**Conclusion:**
With iDL, high segmentation accuracy was achieved across multiple datasets, and cross-center, even without transfer learning. The tool was deemed very usable, and minimum manual corrections were needed for GTVn, and for two out of three observers for GTVt.

**Keywords:** interactive deep-learning, multi-center testing

**References:**
[1] Andrearczyk V, *et al.* "Overview of the HECKTOR Challenge at MICCAI 2021: Automatic Head and Neck Tumor Segmentation and Outcome Prediction in PET/CT Images", in: Head and Neck Tumor Segmentation and Outcome Prediction, 1-37 (2022).

**3762**

Digital Poster

**Zero-shot auto-segmentation of rectal cancer CTV for MRI-guided online adaptive radiotherapy prompted with pre-treatment delineations**

Nicole Ferreira Silverio[1], Alice Couwenberg[1], Aneesh Rangnekar[2], Harini Veeraraghavan[2], Tomas Janssen[1]
[1]Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, Netherlands. [2]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, USA

**Purpose/Objective:**
Auto-segmentation of clinical target volumes (CTV) is a complex task, due to large variation in patient anatomy and implicit clinical decisions that influence the delineations. Online adaptive radiotherapy (OART) has the advantage that pre-treatment delineations are available. [1] showed that a deep-learning mesorectum CTV auto-segmentation model for MRI-guided OART improved substantially when this pre-treatment information is added as input. However, such a model is highly specialized, only applicable for the task it was trained for, limiting its generalizability. Extension to other target sites requires (re)training with extensive and well curated labeled datasets, which is time consuming, labor intensive and impractical to obtain in practice.
Foundation models could potentially overcome the problem by leveraging knowledge to extract universally applicable features through pre-training on large heterogeneous datasets. These models can be used in zero-shot setting through prompting, without additional training.
The aim of this study was to test whether a promptable foundation model, when presented with the pre-treatment delineation, could achieve comparable quality as a specialized deep-learning model for auto-segmentation of the mesorectum CTV for OART.

**Material/Methods:**
Two 2D in-house auto-segmentation models (nnU-Net based) were trained on 476 3D T2 weighted MRI scans and accompanying manual mesorectum delineations of 39 rectal cancer patients treated on a 1.5T MR-Linac. Data was split on a patient level into training-validation-test (20-5-14 patients respectively). One model was trained on only MRI images (MRI-only), another model was trained on a combination of MRI images and pre-treatment delineation based weight maps (MRI+prior) [1].
As zero-shot approach, MedSAM [2] (2D) was applied to the same test-set. Per patient, the pre-treatment delineation was registered to the daily MRI. Per slice, the delineation’s bounding box + 6mm was used to prompt the model.
Segmentations were compared with the ground truth on the Dice similarity coefficient (DSC), 95$^{th}$ percentile Hausdorff distance (95HD) and mean surface distance (MSD).

**Results:**

Median DSC was 0.81, 0.87, 0.91 for MedSAM, MRI-only and MRI+prior respectively. MSD results showed a similar trend, while 95HD was comparable between approaches (figure 1). Figure 2 shows an example patient.

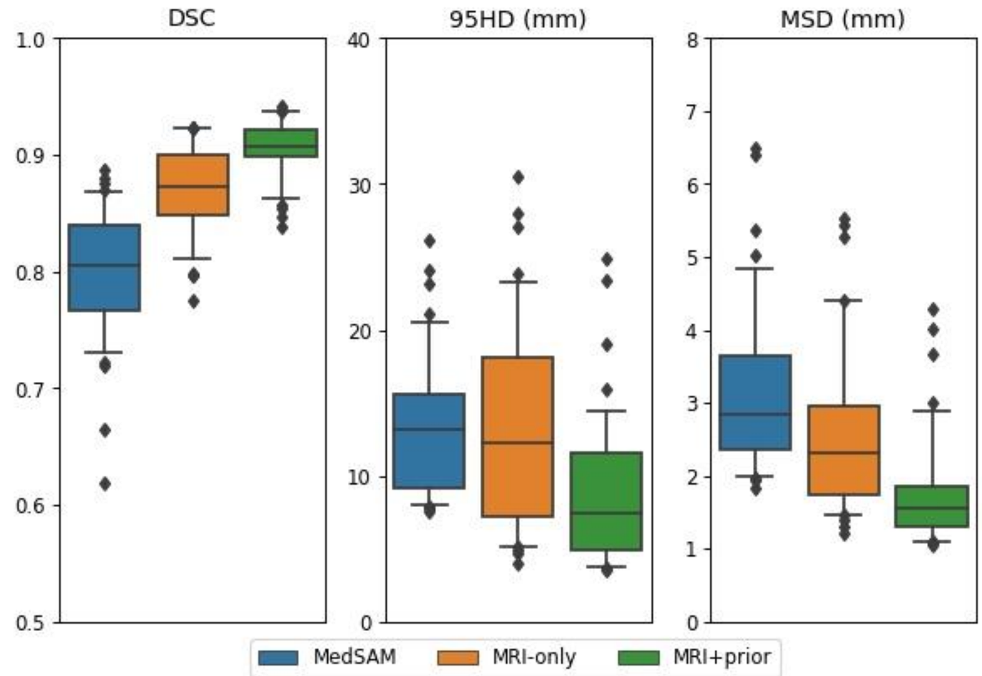


Figure 1. Performance of the models in this study on the segmentation metrics Dice similarity coefficient (DSC), 95th Hausdorff distance (HD95) and mean surface distance (MSD).

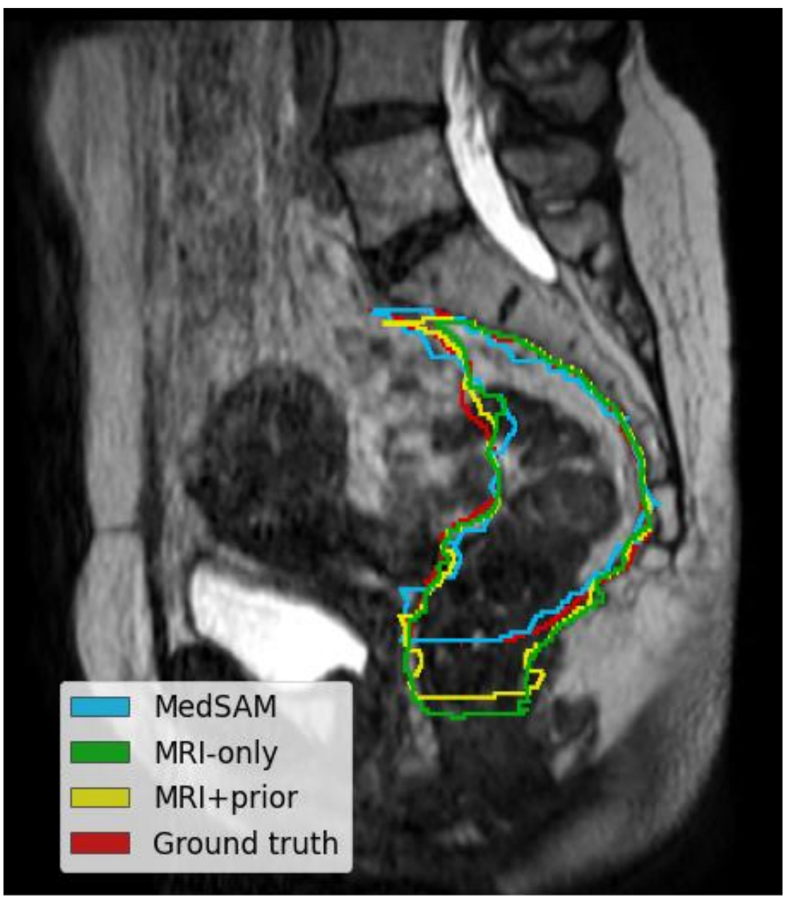


Figure 2. An example of contours generated by the models, after smoothing as a postprocessing step, are shown compared to the ground truth delineation.

**Conclusion:**
Both in-house developed models outperformed zero-shot MedSAM. If data availability is no issue, training a specialized model (preferably with weight maps) results in the best performance. However, with a median DSC of 0.81 and MSD of 3mm MedSAM performs adequate to utilize when data is scarce. Additional fine-tuning of a foundation model on limited data is expected to further improve the results.

**Keywords:** online adaptive RT, target segmentation, AI

**References:**
1. Ferreira Silvério, N., et al., *Evaluation of Deep Learning Clinical Target Volumes Auto-Contouring for Magnetic Resonance Imaging-Guided Online Adaptive Treatment of Rectal Cancer.* Advances in Radiation Oncology, 2024. **9**(6): p. 101483.
2. Ma, J., et al., *Segment anything in medical images.* Nature Communications, 2024. **15**(1): p. 654.

**3763**

Digital Poster

**AI-based contouring in MR gynaecologic brachytherapy workflow: A Practical Evaluation**

Clélie Castex[1], François-Xavier Arnaud[2], Francesca Di Franco[3], Laure Vieillevigne[2], Anne Ducassou[1]
[1]Radiation Oncology, Oncopole Claudius Regaud (OCR), Institut Universitaire du Cancer de Toulouse-Oncopole (IUCTO), Toulouse, France. [2]Medical Physics, Oncopole Claudius Regaud (OCR), Institut Universitaire du Cancer de Toulouse-Oncopole (IUCTO), Toulouse, France. [3]Research department, IRUDIGI SARL, Bayonne, France

**Purpose/Objective:**
Artificial intelligence (AI)-based contouring is widely used in external beam radiotherapy to improve efficiency and consistency. However, its application in brachytherapy, which requires precise delineation of target volumes and organs at risk (OARs) for optimal outcomes, remains underexplored. Magnetic Resonance (MR)-based brachytherapy poses challenges due to the complexity of anatomical structures and the high level of precision required. This study aimed to assess the impact of AI contouring in MR-based gynaecologic brachytherapy by comparing AI-based and reference contours.

**Material/Methods:**
Ten patients, each contoured by a single physician, were included in this study. A commercial AI solution (Limbusv1.8,Radformation,USA) was used to automatically delineate four organs at risk (OARs) on MR images: bladder, bowel, sigmoid colon, and rectum. The auto-contour accuracy was evaluated using a 5-point Likert scale, where a score of 5 indicated that the contour was usable without any modifications, while a score of 1 indicated that the contour was completely unusable. AI-generated contours ($AI_C$) were compared against manual reference contours ($M_C$) using the Dice similarity coefficient (DSC) and the 95th percentile Hausdorff distance (HD95%). Additionally, intra-observer variability (one month later) was assessed by comparing the physician-reviewed AI-based contours ($AI\text{-}R_C$) to the reference $M_C$.

**Results:**
The mean (±SD) Likert score for each OAR was: bladder 3.5 (±0.5), bowel 2.5 (±0.7), sigmoid colon 3.1 (±0.7), and rectum 3.3 (±0.7). Comparing $AI_C$ versus $M_C$ contours, mean DSC values resulted: 0.85±0.06, 0.58±0.09, 0.61±0.10, and 0.74±0.10 for the bladder, bowel, sigmoid colon and rectum, respectively. The HD95% values (mm) were: 7.51±4.61, 21.33±12.04, 29.97±23.80, and 14.44±7.77 for the bladder, bowel, sigmoid colon and rectum, respectively. The mean DSC value between $AI\text{-}R_C$ and $M_C$ contours resulted: 0.87±0.08, 0.61±0.12, 0.73±0.06, and