

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

3
4 David D Yang^{a,b,*}, Leslie K Lee^{b,*}, James MG Tsui^c, Jonathan E Leeman^{a,b}, Katie N
5 Lee^{a,b}, Heather M McClure^a, Atchar Sudhyadhom^{a,b}, Christian V Guthier^{a,b}, Kent W
6 Mouw^{a,b}, Neil E Martin^{a,b}, Peter F Orio^{a,b}, Paul L Nguyen^{a,b}, Anthony V D'Amico^{a,b},
7 Martin T King^{a,b}

8
9 ^aDana-Farber Cancer Institute, Boston, MA, USA

10 ^bBrigham and Women's Hospital, Boston, MA, USA

11 ^cMcGill University, Montreal, Canada

12

13 *Equal contribution

14

15 **Corresponding Author**

16 Martin T King

17 75 Francis St

18 Boston, MA 02115

19 Telephone: +1-617-732-5640

20 Fax: +1-617-632-4247

21 Email: Martin_King@dfci.harvard.edu

22

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

25

26 Abstract word count: 300

27 Text word count: 2794 (including abstract)

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

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

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

37 **Intervention:** RT.

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

41 **Results and Limitations:** The artificial intelligence (AI) model identified DILs with an
42 area under the receiver operating characteristic curve (AUROC) of 0.827 at the patient
43 level. For the 233 patients with available PI-RADS scores, with a median follow-up of
44 5.6 years, AI-defined DIL volume was significantly associated with biochemical failure
45 (adjusted hazard ratio 1.54, 95% confidence interval 1.09-2.17, $p=0.014$) after
46 adjustment for PI-RADS score. Among all 438 patients with a median follow-up of 6.9
47 years, the AUROC for predicting 7-year biochemical failure for AI volume (0.790) was
48 similar to that for an expanded National Comprehensive Cancer Network (NCCN+)
49 category ($p=0.17$). The AUROC for predicting 7-year metastasis for AI volume trended
50 towards being higher compared to NCCN+ categories (0.854 vs 0.769, $p=0.06$).

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

55 **Introduction**

56 Multi-parametric magnetic resonance imaging (mpMRI) has revolutionized the
57 management of localized prostate cancer. mpMRI-based biopsy strategies have been
58 shown to improve the detection of clinically significant disease while reducing the
59 detection of clinically insignificant disease¹⁻⁴. Furthermore, the characteristics of
60 dominant intraprostatic lesions (DILs) on mpMRI, including PI-RADS scores⁵, radiologic
61 T-stage^{6,7}, and lesion size⁸⁻¹¹, have been shown to be prognostic. Recently, a
62 nomogram based on clinical and radiologic parameters was found to be highly
63 prognostic for early biochemical recurrence after radical prostatectomy¹².

64 However, the assignment of various clinical features on mpMRI is subject to
65 significant inter-observer variability. For example, multiple grading systems¹³⁻¹⁵ exist for
66 assigning extracapsular extension (EPE), each with differing sensitivities for the
67 detection of histologic EPE extent¹⁶. Multi-reader studies have reported moderate inter-
68 observer variability in the reporting of Prostate Imaging Reporting and Data Systems
69 (PI-RADS) v2.0 scores¹⁷⁻¹⁹. A recent study demonstrated that the positive predictive
70 value of PI-RADS scores for detecting high-grade prostate cancer was low (35%) and
71 variable across 26 centers²⁰. The varied performance of mpMRI is likely multifactorial
72 and could be partially attributed to technical and interpretative factors, as well as benign
73 conditions that mimic malignancy²¹.

74 Given these limitations, there has been considerable interest in the development
75 of artificial intelligence (AI) algorithms for supporting the radiologist's workflow. Although
76 the performance of AI algorithms has not been shown to match that of radiologists for
77 the detection of clinically significant prostate cancer²², recent work suggests that deep

78 learning algorithms could be used as an adjunct for assisting radiologists²³. However,
79 less is known about the prognostic information provided by deep learning algorithms,
80 particularly in comparison to current staging systems. The purpose of this study is to
81 evaluate whether the mpMRI DIL volume, as determined by a deep learning
82 segmentation algorithm²⁴, could provide prognostic information for patients with
83 localized prostate cancer treated with definitive radiation therapy (RT).

84

85 **Methods**

86 *Clinical datasets*

87 We conducted a retrospective study of 438 patients with cT1cN0M0 prostate
88 cancer who underwent an endorectal coil, high B-value, 3-Tesla mpMRI and were
89 treated with definitive prostate RT, both at our institution, between 2010 and 2017
90 (Table S1). mpMRI were obtained with either General Electric (GE) Signa HDxt (n=237)
91 or DISCOVERY MR750w (n=201) (Table S2). Apparent diffusion coefficient
92 (ADC)/diffusion-weighted imaging (DWI) were obtained at the same resolution (0.70 x
93 0.70 x 3-4 mm) as the T2-weighted images. All research was conducted with approval
94 from the institutional review board.

95 We captured baseline clinical characteristics and treatment characteristics (Table
96 1). We stratified patients into an expanded National Cancer Center Network (NCCN+)
97 classification, based on the 4-tiered NCCN stratification (low, favorable intermediate-
98 risk, unfavorable intermediate-risk, high-risk)²⁵, with the addition of a very high-risk
99 category per eligibility criteria from recent STAMPEDE trials (≥ 2 of the following: cT3-
100 T4, Gleason score 8-10, or PSA ≥ 40 ng/mL)²⁶. We captured radiologic staging
101 parameters, including PI-RADS v2.0 scores²⁷, which were assessed by a subspecialty
102 abdominal radiologist (L.K.L.) for 83 patients who were scanned in 2010-2013 and
103 available for all patients after mid-2015, as well as radiologic T-stage (T1: no visible
104 tumor on mpMRI, T2: organ-confined, T3a: EPE, T3b: seminal vesicle invasion [SVI])
105 based on mpMRI. Lastly, we captured outcomes including biochemical failure (nadir+2
106 ng/mL), development of metastasis (non-regional nodal or distant disease on bone
107 scan, computed tomography, or positron emission tomography), and survival.

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

115

116 *nnUNet algorithm*

117 We utilized the publicly available nnUNet²⁴ deep learning algorithm for automated
118 delineations of the TZ, PZ, and DIL. First, T2, ADC, and DWI images were downloaded
119 onto a research workstation. Images were anonymized and cropped into smaller arrays
120 (128 x 128 x 15 or 20 pixels). T2-weighted images were registered onto the ADC
121 images, utilizing the best performing registration among 4 algorithms (rigid, rigid + B-
122 spline, affine, affine + B-spline) of prostate segmentation distance maps²⁸. The median
123 registration Dice coefficient was 0.92 (interquartile range [IQR] 0.90-0.94).

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

129

130 *Assessment of AI DIL volume performance*

131 We reported the Dice coefficients for the TZ and PZ, as noted by nnUNet. For the
132 DIL, we analyzed the detection performance (patient-level area under the receiver
133 operator characteristic curve [AUROC] and lesion-level specificity at 0.5 false positive
134 per image) utilizing PICA_eval²⁹. To assess whether false positive (FP) or false
135 negative (FN) lesions were smaller and less conspicuous than true positives (TP), we
136 classified AI lesions as TP, FP, or FN, based on whether a reference DIL was present
137 and correctly identified (TP: Dice coefficient $\geq 10\%$ ²⁹), identified but not present (FP), or
138 present but not identified or with insufficient Dice coefficient overlap ($< 10\%$) (FN). We
139 analyzed lesion-level DIL volumes and contrast ratios (ratio of ADC intensities within the
140 DIL versus a surrounding 2mm ring) based on classification status. Finally, we reported
141 the similarity metrics (Dice coefficient) for TP lesions. All processing was performed
142 using Python version 3.8.12 with the SimpleITK toolkit.

143

144 *Comparison of AI and reference DIL volumes for sextant subset*

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

150

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

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

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

163

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

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

175 **Results**

176 *Assessment of AI DIL performance*

177 For the TestPIRADS cohort, median Dice coefficients for the TZ and PZ were
178 0.89 and 0.84, respectively (Table S3). The nnUNet exhibited a patient-level AUROC of
179 0.827 and a lesion-level sensitivity of 72.7% at a false positive rate of 0.5 case per
180 image (Table S4). For patient-level analysis (Table S5), there was no significant
181 difference between V_{AI} (median 1.02cm³, IQR 0.40-2.34) and V_{REF} (median 1.31cm³,
182 IQR 0.37-2.61, $p=0.66$), with Pearson's correlation coefficient of 0.79. For lesion-level
183 analysis (Table S6), the total volume of AI TP lesions (median 1.06cm³, IQR 0.47-1.85)
184 was significantly greater than for FP (median 0.25cm³, IQR 0.09-0.58; $p<0.001$) and FN
185 (median 0.37cm³, IQR 0.21-0.88, $p=0.007$) lesions. The TP lesions were also more
186 conspicuous (median contrast ratio 0.72, IQR 0.68-0.76) than FP (median 0.83, IQR
187 0.79-0.86; $p<0.001$) and FN (median 0.81, IQR 0.79-0.88; $p<0.001$) lesions. The
188 median Dice coefficient for TP lesions was 0.74. Examples of TP, FP, and FN lesions
189 are shown in Fig. S1.

190

191 *Comparison of AI and reference DIL contours for sextant subset*

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

195

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

197 For the 233 patients with PI-RADS scores (TrainPIRADS and TestPIRADS
198 cohorts), with median follow-up of 5.6 years, there were 28 biochemical failures. AI
199 algorithm detected 41.2%, 68.3%, and 87.1% of PI-RADS3, 4, and 5 lesions,
200 respectively (Table S8). PIRADS 5 vs 0-4 (HR 5.90, 95% confidence interval [CI] 2.05-
201 17.03, $p=0.001$) and V_{AI} (HR 1.89 per $\log \text{cm}^3$ increase, 95% CI 1.41-2.53, $p<0.001$)
202 were significantly associated with biochemical failure on univariable analysis (Table 2).
203 However, only V_{AI} retained significance on multivariable analysis (adjusted HR 1.54,
204 95% CI 1.09-2.17; $p=0.014$). The association between V_{AI} and metastasis was similarly
205 significant (adjusted HR 1.71, 95% CI 1.04-2.80, $p=0.04$; Table S9).

206

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

208 Among all 438 patients with a median follow-up of 6.9 years, there were 49
209 biochemical failures, 14 local failures, and 22 metastases. The AUROC for 7-year
210 biochemical failure for V_{AI} (0.790) was similar to that for V_{REF} (0.779; $p=0.42$) and
211 NCCN+ category (0.740; $p=0.17$). The sensitivity and specificity for 7-year biochemical
212 failure at DIL volumes $>0.5 \text{ mL}$ were 90.5% and 45.8%, respectively, and at DIL
213 volumes $>2.0 \text{ mL}$, were 57.7% and 84.4%, respectively. The AUROC for predicting 7-
214 year metastasis for V_{AI} (0.854) was similar to that for V_{REF} (0.817; $p=0.15$) but trended
215 towards being higher compared to NCCN+ category (0.769; $p=0.06$). The sensitivity and
216 specificity for predicting 7-year metastasis at DIL volumes $>0.5 \text{ mL}$ were 94.3% and
217 44.1%, respectively, and at DIL volumes $>2.0 \text{ mL}$ were 68.1% and 82.9%, respectively
218 (Fig. 1).

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

224 Fig. 2 shows cumulative incidences of biochemical failure, local failure, and
225 metastasis for V_{AI} of 0-0.4, 0.5-1.9, and $\geq 2.0 \text{cm}^3$. Respective incidences of 7-year
226 biochemical failure were 2.7%, 9.2%, and 23.7%. Respective incidences of 7-year local
227 failure were 1.0%, 5.8%, and 5.3%. Respective incidences of 7-year metastasis were
228 0.7%, 2.9%, and 11.8%. Table S11 depicts the association of DIL volume ranges with
229 NCCN+ and radiologic factors. Cumulative incidence curves for NCCN+ categories are
230 shown in Fig. S2.

231 **Discussion**

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

240 Though other studies have reported on the prognostic significance of tumor size
241 for biochemical recurrence⁸⁻¹¹, ours is among the first to show that DIL volume can be
242 reliably obtained with an AI algorithm. This is clinically meaningful because DIL volume
243 determination can be time-consuming and is subject to significant inter-observer
244 variability³¹. The AI algorithm, on the other hand, was able to generate the DIL volume
245 in an efficient, automated, and standardized manner. Additionally, this work may be
246 clinically impactful as it suggests that V_{AI} could provide important prognostic information
247 even in the absence of a biopsy. With further validation, V_{AI} could have the potential to
248 provide unique prognostic information in addition to that provided with current clinical
249 and radiologic staging systems.

250 V_{AI} shows exceptional promise as a potential prognostic factor as it is a single,
251 well-defined entity which may be generated in a systematic manner from mpMRI. Unlike
252 NCCN risk categories²⁵, V_{AI} does not require the presence of a systematic biopsy, which
253 may be important as MR-only biopsies are increasingly utilized⁴. Unlike radiomic

254 approaches, V_{AI} does not require the selection and regression of multiple features,
255 which may have limited generalizability due to inter- and intra-scanner reproducibility³².
256 Unlike genomic classifiers³³ or digital pathology approaches³⁴, no specific retrieval or
257 handling of clinical specimens are required. Rather, prognostic information could be
258 obtained directly from the mpMRI.

259 Additionally, this study provides support for extreme dose-escalation to large
260 DILs, either through external beam RT³⁵ or brachytherapy^{36,37} techniques. We found
261 that V_{AI} to be associated with local recurrence risk. Given emerging evidence which
262 strongly suggests that local failures seed distant metastases³⁸, extreme dose-escalation
263 in appropriately selected patients with large DILs may be clinically impactful.

264 Our study showed that FP and FN lesions were smaller and less conspicuous
265 than TP lesions. Our AI algorithm may have retained prognostic significance, despite
266 missing such lesions, because of the weaker association between smaller lesions⁸⁻¹¹
267 and less conspicuous ADC values³⁹ with biochemical recurrence.

268 Key strengths of this study include the utilization of a well-annotated
269 contemporary patient cohort treated with modern RT approaches. Furthermore, this
270 study shows that mpMRI obtained after prior biopsy and with an endorectal coil
271 contains important prognostic information that can be extracted by deep learning
272 approaches. At our institution and likely many others, most mpMRIs acquired before
273 2018 had a pre-existing biopsy¹.

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

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

282 Further research involves extending this work to multi-institutional datasets
283 involving other scanning configurations and treatment types (e.g. surgery). We are very
284 encouraged by the excellent performance of deep learning algorithms across multi-
285 institution datasets^{21,28} and look forward to the results of the PICA-Eval challenge²⁹.
286 Only with additional information can we better understand the full potential of deep
287 learning AI algorithms for providing potentially valuable prognostic information for
288 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

293
 294 Table 2: Cox regression model for biochemical failure for the subset of 233 patients with
 295 available PI-RADS scores.

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

298

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

299

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

301 CI: confidence interval. HR: hazard ratio. AHR: adjusted hazard ratio. SOC: standard of

302 care treatment. Ref: reference. FIR: favorable intermediate-risk. UIR: unfavorable

303 intermediate-risk.

304

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

305

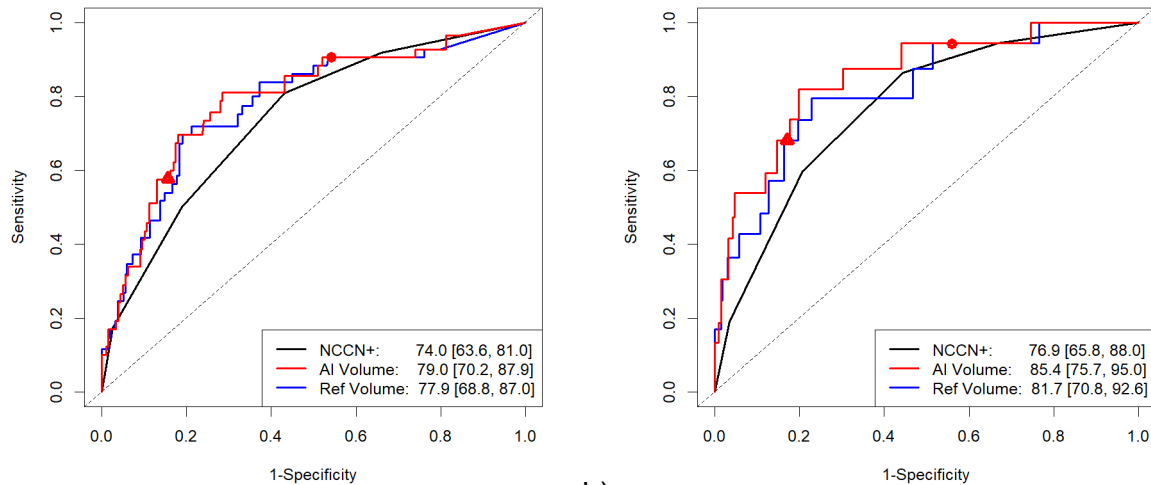
306 Table 4: Cox regression model for metastasis for all patients.

307 CI: confidence interval. HR: hazard ratio. AHR: adjusted hazard ratio. SOC: standard of

308 care treatment. Ref: reference. FIR: favorable intermediate-risk. UIR: unfavorable

309 intermediate-risk.

310



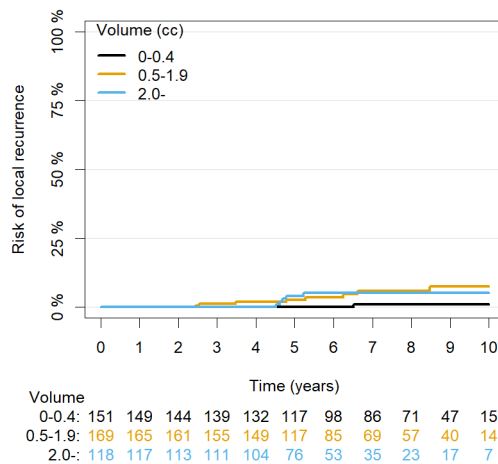
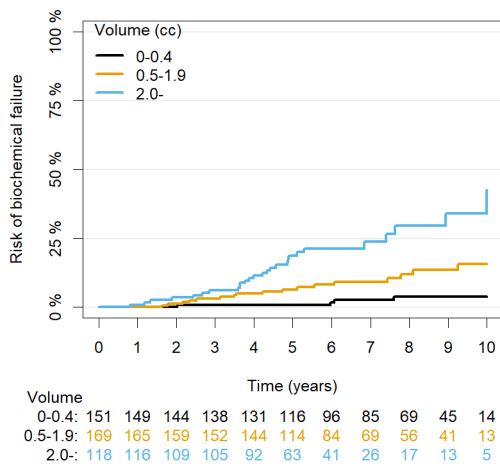
311 a)

b)

312

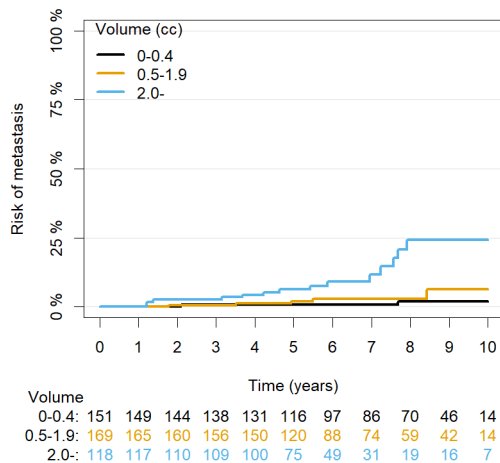
313 Fig 1: Receiver operating characteristics curves for 7-year biochemical failure (a) and
314 metastasis (b), comparing AI DIL volume (red) against reference DIL volume (blue) and
315 NCCN+ stratification (black). Circle and triangle are points on the AI volume receiver
316 operating characteristics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

318



319 a)

b)



320 c)

321 Fig 2: Cumulative incidence curves of DIL volume intervals (0-0.4, 0.5-1.9, 2.0- cc) with

322 (a) biochemical failure, (b) local failure, and (c) metastasis.

323 References

- 324 1. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-
325 parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating
326 confirmatory study. *The Lancet*. 2017;389(10071):815-822. doi:10.1016/S0140-
327 6736(16)32401-1
- 328 2. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy
329 for Prostate-Cancer Diagnosis. *N Engl J Med*. 2018;378(19):1767-1777.
330 doi:10.1056/NEJMoa1801993
- 331 3. Ahdoot M, Wilbur AR, Reese SE, et al. MRI-Targeted, Systematic, and Combined
332 Biopsy for Prostate Cancer Diagnosis. *N Engl J Med*. 2020;382(10):917-928.
333 doi:10.1056/NEJMoa1910038
- 334 4. Hugosson J, Månsson M, Wallström J, et al. Prostate Cancer Screening with PSA
335 and MRI Followed by Targeted Biopsy Only. *N Engl J Med*. 2022;387(23):2126-
336 2137. doi:10.1056/NEJMoa2209454
- 337 5. Rajwa P, Mori K, Huebner NA, et al. The Prognostic Association of Prostate MRI PI-
338 RADS™ v2 Assessment Category and Risk of Biochemical Recurrence after
339 Definitive Local Therapy for Prostate Cancer: A Systematic Review and Meta-
340 Analysis. *J Urol*. 2021;206(3):507. doi:10.1097/JU.0000000000001821
- 341 6. Baboudjian M, Gondran-Tellier B, Touzani A, et al. Magnetic Resonance Imaging–
342 based T-staging to Predict Biochemical Recurrence after Radical Prostatectomy: A
343 Step Towards the iTNM Classification. *Eur Urol Oncol*. Published online October 21,
344 2022. doi:10.1016/j.euo.2022.09.005
- 345 7. Rakauskas A, Peters M, Ball D, et al. The impact of local staging of prostate cancer
346 determined on MRI or DRE at time of radical prostatectomy on progression-free
347 survival: A Will Rogers phenomenon. *Urol Oncol Semin Orig Investig*. Published
348 online December 21, 2022. doi:10.1016/j.urolonc.2022.10.023
- 349 8. Hutten R, Khouri A, Parsons M, et al. The Clinical Significance of Maximum Tumor
350 Diameter on MRI in Men Undergoing Radical Prostatectomy or Definitive
351 Radiotherapy for Locoregional Prostate Cancer. *Clin Genitourin Cancer*.
352 2022;20(6):e453-e459. doi:10.1016/j.clgc.2022.06.010
- 353 9. Woo S, Han S, Kim TH, et al. Prognostic Value of Pretreatment MRI in Patients With
354 Prostate Cancer Treated With Radiation Therapy: A Systematic Review and Meta-
355 Analysis. *Am J Roentgenol*. 2019;214(3):597-604. doi:10.2214/AJR.19.21836
- 356 10. Jambor I, Falagario U, Ratnani P, et al. Prediction of biochemical recurrence in
357 prostate cancer patients who underwent prostatectomy using routine clinical
358 prostate multiparametric MRI and decipher genomic score. *J Magn Reson Imaging*.
359 2020;51(4):1075-1085. doi:10.1002/jmri.26928
- 360 11. Stabile A, Mazzone E, Cirulli GO, et al. Association Between Multiparametric
361 Magnetic Resonance Imaging of the Prostate and Oncological Outcomes after
362 Primary Treatment for Prostate Cancer: A Systematic Review and Meta-analysis.
363 *Eur Urol Oncol*. 2021;4(4):519-528. doi:10.1016/j.euo.2020.11.008
- 364 12. Mazzone E, Gandaglia G, Ploussard G, et al. Risk Stratification of Patients
365 Candidate to Radical Prostatectomy Based on Clinical and Multiparametric Magnetic
366 Resonance Imaging Parameters: Development and External Validation of Novel
367 Risk Groups. *Eur Urol*. 2022;81(2):193-203. doi:10.1016/j.eururo.2021.07.027

- 368 13. Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012.
369 *Eur Radiol.* 2012;22(4):746-757. doi:10.1007/s00330-011-2377-y
- 370 14. Costa DN, Passoni NM, Leyendecker JR, et al. Diagnostic Utility of a Likert Scale
371 Versus Qualitative Descriptors and Length of Capsular Contact for Determining
372 Extraprostatic Tumor Extension at Multiparametric Prostate MRI. *Am J Roentgenol.*
373 2018;210(5):1066-1072. doi:10.2214/AJR.17.18849
- 374 15. Mehralivand S, Shih JH, Harmon S, et al. A Grading System for the Assessment of
375 Risk of Extraprostatic Extension of Prostate Cancer at Multiparametric MRI.
376 *Radiology.* 2019;290(3):709-719. doi:10.1148/radiol.2018181278
- 377 16. Park KJ, Kim M hyun, Kim JK. Extraprostatic Tumor Extension: Comparison of
378 Preoperative Multiparametric MRI Criteria and Histopathologic
379 Correlation after Radical Prostatectomy. *Radiology.* 2020;296(1):87-95.
380 doi:10.1148/radiol.2020192133
- 381 17. Muller BG, Shih JH, Sankineni S, et al. Prostate Cancer: Interobserver Agreement
382 and Accuracy with the Revised Prostate Imaging Reporting and Data System at
383 Multiparametric MR Imaging. *Radiology.* 2015;277(3):741-750.
384 doi:10.1148/radiol.2015142818
- 385 18. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver Reproducibility of
386 the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate
387 Radiologists. *Radiology.* 2016;280(3):793-804. doi:10.1148/radiol.2016152542
- 388 19. Smith CP, Harmon SA, Barrett T, et al. Intra- and interreader reproducibility of PI-
389 RADSv2: A multireader study. *J Magn Reson Imaging.* 2019;49(6):1694-1703.
390 doi:10.1002/jmri.26555
- 391 20. Westphalen AC, McCulloch CE, Anaokar JM, et al. Variability of the Positive
392 Predictive Value of PI-RADS for Prostate MRI across 26 Centers: Experience of the
393 Society of Abdominal Radiology Prostate Cancer Disease-focused Panel.
394 *Radiology.* 2020;296(1):76-84. doi:10.1148/radiol.2020190646
- 395 21. Panebianco V, Giganti F, Kitzing YX, et al. An update of pitfalls in prostate mpMRI: a
396 practical approach through the lens of PI-RADS v. 2 guidelines. *Insights Imaging.*
397 2018;9(1):87-101. doi:10.1007/s13244-017-0578-x
- 398 22. Hosseinzadeh M, Saha A, Brand P, Slootweg I, de Rooij M, Huisman H. Deep
399 learning-assisted prostate cancer detection on bi-parametric MRI: minimum training
400 data size requirements and effect of prior knowledge. *Eur Radiol.* 2022;32(4):2224-
401 2234. doi:10.1007/s00330-021-08320-y
- 402 23. Mehralivand S, Yang D, Harmon SA, et al. Deep learning-based artificial intelligence
403 for prostate cancer detection at biparametric MRI. *Abdom Radiol.* 2022;47(4):1425-
404 1434. doi:10.1007/s00261-022-03419-2
- 405 24. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-
406 configuring method for deep learning-based biomedical image segmentation. *Nat*
407 *Methods.* 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z
- 408 25. Mohler JL, Antonarakis ES, Armstrong AJ, et al. Prostate Cancer, Version 2.2019,
409 NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw JNCCN.*
410 2019;17(5):479-505. doi:10.6004/jnccn.2019.0023
- 411 26. Attard G, Murphy L, Clarke NW, et al. Abiraterone acetate and prednisolone with or
412 without enzalutamide for high-risk non-metastatic prostate cancer: a meta-analysis
413 of primary results from two randomised controlled phase 3 trials of the STAMPEDE

- 414 platform protocol. *The Lancet*. 2022;399(10323):447-460. doi:10.1016/S0140-
415 6736(21)02437-5
- 416 27. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging - Reporting
417 and Data System: 2015, Version 2. *Eur Urol*. 2016;69(1):16-40.
418 doi:10.1016/j.eururo.2015.08.052
- 419 28. Fedorov A, Khallaghi S, Sánchez CA, et al. Open-source image registration for
420 MRI–TRUS fusion-guided prostate interventions. *Int J Comput Assist Radiol Surg*.
421 2015;10(6):925-934. doi:10.1007/s11548-015-1180-7
- 422 29. Saha A, Twilt JJ, Bosma JS, et al. *Artificial Intelligence and Radiologists at Prostate
423 Cancer Detection in MRI: The PI-CAI Challenge (Study Protocol)*. Zenodo; 2022.
424 doi:10.5281/zenodo.6667655
- 425 30. Turchan WT, Kauffmann G, Patel P, Oto A, Liauw SL. PI-RADS score is associated
426 with biochemical control and distant metastasis in men with intermediate-risk and
427 high-risk prostate cancer treated with radiation therapy. *Urol Oncol Semin Orig
428 Investig*. 2020;38(6):600.e1-600.e8. doi:10.1016/j.urolonc.2019.12.015
- 429 31. Steenbergen P, Haustermans K, Lerut E, et al. Prostate tumor delineation using
430 multiparametric magnetic resonance imaging: Inter-observer variability and
431 pathology validation. *Radiother Oncol*. 2015;115(2):186-190.
432 doi:10.1016/j.radonc.2015.04.012
- 433 32. Fedorov A, Vangel MG, Tempny CM, Fennessy FM. Multiparametric Magnetic
434 Resonance Imaging of the Prostate: Repeatability of Volume and Apparent Diffusion
435 Coefficient Quantification. *Invest Radiol*. 2017;52(9):538-546.
436 doi:10.1097/RLI.0000000000000382
- 437 33. Spratt DE, Yousefi K, Deheshi S, et al. Individual Patient-Level Meta-Analysis of the
438 Performance of the Decipher Genomic Classifier in High-Risk Men After
439 Prostatectomy to Predict Development of Metastatic Disease. *J Clin Oncol*.
440 2017;35(18):1991-1998. doi:10.1200/JCO.2016.70.2811
- 441 34. Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via
442 multi-modal deep learning on randomized phase III clinical trials. *Npj Digit Med*.
443 2022;5(1):1-8. doi:10.1038/s41746-022-00613-w
- 444 35. Kerkmeijer LGW, Groen VH, Pos FJ, et al. Focal Boost to the Intraprostatic Tumor in
445 External Beam Radiotherapy for Patients With Localized Prostate Cancer: Results
446 From the FLAME Randomized Phase III Trial. *J Clin Oncol*. Published online
447 January 20, 2021. doi:10.1200/JCO.20.02873
- 448 36. Morris WJ, Tyldesley S, Rodda S, et al. Androgen Suppression Combined with
449 Elective Nodal and Dose Escalated Radiation Therapy (the ASCENDE-RT Trial): An
450 Analysis of Survival Endpoints for a Randomized Trial Comparing a Low-Dose-Rate
451 Brachytherapy Boost to a Dose-Escalated External Beam Boost for High- and
452 Intermediate-risk Prostate Cancer. *Int J Radiat Oncol*. 2017;98(2):275-285.
453 doi:10.1016/j.ijrobp.2016.11.026
- 454 37. Hoskin PJ, Rojas AM, Ostler PJ, Bryant L, Lowe GJ. Randomised trial of external-
455 beam radiotherapy alone or with high-dose-rate brachytherapy for prostate cancer:
456 Mature 12-year results. *Radiother Oncol*. 2021;154:214-219.
457 doi:10.1016/j.radonc.2020.09.047
- 458 38. Ma TM, Chu FI, Sandler H, et al. Local Failure Events in Prostate Cancer Treated
459 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