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

A Geometric and Dosimetric Assessment of Autocontouring for the Delineation of Pelvic Nodal Radiation Therapy Targets

Daniel Hoffman, MS[1,2]; Alexander R. Podgorsak, Ph.D. [1]
[1]University of Rochester; [2]Allegheny Health Network

University of Rochester Medicine

Allegheny Health Network

Abstract

To establish clinical guidelines for auto-contouring (AC) pelvic lymph node target volumes, this retrospective study of 50 patients (47 prostate, 3 bladder) compared two deep-learning AC models (AutoContour, RADformation) against physician-drawn manual contours.

Geometrically, Model 1 demonstrated significantly higher agreement to the ground truth PTV than Model 2 (median DSC: 0.765 vs. 0.669; median MSD: 3.79 mm vs. 6.30 mm).

Dosimetrically, using an auto-planner (RapidPlan, Varian), plans based on Model 1's target volumes achieved significantly superior ground truth dose coverage (median $V_{100\%}$: 85.9% vs. 65.8%), albeit with slightly higher median doses to the bladder and rectum.

Relying solely on either unedited AC model resulted in clinically unacceptable target underdosing. Therefore, while Model 1 provides a superior geometric starting point for delineation, mandatory physician review and manual editing remain essential prior to clinical use.

Surface distance metrics (Mean Surface Distance and Maximum Hausdorff Distance) were significantly better predictors of dosimetric success than traditional volumetric overlap (Dice-Similarity coefficient), suggesting they should be prioritized when evaluating nodal auto-contours.

Introduction

Clinical Background:

- 20–40% post-prostatectomy recurrence rate due to microscopic spread to pelvic lymph nodes (PLNs).[1, 2]
- RTOG 0534 (SPPORT) trial proved expanding radiotherapy to PLNs reduces progression.[3]

Contouring Challenges:

- 2021 NRG Oncology consensus mandates expanded, complex PLN volumes; replacing 2009 RTOG consensus.[4, 5]
- Contouring these altered regions is time-consuming and variable due to lack of clear borders.[4]

Role of Auto-Segmentation:

- Deep-learning auto-segmentation effectively segments Organs at Risk (bladder, rectum) but struggles with diffuse lymphatic regions.[6–8]
- Commercial tools lack validation for applying the expanded 2021 NRG guidelines to post-prostatectomy anatomy.[8]

Methods and Materials

Patient Cohort & Workflow:

- Retrospective analysis of 50 post-prostatectomy PLNRT patients (47 prostate, 3 bladder) across 3 physicians.
- Unedited auto-contours generated on simulation CTs via AutoContour (RadFormation).[8]
- Evaluated Model 1 (2021 NRG) and Model 2 (2009 RTOG) against manual ground truth.

Geometric Evaluation:

- Spatial accuracy (n = 50) assessed in Velocity (Varian) software using success thresholds: Dice-Similarity Coefficient (DSC) ≥ 0.8; Relative Volume Difference (RVD) ideal = 0; Mean Surface Distance (MSD) ≤ 3 mm; Maximum Hausdorff Distance (HD) ≤ 3 mm

Dosimetric Evaluation:

- 47 prostate cases evaluated (5 mm CTV-to-PTV margin).
- Standardized RapidPlan (Varian) model and original beam geometries used to isolate contour variation.
- Clinical goals: Ground truth PTV_{Manual} $V_{100\%}$ ≥ 95%; Bladder $V_{88\%}$ < 70%; Rectum $V_{88\%}$ < 55%.

Statistics:

- Wilcoxon Signed-Rank test and Spearman's Rank correlation.

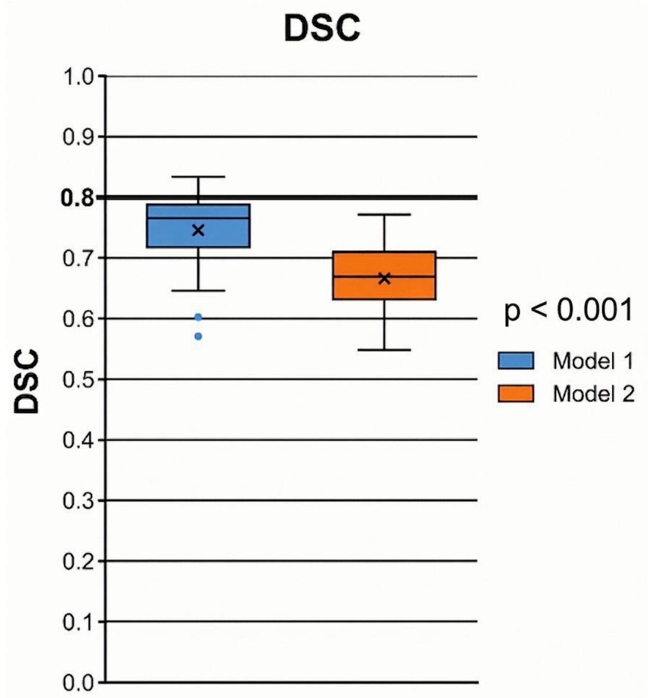

Chart 1. Volume Overlap (DSC)

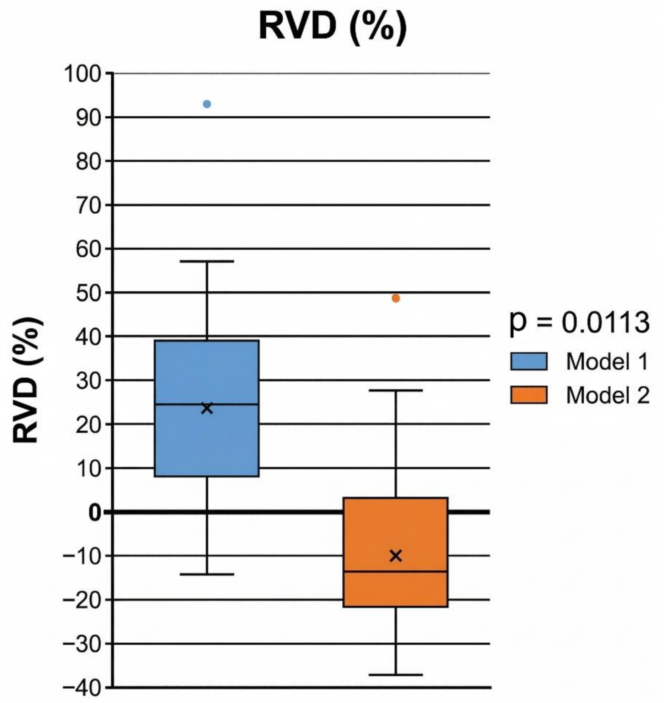

Chart 2. Contouring Bias (RVD)

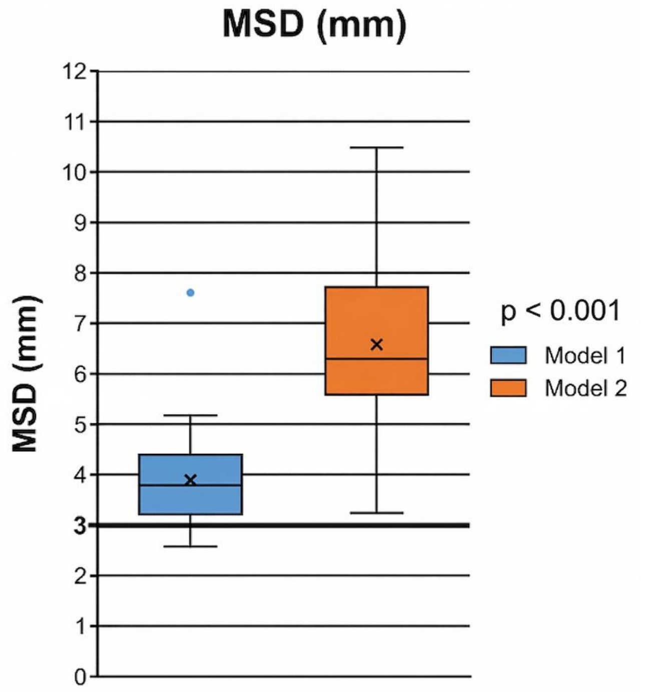

Chart 3. Surface Accuracy (MSD)

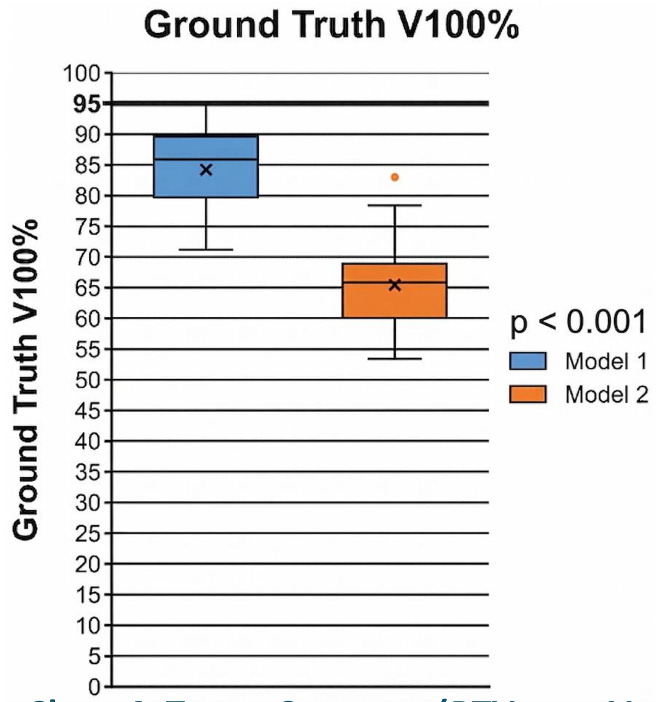

Chart 4. Target Coverage (PTV_{Manual} $V_{100\%}$)

Results

Geometric Accuracy:

- Model 1 (2021 NRG) significantly outperformed Model 2 across all spatial metrics (p < 0.001).
- Model 1 systematically over-contoured and Model 2 under-contoured.

Dosimetric Impact:

- Model 1 provided superior ground truth target coverage.
- Both models met clinical OAR dose constraints, remaining well below failure thresholds.

Correlation:

- Surface metrics (MSD, HD) predicted coverage for both models; DSC only for Model 2.
- OAR sparing was driven by optimization constraints rather than contour boundaries.

Table 1. Key Geometric and Dosimetric Outcomes

Metric	Model 1 (2021 NRG) Median	Model 2 (2009 RTOG) Median	Significance
Volume Overlap (DSC)	0.765	0.669	p < 0.001
Contouring Bias (RVD)	+24.5%	-13.6%	p = 0.0113
Surface Accuracy (MSD)	3.79 mm	6.30 mm	p < 0.001
Target Coverage (PTV_{Manual} $V_{100\%}$)	85.9%	65.8%	p < 0.001

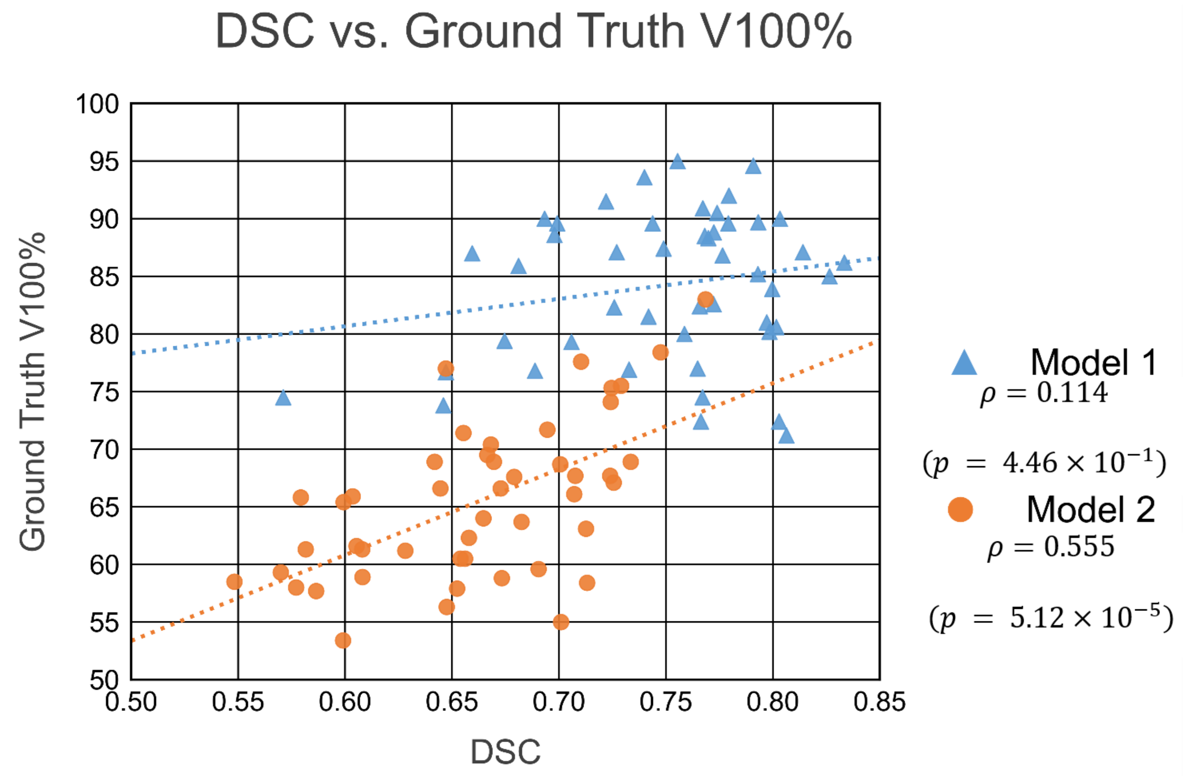

Chart 5. Predictive Value of Geometric Overlap on Target Coverage

Discussion

Quantitative Success vs. Qualitative Blind Spots:

- Model 1 is geometrically superior, establishing Model 2 as obsolete for this population.
- Both models lack clinical intuition, systematically truncating presacral nodes (rigidly stopping at S3) and underestimating superior borders compared to physician ground truth.

Necessity of Manual Intervention:

- Relying on unedited Model 1 contours caused a ~14% underdosage of the intended target volume.
- This underdosage is a compounded error from AI geometric omissions and the optimizer hitting an artificial plateau.
- Automated tools provide an initial draft but require manual clinical editing.

Disconnect Between Geometry and Dosimetry:

- Surface metrics (MSD, HD) are reliable predictors of dose coverage; DSC is a poor predictor for expansive targets.
- OAR toxicity profiles were driven by the optimizer's priorities rather than auto-segmentation spatial accuracy.

Conclusions

Clinical Applicability & Limitations:

- Modern PLN AutoContour model (Model 1) offers geometrically superior baselines vs. legacy model (Model 2).
- Both AutoContour PLN models truncate presacral nodes and superior borders in regions lacking sharp gradients.
- Unedited auto-contours cause unacceptable target underdosing.

Predicting Dosimetric Success:

- DSC failed to predict target coverage (PTV_{Manual} $V_{100\%}$) for expansive volumes.
- Surface metrics (MSD, HD, RVD) are significantly more reliable predictors as they capture localized boundary errors dictating dose fall-off.

Future Workflows:

- Oncologists are transitioning from primary delineators to expert AI editors.
- For complex postoperative targets, AI is a useful starting point, but manual intervention remains a clinical necessity for patient safety.

Contact

Alexander Podgorsak
Alexander_Podgorsak@URMC.Rochester.edu

References

1. Kolodziej M. Am J Manag Care. 2014;20(12 Suppl):S273-S281.
2. Tourinho-Barbosa R, et al. Int Braz J Urol. 2018;44(1):14-21.
3. Pollack A, et al. Lancet. 2022;399(10338):1886-1901.
4. Hall WA, et al. Int J Radiat Oncol Biol Phys. 2021;109(1):174-185.
5. Lawton CA, et al. Int J Radiat Oncol Biol Phys. 2009;74(2):383-387.
6. Kiljunen T, et al. Diagnostics. 2020;10(11):959.
7. Mofid B, et al. J Biomed Phys Eng. 2025;15(6):575-588.
8. Liu C, et al. Radformation. 2023.