Advances in Radiation Oncology (2025) **10**, 101745

advances in radiation oncology

www.advancesradonc.org

Scientific Article

Domain-Adaptive and Per-Fraction Guided Deep Learning Framework for Magnetic Resonance Imaging-Based Segmentation of Organs at Risk in Gynecologic Cancers

Reza Kalantar, PhD,[a,b] Manasi Ingle, FRCR,[a,b] Romelie Rieu, BA, BmBCh,[a,b] Sebastian Curcean, MD,[b,c] Jessica Mary Winfield, PhD,[a,b] Gigin Lin, MD, PhD,[d] Christina Messiou, MD, MRCP, FRCR,[a,b] Susan Lalondrelle, FRCR,[a,b] Dow-Mu Koh, MD, FRCP, FRCR,[a,b,*] and Matthew David Blackledge, PhD[a,b]

[a]The Institute of Cancer Research, London, United Kingdom; [b]The Royal Marsden Hospital, London, United Kingdom; [c]Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania; and [d]Chang Gung Memorial Hospital, Linkou and Chang Gung University, Taoyuan, Taiwan

Received 1 July 2024; accepted 10 February 2025

Purpose: The integration of magnetic resonance imaging into radiation therapy (RT) treatment necessitates automated segmentation algorithms for fast and accurate adaptive interventions, particularly in magnetic resonance imaging-integrated linear accelerator (MR-linac or MRL) treatment systems. However, the scarcity of data hampers the training of these models. This study aimed to address this shortcoming by developing a synthetic MRL-assisted deep learning framework to establish a robust baseline for organ at risk segmentation on MRL images and enable domain adaptation for automatic delineations during adaptive RT treatments.

Methods and Materials: We used a retrospective data set, comprising 158 patients diagnosed with various gynecologic cancers who underwent computed tomography scanning for RT planning and 25 patients with T_2-weighted MRL scans for model fine-tuning, adaptation, and evaluation. A patch-based cycle-consistent generative adversarial network was developed to synthesize MRL images from computed tomography data. Subsequently, a domain-adaptive segmentation network was trained to segment the 6 organs at risk on acquired MRL images. In addition, we employed per-fraction adaptation to enhance anatomical conformity guided by prior treatment fractions of individual patients. A quantitative evaluation and blinded human reader assessment were conducted to establish contour acceptance rates.

Results: The synthetic MRL-assisted model improved organ at risk segmentation accuracy on MRL images, with fraction-adapted contours displaying high anatomical fidelity. Two radiation oncologists reported contour acceptance rates of 100% and 98% for treatment planning after adaptation.

Conclusions: This novel framework holds promise to bridge the semantic gap between computed tomography and magnetic resonance imaging databases, potentially facilitating adaptive RT treatments and reducing treatment times as well as clinician burden. The utility of this framework can extend beyond gynecologic and pelvic cancers.

Sources of support: This study represents independent research funded by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre and the Clinical Research Facility in Imaging at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London. The views expressed are those of the author(s) and not necessarily those of the NIHR or theDepartment of Health and Social Care.

Data sharing statement: The data are not available because of data privacy restrictions.

*Corresponding author: Dow-Mu Koh; Email: dow-mu.koh@icr.ac.uk

https://doi.org/10.1016/j.adro.2025.101745

Introduction

In radiation therapy (RT), the precise and accurate delineation of organs at risk (OARs) is crucial for minimizing radiation exposure to healthy tissue, reducing the likelihood of acute and late toxicity. Conventionally, RT treatment planning and dosimetry are based on computed tomography (CT). Magnetic resonance imaging (MRI) is increasingly employed to guide delineations for treatment planning, primarily because of its superior soft-tissue contrast compared with CT. For gynecologic cancers, MRI is used to accurately delineate the tumor target volumes [gross tumor volume, high- and low-risk clinical target volume [CTV], and OARs). However, manual delineation of these regions of interest is time-consuming and may result in inter- and intraoperator variabilities.[1,2]

Online adaptive RT (ART) is a recent advance in RT delivery,[3] incorporating repeated imaging and revision of the treatment plan throughout the course of therapy. This approach necessitates the regular and rapid definition of regions of interest, which serves as a bottleneck in the ART workflows. Previous studies for abdominal and pelvic treatments have reported that manual segmentation may take more than 30 minutes in some cases.[4] This underscores the need for automatic segmentation technologies to improve the accuracy, efficiency, and generalizability of OARs segmentation models in MRI-guided ART workflows. With the introduction of new MRI-integrated linear accelerator systems (MR-linac or MRL) for treatment delivery, there is a need to develop automated MRL-based segmentation algorithms for OARs, particularly in the context of gynecologic cancers where primary RT treatment is often employed.

Previous studies have shown a diverse range of techniques for OAR segmentation, including atlas-based techniques[5-7] and traditional machine learning[8,9] and deep learning techniques.[10-21] They have demonstrated that deep learning techniques, particularly convolutional neural networks (CNNs), have achieved superior performance in OAR segmentation tasks compared with other methods. Among the existing CNN-based methods, U-Net[22] and its variations are prominent architectures used in segmentation applications. In the context of OAR segmentation in the management of pelvic cancers, these models were employed to delineate regions of interest on CT for gynecologic cancers,[10,11,13,17] prostate cancer,[18,19] and rectal cancer.[20,21] Several studies have investigated the delineation of OARs on MRI for gynecologic cancers. For example, Yoganathan et al[23] used a residual neural network model (InceptionResNetv2) and an ensemble 2.5-dimensional approach to delineate OARs for MRI-based planning of high-dose-rate brachytherapy in cervical cancer. Breto et al[12] used the Mask region-Based convolutional neural network (Mask-RCNN) architecture for per-fraction prediction of the cervical gross tumor volume and OARs for MR-guided RT (MRgRT). The authors also investigated the inclusion of planning MRI for patient-specific segmentation fine-tuning but reported no overall improvement with this approach. Nevertheless, training these models on image data sets from a single disease may limit the diversity and amount of training data, potentially affecting both the accuracy and the generalizability of the models across other pelvic diseases.

In MRgRT studies, the shortage of annotated and high-quality data remains a considerable limitation for training deep learning algorithms.[24,25] Within the field of continuously evolving treatment technologies, for example, MRL,[26] there is a need for adaptive deep learning strategies that tackle these shortcomings and are accurate and can be generalized to different tumor types. One potential solution is domain adaptation and semantic knowledge transfer across domains. Specifically, we hypothesize that undertaking this approach to propagate anatomical knowledge of pelvic OARs from CT can improve the segmentation models for MRI applications, which is of particular importance with the advent of MRL technology. Previous research studies have reported promising results in using synthetic images to boost segmentation performance in other domains.[27-31] However, pelvic CT to MRI synthesis is challenging, and the use of only synthetic MRI images for multiorgan segmentation in MRgRT remains relatively unexplored.

In this study, we present a novel domain-adaptive deep learning framework, enhanced by synthetic T_2-weighted (T_2W) MRI pretraining and per-fraction adaptation for segmentation of 6 OARs on the MRL images, including the bladder, rectum, bowel, sigmoid, and left and right femoral heads. We evaluate our framework by comprehensive quantitative and qualitative analyses, determining the clinical acceptance rates of our baseline and self-adapted models for ART by blinded expert human reader assessment. The key contributions of our study include (1) the development of a versatile synthetic MRI-based OARs segmentation baseline model across pelvic cancers in female patients; (2) a 3-dimensional patch-based training strategy for image synthesis and segmentation for spatial consistency; and (3) a per-fraction guided adaptive technique for patient-specific conformity based on data from preceding treatment fractions.

Methods and Materials

Patient population and imaging protocols

This study included a retrospective cohort of 158 female patients (aged 21-88 years) diagnosed with various gynecologic cancers who underwent routine CT planning scans. The patients were diagnosed with tumor stages ranging from FIGO 1a to FIGO 4, including cases of primary, recurrent, and relapsed diseases. Cancer types

represented in the cohort included cervical, endometrial, vulvar, and vaginal cancers, as well as cases with ovarian involvement. CT images were acquired from multiple vendors, with in-plane resolutions between 1.0 × 1.0 mm^2 and 1.6 × 1.6 mm^2 and slice thicknesses from 2.0 to 2.5 mm. Standard-dose pretreatment protocols were followed for image acquisition, and contours from treatment sessions were directly used for this study.

In addition, a retrospective data set of 25 patients (ages 26-79 years) with gynecologic cancers who had received 3-dimensional T_2W MRI on an MRL system was identified. Among these, OAR contours from delivered treatments were available for 15 patients, with a median of 5 treatment fractions per patient. A board-certified radiation oncologist (M.I.) with >5 years of experience reviewed and, when necessary, revised these contours to ensure segmentation accuracy. The remaining 10 patients, each with 1 pretreatment and 1 posttreatment scan, were reserved for the image synthesis process.

The MRL images were acquired on an Elekta Unity MRL system (Elekta AB) equipped with a Philips Marlin 1.5T MRI scanner (Philips Healthcare). These images had voxel resolutions ranging from 0.5 × 0.5 × 0.6 to 0.8 × 0.8 × 1.0 mm^3, with echo and repetition times ranging from 82 to 278 and 1300 to 1535 ms, respectively. The study received approval from the institutional review board. Participants either provided verbal informed consent for the use of their imaging data across the cohort, or consent was waived as per the terms of the service evaluation across the CT cohort.

Framework overview

Our proposed domain-adaptive framework incorporates unsupervised MRL image synthesis from CT images, enabling the segmentation training of OARs using MRL images from patients diagnosed with gynecologic cancers. We developed a cycle-generative adversarial network[32] to perform patch-based synthesis of MRL images. The resulting trained weights were used to generate synthetic MRL images from CT, whereas OAR contours were propagated to make full use of the preexisting CT data. The annotated MRL images were employed for segmentation training and fine-tuning. Finally, a per-fraction adaptation technique was implemented to use data from previous treatment fractions and enhance the conformity of segmentations to the individual patient's anatomy for each subsequent treatment scan. A description of this technique is provided in section "Per-fraction segmentation adaptation." The flowchart of the proposed pipeline is shown in Fig. 1.

Image synthesis

In the image synthesis stage, the cycle-consistent generative adversarial network architecture was modified to accommodate the 3-dimensional input dimensions. The generator was designed with 2 downsampling blocks, 9 residual blocks, and 2 upsampling blocks, all constructed with 3-dimensional convolution, instance normalization, and rectified linear unit (ReLU) activation layers. For downsampling and upsampling, strided convolution and strided deconvolution operations were performed, respectively. To scale the model output to the range of (−1,1), a final Tanh activation layer was incorporated. The patch-GAN discriminator included 4 convolutional blocks, each comprising strided convolution, instance normalization, and leaky ReLU activation, followed by a final sigmoid activation layer. The generator, discriminator, and cycle losses were quantified as the mean absolute error (L_1), the mean squared error (L_2), and an equally weighted mean of L_1 and structural similarity index metric losses, respectively.

Before training, MRI and CT scans underwent intensity normalization to constrain image intensities within a standardized range of (−1 to 1). For MRI images, intensities were normalized using linear scaling based on values within the 95th percentile, corresponding to a range of 0 to 400, whereas for CT images, intensities between −1000 and 1000 Hounsfield units were retained for normalization. Intensity values exceeding this range after scaling were truncated to the nearest limit (−1 or 1), thereby preserving a consistent intensity distribution across all samples. To ensure alignment with the target MRL distribution, all training images were resampled to a spatial resolution of 0.8 × 0.8 × 1.0 mm^3.

The cycle-consistent generative adversarial network model was trained on patches of size 128 × 128 × 128 voxels using stochastic sampling from 20 pairs of T_2W MRI and CT images for a total of 5000 iterations. The Adam optimizer was used with the initial learning rate of $2e^{-4}$, linearly decayed to $5e^{-5}$. Finally, the trained generator weights were used to predict synthetic MRL images from the remaining 138 gynecologic planning CT images for segmentation training. These images were not used in the synthesis process. The inference was performed using a sliding window algorithm with 85% dense patch overlap and equal weight intensity averaging. All computations in this study were performed on an Nvidia RTX6000 GPU and using the TensorFlow, Keras, and Monai[33] Python libraries.

OAR segmentation framework

Our proposed segmentation model was an enhanced version of the U-Net architecture, with inclusion of residual units in its design. It consisted of encoding and decoding pathways, in which each level was connected by skip connections. It included 5 downsampling and upsampling blocks, using strided 3-dimensional convolutions and 3-dimensional deconvolution operations at the onset of each block (filter kernels = 64, 128, 256, 512, 1024). Each block consisted of 2 residual units, employing 3-

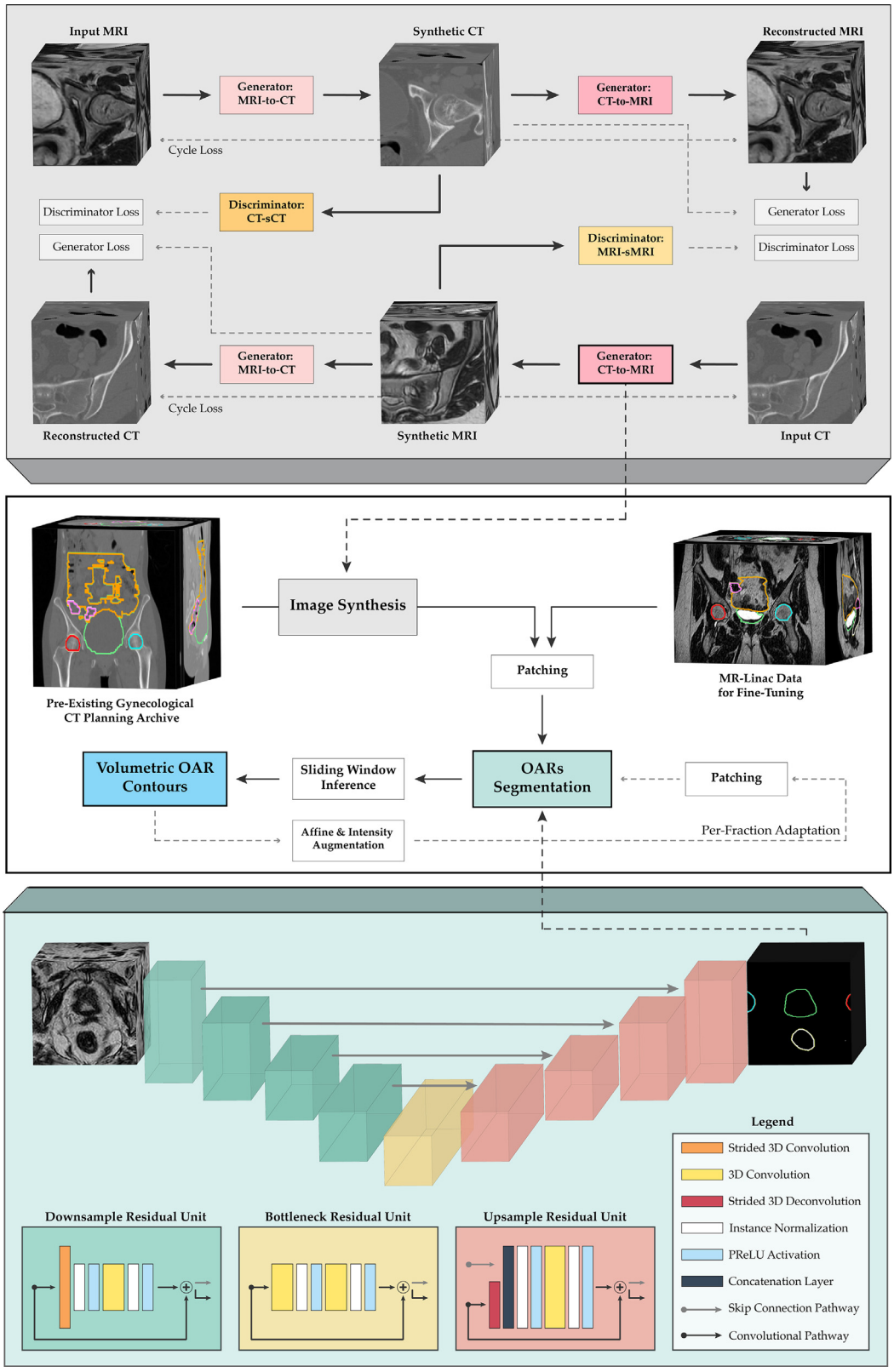

Figure 1 Illustration of the domain-adaptive framework that employs a patch-based cycle-consistent generative adversarial network for MRI synthesis from CT. The framework also features OAR segmentation knowledge propagation for enhanced contouring using a residual U-Net and a per-fraction adaptation technique on MRL images to ensure patient-specific conformity.
Abbreviations: CT = computed tomography; MRI = magnetic resonance imaging; MRL = magnetic resonance imaging-integrated linear accelerator; OAR = organ at risk.

dimensional convolution, followed by instance normalization and the parametric ReLU activation layer. The output of the model was passed through a Softmax activation layer, producing 7 output channels (Fig. 1). The following section provides the experimental configurations evaluated for the segmentation of OARs using model training from only the acquired MRL data as well as knowledge transfer from synthetic MRL images generated from planning CT images.

Experimental training of different segmentation networks

The OAR segmentation performance in our study was evaluated using the following 4 experimental scenarios:

- Scenario 1 (S1): The model was trained exclusively on acquired MRL data.
- Scenario 2 (S2): The model was trained only on synthetic MRL data.
- Scenario 3 (S3): The encoder weights from synthetic MRL training (S2) were frozen, and the decoder weights were fine-tuned on acquired MRL data. Weight freezing reduced the number of trainable parameters from 76.9 to 53.6 million in the model.
- Scenario 4 (S4): Similar to S3, the model was fine-tuned on acquired MRL data following encoder weight freezing from S2, with the exception that weighting was applied to the cross-entropy (CE) component of the loss function for some soft tissues (loss = Dice + CE). This step was implemented to improve the segmentation capabilities of the model for more targeted and more challenging OARs from the acquired MRL data. Weight allocation to each class was set as follows: background (1), bladder (1), small bowel (1), rectum (2), sigmoid (2), left femoral head (0), and right femoral head (0). The weights were arbitrarily selected to demonstrate the impact of custom weighting of organs on overall segmentation performance.

Before training, all MRL images were normalized consistent with techniques described in section "Image synthesis." The patch-based segmentation training for all scenarios was performed using patches of size $96 \times 96 \times 96$ voxels randomly derived from the training data and validated on patches of size $150 \times 150 \times 150$ using a sliding window algorithm to incorporate a larger field of view (75% overlap of adjacent patches). For all experiments, the backpropagation process used the AdamW optimizer[34] with a learning rate of $1e^{-4}$ and weight decay regularization of $1e^{-5}$. The hybrid loss, the sum of combined Dice and CE losses, was used to train the models. The best weights were selected based on the Dice similarity coefficient (DSC) metrics from the validation data. Table 1 shows the breakdown of these scenarios along with the data populations used for each experiment.

The segmentation volumes for the test data were generated using a sliding window inference algorithm with 75% overlap between adjacent patches and a mean inference time of <1 minute. Following this, the test segmentation results underwent quantitative evaluation against ground-truth contours, based on the whole image volumes. Three metrics were applied in this evaluation process: DSC, 95th percentile Hausdorff distance (HD), and

Table 1 Summary of experimental models for OARs segmentation on acquired MRL images trained with and/or without synthetic MRL-based pretraining and fraction adaptation

Scenario	Data set	Train/validation(/test) patients	Comments
S1	Acquired MRL	9/3/3 3-fold cross-validation	The validation folds were pseudo-randomly selected such that no test patients were duplicated across any fold.
S2	Synthetic MRL	113/25	Trained using only the synthetically generated MRL images from planning CT images.
S3	Synthetic MRL + acquired MRL	113/25 + 9/3/3 3-fold cross-validation	Pretrained on synthetic MRL images and the decoder fine-tuned on acquired MRL data.
S4	Synthetic MRL + acquired MRL	113/25 + 9/3/3 3-fold cross-validation	Pretrained on synthetic MRL images and the decoder fine-tuned on acquired MRL data with addition of soft-tissue custom weighting in hybrid loss.
Adapted S4	Synthetic MRL + acquired MRL	113/25 + 9/3/3 3-fold cross-validation + 1 (up to 4 fractions)	The baseline S4 model adapted to individual test patients based on preceding treatment fractions.

Abbreviations: CT = computed tomography; MRL = magnetic resonance imaging-integrated linear accelerator; OAR = organ at risk.
Data preprocessing, normalization, and training hyperparameters across all experiments remained consistent during this study. The S1-S4 models were trained for 100,000 iterations, validating at 500 iteration intervals. The adapted S4 model was trained for 500 intervals, validating at 10 iteration intervals.

mean surface distance. Statistical analysis was performed to compare the performance of all experimental models (the analysis of variance [ANOVA] and paired *t* tests).

Per-fraction segmentation adaptation

To use the preceding fractions to enhance the performance of the segmentation model for individual patients, an adaptive segmentation strategy was implemented. This approach was devised to fine-tune the OAR segmentations, ensuring that the predicted contours conformed to the anatomy of individual patients given the treatment fractions. To simulate the interfraction variations commonly observed in clinical settings, a series of random transformations were applied during training for data augmentation/perturbation. These techniques included 3-dimensional elastic deformations, range-based intensity scaling, and intensity shifts. Elastic deformations involve applying spatial transformations to the image volume using displacement fields. These displacement fields include Gaussian processes and B-spline grids that control how each voxel in the image is displaced, aiming to capture the local distortions and variations that might occur in the tissues between scans. In this experiment, random elastic transformations including Gaussian kernels with standard deviations ranging from 1 to 4, random grid offset between 5 and 15 in magnitude, and small random shearing were applied to accomplish this. Furthermore, small intensity scaling and shifting with probabilities of 25% were applied, to broaden the variability in a dynamic range of patches during training. The same images, without any augmentations, were used for validation. The trained weights from the baseline S4 model, the best performing experimental model, were selected to undertake these analyses (Table 1). Segmentation adaptation training was performed for 500 iterations, with validation intervals set at every 10 iterations. Throughout this process, the encoder weights were frozen and the AdamW optimizer was used, set to a learning rate of $1e^{-5}$.

To evaluate this technique for individual test patients, the relevant set of weights was selected from the validation folds of the trained S4 models. Given the available data, this experiment was confined to contours from up to 4 preceding treatment fractions for each test case. All other training hyperparameters from prior segmentation experiments were unchanged during the adaptation phase.

The quantitative results were computed and averaged across all unseen treatment fractions for test cases. The ANOVA and paired *t* test analyses were performed to determine whether changes in the segmentation of OARs using our adaptive strategy were statistically significant compared with the baseline S4 model without adaptation. In addition, these evaluations were executed for segmentation model adaptation based on fractions 1 to 4.

Human reader assessment

The evaluation of the results was performed independently by 2 radiation oncologists with 6 and 7 years of experience in RT, respectively. The readers were presented with 27 blinded sets of OAR contours. These included the ground truth, contours derived from the baseline automatic segmentation (S4 model), and a set of fraction-adapted contours each for 9 test cases. To minimize bias, no same fractions were presented to the readers. These cases were selected based on the training/test split of data used during the cross-validation of models in previous experiments.

Each set of contours presented to the readers was from a different treatment fraction, selected uniformly from fractions 1 to 5. The readers were given the volumetric contours for the OARs, superimposed on the corresponding acquired T_2W MRL images, with the treated CTV information also available. A categorical scoring system, based on the criteria defined in Table 2, was developed to facilitate the scoring of the contours. Finally, to quantify the level of agreement between the 2 oncologists regarding the clinical acceptance rates for the OARs, Cohen's kappa coefficient was calculated.

Results

OAR segmentation

Our results demonstrated that segmentation pretraining on synthetic MRL images and knowledge transfer after encoder weight freezing to our acquired MRL cohort (embodied in S3 and S4 models), improved the

Table 2 Oncologist reader assessment evaluation criteria for segmented OARs

Score	Grade	Criteria
3	Acceptable	The segmentation is acceptable for treatment without any revisions needed.
2	Minor revision	The segmentation is adequate for treatment but could be improved with minor anatomical revisions.
1	Major revision	The segmentation requires major revisions before it is suitable for treatment.
0	Rejection	The segmentation is unacceptable and must be completely redefined before treatment.

Abbreviation: OAR = organ at risk.
Each OAR was independently evaluated and assigned a score based on the grading criteria, reflecting its clinical acceptance rate.

segmentation of the OARs in DSC and HD metrics. In addition, incorporating soft tissue−specific CE weighting further boosted the segmentation performance of soft-tissue structures, such as small bowel, sigmoid, and rectum. It was observed that the sigmoid structure was the most challenging, exhibiting the largest variability across all quantitative metrics. Statistical analysis of these findings revealed that fine-tuning of the model in S3 and S4 led to significant improvements in the segmentation of small bowel and femoral heads across DSC and HD metrics. The ANOVA tests revealed significant differences in metric values for the left and right femoral heads. The segmentation results across all 4 experimental models are shown in Table 3 and Fig. 2.

Per-fraction segmentation adaptation

The comparative OAR segmentation contours across all experimental scenarios, as well as the fraction-adapted segmentations from 6 out of 9 test cases, are shown in Fig. 3. The segmentation results of synthetic MRL-assisted models exhibited superior performance to those trained only on a limited number of acquired MRL data. For instance, in test case 1, a collapsed bladder, undetected by the model trained exclusively on a limited number of acquired MRL images, was successfully detected in all synthetic MRL-assisted scenarios. Furthermore, the implementation of per-fraction adaptation resulted in a more precise delineation of the rectum and sigmoid, closely matching the contouring standard observed in the ground-truth labels (test cases 1-3, 5). Importantly, the segmentation contours following adaptation did not infringe on the tumor structure (gross tumor volume equivalent). This is pivotal for contour acceptability in treatments, as seen in Fig. 3.

The quantitative metrics after adaptation based on preceding treatment fractions revealed significant improvements in segmentation performance and a decrease in prediction variability for targeted soft tissues, particularly

Table 3 Comparison of the segmentation performance in all 4 scenarios

OARs	S1: MR-only training	S2: sMR-only training	S3: MR finetuning based on sMR training with encoder weight freeze	S4: MR finetuning based on sMR training with encoder weight freeze (CE weighting)
			DSC	
Bladder	0.806 ± 0.083	0.761 ± 0.142	0.813 ± 0.095*	0.802 ± 0.105
Small bowel	0.661 ± 0.096	0.696 ± 0.086	0.721 ± 0.067	0.733 ± 0.078*
Rectum	0.737 ± 0.102	0.761 ± 0.086	0.764 ± 0.081	0.765 ± 0.088*
Sigmoid	0.428 ± 0.209	0.519 ± 0.190	0.487 ± 0.175	0.544 ± 0.147*
Left femoral head	0.882 ± 0.031	0.861 ± 0.039	0.909 ± 0.026*	0.905 ± 0.013
Right femoral head	0.901 ± 0.032	0.883 ± 0.013	0.921 ± 0.019*	0.915 ± 0.011
			HD 95% (mm)	
Bladder	7.511 ± 2.785	9.68 ± 5.876	7.182 ± 5.838*	7.549 ± 4.450
Small bowel	26.517 ± 10.624	20.189 ± 5.059	18.470 ± 6.031*	19.544 ± 8.036
Rectum	19.41 ± 9.728	20.124 ± 9.654	17.442 ± 9.651*	17.535 ± 11.147
Sigmoid	73.256 ± 36.009	58.297 ± 31.514	66.306 ± 48.372	55.040 ± 44.819*
Left femoral head	10.496 ± 11.708	12.186 ± 2.807	5.215 ± 7.949	4.148 ± 1.279*
Right femoral head	12.234 ± 9.277	9.140 ± 1.980	3.667 ± 1.316	3.300 ± 0.913*
			MSD (mm)	
Bladder	2.163 ± 0.712*	3.561 ± 2.114	2.363 ± 1.117	2.620 ± 1.423
Small bowel	4.346 ± 0.957	4.265 ± 0.854*	4.312 ± 0.930	4.637 ± 1.209
Rectum	2.104 ± 0.403	2.188 ± 0.734	1.860 ± 0.402*	2.132 ± 0.506
Sigmoid	5.589 ± 2.892	4.754 ± 2.297*	5.599 ± 2.473	5.616 ± 2.380
Left femoral head	2.77 ± 2.491	2.999 ± 0.716	1.551 ± 1.092	1.416 ± 0.301*
Rright femoral head	2.623 ± 1.865	2.306 ± 0.350	1.260 ± 0.357	1.241 ± 0.334*

The quantitative Dice similarity coefficient (DSC), 95th percentile Hausdorff distance (HD), and mean surface distance (MSD) metrics are shown as mean and standard deviation, comparing the segmentation performance in all 4 scenarios (S1-S4). Values marked with * indicate the best performing scenario for each organ at risk (OAR).

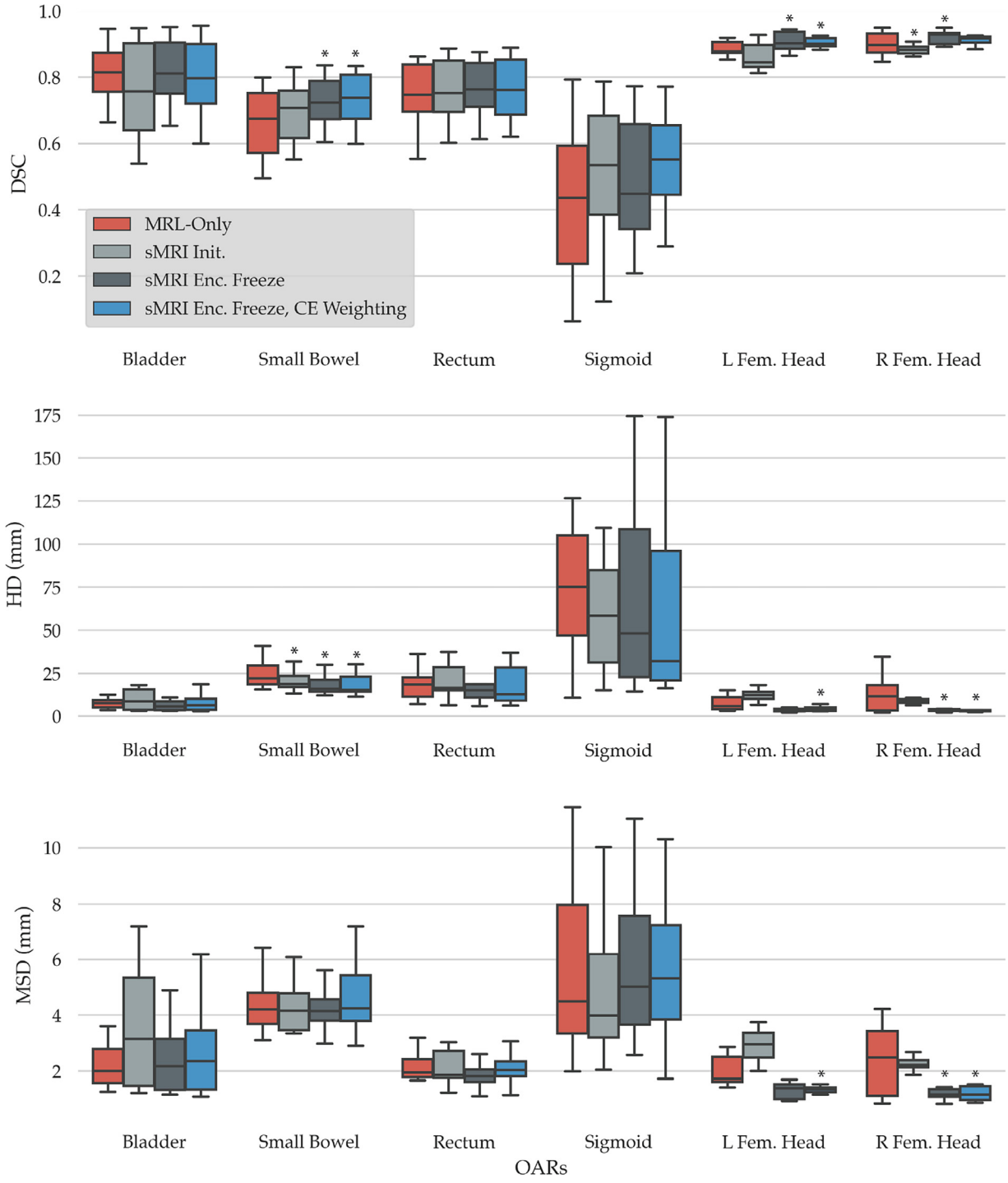

Figure 2 Boxplots of the Dice similarity coefficient (DSC), 95th percentile Hausdorff distance (HD), and mean surface distance (MSD) metrics for all OARs. The metrics were calculated using a 3-fold cross-validation approach, and the segmentation results were averaged over all MRL images from the test cases in each fold. The metrics with significant changes against the S1 model, trained only on acquired MRL data, are marked with a star symbol ($P < .05$).
Abbreviations: MRL = magnetic resonance imaging-integrated linear accelerator; OAR = organ at risk.

the rectum and sigmoid, as shown in the boxplots in Fig. 4. Notably, significant improvements were observed in segmentation of targeted OARs (small bowel, sigmoid, and rectum) after model adaptation from a single preceding treatment fraction. However, no significant changes were observed from adaptations using 1-4 previous treatment fractions. These findings were supported by the ANOVA results from the generated contours, displaying significant changes across all models for DSC and HD metrics for sigmoid and rectum.

Figure 3 Visual comparison of organs at risk (OARs) segmentation results across all scenarios and the adapted contours from 1 preceding fraction, illustrated in axial and sagittal planes for 6 test examples. Key improvements resulting from adaptation, such as enhanced anatomical conformity, are indicated by white arrows (improved collapsed bladder and sigmoid segmentations in examples 1 and 2, respectively) and improved sigmoid delineation (example 2). Notable discrepancies leading to possible contour rejection during treatment planning, such as missing or misidentified structures, are marked by yellow arrows (eg, unrepresented bladder in example 1, tumor misinterpreted as sigmoid in example 6).

Human reader assessment

The results of our blinded reader assessment are presented in Fig. 5. The analysis indicated unanimous agreement between both readers in identifying the ground-truth contours, exhibiting a 100% acceptance rate for treatment either with or without identifying the potential for minor revision (scores 2 and 3). However, we observed a discrepancy between the readers when deciding on structures that were scored as "minor revision" within these contour sets.

The baseline OAR segmentation model from the S4 model displayed a more consistent consensus between the readers. The labels indicating "rejection" or "major revision" (scores 0 and 1) were consistent for the bladder (1 out of 9; Cohen's kappa: 1.0), rectum (1 out of 9; Cohen's

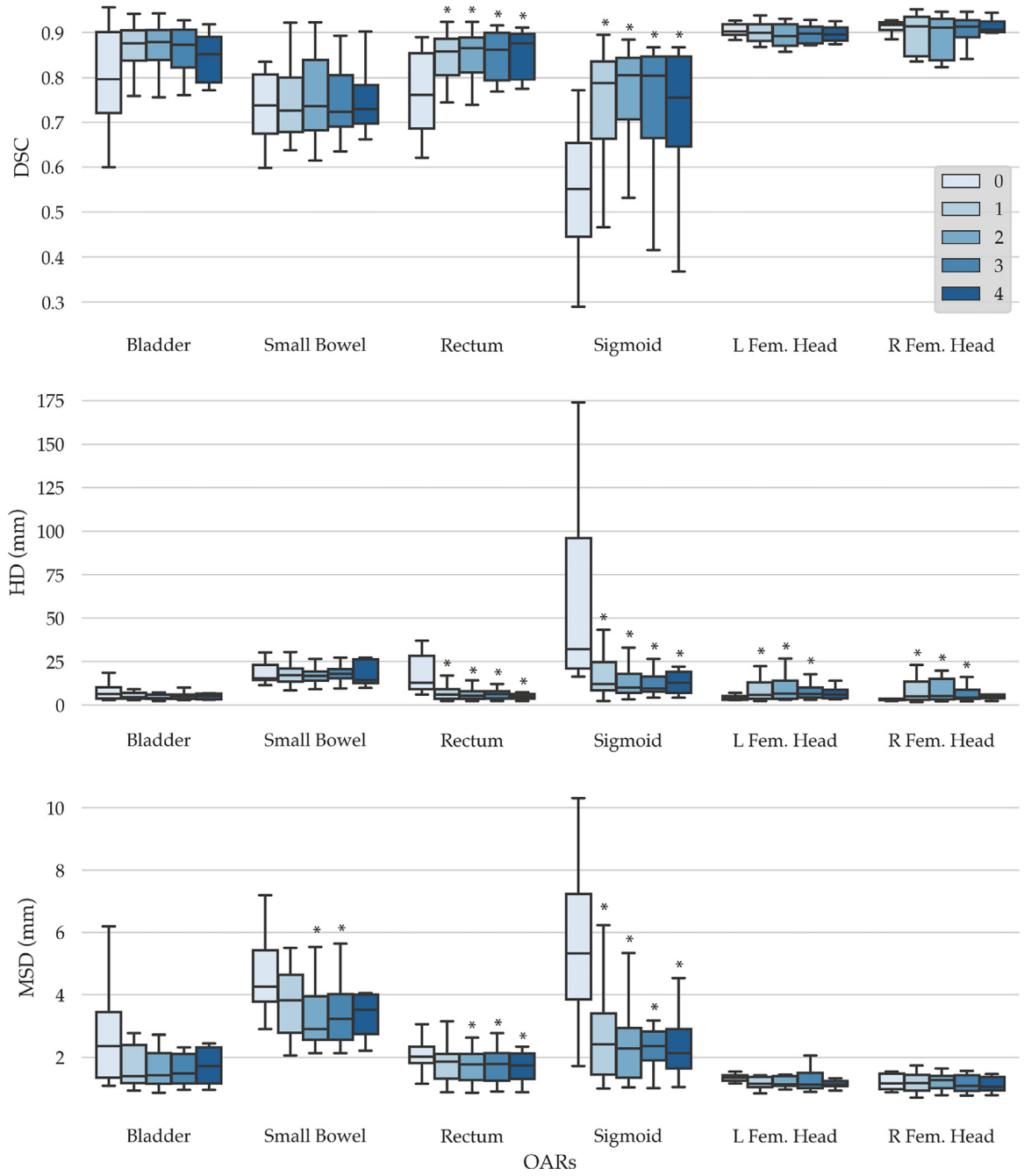

Figure 4 Boxplots illustrating a comparative analysis of the Dice similarity coefficient (DSC), 95th percentile Hausdorff distance (HD), and mean surface distance (MSD) for organs at risk contours generated via the proposed segmentation methodology, with and without the inclusion of fraction adaptation. Metrics indicating a significant improvement ($P < .05$) in transitioning from nonadaptive to single fraction-adaptive segmentations are highlighted with a star symbol.

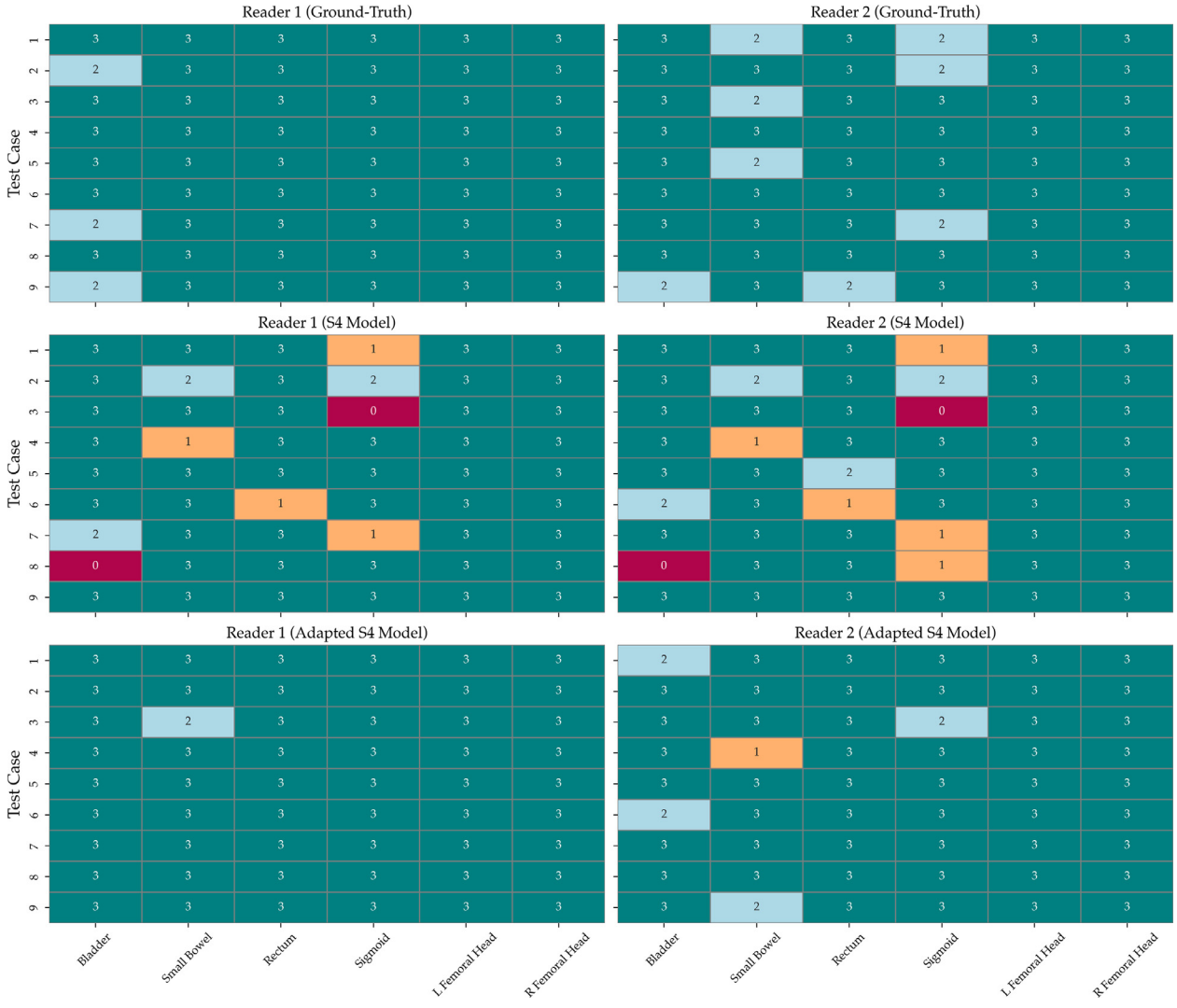

Figure 5 The oncologist reader assessment scores for 9 test patients were evaluated blindly under 3 contouring configurations: ground-truth, synthetic magnetic resonance imaging-assisted model (scenario 4, S4), and the adapted S4 model. Each color and its corresponding score signify a different assessment category: Acceptance (3) in teal, Acceptance with Minor Revision (2) in light blue, Major Revision Required before Acceptance (1) in orange, and Rejection with Mandate for Contour Redefinition (0) in red. The order of test cases is consistent in all figures.

kappa: 1.0), and the femoral heads (0 out of 9; Cohen's kappa: 1.0). For the sigmoid, however, there was a slight divergence in the assessment of reader 1 (3 of 9) and reader 2 (4 of 9), albeit with a Cohen's kappa score of 0.84. This divergence was also observed for the small bowel where a Cohen's kappa score of 0.65 was calculated.

The adaptation technique resulted in a significant improvement in scores. Reader 1 achieved a 100% acceptance rate for all OARs in the test data, whereas reader 2 scored 98%. The only exception was the small bowel in test case 4, which required a major revision before treatment. However, there was a discrepancy between the readers' scored labels as "minor revision" or "acceptable." Despite this divergence in viewpoints, it had no impact on the overall acceptance rates of these contours.

Visual inspection of results, including the information on the treated CTVs, revealed that some structures, although not fully conforming to their anatomical boundaries, were located at a sufficient distance from the CTV to pass the reader assessment. Readers were more lenient for structures, for example, >3-5 cm away from the CTV. In contrast, contours generated by the adapted model not only successfully passed the reader assessment but also demonstrated superior fidelity to the actual anatomy. This can be observed in the delineation of small bowel structures in Examples 1 and 2 of Fig. 6. However, any contours that infringed the boundaries of the CTV were consistently scored as "rejection" or "major revision," depending on the utility of the generated contours (see examples 5 and 7 in Fig. 6). Moreover, the very poor segmentation of critical structures such as the bladder was

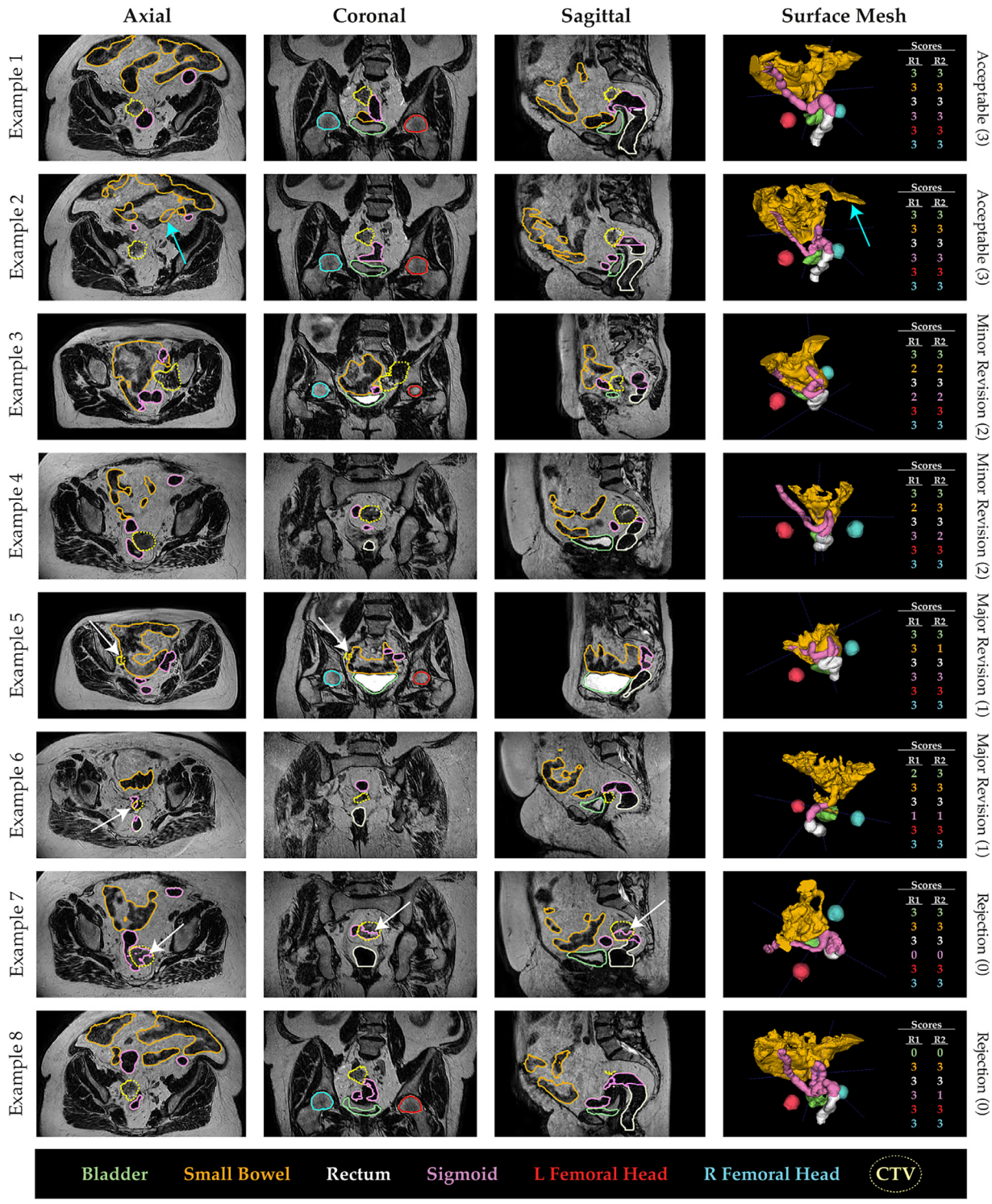

Figure 6 Qualitative representation of 8 examples from 5 test patients from the oncologist assessment. Examples 1, 2, and 8, from test case 8, show contours from an adapted S4 model, ground-truth treatment plan, and baseline S4 model, respectively. Example 3, from test case 2, is derived from the baseline S4 model. Examples 4 and 8, from test case 3, are from adapted and baseline S4 models, respectively. Example 5, from test case 4, is generated from an adapted model. Examples 6 and 7, from adapted and baseline S4 models, display the improved adaptation results from test case 7. Light blue arrows highlight ground-truth contours that, despite anatomical deviations, were accepted by both readers (R1, R2). White arrows denote areas where organs at risk contour infringements onto the clinical target volume led to scores of 0 or 1.

labeled as "rejection," irrespective of their distance from the location of the CTV (see example 8 in Fig. 6).

Discussion

Segmentation of OARs is a crucial task in RT planning because accurate execution of this task correlates to minimized collateral toxicity to healthy tissues during treatment delivery. With advanced treatment systems such as MRL, there is an increasing need for automatic algorithms to perform fast and accurate segmentations for more efficient workflows. However, 1 major limitation remains the lack of adequate and high-quality data that may hinder the performance and generalizability of algorithms.

In this study, we developed a novel domain-adaptive deep learning framework to address this shortcoming for the task of OARs segmentation for gynecologic cancers on MRL. Our technique included the use of preexisting CT repositories and synthetic MRL data to leverage anatomical knowledge from treatment plans across medical imaging domains. We demonstrated that this approach provided a robust segmentation baseline, effectively handling the anatomical complexities of OARs across different gynecologic cancers and outperforming the models trained solely on small acquired MRL data. Although transfer learning approaches can effectively improve the performance of models when facing limited data,[35,36] our approach uses relevant information not only for feature extraction but also for learning the target anatomy of our OARs.

The incorporation of soft tissue weighting into the hybrid loss function for training our segmentation algorithm provided valuable insights into the impact of training strategies on segmentation model performance. However, challenging soft-tissue structures such as small bowel, sigmoid, and rectum still faced poor segmentation outcomes using the baseline segmentation model. Statistical testing revealed that only the femoral heads had significant changes among all models for the nonadapted model, although this was perhaps largely because of protocol differences between the cohorts.

Our fraction guided adaptation strategy using patient-specific anatomical learning from preceding treatment fractions of individual cases drastically improved the segmentation performance across all OARs, leading to acceptance rates of 98% and 100% from 2 radiation oncologists through a blinded human reader assessment. Furthermore, we demonstrated through statistical testing that the adapted model yielded significant improvements in segmentations for the small bowel, rectum, and sigmoid that were targeted during the training of this model. These findings suggested that the use of 1 previous treatment fraction may be sufficient to yield significant improvements for subsequent treatment planning on MRL systems.

To our knowledge, only 1 previous study has performed per-fraction segmentation of OARs for the purpose of adaptive MRgRT.[12] The authors noted that their proposed model performed satisfactorily in delineating the OARs, showing comparable success to previous studies. However, the integration of planning or daily treatment fraction MRL images did not improve their segmentation results. Furthermore, the study leveraged a Mask-RCNN architecture, which used a backbone pretrained on a very large data set comprising natural images. Despite this, their approach architecture relied on 2-dimensional segmentations and required minutes to generate predictions. It may also lack the incorporation of domain-specific anatomical knowledge, limiting the model's adaptability for unseen patients, particularly those with atypical anatomies.

To address these shortcomings, we employed a 3-dimensional patch-based strategy that provided more training examples while still maintaining spatial consistency for contouring. Significantly, our framework generates the segmentation contours in under 1 minute. Although our approach does not completely obviate the need for human expert intervention, it offers great promise to overcome the disadvantages in current MRL practice. We believe that the inclusion of expert clinicians in evaluating and refining this pipeline is crucial to ensure the accuracy and clinical relevance of the generated segmentations. Their involvement can also help establish guidelines for incorporating the automated segmentation results into the clinical workflow, ensuring seamless integration and appropriate decision-making.

The deep learning-generated segmentation contours offer the advantage of generalizability and reduced sensitivity to bias. We propose that the contours produced by our models may be more suitable for training future algorithms than those derived from treatment fractions. ART interventions are highly time-sensitive, with decisions on organ delineation relying on the judgment of the radiation oncologist during the scanning session. Although segmentation contours close to the CTV (<3-5 cm) are meticulously scrutinized, precise conformity to organ boundaries for distant structures is often deprioritized because these variations typically do not impact treatment outcomes. Consequently, some contours in this study were revised by a radiation oncologist before being used for segmentation training.

However, it is essential to recognize that when using real-world clinical data, careful attention must be given to contour definition protocols because inconsistencies can influence quantitative evaluations. For instance, in this study, contour definitions varied for structures such as the femoral heads, which in some cases included the femoral necks, and the bowel, where certain demarcations encompassed the loops while others used more general definitions. Because of the lack of publicly available data sets in this domain, direct quantitative comparisons of performance on these OARs were not conducted.

The application of our technique could extend beyond gynecologic cancers. Such technologies are vital when

more global and standardized efforts are made to perform continual segmentation learning for treatment technologies. Given the inherent versatility of this approach, it could be adapted for use in the treatment process of other pelvic malignancies, presenting a potentially valuable tool in the ongoing drive to automate and improve treatment planning on MRL systems.

This study aimed to develop novel strategies to address the significant challenge of limited data availability for training accurate segmentation models. Access to adequate and high-quality medical data is often hindered by various factors, including limited clinician time for image annotation, the need for specific imaging technologies, protocols, and sequences, and the lack of availability of these technologies across different medical centers. In addition, the prevalence of certain diseases and the adoption rate of MRL treatments further constrain data availability for such research.

Although the findings of this study were based on a relatively small test data set, they offer valuable insights into leveraging preexisting treatment data within our clinic. These strategies could have meaningful implications for reducing treatment times and improving the efficiency of MRL technologies in performing OAR segmentation tasks for gynecologic cancer treatments. Nevertheless, further research is required to evaluate these approaches on larger data sets and within real-world clinical settings.

Several additional limitations of this study should be acknowledged. First, the evaluation was conducted on a relatively small data set from a single institution, which may limit the generalizability of the findings. Furthermore, the exclusion of target volumes during the training process could be considered a drawback because OAR demarcations might inadvertently encroach on tumor sites. Because of the substantial clinician time required, this study did not quantify the exact time oncologists spent revising or redefining contours, nor did it directly measure the time savings achieved by our method. However, qualitative clinician feedback and categorical scoring suggested the significant potential for our approach to providing these benefits. In addition, the quality of synthetic MRL images was not assessed because prior studies have already demonstrated the advantages of such methods.[25,27,28,30] Nonetheless, investigating these images remains critical to ensure that the model is learning appropriate features through this process. Future studies are essential to evaluate the clinical impact of this technique on MRL treatment times and the overall efficiency of MRgRT workflows.

Conclusions

In conclusion, this study introduced a novel framework for rapid and accurate OAR segmentation in MRgRT for gynecologic cancers. By incorporating a domain-adaptive approach, synthetic MRL imaging data were integrated to enhance segmentation performance on real-world treatment scans. Furthermore, a fraction adaptation technique was developed to use preceding treatment fractions, enabling more precise, and patient-specific segmentations. With high clinical acceptance rates, this approach has the potential to improve MRgRT workflows by leveraging preexisting treatment data while minimizing the need for extensive clinician input for annotations. This framework could serve as a valuable benchmark for training future segmentation algorithms in this domain. However, although it streamlines the segmentation process, it is not intended to replace the critical role of expert review. Future studies are essential to validate these findings on larger data sets and to comprehensively evaluate the clinical impact and applicability of this method to other cancer types.

Disclosures

Gigin Lin received research funding from the Ministry of Science and Technology Taiwan (MOST 110-2628-B-182A-018). The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Reza Kalantar was responsible for statistical analysis.

References

1. Nelms BE, Tom EWA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. *Int J Radiat Oncol Biol Phys*. 2012;82:368-378. https://doi.org/10.1016/j.ijrobp.2010.10.019.
2. Brouwer CL, Steenbakkers RJHM, van den Heuvel E, et al. 3D variation in delineation of head and neck organs at risk. *Radiat Oncol*. 2012;7:32. https://doi.org/10.1186/1748-717X-7-32.
3. Pathmanathan AU, van As NJ, Kerkmeijer LGW, et al. Magnetic resonance imaging-guided adaptive radiation therapy: a "game changer" for prostate treatment? *Int J Radiat Oncol Biol Phys*. 2018;100:361-373. https://doi.org/10.1016/J.IJROBP.2017.10.020.
4. Wang J, Chen Y, Tu Y, et al. Evaluation of auto-segmentation for brachytherapy of postoperative cervical cancer using deep learning-based workflow. *Phys Med Biol*. 2023;68:acba76. https://doi.org/10.21203/rs.3.rs-2100195/v1.
5. Schipaanboord B, Boukerroui D, Peressutti D, et al. An evaluation of atlas selection methods for atlas-based automatic segmentation in radiotherapy treatment planning. *IEEE Trans Med Imaging*. 2019;8:2654-2664.
6. Zabel WJ, Conway JL, Gladwish A, et al. Clinical evaluation of deep learning and atlas-based auto-contouring of bladder and rectum for prostate radiation therapy. *Pract Radiat Oncol*. 2021;11:e80-e89. https://doi.org/10.1016/J.PRRO.2020.05.013.

7. Daisne JF, Blumhofer A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. *Radiat Oncol*. 2013;8:154.
8. Chu C, De Fauw J, Tomasev N, et al. Applying machine learning to automated segmentation of head and neck tumour volumes and organs at risk on radiotherapy planning CT and MRI scans. *F1000Res*. 2016;5:2104.
9. Ibragimov B, Pernus F, Strojan P, Xing L. TH-CD-206-05: machine-learning based segmentation of organs at risks for head and neck radiotherapy planning. *Med Phys*. 2016;43:3883.
10. Wang J, Chen Y, Xie H, Luo L, Tang Q. Evaluation of auto-segmentation for EBRT planning structures using deep learning-based workflow on cervical cancer. *Sci Rep*. 2022;12:13650.
11. Liu Z, Liu X, Xiao B, et al. Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network. *Phys Med*. 2020;69:184-191. https://doi.org/10.1016/J.EJMP.2019.12.008.
12. Breto AL, Spieler B, Zavala-Romero O, et al. Deep learning for per-fraction automatic segmentation of gross tumor volume (GTV) and organs at risk (OARs) in adaptive radiotherapy of cervical cancer. *Front Oncol*. 2022;12:1730. https://doi.org/10.3389/FONC.2022.854349.
13. Li Z, Zhu Q, Zhang L, Yang X, Li Z, Fu J. A deep learning-based self-adapting ensemble method for segmentation in gynecological brachytherapy. *Radiat Oncol*. 2022;17:152. https://doi.org/10.1186/S13014-022-02121-3.
14. van Dijk LV, Van den Bosch L, Aljabar P, et al. Improving automatic delineation for head and neck organs at risk by deep learning contouring. *Radiother Oncol*. 2020;142:115-123. https://doi.org/10.1016/J.RADONC.2019.09.022.
15. Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L. Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncol*. 2019;58:257-264. https://doi.org/10.1080/0284186X.2018.1529421.
16. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. *Semin Radiat Oncol*. 2019;29:185-197. https://doi.org/10.1016/J.SEMRADONC.2019.02.001.
17. Mohammadi R, Shokatian I, Salehi M, Arabi H, Shiri I, Zaidi H. Deep learning-based auto-segmentation of organs at risk in high-dose rate brachytherapy of cervical cancer. *Radiother Oncol*. 2021;159:231-240. https://doi.org/10.1016/J.RADONC.2021.03.030.
18. Kazemifar S, Balagopal A, Nguyen D, et al. Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. *Biomed Phys Eng Express*. 2018;4:055003. https://doi.org/10.1088/2057-1976/AAD100.
19. Balagopal A, Kazemifar S, Nguyen D, et al. Fully automated organ segmentation in male pelvic CT images. *Phys Med Biol*. 2018;63:245015. https://doi.org/10.1088/1361-6560/AAF11C.
20. Song Y, Hu J, Wu Q, et al. Automatic delineation of the clinical target volume and organs at risk by deep learning for rectal cancer postoperative radiotherapy. *Radiother Oncol*. 2020;145:186-192. https://doi.org/10.1016/J.RADONC.2020.01.020.
21. Guo H, Wang J, Xia X, et al. The dosimetric impact of deep learning-based auto-segmentation of organs at risk on nasopharyngeal and rectal cancer. *Radiat Oncol*. 2021;16:113. https://doi.org/10.1186/S13014-021-01837-Y.
22. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci*. 2015;9351:234-241. https://doi.org/10.1007/978-3-319-24574-4_28.
23. Yoganathan SA, Paul SN, Paloor S, et al. Automatic segmentation of magnetic resonance images for high-dose-rate cervical cancer brachytherapy using deep learning. *Med Phys*. 2022;49:1571-1584. https://doi.org/10.1002/MP.15506.
24. Kalantar R, Lin G, Winfield JM, et al. Automatic segmentation of pelvic cancers using deep learning: state-of-the-art approaches and challenges. *Diagnostics (Basel)*. 2021;11:1964. https://doi.org/10.3390/DIAGNOSTICS11111964.
25. Bird D, Nix MG, McCallum H, et al. Multicentre, deep learning, synthetic-CT generation for ano-rectal MR-only radiotherapy treatment planning. *Radiother Oncol*. 2021;156:23-28. https://doi.org/10.1016/J.RADONC.2020.11.027.
26. Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol*. 2017;62:L41-L50. doi:10.1088/1361-6560/AA9517.
27. Dong X, Lei Y, Tian S, et al. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. *Radiother Oncol*. 2019;141:192-199. https://doi.org/10.1016/J.RADONC.2019.09.028.
28. Lei Y, Wang T, Tian S, et al. Male pelvic multi-organ segmentation aided by CBCT-based synthetic MRI. *Phys Med Biol*. 2020;65:035013. https://doi.org/10.1088/1361-6560/AB63BB.
29. Kalantar R, Lalondrelle S, Winfield JM, Messiou C, Koh DM, Blackledge MD. Synthetic MRI-assisted multi-wavelet segmentation framework for organs-at-risk delineation on CT for radiotherapy planning. *ISMRM*. 2021. PublishedAccessed June 7, 2023; https://archive.ismrm.org/2021/3747.html.
30. Kalantar R, Messiou C, Winfield JM, et al. CT-based pelvic T1-weighted MR image synthesis using UNet, UNet++, and cycle-consistent generative adversarial Network (Cycle-GAN). *Front Oncol*. 2021;11:665807. https://doi.org/10.3389/FONC.2021.665807.
31. Maspero M, Savenije MHF, Dinkla AM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol*. 2018;63:185001. https://doi.org/10.1088/1361-6560/AADA6D.
32. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV). *IEEE*. 2017:2223-2232.
33. Cardoso MJ, Li W, Brown R, et al. MONAI: an open-source framework for deep learning in healthcare. Published online November 4, 2022. Accessed June 7, 2023. https://arxiv.org/abs/2211.02701v1.
34. Loshchilov I, Hutter F. Decoupled weight decay regularization. *7th International Conference on Learning Representations, ICLR 2019*. November 14, 2017. Published onlineAccessed June 26, 2023; https://arxiv.org/abs/1711.05101v3.
35. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22:1345-1359.
36. Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning. *Proc IEEE*. 2021;109:43-76.