Original Investigation

Improving Prostate Cancer Detection With MRI: A Multi-Reader, Multi-Case Study Using Computer-Aided Detection (CAD)

Mark A. Anderson, MD[#], Sarah Mercaldo, PhD[#], Ryan Chung, MD, Ethan Ulrich, BS, Randall W. Jones, PhD, MBA, Mukesh Harisinghani, MD

Rationale and Objectives: To evaluate whether addition of a computer-aided diagnostic (CAD) generated MRI series improves detection of clinically significant prostate cancer.

Materials and Methods: Nine radiologists retrospectively interpreted 150 prostate MRI examinations without and then with an additional random forest-based CAD model-generated MRI series. Characteristics of biopsy negative versus positive (Gleason $\geq$ 7 adenocarcinoma) groups were compared using the Wilcoxon test for continuous and Pearson's chi-squared test for categorical variables. The diagnostic performance of readers was compared without versus with CAD using MRMC methods to estimate the area under the receiver operator characteristic curve (AUC). Inter-reader agreement was assessed using weighted inter-rater agreement statistics. Analyses were repeated in peripheral and transition zone subgroups.

Results: Among 150 men with median age 67 $\pm$ 7.4 years, those with clinically significant prostate cancer were older (68 $\pm$ 7.6 years vs. 66 $\pm$ 7.0 years; $p < .02$), had smaller prostate volume (43.9 mL vs. 60.6 mL; $p < .001$), and no difference in prostate specific antigen (PSA) levels (7.8 ng/mL vs. 6.9 ng/mL; $p = .08$), but higher PSA density (0.17 ng/mL/cc vs. 0.10 ng/mL/cc; $p < .001$). Inter-rater agreement (IRA) for PI-RADS scores was moderate without CAD and significantly improved to substantial with CAD (IRA = 0.47 vs. 0.65; $p < .001$). CAD also significantly improved average reader AUC (AUC = 0.72, vs. AUC = 0.67; $p = .02$).

Conclusion: Addition of a random forest method-based, CAD-generated MRI image series improved inter-reader agreement and diagnostic performance for detection of clinically significant prostate cancer, particularly in the transition zone.

Key Words: prostate adenocarcinoma; magnetic resonance imaging; bi-parametric prostate MRI; computer-aided diagnosis; random forest model.

Abbreviations: CAD Computer-aided diagnostic, **AUC** area under the receiver operator characteristic curve, **PSA** prostate specific antigen, **bpRF** boosted parallel random forest, **mpMRI** multiparametric MRI, **bpMRI** biparametric MRI, **PI-RADS** Prostate Imaging Reporting & Data System, **TRW** T2-weighted, **DWI** diffusion-weighted imaging, **ADC** apparent diffusion coefficient, **MRMC** multi-reader multi-case, **PPV** positive predictive value, **IQR** interquartile ratio, **IRA** inter-rater agreement, **BPH** benign prostatic hypertrophy

INTRODUCTION

Multiparametric MRI (mpMRI) has a well-established role in diagnostic work up of prostate cancer as it has been shown to provide enhanced detection of clinically significant prostate cancer and thereby allow accurate targeting of these lesions using MRI guided ultrasound (US) fusion biopsy (1–3). The success of this pathway is predicated on consistent assessment of the mpMRI, however, the diagnostic performance of the Prostate Imaging Reporting and Data System (PI-RADS) classification for prostate cancer detection by MRI has been shown to vary widely across institutions, reader experience, care setting, MRI field strength, scan protocol, and version of PI-RADS classification (4–13). Additionally, poor to fair

Acad Radiol 2022; ■:1–10

From the Department of Radiology, Abdominal Imaging Division, Massachusetts General Hospital, 55 Fruit Street, White 270, Boston, MA 02114 (M.A.A., R.C., M.H.); Harvard Medical School, 55 Fruit Street, White 270, Boston, MA 02114 (M.A.A., R.C., M.H.); Department of Radiology, Institute for Technology Assessment, Massachusetts General Hospital, Boston, Massachusetts (S.M.) (S.M.); Bot Image Inc., Omaha, Nebraska (E.U., R.W.J.) (E.U., R.W.J.). Received August 11, 2022; revised September 6, 2022; accepted September 12, 2022. **Address correspondence to:** M.A. A. e-mail: mark.anderson@mgh.harvard.edu

[#] Mark A. Anderson and Sarah Mercaldo contributed equally to this work.

https://doi.org/10.1016/j.acra.2022.09.009

inter-reader agreement has been demonstrated for detection of prostate cancer in the transition zone, without improvement using newer PI-RADS classification schemes (5,14).

Due to this variability, artificial intelligence methods for lesion detection and characterization have been examined for their ability to assist radiologists in interpretation of prostate MRI. Texture and histogram analysis as well as deep learning algorithms for lesion detection and characterization have been used to facilitate prostate MRI interpretation (15–18). One approach which has shown good diagnostic performance for detection of clinically significant prostate cancer is computer-aided diagnosis (CAD) using a random forest (RF) method based on signal intensity, texture features, and spatial analysis from T2-weighted (T2W), apparent diffusion coefficient (ADC) map, and high b-value (>1000 s/mm^2) diffusion-weighted images (DWI) (19,20). This CAD device produces a color map overlaid on the axial T2W series to highlight regions within the prostate that are suspicious for cancer. The device also produces a numeric value for each detected lesion that represents the likelihood of prostate cancer. This CAD device does not rely on contrast-enhanced sequences, so can be utilized as part of an mpMRI or biparametric MRI (bpMRI) protocol, allowing for use in patients who cannot receive intravenous contrast. We hypothesized that the addition of a CAD generated MRI series would improve reader sensitivity, positive predictive value, and AUC. The purpose of this study was to evaluate whether addition of a computer-aided diagnostic generated MRI image series improves detection performance for identification of clinically significant prostate cancer.

MATERIALS AND METHODS

This retrospective study was approved by the local Institutional Review Boards at patient contributing sites and complied with the Health Insurance Portability and Accountability Act guidelines. Subject informed consent was waived because of the retrospective study design.

Model Development and Implementation

There were 958 prostate MRI examinations, independent from those 150 patients reviewed in this study, that were annotated to establish the reference standard for the CAD model that was used in this study. Annotations for cancer, benign, and normal tissues were created using biopsy locations and clinical reports provided by the sources for each examination. Sources for these data were various institutions whose radiologists were academic, abdominal fellowship-trained domain experts with at least 15 years of experience. These radiologists interpreted the studies and then provided targets of suspect lesions for the interventionalist using either

TABLE 1. Descriptive Statistics by Negative and Positive Biopsies

	Biopsy Negative *N* = 83 *N* (%)	Biopsy Positive *N* = 67 *N* (%)	Combined *N* = 150 *N* (%)	*p*[a]
Age, (years) Median (IQR)	66.0 (60.5,70.5)	68.0 (64.0,73.5)	67.0 (62.0,71.0)	.02
PSA, (ng/mL) Median (IQR)[b]	6.9 (4.9,9.4)	7.8 (5.4,14.8)	7.2 (5.1,11.0)	.08
Prostate volume (mL) Median (IQR)	60.6 (44.6, 83.9)	43.9 (34.8,53.8)	50.6 (40.1, 74.1)	<.001
PSA Density Median (IQR)[b]	0.10 (0.07, 0.14)	0.17 (0.11, 0.29)	0.13 (0.08, 0.19)	<.001
Zone				.14
Peripheral zone	46 (55.4%)	45 (67.2%)	91 (60.7%)	
Transition zone	37 (44.6%)	22 (32.8%)	59 (39.3%)	
Scanner manufacturer				.89
GE	16 (19.3%)	13 (19.4%)	29 (19.3%)	
Philips	45 (54.2%)	34 (50.8%)	79 (52.7%)	
Siemens	22 (26.5%)	20 (29.9%)	42 (28.0%)	
Field strength				.38
1.5	33 (39.8%)	22 (32.8%)	55 (36.7%)	
3	50 (60.2%)	45 (67.2%)	95 (63.3%)	
Source				.02
1	2 (2.4%)	4 (6.0%)	6 (4.0%)	
2	25 (30.1%)	18 (26.9%)	43 (28.7%)	
3	13 (15.7%)	0 (0%)	13 (8.7%)	
4	12 (14.5%)	13 (19.4%)	25 (16.7%)	
5	22 (26.5%)	23 (34.3%)	45 (30.0%)	
6	9 (10.8%)	9 (13.4%)	18 (12.0%)	

Wilcoxon Rank-sum Tests were used to compare biopsy negative and positive groups for continuous variables, and Chi-square tests were used to compare groups for categorical variables.

[a] Considered statistically significant when <.05.

[b] PSA missing for 38 patients.

in-bore guided biopsies (N = 597), MRI-US fusion guided biopsies (N = 250), or 12-core transrectal ultrasound plus 2–4 cognitive targets (N = 28). The remaining cases (N = 83) were MRI-negative and did not undergo biopsy. All biopsy locations were clearly annotated in three dimensions on the MRI data sets, and those suspect lesions with positive biopsies (Gleason ≥ 3+4) were then segmented such that the software could then train from these known biopsy locations with non-suspicious, non-biopsied locations considered as normal or noncancerous tissues. A cancer prediction model based on a boosted parallel RF (bpRF) model was trained utilizing 64 different T2W, DWI, and ADC signal intensity and texture features, validation of which is described separately (19,20).

Patients

Our study population included 150 men with 150 prostate MRI examinations performed between 1/2004 and 1/2019, 67 (44.7%) of which had biopsy confirmed prostate adenocarcinoma, 46 (30.7%) with biopsy confirmed absence of cancer, and 37 (24.7%) with MRI-negative scans (Table 1). Biopsies were obtained using MRI-US fusion guided biopsy (N = 61), in-bore MRI-guided biopsy (N = 36), or cognitive fusion as part of systemic transrectal US biopsy (N = 16). Patient inclusion is presented in Figure 1.

From the entire available prostate MRI imaging database (N = 2037), cases were excluded if they had less than a full image set consisting of axial T2W and diffusion weighted images (N = 10), inadequate field of view which excluded any part of the prostate gland (N = 31), inadequate image quality including motion or hardware related artifact (N = 6), prior therapy to the prostate including ablation (N = 25), transurethral resection (N = 2), or radical prostatectomy (N = 6), incomplete follow-up (e.g., pathology report unavailable) (N = 20), or inadequate resolution which did not meet the resolution requirements for the CAD device including field of view < 140 mm, in-plane voxel spacing > 0.688 mm, and slice thickness > 4.6 mm for T2W images and in-plane voxel spacing > 2.26 mm and slice thickness > 5.0 mm for the DWI and ADC images (N = 65).

Prostate studies from the remaining database with sufficient ground truth were sampled using stratified sampling to ensure certain sub-categories were represented in the clinical study. Cases that were not selected as part of the 150 patients reviewed in this study were used for CAD algorithm development and testing described previously. Sufficient ground truth criteria included cases with either 1) clinical follow-up

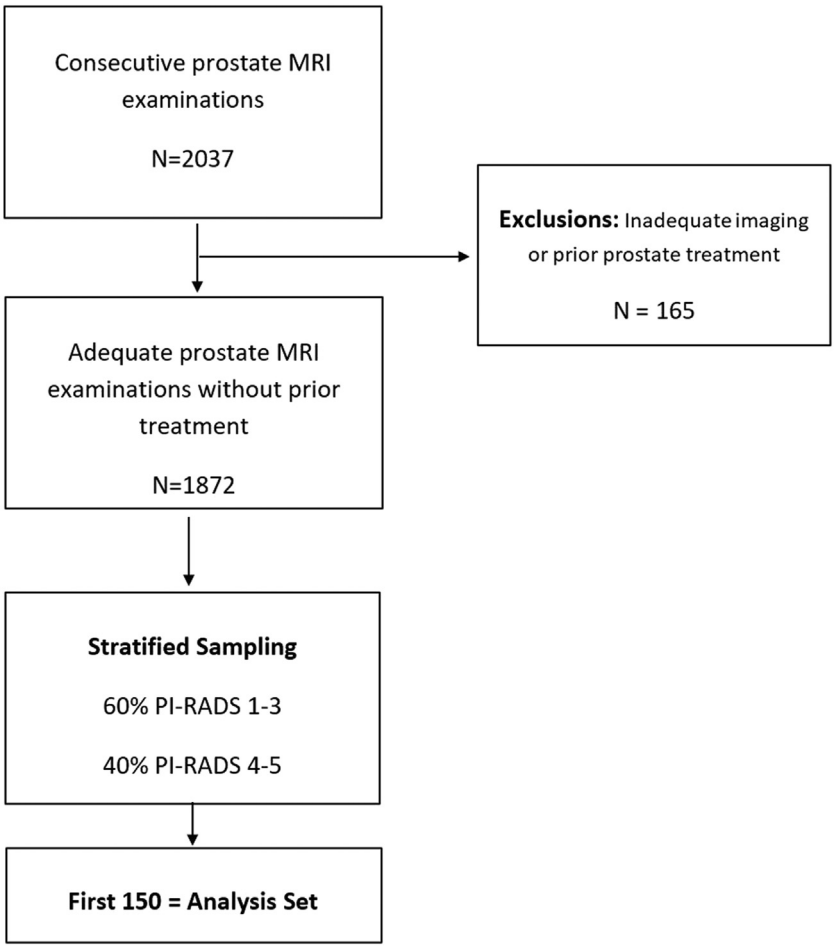

Figure 1. Flow diagram of study design and subject selection. (Color version of figure is available online.)

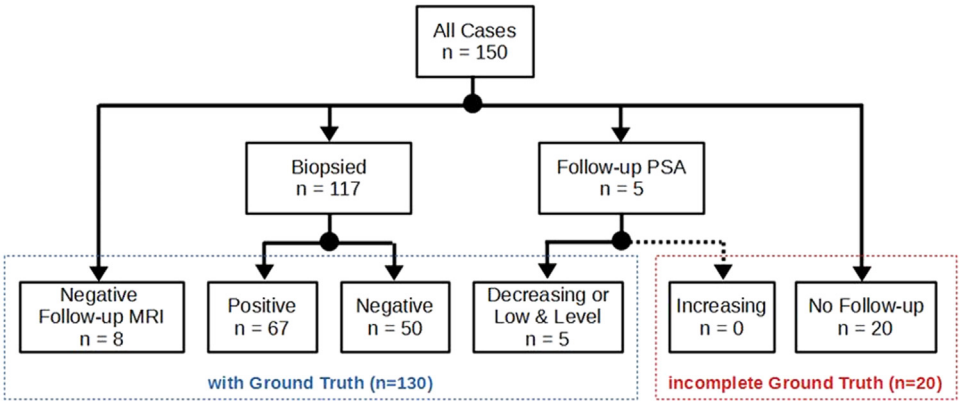

Figure 2. Flow diagram of study participant follow-up. (Color version of figure is available online.)

consisting of biopsy or favorable PSA at follow-up defined as decreasing PSA or level PSA less than 5 ng/mL after at least 5 months after an initial negative prostate MRI, or 2) panel reviewed cases deemed MRI-negative from a panel of five domain experts who had all available clinical information (Fig 2).

Stratified sampling of data was then used among cases with sufficient ground truth and cases were grouped by predetermined subcategories. Then samples were randomly selected from every sub-category until the desired total of 150 was met. For our study, data was grouped into strata defined by scanner manufacturer, field strength, and PI-RADS. We desired to end up with a relatively balanced number (60−40 mix of "normal"/nonsuspicious and suspicious negative versus suspicious positive cases) estimating that these would roughly correlate with groups of PI-RADS 1−3 (60%), and PI-RADS 4−5 (40%). The first 150 cases that fit these criteria were included in the study.

These cases were separate from those used in model development described previously. MRI was performed at both 3T (63.3%) and 1.5T (36.7%) field strength including three scanner vendors (GE, Siemens, and Philips), from six different sites. An endorectal coil was utilized in 37 (24.7%) patients including two scanned on Siemens and 35 on Philips machines.

Readers and Training

Nine fellowship-trained attending radiologists who interpret prostate MRI in their clinical practice with varying levels of experience and practice setting were included (Table 2). Five radiologists were abdominal imaging fellowship trained. Additional fellowships included body imaging, thoracoabdominal imaging, MRI, and combined interventional and diagnostic body imaging. Reader experience interpreting prostate MRI in the clinical setting ranged from 1 to 10 years with a median of 3 years. Practice setting included academic (n = 5) and nonacademic (n = 4) practices.

Each reader was provided training materials included in the User's Manual with an additional one-on-one follow-up call by the clinical coordinator to further clarify and address any questions. Focus was on how the readers were to complete a convenient data sheet wherein they were to identify each suspect lesion by its centroid location within the 36 subvolumes (30 when combining the anatomic zones and their associated axial slice regions) as identified and labeled in PI-RADS.

Readers were given the 150 prostate MRI cases in random order and were blinded to the pathological diagnosis. MRI cases included axial T2W, axial DWI with median maximum b-value 1400 s/mm^2 (range 750−1600 s/mm^2), axial ADC map, with or without axial T1W postcontrast images in random order. Readers marked each focal lesion of ≥ PI-RADS 3 for anatomic location and image number where the lesion was centered. Readers were also provided PSA levels for each patient except for 38 cases for which PSA was unavailable. A PI-RADS score (using PI-RADS version 2.0) for each lesion was assigned and note made as to whether the radiologist recommended targeted biopsy of the lesion (12,13).

Following a washout period of at least 30 days, readers were re-presented with the same 150 examinations with the addition of the CAD generated image series, in a newly randomized order, and again blinded to pathological diagnosis. The same classification system was applied by each reader to each examination. The CAD generated image series was a copy of the T2W axial series overlaid with the colorized probability map or index − an estimate of the likelihood of prostate cancer. Through training on thousands of images with known pathology locations and segmented lesions, the computer generated 100-point scale was scaled to optimally classify prostate tissues.

Image Interpretation

Lesion locations were described in a spreadsheet using a limited set of descriptors to decrease ambiguity of detections. Descriptors included the slice number on axial T2W series, the right-left position (right, middle, or left), the anterior-posterior position (anterior, middle, or posterior), and the prostate subregion (peripheral or transition). The reference standard was used to establish a truth table for matching

TABLE 2. Study Reader Characteristics

Reader	Practice Setting	Location (USA)	Experience (years)
1	Nonacademic	New York	0.5
2	Nonacademic	New York	1.5
3	Academic	Minnesota	2
4	Academic	Massachusetts	2
5	Academic	Minnesota	3
6	Nonacademic	California	4
7	Academic	Massachusetts	4
8	Nonacademic	California	9
9	Academic	Colorado	10

detections to lesions and their biopsy-proven outcomes. Some leeway was allowed in the truth table to further reduce ambiguity of detections. For example, if a biopsy-proven lesion existed on the border of the peripheral zone and the transition zone, then either descriptor would be considered a match. Similarly, a large lesion may occupy multiple slices in the T2W series, so identifying any of the slices near the center of the lesion would be considered a match.

The detections of the nine readers were compared to the truth table. Patients with a Gleason score of 7 or higher were considered biopsy positive, and patients with a Gleason score of 6 or below or no cancer were considered biopsy negative. For each detection that matched a biopsy-proven positive lesion in the truth table, a true positive was assigned with the corresponding PI-RADS score. All missed detections were artificially assigned a PI-RADS 1 score. If a detection matched a biopsy-proven benign lesion or did not match a lesion described in the truth table, then a false positive was assigned with the corresponding PI-RADS score. Detections of the CAD method were manually inspected and matched to the lesions of the truth table. Instead of assigning the PI-RADS score, the detection rating by the CAD method was the probability score. In measuring the CAD performance, the CAD index value (between 1% and 100% correlating to cancer probability) at each ground truth biopsy point was compared to the Gleason score from that anatomic location. A benign biopsy was assigned a CAD index of 0. A missed lesion was a lesion with Gleason grade $\geq$7 and with CAD index less than 50%. If the CAD output was <50% in a non-cancerous area, this was deemed a true negative while a CAD output of $\geq$ 50% in a negative area as deemed a false positive.

Statistical Analysis

We compared the distribution of case characteristics in the biopsy negative group versus the biopsy positive group using the Wilcoxon test (for continuous variables) and the Pearson's chi-squared test (for categorical variables). For all cases, all nine readers reported a PI-RADS score both without and with the assistance of CAD.

Inter-reader agreement of PI-RADS scores without and with CAD was assessed using a quadratic weighted inter-rater agreement approach for ordinal responses with bootstrapped 95% confidence intervals (CIs). The strength of agreement of inter-rater agreement values can be interpreted as: <0 poor, 0.01−0.20 slight, 0.21−0.40 fair, 0.41−0.60 moderate, 0.61−0.80 substantial, 0.81−1.00 almost perfect (21,22).

Based upon PI-RADS scores, multi-reader multi-case (MRMC) analysis was used to compare the reader performance without and with CAD, as measured by the difference in the average area under the receiver operator characteristic curve (AUC) and 95% CIs. Pooled results were calculated using a random effect for both readers and cases using Obuchowski-Rockette methods (23).

Using each possible PI-RADS score (1−5) as a threshold for a positive case determination, sensitivity, and positive predictive value (PPV) were calculated using generalized estimation equation analyses to compare pooled data corrected for repeated measurements by multiple readers.

Analyses were repeated within peripheral and transition zonal subgroups. For cases with only one lesion, the primary zone designation was the single lesion location. For cases with multiple lesions, the zone designation was made for the zone with the most lesions and then the lesion with the highest Gleason grade or the primary lesion identified.

We used a type-I error of 5% for all confidence intervals and two-sided hypothesis tests. We analyzed data using statistical software R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria, 2019). The packages "rms," "geepack," and "MRMCaov" were used in this analysis (24,25).

RESULTS

One hundred fifty men were included with a median age 67 years (interquartile ratio (IQR) 62−71 years). Sixty-seven (44.7%) were biopsy positive (presence of clinically significant prostate adenocarcinoma of Gleason grade $\geq$ 7), 46 (30.7%) were biopsy negative (absence of clinically significant prostate adenocarcinoma, Gleason grade $\geq$7), and 37 (24.7%) were MRI-negative. Men with clinically significant prostate cancer were older than men with biopsy negative for clinically significant cancer (68 years, IQR 64−73.5 vs. 66 years, IQR 60.5−70.5; p = .02) and had smaller prostate gland volume (43.9 mL, IQR 34.8−53.8 vs. 60.6 mL, IQR 44.6−83.9; p < .001). There was no evidence of a significant difference in prostate specific antigen (PSA) level between biopsy positive and negative subjects (7.8 ng/mL, IQR 5.4−14.8 vs. 6.9 ng/mL, IQR 4.9−9.4; p = .08). However, PSA density was significantly higher for patients with clinically significant cancer versus without (0.17 ng/mL/cc, IQR 0.08−0.19 vs. 0.10 ng/mL/cc, IQR 0.07−0.14; p < .001).

The weighted inter-rater agreement (IRA) statistics between the nine readers for PI-RADS scores without CAD was moderate (IRA = 0.47, 95% CI 0.41−0.52). Inter-rater agreement for PI-RADS scores with CAD showed substantial agreement and was significantly higher than without CAD (p < .001; IRA = 0.65; 95% CI 0.60−0.69), (Table 3).

TABLE 3. Multireader Multicase and Kappa Analysis Comparing Reads Without and With CAD

	Without CAD	With CAD	Difference Δ (95% CI)	p^a
AUC				
Reader 1 (0.5 years)	0.42 (0.33–0.51)	0.51 (0.42–0.60)	0.09 (-0.04 to 0.21)	.17
Reader 2 (1.5 years)	0.72 (0.64–0.80)	0.74 (0.66–0.82)	0.02 (-0.05 to 0.09)	.60
Reader 3 (2 years)	0.74 (0.66–0.82)	0.74 (0.66–0.81)	-0.002 (-0.07 to 0.06)	.95
Reader 4 (2 years)	0.76 (0.68–0.84)	0.74 (0.66 to 0.82)	-0.02 (-0.08 to 0.20)	.54
Reader 5 (3 years)	0.74 (0.67–0.82)	0.79 (0.72–0.86)	0.04 (-0.02 to 0.11)	.20
Reader 6 (4 years)	0.72 (0.63–0.80)	0.78 (0.70–0.85)	0.06 (-0.02 to 0.14)	.13
Reader 7 (4 years)	0.63 (0.54–0.72)	0.71 (0.63–0.78)	0.08 (-0.02 to 0.18)	.13
Reader 8 (9 years)	0.64 (0.55–0.72)	0.73 (0.65–0.81)	0.10 (-0.01 to 0.20)	.07
Reader 9 (10 years)	0.70 (0.61–0.78)	0.74 (0.66–0.82)	0.04 (-0.04 to 0.12)	.28
Average (All)	0.67 (0.58–0.76)	0.72 (0.64–0.80)	0.05 (0.01–0.08)	.02
Average (PZ subgroup)	0.67 (0.56–0.77)	0.71 (0.62–0.81)	0.05 (-0.003 to 0.10)	.07
Average (TZ subgroup)	0.67 (0.57–0.77)	0.73 (0.63–0.84)	0.06 (0.02–0.10)	.005
Weighted Inter-rater Agreement				
PI-RADS (All)	0.47 (0.41–0.52)	0.65 (0.60–0.69)	0.18 (0.11–0.25)	<.001
PI-RADS (PZ subgroup)	0.51 (0.45–0.57)	0.67 (0.63–0.72)	0.16 (0.09–0.23)	<.001
PI-RADS (TZ subgroup)	0.39 (0.30–0.48)	0.60 (0.52–0.67)	0.21 (0.10–0.32)	<.001

Note—Values expressed as estimate, with 95% CI in parentheses unless otherwise specified.
a Considered statistically significant when <.05.

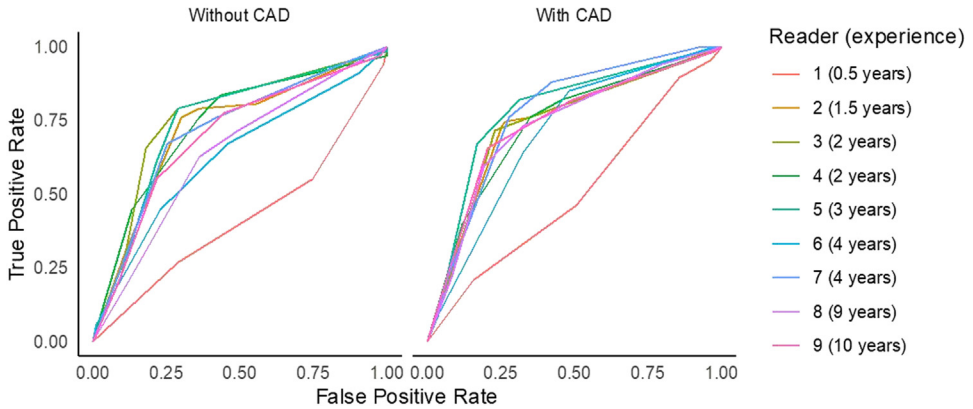

Figure 3. ROC curves for readers without and with CAD. (Color version of figure is available online.)

MRMC analysis reader performance showed that average AUC for all nine readers without CAD (AUC = 0.67, 95% CI, 0.58–0.76, p = .02) significantly increased with CAD (AUC = 0.72, 95% CI, 0.64–0.80) (Table 3, Figs 3 and 4).

The overall distribution of cancers were 64% peripheral zone and 36% transition zone. When comparing the AUC for the average of all readers without versus with CAD by zone, there was no significant difference in peripheral zone cancers but there was a significant increase in AUC for transition zone cancers with addition of CAD (0.67, 95% CI, 0.57–0.77 vs. 0.73, 95% CI 0.63–0.84), p = .004). Inter-rater agreement was also improved to a greater degree in the transition zone with the addition of CAD (p < .001) but also significantly improved agreement in the peripheral zone (p < .001).

Sensitivity for identifying clinically significant cancer for PI-RADS 3 lesions in the transition zone without CAD significantly increased with addition of CAD (0.70, 95% CI, 0.63–0.76; p = .003 vs. .83, 95% CI, 0.77–0.87) and pooling all PI-RADS 3 lesions in any zone (0.76, 95% CI 0.73–0.80; p = .02 vs. .82, 95% CI, 0.79–0.85). Sensitivities and PPVs for other PI-RADS classifications were not significantly different without versus with CAD (Table 4).

DISCUSSION

Our results show that the addition of a boosted parallel random forest model-based CAD generated MRI image series improves inter-reader agreement for PI-RADS classification and the average detection performance for identification of

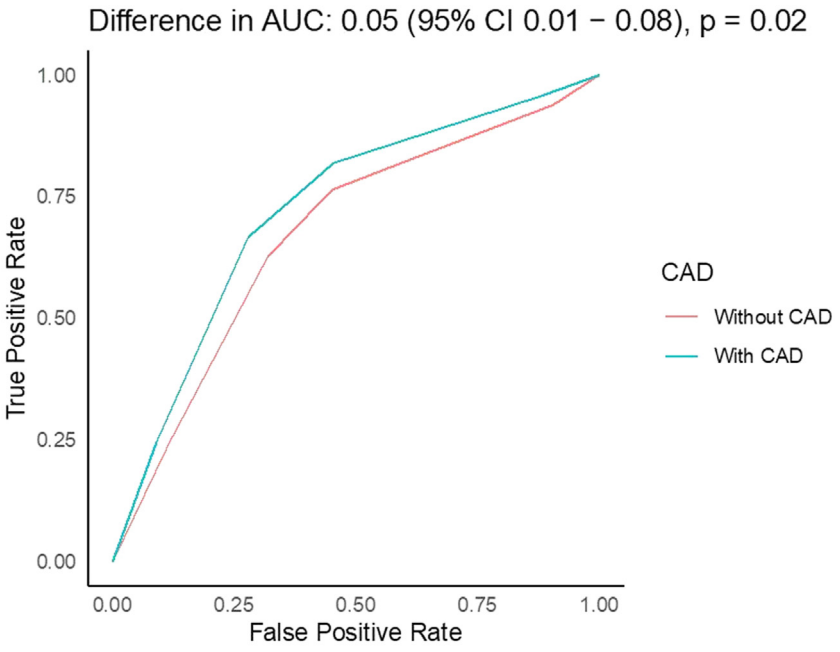

Figure 4. ROC curves for average of readers without versus with CAD. (Color version of figure is available online.)

TABLE 4. Comparison of Sensitivity and PPV Without and With CAD for Each Possible PI-RADS Score Threshold, Overall and Within the Peripheral and Transition Zone Subgroups

PI-RADS Threshold for Performance Metric Evaluation	Overall Without vs. With CAD		Peripheral Zone Without vs. With CAD		Transition Zone Without vs. With CAD	
PI-RADS ≥ 1						
Sensitivity (%)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	1.0 (1.0–1.0)
PPV (%)	0.45 (0.42–0.47)	0.45 (0.42–0.47)	0.49 (0.46–0.53)	0.49 (0.39–0.60)	0.37 (0.33–0.41)	0.37 (0.33–0.41)
PI-RADS ≥ 2						
Sensitivity (%)	0.94 (0.92–0.96)	0.96 (0.94, 0.97)	0.95 (0.93–0.97)	0.96 (0.93, 0.98)	0.91 (0.86–0.94)	0.94 (0.90, 0.97)
PPV (%)	0.46 (0.43–0.48)	0.47 (0.44–0.50)	0.51 (0.47–0.54)	0.51 (0.41–0.61)	0.38 (0.26–0.51)	0.39 (0.28–0.52)
PI-RADS ≥ 3						
Sensitivity (%)	0.76 (0.73–0.80)[a]	0.82 (0.79–0.85)[a]	0.80 (0.75–0.83)	0.81 (0.72–0.88)	0.70 (0.63–0.76)[a]	0.83 (0.77–0.87)[a]
PPV (%)	0.58 (0.54–0.61)	0.59 (0.56–0.63)	0.62 (0.57–0.66)	0.65 (0.53–0.75)	0.50 (0.44–0.56)	0.51 (0.45–0.56)
PI-RADS ≥ 4						
Sensitivity (%)	0.63 (0.59–0.66)	0.67 (0.63–0.70)	0.65 (0.60–0.69)	0.67 (0.56–0.76)	0.58 (0.51–0.65)	0.66 (0.59–0.72)
PPV (%)	0.61 (0.57–0.65)	0.66 (0.62–0.70)	0.65 (0.60–0.69)	0.70 (0.65–0.74)	0.55 (0.48–0.61)	0.59 (0.53–0.66)
PI-RADS ≥ 5						
Sensitivity (%)	0.25 (0.21–0.28)	0.25 (0.21–0.28)	0.18 (0.14–0.22)	0.18 (0.15–0.22)	0.39 (0.32–0.46)	0.38 (0.31–0.45)
PPV (%)	0.63 (0.56–0.69)	0.69 (0.62–0.75)	0.60 (0.51–0.69)	0.70 (0.60–0.78)	0.66 (0.56–0.74)	0.68 (0.59–0.75)

Note—Values expressed as estimate, with 95% CI in parentheses unless otherwise specified.
[a] Performance with CAD is significantly higher than without CAD, $p < .05$. All other comparisons $p > .05$.

clinically significant prostate cancer among nine readers with varying experience levels in both academic and nonacademic care settings.

Improving the reproducibility of prostate MRI interpretations remains a major challenge for radiologists. In a multicenter cross-sectional study of the PI-RADS classification in men with suspected or biopsy-proven untreated prostate cancer, the detection rate of clinically significant (Gleason score 3+4 or higher) disease varied widely across sites, with interquartile range of positive predictive value (PPV) for PI-RADS ≥3 ranging from 27% to 48% (4). Factors that contributed to PPV variability included anatomical location, with lower PPV for transition zone versus peripheral zone cancers. Conversely, magnet field strength (1.5T versus 3T) and presence of endorectal coil did not significantly affect PPV. Additionally, PPV was not significantly better for men undergoing MRI-fusion prostate biopsy with a history of prior biopsy and cancer under active surveillance versus

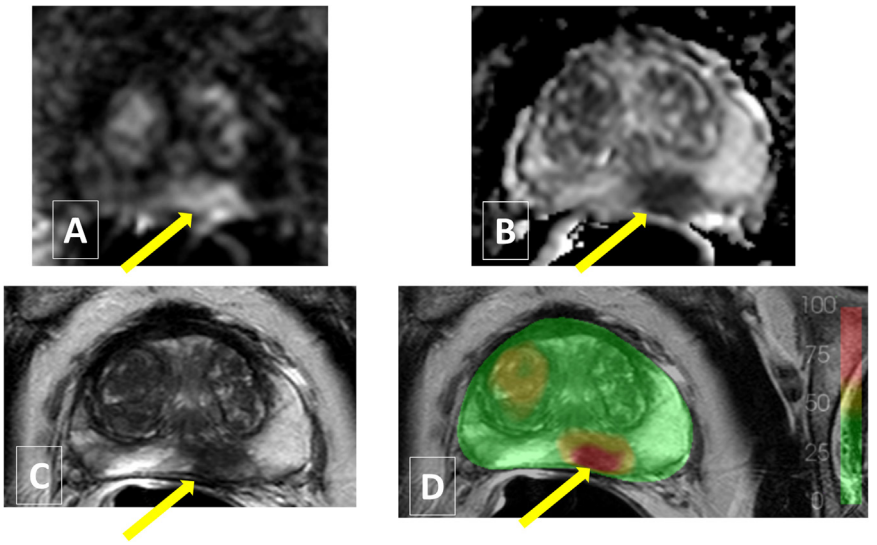

Figure 5.A. 64-year-old male with unknown PSA and no history of prostate cancer. Axial diffusion weighted (b-value 1400 s/mm^2) (A), ADC map (B), T2W (C), and CAD color map overlaid on T2W (D) MR images. There is an ill-defined region of marked diffusion restriction with associated T2 hypointensity in the left posterior peripheral zone. This lesion was classified as PI-RADS 4 by eight of nine readers. MRI-US fusion guided biopsy showed Gleason 8 adenocarcinoma at the left paramedian posterior peripheral zone mid gland. (Color version of figure is available online.)

biopsy naïve men, despite the former patients presumably having a higher pretest probability of clinically significant cancer (4).

The zonal distribution of cancers in our series was in line with known prevalence, with 64% occurring in the peripheral zone (Fig 5) and 36% in the transition zone (Figs 6 and 7) (4). Diagnostic performance within the transition zone, where approximately 30% of cancers occur, is of particular interest since the variable appearance of stromal benign prostatic hypertrophy (BPH) elements can be challenging to discern from sites of cancer, likely contributing to lower PPV for PI-RADS scores within the transition zone (4). Inter-reader agreement for PI-RADS 4 or higher lesions in the transition zone is no higher using the latest PI-RADS classification system (version 2.1) versus older PI-RADS classification schemes and has been shown to be poor (5,14). Tools

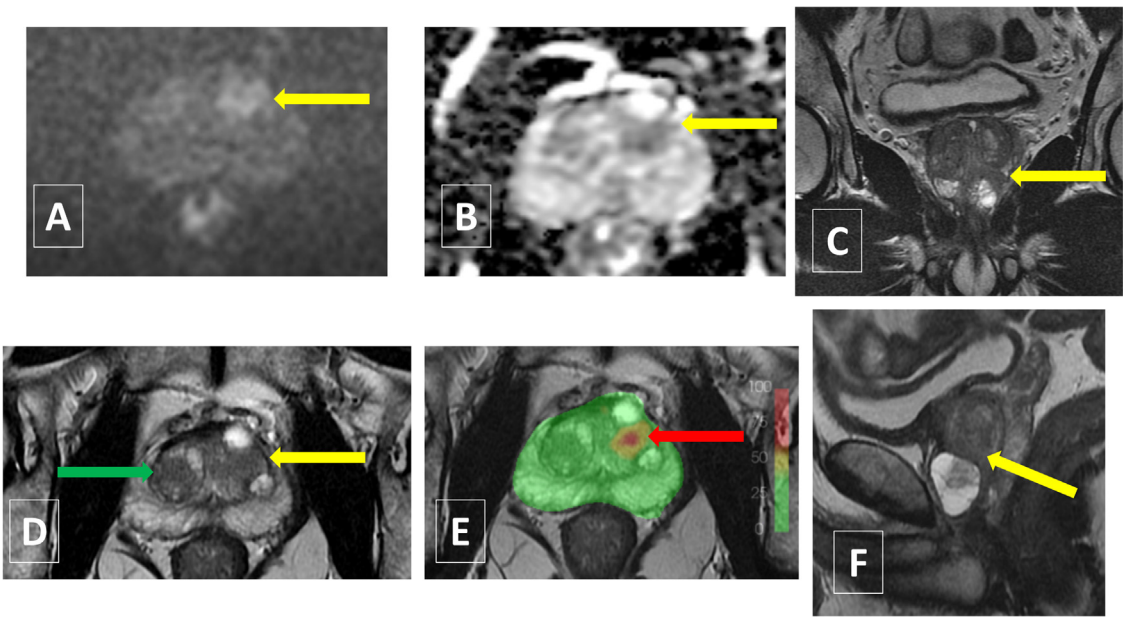

Figure 6.A. 63-year-old male with unknown PSA and no history of prostate cancer. Axial diffusion weighted (b-value 1400 s/mm^2) (A), ADC map (B), coronal T2W (C), axial T2W (D), CAD color map overlaid on axial T2W (E), and sagittal T2W (F) MR images. Within the left mid transition zone, there is a homogeneously T2 hypointense region which is DWI hyperintense and ADC hypointense (yellow arrows, A-D, F). It is challenging to tell whether this is a region within a larger encapsulated BPH nodule (similar finding is present in contralateral right mid transition zone, green arrow, D) or a separate lesion. No reader assigned a PI-RADS 3 or higher or recommended a biopsy for this examination. Addition of CAD overlay on T2W image shows high probability of cancer (red arrow, E) but low probability in the similar appearing right transition zone at the same level, showing the added value of this sequence in the transition zone that is enlarged by BPH nodules of variable T2W signal intensity. TRUS cognitive fusion biopsy demonstrated Gleason 7 adenocarcinoma at the left anterior transition zone mid gland. (Color version of figure is available online.)

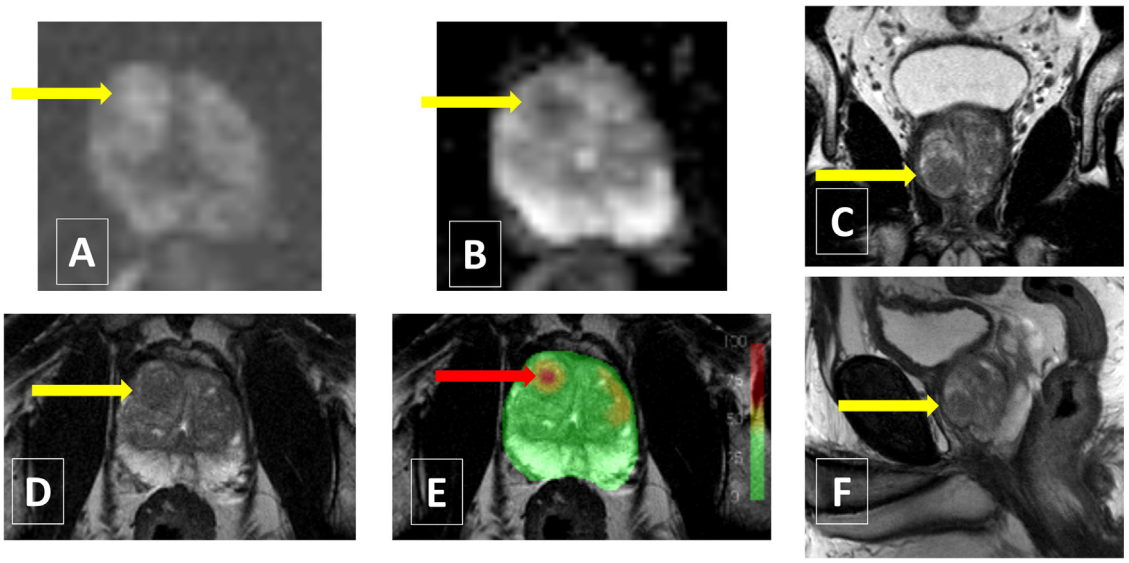

Figure 7. A. 70-year-old male with a history of a prior negative transrectal ultrasound-guided biopsy and PSA 10.3 ng/mL (PSA density 0.18 ng/mL/cc). Axial diffusion weighted (b-value 1400 s/mm^2) (A), ADC map (B), coronal T2W (C), axial T2W (D), CAD color map overlaid on axial T2W (E), and sagittal T2W (F) MR images. Within the right anterior transition zone, there is a homogeneously T2 hypointense region which is DWI hyperintense and ADC hypointense (yellow arrows, A-D, F). This lesion was classified as PI-RADS 4 by six of nine readers. It is challenging to tell whether this is a region within a larger encapsulated BPH nodule or a separate lesion. Addition of CAD overlay on T2W image shows high probability of cancer (red arrow, E) but low probability in elsewhere in the transition zone at the same level, showing the added value of this sequence in the transition zone that is enlarged by BPH nodules of variable T2W signal intensity. MRI-US fusion guided biopsy showed Gleason 7 adenocarcinoma at the right anterior transition zone mid gland. (Color version of figure is available online.)

which can increase the reproducibility of PI-RADS scoring while improving accuracy, particularly in the transition zone, are highly desirable. Our findings that the addition of CAD significantly improved AUC for detection of transition zone cancers and inter-rater agreement for transition zone lesions, bringing it on par with agreement in the peripheral zone, shows that this tool can narrow the gap in diagnostic performance between the peripheral and transition zones (Figs 6 and 7). These findings are reflected by our results of increased sensitivity for intermediate suspicion (PI-RADS 3) lesions in the transition zone with versus without CAD.

Our study subjects included MRI examinations from multiple institutions, scanner vendors, acquired at 3T and 1.5T field strengths, presence and absence of endorectal coils, and variable maximum b-value (including below the suggested minimum of 1400 s/mm^2 according to PI-RADS v2.1 (13), which mimics the variability that occurs in clinical practices of different locations and settings (4). Additionally, we included a range of practice settings and experience for readers. This CAD device relies on noncontrast sequences, which allows it to be employed as part of both bpMRI and mpMRI protocols, which allows for the inclusion of patients who cannot receive intravenous contrast for reasons of renal insufficiency, ordering physician or patient preference, or allergy. These factors may indicate that our results are translatable to a larger set of radiologists and patients, but would need to be verified with further, larger cohorts. The range of parameters of the image data sets used in training and validating the CAD device allow its use across the range of scanner vendor, field strength, maximum b-value, and presence of endorectal coil included in this clinical study, potentially increasing its deployment range.

Limitations of the study include that the ROC analysis was run on per patient basis, meaning that any patient positive for cancer anywhere in the gland was considered a true positive if any PI-RADS 3 or greater lesion was called by a reviewer. Although whole mount MRI prostatectomy specimens would have been a more optimal method for annotation of ground truth for the pre-resection MRIs used in CAD model development, the heterogeneity in contributing institutions and temporal range of the data acquisition preclude this. Relying on targeted and random biopsies for annotations could result in suboptimal annotations of ground truth. There was variability in the MRI examination parameters including field strength and maximum b-value for DWI, presence of endorectal coils, and available clinical information during image review including no available PSA for 25% of patients (38/150). However, this nonuniformity mimics the variability in scan parameters and protocols seen in clinical practice, as does the frequent absence of laboratory information when interpreting images. We did not feel we had the subject numbers or reasonably even distribution across these variables to support a subgroup analysis comparing across these variables. In an era of increasing number of sequences and decreasing slice thickness where radiologists are challenged to look at more images per examination, adding additional CAD generated image series may present challenges to daily workflow. However, an automated process whereby the additional

CAD generated image series is sent directly to PACS is feasible and adds a single axial series to the existing protocols, taking minimal extra time to review.

The addition of a bpRF method-based, CAD generated MRI image series improved inter-rater agreement and diagnostic performance for detection of clinically significant prostate cancer, particularly in the transition zone. This should be considered for clinical use.

FUNDING

None.

REFERENCES

1. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 2017 Feb 25; 389 (10071):815–822.
2. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med 2018 May 10; 378(19):1767–1777.
3. Schoots IG, Roobol MJ, Nieboer D, Bangma CH, et al. Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. Eur Urol 2015 Sep; 68(3):438–450.
4. 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 Jul; 296(1):76–84.
5. Bhayana R, O'Shea A, Anderson MA, et al. PI-RADS Versions 2 and 2.1: interobserver agreement and diagnostic performance in peripheral and transition zone lesions among six radiologists. AJR Am J Roentgenol 2021 Jul; 217(1):141–151.
6. Xu L, Zhang G, Shi B, et al. Comparison of biparametric and multiparametric MRI in the diagnosis of prostate cancer. Cancer Imaging 2019 Dec 21; 19(1):90.
7. Gatti M, Faletti R, Calleris G, et al. Prostate cancer detection with biparametric magnetic resonance imaging (bpMRI) by readers with different experience: performance and comparison with multiparametric (mpMRI). Abdom Radiol (NY) 2019 May; 44(5):1883–1893.
8. Tamada T, Kido A, Yamamoto A, et al. Comparison of biparametric and multiparametric MRI for clinically significant prostate cancer detection with PI-RADS Version 2.1. J Magn Reson Imaging 2021 Jan; 53(1):283–291.
9. Ullrich T, Quentin M, Oelers C, et al. Magnetic resonance imaging of the prostate at 1.5 versus 3.0T: a prospective comparison study of image quality. Eur J Radiol 2017 May; 90:192–197.
10. Li JL, Phillips D, Towfighi S, Wong A, et al. Second-opinion reads in prostate MRI: added value of subspecialty interpretation and review at multidisciplinary rounds. Abdom Radiol (NY) 2022 Feb; 47(2):827–837.
11. Salka BR, Shankar PR, Troost JP, et al. Effect of Prostate MRI Interpretation Experience on PPV Using PI-RADS Version 2: A 6-Year Assessment Among Eight Fellowship-Trained Radiologists. Am. J. Roentgenol 2022 Mar 23: 1–8.
12. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS prostate imaging - reporting and data system: 2015, Version 2. Eur Urol 2016 Jan; 69(1):16–40.
13. Weinreb JC, Barentsz JO, et al. PIRADS-V2-1.pdf [Internet]. American College of Radiology: PI-RADS prostate imaging-reporting and data system, v2.1. 2019 [cited 2022 Jun 13]. Available from: https://www.acr.org/-/media/ACR/Files/RADS/Pi-RADS/PIRADS-V2-1.pdf. Accessed January 5, 2022
14. Kim N, Kim S, Prabhu V, et al. Comparison of prostate imaging and reporting data system V2.0 and V2.1 for evaluation of transition zone lesions: a 5-reader 202-patient analysis. J Comput Assist Tomogr 2022 Apr 8.
15. 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 Nov; 52(5):1499–1507.
16. Mehralivand 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. Academic Radiology [Internet] 2021 Sep 28. [cited 2021 Oct 4]; Available from: https://www.sciencedirect.com/science/article/pii/S1076633221003779. Accessed January 5, 2022.
17. 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 Nov 14; 10(11):E951.
18. Xing P, Chen L, Yang Q, et al. Differentiating prostate cancer from benign prostatic hyperplasia using whole-lesion histogram and texture analysis of diffusion- and T2-weighted imaging. Cancer Imaging 2021 Sep 27; 21(1):54.
19. Lay N, Tsehay Y, Greer MD, et al. Detection of prostate cancer in multiparametric MRI using random forest with instance weighting. J Med Imaging (Bellingham) 2017 Apr; 4(2):024506.
20. Ulrich EJ, Dhaouadi J, Schat R, et al. Comparison of machine learning methods for detection of prostate cancer using bpMRI radiomics features. In: In: Proc Intl Soc Mag Reson Med, London, UK; 2022.
21. Marasini D, Quatto P, Ripamonti E. Assessing the inter-rater agreement for ordinal data through weighted indexes. Stat Methods Med Res 2016 Dec; 25(6):2611–2633.
22. Falotico R, Quatto P. On avoiding paradoxes in assessing inter-reader agreement. Italian J. Appl. Statistics 2010; 22(2):151–161.
23. Obuchowski NA, Rockette HE. Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests an anova approach with dependent observations. Commun Statistics - Simulation and Computation 1995 Jan 1; 24(2):285–308.
24. Smith BJ, Hillis SL. Multi-reader multi-case analysis of variance software for diagnostic performance comparison of imaging modalities. Proc SPIE Int Soc Opt Eng 2020 Feb; 11316:113160K.
25. Smith BJ, Hillis SL, Pesce LL. MCMCaov Multi-reader multi-case analysis of variance. [Internet]. Available from: https://github.com/brian-j-smith/MRMCaov. Accessed January 7, 2022.