Multi-anatomy geometric accuracy study for MRI based deep learning segmentation in MR-only radiotherapy planning

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

A. Amjad, PhD[1]; X. Jiofeng, PhD[2]; D Thill, MS[2]; R. Geoffrey, BS[1]; W. Hall, MD[1]; E. Paulson, PhD[1]

[1]Department of Radiation Oncology, Medical College of Wisconsin, Milwaukee, Wisconsin, United States

[2]Elekta Instruments AB, Stockholm, Sweden

Froedtert & Medical College of Wisconsin

INTRODUCTION

In this work we evaluated the accuracy and usability of deep learning based automatic segmentation (DLAS) models for MR-only radiotherapy (RT) planning. One can either approach DLAS training with clinic-based, sequence-based models, or via vendor generated DLAS trained on generalized/multi-institute data. Our work evaluates the accuracy of three MR only DLAS models, HeadNeck (HN), Male Pelvis (MP), and Female Pelvis (FP) generated by Elekta Inc.

MATERIAL & METHODS

In this work the deep learning based automatic segmentation (DLAS) models were trained using a ResUNet3D network whose input and output size was 112x112x112 at 1.5mm isotropic resolution.

- ❖The training data for HeadNeck (HN), Male Pelvis (MP), and Female Pelvis (FP) comprised of images provided by MOMENTUM clinical trial (NCT04075303). which were acquired from a variety of institutions using the Elekta Unity system. The summary of the training and testing data, OARs and MRI sequences is shown in table 1 and 2.
- ❖Ground truth images were recontoured by Elekta personnel following a common clinical guideline and were reviewed by experienced clinicians.
- ❖During training, the objective function was set to minimize a weighted cross-entropy and DICE loss function with PyTorch's ADAM optimizer. On-line data augmentation included random applications of linear intensity transformation, affine deformation, and bias field correction.
- ❖Each model was trained for 500 iterations of 100 batches on a single NVIDIA V100 GPU, taking approximately 2.5 days to complete.
- ❖Models were then executed by the research tool (ADMIRE v4.2, Elekta).

Table 1. Details of the MRI training data used in the DLAS training.

Anatomical site	# of OARs	# of training data	Primary sequence	Secondary sequence
Head Neck	13	85	2min T2w	6min T2w, Institution-specific 3min T2w & 2min T2w
Male Pelvis	38	60	6min T2w	Spair, Stir, Dixon
Female Pelvis	15	64	2min T2w	Institution-specific 2min T2w

Table 2. Details of the MRI testing data used in the DLAS evaluation.

Anatomical site	# of OARs	# of Testing data	Primary sequence	Secondary sequence
Head Neck	10	10	T1w Tra 1.5T	T1w Radia 3T
Male Pelvis	5	10	T2w Tra 1.5T	T2w AX & SPACE 3T
Female Pelvis	3	10	T2w SPACE 3T	NA

RESULTS

❖Segmentation generation time across the three models was <30 sec for all datasets.

Figure 1. Box plots of 3D DSC for female and male pelvis OARs.

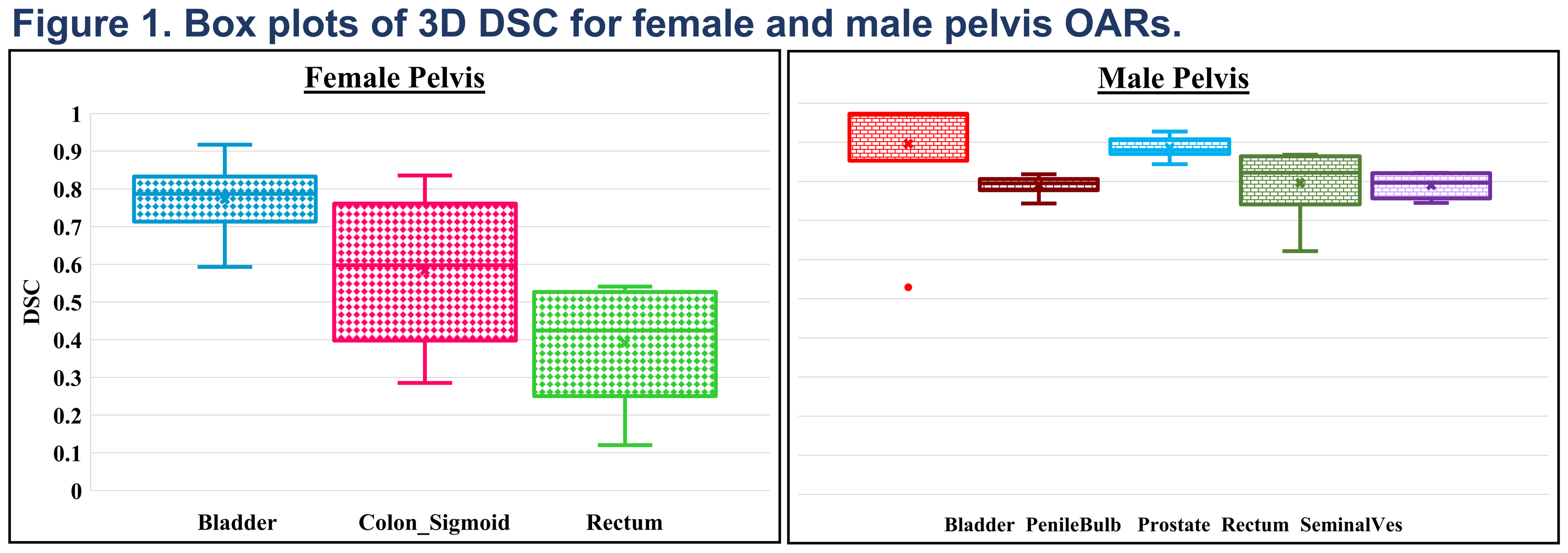


Figure 2. Slice-based percentage of acceptably of the head neck OARs.

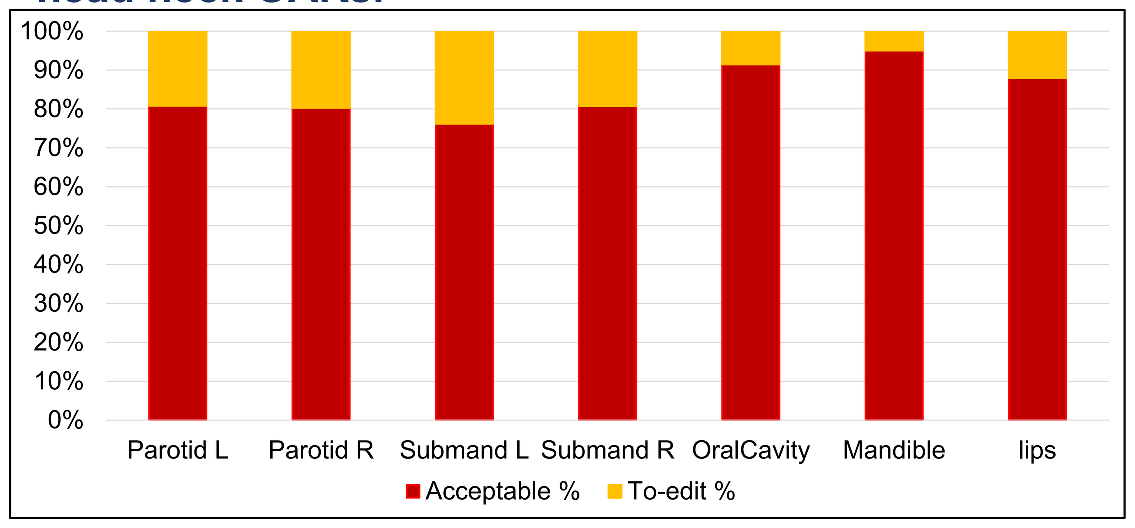
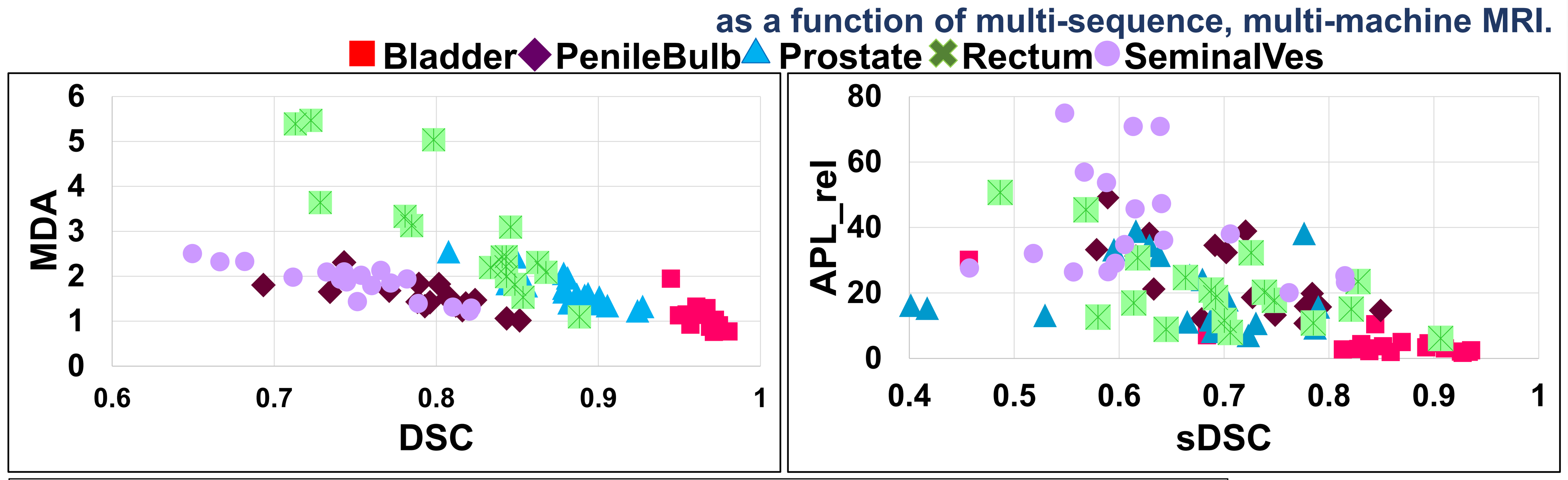


Figure & Table 3. Multi-accuracy metric analysis and scatter plots to demonstrate segmentation capability of the male pelvis organs at risk as a function of multi-sequence, multi-machine MRI.



		Bladder	PenileBulb	Prostate	Rectum	SeminalVes
V-Diff	_Space	-0.03	0.18	0.04	-0.09	-0.02
	_Ax T2	0.00	-0.03	0.16	-0.06	-0.08
	_Tra	-0.14	0.00	0.05	-0.06	0.04
DSC	_Space	0.97	0.80	0.89	0.82	0.74
	_Ax T2	0.96	0.72	0.87	0.80	0.72
	_Tra	0.90	0.79	0.89	0.80	0.79
MDA	_Space	1.03	1.44	1.43	2.34	1.82
	_Ax T2	1.16	2.02	1.97	3.21	2.19
	_Tra	3.42	1.62	1.68	4.94	1.57
HD95	_Space	2.70	4.12	3.88	10.59	5.32
	_Ax T2	3.38	5.65	4.57	16.05	6.34
	_Tra	13.34	4.39	4.34	26.29	4.75
sDSC	_Space	0.86	0.70	0.69	0.70	0.59
	_Ax T2	0.84	0.67	0.57	0.71	0.57
	_Tra	0.83	0.70	0.69	0.67	0.72
APT_rel	_Space	5.2	26.3	19.9	28.3	58.3
	_Ax T2	2.8	38.1	12.0	14.7	36.3
	_Tra	7.6	28.5	14.2	19.4	28.6

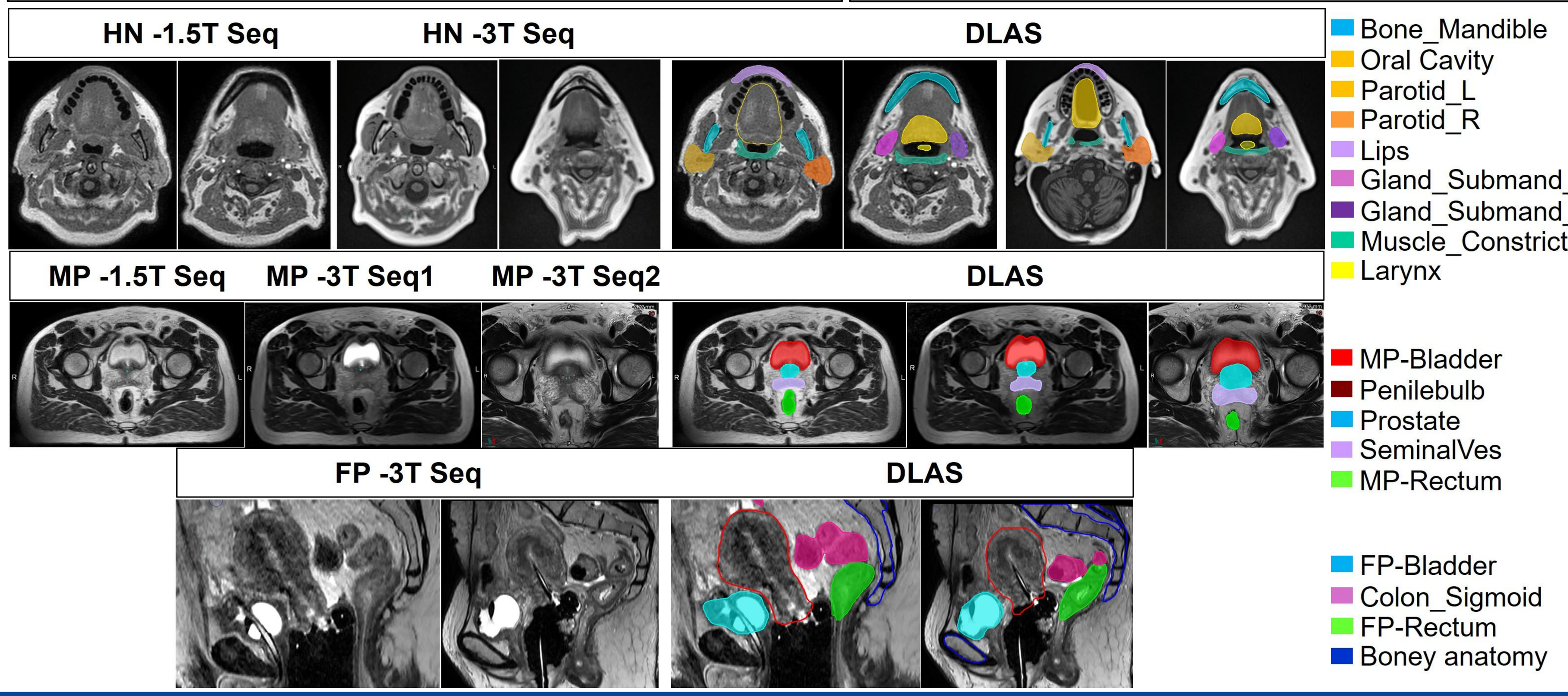


Figure 4. Slice based description of segmentation on HN, MP, FP MRI.

Table 4. 3D Mutil-accuracy metrics for female (3) and male pelvis (5 OARs).

	V-Diff	DSC	MDA	HD95	sDSC	APL_rel
Bladder	-0.24	0.77	2.39	7.7	0.56	60.5
Colon_Sigmoid	-0.35	0.58	6.15	22.46	0.54	159
Rectum	1.33	0.39	6.95	23.05	0.51	17.3
Bladder	-0.14	0.9	3.42	13.34	0.83	7.6
PenileBulb	0	0.79	1.62	4.39	0.7	28.5
Prostate	0.05	0.89	1.68	4.34	0.69	14.2
Rectum	-0.06	0.8	4.94	26.29	0.67	19.4
SeminalVes	0.04	0.79	1.57	4.75	0.72	28.6

DISCUSSION

- ❖Rectum segmentation accuracy can improve 1-28% when DLAS is edited to match clinical rectum length, for both male and female pelvis, Figure 5.
- ❖While the commercial models are a good starting point for clinical segmentation needs, clinic-specific, MRI sequences trained DLAS can further improve/decrease segmentation accuracy/editing, Figure 6.
- ❖A few limitation of the study include, 1) MRI sequence dependence. 2) inaccurate bowel segmentation (DSC average 0.58). 3) incomplete segmentation of small/long thin structures e.g., optic nerves in HN). 4) inaccurate labeling of gauze packing/ring/ovoid applicator as rectum in female pelvis MRI.

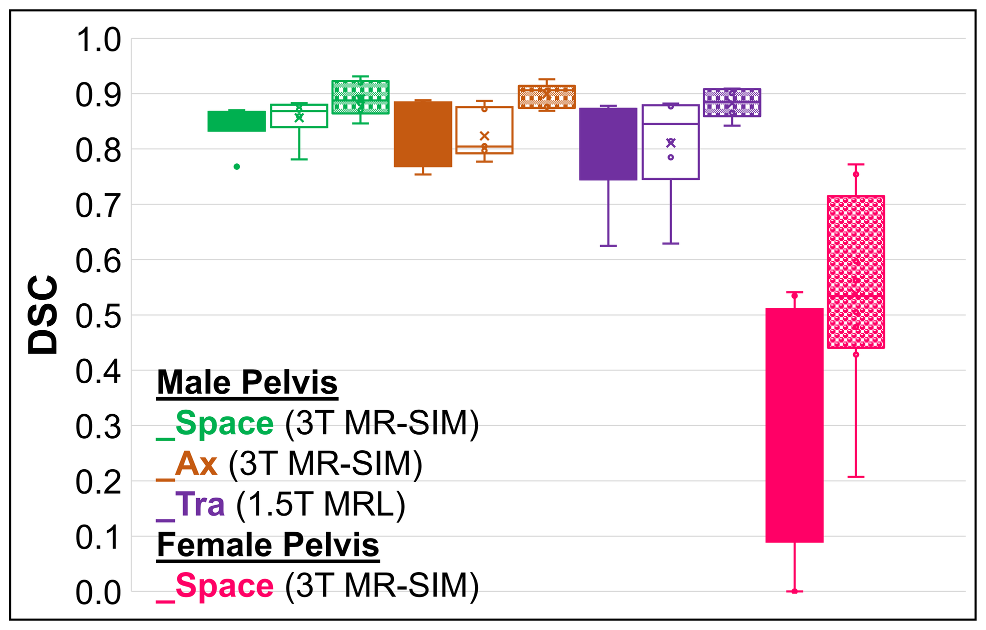
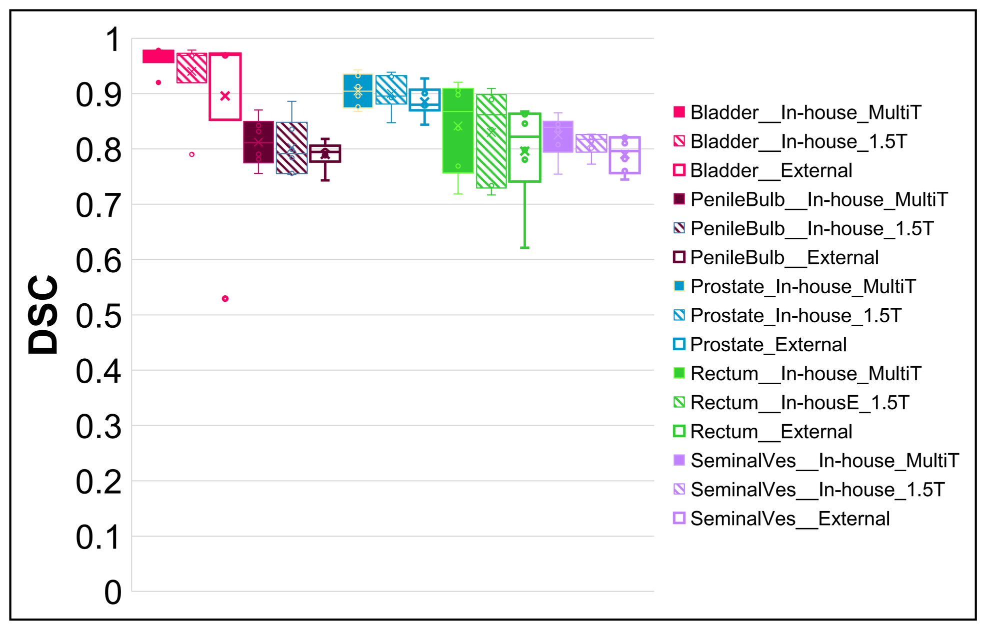


Figure 5. Box plots of rectum segmentation as function of clinical length. Left to right box plots; MP as is segmentation, MP Inferior border matching only, MP superior-inferior length matching, FP as is segmentation, FP superior-inferior length matching.

Figure 6. Box plots of male pelvis segmentation on T2 Tra MRI as function of in-house trained DLAS vs commercial/external DLAS.



CONCLUSIONS

The study demonstrate that for MR only planning clinical workflows, commercial-vendor trained MRI based DLAS models trained with non-clinic-specific MRI sequences can generate reasonable contours, for major head neck and female pelvis organs at risk. Whereas highly accurate segmentation can be achieved for multi-sequence MRI for male pelvis organs at risk.