Geometric Comparison of Auto Segmentation Accuracy of Web-Based AI Model for Efficient Pelvic Radiotherapy

Silambarasan Anbumani[1], Milan Bimali[2], Sanjay Maraboyina[1], and Ganesh Narayanasamy[1]

[1]Department of Radiation Oncology, University of Arkansas for Medical Sciences, Little Rock, AR.

[2]Department of BioStatistics, University of Arkansas for Medical Sciences, Little Rock, AR.

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

ABSTRACT

Purpose: This study evaluates a commercial web-based AI model for auto-segmentation of thoracic and pelvic sites by comparing its contours to expert ground truth and an alternative AI model, emphasizing its accuracy and clinical relevance for radiotherapy planning.

Results: TP ART-Plan took an average of 3-5 minutes to generate auto-segmented contours, while ProtégéAI took 5-8 minutes. Robust auto-segmentation was observed for the , bladder, and femoral heads, with a mean DSC >0.8 and MDA <3 mm, compared with ground-truth contours. The mean DSC and MDA values for the bowel (DSC: 0.61±0.12, MDA: 18.98±20.8), and rectum (DSC: 0.73±0.11, MDA: 4.86±2.67). The substantial differences in the bowel, and rectum, were mainly due to interobserver and institutional margin variability, with image quality likely playing a role. Auto-segmentation performance for ART-Plan vs ProtégéAI was comparable across thoracic and pelvic structures, with ART-Plan outperforming ProtégéAI for the rectum and protégéAI slightly outperforming for the bowel region.

Conclusion: The AI-based TP ART-Plan and MIM ProtégéAI generated thoracic and pelvic contours were evaluated against the physician-approved ground truth. Geometric evaluation of the auto-segmented structures against ground truth contours revealed clinically relevant differences in fidelity which may impact treatment outcomes.

CONTACT

Ganesh Narayanasamy
ganesh@uams.edu
Dept of Radiation Oncology, UAMS
3900 W Capitol Ave, Little Rock, AR
University of Arkansas for Medical Sciences

INTRODUCTION

- Accurate delineation of organs at risk (OARs) is crucial for ensuring effective radiotherapy outcomes, reliable plan evaluation, and enhanced treatment precision.
- This study systematically compares benchmark AI-based auto-segmentation with institution-specific clinical standards and commercial vendor solutions, establishing a reproducible, clinically impactful validation framework.

METHODS AND MATERIALS

- The ART-Plan™ (v2.3.1, TheraPanacea Inc., France) web-based software was employed for the validation of auto-segmentation, integrated within Brainlab Elements (v4.0, Brainlab AG, Germany).
- A retrospective study was conducted on 15 simulation CT datasets. The pelvic OARs included bladder, bowel bag, rectum, and femoral heads. The auto-segmented contours were analyzed and compared with the physician-approved contours, and also the auto-segmented contours developed by Contour ProtégéAI v1.3.1 (MIM Software, OH).
- This analysis was selected from a time frame when no auto-contouring tools were used in our clinic, and the contours were explicitly created by a dosimetrist and approved by a physician.
- Auto-segmentation performance was quantified using Dice Similarity Coefficient (DSC), Mean Distance to Agreement (MDA), and Hausdorff distance (HD).
- The results were assessed to determine whether contouring uncertainty remained within the recommended thresholds, with DSC values above 0.8 and MDA below 3mm, consistent with the guidelines outlined in AAPM Task Group 132 on image registration.
- The geometric results across the models were compared using a paired two-tailed t-test, with statistical significance set at $p < 0.05$.

RESULTS

TP ART-Plan took an average of 3-5 minutes to generate auto-segmented contours, while ProtégéAI took 5-8 min.

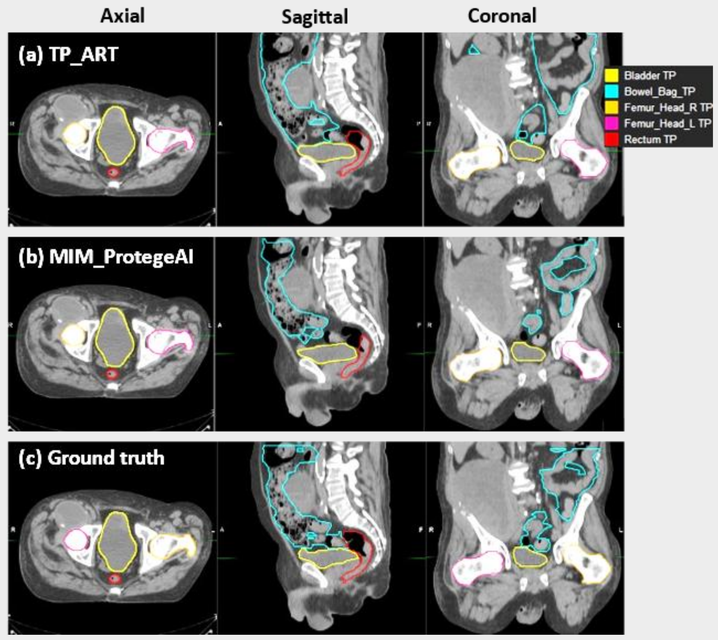

Figure 1: Representative CTSim images with auto-segmented contours TheraPanacea (TP), (b) MIM ProtegeAI and (c) the MD-approved ground truth (Gold standard) contours used for the pelvic radiotherapy

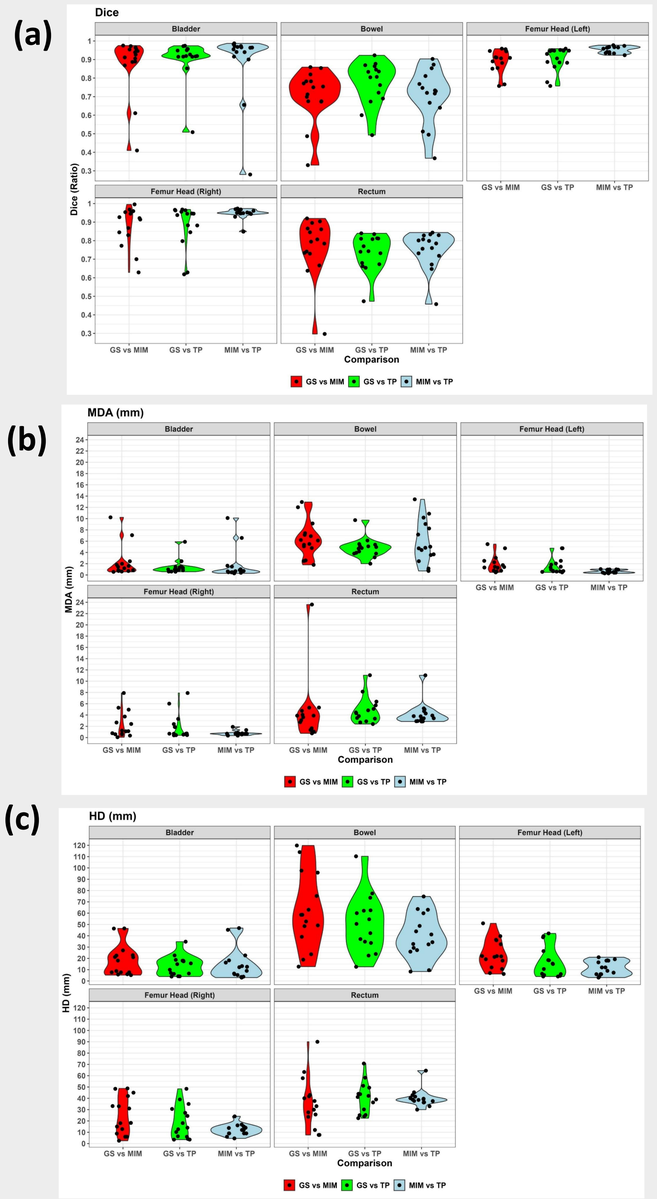

Figure 2: Comparison of (a) DSC, (b) MDA, and (c) HD for the TheraPanacea (TP) plan and MIM ProtegeAI, with the gold standard (GS) contours, along with a comparison between the TP ART plan and MIM ProtegeAI autosegmented contours (MIM vs TP).

DISCUSSION

- Bladder and femoral heads show robust auto-segmentation with mean DSC >0.9 and MDA <2 mm.
- The mean DSC and MDA values for the bowel (DSC: 0.8±0.1, MDA: 4.82 ± 1.65), and rectum (DSC: 0.7±0.1, MDA: 4.75 ± 2.27).
- The relatively higher differences in the bowel and rectum were due to interobserver and institutional margin variability, with image quality likely playing a role.
- However, no statistical significance in geometric metrices for ART-Plan and ProtégéAI was observed for pelvic structures, except for the femoral head.
- The results were assessed to determine whether contouring uncertainty remained within the recommended thresholds, with DSC values above 0.8 and MDA below 3mm, consistent with the guidelines outlined in AAPM Task Group 132 on image registration.
- The study suffers from low statistical power, and a future study will investigate large patient cohorts.

CONCLUSIONS

- The pelvic contours from the AI-based TP ART-Plan and MIM ProtégéAI were compared to the physician contours.
- Geometric analysis of auto-segmented structures showed clinically relevant accuracy, potentially improving treatment planning efficiency.
- Overall, these findings underscore the potential planning of TP ART-Plan, an AI-driven segmentation technique, to reduce manual editing effort, enhance consistency in OAR delineation, and improve the efficiency and reliability of pelvic and thoracic radiotherapy .

REFERENCES

1. Young T, Butterworth V, Misson S, et al.. Br J Radiol. 2025;98(1174):1632-1641. doi:10.1093/bjr/tqaf098Sherer
2. MV, Lin D, Elguindi S, et al., Radiother Oncol. 2021; 160:185-191. doi:10.1016/j.radonc.2021.05.003.