Association Between Artificial Intelligence-Derived Tumor Volume and Oncologic Outcomes for Localized Prostate Cancer Treated with Radiation Therapy

David D Yang[a,b,*], Leslie K Lee[b,*], James MG Tsui[c], Jonathan E Leeman[a,b], Katie N Lee[a,b], Heather M McClure[a], Atchar Sudhyadhom[a,b], Christian V Guthier[a,b], Kent W Mouw[a,b], Neil E Martin[a,b], Peter F Orio[a,b], Paul L Nguyen[a,b], Anthony V D'Amico[a,b], Martin T King[a,b]

[a]Dana-Farber Cancer Institute, Boston, MA, USA
[b]Brigham and Women's Hospital, Boston, MA, USA
[c]McGill University, Montreal, Canada

*Equal contribution

Corresponding Author
Martin T King
75 Francis St
Boston, MA 02115
Telephone: +1-617-732-5640
Fax: +1-617-632-4247
Email: Martin_King@dfci.harvard.edu

Keywords: artificial intelligence; deep learning; dominant intraprostatic lesion; magnetic resonance imaging; PI-RADS; prostate cancer; radiomics; radiation therapy

Abstract word count: 300
Text word count: 2794 (including abstract)

Background: Although clinical features of multi-parametric magnetic resonance imaging (mpMRI) have been associated with biochemical recurrence in localized prostate cancer, such features are subject to inter-observer variability.

Objective: To evaluate whether the volume of the dominant intraprostatic lesion (DIL), as provided by a deep learning segmentation algorithm, could provide prognostic information for patients treated with definitive radiation therapy (RT).

Design, Setting, and Participants: Retrospective study of 438 patients with localized prostate cancer who underwent an endorectal coil, high B-value, 3-Tesla mpMRI and were treated with RT between 2010 and 2017.

Intervention: RT.

Outcome Measurements and Statistical Analysis: Biochemical recurrence and metastasis risk, assessed with a cause-specific Cox regression and time-dependent receiver operating characteristic analysis.

Results and Limitations: The artificial intelligence (AI) model identified DILs with an area under the receiver operating characteristic curve (AUROC) of 0.827 at the patient level. For the 233 patients with available PI-RADS scores, with a median follow-up of 5.6 years, AI-defined DIL volume was significantly associated with biochemical failure (adjusted hazard ratio 1.54, 95% confidence interval 1.09-2.17, p=0.014) after adjustment for PI-RADS score. Among all 438 patients with a median follow-up of 6.9 years, the AUROC for predicting 7-year biochemical failure for AI volume (0.790) was similar to that for an expanded National Comprehensive Cancer Network (NCCN+) category (p=0.17). The AUROC for predicting 7-year metastasis for AI volume trended towards being higher compared to NCCN+ categories (0.854 vs 0.769, p=0.06).

Conclusions: A deep learning algorithm could identify the DIL with good performance. AI-defined DIL volume may be able to provide prognostic information independent of the NCCN+ risk group or other radiologic factors for patients with localized prostate cancer treated with RT.

Introduction

Multi-parametric magnetic resonance imaging (mpMRI) has revolutionized the management of localized prostate cancer. mpMRI-based biopsy strategies have been shown to improve the detection of clinically significant disease while reducing the detection of clinically insignificant disease[1–4]. Furthermore, the characteristics of dominant intraprostatic lesions (DILs) on mpMRI, including PI-RADS scores[5], radiologic T-stage[6,7], and lesion size[8–11], have been shown to be prognostic. Recently, a nomogram based on clinical and radiologic parameters was found to be highly prognostic for early biochemical recurrence after radical prostatectomy[12].

However, the assignment of various clinical features on mpMRI is subject to significant inter-observer variability. For example, multiple grading systems[13–15] exist for assigning extracapsular extension (EPE), each with differing sensitivities for the detection of histologic EPE extent[16]. Multi-reader studies have reported moderate inter-observer variability in the reporting of Prostate Imaging Reporting and Data Systems (PI-RADS) v2.0 scores[17–19]. A recent study demonstrated that the positive predictive value of PI-RADS scores for detecting high-grade prostate cancer was low (35%) and variable across 26 centers[20]. The varied performance of mpMRI is likely multifactorial and could be partially attributed to technical and interpretative factors, as well as benign conditions that mimic malignancy[21].

Given these limitations, there has been considerable interest in the development of artificial intelligence (AI) algorithms for supporting the radiologist's workflow. Although the performance of AI algorithms has not been shown to match that of radiologists for the detection of clinically significant prostate cancer[22], recent work suggests that deep

learning algorithms could be used as an adjunct for assisting radiologists[23]. However, less is known about the prognostic information provided by deep learning algorithms, particularly in comparison to current staging systems. The purpose of this study is to evaluate whether the mpMRI DIL volume, as determined by a deep learning segmentation algorithm[24], could provide prognostic information for patients with localized prostate cancer treated with definitive radiation therapy (RT).

Methods

Clinical datasets

We conducted a retrospective study of 438 patients with cT1cN0M0 prostate cancer who underwent an endorectal coil, high B-value, 3-Tesla mpMRI and were treated with definitive prostate RT, both at our institution, between 2010 and 2017 (Table S1). mpMRI were obtained with either General Electric (GE) Signa HDxt (n=237) or DISCOVERY MR750w (n=201) (Table S2). Apparent diffusion coefficient (ADC)/diffusion-weighted imaging (DWI) were obtained at the same resolution (0.70 x 0.70 x 3-4 mm) as the T2-weighted images. All research was conducted with approval from the institutional review board.

We captured baseline clinical characteristics and treatment characteristics (Table 1). We stratified patients into an expanded National Cancer Center Network (NCCN+) classification, based on the 4-tiered NCCN stratification (low, favorable intermediate-risk, unfavorable intermediate-risk, high-risk)[25], with the addition of a very high-risk category per eligibility criteria from recent STAMPEDE trials (≥2 of the following: cT3-T4, Gleason score 8-10, or PSA ≥40 ng/mL)[26]. We captured radiologic staging parameters, including PI-RADS v2.0 scores[27], which were assessed by a subspecialty abdominal radiologist (L.K.L.) for 83 patients who were scanned in 2010-2013 and available for all patients after mid-2015, as well as radiologic T-stage (T1: no visible tumor on mpMRI, T2: organ-confined, T3a: EPE, T3b: seminal vesicle invasion [SVI]) based on mpMRI. Lastly, we captured outcomes including biochemical failure (nadir+2 ng/mL), development of metastasis (non-regional nodal or distant disease on bone scan, computed tomography, or positron emission tomography), and survival.

We divided patients into 3 cohorts. Of the 233 patients with available PI-RADS scores, we randomly allocated them into training (TrainPIRADS) and testing (TestPIRADS) sets. The remaining 205 patients were allocated into a second testing (TestNoPIRADS) set. For the TrainPIRADS and TestPIRADS sets, a radiation oncologist contoured the prostate transitional zone (TZ), peripheral zone (PZ), and DIL based on the presence of PI-RADS 3-5 scores. For the TestNOPIRADS set, the radiation oncologist contoured only the DIL based on radiology reports and biopsy data.

nnUNet algorithm

We utilized the publicly available nnUNet[24] deep learning algorithm for automated delineations of the TZ, PZ, and DIL. First, T2, ADC, and DWI images were downloaded onto a research workstation. Images were anonymized and cropped into smaller arrays (128 x 128 x 15 or 20 pixels). T2-weighted images were registered onto the ADC images, utilizing the best performing registration among 4 algorithms (rigid, rigid + B-spline, affine, affine + B-spline) of prostate segmentation distance maps[28]. The median registration Dice coefficient was 0.92 (interquartile range [IQR] 0.90-0.94).

We trained a 3D full-resolution nnUNet model utilizing 5-fold cross-validation on the TrainPIRADS set. We then applied the selected configuration onto the TestPIRADS and TestNoPIRADS sets. With the utilization of 5-fold cross-validation on the training set and application of the model on the 2 testing sets, we generated TZ, PZ, and DIL segmentations for all 438 patients.

Assessment of AI DIL volume performance

We reported the Dice coefficients for the TZ and PZ, as noted by nnUNet. For the DIL, we analyzed the detection performance (patient-level area under the receiver operator characteristic curve [AUROC] and lesion-level specificity at 0.5 false positive per image) utilizing PICAI_eval[29]. To assess whether false positive (FP) or false negative (FN) lesions were smaller and less conspicuous than true positives (TP), we classified AI lesions as TP, FP, or FN, based on whether a reference DIL was present and correctly identified (TP: Dice coefficient ≥10%[29]), identified but not present (FP), or present but not identified or with insufficient Dice coefficient overlap (<10%) (FN). We analyzed lesion-level DIL volumes and contrast ratios (ratio of ADC intensities within the DIL versus a surrounding 2mm ring) based on classification status. Finally, we reported the similarity metrics (Dice coefficient) for TP lesions. All processing was performed using Python version 3.8.12 with the SimpleITK toolkit.

Comparison of AI and reference DIL volumes for sextant subset

For the 303 patients with sextant biopsies among all 3 cohorts, we compared the ability of the AI and reference DIL contours to detect clinically significant (Gleason grade group ≥2) disease within each sextant. We compared area under the curve (AUC) values utilizing ROC analysis (R package pROC). AI and reference DIL contours were allocated according to sextants based on the reference prostate contour.

Comparison of DIL volume with radiologic staging for PI-RADS subset

For the 233 patients (TrainPIRADS and TestPIRADS) with available PI-RADS scores among all 3 cohorts, we conducted cause-specific Cox regression models (R

package coxph) for predicting biochemical failure with log-transformed AI DIL volume in cm^3 (V_{AI}), clinical T-stage, radiologic T-stage, PI-RADS scores (0-4 vs 5[30]), age, whether patients received standard of care (SOC, i.e. ≥75.6 Gy if receiving external beam RT and androgen deprivation therapy [ADT] duration of 6 months for unfavorable intermediate-risk disease and ≥18 months for high/very high-risk disease), year of treatment, and scanner model (Signa HDxt vs DISCOVERY MR750w) as independent variables. An additional Cox regression model for metastasis was created using the above independent variables with the addition of salvage ADT use as a time-dependent variable.

Comparison of DIL volume with NCCN+ risk category for entire cohort

For the entire cohort, we created ROC curves for NCCN+, reference DIL volume (V_{REF}), and V_{AI} for predicting 7-year biochemical failure and metastasis, utilizing the timeROC package. We also created cause-specific Cox regression models (R package coxph) for predicting risk of biochemical failure, local failure, and metastasis. Independent variables included log-transformed V_{AI}, NCCN+ classification, age, receipt of SOC, year of treatment, scanner model, and for local failure and metastasis models, salvage ADT use as a time-dependent variable. Finally, we estimated cumulative incidences of biochemical failure, local failure, and metastasis, based on partitions of V_{AI} intervals, utilizing the R package prodlim. All statistical analysis was conducted with R version 4.1.2.

Results

Assessment of AI DIL performance

For the TestPIRADS cohort, median Dice coefficients for the TZ and PZ were 0.89 and 0.84, respectively (Table S3). The nnUNet exhibited a patient-level AUROC of 0.827 and a lesion-level sensitivity of 72.7% at a false positive rate of 0.5 case per image (Table S4). For patient-level analysis (Table S5), there was no significant difference between V_{AI} (median 1.02cm^3, IQR 0.40-2.34) and V_{REF} (median 1.31cm^3, IQR 0.37-2.61, p=0.66), with Pearson's correlation coefficient of 0.79. For lesion-level analysis (Table S6), the total volume of AI TP lesions (median 1.06cm^3, IQR 0.47-1.85) was significantly greater than for FP (median 0.25cm^3, IQR 0.09-0.58; p<0.001) and FN (median 0.37cm^3, IQR 0.21-0.88, p=0.007) lesions. The TP lesions were also more conspicuous (median contrast ratio 0.72, IQR 0.68-0.76) than FP (median 0.83, IQR 0.79-0.86; p<0.001) and FN (median 0.81, IQR 0.79-0.88; p<0.001) lesions. The median Dice coefficient for TP lesions was 0.74. Examples of TP, FP, and FN lesions are shown in Fig. S1.

Comparison of AI and reference DIL contours for sextant subset

For the subset of 303 patients with sextant biopsies among all 3 cohorts, we did not detect a difference in AUROC values associated with AI versus reference contours for any sextant (Table S7).

Comparison of DIL volume with radiologic staging for PI-RADS subset

For the 233 patients with PI-RADS scores (TrainPIRADS and TestPIRADS cohorts), with median follow-up of 5.6 years, there were 28 biochemical failures. AI algorithm detected 41.2%, 68.3%, and 87.1% of PI-RADS3, 4, and 5 lesions, respectively (Table S8). PIRADS 5 vs 0-4 (HR 5.90, 95% confidence interval [CI] 2.05-17.03, p=0.001) and V_{AI} (HR 1.89 per log cm^3 increase, 95% CI 1.41-2.53, p<0.001) were significantly associated with biochemical failure on univariable analysis (Table 2). However, only V_{AI} retained significance on multivariable analysis (adjusted HR 1.54, 95% CI 1.09-2.17; p=0.014). The association between V_{AI} and metastasis was similarly significant (adjusted HR 1.71, 95% CI 1.04-2.80, p=0.04; Table S9).

Comparison of DIL volume with NCCN+ risk category for the entire cohort

Among all 438 patients with a median follow-up of 6.9 years, there were 49 biochemical failures, 14 local failures, and 22 metastases. The AUROC for 7-year biochemical failure for V_{AI} (0.790) was similar to that for V_{REF} (0.779; p=0.42) and NCCN+ category (0.740; p=0.17). The sensitivity and specificity for 7-year biochemical failure at DIL volumes >0.5 mL were 90.5% and 45.8%, respectively, and at DIL volumes >2.0 mL, were 57.7% and 84.4%, respectively. The AUROC for predicting 7-year metastasis for V_{AI} (0.854) was similar to that for V_{REF} (0.817; p=0.15) but trended towards being higher compared to NCCN+ category (0.769; p=0.06). The sensitivity and specificity for predicting 7-year metastasis at DIL volumes >0.5 mL were 94.3% and 44.1%, respectively, and at DIL volumes >2.0 mL were 68.1% and 82.9%, respectively (Fig. 1).

V_{AI} was significant on both univariable and multivariable analyses for predicting biochemical failure (Table 3; adjusted HR 1.61, 95% CI 1.23-2.12, p<0.001). For metastasis, only V_{AI} was significant on univariable and multivariable analysis (Table 4; adjusted HR, 1.71, 95% CI 1.08-2.70, p=0.02). V_{AI} was also associated with local recurrence risk on univariable analysis (HR 1.41, 95% CI 1.00-1.98, p=0.05; Table S10).

Fig. 2 shows cumulative incidences of biochemical failure, local failure, and metastasis for V_{AI} of 0-0.4, 0.5-1.9, and ≥2.0cm^3. Respective incidences of 7-year biochemical failure were 2.7%, 9.2%, and 23.7%. Respective incidences of 7-year local failure were 1.0%, 5.8%, and 5.3%. Respective incidences of 7-year metastasis were 0.7%, 2.9%, and 11.8%. Table S11 depicts the association of DIL volume ranges with NCCN+ and radiologic factors. Cumulative incidence curves for NCCN+ categories are shown in Fig. S2.

Discussion

Our study demonstrates that the prostate and DIL can be accurately delineated on mpMRI using nnUNet,[24] an out-of-the-box self-configuring deep learning segmentation algorithm. Our findings suggest that AI-determined DIL volume provides prognostic information comparable to both NCCN+ risk category and human-generated DIL volume for estimating biochemical recurrence and metastasis risk for localized prostate cancer treated with RT. Furthermore, for the subset of patients with PI-RADS scores, V_{AI} was more strongly associated with biochemical failure than other mpMRI parameters, including radiologic staging and PI-RADS score.

Though other studies have reported on the prognostic significance of tumor size for biochemical recurrence[8–11], ours is among the first to show that DIL volume can be reliably obtained with an AI algorithm. This is clinically meaningful because DIL volume determination can be time-consuming and is subject to significant inter-observer variability[31]. The AI algorithm, on the other hand, was able to generate the DIL volume in an efficient, automated, and standardized manner. Additionally, this work may be clinically impactful as it suggests that V_{AI} could provide important prognostic information even in the absence of a biopsy. With further validation, V_{AI} could have the potential to provide unique prognostic information in addition to that provided with current clinical and radiologic staging systems.

V_{AI} shows exceptional promise as a potential prognostic factor as it is a single, well-defined entity which may be generated in a systematic manner from mpMRI. Unlike NCCN risk categories[25], V_{AI} does not require the presence of a systematic biopsy, which may be important as MR-only biopsies are increasingly utilized[4]. Unlike radiomic

approaches, V_{AI} does not require the selection and regression of multiple features, which may have limited generalizability due to inter- and intra-scanner reproducibility[32]. Unlike genomic classifiers[33] or digital pathology approaches[34], no specific retrieval or handling of clinical specimens are required. Rather, prognostic information could be obtained directly from the mpMRI.

Additionally, this study provides support for extreme dose-escalation to large DILs, either through external beam RT[35] or brachytherapy[36,37] techniques. We found that V_{AI} to be associated with local recurrence risk. Given emerging evidence which strongly suggests that local failures seed distant metastases[38], extreme dose-escalation in appropriately selected patients with large DILs may be clinically impactful.

Our study showed that FP and FN lesions were smaller and less conspicuous than TP lesions. Our AI algorithm may have retained prognostic significance, despite missing such lesions, because of the weaker association between smaller lesions[8–11] and less conspicuous ADC values[39] with biochemical recurrence.

Key strengths of this study include the utilization of a well-annotated contemporary patient cohort treated with modern RT approaches. Furthermore, this study shows that mpMRI obtained after prior biopsy and with an endorectal coin contains important prognostic information that can be extracted by deep learning approaches. At our institution and likely many others, most mpMRIs acquired before 2018 had a pre-existing biopsy[1].

It should be noted that the algorithm developed within this study may not be generalizable to MRIs obtained using different scanning parameters or without endorectal coils. External validation would be necessary to establish generalizability. It

is our hope that deep learning algorithms with similar levels of performance may be built using nnUNet (or other deep learning segmentation approaches) for MRIs obtained under different conditions. Additionally, the AI algorithm missed significant disease in a subset of cases, including 12.9% of PI-RADS 5 lesions. As a result, AI DIL volume should not be used as a standalone entity.

Further research involves extending this work to multi-institutional datasets involving other scanning configurations and treatment types (e.g. surgery). We are very encouraged by the excellent performance of deep learning algorithms across multi-institution datasets[21,28] and look forward to the results of the PICAI-Eval challenge[29]. Only with additional information can we better understand the full potential of deep learning AI algorithms for providing potentially valuable prognostic information for patients.

Characteristics	TrainPIRADS (n=150)	TestPIRADS (n=83)	TestNoPIRADS (n= 205)
Age			
Median (IQR)	69.5 (64.0, 73.0)	68.0 (65.0, 73.0)	66 (61.0, 72.0)
Gleason grade group			
1	18 (12.0%)	17 (20.5%)	90 (43.9%)
2	62 (41.3%)	32 (38.6%)	70 (34.2%)
3	30 (20.0%)	11 (13.3%)	23 (11.2%)
4	22 (14.7%)	9 (10.8%)	9 (4.4%)
5	18 (12.0%)	14 (16.9%)	13 (6.3%)
Clinical T-stage			
T1-T2a	122 (81.3%)	69 (83.1%)	182 (88.8%)
T2b-T2c	12 (8.0%)	10 (12.1%)	17 (8.3%)
T3a	8 (5.3%)	3 (3.6%)	5 (2.4%)
T3b	8 (5.3%)	1 (1.2%)	1 (0.5%)
PSA			
0-9.9	107 (71.3%)	63 (75.9%)	179 (87.3%)
10.0-19.9	27 (18.0%)	12 (14.5%)	19 (9.3%)
20.0+	16 (10.7%)	8 (9.6%)	7 (3.4%)
PPC			
0-49%	80 (53.3%)	49 (59.0%)	161 (78.5%)
50-100%	65 (43.3%)	33 (39.8%)	41 (20.0%)
Missing	5 (3.3%)	1 (1.2%)	3 (1.5%)
NCCN+ category			
Low	13 (8.7%)	14 (16.9%)	80 (39.0%)
FIR	34 (22.7%)	14 (16.9%)	57 (27.8%)
UIR	52 (34.7%)	24 (28.9%)	38 (18.5%)
High	37 (24.7%)	28 (33.7%)	26 (12.7%)
Very High	14 (9.3%)	3 (3.61)	4 (2.0%)
Model			
Signa HDxt	52 (34.7%)	31 (37.3%)	154 (75.1%)
DISCOVERY MR750w	98 (65.3%)	52 (62.7%)	51 (24.9%)
PIRADS			
0-2	6 (4.0%)	10 (12.1%)	NA
3	15 (10.0%)	8 (9.6%)	NA
4	48 (32.0%)	24 (28.9%)	NA
5	81 (54.0%)	41 (49.4%)	NA
Radiologic T-stage			
T1	6 (4.00%)	10 (12.05%)	N/A
T2	83 (55.33%)	49 (59.04%)	N/A
T3a	45 (30.00%)	20 (24.10%)	N/A
T3b	16 (10.67%)	4 (4.82%)	N/A
Hemorrhage			
Yes	56 (37.33%)	28 (33.73%)	N/A

No	94 (62.67%)	55 (66.27%)	N/A
Time between biopsy and MRI (days)			
Patients with available data	148	82	204
Median (IQR)	46.5 (24.0, 72.3)	48.5 (22.5, 101.3)	36.0 (19.0, 59.3)
Time between MRI and Treatment (days)			
Median (IQR)	23.0 (7.3, 50.3)	44.0 (15.5, 66.0)	54.0 (27.0, 83.0)
Radiation Therapy			
EBRT	110 (73.3%)	54 (65.1%)	83 (40.5%)
Brachytherapy	21 (14.00%)	19 (22.9%)	117 (57.1%)
Combination	19 (12.7%)	10 (12.0%)	5 (2.4%)
Radiation Nodal Coverage			
Prostate/SV	132 (88.0%)	71 (85.5%)	199 (97.1%)
Pelvic Nodes	18 (12.0%)	12 (14.5%)	6 (2.9%)
Prostate EBRT Dose (cGy)			
7020	8 (5.33%)	5 (6.02%)	8 (3.90%)
7560-7920	100 (66.67%)	49 (59.04%)	75 (36.59%)
6000	2 (1.33%)	0 (0.00%)	0 (0.00%)
ADT Duration (mo)			
0-2.9	36 (24.0%)	28 (33.7%)	126 (61.4%)
3.0-6.0	71 (47.3%)	31 (37.4%)	53 (25.9%)
6.1-18.0	7 (4.7%)	4 (4.8%)	2 (1.0%)
>18.0	36 (24.0%)	20 (24.1%)	24 (11.7%)
Systemic therapy Intensification			
None	138 (92.0%)	77 (92.8%)	198 (96.5%)
Docetaxel	2 (1.3%)	1 (1.2%)	4 (2.0%)
Enzalutamide	10 (6.7%)	5 (6.0%)	3 (1.5%)
Year of Treatment Initiation			
2010-2014	52 (34.7%)	31 (37.3%)	187 (91.2%)
2015-2017	98 (65.3%)	52 (62.7%)	18 (8.8%)

289

290 Table 1: Baseline clinical, radiologic, and treatment factors.

291 SV: seminal vesicles. EBRT: external beam radiation therapy. Radiologic T stage and

292 hemorrhage data were not available for the TestNoPIRADS cohort.

Characteristic	HR (95% CI)	p-value	AHR (95% CI)	p-value
Clinical T-stage				
T1c-T2a (ref)	-	-		
T2b-T2c	0.88 [0.20, 3.76]	0.86		
T3a	2.51 [0.85, 7.38]	0.1		
T3b	1.09 [0.15, 8.14]	0.93		
Radiologic T-stage				
T1 (ref)	-	-		
T2	0.77 [0.10, 6.18]	0.81		
T3a	2.33 [0.30, 17.95]	0.42		
T3b	4.18 [0.51, 34.38]	0.18		
PIRADS				
0-4 (ref)	-	-	-	-
5	5.90 [2.05, 17.03]	0.001	2.47 [0.72, 8.29]	0.14
Age	0.91 [0.86, 0.96]	<0.001	0.93 [0.88, 0.98]	0.005
SOC	0.90 [0.40, 2.06]	0.81		
Year of Treatment	1.04 [0.88, 1.24]	0.62		
MRI model	1.07 [0.43, 2.62]	0.89		
AI Volume	1.89 [1.41, 2.53]	<0.001	1.54 [1.09, 2.17]	0.014

Table 2: Cox regression model for biochemical failure for the subset of 233 patients with available PI-RADS scores.

CI: confidence interval. HR: hazard ratio. AHR: adjusted hazard ratio. SOC: standard of care treatment. Ref: reference.

Characteristic	HR	p-value	AHR	p-value
NCCN+				
Low (ref)	-	-	-	-
FIR	1.72 [0.38, 7.71]	0.48	1.58 [0.34, 7.21]	0.56
UIR	6.59 [1.91, 22.73]	0.003	4.52 [1.26, 16.22]	0.02
High	9.51 [2.79, 32.49]	<0.001	4.64 [1.23, 17.52]	0.02
Very high	20.99 [5.50, 80.08]	<0.001	4.35 [0.89, 21.15]	0.07
Age	0.95 [0.91, 0.98]	0.006	0.93 [0.90, 0.97]	<0.001
SOC	0.89 [0.44, 1.78]	0.74		
Year of Treatment	1.15 [1.00, 1.33]	0.05	0.99 [0.74, 1.33]	0.96
MRI model	2.86 [1.47, 5.58]	0.002	0.64 [0.17, 2.49]	0.52
AI Volume	2.00 [1.62, 2.46]	<0.001	1.61 [1.23, 2.12]	<0.001

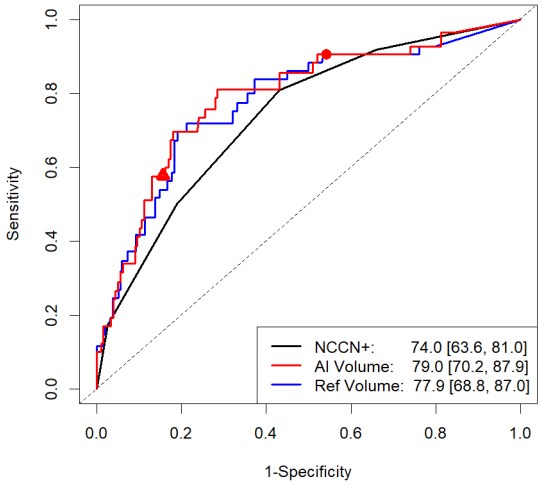
Table 3: Cox regression model for biochemical failure for all patients.

CI: confidence interval. HR: hazard ratio. AHR: adjusted hazard ratio. SOC: standard of care treatment. Ref: reference. FIR: favorable intermediate-risk. UIR: unfavorable intermediate-risk.

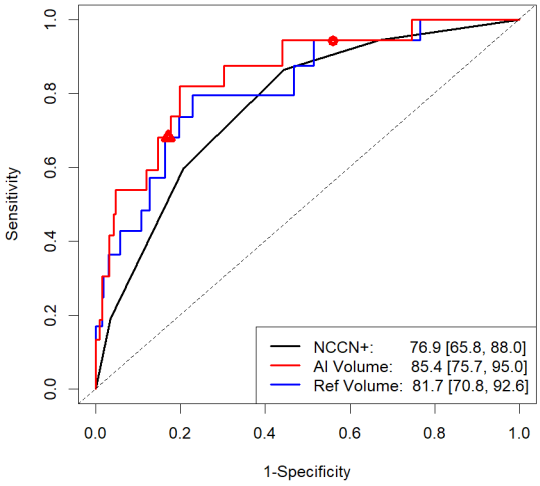
Characteristic	HR (95% CI)	p-value	AHR (95% CI)	p-value
NCCN+				
Low/FIR (ref)	-	-	-	-
UIR	3.07 [0.68, 13.76]	0.14	1.33 [0.27, 6.48]	0.73
High	10.63 [2.95, 38.29]	<0.001	3.18 [0.75, 13.60]	0.12
Very high	17.84 [3.94, 80.71]	<0.001	1.47 [0.21, 10.27]	0.7
Age	0.96 [0.90, 1.02]	0.016		
SOC	1.04 [0.35, 3.08]	0.94		
Year of Treatment	1.23 [0.97, 1.55]	0.08		
MRI model	5.36 [1.70, 16.86]	0.004	0.53 [0.17, 1.66]	0.27
Salvage ADT	29.53 [11.19, 77.92]	<0.001	12.58 [4.51, 31.11]	<0.001
AI Volume	2.29 [1.67, 3.15]	<0.001	1.71 [1.08, 2.70]	0.02

Table 4: Cox regression model for metastasis for all patients.

CI: confidence interval. HR: hazard ratio. AHR: adjusted hazard ratio. SOC: standard of care treatment. Ref: reference. FIR: favorable intermediate-risk. UIR: unfavorable intermediate-risk.

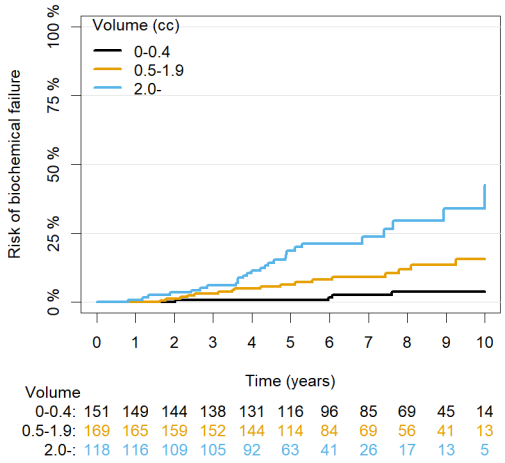


a)

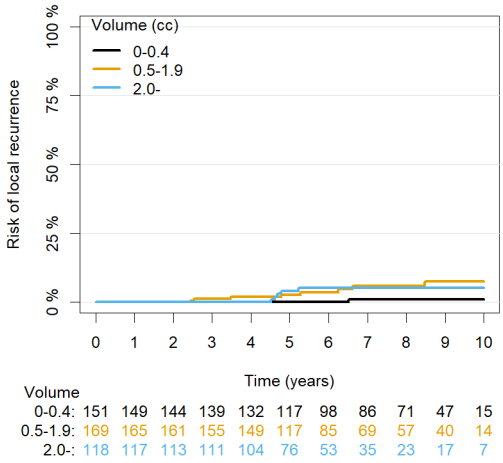


b)

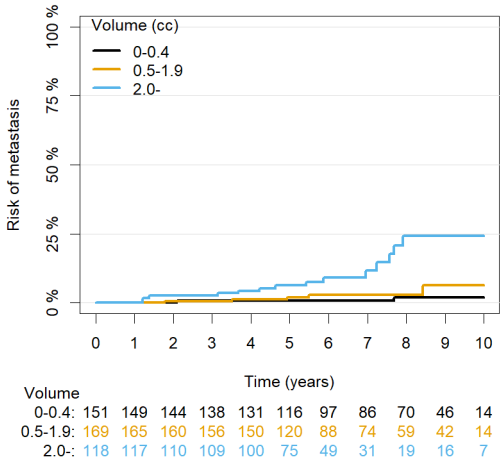
Fig 1: Receiver operating characteristics curves for 7-year biochemical failure (a) and metastasis (b), comparing AI DIL volume (red) against reference DIL volume (blue) and NCCN+ stratification (black). Circle and triangle are points on the AI volume receiver operating characteristics curve corresponding to thresholds of 0.5 cc and 2.0 cc, respectively.



a)



b)



c)

Fig 2: Cumulative incidence curves of DIL volume intervals (0-0.4, 0.5-1.9, 2.0- cc) with (a) biochemical failure, (b) local failure, and (c) metastasis.

References

1. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *The Lancet*. 2017;389(10071):815-822. doi:10.1016/S0140-6736(16)32401-1
2. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med*. 2018;378(19):1767-1777. doi:10.1056/NEJMoa1801993
3. Ahdoot M, Wilbur AR, Reese SE, et al. MRI-Targeted, Systematic, and Combined Biopsy for Prostate Cancer Diagnosis. *N Engl J Med*. 2020;382(10):917-928. doi:10.1056/NEJMoa1910038
4. Hugosson J, Månsson M, Wallström J, et al. Prostate Cancer Screening with PSA and MRI Followed by Targeted Biopsy Only. *N Engl J Med*. 2022;387(23):2126-2137. doi:10.1056/NEJMoa2209454
5. Rajwa P, Mori K, Huebner NA, et al. The Prognostic Association of Prostate MRI PI-RADS™ v2 Assessment Category and Risk of Biochemical Recurrence after Definitive Local Therapy for Prostate Cancer: A Systematic Review and Meta-Analysis. *J Urol*. 2021;206(3):507. doi:10.1097/JU.0000000000001821
6. Baboudjian M, Gondran-Tellier B, Touzani A, et al. Magnetic Resonance Imaging–based T-staging to Predict Biochemical Recurrence after Radical Prostatectomy: A Step Towards the iTNM Classification. *Eur Urol Oncol*. Published online October 21, 2022. doi:10.1016/j.euo.2022.09.005
7. Rakauskas A, Peters M, Ball D, et al. The impact of local staging of prostate cancer determined on MRI or DRE at time of radical prostatectomy on progression-free survival: A Will Rogers phenomenon. *Urol Oncol Semin Orig Investig*. Published online December 21, 2022. doi:10.1016/j.urolonc.2022.10.023
8. Hutten R, Khouri A, Parsons M, et al. The Clinical Significance of Maximum Tumor Diameter on MRI in Men Undergoing Radical Prostatectomy or Definitive Radiotherapy for Locoregional Prostate Cancer. *Clin Genitourin Cancer*. 2022;20(6):e453-e459. doi:10.1016/j.clgc.2022.06.010
9. Woo S, Han S, Kim TH, et al. Prognostic Value of Pretreatment MRI in Patients With Prostate Cancer Treated With Radiation Therapy: A Systematic Review and Meta-Analysis. *Am J Roentgenol*. 2019;214(3):597-604. doi:10.2214/AJR.19.21836
10. Jambor I, Falagario U, Ratnani P, et al. Prediction of biochemical recurrence in prostate cancer patients who underwent prostatectomy using routine clinical prostate multiparametric MRI and decipher genomic score. *J Magn Reson Imaging*. 2020;51(4):1075-1085. doi:10.1002/jmri.26928
11. Stabile A, Mazzone E, Cirulli GO, et al. Association Between Multiparametric Magnetic Resonance Imaging of the Prostate and Oncological Outcomes after Primary Treatment for Prostate Cancer: A Systematic Review and Meta-analysis. *Eur Urol Oncol*. 2021;4(4):519-528. doi:10.1016/j.euo.2020.11.008
12. Mazzone E, Gandaglia G, Ploussard G, et al. Risk Stratification of Patients Candidate to Radical Prostatectomy Based on Clinical and Multiparametric Magnetic Resonance Imaging Parameters: Development and External Validation of Novel Risk Groups. *Eur Urol*. 2022;81(2):193-203. doi:10.1016/j.eururo.2021.07.027

13. Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012. *Eur Radiol*. 2012;22(4):746-757. doi:10.1007/s00330-011-2377-y
14. Costa DN, Passoni NM, Leyendecker JR, et al. Diagnostic Utility of a Likert Scale Versus Qualitative Descriptors and Length of Capsular Contact for Determining Extraprostatic Tumor Extension at Multiparametric Prostate MRI. *Am J Roentgenol*. 2018;210(5):1066-1072. doi:10.2214/AJR.17.18849
15. Mehralivand S, Shih JH, Harmon S, et al. A Grading System for the Assessment of Risk of Extraprostatic Extension of Prostate Cancer at Multiparametric MRI. *Radiology*. 2019;290(3):709-719. doi:10.1148/radiol.2018181278
16. Park KJ, Kim M hyun, Kim JK. Extraprostatic Tumor Extension: Comparison of Preoperative Multiparametric MRI Criteria and Histopathologic Correlation after Radical Prostatectomy. *Radiology*. 2020;296(1):87-95. doi:10.1148/radiol.2020192133
17. Muller BG, Shih JH, Sankineni S, et al. Prostate Cancer: Interobserver Agreement and Accuracy with the Revised Prostate Imaging Reporting and Data System at Multiparametric MR Imaging. *Radiology*. 2015;277(3):741-750. doi:10.1148/radiol.2015142818
18. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists. *Radiology*. 2016;280(3):793-804. doi:10.1148/radiol.2016152542
19. Smith CP, Harmon SA, Barrett T, et al. Intra- and interreader reproducibility of PI-RADSv2: A multireader study. *J Magn Reson Imaging*. 2019;49(6):1694-1703. doi:10.1002/jmri.26555
20. Westphalen AC, McCulloch CE, Anaokar JM, et al. Variability of the Positive Predictive Value of PI-RADS for Prostate MRI across 26 Centers: Experience of the Society of Abdominal Radiology Prostate Cancer Disease-focused Panel. *Radiology*. 2020;296(1):76-84. doi:10.1148/radiol.2020190646
21. Panebianco V, Giganti F, Kitzing YX, et al. An update of pitfalls in prostate mpMRI: a practical approach through the lens of PI-RADS v. 2 guidelines. *Insights Imaging*. 2018;9(1):87-101. doi:10.1007/s13244-017-0578-x
22. Hosseinzadeh M, Saha A, Brand P, Slootweg I, de Rooij M, Huisman H. Deep learning–assisted prostate cancer detection on bi-parametric MRI: minimum training data size requirements and effect of prior knowledge. *Eur Radiol*. 2022;32(4):2224-2234. doi:10.1007/s00330-021-08320-y
23. Mehralivand S, Yang D, Harmon SA, et al. Deep learning-based artificial intelligence for prostate cancer detection at biparametric MRI. *Abdom Radiol*. 2022;47(4):1425-1434. doi:10.1007/s00261-022-03419-2
24. 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(2):203-211. doi:10.1038/s41592-020-01008-z
25. Mohler JL, Antonarakis ES, Armstrong AJ, et al. Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw JNCCN*. 2019;17(5):479-505. doi:10.6004/jnccn.2019.0023
26. Attard G, Murphy L, Clarke NW, et al. Abiraterone acetate and prednisolone with or without enzalutamide for high-risk non-metastatic prostate cancer: a meta-analysis of primary results from two randomised controlled phase 3 trials of the STAMPEDE

platform protocol. *The Lancet*. 2022;399(10323):447-460. doi:10.1016/S0140-6736(21)02437-5

27. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol*. 2016;69(1):16-40. doi:10.1016/j.eururo.2015.08.052
28. Fedorov A, Khallaghi S, Sánchez CA, et al. Open-source image registration for MRI–TRUS fusion-guided prostate interventions. *Int J Comput Assist Radiol Surg*. 2015;10(6):925-934. doi:10.1007/s11548-015-1180-7
29. Saha A, Twilt JJ, Bosma JS, et al. *Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI: The PI-CAI Challenge (Study Protocol)*. Zenodo; 2022. doi:10.5281/zenodo.6667655
30. Turchan WT, Kauffmann G, Patel P, Oto A, Liauw SL. PI-RADS score is associated with biochemical control and distant metastasis in men with intermediate-risk and high-risk prostate cancer treated with radiation therapy. *Urol Oncol Semin Orig Investig*. 2020;38(6):600.e1-600.e8. doi:10.1016/j.urolonc.2019.12.015
31. Steenbergen P, Haustermans K, Lerut E, et al. Prostate tumor delineation using multiparametric magnetic resonance imaging: Inter-observer variability and pathology validation. *Radiother Oncol*. 2015;115(2):186-190. doi:10.1016/j.radonc.2015.04.012
32. Fedorov A, Vangel MG, Tempany CM, Fennessy FM. Multiparametric Magnetic Resonance Imaging of the Prostate: Repeatability of Volume and Apparent Diffusion Coefficient Quantification. *Invest Radiol*. 2017;52(9):538-546. doi:10.1097/RLI.0000000000000382
33. Spratt DE, Yousefi K, Deheshi S, et al. Individual Patient-Level Meta-Analysis of the Performance of the Decipher Genomic Classifier in High-Risk Men After Prostatectomy to Predict Development of Metastatic Disease. *J Clin Oncol*. 2017;35(18):1991-1998. doi:10.1200/JCO.2016.70.2811
34. Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *Npj Digit Med*. 2022;5(1):1-8. doi:10.1038/s41746-022-00613-w
35. Kerkmeijer LGW, Groen VH, Pos FJ, 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*. Published online January 20, 2021. doi:10.1200/JCO.20.02873
36. Morris WJ, Tyldesley S, Rodda S, et al. Androgen Suppression Combined with Elective Nodal and Dose Escalated Radiation Therapy (the ASCENDE-RT Trial): An Analysis of Survival Endpoints for a Randomized Trial Comparing a Low-Dose-Rate Brachytherapy Boost to a Dose-Escalated External Beam Boost for High- and Intermediate-risk Prostate Cancer. *Int J Radiat Oncol*. 2017;98(2):275-285. doi:10.1016/j.ijrobp.2016.11.026
37. Hoskin PJ, Rojas AM, Ostler PJ, Bryant L, Lowe GJ. Randomised trial of external-beam radiotherapy alone or with high-dose-rate brachytherapy for prostate cancer: Mature 12-year results. *Radiother Oncol*. 2021;154:214-219. doi:10.1016/j.radonc.2020.09.047
38. Ma TM, Chu FI, Sandler H, et al. Local Failure Events in Prostate Cancer Treated with Radiotherapy: A Pooled Analysis of 18 Randomized Trials from the Meta-

460 analysis of Randomized Trials in Cancer of the Prostate Consortium (LEVIATHAN).
461 *Eur Urol*. 2022;82(5):487-498. doi:10.1016/j.eururo.2022.07.011
462 39. Chatterjee A, Turchan WT, Fan X, et al. Can Pre-treatment Quantitative Multi-
463 parametric MRI Predict the Outcome of Radiotherapy in Patients with Prostate
464 Cancer? *Acad Radiol*. 2022;29(7):977-985. doi:10.1016/j.acra.2021.09.012
465