

Original Research Article

Automatic segmentation of the urethra and prostate zones with deep learning on T2-weighted magnetic resonance imaging



William Holmlund^{a,*}, Attila Simkó^a, Karin Söderkvist^a, Péter Palásti^b, Szilvia Tótin^b, Kamilla Kalmár^b, Zsófia Domoki^b, Zsuzsanna Fejes^b, Tamás Z. Kincses^b, Patrik Brynolfsson^{a,c}, Tufve Nyholm^a

^a Umeå University, Department of Diagnostics and Intervention, Umeå, Sweden

^b University of Szeged, Albert Szent-Györgyi Medical School, Department of Radiology, Szeged, Hungary

^c Skåne University Hospital, Department of Haematology, Oncology and Radiation Physics, Lund, Sweden

ARTICLE INFO

Keywords:

Automatic Segmentation

Urethra

Prostate Zones

ABSTRACT

Background and purpose: Accurate segmentation of the urethra is crucial for safe focal dose escalated radiotherapy, while prostate zone identification is important for prostate cancer diagnosis. Manual delineations on magnetic resonance imaging (MRI) are labour-intensive and variable, and while deep learning offers promise in automating this process, no available solution currently exists. This study aimed to develop and evaluate a deep learning model for automatic segmentation of the urethra, prostate and all prostate zones and benchmark its performance against inter-reader variability and assess generalisability to external data from a different MRI vendor.

Materials and methods: The public datasets ProstateZones and PROSTATEx included 200 magnetic resonance images with manual delineations, with 160 used for training/validation and 40 with independent duplicate segmentations used as a test set. A nnU-Net deep learning model was evaluated on the unseen test set and externally validated on a dataset with 55 samples. Performance was assessed using Dice Similarity Coefficient (DSC), Surface DSC, percentile Symmetric Surface Distance, and Center Line Distance (CLD) metrics.

Results: The model outperformed the inter-reader variability on multiple structures, and notably on all metrics for the urethra, with median CLD values of 2.8 and 2.9 mm compared to 3.6 mm for inter-reader variability. External validation showed robust generalisability to a dataset collected from a different vendor.

Conclusions: This study demonstrated that a deep learning model can achieve expert-level performance in automated segmentation of the urethra, prostate, and prostate zones. Robust performance on external data highlighted potential as a decision support solution.

1. Introduction

Prostate cancer (PCa) is the most prevalent malignancy in males worldwide [1], requiring accurate diagnostic and treatment approaches. Magnetic resonance imaging (MRI) is recommended for diagnosis of clinically significant PCa [2], with the global prostate imaging reporting and data system (PI-RADS) guidelines aiming to improve standardisation and interpretation [3]. The guidelines specify that the assessment of prostatic lesions on multiparametric MRI (mpMRI) should use different primary sequences based on their zonal location. These

anatomical zones of the prostate are defined as the peripheral zone (PZ), central zone (CZ), transitional zone (TZ), and anterior fibromuscular stroma (AFS), each presenting different characteristics and histological features [4]. Additionally, the guidelines recommend reporting prostate specific antigen density (PSAd), which has been linked to the presence of clinically significant PCa [5]. Further, TZ-specific PSAd has been reported to improve discrimination of clinically significant PCa and reduce false-positive MRI findings [6]. Moreover, accurate identification of lesion location is important for targeted biopsies, which has been shown to reduce the number of unnecessary biopsies and reduce diagnosis of

Abbreviations: DSC, Dice Similarity Coefficient; SDSC, Surface Dice Similarity Coefficient; CLD, Center Line Distance; pSSD, percentile Symmetric Surface Distance; PZ, Peripheral Zone; CZ, Central Zone; TZ, Transitional Zone; AFS, Anterior Fibromuscular Stroma.

* Corresponding author.

E-mail address: william.holmlund@umu.se (W. Holmlund).

<https://doi.org/10.1016/j.phro.2026.100964>

Received 4 December 2025; Received in revised form 24 March 2026; Accepted 4 April 2026

Available online 7 April 2026

2405-6316/© 2026 The Authors. Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

clinically insignificant cancers [7,8].

Traditionally, radiotherapy of PCa has been delivered with a homogeneous dose to the entire prostate, although the efficacy of local dose escalation to a sub volume of the prostate has been demonstrated [9]. The risk of an increased dose to the prostatic urethra is a clinical concern when escalating the radiotherapy dose inside the prostate. A combined analysis by Leeman et al. [10] demonstrated a significant association between urethral dose and genitourinary (GU) toxicity, including acute grade ≥ 2 and late ≥ 3 GU toxicity. The risk for urethral side effects is mitigated by using dose constraints, hence requiring delineation of the urethral structure [11]. Moreover, the effects of a local dose escalation could potentially be enhanced by utilising zonal information to individualise treatment and stratify risk, as location of the cancer in different zones have different incidence, prognosis, and outcome, making treatment zonal-dependent instead of zonal-agnostic [12].

Manual delineations of these structures on MRI are tedious and time-consuming. Moreover, they are subject to inter- and intra-reader variability due to low image contrast and patient-specific differences in prostate anatomy [13]. Deep learning models have been well-documented in addressing challenges in medical image analysis overall [14,15], and particularly for segmentations of prostate MRI [16,17].

Despite the emerging clinical importance of accurately delineating the prostatic urethra, only a limited number of studies have been published. Multiple studies used models that relied on manual delineations with a Foley catheter in place [18–21]. However, the catheter impacts the anatomical configuration of the prostate and urethra, which makes the results unreliable for a non-catheter scenario [22,23]. MRI-based approaches offer a non-invasive alternative, leveraging its superior soft tissue contrast. Nonetheless, only a few studies have explored automatic segmentation of the urethra on MRI [24–27]. For all these studies, it is important to note that their methods were developed using proprietary datasets, and one study optimised image sequences for visualisation of the urethra [26]. Furthermore, all delineations were carried out by supervised medical students [25] or a single expert [24,26,27], and none of the models have been made publicly available, with only one using an external test set [27].

For the prostate zones, previous studies have investigated multiple variants of the U-Net architecture for segmentation of the prostate gland and zonal structures, with most differentiating only between PZ and non-PZ regions [28–38]. For example, Aldoj et al. [28] developed a Dense U-Net for automatic segmentation of the prostate, PZ and CZ. Bardis et al. [29] utilised a combination of three customised hybrid 3D/2D U-Nets to localise and segment the prostate and subsequently classify the PZ and TZ. Xu et al. [35] used a 3D U-Net model to segment the PZ and central gland (CG). Jimenez-Pastor et al. [33] employed a U-Net with deep supervision to automatically segment the prostate, PZ, a combined CZ-TZ, and seminal vesicles. All these studies segmented the prostate into sub-regions, described as PZ and non-PZ regions, with a clear inconsistency in terminology. This varying terminology and definition of delineated structures in the existing literature is a major limitation for clinical applicability, as was highlighted in a topical review [39]. Only a study by Meyer et al. [38], which delineated the PZ, CG (a combined delineation of CZ and TZ), AFS, and the distal prostatic urethra, included more than two intra-prostatic structures. Furthermore, the topical review highlighted a general lack of inter-reader variability data and that a vast majority of data were private, making comparisons of the performance between models difficult.

This study aimed to implement a deep learning model for automatic segmentation of the complete intra-prostatic urethra, prostate, and all four prostate zones, using structure terminology that adheres to the PI-RADS guidelines. The model was evaluated relative to inter-reader variability between a pair of radiologists, and its generalisability was assessed on an external dataset acquired from a different MRI vendor. Additionally, this work established a benchmark for a publicly available segmentation dataset with open code repositories to enable reproducible

and continued research.

2. Materials and methods

2.1. Data

In this work, the publicly available PROSTATEx [40,41] dataset was used, along with our previously published ProstateZones [42,43] segmentation dataset, which included segmentations for 200 randomly selected patients from PROSTATEx. The segmentations were manually delineated pixel-wise on axial T2w MRIs by two experienced radiologists (P.P., Z.F.) in collaboration with three junior colleagues (K.K., S.T., Z.D.). Prostate zones, including the PZ, CZ, TZ, and AFS, were delineated, and the urethra was defined as a six mm diameter circular region in each slice where present [43].

Out of the 200 segmentations, 160 samples had one segmentation per structure and were used for training and validating the deep learning model, while 40 samples had two segmentations per structure and were used as a test set. The duplicate delineations were performed independently by two experienced radiologists, which provided an inter-reader variability baseline for assessing model performance.

2.2. Deep learning model

A nnU-Net (v.2.2) [44] was trained on an NVIDIA GeForce RTX 3090 GPU with five-fold cross-validation on the 160 training samples of the dataset. The model performed semantic segmentations across all six classes (background, urethra and four prostate zones) using the image volumes. Code, weights and training details of the model are available with further details (<https://github.com/UMU-DDI/prostate-segmentation>).

2.3. Post-processing

Post-processing was used to align the model output with the manual delineation structure of the urethra, represented as a single circle in each slice with a fixed diameter of six mm.

First, the voxel-wise probability maps were used to extract the prostate volume as the largest connected component from the inverse of the background. Second, the coordinates of the voxel with the highest probability of being the urethra was extracted for each slice. If multiple voxels had the same probability, the centre of mass of the voxels was used. Third, a three-dimensional 2nd degree polynomial was fitted to the coordinates, with the initial probabilities used as weights. The polynomial fit output was used as the centre around which a new urethra was drawn with the predetermined circular shape and diameter of six mm, restricted by the prostate volume. Then, the largest connected components for each of the prostate zones were determined. Finally, any remaining voxels within the predicted prostate volume were assigned based on minimum Euclidean distance to the zones, emphasising anatomical correctness. This fully automated post-processing setup created a structure priority scheme, which emphasised the prostate boundary and urethra.

Example predictions with and without post-processing are displayed in [supplementary Figs. S1-S2](#).

2.4. Evaluation metrics

The performance of the model was evaluated with the Dice Similarity Coefficient (DSC), Surface DSC (SDSC) [45], percentile Symmetric Surface Distance (pSSD), and Center Line Distance (CLD). DSC quantified volumetric overlap, while SDSC captured boundary agreement within a specified tolerance to reflect typical inter-reader variability. The pSSD provided the symmetric surface distance required to include a given percentile of all boundary points, offering a tolerance-free, distribution-based measure directly related to the SDSC. Finally, the CLD evaluated

the average symmetric distance between two skeletonised versions of the urethra.

Difference between model and reader variation was tested for statistical significance using the related-samples Wilcoxon Signed Rank Test, as the data was determined to be non-Gaussian distributed based on the Shapiro-Wilk test. The same non-parametric test was also applied to quantify the effect of the post-processing on segmentation performance. Dataset differences in prostate volume and zonal proportions were tested using Welch’s t-test on transformed data, while associations between CLD and anatomical structures were examined using correlation and linear regression on log10-transformed volumes. No correction for multiple testing was performed. All evaluations were performed in Hero v.2025.2 (Hero Imaging AB, Umeå, Sweden; <https://www.heroimaging.com/>) with subsequent statistical analysis performed in SPSS Statistics v.29.0.1.0 (IBM, Armonk, NY, USA).

2.5. External test set

To assess the generalisability of the model, external validation was performed on a private dataset, which consisted of a prospective collection of 55 intermediate and high-risk prostate cancer patients from an ethically approved in-house study at the University Hospital of Umeå (Regional Ethics Board approval: Dnr 2016-220-31 M) [46].

Segmentations of the urethra and prostate zones were delineated slice-wise on the axial T2w MRI in the clinical software RayStation v.2023B (RaySearch Laboratories, Stockholm, Sweden). They followed the same delineation procedure as the ProstateZones dataset, including a fixed 6 mm circular urethra, and were performed by a medical physicist (W.H.) and subsequently controlled and corrected by an experienced radiation oncologist (K.S.). Images were collected with a Signa PET/MR 3 T scanner (GE Healthcare, Chicago, IL, USA) and had a reconstructed in-plane resolution of 0.41 mm and a slice thickness of 2.5 mm.

The pre- and post-processing of the external data was performed in the same manner as the other test data.

3. Results

Inter-reader variability on the test set showed a median CLD of 3.6 mm, with all metrics summarized in Table 1. Model comparisons showed a median CLD of 2.9 and 2.8 mm compared to Reader 1 and 2, respectively, as detailed in Table 2, with representative examples

Table 1

Inter-reader variability for the test data (n = 40). Metrics are presented as the median and inter-quartile ranges.

	DSC	SDSC 1 mm	pSSD (mm)			CLD (mm)
			P ₅₀	P ₈₀	P ₉₅	
Prostate	0.92	0.71	0.50	1.5	3.0	–
	[0.90, 0.93]	[0.66, 0.73]	[0.50, 0.52]	[1.4, 1.8]	[3.0, 3.0]	
PZ	0.75	0.67	0.50	1.8	3.7	–
	[0.72, 0.80]	[0.65, 0.71]	[0.50, 0.56]	[1.6, 2.1]	[3.2, 4.7]	
CZ	0.44	0.41	2.15	6.0	10.4	–
	[0.35, 0.56]	[0.30, 0.50]	[1.12, 3.00]	[3.4, 7.5]	[7.7, 12.4]	
TZ	0.84	0.59	0.71	2.4	3.6	–
	[0.79, 0.87]	[0.56, 0.62]	[0.56, 1.00]	[2.1, 2.7]	[3.3, 3.9]	
AFS	0.39	0.52	1.00	3.2	6.1	–
	[0.26, 0.48]	[0.44, 0.57]	[0.71, 1.52]	[3.0, 4.2]	[4.5, 8.9]	
Urethra	0.33	0.42	1.54	3.9	6.2	3.6
	[0.25, 0.43]	[0.33, 0.49]	[1.12, 2.50]	[2.9, 4.8]	[4.1, 7.2]	[2.8, 4.4]

DSC: Dice Similarity Coefficient, SDSC: Surface DSC, pSSD: Percentile Symmetric Surface Distance, CLD: Center Line Distance, PZ: Peripheral zone, CZ: Central zone, TZ: Transitional zone, AFS: Anterior fibromuscular stroma.

Table 2

Performance metrics for the model stratified for both readers, with comparisons towards Reader 1 presented in the upper half and Reader 2 in the lower half. Metrics are presented as the median and inter-quartile ranges.

Model		DSC	SDSC 1 mm	pSSD (mm)			CLD (mm)
				P ₅₀	P ₈₀	P ₉₅	
Reader 1	Prostate	0.91	0.69	0.50	1.6	3.2 *	–
		[0.90, 0.92]	[0.63, 0.73]	[0.50, 0.56]	[1.5, 2.1]	[3.0, 3.6]	
	PZ	0.78	0.70	0.50	1.7	3.5	–
		[0.72, 0.80]	[0.63, 0.74]	[0.50, 0.52]	[1.4, 2.1]	[3.0, 4.4]	
	CZ	0.50 *	0.47 *	1.31 *	4.7	9.5	–
		[0.43, 0.56]	[0.40, 0.54]	[1.00, 1.58]	[3.9, 6.1]	[8.4, 11.1]	
	TZ	0.86 *	0.61 *	0.64	2.1 *	3.2 *	–
		[0.83, 0.88]	[0.57, 0.65]	[0.50, 0.85]	[1.8, 2.4]	[3.0, 3.6]	
	AFS	0.42	0.59 *	0.71	3.0 *	4.5 *	–
		[0.34, 0.52]	[0.49, 0.67]	[0.50, 1.12]	[1.9, 3.4]	[3.4, 7.1]	
	Urethra	0.41 *	0.49 *	1.10 *	3.0 *	4.0 *	2.9 *
		[0.36, 0.52]	[0.41, 0.56]	[0.71, 1.41]	[2.7, 3.5]	[3.5, 5.5]	[2.3, 3.2]
Reader 2	Prostate	0.92	0.72	0.50	1.5	3.0	–
		[0.91, 0.93]	[0.67, 0.76]	[0.50, 0.52]	[1.2, 1.8]	[3.0, 3.2]	
	PZ	0.75	0.71 *	0.50	1.6	3.5	–
		[0.73, 0.80]	[0.66, 0.74]	[0.50, 0.52]	[1.4, 2.1]	[3.0, 5.1]	
	CZ	0.54 *	0.49 *	1.12 *	3.0 *	5.8 *	–
		[0.45, 0.61]	[0.42, 0.59]	[0.68, 1.50]	[2.7, 3.8]	[4.3, 7.8]	
	TZ	0.85 *	0.59 *	0.56 *	2.1 *	3.4 *	–
		[0.82, 0.87]	[0.57, 0.64]	[0.50, 0.71]	[2.0, 2.5]	[3.1, 3.7]	
	AFS	0.53 *	0.60 *	0.71	3.0 *	5.0	–
		[0.40, 0.58]	[0.46, 0.64]	[0.50, 1.41]	[2.4, 3.2]	[3.6, 6.8]	
	Urethra	0.41 *	0.49 *	1.12 *	3.0 *	4.0 *	2.8 *
		[0.30, 0.52]	[0.39, 0.62]	[0.67, 1.80]	[2.2, 3.6]	[3.2, 5.1]	[2.3, 3.7]

DSC: Dice Similarity Coefficient, SDSC: Surface DSC, pSSD: Percentile Symmetric Surface Distance, CLD: Center Line Distance, PZ: Peripheral zone, CZ: Central zone, TZ: Transitional zone, AFS: Anterior fibromuscular stroma.
*: Statistically significant difference relative to the inter-reader variability (p < 0.05).

displayed in Fig. 1.

Prediction and post-processing of all 40 cases required less than 10 minutes, for an average of 14.7 seconds per sample.

All statistical differences indicated lower model-reader variation compared to the inter-reader variation, except for the 95th pSSD for the prostate compared to Reader 1 (Table 2). Compared to the initial prediction, the applied post-processing mainly influenced the urethra, with a mean improvement for the CLD of -0.71 mm, as shown in supplementary Table S1.

On the external test set, the model achieved a median CLD of 2.3 mm, demonstrating robust generalisability as summarized in Table 3, with representative examples displayed in Fig. 2.

For the external dataset, prediction and post-processing of all 55 cases required just over 15 minutes, for an average of 16.6 seconds per sample on the same system as used during training.

Systematic anatomical differences were observed between the test set and the external validation dataset, as detailed in supplementary Tables S2–S3. The test set exhibited significantly larger prostate volumes (mean 58.5 cm³ vs. 40.8 cm³; p<0.001) and altered zonal composition, including a lower proportion of CZ (Odds Ratio (OR) = 0.48; 95% CI: 0.40-0.58; p<0.001) and a higher proportion of TZ (OR = 1.45; 95% CI: 1.16-1.81; p=0.001). Additionally, model performance for the urethra segmentation showed correlation with anatomical factors, with the

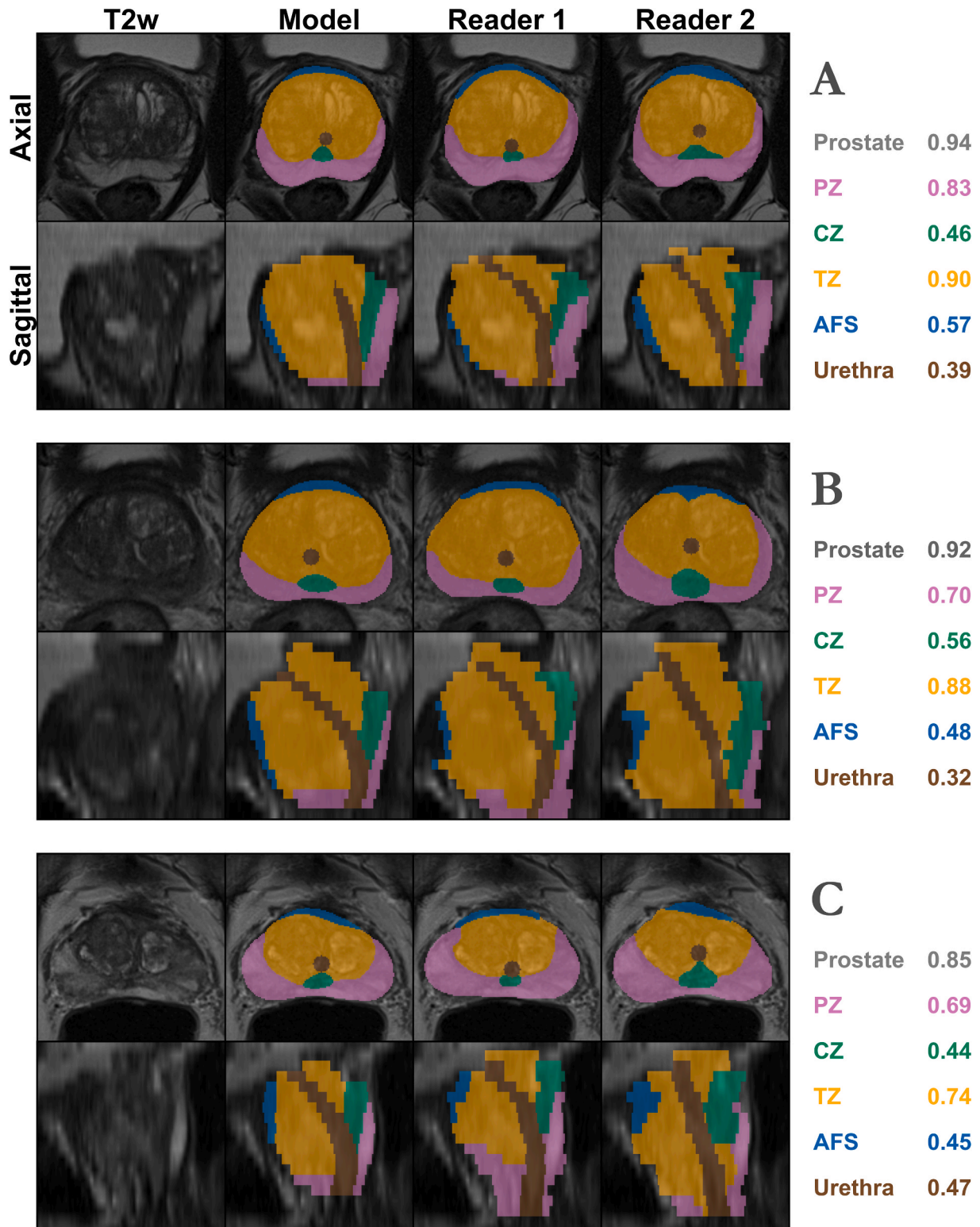


Fig. 1. Example segmentations on the test data. Examples with the smallest (panel A), average (panel B), and largest (panel C) differences in DSC between manual and automatic segmentations. The average model DSC for each structure is presented for each case to the right in each panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

strongest association for the CLD observed with increased TZ volume for both datasets (Pearson $r=0.54-0.56$), as detailed in supplementary Table S4.

4. Discussion

In this study, we implemented a deep learning model for the segmentation of the urethra, prostate, and prostate zones, with evaluation on unseen test data containing segmentations from multiple readers and an external dataset to assess generalisability. Overall, performance on

Table 3

Performance metrics for the model on the external test set (n = 55). Metrics are presented as the median and inter-quartile ranges.

	DSC	SDSC 1 mm	pSSD (mm)			CLD (mm)
			P ₅₀	P ₈₀	P ₉₅	
Prostate	0.91 [0.89, 0.92]	0.74 [0.71, 0.78]	0.41 [0.41, 0.41]	1.5 [1.2, 1.7]	2.9 [2.5, 3.6]	–
PZ	0.80 [0.76, 0.83]	0.75 [0.70, 0.78]	0.41 [0.41, 0.41]	1.2 [1.2, 1.9]	2.8 [2.5, 3.7]	–
CZ	0.41 [0.35, 0.46]	0.35 [0.27, 0.41]	2.32 [1.72, 2.58]	6.5 [5.2, 7.7]	11.1 [9.6, 12.6]	–
TZ	0.81 [0.79, 0.86]	0.64 [0.59, 0.69]	0.58 [0.41, 0.82]	2.1 [1.7, 2.2]	3.5 [2.9, 4.8]	–
AFS	0.56 [0.45, 0.61]	0.69 [0.59, 0.77]	0.41 [0.41, 0.82]	1.6 [1.2, 2.5]	3.3 [2.7, 5.0]	–
Urethra	0.49 [0.37, 0.68]	0.53 [0.42, 0.73]	0.92 [0.58, 1.30]	2.5 [1.3, 3.0]	3.7 [2.7, 5.0]	2.3 [1.5, 3.2]

DSC: Dice Similarity Coefficient, SDSC: Surface DSC, pSSD: Percentile Symmetric Surface Distance, CLD: Center Line Distance,
PZ: Peripheral zone, CZ: Central zone, TZ: Transitional zone, AFS: Anterior fibromuscular stroma.

the test set was at the level of, or better than, the inter-reader variability, with robust predictions on the external test set. Notably, when compared to the inter-reader variability, urethra segmentations were significantly closer to each radiologist than the radiologists were to each other, indicating that the model output tended towards an average of both on unseen test data.

In the existing literature, the urethra segmentation of Czipczer et al. [25], achieved a DSC of 0.39 and a 1 mm SDSC of 0.43 when segmenting the urethra as an organ at risk for radiotherapy planning. In a study with over 900 total patients, Belue et al. [24] achieved a DSC of 0.59 and a CLD of 2.56 mm, although their CLD was computed slice-wise and only included slices with both segmentations present, meaning that any slice with only one segmentation was not penalised. Furthermore, all segmentations were performed by a single delineator, a known limitation due to potential bias introduced in predictions when compared to the same delineator [38]. In contrast, our evaluations included all slices, meaning that any slice with only one segmentation would register a distance of at least the slice thickness. Additionally, a strength of our study was that it involved multiple delineators in the delineation process for the training data, and that evaluations were performed against three different delineators in total. Xu et al. [26] reported that manual delineations of the urethra typically required 5 minutes per sample, which is substantially longer than our complete pipeline, which processed each sample in, on average less than 17 seconds in both the test and external validation datasets.

Comparing segmentations of prostate zones is challenging due to varying terminology and structure definitions used in previous studies. However, the literature reports mean or median DSC values for the prostate, PZ, and TZ/CZ/CZ-TZ/CG in the ranges of 0.84-0.95, 0.70-0.87, and 0.85-0.94, respectively [25,28–38]. A single study that delineated the AFS reported a DSC of 0.41 [38]. This was in line with our results; however, differentiating the prostate into a larger number of structures is inherently more challenging. Similarly, as far as a comparison is possible, the inter-reader variability, presented in Table 1, was comparable with previous studies [13,28,31,47,48].

This study included limitations. First, the model used did not reflect the contemporary state-of-the-art, such as transformer-inspired models like MedNeXt [49]. However, rather than optimising for state-of-the-art performance, our primary goal was to establish a publicly accessible and reproducible benchmark for segmentation of the urethra and prostate zones. The focus was therefore on evaluation against inter-reader

variability and external validation across vendors, and open access to data and code, rather than model innovation. Still, the fact that a lightweight model like the nnU-Net could reach segmentation performance at the level of the inter-reader variability demonstrated that this task can be achieved without requiring large computational resources.

Second, the resolution of the images from PROSTATEx did not comply with the recommendations of the PI-RADS v2.1 guidelines, where adherence has been shown to make models more generalisable [50]. Additionally, training data consisted solely of images from a single vendor (Siemens), which has been shown to degrade performance when evaluated on data from a different vendor (GE), compared to models trained on data from multiple vendors [37]. Although we did not compare our model with one trained on data from multiple vendors, our results were somewhat contrary to these findings. Several factors may have contributed to this outcome. For example, the external data had higher resolution, which likely supported clearer depiction of both zonal boundaries and the urethra. Additionally, the datasets differed anatomically, including variations in prostate size and zonal proportions. These anatomical differences may partly reflect variation in reader background and experience, particularly in how anatomical boundaries were interpreted. In our analysis, larger prostates and, in particular, a larger TZ were associated with greater urethral variability, reflected by increased CLD values. Enlarged TZ volumes are often linked to benign prostatic hyperplasia, a condition known to distort gland morphology and complicate delineations of the urethra [3,24]. Notably, both total prostate and TZ volume were larger in the test set than in the external data, which may partially explain why the deviation in urethral segmentations were more pronounced there.

Third, the urethral ground truth relied on a heuristic delineation protocol, in which the centre of the urethra was identified on each slice and represented using a fixed 6 mm circular contour, with interpolation applied in low-visibility slices. While this approach is clinically practical and consistent across datasets, it is not an accurate representation of the true urethral anatomy, something that should be considered in future studies. However, as the post-processing worked under the same heuristics, comparison within the study remained valid.

Fourth, while this study included post-processing that emphasised the urethra and prostate boundary for its clear importance in local dose escalated radiotherapy, it did not evaluate the clinical applicability of these segmentations within a radiotherapy workflow.

In conclusion, this study implemented a deep learning model, the nnU-Net, for segmentations of the urethra, prostate, and prostate zones with performance comparable to the inter-reader variability of two radiologists on unseen test data. Notably, the model provided results significantly better than the inter-reader variability for multiple structures and metrics, especially for the urethra, with only a single comparison in total being significantly worse. The segmentation of the urethra approached an average of the manual delineations, as the output was on average closer to each delineator than they were to each other. Additionally, performance was robust when generalisability was assessed on external test data collected from a different vendor. These findings showed the potential to reduce time and variability associated with manual delineations of the urethra, prostate, and prostate zones. By achieving segmentation quality comparable to manual delineations in seconds instead of minutes, this approach showed promise as a decision support solution, although further validation is needed in clinical settings. Furthermore, this work addressed an important gap in the field by presenting a fully automatic solution that theoretically impacts both diagnostic and therapeutic applications. Finally, it established a benchmark for a publicly available segmentation dataset with open code repositories to enable reproducible and continued research.

CRedit authorship contribution statement

William Holmlund: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Attila Simkó:**

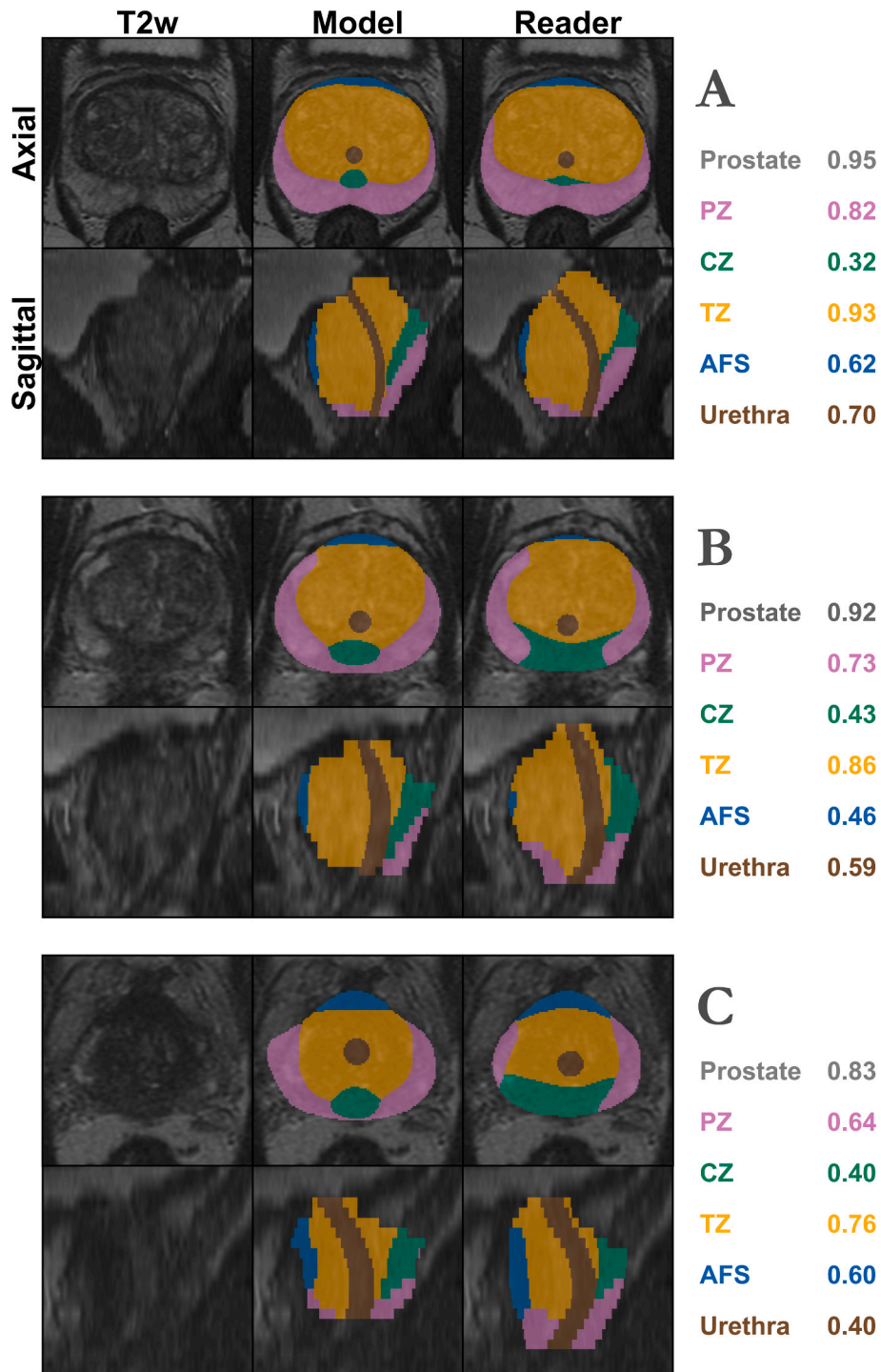


Fig. 2. Example segmentations on the external validation data. Examples with the smallest (panel A), average (panel B), and largest (panel C) differences in DSC between manual and automatic segmentations. The model DSC for each structure is presented for each case to the right in each panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Writing – review & editing, Methodology, Conceptualization. **Karin Söderkvist:** Writing – review & editing, Data curation, Conceptualization. **Péter Palásti:** Writing – review & editing, Data curation. **Szilvia Tóti:** Writing – review & editing, Data curation. **Kamilla Kalmár:** Writing – review & editing, Data curation. **Zsófia Domoki:** Writing – review & editing, Data curation. **Zsuzsanna Fejes:** Writing – review & editing, Data curation. **Tamás Z. Kincses:** Supervision and Writing – review & editing. **Patrik Brynolfsson:** Writing – review & editing, Conceptualization. **Tufve Nyholm:** Writing – review & editing,

Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: TN, PB and AS are co-owners of Hero Imaging AB, developing the software used during evaluations. All other authors declare no conflict of interest.

Acknowledgment

This project was funded by the Cancer Research Foundation in Northern Sweden and Cancerfonden.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.phro.2026.100964>.

References

- [1] International Agency for Research on Cancer. Global Cancer Observatory: Cancer Today. <https://gco.iarc.who.int/today/en>; 2024 [accessed 10 February 2026].
- [2] Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer—2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol* 2021;79:243–62. <https://doi.org/10.1016/j.eururo.2020.09.042>.
- [3] Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. update of Prostate Imaging Reporting and Data System version 2. *Eur Urol* 2019;2019(76):340–51. <https://doi.org/10.1016/j.eururo.2019.02.033>.
- [4] McNeal JE. Normal histology of the prostate. *Am J Surg Pathol*. 1988;12:619–33. <https://doi.org/10.1097/00000478-198808000-00003>.
- [5] Haj-Mirzaian A, Burk KS, Lacsos R, Glazer DI, Saini S, Kibel AS, et al. Magnetic Resonance Imaging, Clinical, and Biopsy Findings in Suspected Prostate Cancer. A Systematic Review and Meta-Analysis. *JAMA Netw Open*. 2024;7(3):e244258. <https://doi.org/10.1001/jamanetworkopen.2024.4258>.
- [6] Hamm CA, Baumgärtner GL, Padhani AR, Froböse KP, Dräger F, Beetz NL, et al. Reduction of false positives using zone-specific prostate-specific antigen density for prostate MRI-based biopsy decision strategies. *Eur Radiol*. 2024;34:6229–40. <https://doi.org/10.1007/s00330-024-10700-z>.
- [7] Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med*. 2018;378:1767–77. <https://doi.org/10.1056/NEJMoa1801993>.
- [8] Eklund M, Jäderling F, Discacciati A, Bergman M, Annerstedt M, Aly M, et al. MRI-Targeted or Standard Biopsy in Prostate Cancer Screening. *N Engl J Med*. 2021;385:908–20. <https://doi.org/10.1056/NEJMoa2100852>.
- [9] Kerkmeijer LGW, Groen VH, Pos FJ, Haustermans K, Monninkhof EM, Smeenk RJ, et al. Focal Boost to the Intraprostatic Tumor in External Beam Radiotherapy for Patients With Localized Prostate Cancer: Results From the FLAME Randomized Phase III Trial. *J Clin Oncol*. 2021;39:787–96. <https://doi.org/10.1200/JCO.20.02873>.
- [10] Leeman JE, Chen YH, Catalano P, Bredfeldt J, King M, Mouw KW, et al. Radiation Dose to the Intraprostatic Urethra Correlates Strongly With Urinary Toxicity After Prostate Stereotactic Body Radiation Therapy: A Combined Analysis of 23 Prospective Clinical Trials. *Int J Radiat Oncol Biol Phys*. 2022;112:75–82. <https://doi.org/10.1016/j.ijrobp.2021.06.037>.
- [11] Groen VH, van Schie M, Zuihthoff NPA, Monninkhof EM, Kunze-Busch M, de Boer JCJ, et al. Urethral and bladder dose-effect relations for late genitourinary toxicity following external beam radiotherapy for prostate cancer in the FLAME trial. *Radiother Oncol*. 2022;167:127–32. <https://doi.org/10.1016/j.radonc.2021.12.027>.
- [12] Ali A, Du Feu A, Oliveira P, Choudhury A, Bristow RG, Baena E. Prostate zones and cancer: lost in transition? *Nat Rev Urol*. 2022;19:101–15. <https://doi.org/10.1038/s41585-021-00524-7>.
- [13] Montagne S, Hamzaoui D, Allera A, Ezziane M, Luzurier A, Quint R, et al. Challenge of prostate MRI segmentation on T2-weighted images: inter-observer variability and impact of prostate morphology. *Insights Imaging*. 2021;12:71. <https://doi.org/10.1186/s13244-021-01010-9>.
- [14] Litjens G, Kooi T, Beijndorf BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88. <https://doi.org/10.1016/j.media.2017.07.005>.
- [15] Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: A survey. *IET Image Process*. 2022;16:1243–67. <https://doi.org/10.1049/ipr2.12419>.
- [16] Turkbey B, Haider MA. Deep learning-based artificial intelligence applications in prostate MRI: brief summary. *Br J Radiol*. 2022;95:20210563. <https://doi.org/10.1259/bjr.20210563>.
- [17] Khan Z, Yahya N, Alsaih K, Al-Hiyali MI, Meriaudeau F. Recent Automatic Segmentation Algorithms of MRI Prostate Regions: A Review. *IEEE Access*. 2021;9:97878–905. <https://doi.org/10.1109/ACCESS.2021.3090825>.
- [18] Elguindi S, Zelefsky MJ, Jiang J, Veeraraghavan H, Deasy JO, Hunt MA, et al. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. *Phys Imaging Radiat Oncol*. 2019;12:80–6. <https://doi.org/10.1016/j.phro.2019.11.006>.
- [19] Acosta O, Mylona E, Le Dain M, Voisin C, Lizée T, Rigaud B, et al. Multi-atlas-based segmentation of prostatic urethra from planning CT imaging to quantify dose distribution in prostate cancer radiotherapy. *Radiother Oncol*. 2017;125:492–9. <https://doi.org/10.1016/j.radonc.2017.09.015>.
- [20] Takagi H, Kadoya N, Kajikawa T, Tanaka S, Takayama Y, Chiba T, et al. Multi-atlas-based auto-segmentation for prostatic urethra using novel prediction of deformable image registration accuracy. *Med Phys*. 2020;47:3023–31. <https://doi.org/10.1002/mp.14154>.
- [21] Cubero L, Garcia-Elcano L, Mylona E, Boue-Rafle A, Cozzarini C, Ubeira Gabellini MG, et al. Deep learning-based segmentation of prostatic urethra on computed tomography scans for treatment planning. *Phys Imaging Radiat Oncol*. 2023;26:100431. <https://doi.org/10.1016/j.phro.2023.100431>.
- [22] Litzenberg DW, Muenz DG, Archer PG, Jackson WC, Hamstra DA, Hearn JW, et al. Changes in prostate orientation due to removal of a Foley catheter. *Med Phys*. 2018;45:1369–78. <https://doi.org/10.1002/mp.12830>.
- [23] Dekura Y, Nishioka K, Hashimoto T, Miyamoto N, Suzuki R, Yoshimura T, et al. The urethral position may shift due to urethral catheter placement in the treatment planning for prostate radiation therapy. *Radiat Oncol*. 2019;14:226. <https://doi.org/10.1186/s13014-019-1424-8>.
- [24] Belue MJ, Harmon SA, Patel K, Daryanani A, Yilmaz EC, Pinto PA, et al. Development of a 3D CNN-based AI Model for Automated Segmentation of the Prostatic Urethra. *Acad Radiol*. 2022;29:1404–12. <https://doi.org/10.1016/j.acra.2022.01.009>.
- [25] Czipczer V, Kolozsvári B, Deák-Karancsi B, Capala ME, Pearson RA, Borzási E, et al. Comprehensive deep learning-based framework for automatic organs-at-risk segmentation in head-and-neck and pelvis for MR-guided radiation therapy planning. *Front Phys*. 2023;11:1236792. <https://doi.org/10.3389/fphy.2023.1236792>.
- [26] Xu D, Ma TM, Savjani R, Pham J, Cao M, Yang Y, et al. Fully automated segmentation of prostatic urethra for MR-guided radiation therapy. *Med Phys*. 2023;50:354–64. <https://doi.org/10.1002/mp.15983>.
- [27] Song Y, Nguyen L, Dornisch AM, Baxter MT, Barrett T, Dale AM, et al. Urethra contours on MRI: Multidisciplinary consensus educational atlas and reference standard for artificial intelligence benchmarking. *Radiother Oncol* 2026;1112131. <https://doi.org/10.1016/j.radonc.2025.1112131>.
- [28] Aldo J, Biavati F, Michallek F, Stober S, Dewey M. Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. *Sci Rep*. 2020;10:14315. <https://doi.org/10.1038/s41598-020-71080-0>.
- [29] Bardis M, Houshyar R, Chantaduly C, Tran-Harding K, Ushinsky A, Chahine C, et al. Segmentation of the Prostate Transition Zone and Peripheral Zone on MR Images with Deep Learning. *Radiol Imaging Cancer*. 2021;3:e200024. <https://doi.org/10.1148/rycan.2021200024>.
- [30] Cuocolo R, Comelli A, Stefano A, Benfante V, Dahiya N, Stanzione A, et al. Deep Learning Whole-Gland and Zonal Prostate Segmentation on a Public MRI Dataset. *J Magn Reson Imaging*. 2021;54:452–9. <https://doi.org/10.1002/jmri.27585>.
- [31] Adams LC, Makowski MR, Engel G, Rattunde M, Busch F, Asbach P, et al. Prostate158 - An expert-annotated 3T MRI dataset and algorithm for prostate cancer detection. *Comput Biol Med*. 2022;148:105817. <https://doi.org/10.1016/j.combiomed.2022.105817>.
- [32] Hamzaoui D, Montagne S, Renard-Penna R, Ayache N, Delingette H. Automatic zonal segmentation of the prostate from 2D and 3D T2-weighted MRI and evaluation for clinical use. *J Med Imaging*. 2022;9:024001. <https://doi.org/10.1117/1.JMI.9.2.024001>.
- [33] Jimenez-Pastor A, Lopez-Gonzalez R, Fos-Guarinos B, Garcia-Castro F, Wittenberg M, Torregrosa-Andres A, et al. Automated prostate multi-regional segmentation in magnetic resonance using fully convolutional neural networks. *Eur Radiol*. 2023;33:5087–96. <https://doi.org/10.1007/s00330-023-09410-9>.
- [34] Rodrigues NM, Almeida JG, Verde ASC, Gaivao AM, Bilreiro C, Santiago I, et al. Analysis of domain shift in whole prostate gland, zonal and lesions segmentation and detection, using multicentric retrospective data. *Comput Biol Med*. 2024;171:108216. <https://doi.org/10.1016/j.combiomed.2024.108216>.
- [35] Xu L, Zhang G, Zhang D, Zhang J, Zhang X, Bai X, et al. Development and clinical utility analysis of a prostate zonal segmentation model on T2-weighted imaging: a multicenter study. *Insights Imaging*. 2023;14:44. <https://doi.org/10.1186/s13244-023-01394-w>.
- [36] Zabiollahy F, Schieda N, Krishna Jeyaraj S, Ukwatta E. Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets. *Med Phys*. 2019;46:3078–90. <https://doi.org/10.1002/mp.13550>.
- [37] Zavala-Romero O, Breto AL, Xu IR, Chang YC, Gautney N, Dal Pra A, et al. Segmentation of prostate and prostate zones using deep learning: A multi-MRI vendor analysis. *Strahlenther Onkol*. 2020;196:932–42. <https://doi.org/10.1007/s00066-020-01607-x>.
- [38] Meyer A, Rakr M, Schindele D, Blaschke S, Schostak M, Fedorov A, et al. Towards patient-individual PI-Rads v2 sector map: CNN for automatic segmentation of prostatic zones from T2-weighted MRI. In: 2019 IEEE 16th Int Symp Biomed Imaging (ISBI); 2019. p. 696–700. <https://doi.org/10.1109/ISBI.2019.8759572>.
- [39] Wu C, Montagne S, Hamzaoui D, Ayache N, Delingette H, Renard-Penna R. Automatic segmentation of prostate zonal anatomy on MRI: a systematic review of the literature. *Insights Imaging*. 2022;13:202. <https://doi.org/10.1186/s13244-022-01340-2>.
- [40] Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. SPIE-AAPM PROSTATEx Challenge Data, The Cancer Imaging Archive, v2; 2017. <https://doi.org/10.7937/k9TClA.2017.MURS5CL>.
- [41] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045–57. <https://doi.org/10.1007/s10278-013-9622-7>.
- [42] Holmlund W, Simkó A, Söderkvist K, Palásti P, Tótin S, Kalmár K, et al. ProstateZones – Segmentations of the prostatic zones and urethra for the PROSTATEx dataset. *Zenodo* 2024. <https://doi.org/10.5281/zenodo.10718469>.

- [43] Holmlund W, Simkó A, Söderkvist K, Palásti P, Tótin S, Kalmár K, et al. ProstateZones – Segmentations of the prostatic zones and urethra for the PROSTATEx dataset. *Sci Data*. 2024;11:1097. <https://doi.org/10.1038/s41597-024-03945-2>.
- [44] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203–11. <https://doi.org/10.1038/s41592-020-01008-z>.
- [45] Nikolov S, Blackwell S, Zverovitch A, Mendes R, Livne M, De Fauw J, et al. Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study. *J Med Internet Res*. 2021;23:e26151. <https://doi.org/10.2196/26151>.
- [46] Nilsson E, Sandgren K, Grefve J, Jonsson J, Axelsson J, Keeratijarut Lindberg A, et al. The grade of individual prostate cancer lesions predicted by magnetic resonance imaging and positron emission tomography. *Commun Med*. 2023;3:164. <https://doi.org/10.1038/s43856-023-00394-7>.
- [47] Becker AS, Chaitanya K, Schawkat K, Muehlematter UJ, Hötter AM, Konukoglu E, et al. Variability of manual segmentation of the prostate in axial T2-weighted MRI: A multi-reader study. *Eur J Radiol*. 2019;121:108716. <https://doi.org/10.1016/j.ejrad.2019.108716>.
- [48] Padgett KR, Swallen A, Pirozzi S, Piper J, Chinae FM, Abramowitz MC, et al. Towards a universal MRI atlas of the prostate and prostate zones. *Strahlenther Onkol*. 2019;195:121–30. <https://doi.org/10.1007/s00066-018-1348-5>.
- [49] Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jäger, P. F., & Maier-Hein, K. H. MedNeXt: Transformer-Driven Scaling of ConvNets for Medical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2023*. Lecture Notes in Computer Science, vol. 14223. 2023:405–415. https://doi.org/10.1007/978-3-031-43901-8_39.
- [50] Fernandez-Quilez A, Nordström T, Eftestøl T, Alvestad AB, Jäderling F, Kjosavik SR, et al. Revisiting prostate segmentation in magnetic resonance imaging (MRI): On model transferability, degradation and PI-RADS adherence. [preprint] medRxiv. 2023:2023.08.21.23294376. <https://doi.org/10.1101/2023.08.21.23294376>.