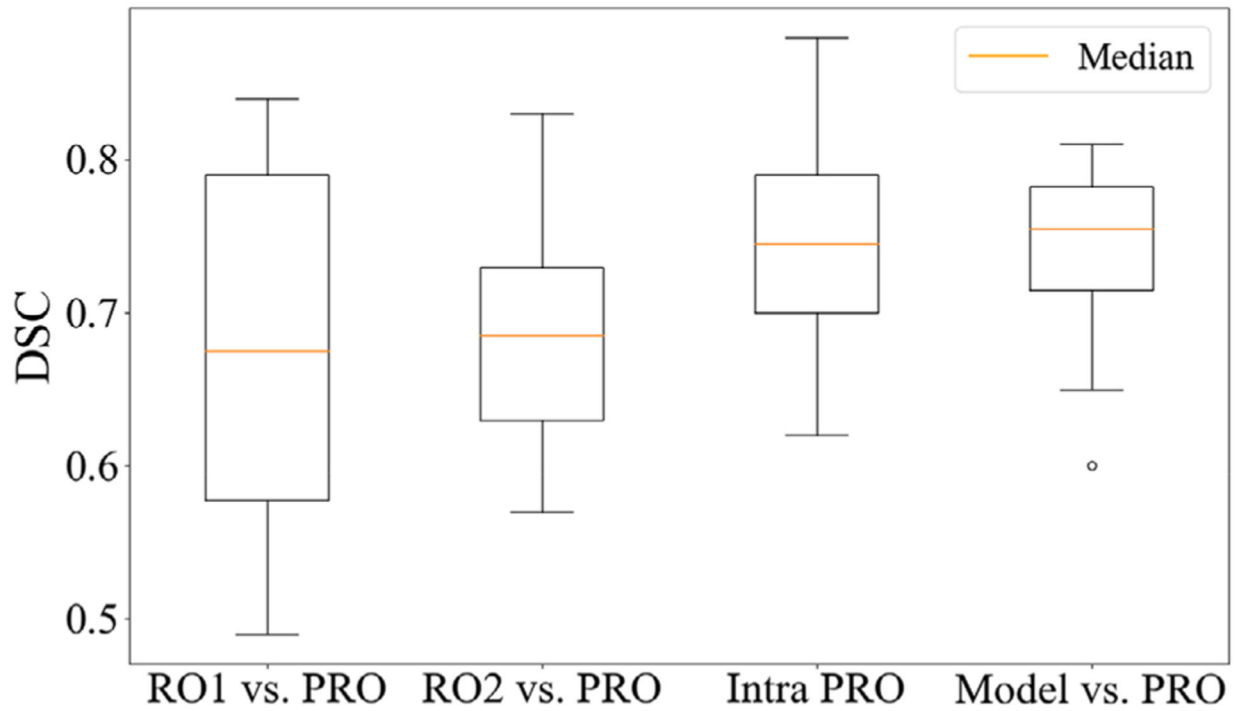
FIGURE 3.

(a) The volume distribution of high-risk clinical target volume (HR-CTV$_{CT}$) and HR-CTV$_{MR}$. (b) Boxplots of HR-CTV$_{CT}$ in computed tomography (CT) and HR-CTV$_{MR}$ in MR. The difference in volumes is not statistically significant ($p = 0.06$)
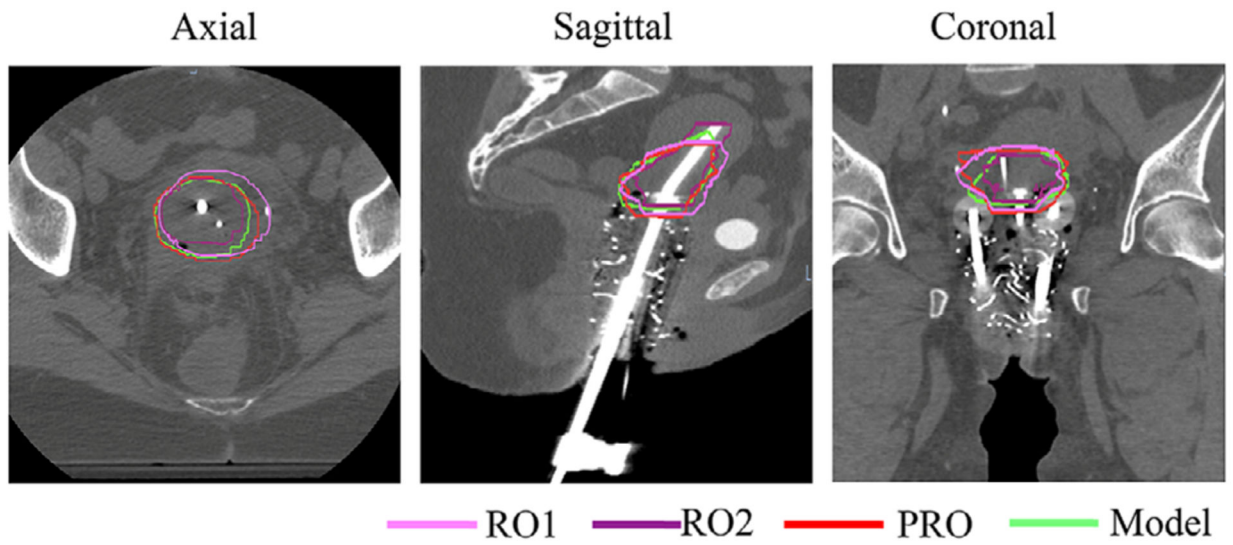
FIGURE 4.
(a) Receiver operating characteristic (ROC) curves from the four models, including dual-path with asymmetric weighting (DP_AW), dual-path with symmetric weighting (DP_SW), single-path with input computed tomography (CT) + MR (SP_CTMR), and single-path with input CT (SP_CT). (b) The objective loss as a function of the epoch for the four convolutional neural network (CNN) models. (c) The objective loss as a function of epoch for training and validation (DP_AW_VAL) cohorts for dual path with asymmetric weighting

FIGURE 5.
Comparison of two additional radiation oncologists (RO1 and RO2), the primary brachytherapy radiation oncologist (PRO), and our model

FIGURE 6.

Comparison of manual PRO (red), RO1 (pink), RO2 (purple), and automatic dual-path
(green) contours in axial, sagittal, and coronal views for a representative patient

TABLE 1

Summary of segmentation performance using different networks

Model type	Volume group	DSC$_V$ Mean ± SD (minimum, maximum) 95% CI	S$_V$ Mean ± SD (minimum, maximum) 95% CI	P$_V$ Mean ± SD (minimum, maximum) 95% CI	95-HD (mm) Mean ± SD (minimum, maximum) 95% CI
SP$_{CT}$	Small	0.58 ± 0.04 (0.52, 0.65) (0.56, 0.60)	0.71 ± 0.03 (0.65, 0.75) (0.69, 0.73)	0.68 ± 0.03 (0.63, 0.72) (0.66, 0.70)	9.34 ± 1.62 (7.68, 14.62) (8.46, 10.22)
	Medium	0.72 ± 0.03 (0.66, 0.81) (0.71, 0.73)	0.75 ± 0.02 (0.72, 0.82) (0.74, 0.76)	0.74 ± 0.02 (0.71, 0.83) (0.73, 0.75)	6.78 ± 1.25 (4.57, 9.68) (6.39, 7.17)
	Large	0.67 ± 0.05 (0.62, 0.74) (0.64, 0.70)	0.72 ± 0.03 (0.69, 0.76) (0.70, 0.74)	0.73 ± 0.02 (0.69, 0.77) (0.72, 0.74)	8.76 ± 1.88 (5.26, 13.31) (7.74, 9.78)
	All	**0.68 ± 0.12 (0.52, 0.81) (0.65, 0.71)**	**0.74 ± 0.06 (0.65, 0.82) (0.73, 0.75)**	**0.72 ± 0.07 (0.63, 0.83) (0.70, 0.74)**	**7.65 ± 2.86 (4.57, 14.62) (6.95, 8.35)**
SP$_{CTMR}$	Small	0.58 ± 0.04 (0.54, 0.65) (0.56, 0.60)	0.72 ± 0.02 (0.70, 0.76) (0.71, 0.73)	0.70 ± 0.02 (0.67, 0.72) (0.69, 0.71)	9.21 ± 1.58 (7.35, 13.45) (8.35, 10.07)
	Medium	0.75 ± 0.03 (0.69, 0.81) (0.74, 0.76)	0.81 ± 0.02 (0.78, 0.84) (0.80, 0.82)	0.78 ± 0.02 (0.76, 0.84) (0.77, 0.79)	6.32 ± 1.19 (4.26, 9.67) (5.95, 6.69)
	Large	0.71 ± 0.03 (0.68, 0.77) (0.69, 0.73)	0.78 ± 0.03 (0.74, 0.81) (0.76, 0.80)	0.75 ± 0.02 (0.73, 0.81) (0.74, 0.76)	7.67 ± 1.43 (6.14, 11.37) (6.89, 8.45)
	All	**0.71 ± 0.09 (0.54, 0.81) (0.69, 0.73)**	**0.79 ± 0.05 (0.70, 0.84) (0.78, 0.80)**	**0.76 ± 0.06 (0.67, 0.84) (0.75, 0.77)**	**7.16 ± 2.67 (4.26, 13.45) (6.51, 7.81)**
DP$_{SW}$	Small	0.60 ± 0.04 (0.54, 0.66) (0.58, 0.62)	0.72 ± 0.03 (0.68, 0.76) (0.70, 0.74)	0.68 ± 0.02 (0.65, 0.71) (0.67, 0.69)	7.67 ± 1.24 (6.31, 10.55) (6.99, 8.34)
	Medium	0.75 ± 0.02	0.78 ± 0.02	0.81 ± 0.02	6.21 ± 1.14

Model type	Volume group	DSC$_V$ Mean ± SD (minimum, maximum) 95% CI	S$_V$ Mean ± SD (minimum, maximum) 95% CI	P$_V$ Mean ± SD (minimum, maximum) 95% CI	95-HD (mm) Mean ± SD (minimum, maximum) 95% CI
		(0.71, 0.82) (0.74, 0.76)	(0.75, 0.81) (0.77, 0.79)	(0.78, 0.83) (0.80, 0.82)	(4.61, 9.23) (5.85, 6.57)
	Large	0.73 ± 0.04 (0.66, 0.77) (0.71, 0.75)	0.75 ± 0.02 (0.73, 0.79) (0.74, 0.76)	0.79 ± 0.02 (0.77, 0.82) (0.78, 0.80)	7.45 ± 1.28 (5.76, 10.31) (5.06, 6.46)
	All	**0.72 ± 0.08 (0.54, 0.82) (0.70, 0.74)**	**0.76 ± 0.05 (0.68, 0.83) (0.75, 0.77)**	**0.78 ± 0.05 (0.65, 0.83) (0.77, 0.79)**	**6.73 ± 2.06 (4.61, 10.55) (6.23, 7.23)**
DP$_{aw}$	Small	0.65 ± 0.03 (0.61, 0.70) (0.63, 0.67)	0.76 ± 0.02 (0.72, 0.78) (0.75, 0.77)	0.74 ± 0.02 (0.71, 0.77) (0.73, 0.75)	7.34 ± 1.19 (5.35, 10.45) (6.69, 7.99)
	Medium	0.79 ± 0.02 (0.74, 0.85) (0.78, 0.80)	0.83 ± 0.02 (0.79, 0.86) (0.82, 0.84)	0.86 ± 0.02 (0.83, 0.90) (0.85, 0.87)	5.48 ± 1.01 (3.67, 8.43) (5.16, 5.79)
	Large	0.75 ± 0.04 (0.68, 0.79) (0.73, 0.77)	0.78 ± 0.02 (0.76, 0.83) (0.77, 0.79)	0.81 ± 0.02 (0.79, 0.85) (0.80, 0.82)	6.21 ± 1.15 (4.76, 9.32) (5.58, 6.83)
	All	**0.76 ± 0.06 (0.61, 0.85) (0.75, 0.77)**	**0.81 ± 0.04 (0.72, 0.86) (0.80, 0.82)**	**0.83 ± 0.04 (0.71, 0.90) (0.82, 0.84)**	**5.99 ± 1.68 (3.67, 10.45) (5.58, 6.39)**

Note: SPCT, SPCTMR, DPSW, and DPAW stand for single-path with input of CT, single-path with combined inputs of CT and MR, dual-path with symmetric weighting, and dual-path with asymmetric weighting, respectively.

Abbreviations: 95-HD, 95% Hausdorff distance; CI, confidence interval; DSC$_V$, voxel-based dice similarity coefficient; SD, standard deviation.

TABLE 2

Auto-segmentation dosimetric evaluation compared to intra- and interoperator variability

	D90% median (range)	PRO versus PRO2 (RSD%)	PRO versus RO1 (RSD%)	PRO versus RO2 (RSD%)	PRO versus model (RSD%)
1	8.9 (7.8–11.8)	3.4	3.5	17.4	4.4
2	8.5 (8.2–10.1)	1.6	1.0	7.3	8.6
3	8.0 (7.0–9.7)	1.3	6.4	9.5	4.1
4	7.3 (5.0–9.3)	12.6	28.3	2.1	9.9
5	8.5 (7.1–9.9)	2.2	8.7	7.9	5.6
6	9.4 (7.3–10.5)	5.4	12.5	4.1	7.2
7	8.3 (7.3–9.5)	13.1	13.1	1.7	6.8
8	8.1 (7.6–11.9)	1.8	3.4	18.9	0.3
9	7.2 (5.7–8.3)	7.3	4.7	2.4	15.8
10	8.3 (5.7–8.6)	0.4	18.9	1.6	4.8
11	8.1 (6.5–8.2)	5.1	0.2	0.7	11.1
12	7.4 (5.5–9.1)	1.5	17.9	10.0	24.8
	Median (range)	2.8 (0.4–13.1)	7.6 (0.2–28.3)	5.7 (0.7–18.9)	7.0 (0.3–24.8)
	p-Value	0.07	0.002	0.034	0.126

Abbreviations: PRO, primary radiation oncologist; RSD, relative standard deviation.