Comparison of the impact of rectal susceptibility artifacts in prostate magnetic resonance imaging on subjective evaluation and deep learning: a two-center retrospective study

Zheng Wang[1,2], Peng Lu[1,2], Song Liu[1,2], Chengzhi Fu[1,2], Yong Ye[1,2], Chengxin Yu[1,2*] and Lei Hu[3,4*]

Abstract

Background To compare the influence of rectal susceptibility artifacts on the subjective evaluation and deep learning (DL) in prostate cancer (PCa) diagnosis.

Methods This retrospective two-center study included 1052 patients who underwent MRI and biopsy due to clinically suspected PCa between November 2019 and November 2023. The extent of rectal artifacts in these patients' images was evaluated using the Likert four-level method. The PCa diagnosis was performed by six radiologists and an automated PCa diagnosis DL method. The performance of DL and radiologists was evaluated using the area under the receiver operating characteristic curve (AUC) and the area under the multi-reader multi-case receiver operating characteristic curve, respectively.

Results Junior radiologists and DL demonstrated statistically significantly higher AUCs in patients without artifacts compared to those with artifacts (R1: 0.73 vs. 0.64; $P = 0.01$; R2: 0.74 vs. 0.67; $P = 0.03$; DL: 0.77 vs. 0.61; $P < 0.001$). In subgroup analysis, no statistically significant differences in the AUC were observed among different grades of rectal artifacts for both all radiologists ($0.08 \leq P \leq 0.90$) and DL models ($0.12 \leq P \leq 0.96$). The AUC for DL without artifacts significantly exceeded those with artifacts in both the peripheral zone (PZ) and transitional zone (TZ) (DL_{PZ}: 0.78 vs. 0.61; $P = 0.003$; DL_{TZ}: 0.73 vs. 0.59; $P = 0.011$). Conversely, there were no statistically significant differences in AUC with and without artifacts for all radiologists in PZ and TZ ($0.08 \leq P \leq 0.98$).

Conclusions Rectal susceptibility artifacts have significant negative effects on subjective evaluation of junior radiologists and DL.

Clinical trial number Not applicable.

*Correspondence:
Chengxin Yu
ycyucx@163.com
Lei Hu
hulei@gdph.org.cn
Full list of author information is available at the end of the article

Background

Prostate cancer (PCa) stands as the most prevalent malignancy within the male genitourinary system. It ranks second among malignant tumors afflicting men worldwide in terms of its incidence rate [1]. Precision diagnosis of PCa contributes to the reduction of overtreatment in low-risk patients. It also extends the survival time of high-risk patients and enhances their quality of life [2–5].

Magnetic resonance imaging (MRI) serves as an effective non-invasive method for diagnosing PCa [6–8]. However, MRI is often hindered by rectal susceptibility artifacts, which are the most common types of artifacts encountered. These artifacts typically manifest as local signal loss or decreased signal intensity in prostate tissue, potentially resulting in misdiagnosis or overlooked detection of PCa [9–11]. However, the influence of rectal artifacts on PCa diagnosis has not been fully elucidated.

Previous studies have not reached a consensus on the diagnostic impact of rectal susceptibility artifacts on subjective evaluation. Plodeck et al. [12] found that susceptibility artifacts due to rectal dilation could affect subjective evaluation. Conversely, Coskun et al. [13] found that reducing rectal artifacts did not lead to an improvement in subjective evaluation by radiologists and urologists. Similarly, the impact of rectal artifacts on MRI-based deep learning (DL) has not been fully clarified. Although numerous MRI-based DL models using deep neural networks (DNNs) have been proposed and have demonstrated comparable or even superior performance in PCa diagnosis compared to subjective evaluation in recent years [14–16], these DL models may be more susceptible to being misled by rectal susceptibility artifacts. Previous studies indicated that DNNs are extremely sensitive to detect noise changes and angular distortions that occur within the images [17, 18]. Another clinical study also indicates that rectal susceptibility artifacts are independent risk factors that lead to erroneous predictions of DL [19]. As far as we know, there have been no clinical studies comparing the impact of rectal susceptibility artifacts on subjective evaluation and DL in prostate MRI.

Clarifying the influence of rectal susceptibility artifacts is crucial for devising follow-up scanning strategies aimed at enhancing diagnostic accuracy. This study aims to investigate and compare the effects of rectal susceptibility artifacts on subjective evaluation and DL.

Methods

This retrospective study was approved by the Institutional Ethical Committee of the local hospital [Approve No: 2023-001-01]. The requirement for written informed consent was waived due to the retrospective and anonymous analysis of the data. All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki Declaration and its later amendments.

Patient cohort

Between November 2019 and November 2023, participants with clinically suspected PCa who underwent MRI examinations were consequently enrolled from two medical centers (Yichang Central People's Hospital and Guangdong Provincial People's Hospital). The inclusion criteria included the following: (a) ultrasound-guided prostate biopsy confirming PCa; (b) 3.0T prostate biparametric MRI (bpMRI) examination according to the Prostate Imaging Reporting and Data System (PI-RADS) version 2.1; and (c) no prior history of PCa treatment. The exclusion criteria were as follows: (a) patients with incomplete MRI information or missing sequences, which may introduce selection bias by excluding cases with suboptimal imaging quality; and (b) unable to obtain a conclusive histopathological diagnosis.

The study finally enrolled 1052 patients, comprising 427 with PCa and 625 without PCa. Details of patient selection are shown in Fig. 1. Clinical and pathological characteristics of the patients, including age, prostate-specific antigen (PSA) levels, malignant lesion location, Gleason grading groups (GGG), and pathological stage, were obtained from pathology reports. The clinicopathological characteristics of patients are presented in Table 1.

MRI acquisition

All images were acquired using one of two 3.0T MRI scanners (Philips Ingenia CX, The Netherlands; United Imaging uMR 790, China) and a pelvic phased array coil. The imaging protocol included axial T2-weighted imaging (T2WI) and diffusion-weighted imaging (DWI). Apparent diffusion coefficient (ADC) maps were calculated inline by the scanner software using linear fitting based on a mono-exponential model. The scan parameters are detailed in Table 2.

MRI-based rectal artifact subjective evaluation

Two radiologists with 5–10 years of experience in prostate imaging, who were blinded to all clinical details and biopsy results, independently performed subjective evaluations of rectal artifacts. A senior radiologist with 15 years of experience in prostate imaging reviewed and resolved discrepancies in scoring between the two radiologists. To assess the varying levels of rectal artifacts, we employed a scoring method to evaluate their impact
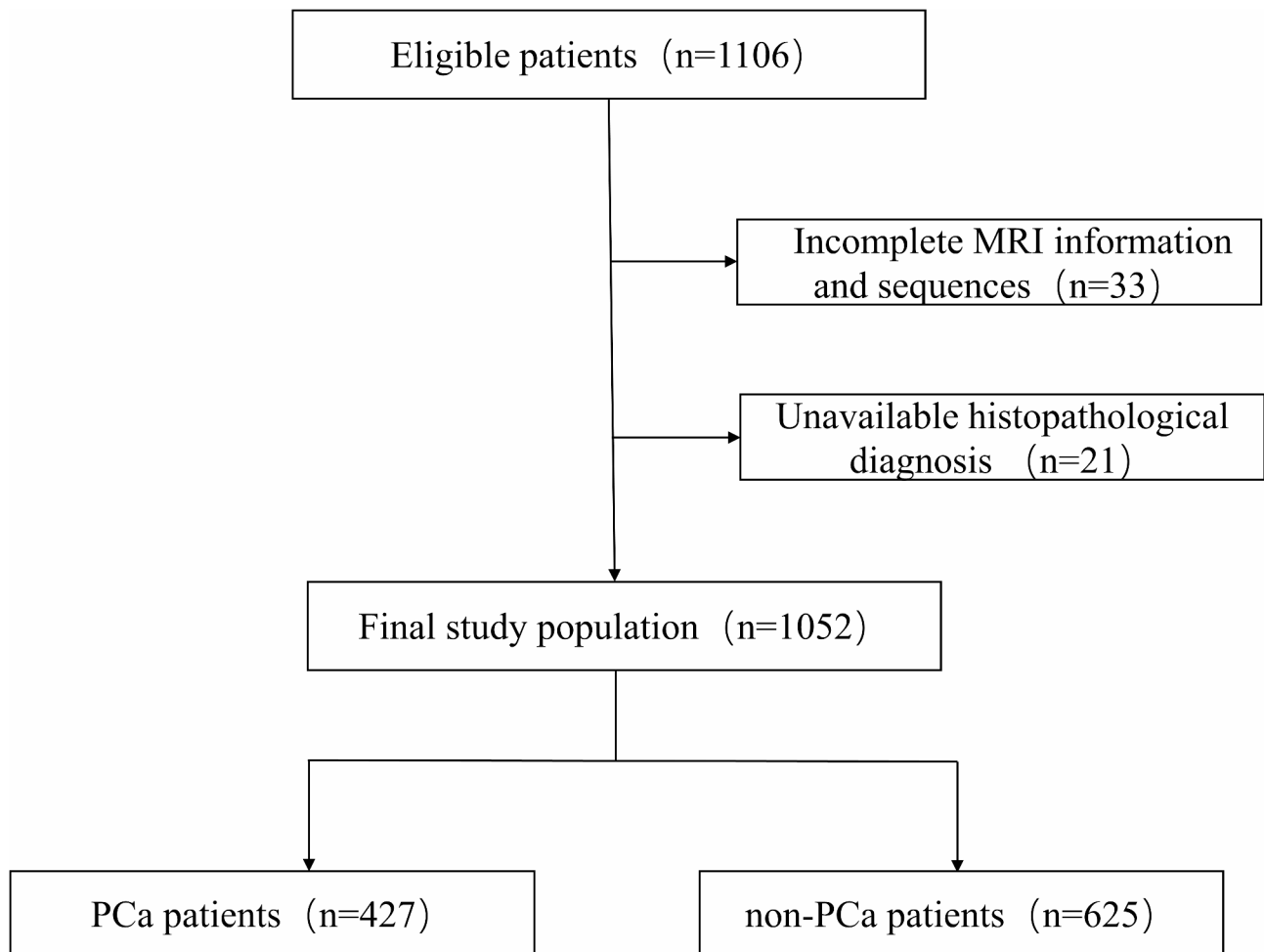
Fig. 1 Flow chart of participant inclusion and exclusion

on the prostate gland region. Images were scored using a 4-point Likert scale: Score 1 indicated no artifact; Score 2 represented mild artifact, when less than 50% of the peripheral zone (PZ) next to the rectum was involved; Score 3 indicated moderate artifact, when 51–100% of the PZ was affected without involving the transition zone (TZ); and Score 4 indicated severe artifact, when the artifact extended into the TZ [20]. For cases where there are obvious gas artifacts on T2WI but not on DWI, or where artifacts are significant on DWI but less pronounced on T2WI, the final degree of artifact was determined based on the sequence that demonstrated the most significant interference with prostate visualization. This approach ensures that the most clinically relevant artifacts are accounted for in the grading process. In situations where discrepancies between T2WI and DWI artifact severity were observed, consensus was reached through a joint review by experienced radiologists.

For patients in whom the PZ was not visualized at all due to severe artifacts, the artifact was classified as "severe". This classification reflects the maximum

potential interference with diagnostic evaluation, recognizing that complete obscuration of the PZ significantly impairs accurate lesion detection and characterization. Examples of varying levels of rectal artifacts are shown in Fig. 2.

MRI-based DL diagnostic model

To assess the impact of rectal artifacts on the DL diagnostic model, we employed the 3D-nnUNet [21] as the network framework to construct the prostate MRI diagnostic model in this study. The model was trained using annotated data from the public dataset "Prostate Imaging: Cancer AI (Version 1.1)" [22]. This dataset comprises 1500 anonymous prostate MRI scans from 1476 patients across three centers (Radboud University Medical Center, University Medical Center Groningen, and Ziekenhuis Groep Twente).

Initially, a prostate gland segmentation model was employed to generate segmentation masks for the central and peripheral glands using T2WI images. Subsequently, a prostate lesion detection model was implemented,

Wang *et al. BMC Medical Imaging*   (2025) 25:61

Page 4 of 11

Table 1 Clinicopathological characteristics of patients

Factors [a]	PCa (n = 427)	non-PCa (n = 625)	P value
Age (y)	72 ± 8	69 ± 8	<0.001
PSA level (ng/ml)	32.18 (9.48, 78.37)	6.46 (3.16, 11.72)	<0.001
Malignant lesion			
Diameter (mm)	23.0 (20.2, 26.5)	NA	NA
Location			
PZ	281 (65.8)	NA	NA
TZ	146 (34.2)	NA	NA
GGG			
GGG 1(GS 3 + 3)	54 (12.6)	NA	NA
GGG 2(GS 3 + 4)	50 (11.7)	NA	NA
GGG 3(GS 4 + 3)	76 (17.8)	NA	NA
GGG 4(GS=8)	124 (29.1)	NA	NA
GGG 5(GS>8)	123 (28.8)	NA	NA
TNM			
T1	49 (11.5)	NA	NA
T2	190 (44.5)	NA	NA
T3	136 (31.8)	NA	NA
T4	52 (12.2)	NA	NA

[a] Data are presented as n (%) or median (interquartile range). PSA prostate-specific antigen, PZ peripheral zone, TZ transitional zone, GGG Gleason grading group, TNM tumor node metastasis, NA not applicable

utilizing segmented masks of the central and peripheral glands along with T2WI, DWI, and ADC images as inputs. The model generated a lesion confidence map representing the confidence level of lesions in different prostate regions. Using the lesion confidence map, corresponding lesion candidate regions and their probabilities of PCa detection were generated. In previous study, this model achieved an accuracy of 0.87 on a quality-filtered open-source dataset, and it performed well in multicenter external validation [23].

Lesion annotation and histopathologic matching

A radiologist with more than 15 years of experience in prostate MRI reviewed the images. When MRI suspected PCa and the target lesion for biopsies was in the same sector, the slice of the image containing the largest lesion extent was selected and marked. The lesions were classified as originating from either the TZ or PZ. Lesions covering both PZ and TZ were termed TZ lesions if they covered at least 70% of the TZ [24]; otherwise, they were considered PZ lesions.

Diagnostic assessment

Six radiologists, consisting of two junior radiologists (R1-2) with 5 years of experience, two intermediate-level radiologists (R3-4) with 10 years of experience, and two senior radiologists (R5-6) with 15 years of experience, were recruited. They were blinded to the clinical information and analyzed the prostate MR images independently. Each identified lesion was assigned a score according to the PI-RADS version 2.1 [6], based on T2WI and DWI.

To compare the differences in the effect of rectal susceptibility artifacts on the diagnostic performance of radiologists and DL, we initially assessed the impact of rectal susceptibility artifacts on both radiologists' diagnostic performance and that of the DL model at the patient level. Furthermore, we conducted a subgroup analysis to compare the performance of radiologists and the DL model across the absence of artifacts and varying grades of artifacts, as well as among different lesion zones.

Statistical analysis

Statistical analysis was performed using R version 4.10 (R Foundation for Statistical Computing, Vienna, Austria; https://www.R-project.org/). The normality of data distribution was assessed using the Kolmogorov-Smirnov test for a single sample. For continuous variables following a normal distribution, differences were analyzed using an independent sample t-test, while the Mann-Whitney U test was employed for non-normally distributed continuous variables. Comparison of differences in ordinal data was evaluated using the Chi-square Test. The

Table 2 Magnetic resonance sequence parameters

Parameter	Ingenia CX		uMR 790	
	T2WI	DWI	T2WI	DWI
Sequence	TSE	SE-EPI	FSE	EPI
Echo time(ms)	110	58	130	52
Repetition time(ms)	4000	6259	4210	2329
Field of view(mm^2)	250×250	250×250	432×432	168×100
Scan matrix	416×284	100×82	288×288	112×67
PE Dir.	Left → Right	Anterior → Posterior	Right →Left	Anterior → Posterior
Layer thickness(mm)	4	4	3	4
Layer spacing(mm)	4.4	4.4	3.3	4.4
Bandwidth (Hz/px)	141	3152	250	1500
bvalue(s/mm^2)	NA	0,800,1000,1200	NA	5,080,010,001,300
Acquisition time(min)	1:04	2:49	0:55	2:29

TSE turbo spin echo, FSE fast spin echo, EPI echo planar imaging, SE-EPI spin echo echo-planar imaging, PE Dir. phase encoding direction, NA not applicable
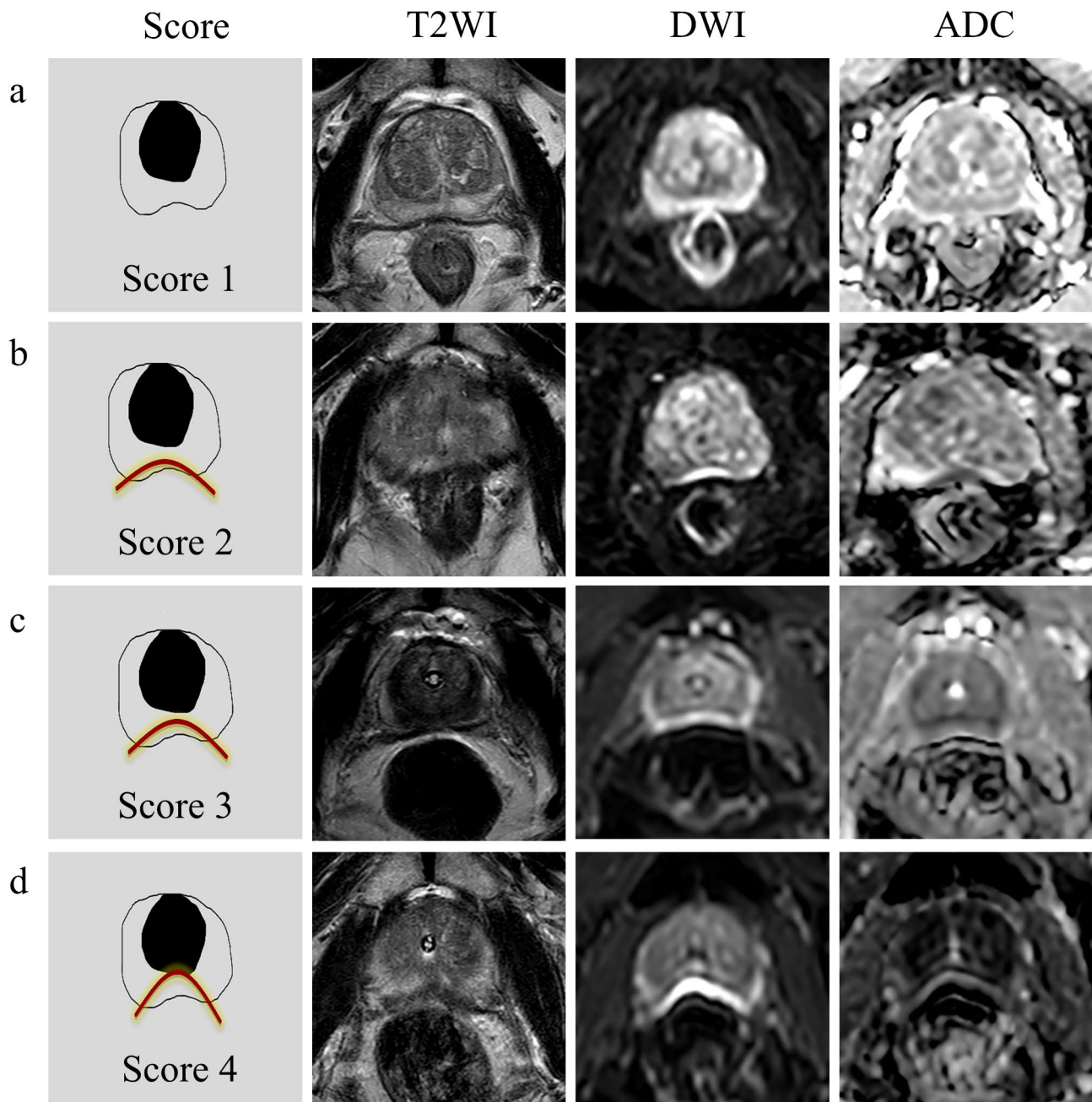
Fig. 2 Examples of MRI-based rectal artifact subjective evaluation. Selected images used to show Score 1 (First line **a**): no artifact; Score 2 (Second line **b**): Mild artifact when < 50% of the PZ next to the rectum was involved; Score 3 (Third line **c**): moderate artifact when 51–100% of the PZ was affected without involving the TZ; and Score 4 (Fourth line **d**): Severe when the artifact extends into the TZ

inter-radiologist agreement of rectal artifact evaluation was assessed using Cohen's weighted kappa, with the following interpretation: 0.01 to 0.20 indicating poor consistency; 0.21 to 0.40 indicating fair consistency; 0.41 to 0.60 indicating moderate consistency; 0.61 to 0.80 indicating substantial consistency; and 0.81 to 0.99 indicating almost perfect consistency. The performance of DL and radiologists was evaluated using the area under the receiver operating characteristic curve (AUC) and the

area under the multi-reader multi-case receiver operating characteristic curve, respectively. Differences in AUCs were compared using DeLong's test. A significance level of $P < 0.05$ was considered statistically significant.

Results

Patient characteristics

A total of 1052 patients participated in this study, with 427 patients in the PCa group and 625 patients in

Wang *et al. BMC Medical Imaging*   (2025) 25:61

Page 6 of 11

Table 3 Rectal artifacts subjective evaluation

Artifact grade [a]	All cases(n = 1052)	non-PCa (n = 625)	PCa (n = 427)
no artifact (Score 1)	647 (61.5)	372 (59.5)	275 (64.4)
mild artifact (Score 2)	177 (16.8)	89 (14.2)	88 (20.6)
moderate artifact (Score 3)	132 (12.6)	93 (14.9)	39 (9.1)
severe artifact (Score 4)	96 (9.1)	71 (11.4)	25 (5.9)

[a] Data in parentheses are percentages

the non-PCa group. Patients in the PCa group (mean age: 72 ± 8 years, median PSA: 32.18 ng/ml [IQR 9.48, 78.37]) differed significantly from those in the non-PCa group (mean age: 69 ± 8 years, median PSA: 6.46 ng/ml [IQR 3.16, 11.72]) in terms of age and PSA levels (both $P < 0.001$). Regarding the PCa group, the diameter of cancer lesions without rectal artifacts was approximately 23.2 mm (IQR 20.6, 26.5). In the presence of rectal artifacts, the diameter was slightly lower, measuring around 22.5 mm (IQR 19.4, 26.6). When stratified by artifact severity (mild, moderate, and severe), cancer lesion diameters were similar: approximately 22.5 mm (IQR 19.4, 28.1) for mild artifacts, 22.3 mm (IQR 19.1, 26.3) for moderate artifacts, and 22.9 mm (IQR 19.7, 27.9) for severe artifacts, with no significant differences detected ($0.43 \le P \le 0.81$). Table 1 presents baseline epidemiological and clinicopathological patient characteristics.

Rectal artifact subjective evaluation

Inter-radiologist agreement on rectal artifact evaluation was perfect ($\kappa = 0.822$). Of the total, 61.5% (647/1052) of patients showed no rectal artifacts, while 16.8% (177/1052), 12.6% (132/1052), and 9.1% (96/1052)

exhibited mild, moderate, and severe rectal artifacts, respectively. The differences between the PCa group and non-PCa group in different grades of rectal artifacts were statistically significant ($P < 0.001$). Details are presented in Table 3.

Diagnostic accuracy assessment

Figure 3; Table 4 illustrate significant variations in diagnostic efficacy among radiologists based on subjective evaluation. Junior radiologists (R1: AUC 0.70, 95% CI: 0.67–0.72; R2: AUC 0.72, 95% CI: 0.69–0.75) showed relatively lower diagnostic efficacy, while intermediate-level radiologists (R3: AUC 0.77, 95% CI: 0.75–0.80; R4: AUC 0.78, 95% CI: 0.75–0.80) demonstrated better performance. Senior radiologists (R5: AUC 0.79, 95% CI: 0.76–0.81; R6: AUC 0.82, 95% CI: 0.79–0.84) exhibited the highest diagnostic performance. The average AUC across all radiologists was 0.76 (95% CI: 0.75–0.78). In terms of DL performance, the AUC was 0.71 (95% CI: 0.68–0.74).

As shown in Table 5; Fig. 4, the diagnostic performance without artifacts for junior radiologists was much higher than that with artifacts (R1: AUC 0.73, 95%CI: 0.69–0.76 versus 0.64, 95%CI: 0.59–0.69; $P = 0.01$; R2: AUC 0.74, 95%CI: 0.71–0.78 versus 0.67, 95%CI: 0.63–0.72; $P = 0.03$). The difference was not statistically significant ($0.10 \le P \le 0.23$) between the diagnostic performance of the intermediate-level (R3: AUC 0.80, 95%CI: 0.76–0.82 versus 0.74, 95%CI: 0.70–0.79; $P = 0.10$; R4: AUC 0.80, 95%CI: 0.76–0.82 versus 0.75, 95%CI: 0.71–0.80; $P = 0.13$) and senior (R5: AUC 0.81, 95%CI: 0.77–0.83 versus 0.77, 95%CI: 0.72–0.81; $P = 0.23$; R6: AUC 0.83, 95%CI: 0.81–0.86 versus 0.79, 95%CI: 0.74–0.83; $P = 0.10$) radiologists. In the realm of DL, it was noted that the
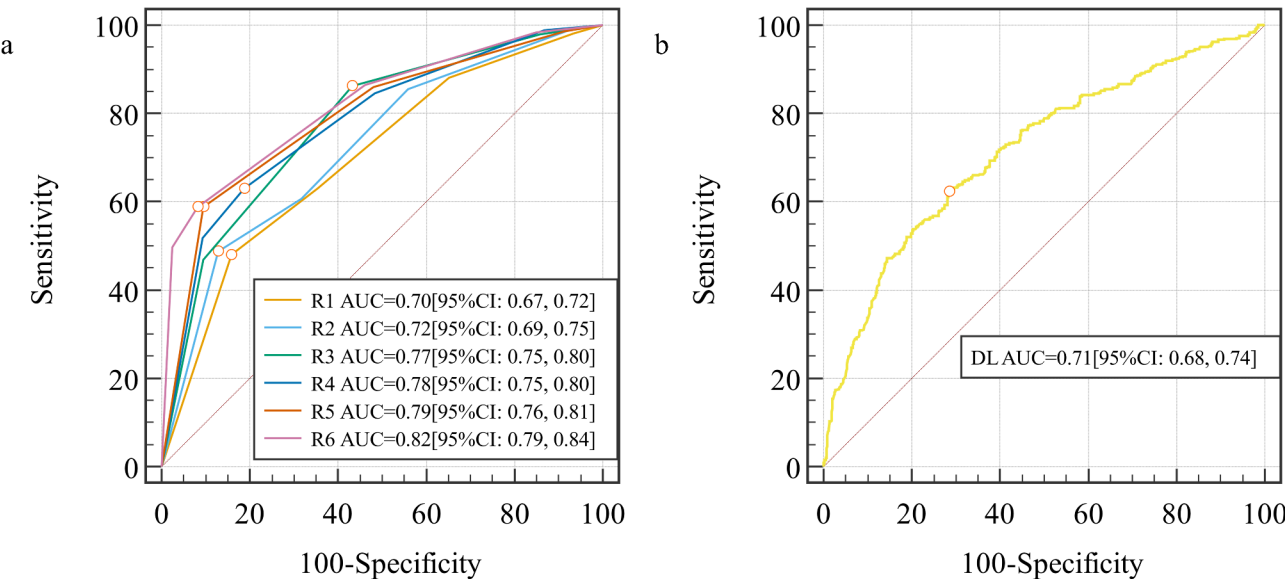

Fig. 3 Overall diagnostic performance of radiologists (**a**) and DL (**b**). Data inside parentheses represent 95% confidence intervals

Table 4 PI-RADS scores of radiologists with different experience [a]

PI-RADS [a]	PCa (n=427)						non-PCa (n=625)					
	Junior radiologists		Intermediate-level radiologists		Senior radiologists		Junior radiologists		Intermediate-level radiologists		Senior radiologists	
	R1	R2	R3	R4	R5	R6	R1	R2	R3	R4	R5	R6
1	8 (1.9)	5 (1.2)	9 (2.1)	5 (1.2)	6 (1.4)	7 (1.7)	41 (6.6)	48 (7.7)	89 (14.3)	83 (13.3)	54 (8.6)	94 (15.0)
2	43 (10.1)	57 (13.4)	50 (11.7)	61 (14.3)	54 (12.6)	51 (11.9)	177 (28.3)	228 (36.5)	266 (42.6)	240 (38.4)	271 (43.4)	243 (38.9)
3	107 (25.1)	106 (24.8)	117 (27.4)	92 (21.5)	69 (16.2)	118 (27.6)	186 (29.8)	151 (24.1)	147 (23.5)	185 (29.6)	144 (23.1)	237 (37.9)
4	64 (14.9)	51 (11.9)	51 (11.9)	48 (11.2)	47 (11.0)	39 (9.1)	122 (19.5)	118 (18.9)	64 (10.2)	59 (9.4)	97 (15.5)	36 (5.8)
5	205 (48.0)	208 (48.7)	200 (46.9)	221 (51.8)	251 (58.8)	212 (49.7)	99 (15.8)	80 (12.8)	59 (9.4)	58 (9.3)	59 (9.4)	15 (2.4)

[a] Data in parentheses are percentages

diagnostic performance without artifacts was significantly higher compared to the performance with artifacts (DL: AUC 0.77, 95%CI: 0.73–0.80, versus 0.61, 95%CI: 0.57–0.66; $P < 0.001$). There were no statistically significant differences in the AUC with different grades of rectal artifacts for all radiologists ($0.08 \leq P \leq 0.90$) and DL ($0.12 \leq P \leq 0.96$).

As illustrated in Table 6, there were no statistically significant differences in the AUC with artifacts and without artifacts for all radiologists in the PZ and TZ ($0.08 \leq P \leq 0.98$). However, for DL, the AUC without artifacts was notably higher than those with artifacts in both PZ and TZ (DL$_{PZ}$: AUC 0.78, 95%CI: 0.72–0.83, versus 0.61, 95%CI: 0.53–0.69; $P = 0.003$; DL$_{TZ}$: AUC 0.73, 95%CI: 0.67–0.77, versus 0.59, 95%CI: 0.52–0.65; $P = 0.011$).

Discussion

This study assessed the effects of rectal susceptibility artifacts on both subjective evaluation and DL in PCa diagnosis. The findings revealed a notable impact of rectal artifacts on diagnostic accuracy, particularly evident among junior radiologists with limited experience and the DL algorithm.

Previous studies have explored the impact of rectal artifacts on the subjective evaluation of PCa. The study by Antunes [20] indicated that, due to the presence of rectal artifacts, 36.8% of malignant lesions were underestimated. Similarly, Caglic et al. [25] found that rectal distension had a significant negative impact on the quality of T2WI and DWI images. Other relevant studies [26–28] have also indicated that reducing rectal artifacts can significantly improve diagnostic accuracy for interpreters. However, these aforementioned studies did not include reader experience as an analyzed factor when assessing the impact of rectal artifacts on diagnostic performance. Our study found that the impact of rectal artifacts on diagnostic accuracy varied among radiologists with different levels of experience. This discrepancy is possibly due to the fact that intermediate-level and senior radiologists possess richer clinical experience, familiarity with various artifacts, and the ability to distinguish between true lesions and artifacts, thereby enabling them to better interpret prostate MRI images. In contrast, junior radiologists lack sufficient experience, which may lead them to be influenced by rectal artifacts, thereby affecting their diagnostic accuracy.

Our study found no significant differences in diagnostic performance between radiologists when assessing various scores of rectal artifacts. This observation may be attributed to the inherent characteristics of PCa and the anatomy of the prostate. Approximately 70–75% of PCa cases occur in the PZ [6], which is where the rectum is situated behind the prostate. Moreover, the most

Table 5 Impact of rectal artifacts with different grades on subjective evaluation and DL

	AUC		P value	AUC			P value		
	no artifact	Artifact	no artifact vs. artifact	mild	moderate	severe	mild vs. moderate	moderate vs. severe	mild vs. severe
R1	0.73(0.69–0.76)	0.64(0.59–0.69)	0.01	0.71(0.61–0.80)	0.66(0.58–0.73)	0.58(0.49–0.67)	0.52	0.25	0.13
R2	0.74(0.71–0.78)	0.67(0.63–0.72)	0.03	0.76(0.66–0.84)	0.67(0.59–0.73)	0.63(0.54–0.71)	0.15	0.6	0.09
R3	0.80(0.76–0.82)	0.74(0.70–0.79)	0.1	0.82(0.73–0.89)	0.74(0.66–0.81)	0.73(0.66–0.80)	0.18	0.78	0.08
R4	0.80(0.76–0.82)	0.75(0.71–0.80)	0.13	0.77(0.70–0.83)	0.73(0.64–0.82)	0.72(0.65–0.79)	0.64	0.85	0.44
R5	0.81(0.77–0.83)	0.77(0.72–0.81)	0.23	0.82(0.73–0.89)	0.75(0.67–0.82)	0.74(0.67–0.80)	0.3	0.9	0.21
R6	0.83(0.81–0.86)	0.79(0.74–0.83)	0.1	0.84(0.75–0.90)	0.79(0.71–0.85)	0.77(0.70–0.83)	0.43	0.74	0.25
DL	0.77(0.73–0.80)	0.61(0.57–0.66)	< 0.001	0.65(0.58–0.72)	0.66(0.57–0.74)	0.53(0.42–0.63)	0.96	0.14	0.12

Data inside parentheses represent 95% confidence intervals. AUC area under the receiver operating characteristic curve

common location for the accumulation of rectal artifacts is also in the PZ [25, 29]. In our study, rectal artifact scoring was standardized based on the extent of involvement, following the established method [20]. Notably, rectal artifact scores of 2–4 predominantly affected the PZ, while artifacts originating from the rectal region and radiating outward mainly impacted the PZ, with the central zone experiencing lighter accumulation of rectal artifacts. This consistency in artifact distribution across the PZ may explain the lack of statistically significant differences in diagnostic performance across different scores of rectal artifacts.

Similarly to subjective evaluation, the diagnostic performance of DL also indicates that rectal artifacts interfere with its effectiveness, resulting in reduced diagnostic efficacy, and there is no notable variance in DL diagnostic performance across patients with varying degrees of rectal artifacts. This emphasizes the need for cautious interpretation of diagnostic results provided by DL models for patients with poor image quality in clinical practice. However, the impact of rectal artifacts on DL is more complex and challenging to explain. Prior studies [30] have shown that DL is more sensitive to small perturbations in images compared to human visual evaluation. Even slight changes in angle, the addition of local noise, or the alteration of a single pixel can lead to misdiagnosis or missed diagnosis in well-trained models.

Diagnostic errors caused by rectal susceptibility artifacts can have significant implications for patient treatment strategies, particularly when distinguishing between low-risk and high-risk PCa. For low-risk PCa, misdiagnosis or underestimation due to artifacts may lead to unnecessary interventions, such as overtreatment with surgery or radiation, which can negatively impact the patient's quality of life. Conversely, in high-risk PCa, missed or inaccurately characterized lesions can result in delayed or insufficient treatment, potentially leading to disease progression and poorer prognoses. Therefore, understanding and mitigating the influence of rectal artifacts is essential to optimizing patient outcomes and ensuring appropriate treatment stratification.

Our study suggested the importance of developing specific MRI scanning and diagnostic strategies to mitigate the impact of rectal artifacts in subsequent clinical practice. Currently, efforts to mitigate rectal artifacts in prostate imaging primarily focus on clinical strategies and imaging techniques. For instance, pre-MRI bowel preparation has been shown to effectively reduce the occurrence of rectal artifacts [12]. Furthermore, more advanced imaging technologies, such as reduced field of view imaging [19], parallel imaging, and segmented readout echo planar imaging (EPI) sequences [31], have also been demonstrated to reduce the severity of macroscopic rectal artifacts to some extent, thus reducing the impact of rectal artifacts on the diagnosis of PCa. However, since rectal artifacts cannot be completely eliminated at present, their influence on DL remains. DL models often struggle with generalization when exposed to imaging artifacts not present in the training data, leading to decreased diagnostic performance. To enhance robustness, future work should focus on training DL models with artifact-augmented datasets. By artificially introducing rectal artifacts during the training phase, models can learn to identify and compensate for these distortions, thereby improving their reliability in real-world clinical settings. Additionally, incorporating advanced techniques may further bolster the resilience of DL models against artifacts.

Recent advancements in DL have introduced techniques such as domain adaptation and adversarial training to address the issue of imaging artifacts [32]. Domain adaptation methods allow DL models to generalize better across datasets with varying artifact characteristics by aligning feature distributions from different domains, thereby enhancing model robustness. Adversarial training, on the other hand, involves exposing the model to intentionally perturbed inputs during the training phase, enabling it to learn features that are less sensitive to artifacts and noise. Our recent study [23] highlighted the efficacy of adversarial training in enhancing model resilience against common imaging artifacts, thus reducing diagnostic errors. Despite these advancements, a certain gap
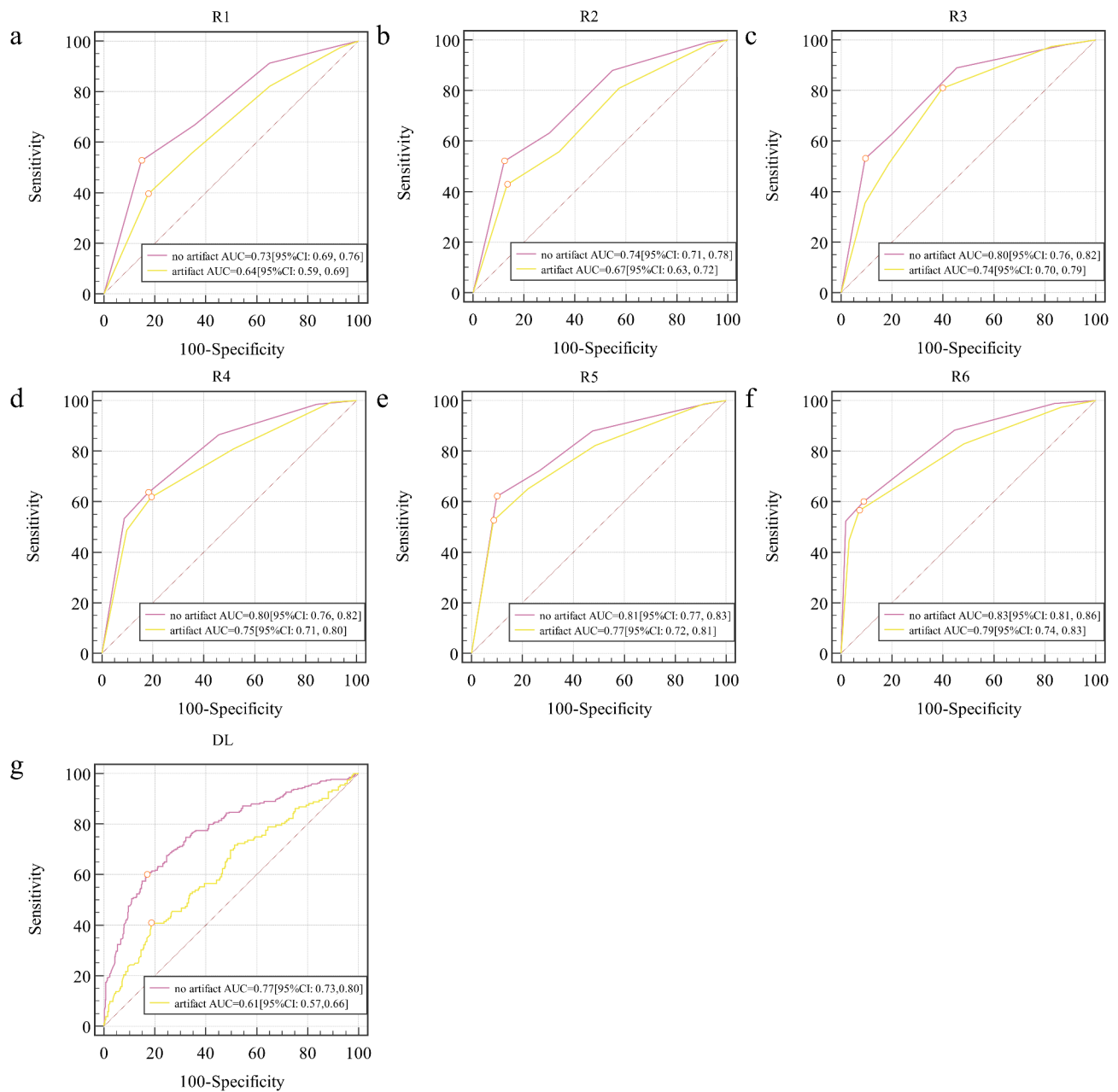
Fig. 4 Diagnostic performance of radiologists (**a-f**) and DL (**g**) in patients with artifact and without artifact. Data inside parentheses represent 95% confidence intervals

in diagnostic performance still exists between patients with and without rectal artifacts, even when using these enhanced DL techniques. This underscores the need for continuous development and validation of robust algorithms capable of effectively mitigating the impact of artifacts in clinical settings.

The present study is subject to several limitations. Firstly, its retrospective design incorporates data from only two medical centers, which may introduce selection bias and limit the generalizability of the findings. In our future studies, we plan to include more diverse populations across multiple centers and expand the number of readers. Additionally, we will conduct prospective validation to further substantiate the conclusions of this study. Currently, assessment of rectal artifact severity relies on subjective scoring, which introduces the possibility of evaluation bias. To overcome this limitation, future research should explore the development of automated scoring systems or DL-based artifact detection methods. Automated approaches can provide consistent, objective evaluations of artifact severity, reducing inter-observer variability and enhancing reproducibility.

Table 6 Impact of rectal artifacts with lesion location on subjective evaluation and DL

	PZ			TZ		
	AUC		*P* value	AUC		*P* value
	no artifact	artifact	no artifact vs. artifact	no artifact	artifact	no artifact vs. artifact
R1	0.64(0.57–0.70)	0.60(0.51–0.68)	0.43	0.66(0.60–0.71)	0.59(0.52–0.65)	0.23
R2	0.66(0.58–0.72)	0.61(0.52–0.69)	0.5	0.73(0.67–0.77)	0.64(0.57–0.70)	0.08
R3	0.75(0.69–0.81)	0.72(0.64–0.79)	0.49	0.72(0.66–0.76)	0.70(0.64–0.76)	0.93
R4	0.77(0.69–0.83)	0.71(0.65–0.77)	0.32	0.75(0.69–0.79)	0.68(0.62–0.74)	0.25
R5	0.76(0.67–0.82)	0.73(0.66–0.78)	0.62	0.75(0.69–0.79)	0.72(0.66–0.76)	0.63
R6	0.84(0.78–0.88)	0.76(0.68–0.83)	0.09	0.76(0.70–0.80)	0.75(0.69–0.79)	0.98
DL	0.78(0.72–0.83)	0.62(0.53–0.70)	0.003	0.73(0.67–0.77)	0.60(0.53–0.66)	0.011

Data inside parentheses represent 95% confidence intervals. PZ peripheral zone, TZ transitional zone, AUC area under the receiver operating characteristic curve

Implementing DL algorithms trained to recognize and quantify rectal artifacts could streamline the diagnostic process, minimize human error, and improve the accuracy of PCa detection in clinical settings. Finally, while this study primarily addresses the impact of artifacts on individual diagnosticians, including both radiologists and DL models, it does not delve into the potential for collaboration between human expertise and artificial intelligence. The integration of radiologists' experience with the computational power of DL could be a promising approach to mitigate the effects of artifacts on diagnostic accuracy. This collaborative approach, where human intuition and model predictions complement each other, remains an important avenue for future research to explore. In our subsequent research, we aim to investigate this collaboration in more detail to further enhance diagnostic outcomes.

Conclusions

In conclusion, rectal susceptibility artifacts significantly impact both subjective evaluation and DL-based models for PCa classification, with noticeable differences between the two methods.

Abbreviations

PCa	Prostate cancer
MRI	Magnetic resonance imaging
DL	Deep learning
DNNs	Deep neural networks
bpMRI	Biparametric Magnetic resonance imaging
PI-RADS	Prostate Imaging Reporting and Data System
GGG	Gleason grading groups
T2WI	T2-weighted imaging
DWI	Diffusion-weighted imaging
ADC	Apparent diffusion coefficient
PZ	Peripheral zone
TZ	Transitional zone
PSA	Prostate-specific antigen
TNM	Tumor node metastasis
TSE	Turbo spin echo
FSE	Fast spin echo
EPI	Echo planar imaging
SE-EPI	Spin echo echo-planar imaging
PE	Dir Phase encoding direction
IQR	Interquartile range
AUC	Area under the receiver operating characteristic curve

Declarations

Ethics approval and consent to participate
This retrospective study was approved by the Ethics Committee of Yichang Central People's Hospital Affiliated to the First Clinical Medical College of China Three Gorges University [Approve No: 2023-001-01] and the need for informed patient consent for inclusion was waived.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Radiology, The First College of Clinical Medical Science, China Three Gorges University, Yichang Central People's Hospital, No. 183, Yiling Avenue, Yichang 443000, China
[2]Institute of Medical Imaging, China Three Gorges University, No. 183, Yiling Avenue, Yichang 443000, China
[3]Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Southern Medical University, No. 106, Zhongshan Er Road, 510080 Guangzhou, China
[4]Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, No. 106, Zhongshan Er Road, 510080 Guangzhou, China

References
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49. https://doi.org/10.3322/caac.21660.

2. Hectors SJ, Cherny M, Yadav KK, Beksac AT, Thulasidass H, Lewis S, et al. Radiomics features measured with multiparametric magnetic resonance imaging predict prostate Cancer aggressiveness. J Urol. 2019;202:498–505. https://doi.org/10.1097/JU.0000000000000272.

3. Turkbey B, Brown AM, Sankineni S, Wood BJ, Pinto PA, Choyke PL. Multiparametric prostate magnetic resonance imaging in the evaluation of prostate cancer. CA Cancer J Clin. 2016;66:326–36. https://doi.org/10.3322/caac.21333.

4. Gomez SL, Washington SR, Cheng I, Huang FW, Cooperberg MR. Monitoring prostate Cancer incidence trends: value of multiple imputation and delay adjustment to discern disparities in Stage-specific trends. Eur Urol. 2021;79:42–3. https://doi.org/10.1016/j.eururo.2020.10.022.

5. Pecoraro M, Turkbey B, Purysko AS, Girometti R, Giannarini G, Villeirs G, et al. Diagnostic accuracy and observer agreement of the MRI prostate imaging for recurrence reporting assessment score. Radiology. 2022;304:342–50. https://doi.org/10.1148/radiol.212252.

6. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. Eur Urol. 2019;76:340–51. https://doi.org/10.1016/j.eururo.2019.02.033. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2.

7. O'Connor L, Wang A, Walker SM, Yerram N, Pinto PA, Turkbey B. Use of multiparametric magnetic resonance imaging (mpMRI) in localized prostate cancer. Expert Rev Med Devices. 2020;17:435–42. https://doi.org/10.1080/17434440.2020.1755257.

8. Brembilla G, Giganti F, Sidhu H, Imbriaco M, Mallett S, Stabile A, et al. Diagnostic accuracy of abbreviated Bi-Parametric MRI (a-bpMRI) for prostate Cancer detection and screening: A Multi-Reader study. Diagnostics (Basel). 2022;12. https://doi.org/10.3390/diagnostics12020231.

9. Mazaheri Y, Vargas HA, Nyman G, Akin O, Hricak H. Image artifacts on prostate diffusion-weighted magnetic resonance imaging: trade-offs at 1.5 Tesla and 3.0 Tesla. Acad Radiol. 2013;20:1041–7. https://doi.org/10.1016/j.acra.2013.04.005.

10. Rosenkrantz AB, Taneja SS. Radiologist, be aware: ten pitfalls that confound the interpretation of multiparametric prostate MRI. AJR Am J Roentgenol. 2014;202:109–20. https://doi.org/10.2214/AJR.13.10699.

11. Ahlawat S, Fayad LM. Diffusion weighted imaging demystified: the technique and potential clinical applications for soft tissue imaging. Skeletal Radiol. 2018;47:313–28. https://doi.org/10.1007/s00256-017-2822-3.

12. Plodeck V, Radosa CG, Hübner H, Baldus C, Borkowetz A, Thomas C, et al. Rectal gas-induced susceptibility artefacts on prostate diffusion-weighted MRI with Epi read-out at 3.0 T: does a preparatory micro-enema improve image quality? Abdom Radiol (NY). 2020;45:4244–51. https://doi.org/10.1007/s00261-020-02600-9.

13. Coskun M, Mehralivand S, Shih JH, Merino MJ, Wood BJ, Pinto PA, et al. Impact of bowel Preparation with fleet's™ enema on prostate MRI quality. Abdom Radiol (NY). 2020;45:4252–9. https://doi.org/10.1007/s00261-020-02487-6.

14. Hiremath A, Shiradkar R, Merisaari H, Prasanna P, Ettala O, Taimen P, et al. Test-retest repeatability of a deep learning architecture in detecting and segmenting clinically significant prostate cancer on apparent diffusion coefficient (ADC) maps. Eur Radiol. 2021;31:379–91. https://doi.org/10.1007/s00330-020-07065-4.

15. Winkel DJ, Tong A, Lou B, Kamen A, Comaniciu D, Disselhorst JA, et al. A novel deep learning based Computer-Aided diagnosis system improves the accuracy and efficiency of radiologists in reading biparametric magnetic resonance images of the prostate: results of a multireader, multicase study. Invest Radiol. 2021;56:605–13. https://doi.org/10.1097/RLI.0000000000000780.

16. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol. 2020;21:233–41. https://doi.org/10.1016/S1470-2045(19)30739-9.

17. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. Science. 2019;363:1287–9. https://doi.org/10.1126/science.aaw4399.

18. Heaven D. Why deep-learning AIs are so easy to fool. Nature. 2019;574:163–6. https://doi.org/10.1038/d41586-019-03013-5.

19. Hu L, Fu C, Song X, Grimm R, von Busch H, Benkert T, et al. Automated deep-learning system in the assessment of MRI-visible prostate cancer: comparison of advanced zoomed diffusion-weighted imaging and conventional technique. Cancer Imaging. 2023;23:6. https://doi.org/10.1186/s40644-023-00527-0.

20. Antunes N, Vas D, Sebastia C, Salvador R, Ribal MJ, Nicolau C. Susceptibility artifacts and PIRADS 3 lesions in prostatic MRI: how often is the dynamic contrast-enhance sequence necessary? Abdom Radiol (NY). 2021;46:3401–9. https://doi.org/10.1007/s00261-021-03011-0.

21. Isensee F, Jaeger PF, Kohl S, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18:203–11. https://doi.org/10.1038/s41592-020-01008-z.

22. Saha A, Twilt J, Bosma J, van Ginneken B, Yakar D, Elschot M et al. Artificial intelligence and radiologists at prostate Cancer detection in MRI: the PI-CAI challenge (Study Protocol). 2022. https://doi.org/10.5281/zenodo.6667655

23. Hu L, Guo X, Zhou D, Wang Z, Dai L, Li L, et al. Development and validation of a deep learning model to reduce the interference of rectal artifacts in MRI-based prostate Cancer diagnosis. Radiol Artif Intell. 2024;6:e230362. https://doi.org/10.1148/ryai.230362.

24. Hu L, Wei L, Wang S, Fu C, Benker T, Zhao J. Better lesion conspicuity translates into improved prostate cancer detection: comparison of non-parallel-transmission-zoomed-DWI with conventional-DWI. Abdom Radiol (NY). 2021;46:5659–68. https://doi.org/10.1007/s00261-021-03268-5.

25. Caglic I, Hansen NL, Slough RA, Patterson AJ, Barrett T. Evaluating the effect of rectal distension on prostate multiparametric MRI image quality. Eur J Radiol. 2017;90:174–80. https://doi.org/10.1016/j.ejrad.2017.02.029.

26. Arnoldner MA, Polanec SH, Lazar M, Noori KS, Clauser P, Potsch N, et al. Rectal Preparation significantly improves prostate imaging quality: assessment of the PI-QUAL score with visual grading characteristics. Eur J Radiol. 2022;147:110–45. https://doi.org/10.1016/j.ejrad.2021.110145.

27. Huang YH, Ozutemiz C, Rubin N, Schat R, Metzger GJ, Spilseth B. Impact of 18-French rectal tube placement on image quality of multiparametric prostate MRI. AJR Am J Roentgenol. 2021;217:919–20. https://doi.org/10.2214/AJR.21.25732.

28. Reischauer C, Cancelli T, Malekzadeh S, Froehlich JM, Thoeny HC. How to improve image quality of DWI of the prostate-enema or catheter preparation? Eur Radiol. 2021;31:6708–16. https://doi.org/10.1007/s00330-021-07842-9.

29. Dietrich O, Reiser MF, Schoenberg SO. Artifacts in 3-T MRI: physical background and reduction strategies. Eur J Radiol. 2008;65:29–35. https://doi.org/10.1016/j.ejrad.2007.11.005.

30. Hirano H, Minagi A, Takemoto K. Universal adversarial attacks on deep neural networks for medical image classification. BMC Med Imaging. 2021;21:9. https://doi.org/10.1186/s12880-020-00530-y.

31. Li L, Wang L, Deng M, Liu H, Cai J, Sah VK, et al. Feasibility study of 3-T DWI of the prostate: Readout-Segmented versus Single-Shot Echo-Planar imaging. AJR Am J Roentgenol. 2015;205:70–6. https://doi.org/10.2214/AJR.14.13489.

32. Hu L, Zhou D, Xu J, Lu C, Han C, Shi Z, et al. Protecting prostate Cancer classification from rectal artifacts via targeted adversarial training. IEEE J Biomed Health Inf. 2024;28:3997–4009. https://doi.org/10.1109/JBHI.2024.3384970.

Publisher's note