


ORIGINAL ARTICLE

Open Access



# Automated pelvic MRI measurements associated with urinary incontinence for prostate cancer patients undergoing radical prostatectomy

Ingeborg van den Berg<sup>1,2\*</sup> , Robert N. Spaans<sup>2,3</sup>, Frank J. Wessels<sup>4</sup>, Erik J. R. J. van der Hoeven<sup>5</sup>, Charlotte J. Tutein Nolthenius<sup>5</sup>, Roderick C. N. van den Bergh<sup>2</sup>, Jochem R. N. van der Voort van Zyp<sup>1</sup>, Cornelis A. T. van den Berg<sup>1</sup> and Harm H. E. van Melick<sup>2</sup>

## Abstract

**Background** Pelvic morphological parameters on magnetic resonance imaging (MRI), such as the membranous urethral length (MUL), can predict urinary incontinence after radical prostatectomy but are prone to interobserver disagreement. Our objective was to improve interobserver agreement among radiologists in measuring pelvic parameters using deep learning (DL)-based segmentation of pelvic structures on MRI scans.

**Methods** Preoperative MRI was collected from 167 prostate cancer patients undergoing radical prostatectomy within our regional multicentric cohort. Two DL networks (nnU-Net) were trained on coronal and sagittal scans and evaluated on a test cohort using an 80/20% train-test split. Pelvic parameters were manually measured by three abdominal radiologists on raw MRI images and with the use of DL-generated segmentations. Automated measurements were also performed for the pelvic parameters. Interobserver agreement was evaluated using the intraclass correlation coefficient (ICC) and the Bland–Altman plot.

**Results** The DL models achieved median Dice similarity coefficient (DSC) values of 0.85–0.97 for coronal structures and 0.87–0.98 for sagittal structures. When radiologists used DL-generated segmentations of pelvic structures, the interobserver agreement for sagittal MUL improved from 0.64 (95% confidence interval 0.28–0.83) to 0.91 (95% CI 0.84–0.95). Furthermore, there was an increase in ICC values for the obturator internus muscle from 0.74 (95% CI 0.42–0.87) to 0.86 (95% CI 0.75–0.92) and for the levator ani muscle from 0.40 (95% CI 0.05–0.66) to 0.61 (95% CI 0.31–0.78).

**Conclusions** DL-based automated segmentation of pelvic structures improved interobserver agreement in measuring pelvic parameters on preoperative MRI scans.

**Relevance statement** The implementation of deep learning segmentations allows for more consistent measurements of pelvic parameters by radiologists. Standardized measurements are crucial for incorporating these parameters into urinary continence prediction models.

## Key points

- DL-generated segmentations improve interobserver agreement for pelvic measurements among radiologists.
- Membranous urethral length measurement improved from substantial to almost perfect agreement.

\*Correspondence:

Ingeborg van den Berg  
I.vandenBerg-8@umcutrecht.nl

Full list of author information is available at the end of the article




© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

- Artificial intelligence enhances objective pelvic parameter assessment for continence prediction models.

**Keywords** Artificial intelligence, Deep learning, Membranous urethral length, Prostate cancer, Urinary incontinence

### Graphical Abstract

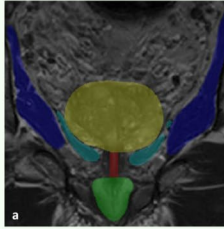
## Automated pelvic MRI measurements associated with urinary incontinence for prostate cancer patients undergoing radical prostatectomy



- Deep learning-generated segmentations improved interobserver agreement for pelvic measurements among radiologists.
- Membranous urethral length measurement improved from substantial to almost perfect agreement.
- AI enhances objective pelvic parameter assessment for continence prediction models.

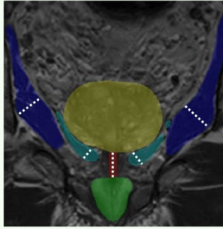
**Deep learning segmentations allow for more consistent measurements of pelvic parameters by radiologists. Standardized measurements are crucial for incorporating these parameters into urinary continence prediction models.**

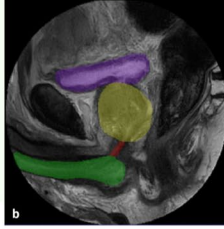
**Pelvic structures**



**a**

**Pelvic measurements**






**b**

*Pelvic structures measured: bladder (purple), corpus spongiosum (green), prostate (yellow), membranous urethra (red), obturator internus muscle (dark blue), and levator ani muscle (cyan)*

*Coronal direction (a); sagittal direction (b).*



Eur Radiol Exp (2023) van den Berg I, Spaans RN, Wessels FJ et al.  
DOI: 10.1186/s41747-023-00402-4

### Background

Urinary incontinence is a potential complication in patients with prostate cancer (PCa) undergoing robot-assisted radical prostatectomy (RARP). The 12-month postoperative prevalence of urinary leakage and its interference with daily life were reported by 36% and 17% of men, respectively [1]. Several patient-related variables such as age, body mass index, baseline urinary function, surgical factors, and magnetic resonance imaging (MRI)-based measurements have been found to be associated with post-prostatectomy continence recovery [2–4]. MRI-based measurements include parameters related to the prostate, the urethra, and the pelvic musculoskeletal system, such as the membranous urethral length (MUL), the levator ani muscle (LAM) thickness, and the intravesical prostatic protrusion length (IPPL). Among these measurements, the MUL has demonstrated the highest predictive value, with greater preoperative MUL serving as an independent prognostic factor for continence recovery [2, 5].

The measurement of the MUL is typically performed manually on T2-weighted prostate MRI. It involves determining the distance of the membranous urethra from the inferior border of the prostate apex to the superior border of the penile bulb in either the coronal or sagittal plane. However, manual assessment is prone to considerable interobserver disagreement, with varying intra-class correlation coefficient (ICC) values from 0.37 to 0.57 [6]. To improve interobserver agreement, Veerman et al. [7] investigated the use of standardized anatomical landmarks for MUL measurements, resulting in improved interobserver agreement with ICC values ranging from 0.63 to 0.84 among three radiologists within a single center.

Despite this improvement, Boellaard et al. [8] emphasized the need for artificial intelligence (AI) to further improve interobserver agreement through automated measurements of pelvic parameters. AI models can be used to segment relevant structures and improve the consistency and efficiency of radiologists on prostate MRI [9]. Small deviations in pelvic measurements can

have a significant impact on the predictive probability of continence recovery when incorporating them into urinary incontinence prediction models and can influence the treatment choices of PCa patients [10].

Therefore, our study aims to standardize pelvic floor measurements using AI-driven segmentations, enabling automated measurements of the prostate, membranous urethra, and pelvic musculoskeletal structures. We assessed interobserver agreement among radiologists using both manual assessment on raw MRI images and AI-aided visualization of relevant pelvic structures. Additionally, we compared the pelvic measures obtained through the fully automated workflow with those acquired using the manual approach.

**Methods**

**Data collection**

MRI data for this study was retrospectively extracted from a prospective registry (NCT04228211) where

patients had provided informed consent for data sharing. This prospective registry received approval from the institutional review and ethics board of the University Medical Center Utrecht (19–692/M). The study population included 167 PCa patients who underwent RARP within our regional multicentric prospective registry between March 2020 and November 2022. Exclusion criteria included patients with T4 tumors ( $n=0$ ). We collected preoperative 1.5-T or 3-T prostate MRI data from five Dutch hospitals within our region. Furthermore, MRI scans of three additional patients from three hospitals outside our region were also included in the study. We included the T2-weighted turbo-echo sequences in the coronal and sagittal directions. One sagittal MRI scan had to be excluded due to the presence of motion artifacts. The MRI sequence parameters for each regional hospital are listed in Table 1.

The following data were extracted from the Utrecht Prostate Cohort database: age at diagnosis, prostate-specific

**Table 1** MRI sequence parameters of coronal and sagittal T2-weighted imaging within our regional multicentric cohort

Sequence parameter	St. Antonius Hospital, Utrecht (n=61)	St. Antonius Hospital, Nieuwegein (n=46)	Diakonessenhuis (n=24)	UMCU (n=17)	Rivierenland (n=8)	Rivas Zorggroep (n=8)
MRI system	Siemens 3 T	Philips 3 T	Siemens 3 T	Philips 3 T	Siemens 1.5 T	Siemens 1.5 T
Scan orientation	Coronal (oblique), sagittal	Coronal (oblique), sagittal	Coronal, sagittal	Coronal, sagittal	Coronal (oblique), sagittal	Coronal, sagittal
TR (ms)						
Coronal	5,000	2,915	7,500	3,000–5,500	4,640	7,500
Sagittal	5,000	4,985	7,400	2,900–5,500	5,060	7,500
TE (ms)						
Coronal	103	95	104	140	108	108
Sagittal	103	95	104	110	87	108
Flip angle (°)						
Coronal	150	90	160	90	160	160
Sagittal	140	90	160	90	150	160
Slice thickness (mm)	3.0	3.0	3.0	3.0	3.0	3.5
Number of average						
Coronal	2	2	3	2	2	3
Sagittal	2	1	2	1	1	3
Number of slices						
Coronal	24	26	25	38	30	20
Sagittal	24	40	25	23	28	24
Reconstructed pixel size (mm × mm)						
Coronal	0.63 × 0.63	0.39 × 0.39	0.27 × 0.27	0.47 × 0.47	0.68 × 0.68	0.63 × 0.63
Sagittal	0.63 × 0.63	0.63 × 0.63	0.27 × 0.27	0.30 × 0.30	0.86 × 0.86	0.63 × 0.63
FOV (mm × mm)						
Coronal	200 × 200	200 × 200	199 × 199	180 × 180	220 × 220	200 × 200
Sagittal	200 × 200	180 × 180	199 × 199	260 × 260	220 × 220	200 × 200

MRI Magnetic resonance imaging, TR Repetition time, TE Echo time, FOV Field of view

antigen (PSA) at diagnosis, biopsy grade group, Prostate Imaging Data and Reporting System (PI-RADS) version 2 score, clinical stage, and the pathological stage after RARP (Table 2).

**Pelvic structures**

In the coronal plane, the delineated structures encompassed the prostate, corpus spongiosum, membranous urethra, left and right obturator internus muscle (OIM), and the left and right levator ani muscle (LAM). In the sagittal plane, the delineated pelvic structures included the prostate, bladder, corpus spongiosum, and the membranous urethra. The anatomical definition of the pelvic structures was first discussed in a group meeting with three abdominal radiologists (F.W., E.H., and C.T.N., with nine, nine, and seven years of experience, respectively) based on clinical relevance and current literature. The membranous urethra was defined to be delineated from the lower border of the peripheral zone of the prostate to the upper border of the corpus spongiosum. The lumen

of the urethra was delineated in the midsagittal and mid-coronal slices for each MRI scan.

All MRI scans were randomly divided into a training and test cohort at an 8:2 ratio after stratification by MRI location to ensure a balanced representation in both cohorts. A single annotator (I.B.) performed the delineation of pelvic structures and controlled 70 scans of the training cohort. These 70 scans were divided into three subsets, with each subset being controlled by one radiologist. The feedback received from these radiologists was then integrated into the remaining cases of both the training and test cohorts.

**Model development**

Two three-dimensional convolutional neural networks (CNNs) were employed for training with a fivefold cross-validation for the coronal and sagittal directions, enabling automated segmentation of pelvic structures. We selected the state-of-the-art network architecture nnU-Net due to its automated pipeline that encompasses preprocessing to postprocessing. nnU-Net is an open-source model and has been successfully applied on many MRI-based anatomical sites [11], including small elongated neurovascular structures on prostate MRI [12]. Both nnU-Net models were trained for 1,000 epochs with a batch size of 2, utilizing a patch size of 24×256×256 and pooling operations of 2 along the first axis and 6 along the second and third axes.

**Pelvic measurements**

Post-processing steps were performed to automate pelvic measurements on T2-weighted coronal and sagittal scans using a set of functions in Python 3.9, collecting the distances in millimeters (mm). The pelvic structures and measurements are shown in Fig. 1.

1. Coronal MUL: The largest segmentation in coronal plane was selected to determine the coronal MUL. We located the centroid of the segmentations in each axial slice and generated a line positioned at the center of the MUL for each axial slice to calculate the length in coronal direction. In cases where the membranous urethra was divided into two parts across adjacent slices, with one part connected to the prostate and the other part connected to the corpus spongiosum, both parts were considered for calculating the MUL.
2. Sagittal MUL: The largest segmentation in sagittal plane was selected to determine the sagittal MUL. The length was calculated with the same post-processing steps as for the coronal MUL.
3. IPPL: The length of the intravesical prostatic protrusion (IPPL) was determined by finding the intersection between the bladder and the prostate. This intersection was obtained by performing a one-pixel

**Table 2** Patient characteristics

Characteristic	Training cohort (n = 134)	Testing cohort (n = 33)
Age (years), median (IQR)	69 (65–72)	68 (59–72)
PSA (ng/mL), median (IQR)	7.1 (5.1–12.7)	8.1 (5.3–12.9)
Clinical T stage, n (%)		
T1	66 (49.3)	18 (54.5)
T2	49 (36.6)	10 (30.3)
T3	17 (12.7)	3 (9.1)
Unknown	2 (1.5)	2 (6.1)
Biopsy Gleason Grade Group, n (%)		
1	18 (13.4)	3 (9.1)
2	55 (41.0)	15 (45.5)
3	28 (20.9)	7 (21.2)
≥ 4	33 (24.6)	8 (24.2)
PI-RADS version 2 score, n (%)		
2	6 (4.5)	2 (6.1)
3	7 (5.2)	0 (0.0)
4	52 (38.8)	16 (48.5)
5	67 (50.0)	15 (45.5)
Not reported	2 (1.5)	0 (0.0)
D’Amico risk group, n (%)		
Low-risk	13 (9.7)	3 (9.1)
Intermediate-risk	70 (52.2)	19 (57.6)
High-risk	51 (38.1)	11 (33.3)
Pathological T stage, n (%)		
T2	85 (63.4)	20 (60.6)
T3	48 (35.8)	13 (39.4)
T4	1 (0.7)	0 (0.0)

PSA Prostate-specific antigen, IQR Interquartile range, PI-RADS Prostate Imaging Data and Reporting System

dilation of the bladder. A line was then formed connecting the outermost points of the intersection. The IPPL was computed as the maximum perpendicular distance between this line and the contour of the prostate.

4. LAM and OIM thickness: The maximum thickness of the left and right LAM and OIM was determined by compiling a list of the shortest distances between each point of the muscles' centerline and its border in the coronal direction. The maximum distance from this list was selected and doubled to obtain the muscles' diameter.

### Model evaluation

The DL-generated segmentations of the pelvic structures were evaluated on the test cohort using volumetric Dice similarity coefficient (DSC), mean surface distances (MSD), and Hausdorff distances (HD95), utilizing the DeepMind Python package (<https://github.com/deepmind/surface-distance>). Volumetric DSC measures spatial overlap, MSD represent average distances, and HD95 capture maximum surface distances (95th percentile).

The pelvic MRI parameters (*i.e.*, coronal MUL, sagittal MUL, IPPL, OIM thickness and LAM thickness) were evaluated on the test cohort in two sessions by three abdominal radiologists (F.W., E.H., C.T.N.). The first session comprised manual measurements on raw MRI images in coronal and sagittal direction and the second session involved manual measurements on MRI images with DL-generated segmentations of pelvic structures in ITK-SNAP (version 3.6.0; <http://www.itksnap.org>). The evaluation of the second session was conducted at least seven days after the first evaluation, and the patient order was mixed to avoid recall bias. The radiologists were blinded to patients' clinical data.

The absolute interobserver agreement for each measurement was determined by applying a two-way random effect analysis with the intraclass correlation coefficient (ICC). ICC values  $\leq 0$  indicate no agreement, 0.01–0.20 as none to slight agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1.00 as almost perfect agreement [13]. The interobserver agreement for each measurement was visualized in Bland–Altman plots between the manual and the AI-aided measurements (Fig. 2). Non-normally distributed data were presented as median with interquartile range (IQR). Wilcoxon-signed rank tests were conducted and  $p < 0.050$  was considered statistically significant.

## Results

### Segmentation performance

Table 3 presents the DL segmentation performances in the coronal and sagittal planes. The median DSC values ranged from 0.85 to 0.97 for coronal structures and from 0.87 to 0.98 for sagittal structures. The highest DSC value was observed for the bladder (DSC 0.98; IQR 0.97–0.98), while the lowest DSC value was found for the membranous urethra (DSC 0.85; IQR 0.76–0.92) in the coronal direction. The median MSD values ranged from 0.10 to 0.31 mm, with the lowest value observed for the membranous urethra in the sagittal direction (MSD 0.10 mm; IQR 0.05–0.14). Additionally, the DL network prediction time for the sagittal test cohort was 2 min and 40 s (approximately 5 s per scan) and for the coronal test cohort 3 min and 5 s (approximately 6 s per scan).

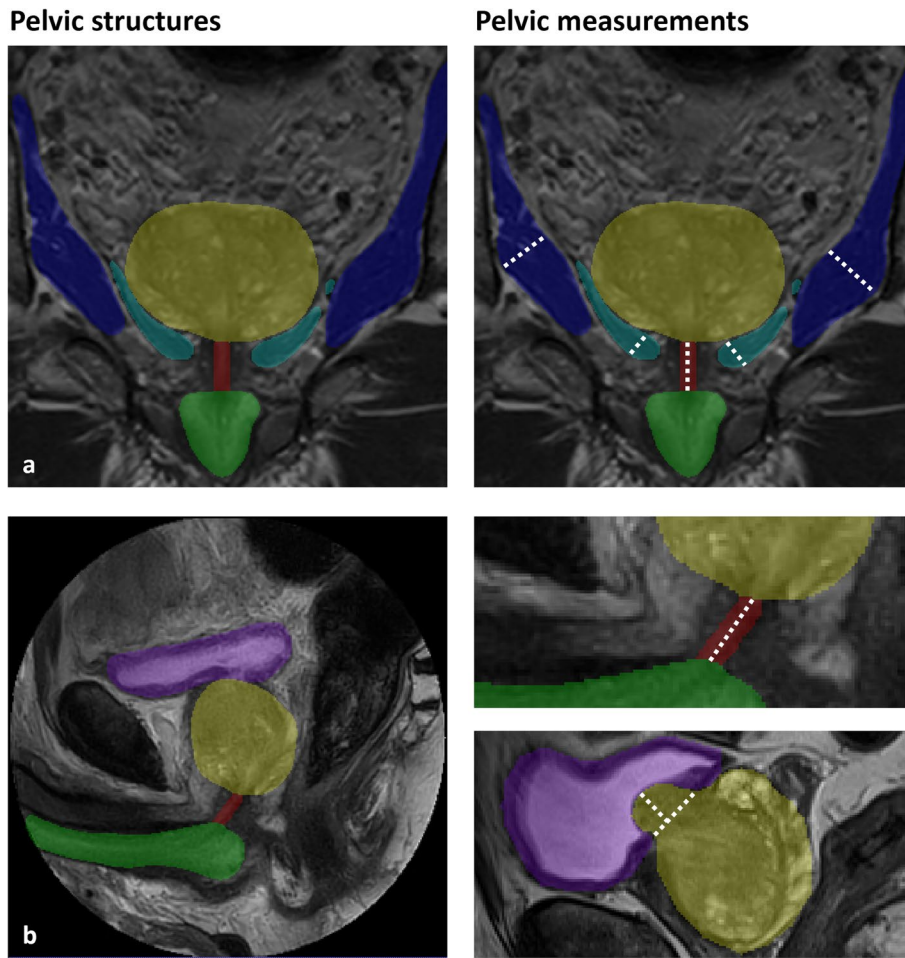
### Pelvic measurements

The pelvic measurements are presented in Table 4 for the three different measurement approaches: manual, AI-aided, and automated. The automated approach resulted in significantly higher values for the IPPL, LAM thickness, and OIM thickness ( $p < 0.001$ ) and showed comparable values for the coronal MUL ( $p = 0.091$ ) and sagittal MUL ( $p = 0.606$ ) in comparison to the AI-aided approach. Sagittal MUL was highest for the manual approach (14.94 mm; IQR 12.08–17.73), followed by the automated approach (14.32 mm; IQR 12.50–17.12), and the AI-aided approach (14.04; IQR 11.63–16.97). Coronal MUL measurements were found to be lower, with a median value of 13.78 mm for the manual and AI-aided approach, and 14.06 for the automated approach. The median LAM thickness ranged from 8.56 to 11.14 mm, and the OIM thickness ranged from 17.92 to 19.77 mm, with the automated approach yielding the highest median values for both measurements.

### Interobserver agreement

The ICC values ranged from 0.40 to 0.88 for the manual approach and from 0.61 to 0.91 for the AI-aided approach (Table 4). The highest interobserver agreement was observed for sagittal MUL (ICC 0.91, 95% CI 0.84–0.95) with the AI-aided approach, indicating almost perfect agreement. The manual approach showed substantial agreement (ICC 0.64, 95% CI 0.28–0.83). The lowest interobserver agreement was found for LAM thickness, with the manual approach demonstrating moderate agreement (ICC 0.40, 95% CI 0.05–0.66) and the AI-aided approach showing substantial agreement (ICC 0.61, 95% CI 0.31–0.78). A Bland–Altman plot of sagittal MUL is presented in Fig. 2, showing the median





**Fig. 1** Pelvic structures and pelvic measurements in coronal direction (a) and sagittal direction (b) on T2-weighted images. Pelvic structures include the bladder (purple), corpus spongiosum (green), prostate (yellow), membranous urethra (red), obturator internus muscle (dark blue), and levator ani muscle (cyan)

**Table 3** Deep learning segmentation performances

	DSC	MSD [mm]	HD95 [mm]	Volume [cc]
Coronal structures				
Prostate	0.97 (0.96–0.97)	0.22 (0.16–0.32)	1.25 (0.87–1.75)	40.93 (36.30–54.02)
Urethra	0.85 (0.76–0.92)	0.14 (0.07–0.32)	1.25 (0.60–3.01)	0.20 (0.17–0.24)
Corpus spongiosum	0.96 (0.96–0.97)	0.10 (0.09–0.12)	0.60 (0.42–0.63)	9.73 (8.04–12.47)
OIM	0.95 (0.92–0.96)	0.25 (0.15–0.36)	1.25 (0.63–3.00)	38.17 (33.80–43.21)
LAM	0.90 (0.86–0.91)	0.24 (0.17–0.33)	2.99 (0.88–3.00)	5.92 (5.08–6.83)
Sagittal structures				
Bladder	0.98 (0.97–0.98)	0.15 (0.11–0.23)	0.63 (0.63–1.20)	100.64 (73.47–148.26)
Prostate	0.96 (0.94–0.97)	0.31 (0.22–0.47)	1.25 (1.12–1.90)	37.23 (32.52–46.61)
Urethra	0.87 (0.77–0.90)	0.10 (0.05–0.14)	0.82 (0.63–1.88)	0.21 (0.18–0.24)
Corpus spongiosum	0.94 (0.90–0.95)	0.22 (0.16–0.31)	1.88 (0.88–3.00)	10.84 (9.57–13.04)

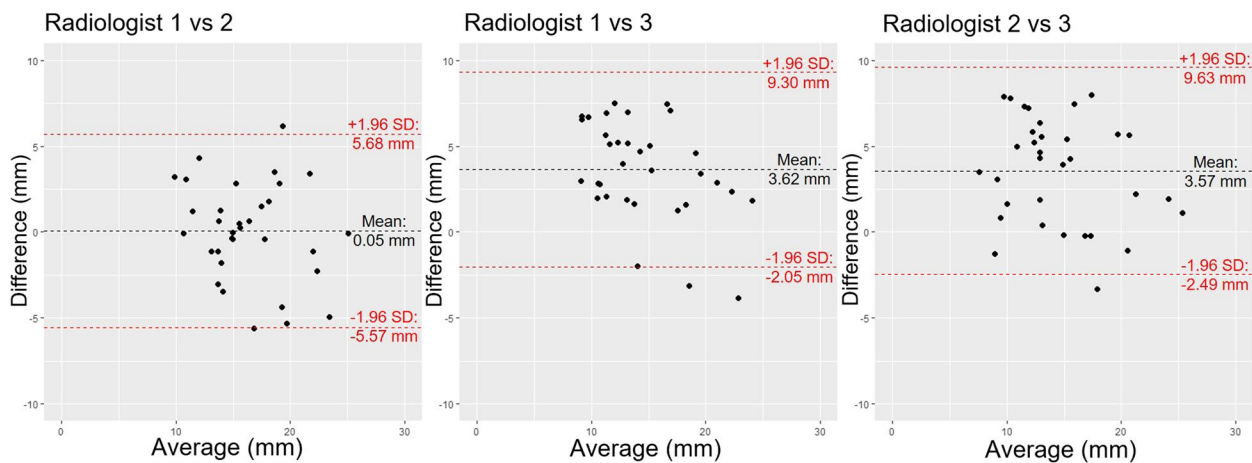
For the LAM and IOM: left and right are combined. Data are expressed as median (interquartile range). DSC Dice similarity coefficient, MSD Mean surface distance, HD95 95% boundary Hausdorff distance, LAM Levator ani muscle, OIM Obturator internus muscle

**Table 4** Comparison of pelvic measurements on T2-weighted MRI: manual, AI-aided and automated workflows with intraclass correlation coefficients

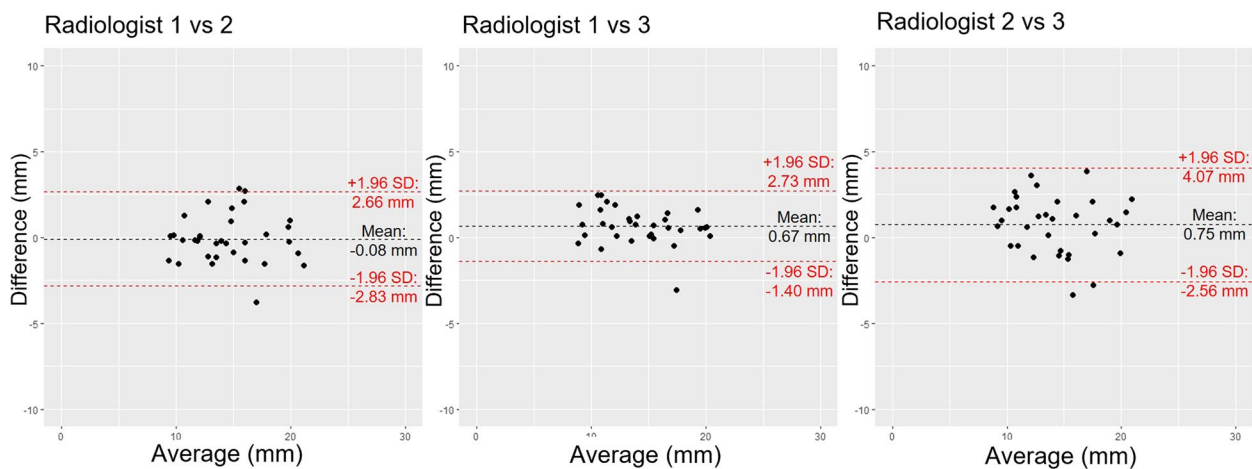
Measurements	Manual [mm]	AI-aided [mm]	Automated [mm]	ICC (Manual)	ICC (AI-aided)
Coronal MUL	13.78 (11.55–16.10)	13.78 (12.20–15.67)	14.06 (13.12–16.41)	0.69 (0.51–0.82)	0.90 (0.82–0.94)
Sagittal MUL	14.94 (12.08–17.73)	14.04 (11.63–16.97)	14.32 (12.50–17.12)	0.64 (0.28–0.83)	0.91 (0.84–0.95)
IPPL	0.00 (0.00–0.00)	0.00 (0.00–0.00)	2.40 (1.09–3.26)	0.88 (0.81–0.94)	0.86 (0.77–0.93)
LAM thickness	8.75 (7.27–10.08)	8.56 (7.35–10.05)	11.14 (10.18–12.31)	0.40 (0.05–0.66)	0.61 (0.31–0.78)
OIM thickness	18.05 (16.23–20.28)	17.92 (15.96–20.31)	19.77 (17.93–21.93)	0.74 (0.42–0.87)	0.86 (0.75–0.92)

For the LAM and OIM: left and right are combined. Data are expressed as median (interquartile range). AI Artificial intelligence, ICC Intraclass correlation coefficient, MUL Membranous urethral length, IPPL Intravesical prostatic protrusion length, LAM Levator ani muscle, OIM Obturator internus muscle

### Manual assessment sagittal MUL



### AI-aided assessment sagittal MUL



**Fig. 2** Bland–Altman plots of measured membranous urethral length (MUL) in the sagittal direction for manual assessment and artificial intelligence (AI)-aided assessment. The black dashed line represents the mean difference and the red dashed lines represent the upper and lower 95% limits of agreement

difference between the manual and AI-aided approaches. The Bland–Altman plots of the other pelvic measurements are shown in Supplementary Figs. 1–4.

### Discussion

To our knowledge, this is the first study that developed and evaluated an automated workflow to measure pelvic floor

parameters associated with post-prostatectomy incontinence. The implementation of DL-generated segmentations of pelvic structures resulted in higher interobserver agreement among radiologists. Both coronal and sagittal MUL measurements demonstrated improved interobserver agreement, with the coronal MUL increasing from an ICC value of 0.69 (95% CI 0.51–0.82) to 0.90 (95% CI 0.82–0.94) and the sagittal MUL improving from 0.64 (95% CI 0.28–0.83) to 0.91 (95% CI 0.84–0.95), indicating a shift from substantial to almost perfect agreement. Veerman et al. [7] reported a comparable ICC value of 0.63 (95% CI 0.28–0.81) for the sagittal MUL measurements on raw images but achieved an improved ICC value of 0.84 (95% CI 0.66–0.91) with the implementation of consistent anatomical definitions. In our study, DL-generated segmentations resulted in a higher ICC value of 0.91 (95% CI 0.84–0.95), demonstrating the potential of AI to further improve consistency in the measurements.

Additionally, the use of DL-generated segmentations improved interobserver agreement for the musculoskeletal structures. The OIM thickness had an ICC value of 0.74 (95% CI 0.42–0.87) for the manual assessment and 0.86 (95% CI 0.75–0.92) for the AI-aided assessment, indicating an improvement from substantial agreement to almost perfect agreement. The LAM thickness showed moderate agreement for the manual approach (ICC 0.40; 95% CI 0.05–0.66) and improved to substantial agreement for the AI-aided approach (ICC 0.61; 95% CI 0.31–0.78). However, IPPL measurements demonstrated no improvement with DL-generated segmentations, with ICC values of 0.88 (95% CI 0.81–0.94) and 0.86 (95% CI 0.77–0.93) for the manual and AI-aided approach, respectively.

In comparison with the manual and AI-aided approaches, the automated approach yielded significantly higher values for the LAM thickness, OIM thickness, and the IPPL. The higher automated muscle thickness could be explained by the fact that the segmentations around the muscle borders were slightly larger in some cases. In addition, variations in muscle diameters may be attributed to the computer's capability to accurately compute the maximum diameter, while variations in angle or slice positioning during manual measurements by radiologists may affect the measured diameter. In future studies, it may be beneficial to present the computed thickness to radiologists, enabling them to verify the measurements and potentially minimize the need for manual measurements. The automated IPPL measurements also demonstrated higher values compared to the manual and AI-aided measurements, often resulting in an overestimation of the protrusion. Consequently, manual assessments of DL-generated segmentations remain important, especially in cases where segmentation errors may arise.

The automated sagittal and coronal MUL measurements were not statistically different from the AI-aided measurements. The automated sagittal MUL measurements were 14.32 mm, the AI-aided measurements were 14.04 mm, and the manual measurements were 14.94 mm. The larger manual MUL measurements can be attributed to the different positioning of the superior point of the membranous urethra. During the first session, two radiologists defined the superior point in some patients differently, placing it more superiorly where the urethra was still visible in the caudal portion of the prostate apex. This superior point may align more closely with the surgical section during RARP, where urologists aim to preserve the membranous urethra for the vesicourethral anastomosis [14], except in cases involving apical tumors. In the second session, the superior point was consistently positioned at the inferior border of the prostate due to the DL-generated segmentations, which contributed to increased consistency as indicated by the AI-aided measurements. This definition of the MUL measurement was comparable to previous studies [3, 6, 7].

The MUL measurements in our study were comparable with Kim et al. [3], who reported mean values of 14.6 mm in the coronal direction and 14.2 mm in the sagittal direction. Veerman et al. [7] found higher MUL measurements of 17 mm, which could potentially be attributed to the inclusion of patients with higher MUL values since the introduction of a risk prediction model of urinary incontinence in their institution [15]. LAM measurements were comparable to the findings of Sadahira et al. [16] but higher than the measurements reported by Muñoz-Calahorra et al. [6], which ranged between 4.22 and 6.87 mm. These differences in measurements could be attributed to variances in anatomical definition, as Muñoz-Calahorra et al. [6] measured the thickness before the insertion of the muscle puborectalis fibers, while we measured the largest thickness regardless of its location.

More consistent pelvic measurements are especially clinically relevant when including them in urinary incontinence prediction models for patients undergoing RARP. Multiple studies have already created prediction models for continence outcome after RARP and incorporated patient-related variables and MRI-based variables like the MUL [10, 17–19]. Standardized measurements are essential since the continence prediction tool described in Tillier et al. [10] demonstrates that a 1-mm increase in the MUL corresponds to an average increase of 8.1 percentage points in the chance of continence recovery after six months of RARP [7]. These findings can influence the treatment choices of PCa patients. Therefore, standardizing prostate MRI measurements using DL is an unmet need for continence prediction models. Additionally, DL-generated



segmentations provide an opportunity to assess additional geometric variables, including shape-based features, surface areas, and volume-based features.

This study had some limitations. First, not all training cases were verified by three radiologists but a subset of 70 patients was evaluated and the feedback of the three different radiologists was used. Second, the manual assessment of the pelvic measurements was performed in sagittal and coronal plane. It was not possible to visualize the axial, coronal, and sagittal planes simultaneously. The visualization of the prostate in three directions would have enhanced the ease of evaluating MUL measurements. Third, external validation of our DL model and automated measurements in other centers is essential because only a part of the T2-weighted images adhered to the PI-RADSv2 MRI acquisition parameter guideline [20]. We believe that our DL model has the potential for generalizability, as it has been trained using data from five regional hospitals with various magnetic field strengths, vendors, and sequence protocols. Additional training cases with variations in anatomy, such as more patients with IPP and patients with hip prostheses, could contribute to improving the generalization ability of our DL model. Finally, the clinical impact of AI-aided assessment of pelvic parameters could not be retrospectively assessed. Future studies are needed to evaluate the relationship between objective pelvic measurements and postoperative outcomes.

In conclusion, our study demonstrates the potential of deep learning (DL)-generated segmentations to improve interobserver agreement of pelvic measurements among radiologists. The membranous urethral length (MUL) measurement improved from substantial agreement to almost perfect agreement among radiologists with DL-generated segmentations of pelvic structures. Standardized pelvic measurements can be incorporated into urinary incontinence prediction models for patients undergoing RARP.

#### Abbreviations

AI	Artificial intelligence
CI	Confidence interval
DL	Deep learning
DSC	Dice similarity coefficient
ICC	Intraclass correlation coefficient
IPPL	Intravesical prostatic protrusion length
IQR	Interquartile range
LAM	Levator ani muscle
MRI	Magnetic resonance imaging
MSD	Mean surface distances
MUL	Membranous urethral length
OIM	Obturator internus muscle
PCa	Prostate cancer
RARP	Robot-assisted radical prostatectomy

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41747-023-00402-4>.

**Additional file 1: Supplementary Figure 1.** Bland-Altman plots of measured membranous urethral length (MUL) in the coronal direction for manual assessment and AI-aided assessment. **Supplementary Figure 2.** Bland-Altman plots of measured intravesical prostatic protrusion length (IPPL) in the sagittal direction for manual assessment and AI-aided assessment. **Supplementary Figure 3.** Bland-Altman plots of measured obturator internus muscle (OIM) thickness in the coronal direction for manual assessment and AI-aided assessment. **Supplementary Figure 4.** Bland-Altman plots of measured levator ani muscle (LAM) thickness in the coronal direction for manual assessment and AI-aided assessment.

#### Authors' contributions

IB trained and evaluated the deep learning models. IB and RN developed the post-processing steps. FW, EH, and CN performed a clinical evaluation of the deep learning models. All authors read and approved the final manuscript.

#### Funding

The authors state that this work has not received any funding.

#### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

This study received approval from the institutional review and ethics board of the University Medical Center Utrecht (19–692/M) and patients provided informed consent for sharing their data within the Utrecht Prostate Cohort (UPC) study (NCT04228211).

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Department of Radiation Oncology, Division of Imaging & Oncology, University Medical Center Utrecht, Utrecht, The Netherlands. <sup>2</sup>Department of Urology, St. Antonius Hospital, Nieuwegein, Utrecht, The Netherlands. <sup>3</sup>Technical Medicine, University of Twente, Enschede, The Netherlands. <sup>4</sup>Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands. <sup>5</sup>Department of Radiology, St. Antonius Hospital, Nieuwegein, Utrecht, The Netherlands.

Received: 31 August 2023 Accepted: 23 October 2023

Published online: 02 January 2024

#### References

1. Lane JA, Donovan JL, Young GJ et al (2022) Functional and quality of life outcomes of localised prostate cancer treatments (Prostate Testing for Cancer and Treatment [ ProtecT ] study). *BJU Int* 130:370–380. <https://doi.org/10.1111/bju.15739>
2. van Dijk-de Haan MC, Boellaard TN, Tissier R et al (2022) Value of different magnetic resonance imaging-based measurements of anatomical structures on preoperative prostate imaging in predicting urinary continence after radical prostatectomy in men with prostate cancer: a systematic review and meta-analysis. *Eur Urol Focus* 8:1211–1225. <https://doi.org/10.1016/j.euf.2022.01.015>

3. Kim LHC, Patel A, Kinsella N, Sharabiani MTA, Ap Dafydd D, Cahill D (2020) Association between preoperative magnetic resonance imaging–based urethral parameters and continence recovery following robot-assisted radical prostatectomy. *Eur Urol Focus* 6:1013–1020. <https://doi.org/10.1016/j.euf.2019.01.011>
4. Lardas M, Grivas N, Debray TPA et al (2022) Patient- and tumour-related prognostic factors for urinary incontinence after radical prostatectomy for nonmetastatic prostate cancer: a systematic review and meta-analysis. *Eur Urol Focus* 8:674–689. <https://doi.org/10.1016/j.euf.2021.04.020>
5. Lamberg H, Shankar PR, Singh K et al (2022) Preoperative prostate MRI predictors of urinary continence following radical prostatectomy. *Radiology* 303:99–109. <https://doi.org/10.1148/RADIOL.210500>
6. Muñoz-Calahorra C, Parada-Blázquez MJ, García-Sánchez C, López-Arellano L, Vizcaíno-Velázquez P, Medina-López RA (2023) Inter-observer variability in male pelvic-floor MRI measurements that might predict post-prostatectomy incontinence. *World J Urol* 41:1147–1155. <https://doi.org/10.1007/s00345-023-04320-3>
7. Veerman H, Hagens MJ, Hoeks CM et al (2022) A standardized method to measure the membranous urethral length (MUL) on MRI of the prostate with high inter- and intra-observer agreement. *Eur Radiol* 33:3295–3302. <https://doi.org/10.1007/s00330-022-09320-2>
8. Boellaard TN, Hagens MJ, Veerman H et al (2023) Prostate MRI for improving personalized risk prediction of incontinence and surgical planning: the role of membranous urethral length measurements and the use of 3D models. *Life* 13:830. <https://doi.org/10.3390/life13030830>
9. Liu G, Pan S, Zhao R et al (2023) The added value of AI-based computer-aided diagnosis in classification of cancer at prostate MRI. *Eur Radiol* 33:5118–5130. <https://doi.org/10.1007/s00330-023-09433-2>
10. Tillier CN, Vromans RD, Boekhout AH et al (2021) Individual risk prediction of urinary incontinence after prostatectomy and impact on treatment choice in patients with localized prostate cancer. *Neurourol Urodyn* 40:1550–1558. <https://doi.org/10.1002/nau.24703>
11. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 18:203–211. <https://doi.org/10.1038/s41592-020-01008-z>
12. van den Berg I, Savenije MHF, Teunissen FR et al (2023) Deep learning for automated contouring of neurovascular structures on magnetic resonance imaging for prostate cancer patients. *Phys Imaging Radiat Oncol* 26:100453. <https://doi.org/10.1016/j.phro.2023.100453>
13. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174. <https://doi.org/10.2307/2529310>
14. Schlomm T, Heinzer H, Steuber T et al (2011) Full functional-length urethral sphincter preservation during radical prostatectomy. *Eur Urol* 60:320–329. <https://doi.org/10.1016/j.eururo.2011.02.040>
15. Grivas N, van der Roest R, Schouten D et al (2018) Quantitative assessment of fascia preservation improves the prediction of membranous urethral length and inner levator distance on continence outcome after robot-assisted radical prostatectomy. *Neurourol Urodyn* 37:417–425. <https://doi.org/10.1002/nau.23318>
16. Sadahira T, Mitsui Y, Araki M et al (2019) Pelvic magnetic resonance imaging parameters predict urinary incontinence after robot-assisted radical prostatectomy. *Low Urin Tract Symptoms*. 11:122–126. <https://doi.org/10.1111/luts.12245>
17. Gu Z, Zheng Z, Zhang W et al (2023) The development and assessment of a predicting nomogram for the recovery of immediate urinary continence following laparoscopic radical prostatectomy. *Front Surg* 9:1071093. <https://doi.org/10.3389/fsurg.2022.1071093>
18. Zhang F, Chu H, Hao Y et al (2023) Preoperative predictive model of early urinary continence recovery after laparoscopic radical prostatectomy. *World J Urol* 41:59–65. <https://doi.org/10.1007/s00345-022-04198-7>
19. Onishi T, Sekito S, Kajiwara S, Shibahara T (2022) Simple and reliable predictive factor for early recovery of urinary continence after non-nerve-sparing robot-assisted laparoscopic radical prostatectomy. *J Robot Surg* 17:1077–1083. <https://doi.org/10.1007/s11701-022-01516-4>
20. Cuocolo R, Stanzione A, Ponsiglione A et al (2019) Prostate MRI technical parameters standardization: a systematic review on adherence to

PI-RADSV2 acquisition protocol. *Eur J Radiol* 120:108662. <https://doi.org/10.1016/j.ejrad.2019.108662>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---