

Evaluation of a Cascaded Deep Learning–based Algorithm for Prostate Lesion Detection at Biparametric MRI

Yue Lin, AB • Enis C. Yilmaz, MD • Mason J. Belue, BS • Stephanie A. Harmon, PhD • Jesse Tetreault, MS • Tim E. Phelps, PhD • Katie M. Merriman, BS • Lindsey Hazen, RN • Charisse Garcia, RN • Dong Yang, PhD • Ziyue Xu, PhD • Nathan S. Lay, PhD • Antoun Toubaji, MD • Maria J. Merino, MD • Daguang Xu, PhD • Yan Mee Law, MD • Sandeep Gurram, MD • Bradford J. Wood, MD • Peter L. Choyke, MD • Peter A. Pinto, MD • Baris Turkbey, MD

From the Molecular Imaging Branch (Y.L., E.C.Y., M.J.B., S.A.H., T.E.P., K.M.M., N.S.L., P.L.C., B.T.), Center for Interventional Oncology (L.H., C.G., B.J.W.), Laboratory of Pathology (A.T., M.J.M.), and Urologic Oncology Branch (S.G., P.A.P.), National Cancer Institute, National Institutes of Health, 10 Center Dr, MSC 1182, Bldg 10, Rm B3B85, Bethesda, MD 20892; NVIDIA, Santa Clara, Calif (J.T., D.Y., Z.X., D.X.); Department of Radiology, Clinical Center, National Institutes of Health, Bethesda, Md (L.H., C.G., B.J.W.); and Department of Radiology, Singapore General Hospital, Singapore (Y.M.L.). Received March 25, 2023; revision requested May 15; final revision received January 24, 2024; accepted March 18. **Address correspondence to** B.T. (email: turbeyi@mail.nih.gov).

National Institutes of Health does not endorse or recommend any commercial products, processes, or services. The views and personal opinions of authors expressed herein do not necessarily state or reflect those of the US government, nor reflect any official recommendation nor opinion of the National Institutes of Health nor National Cancer Institute.

Conflicts of interest are listed at the end of this article.

Radiology 2024; 311(2):e230750 • <https://doi.org/10.1148/radiol.230750> • Content codes: **GU** **MR** **AI**

Background: Multiparametric MRI (mpMRI) improves prostate cancer (PCa) detection compared with systematic biopsy, but its interpretation is prone to interreader variation, which results in performance inconsistency. Artificial intelligence (AI) models can assist in mpMRI interpretation, but large training data sets and extensive model testing are required.

Purpose: To evaluate a biparametric MRI AI algorithm for intraprostatic lesion detection and segmentation and to compare its performance with radiologist readings and biopsy results.

Materials and Methods: This secondary analysis of a prospective registry included consecutive patients with suspected or known PCa who underwent mpMRI, US-guided systematic biopsy, or combined systematic and MRI/US fusion–guided biopsy between April 2019 and September 2022. All lesions were prospectively evaluated using Prostate Imaging Reporting and Data System version 2.1. The lesion- and participant-level performance of a previously developed cascaded deep learning algorithm was compared with histopathologic outcomes and radiologist readings using sensitivity, positive predictive value (PPV), and Dice similarity coefficient (DSC).

Results: A total of 658 male participants (median age, 67 years [IQR, 61–71 years]) with 1029 MRI-visible lesions were included. At histopathologic analysis, 45% (294 of 658) of participants had lesions of International Society of Urological Pathology (ISUP) grade group (GG) 2 or higher. The algorithm identified 96% (282 of 294; 95% CI: 94%, 98%) of all participants with clinically significant PCa, whereas the radiologist identified 98% (287 of 294; 95% CI: 96%, 99%; $P = .23$). The algorithm identified 84% (103 of 122), 96% (152 of 159), 96% (47 of 49), 95% (38 of 40), and 98% (45 of 46) of participants with ISUP GG 1, 2, 3, 4, and 5 lesions, respectively. In the lesion-level analysis using radiologist ground truth, the detection sensitivity was 55% (569 of 1029; 95% CI: 52%, 58%), and the PPV was 57% (535 of 934; 95% CI: 54%, 61%). The mean number of false-positive lesions per participant was 0.61 (range, 0–3). The lesion segmentation DSC was 0.29.

Conclusion: The AI algorithm detected cancer-suspicious lesions on biparametric MRI scans with a performance comparable to that of an experienced radiologist. Moreover, the algorithm reliably predicted clinically significant lesions at histopathologic examination.

ClinicalTrials.gov Identifier: NCT03354416

© RSNA, 2024

Supplemental material is available for this article.

Multiparametric MRI (mpMRI) of the prostate allows visualization of cancers and enables MRI/US fusion–guided targeted biopsy. Prostate mpMRI improves the diagnosis of clinically significant prostate cancer (csPCa) through targeted biopsy (1) and reduces the upgrading rate of radical prostatectomy specimens compared with US-guided systematic biopsy (2). Additionally, mpMRI is useful for preventing unnecessary biopsies by allowing radiologists to rule out csPCa when the mpMRI results are negative (3). In response to the growth in mpMRI use, the Prostate Imaging Reporting and Data System (PI-RADS) guidelines

were introduced to standardize prostate mpMRI acquisition, interpretation, and reporting (4). Even with standardized guidelines, mpMRI evaluations are subject to significant variations in intra- and interreader reproducibility and interpretative performance (5–7). In addition, targeted biopsy requires manual lesion segmentation, which can introduce further variation in mpMRI examination performance (8).

Artificial intelligence (AI) algorithms that are based on deep learning approaches can augment radiologist performance at multiple steps in the prostate cancer (PCa) diagnostic pathway (9,10). AI can provide an

Abbreviations

AI = artificial intelligence, csPCa = clinically significant PCa, DSC = Dice similarity coefficient, GG = grade group, ISUP = International Society of Urological Pathology, mpMRI = multiparametric MRI, PCa = prostate cancer, PI-RADS = Prostate Imaging Reporting and Data System, PPV = positive predictive value

Summary

A deep learning–based algorithm automatically detected cancerous lesions on prostate biparametric MRI scans with reasonable performance and reliably predicted clinically significant prostate cancer at biopsy.

Key Results

- In a secondary analysis of a prospective registry of 658 participants who underwent multiparametric MRI (mpMRI) and biopsy, a deep learning algorithm detected clinically significant prostate cancer with participant-level sensitivity similar to that of a radiologist (96% vs 98%; $P = .23$), based on biopsy results.
- The lesion-level detection sensitivity and positive predictive value were 55% and 57%, respectively, with radiologist readings as ground truth.
- Algorithm-based lesion segmentations partially agreed with radiologist segmentations on mpMRI scans (Dice similarity coefficient = 0.29).

objective assessment of prostate MRI scans and standardize intraprostatic lesion detection. Deep learning algorithms have improved dramatically in recent years due to the availability of larger labeled data sets, accessibility of advanced computing hardware, and improvements in algorithm training techniques and architectures (11,12). Several studies have been published showcasing the diagnostic value of deep learning algorithms in prostate imaging (13–15). Recently, Mehravand et al (16) reported on the development phase of a cascaded AI model that can detect and segment lesions suspicious for cancer on biparametric prostate MRI scans. This AI model showed promising initial performance metrics during the development phase; however, similar to many other prostate MRI AI models reported in the literature, its diagnostic performance in a larger patient sample has not been evaluated. Therefore, the purpose of this study was to evaluate this cascaded deep learning–based AI algorithm for PCa detection and segmentation in a prospective PI-RADS version 2.1–evaluated sample and to compare its performance against that of an expert radiologist and systematic or targeted biopsy results.

Materials and Methods

Study Participants

This secondary analysis of a prospective registry (ClinicalTrials.gov: NCT03354416) was approved by the institutional review board of the National Institutes of Health, and written informed consent was obtained from all

participants. This Health Insurance Portability and Accountability Act–compliant study included consecutive patients with increased risk of PCa (ie, strong family history) or with suspected or known PCa who underwent mpMRI and subsequent biopsy between April 2019 and September 2022 at an academic medical center. Patients with PI-RADS category 1 scans underwent US-guided systematic biopsy, while patients with PI-RADS category 2–5 lesions underwent combined US-guided systematic biopsy and MRI/US fusion–guided targeted biopsy. Patients with a prior history of prostate cancer treatment, such as androgen deprivation therapy, focal therapy, radiation therapy, or radical prostatectomy, were excluded. Patients with incomplete biopsy results were excluded as well. The final study population included treatment-naïve participants with available biopsy results. The results from a subset ($n = 454$) of this sample were previously published in a study that aimed to evaluate PI-RADS version 2.1 prospectively (17).

Image Acquisition

All participants were scanned with a 3-T MRI scanner (Ingenuia Elition X, Class IIa; Philips Healthcare) using a 16-channel surface coil (Sense; Philips Healthcare). An endorectal coil (BPX-30; Medrad) was used in 121 participants. T2-weighted images, high- b -value diffusion-weighted images, apparent diffusion coefficient maps, and dynamic contrast-enhanced sequences were obtained. Full image acquisition parameters of the current study and the original training sample are summarized in Tables S1 and S2.

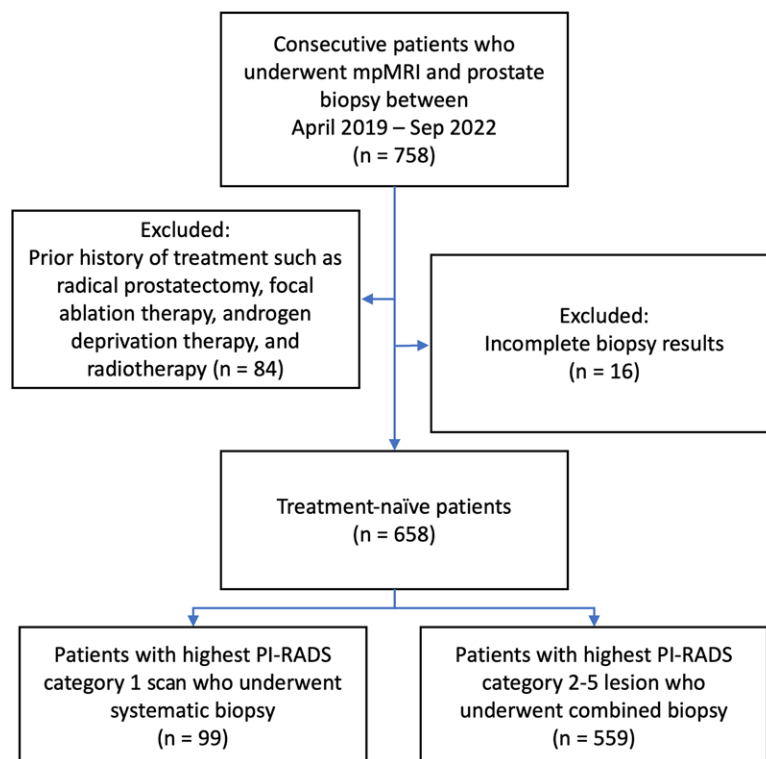


Figure 1: Participant flow diagram. mpMRI = multiparametric MRI, PI-RADS = Prostate Imaging Reporting and Data System.

Image Analysis

All scans were prospectively evaluated during the clinical reading, and intraprostatic lesions were categorized using PI-RADS version 2.1 (4) by one genitourinary radiologist (B.T., with >15 years of experience in prostate imaging, 1000 scans per year). The whole prostate and intraprostatic lesions were volumetrically contoured by the same genitourinary radiologist (B.T.) with a commercial biopsy preparation system (DynaCAD 5.0; Invivo) using all mpMRI sequences and anatomic planes at the time of the prospective reading. Details regarding biopsy and pathologic evaluation are provided in Appendix S1. International Society of Urological Pathology (ISUP) grade group (GG) 2 and higher lesions were considered csPCa (18).

A multireader validation study was conducted to assess interreader variability with the radiologist ground truth. A body radiologist from a different institution (Y.M.L., with >10 years of experience in prostate imaging) evaluated prostate mpMRI scans from a subset of participants randomly selected from the study sample. The reader was blinded to clinical and pathologic details. The reader was asked to detect and assign a category to the index lesion using PI-RADS version 2.1. Both readers identified lesion location using a standardized reporting template adapted from the PI-RADS version 2.1 sector map (4). A prostate mpMRI-focused research fellow (Y.L.) mapped and correlated those lesions based on the reader report, lesion morphologic characteristics, and landmarks (eg, benign prostatic hyperplasia nodules, urethra, prostatic capsule). Interreader agreement regarding the overall PI-RADS index lesion category assignment was examined.

AI Model Evaluation

A previously developed biparametric MRI-based cascaded deep learning AI model was used for PCa detection in this study (16). The algorithm inputs included biparametric MRI sequences (T2-weighted images, high-*b*-value diffusion-weighted images, and apparent diffusion coefficient maps), and the outputs included prostate organ and intraprostatic lesion segmentations. The AI model used in this study is accessible at <https://github.com/Project-MONAI/research-contributions/tree/main/prostate-mri-lesion-seg>. Comparison of both radiologist and AI algorithm findings to the biopsy results, as the reference ground truth, was performed at the lesion and participant levels. For lesion-level analysis, the lesion detection performance of the algorithm was evaluated using histopathologic examination combining both targeted biopsy and systematic biopsy results as the reference ground truth (Fig S1). AI predictions were first correlated with targeted biopsy results when matching to a radiologist-defined lesion. Twelve-core transrectal and transperineal systematic biopsy results were then reconstructed and spatially matched to the remaining AI predictions that did not correspond to any targeted biopsy results (ie, false positives based on radiologist ground truth) using in-house software. Details regarding systematic biopsy reconstruction are provided in Appendix S1. A prediction

Table 1: Participant and Lesion Characteristics (n = 658 Participants)

Characteristic	Value
Age (y)*	67 (61–71)
Age range (y)	35–82
PSA level (ng/mL)*	6.7 (4.7–9.8)
Prostate volume (cm ³)*	60 (43–86)
PSA density (ng/mL ²)*	0.10 (0.06–0.16)
mpMRI-biopsy interval (d)*	55 (19–93)
Family history of PCa	
Negative	339 (52)
Positive	265 (40)
Unknown	54 (8)
Prior biopsy	
Biopsy naive	203 (31)
Prior negative biopsy	126 (19)
Prior positive biopsy	329 (50)
Combined biopsy results	
Negative	242 (37)
Positive for PCa	416 (63)
Positive for csPCa	294 (45)
Highest PI-RADS category	
1	99 (15)
2	69 (10)
3	110 (17)
4	220 (33)
5	160 (24)
Highest ISUP grade group	
Benign	242 (37)
1	122 (19)
2	159 (24)
3	49 (7)
4	40 (6)
5	46 (7)
Number of lesions at mpMRI	
0	99 (15)
1	245 (37)
2	196 (30)
3	85 (13)
4	29 (4)
5	4 (1)

Note.—Unless otherwise specified, data are numbers of participants, with percentages in parentheses. csPCa = clinically significant PCa, ISUP = International Society of Urological Pathology, mpMRI = multiparametric MRI, PCa = prostate cancer, PI-RADS = Prostate Imaging Reporting and Data System, PSA = prostate-specific antigen.

* Data are medians, with IQRs in parentheses.

that corresponded to a cancerous lesion at targeted biopsy or systematic biopsy was considered to be true positive. Cancer detection rates for the radiologist were calculated based on targeted biopsy results, whereas both the targeted biopsy and spatially matched systematic biopsy results were used to calculate cancer detection rates for the AI algorithm. At the participant level, the sensitivity and positive predictive value (PPV) of radiologist and AI findings were calculated

Table 2: Lesion-Level Comparisons of Radiologist and AI Cancer Detection Rates Based on Histopathologic Ground Truth

Biopsy Result	Cancer Detection Rate (%)		P Value
	Radiologist	AI	
Overall			
PCa	49 (45, 53) [507/1029]	44 (40, 47) [424/968]	<.001
csPCa	32 (28, 35) [327/1029]	28 (25, 31) [273/968]	<.001
ISUP grade group			
Benign	51 (48, 54) [522/1029]	56 (53, 59) [544/968]	
1	18 (15, 20) [180/1029]	16 (13, 18) [151/968]	
2	19 (17, 21) [195/1029]	17 (15, 20) [168/968]	
3	5 (4, 7) [54/1029]	4 (3, 6) [42/968]	
4	4 (3, 5) [42/1029]	3 (2, 4) [32/968]	
5	4 (2, 5) [36/1029]	3 (2, 4) [31/968]	

Note.—Data in parentheses are 95% CIs, and data in brackets are numbers of lesions. AI = artificial intelligence, csPCa = clinically significant PCa, PCa = prostate cancer, ISUP = International Society of Urological Pathology.

using the highest ISUP GGs from both the systematic biopsy and targeted biopsy results as the ground truth.

AI detection performance was also examined based on radiologist ground truth, and data were analyzed at both the participant and lesion levels. True positives, false positives, and false negatives were identified to evaluate the sensitivity, PPV, and false discovery rate. For lesion-level analysis, a prediction was considered a true positive when it had any overlapping area with the radiologist ground truth. Given the use of a radiologist ground truth, true negatives could not be evaluated at the lesion level. At the participant level, a true positive was defined as a positive AI prediction in a PI-RADS category 2 or higher scan. A true negative was defined as a PI-RADS category 1 scan correctly predicted to contain no lesions by the AI algorithm. The segmentation accuracy was measured with the Dice similarity coefficient (DSC).

Statistical Analysis

To assess interreader agreement, three metrics were used: the unweighted κ statistic, the quadratically weighted κ statistic, and the proportion of agreement, defined as the proportion of lesions detected by both readers and given the same assignment. κ values were categorized as indicating slight (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), or excellent (0.81–1.00) agreement. To

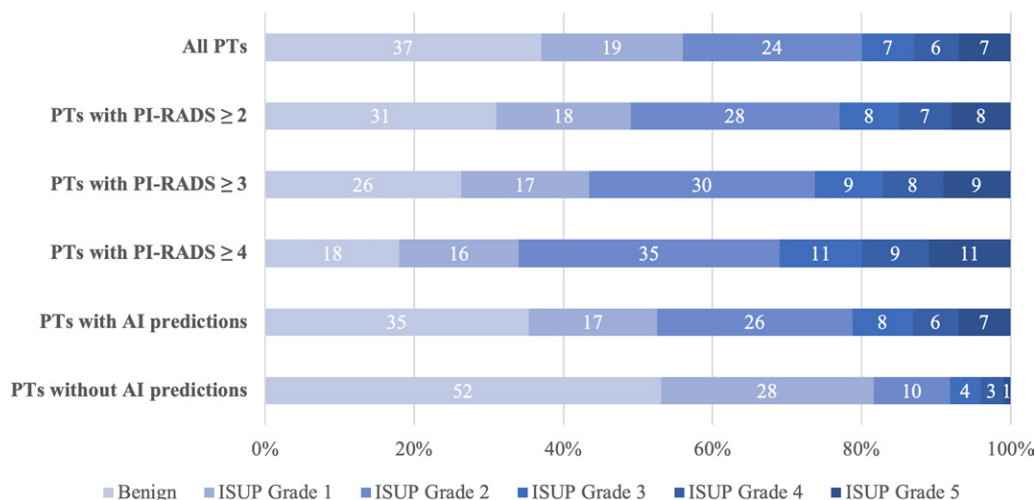


Figure 2: Distribution of combined biopsy outcomes (percentage) based on highest International Society of Urological Pathology (ISUP) grade group per participant (PT). AI = artificial intelligence, PI-RADS = Prostate Imaging Reporting and Data System.

account for intra- and interlesion correlations, 95% CIs were estimated from 2000 bootstrap samples by random sampling at the participant level (19). The Wald test was used to compare the detection performance of the algorithm with that of the radiologist at both the lesion level (cancer detection rate) and participant level (sensitivity and PPV) (20). The Wald test used the bootstrap standard error to estimate the difference in cancer detection rate, sensitivity, and PPV between the radiologist and the AI model. Descriptive statistics are reported for radiologist interpretations considered positive calls at different PI-RADS thresholds (participant-level highest PI-RADS category ≥ 2 , ≥ 3 , or ≥ 4) and lesion-level cancer detection rates within ISUP GGs. All P values are two-sided, and $P < .05$ was considered statistically significant. Statistical analyses were performed using R software (version 4.1.2; R Foundation for Statistical Computing).

Table 3: Participant-Level Sensitivity and PPV of AI versus Radiologist Detection of Prostate Cancer Based on Combined Biopsy Results

Reader	Sensitivity (%)	<i>P</i> Value*	PPV (%)	<i>P</i> Value*
Radiologist				
PI-RADS ≥ 2	93 (90, 95) [387/416]		69 (65, 73) [387/559]	
PI-RADS ≥ 3	87 (84, 90) [361/416]		74 (70, 77) [361/490]	
PI-RADS ≥ 4	75 (70, 78) [310/416]		82 (78, 85) [310/380]	
AI	92 (90, 95) [385/416]	.75	65 (61, 69) [385/594]	<.001

Note.—Data in parentheses are 95% CIs, and data in brackets are numbers of participants. AI = artificial intelligence, PI-RADS = Prostate Imaging Reporting and Data System, PPV = positive predictive value.

* *P* values reported for AI-detected lesions versus radiologist assessment of PI-RADS category 2 or higher, to reflect training conditions of AI algorithm.

Table 4: Participant-Level Sensitivity and PPV of AI versus Radiologist Detection of Clinically Significant Prostate Cancer Based on Combined Biopsy Results

Reader	Sensitivity (%)	<i>P</i> Value*	PPV (%)	<i>P</i> Value*
Radiologist				
PI-RADS ≥ 2	98 (96, 99) [287/294]		51 (47, 55) [287/559]	
PI-RADS ≥ 3	95 (92, 97) [278/294]		57 (52, 61) [278/490]	
PI-RADS ≥ 4	85 (81, 89) [251/294]		66 (61, 71) [251/380]	
AI	96 (94, 98) [282/294]	.23	47 (44, 52) [282/594]	<.001

Note.—Data in parentheses are 95% CIs, and data in brackets are numbers of participants. AI = artificial intelligence, PI-RADS = Prostate Imaging Reporting and Data System, PPV = positive predictive value.

* *P* values reported for AI-detected lesions versus radiologist assessment of PI-RADS category 2 or higher, to reflect training conditions of AI algorithm.

Results

Participant Characteristics

Among 758 patients initially included in the study, 84 and 16 patients were excluded because they had a history of prior treatment or incomplete biopsy results, respectively (Fig 1). Therefore, the final study sample consisted of 658 male participants (median age, 67 years [IQR, 61–71 years]). The median prostate-specific antigen level was 6.7 ng/mL (IQR, 4.7–9.8 ng/mL), and the median prostate-specific antigen density was 0.10 ng/mL² (IQR, 0.06–0.16 ng/mL²). Of 1029 MRI-visible lesions, three (0.3%) were PI-RADS category 1, 219 (21%) were category 2, 247 (24%) were category 3, 381 (37%) were category 4, and 179 (17%) were category 5. Overall, 49% (507 of 1029) of the lesions were positive for PCa, and 32% (327 of 1029) of the lesions were positive for csPCa at targeted biopsy. A total of 416 of 658 (63%) participants had PCa and 294 of 658 (45%) participants had csPCa at combined biopsy. Detailed participant and lesion characteristics are presented in Table 1. Participant and lesion characteristics of the original development sample (16) are summarized in Table S3.

Multireader Analysis

The subset for the multireader study included mpMRI scans in 51 participants. The median age of participants in the subset was 69 years (IQR, 64–75 years), and median

prostate-specific antigen level was 7.2 ng/mL (IQR, 5.5–9.0 ng/mL). Both the unweighted κ (0.41) and the quadratically weighted κ (0.48) indicated moderate agreement for category assignment of index lesions. Regarding the specific PI-RADS cutoff, the unweighted κ was 0.45 and the quadratically weighted κ was 0.55 for a cutoff of PI-RADS category 3 or higher. Readers had fair agreement for PI-RADS category 3 assignment (proportion of agreement, 0.33) and moderate agreement for PI-RADS category 4 (0.42) and category 5 (0.59) assignment.

AI Detection Performance Based on Histopathologic Results

Lesion-level histopathologic analysis revealed that of 1029 intraprostatic lesions defined by the radiologist, 49% (507 of 1029) were PCa and 32% (327 of 1029) were csPCa at targeted biopsy (Table 2). In comparison, the AI model predicted 968 lesions, 44% (424 of 968) and 28% (273 of 968) of which were PCa and csPCa, respectively ($P < .001$ for both comparisons of detection rate between radiologist and AI). For ISUP GG 2, 3, 4, and 5 lesions, respectively, sensitivity for the radiologist and the AI model was 70% (195 of 277) versus 61% (168 of 277), 59% (54 of 92) versus 46% (42 of 92), 55% (42 of 77) versus 42% (32 of 77), and 54% (36 of 67) versus 46% (31 of 67). Additionally, the AI model predicted 106 lesions in 99 participants with PI-RADS category 1 MRI scans. Of these predicted

Table 5: Lesion- and Participant-Level Detection Performance Metrics Based on Radiologist Ground Truth

Performance Statistic	Value
Lesion level*	
True positive	569
False negative	466
False positive	399
True negative	NA
Sensitivity (%) [†]	55 (52, 58) [569/1029]
PPV (%) [‡]	57 (54, 61) [535/934]
FDR (%) [‡]	43 (39, 46) [399/934]
Specificity (%)	NA
DSC	0.29
Participant level	
True positive	519
False negative	41
False positive	75
True negative	23
Sensitivity (%) [†]	93 (90, 95) [519/559]
PPV (%) [‡]	87 (85, 90) [519/594]
FDR (%) [‡]	13 (10, 15) [75/594]
Specificity (%) [‡]	23 (15, 32) [23/98]
DSC	0.34

Note.—Unless otherwise specified, data are numbers of lesions for lesion-level metrics and numbers of participants for participant-level metrics. DSC = Dice similarity coefficient, FDR = false discovery rate, NA = not applicable, PPV = positive predictive value.

* The AI algorithm identified 934 unique lesions, and 535 of these lesions corresponded to 569 radiologist ground truth lesions.

[†] Data in parentheses are 95% CIs, and data in brackets are numbers of lesions.

[‡] Data in parentheses are 95% CIs, and data in brackets are numbers of participants.

lesions, 8% (nine of 106) were PCa and 5% (five of 106) were csPCa.

Of the 658 participants, 45% (294 of 658) had csPCa at combined biopsy (Fig 2). The AI model detected 84% (103 of 122), 96% (152 of 159), 96% (47 of 49), 95% (38 of 40), and 98% (45 of 46) of all participants with ISUP GG 1, 2, 3, 4, and 5 disease, respectively. In participants without any AI lesion predictions, 52% (35 of 67) had benign biopsy results, and 28% (19 of 67) had ISUP GG 1 disease. The AI model identified 92% (385 of 416) of all participants with PCa, whereas the radiologist identified 93% (387 of 416) of all participants at combined biopsy ($P = .75$) (Table 3). The AI model identified 96% (282 of 294) of all participants with csPCa, whereas the radiologist identified 98% (287 of 294) of participants ($P = .23$) (Table 4). The PPV for detecting PCa and csPCa based on radiologist assessment of PI-RADS category 2 or higher lesions was 69% (387 of 559) and 51% (287 of 559), respectively. In comparison, the AI model had a PPV of 65% (385 of 594) for detecting PCa and 47% (282 of 594) for detecting csPCa (both $P < .001$). A total of 99 participants had highest PI-RADS category 1 MRI scans.

Negative predictive value and specificity for the AI model and radiologist are summarized in Table S4. Systematic biopsy for these participants (targeted biopsy was not performed) revealed that 29% (29 of 99) of the participants had PCa and 7% (seven of 99) had csPCa. The AI model identified lesions in 76 of these participants, of whom 25% (19 of 76) had PCa and 8% (six of 76) had csPCa. Overall, the AI model correctly identified six of the seven participants (86%) who were assigned PI-RADS category 1 and had csPCa.

AI Detection Performance Based on Radiologist Interpretation as the Ground Truth

The AI algorithm identified 934 lesions. Of these, 535 predictions corresponded to 569 radiologist-defined lesions (ie, 569 true positives) (Table 5). A total of 399 AI lesions were not radiologist defined (false positives), and this occurred at a mean rate of 0.61 lesions (range, 0–3) per participant, while 466 radiologist-defined lesions were not detected by AI (false negatives). In the lesion-level analysis where the radiologist reading was used as the ground truth, the algorithm had a sensitivity of 55% (569 of 1029; 95% CI: 52%, 58%), a PPV of 57% (535 of 934; 95% CI: 54%, 61%), and a false discovery rate of 43% (399 of 934; 95% CI: 39%, 46%). The AI algorithm detected 30% (65 of 219), 42% (104 of 247), 61% (232 of 381), and 92% (165 of 179) of PI-RADS category 2, 3, 4, and 5 lesions, respectively (Fig S2).

In the participant-level analysis, the algorithm demonstrated a sensitivity of 93% (519 of 559; 95% CI: 90%, 95%), a PPV of 87% (519 of 594; 95% CI: 85%, 90%), a false discovery rate of 13% (75 of 594; 95% CI: 10%, 15%), and a specificity of 23% (23 of 98; 95% CI: 15%, 32%) (Table 5). The algorithm detected 36% (25 of 69), 49% (54 of 110), 74% (162 of 220), and 93% (149 of 159) of participants with highest PI-RADS category 2, 3, 4, and 5 lesions, respectively (Fig S3). Examples of algorithm lesion predictions and probability maps in participants are shown in Figures 3–6.

DSC Analysis

The DSC of the AI model for lesion segmentation was 0.29 for all lesions. For PI-RADS category 2, 3, 4, and 5 lesions, the DSC was 0.13, 0.21, 0.29, and 0.58, respectively. For benign and ISUP GG 1, 2, 3, 4, and 5 lesions, the DSC was 0.14, 0.32, 0.39, 0.40, 0.49, and 0.50, respectively. The DSC for lesion segmentation was 0.34 in the participant-level analysis. For PI-RADS category 2, 3, 4, and 5 index lesions, the DSC was 0.09, 0.21, 0.31, and 0.58, respectively. When the highest GG in each participant was considered, the DSC was 0.17, 0.24, 0.39, 0.39, 0.45, and 0.56 for benign and ISUP GG 1, 2, 3, 4, and 5 lesions, respectively. Detailed DSC results are summarized in Table S5.

Discussion

Multiparametric MRI improves the detection of prostate cancer (PCa) through targeted biopsy, but its interpretation is prone to substantial interreader variation, which results in performance inconsistency in the PCa diagnosis pathway.

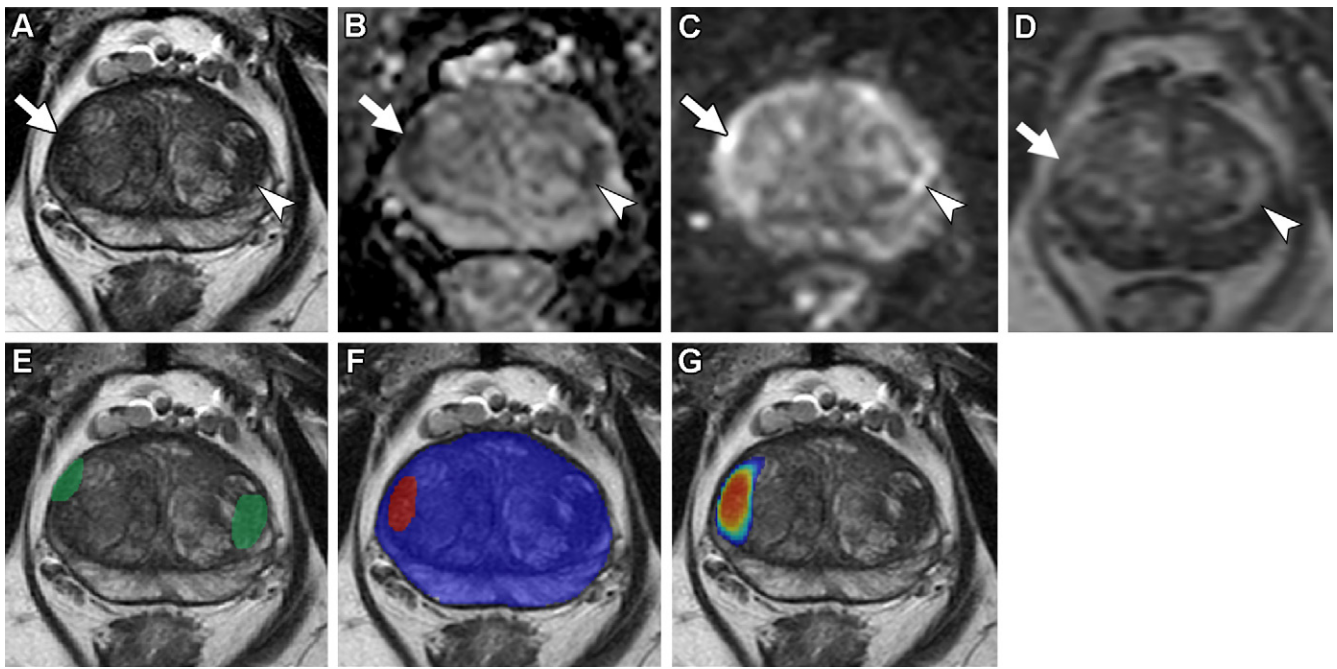


Figure 3: Axial multiparametric MRI scans in a 72-year-old male participant with a serum prostate-specific antigen level of 9.1 ng/mL: **(A)** T2-weighted image, **(B)** apparent diffusion coefficient map, **(C)** high-*b*-value diffusion-weighted image ($b = 1500 \text{ sec/mm}^2$), **(D)** dynamic contrast-enhanced image (frame 17 of 54 acquired at 5.6-second intervals), **(E)** T2-weighted image with radiologist-segmented lesions (green contours) overlaid, **(F)** T2-weighted image with artificial intelligence (AI) prediction map overlaid (red contour is positive prediction; blue contour is AI prostate organ segmentation), and **(G)** T2-weighted image with AI probability map overlaid (red indicates higher probability). Two distinct lesions were detected by the radiologist and represented the ground truth. Lesion 1 (1.6 cm; arrow in **A–D**) was in the right midgland transition zone and was designated Prostate Imaging Reporting and Data System (PI-RADS) category 4. Lesion 2 (1.5 cm; arrowhead in **A–D**) was in the left midgland transition zone and was designated PI-RADS category 3. Lesion 1 was correctly detected (true positive), while lesion 2 was missed by the AI algorithm (false negative). Based on targeted biopsy samples, lesion 1 was positive for Gleason score 7 (3 + 4) prostate adenocarcinoma, and lesion 2 was benign.

AI has the potential to assist radiologists by standardizing intraprostatic lesion detection and reducing variability. Our study employed a fully automated deep learning–based algorithm for prostate lesion detection and segmentation in a large sample that was clinically managed prospectively using Prostate Imaging Reporting and Data System version 2.1. The radiologist identified 1029 MRI-visible lesions, and 32% (327 of 1029) had clinically significant PCa (csPCa) at targeted biopsy, whereas the algorithm predicted 968 cancerous foci, and 28% (273 of 968) of these were csPCa. The algorithm correctly detected 96% (282 of 294) of all participants with csPCa, whereas the radiologist identified 98% (287 of 294) ($P = .23$). The lesion- and participant-level cancer detection rates achieved by the AI model are within the range of the performance metrics of MRI-guided biopsy reported in prospective studies (1,21–25) and a meta-analysis (26). Last, in the lesion-level analysis, the algorithm demonstrated adequate detection performance compared with the prospective reading of the radiologist, with a sensitivity of 55% (569 of 1029; 95% CI: 52%, 58%) and a positive predictive value of 57% (535 of 934; 95% CI: 54%, 61%).

Several other studies have explored the use of deep learning–based algorithms for prostate lesion detection and prostate MRI segmentation (27–30). In a retrospective study, Hamm et al (31) showcased an algorithm with lesion-level accuracy of 86% and sensitivity of 90% in an external testing set comprising 330 lesions. In another study, Winkel

et al (32) prospectively tested an AI model on 48 biopsy-naïve men and showed a 60% lesion-level sensitivity and a mean false-positive rate per participant of 0.88 based on radiologist ground truth. These studies relied solely on radiologist-defined lesions; thus, false-negative or MRI-invisible lesions could not be confidently ruled out. A potential advantage of deep learning algorithms is that they can detect lesions in regions that were considered lesion-negative at MRI by human readers. Using both targeted biopsy and spatially matched systematic biopsy results, we were able to examine the histopathologic outcomes of AI-predicted lesions that were not biopsied at targeted biopsy.

Regarding prostate lesion segmentation, previous research has demonstrated moderate agreement among radiologists, with a DSC ranging from 0.48 to 0.52 (33). The original development article (16) for the AI algorithm evaluated here reported a DSC of 0.36 for lesion-level segmentation in the testing sample, and our study achieved a DSC of 0.29 employing this algorithm in a new prospective sample. The DSC for the AI model in our study is consistent with that of other published AI models (30,34). For benign lesions the DSC was 0.14, whereas for csPCa the DSC reached 0.4 or higher. Moreover, the algorithm exhibited a tendency toward prioritizing the central region of the lesion, while radiologists incorporated additional cues to include peripheral aspects. The algorithm employed a benign prostatic hyperplasia filter to achieve a balance between segmenting a

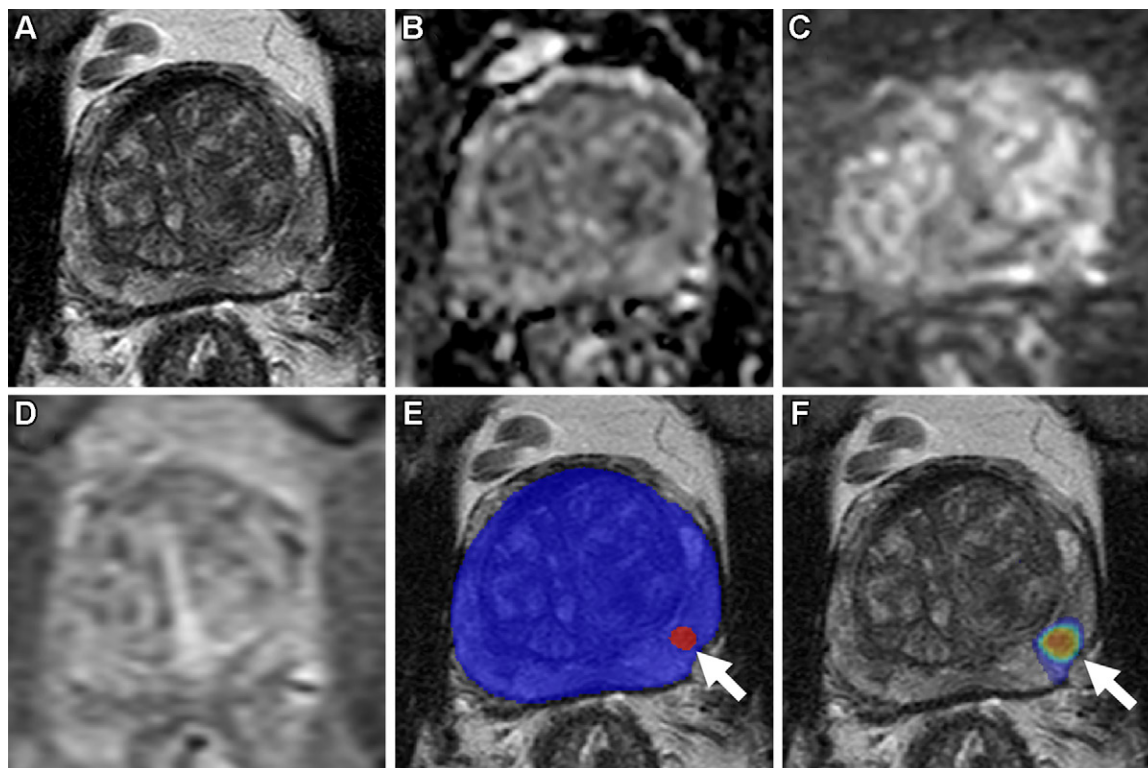


Figure 4: Axial multiparametric MRI scans in a 64-year-old male participant with a serum prostate-specific antigen level of 8.1 ng/mL: **(A)** T2-weighted image, **(B)** apparent diffusion coefficient map, **(C)** high-b-value diffusion-weighted image ($b = 1500 \text{ sec/mm}^2$), **(D)** dynamic contrast-enhanced image (frame 45 of 54 acquired at 5.6-second intervals), **(E)** T2-weighted image with artificial intelligence (AI) prediction map overlaid (red contour is positive prediction; blue contour is AI prostate organ segmentation), and **(F)** T2-weighted image with AI probability map overlaid (red indicates higher probability). No distinct lesion was detected by the radiologist (Prostate Imaging Reporting and Data System category 1). One lesion was called by the AI algorithm in the left midgland peripheral zone (arrow in **E** and **F**), representing a false positive based on the radiologist ground truth. Systematic biopsy obtained from this site (left midgland lateral) was positive for Gleason score 7 (3 + 4) prostate adenocarcinoma.

substantial portion of the lesion area and minimizing false-positive findings beyond tumor foci (16). Additionally, our results showed that the predictive capability of the AI algorithm was comparable to clinical PI-RADS interpretation. This suggests that identifying the position and existence of a lesion, even if it does not align entirely with the radiologist-defined extent, may be helpful for supporting diagnostic decision-making in cSPCa detection and patient selection for biopsy. Of note, the performance of this AI model was higher in higher PI-RADS category lesions; this model may be useful for PI-RADS category 5 lesion biopsy planning, as it detected 92% (165 of 179) of the lesions and achieved a DSC of greater than 0.50.

In the near future, this AI algorithm could be clinically deployed as a companion system implemented in the clinical picture archiving and communication system. This AI algorithm can act as a second reader opinion by providing binary prediction masks and probability maps for radiologists during their readings. It is important to note that for participant-level analysis, the algorithm had positive prediction in these participants, and this does not necessarily indicate the lesion itself was localized. Thus, the algorithm could function as a patient triage tool for biopsy as the AI model was able to identify the majority of participants with cSPCa.

Additionally, not only does the algorithm offer decision support by highlighting lesions and additional areas of concern for radiologists, but its exportable lesion segmentation masks also present an opportunity for automating the labor-intensive processes of lesion segmentation (especially for PI-RADS 5 lesions), streamlining tasks related to biopsy and radiation therapy planning.

Our study has several limitations. All scans were evaluated and annotated by one expert genitourinary radiologist. This might limit the reproducibility of this study; however, this also serves as an ideal situation for level 2 analysis per Fryback and Thornbury (35), addressing the diagnostic performance of a new diagnostic imaging method. Achieving results similar to those of a highly experienced expert radiologist is a desirable outcome for an AI algorithm. Another limitation is that this AI model does not incorporate dynamic contrast-enhanced imaging due to inconsistent image acquisition in the initial study population. Model performance might have been impacted by this lack of information related to dynamic contrast-enhanced imaging. However, biparametric MRI is becoming more popular as the current trend moves toward prostate MRI without the use of contrast agents (36,37).

Variations in acquisition parameters, specifically diffusion imaging protocols, could impact the algorithm's

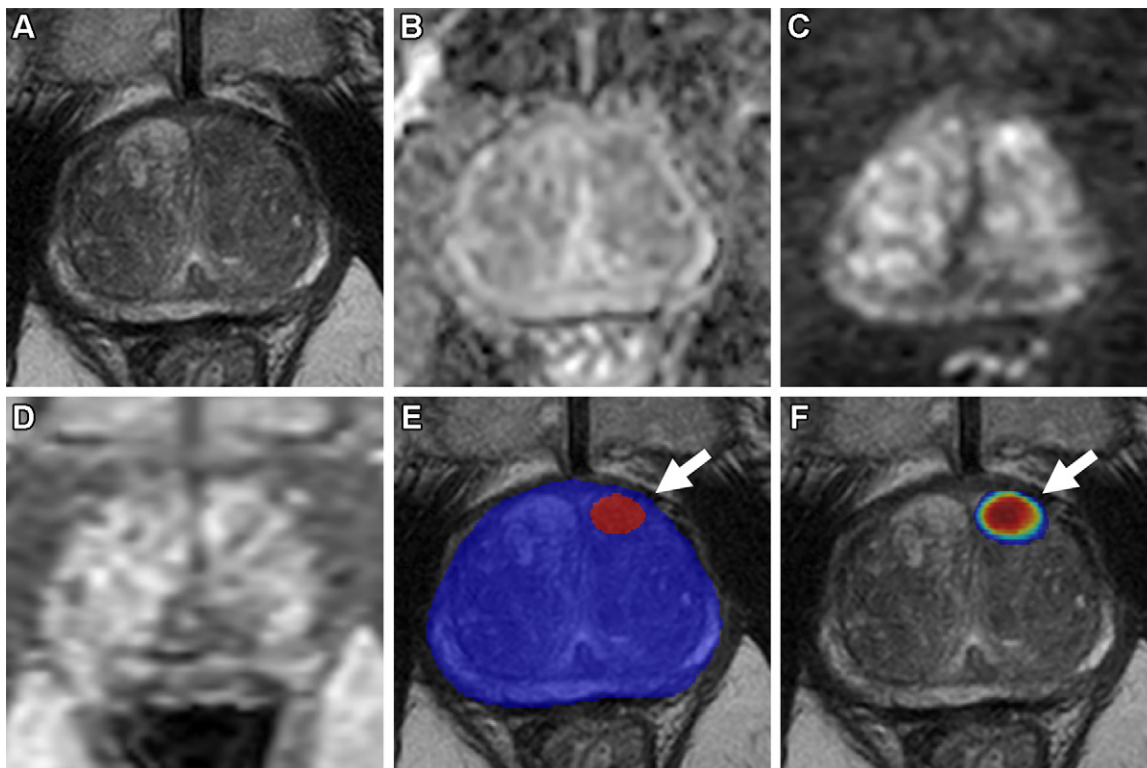


Figure 5: Axial multiparametric MRI scans in a 69-year-old male participant with a serum prostate-specific antigen level of 7.3 ng/mL: **(A)** T2-weighted image, **(B)** apparent diffusion coefficient map, **(C)** high-b-value diffusion-weighted image ($b = 1500 \text{ sec/mm}^2$), **(D)** dynamic contrast-enhanced image (frame 25 of 54 acquired at 5.6-second intervals), **(E)** T2-weighted image with artificial intelligence (AI) prediction map overlaid (red contour is positive prediction; blue contour is AI prostate organ segmentation), and **(F)** T2-weighted image with AI probability map overlaid (red indicates higher probability). One lesion was called by the AI algorithm in the left midgland anterior transition zone (arrow in **E** and **F**), representing a false positive based on the radiologist ground truth. A systematic biopsy sample obtained from this site (left midgland medial) was benign.

performance (38). Further investigation into the performance of the algorithm across diverse imaging conditions is warranted to ensure its generalizability. For AI inference, diffusion-weighted imaging series were spatially aligned and resampled to the T2-weighted image space using Digital Imaging and Communications in Medicine spatial and affine matrix information. Nonetheless, distortion could still lead to a misalignment in per-pixel information between sequences. This represents a limitation of the present algorithm, and major distortions would be expected to inherently decrease performance.

While the algorithm demonstrated a commendable detection rate in participants with ISUP GG 1 tumors, concerns arise regarding the potential for overdiagnosis and subsequent overtreatment. The study sample exhibited a high prevalence of cancer, and many participants had prior positive biopsy. Additionally, certain lesions might be undetected by both the AI and the radiologist, potentially affecting the sensitivity of their predictions. However, this would not substantially influence the comparative analysis, as such undetected lesions would uniformly affect the performance of both. Finally, an experimental reading that included interaction of the radiologist and AI model was not performed in this study. In the future, this algorithm needs to be further validated in a large prospective multicenter study to evaluate its generalizability in an even more heterogeneous

sample and to compare its performance to that of readers with different experience levels.

In conclusion, this cascaded deep learning–based artificial intelligence (AI) model automatically and objectively detected cancerous lesions with reasonable detection performance on prostate biparametric MRI scans in a sample prospectively evaluated using Prostate Imaging Reporting and Data System version 2.1. Moreover, the lesions detected by the algorithm correlated with clinically significant lesions at histopathologic examination. These results demonstrate that this AI model may be useful as an adjunct tool to assist radiologists during prostate MRI readings.

Deputy Editor: Vicky Goh

Scientific Editor: Sarah Atzen

Acknowledgements: This project has been funded in whole or in part with federal funds from the National Cancer Institute of the National Institutes of Health. Research support was provided by the National Institutes of Health Medical Research Scholars Program, a public-private partnership supported jointly by the National Institutes of Health and contributions to the Foundation for the National Institutes of Health from the American Association for Dental, Oral, and Craniofacial Research and Colgate-Palmolive.

Author contributions: Guarantors of integrity of entire study, Y.L., M.J.B., L.H., P.A.P., B.T.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agree to ensure any questions related to the work

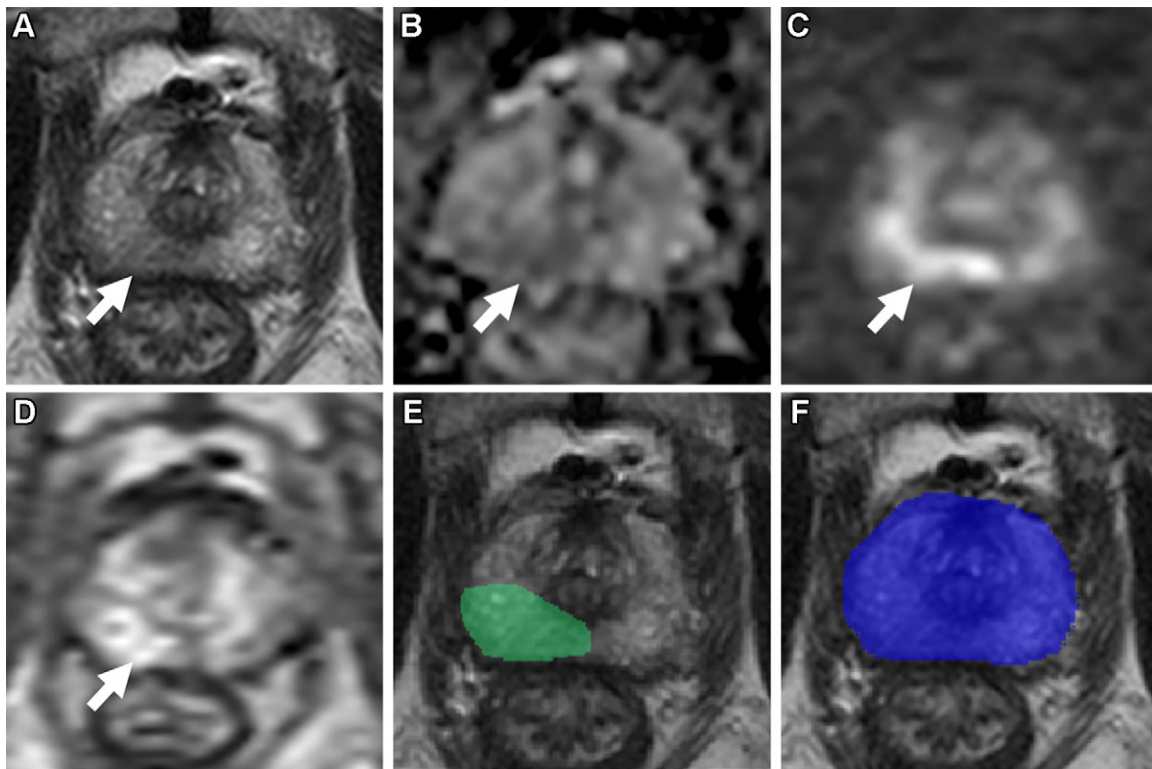


Figure 6: Axial multiparametric MRI scans in a 74-year-old male participant with a serum prostate-specific antigen level of 12.9 ng/mL: (A) T2-weighted image, (B) apparent diffusion coefficient map, (C) high-b-value diffusion-weighted image ($b = 1500 \text{ sec/mm}^2$), (D) dynamic contrast-enhanced image (frame 16 of 54 acquired at 5.6-second intervals), (E) T2-weighted image with radiologist-segmented lesion (green contour) overlaid, and (F) T2-weighted image with artificial intelligence (AI) prediction map overlaid (no positive prediction; blue contour is AI prostate organ segmentation). One lesion was detected by the radiologist and represented the ground truth. The lesion (1.9 cm; arrow in A–D) was in the right apical midgland peripheral zone and was designated Prostate Imaging Reporting and Data System category 4. This lesion was missed by the AI algorithm, representing a false negative. A targeted biopsy sample obtained from the lesion was positive for Gleason score 7 (3 + 4) prostate adenocarcinoma.

are appropriately resolved, all authors; literature research, Y.L., E.C.Y., M.J.B., D.Y., Z.X., B.J.W., B.T.; clinical studies, Y.L., E.C.Y., M.J.B., L.H., C.G., D.Y., A.T., M.J.M., Y.M.L., S.G., B.J.W., P.L.C., P.A.P., B.T.; experimental studies, Y.L., E.C.Y., M.J.B., S.A.H., J.T., D.Y., Z.X., Y.M.L., B.J.W., B.T.; statistical analysis, Y.L., M.J.B., S.A.H., J.T., D.Y., Z.X., Y.M.L., P.A.P., B.T.; and manuscript editing, Y.L., E.C.Y., M.J.B., S.A.H., T.E.P., K.M.M., L.H., Z.X., N.S.L., M.J.M., D.X., Y.M.L., S.G., B.J.W., P.L.C., P.A.P., B.T.

Data sharing: Data generated or analyzed during the study are available from the corresponding author by request.

Disclosures of conflicts of interest: Y.L. No relevant relationships. E.C.Y. Associate editor on the *Radiology* In Training Editorial Board. M.J.B. No relevant relationships. S.A.H. Institution (National Cancer Institute) has a cooperative research and development agreement with NVIDIA. J.T. This work was completed under a cooperative research and development agreement between the National Institutes of Health (NIH) and NVIDIA. T.E.P. No relevant relationships. K.M.M. No relevant relationships. L.H. No relevant relationships. C.G. No relevant relationships. D.Y. No relevant relationships. Z.X. No relevant relationships. N.S.L. This research supported by NIH Intramural Research Program; receives shared royalties from NIH for an unrelated prostate cancer detection system patent (US patent no. US11200667B2). A.T. No relevant relationships. M.J.M. No relevant relationships. D.X. No relevant relationships. Y.M.L. No relevant relationships. S.G. No relevant relationships. B.J.W. Institution (NIH) and NVIDIA have a cooperative research and development agreement; NIH and Philips have a cooperative research and development agreement that includes work in the space of prostate cancer and also artificial intelligence; principal investigator on cooperative research and development agreements between NIH and Philips, Celsion, Biocompatibles, Boston Scientific, Siemens, NVIDIA, XACT Robotics, and Promaxo; support for research from 3T Technologies (devices), Exact Imaging (data), AngioDynamics (equipment), AstraZeneca (pharmaceuticals, National Cancer Institute cooperative research and development agreement), ArciTrax (devices and equipment), Johnson

& Johnson (equipment), Medtronic (equipment), Theromix (supplies), Profound Medical (equipment and supplies), QT Imaging (equipment and supplies), Boston Scientific (equipment), Varian (materials), Combat Medical (equipment), Clinical LaserThermia Systems (equipment), MediView (equipment), CIVCO Medical Solutions (equipment), Galvanize (equipment); royalties from a patent licensing agreement between Philips and NIH; licensing agreements between NIH and NVIDIA and Canon; NIH and CIVCO are coinventors of intellectual property; NIH and ArciTrax are coinventors of intellectual property; NIH and Philips are coinventors of intellectual property; NIH and Boston Scientific are coinventors of intellectual property; NIH and NVIDIA are coinventors of intellectual property; NIH may own intellectual property in the field; another nonlicensed family of patents available upon request (includes 50 patents); editor of Springer book, *Interventional Urology, 2nd Edition*, which may contain indirect mention of the topic of this work; NIH liaison to Academy for Radiology & Biomedical Imaging Research; and NIH liaison to Society of Interventional Oncology. P.L.C. Patents planned, issued, or pending with Philips Medical Systems and NVIDIA (author does not receive royalty payments). P.A.P. Royalties from a licensing agreement between Philips and institution (NIH); cooperative research and development agreement between NIH and Philips; NIH has intellectual property in the field including, among other patents and patent applications, System and Method for Prostate Cancer Detection and Distribution Mapping (US patent no. 8,447,384) and System and Method for Computer Aided Cancer Detection Using T2-weighted and High-Value Diffusion-weighted Magnetic Resonance Imaging (US patent no. 10,215,830); and licensing agreement with NIH and Philips (Invivo). B.T. Cooperative research and development agreement with NVIDIA and Philips and receives royalties from NIH.

References

1. 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.
2. 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.

3. 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. *Lancet* 2017;389(10071):815–822.
4. Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate Imaging Reporting and Data System version 2.1: 2019 update of Prostate Imaging Reporting and Data System version 2. *Eur Urol* 2019;76(3):340–351.
5. Greer MD, Shih JH, Lay N, et al. Interreader variability of Prostate Imaging Reporting and Data System version 2 in detecting and assessing prostate cancer lesions at prostate MRI. *AJR Am J Roentgenol* 2019;212(6):1197–1205.
6. 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.
7. 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.
8. Turkbey B, Oto A. Factors impacting performance and reproducibility of PI-RADS. *Can Assoc Radiol J* 2021;72(3):337–338.
9. Yilmaz EC, Belue MJ, Turkbey B, Reinhold C, Choyke PL. A brief review of artificial intelligence in genitourinary oncological imaging. *Can Assoc Radiol J* 2023;74(3):534–547.
10. Barrett T, de Rooij M, Giganti F, Allen C, Barentsz JO, Padhani AR. Quality checkpoints in the MRI-directed prostate cancer diagnostic pathway. *Nat Rev Urol* 2023;20(1):9–22.
11. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med* 2021;4(1):65.
12. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput Sci* 2021;2(6):420.
13. Schelb P, Kohl S, Radtke JP, et al. Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment. *Radiology* 2019;293(3):607–617.
14. Mehravand S, Harmon SA, Shih JH, et al. Multicenter multireader evaluation of an artificial intelligence-based attention mapping system for the detection of prostate cancer with multiparametric MRI. *AJR Am J Roentgenol* 2020;215(4):903–912.
15. Cao R, Zhong X, Afshari S, et al. Performance of deep learning and genitourinary radiologists in detection of prostate cancer using 3-T multiparametric magnetic resonance imaging. *J Magn Reson Imaging* 2021;54(2):474–483.
16. Mehravand S, Yang D, Harmon SA, et al. A cascaded deep learning-based artificial intelligence algorithm for automated lesion detection and classification on biparametric prostate magnetic resonance imaging. *Acad Radiol* 2022;29(8):1159–1168.
17. Yilmaz EC, Shih JH, Belue MJ, et al. Prospective evaluation of PI-RADS version 2.1 for prostate cancer detection and investigation of multiparametric MRI-derived markers. *Radiology* 2023;307(4):e221309.
18. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol* 2016;40(2):244–252.
19. Vollset SE. Confidence intervals for a binomial proportion. *Stat Med* 1993;12(9):809–824.
20. Hauck WW, Donner A. Wald's test as applied to hypotheses in logit analysis. *J Am Stat Assoc* 1977;72(360):851–853.
21. 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.
22. 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.
23. Klotz L, Chin J, Black PC, et al. Comparison of multiparametric magnetic resonance imaging-targeted biopsy with systematic transrectal ultrasonography biopsy for biopsy-naïve men at risk for prostate cancer: a phase 3 randomized clinical trial. *JAMA Oncol* 2021;7(4):534–542.
24. van der Leest M, Cornel E, Israël B, et al. Head-to-head comparison of transrectal ultrasound-guided prostate biopsy versus multiparametric prostate resonance imaging with subsequent magnetic resonance-guided biopsy in biopsy-naïve men with elevated prostate-specific antigen: a large prospective multicenter clinical study. *Eur Urol* 2019;75(4):570–578.
25. Ahdoot M, Lebastchi AH, Long L, et al. Using Prostate Imaging-Reporting and Data System (PI-RADS) scores to select an optimal prostate biopsy method: a secondary analysis of the trio study. *Eur Urol Oncol* 2022;5(2):176–186.
26. Oerther B, Engel H, Bamberg F, Sigle A, Gratzke C, Benndorf M. Cancer detection rates of the PI-RADSv2.1 assessment categories: systematic review and meta-analysis on lesion level and patient level. *Prostate Cancer Prostatic Dis* 2022;25(2):256–263.
27. Aldoj N, Lukas S, Dewey M, Penzkofer T. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. *Eur Radiol* 2020;30(2):1243–1253.
28. Sanford T, Harmon SA, Turkbey EB, et al. Deep-learning-based artificial Intelligence for PI-RADS classification to assist multiparametric prostate MRI interpretation: a development study. *J Magn Reson Imaging* 2020;52(5):1499–1507.
29. Arif M, Schoots IG, Castillo Tovar J, et al. Clinically significant prostate cancer detection and segmentation in low-risk patients using a convolutional neural network on multi-parametric MRI. *Eur Radiol* 2020;30(12):6582–6592.
30. Schelb P, Wang X, Radtke JP, et al. Simulated clinical deployment of fully automatic deep learning for clinical prostate MRI assessment. *Eur Radiol* 2021;31(1):302–313.
31. Hamm CA, Baumgärtner GL, Biessmann F, et al. Interactive explainable deep learning model informs prostate cancer diagnosis at MRI. *Radiology* 2023;307(4):e22276.
32. Winkel DJ, Wetterauer C, Matthias MO, et al. Autonomous detection and classification of PI-RADS lesions in an MRI screening population incorporating multicenter-labeled deep learning and biparametric imaging: proof of concept. *Diagnostics (Basel)* 2020;10(11):951.
33. Schelb P, Tavakoli AA, Tubtawee T, et al. Comparison of prostate MRI lesion segmentation agreement between multiple radiologists and a fully automatic deep learning system. *Rofo* 2021;193(5):559–573.
34. Dai Z, Carver E, Liu C, et al. Segmentation of the prostatic gland and the intraprostatic lesions on multiparametric magnetic resonance imaging using mask region-based convolutional neural networks. *Adv Radiat Oncol* 2020;5(3):473–481.
35. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11(2):88–94.
36. Belue MJ, Yilmaz EC, Daryanani A, Turkbey B. Current status of biparametric MRI in prostate cancer diagnosis: literature analysis. *Life (Basel)* 2022;12(6):804.
37. Woo S, Suh CH, Kim SY, Cho JY, Kim SH, Moon MH. Head-to-head comparison between biparametric and multiparametric MRI for the diagnosis of prostate cancer: a systematic review and meta-analysis. *AJR Am J Roentgenol* 2018;211(5):W226–W241.
38. Netzer N, Eith C, Bethge O, et al. Application of a validated prostate MRI deep learning system to independent same-vendor multi-institutional data: demonstration of transferability. *Eur Radiol* 2023;33(11):7463–7476.