

ORIGINAL ARTICLE / *Cancer imaging*

Computer-aided diagnosis system for characterizing ISUP grade ≥ 2 prostate cancers at multiparametric MRI: A cross-vendor evaluation



S. Transin^a, R. Souchon^b, C. Gonindard-Melodelima^c,
R. de Rozario^{b,d}, P. Walker^{e,f}, M. Funes de la Vega^g,
R. Loffroy^{a,e}, L. Cormier^h, O. Rouvière^{b,i,j,*}

^a Department of Radiology, University Hospital François-Mitterrand, 21000 Dijon, France

^b Inserm, U1032, LabTau, 69003 Lyon, France

^c CNRS, UMR 5553, Laboratoire d'Ecologie Alpine, Université Grenoble-Alpes, 38041 Grenoble, France

^d Département Informatique, Faculté des Sciences et Technologies, Université Lyon 1, 69000 Lyon, France

^e Medical Imaging Group, Laboratory of Electronics, Computer Science and Imaging (Le2I), CNRS 6306, University of Burgundy, 21000 Dijon, France

^f Department of MR Spectroscopy, University Hospital François-Mitterrand, 21000 Dijon, France

^g Department of Pathology, University Hospital François-Mitterrand, 21000 Dijon, France

^h Department of Urology, University Hospital François-Mitterrand, 21000 Dijon, France

ⁱ Department of Urinary and Vascular Radiology, Hôpital Édouard-Herriot, Hospices Civils de Lyon, 69437 Lyon, France

^j Université de Lyon, Université Lyon 1, faculté de médecine Lyon Est, Lyon, France

KEYWORDS

Male;
Prostatic neoplasms;
Magnetic resonance
imaging (MRI);
Diagnosis;
Computer-assisted
diagnosis

Abstract

Purpose: To assess the performance of a computer-aided diagnosis (CADx) system trained at characterizing International Society of Urological Pathology (ISUP) grade ≥ 2 peripheral zone (PZ) prostate cancers on multiparametric magnetic resonance imaging (mpMRI) examinations from a different institution and acquired on different scanners than those used for the training database.

Patients and methods: Preoperative mpMRIs of 74 men (median age, 65.7 years) treated by prostatectomy between 2014 and 2017 were retrospectively selected. One radiologist

* Corresponding author at: Service d'imagerie, pavillon B, hôpital Edouard-Herriot, 5, place d'Arsonval, 69437 Lyon, France.
E-mail address: olivier.rouviere@netcourrier.com (O. Rouvière).

outlined suspicious lesions and scored them using Prostate Imaging-Reporting and Data System version 2 (PI-RADSv2); their CADx score was calculated using a classifier trained on an independent database of 106 patients treated by prostatectomy in another institution. The lesions' nature was assessed by comparison with prostatectomy whole-mounts. Diagnostic accuracy was estimated with areas under receiver operating characteristic curves (AUCs). Sensitivity and specificity were calculated using a CADx threshold (≥ 0.21) that yielded 95% sensitivity in the training database, and a PI-RADSv2 ≥ 3 threshold.

Results: A total of 127 lesions (PZ, $n = 104$; transition zone [TZ], $n = 23$) were described. In PZ, CADx and PI-RADSv2 scores had similar AUCs for characterizing ISUP grade ≥ 2 cancers (0.78 [95% confidence interval (CI): 0.69–0.87] vs. 0.74 [95%CI: 0.62–0.82], respectively) ($P = 0.59$). Sensitivity and specificity were respectively 89% (95%CI: 82–97%) and 42% (95%CI: 26–58%) for the CADx score, and 97% (95%CI: 93–100%) and 37% (95%CI: 22–52%) for the PI-RADSv2 score. In TZ, both scores showed poor specificity.

Conclusion: In this external cohort, the CADx and PI-RADSv2 scores showed similar performances in characterizing ISUP grade ≥ 2 cancers.

© 2019 Société française de radiologie. Published by Elsevier Masson SAS. All rights reserved.

Multiparametric magnetic resonance imaging (mpMRI) has high sensitivity for detecting prostate cancers with an International Society of Urological Pathology (ISUP) grade ≥ 2 [1–3]. Thus, targeting suspicious lesions seen on mpMRI improves the detection of these cancers as compared to the classical diagnostic pathway that uses systematically distributed biopsies [4–7]. As a result, mpMRI is increasingly obtained before biopsy [8,9].

However, mpMRI interpretation needs expertise and shows moderate inter-reader reproducibility, even with the use of the Prostate Imaging-Reporting and Data System version 2 (PI-RADSv2) [10–12]. In order to improve mpMRI interpretation, many research groups have developed computer-aided detection and diagnosis (CAD) systems that combine various image features [13–15]. There are two types of CAD systems. Computer-aided detection (CADe) systems provide probability maps highlighting regions of the gland that may contain cancer. They aim at improving both the detection (sensitivity) and the characterization (specificity) of suspicious areas. Computer-aided diagnosis (CADx) systems, only provide a probability score for disease in a region of interest (ROI) delineated by the radiologist. Some CAD systems [16–18] have been shown to improve human interpretation in so-called 'internal' test populations, (*i.e.*, on patients imaged in the same institution and on the same MRI scanners). Evaluation in 'internal' test population is optimistically biased [19] because of overfitting [20] and because intensity and texture features derived from MR images are prone to substantial variability across institutions and scanners. The use of quantitative MRI [21] is expected to reduce variability but some residual variability may exist [22–26]. Only a few CAD systems gave good results when tested on datasets from MRI scanners that were not those used for training [27–29]; and yet, in these cases, test populations were imaged on MRI scanners from the same institutions, using similar imaging parameters. To our knowledge, only one CADe system has been evaluated on a true

'external' database (*i.e.*, on data coming from completely different institutions) [30]. However, its impact on human interpretation was moderate, with only a trend towards better sensitivity in detecting index lesions in the transition zone (TZ), for less-experienced readers.

Developing a CADe is difficult because it needs automatic segmentation between the peripheral zone (PZ) and TZ, and robust co-registration between the pulse sequences despite image distortion or prostate motion. A CADx fulfils a less ambitious goal but provides less noisy quantifications and is not limited by segmentation or co-registration issues.

We previously trained a CADx system to characterize ISUP grade ≥ 2 cancers on mpMRIs from two different manufacturers obtained at a single institution in patients treated by radical prostatectomy [29]. It outperformed human reading when tested in an independent 'internal' cohort of patients referred for pre-biopsy mpMRI [31].

The purpose of this study was to evaluate the performances of the CADx system in characterizing ISUP grade ≥ 2 cancers in an 'external' dataset obtained from a different institution and different MRI scanners than the two used during the training phase.

Materials and methods

CADx training database

The CADx was trained to characterize ISUP grade ≥ 2 cancers in PZ using a database of 106 mpMRIs obtained before prostatectomy at Institution 1 (Hospices Civils de Lyon). All patients gave written informed consent and signed the institutional review board-approved inclusion form. mpMRIs included T2-weighted, diffusion-weighted and dynamic contrast-enhanced (DCE) imaging. They were obtained on one of three 3T imagers (MR 750[®], General

Electric Healthcare; Achieva X-series[®] and Ingenia[®], Philips Healthcare). Suspicious lesions were delineated by two radiologists and their nature was assessed after comparison with prostatectomy whole-mounts. The model providing the best discrimination of ISUP grade ≥ 2 cancers in PZ combined the 10th percentile of the ADC distribution (ADC_{10th}) and the time to the peak of enhancement (TTP). The CADx score providing 95% sensitivity for detecting ISUP grade ≥ 2 cancers in the training database was 0.21. Using this score as threshold, the specificity was 57% (95%CI: 24–69) [29,31].

Test database

We retrospectively selected the patients who underwent radical prostatectomy in the Department of Urology of Institution 2 (Centre hospitalo-universitaire Dijon Bourgogne) between January 2014 and December 2017, and who had undergone preoperative prostate mpMRI in the Departments of Radiology of Institution 2 or of Institution 3 (Centre de lutte contre le cancer Georges-François-Leclerc), these two departments sharing the same MRI scanners and the same Picture Archiving and Communication System (PACS). There were no exclusion criteria based on the imaging field strength nor on MR image parameters, provided the protocol included T2-weighted, diffusion-weighted and DCE imaging. The retrospective analysis of the test database was approved by the Ethics Committee of the Hospices Civils de Lyon; all patients received a letter detailing the study and offering them the opportunity to withdraw. No patient has withdrawn.

Three different MRI scanners from the same manufacturer (Siemens Healthineers) were used during the study period: Magnetom Aera[®] (1.5 T), Magnetom Trio-TIM[®] (3 T, from September 2013 to August 2016) and Magnetom Skyra[®] (3 T, from August 2016 to December 2017). An eight-channel (Aera[®] and Trio-TIM[®]) or 18-channel (Skyra[®]) pelvic phased-array coil was used without any endorectal coil. There was a large heterogeneity in the image parameters used (Table 1). For DCE imaging, an intravenous injection of 0.2 mL/kg of gadoterate meglumine (Dotarem[®], Guerbet) or gadobenate dimeglumine (Multihance[®], Bracco) was performed at 5 mL/s [32].

Image interpretation and calculation of the CAD scores

MR images of the test database were retrieved from the PACS of Institution 2 and anonymized. They were reviewed by an urologist (O.R.) with 20 years of experience in prostate imaging, blinded to clinical and pathological data. The radiologist noted all visible lesions whatever their degree of suspicion for malignancy. In PZ, all lesions with low-signal intensity at T2-weighted imaging and/or low-signal intensity on apparent diffusion coefficient (ADC) maps, and/or focal early or intense enhancement at DCE imaging were taken into consideration. In TZ, only homogeneous, low-signal intensity areas at T2-weighted imaging, with ill-defined margins, no visible capsule, and no cystic component were noted [33].

For each lesion, the radiologist assessed its location (PZ/TZ), its PI-RADSv2 score and the presence of post-biopsy bleeding artefact (none, mild, moderate, marked).

Because high *b*-value (>1400 s/mm²) images were not available, PI-RADSv2 assessment for diffusion-weighted imaging was made only on ADC maps (1: No abnormality on ADC; 2: Indistinct hypointense on ADC; 3: Focal mildly/moderately hypointense on ADC; 4: Focal markedly hypointense on ADC < 1.5 cm; 5: Focal markedly hypointense on ADC ≥ 1.5 cm).

Finally, the radiologist delineated the lesions on the three pulse sequences using Osirix[®] software (Pixmeo). ROIs were delineated only on the section level considered the most representative of the lesion (i.e., the one showing the most marked abnormality). For DCE images, ROIs were delineated only on the phase that best showed the lesion and automatically copied on the other phases. ROIs delineated on the three pulse sequences could have different size and shape and could be on slightly different slice levels if the lesion encompassed several slice levels. The model defined in the training population was used to calculate the CADx scores in the ROIs.

Determination of ground truth

Prostatectomy specimens were processed according to ISUP guidelines [34]. Tumor foci were given an individual ISUP grade and were delineated on whole-mount specimens by a pathologist (M.F.V.) with 15 years of experience in uropathology.

The ROIs were compared to the prostatectomy whole-mounts by the radiologist who delineated them, a researcher with 16 years of experience in prostate imaging (R.S.) and another radiologist with one year of experience (S.T.). By means of consensus, they assessed whether the ROIs matched the position of a histologic cancer or not. They also contoured, by consensus, histological cancers missed at mpMRI interpretation but that could be retrospectively seen on MR images.

Statistical analysis

Quantitative characteristics were described using medians and interquartile ranges (IQRs). PI-RADSv2 and CADx scores were compared using areas under the receiver operating characteristic curves (AUCs). AUCs were estimated using binormal smoothing and compared by the bootstrap test. Sensitivities and specificities were calculated using prospectively chosen thresholds (≥ 3 for the PI-RADSv2 score and ≥ 0.21 for the CADx score) that were defined before analyzing the data.

Statistical analysis was performed using R software (<http://www.cran.r-project.org>). Multiple testing was accounted for by using Bonferroni correction. $P < 0.05$ was considered statistically significant. All interval estimations are 95% confidence intervals (95%CI). The study is registered on ClinicalTrials.gov (identifier NCT03687918).

Results

Test population

Seventy-four men were included (Fig. 1). Their median age and prostate-specific antigen level were 65.7 years (IQR, 62.1–68.7) and 7.4 ng/ml (IQR, 5.7–11), respectively. Patients underwent MR imaging between September 2013

Table 1 Parameters of prostate multiparametric magnetic resonance imaging.

	Sequence name	TR (ms)	TE (ms)	FOV (mm ²)	AM	Flip angle (°)	ST (mm)	bV (s/mm ²)	TP	TPRE (s)
3 T	T2/3D-1	3500	133 to 135	303 × 249	135 × 231 × 384 to 167 × 289 × 384	140	1.25			
	T2/3D-2	2350 to 3100	101 to 134	243 × 303 to 263 × 320	137 × 223 × 384 to 151 × 259 × 448	130 to 160	1.25			
	T2/3D-3	2500	136	320 × 320	312 × 430 × 448	100 to 110	1.25			
	T2/2D	4690 to 6450	139 to 151	200 × 200 to 220 × 220	315 × 320	120	2,5			
	DWI/600	2300-330	74	175 × 175 to 230 × 230	96 × 96	90	4	100–600		
	DWI/800-1	3500	67	250 × 250	106 × 156	90	4	200–800		
	DWI/800-2	3100	59	175 × 175	96 × 96	90	4	200–800		
1.5 T	DCE-1	3.24 to 3.95	1.03 to 1.29	210 × 280 to 263 × 350	125 × 256	10	3.5		23 to 40	6.8 to 17.6
	DCE-2	3.92	1.43	260 × 260	205 × 256	12	3.5		23 to 35	8.1
	T2	2780	151	250 × 250	320 × 320	132	3,2			
	DWI	3100	85	198 × 198	96 × 96	90	4	0-600		
	DCE	5.35	1.66	225 × 300	158 × 320	15	3		23	11.3 to 22.5

TR: repetition time; TE: echo time; FOV: field of view; AM: acquisition matrix; ST: slice thickness; bV: *b*-values; TP: time points; TPRE: temporal resolution. Before August 2016, three 3 T protocols were used, associating the following pulse sequences: (a) T2/2D, DWI/600 or DWI/800-1, DCE-1; (b) T2/3D-1, DWI/600 or DWI/800-1, DCE-1; (c) T2/3D-2, DWI-600, DCE-1. After August 2016, only one 3 T protocol was used, associating T2/3D-3, DWI/800-2 and DCE-2. The same 1.5 T imaging protocol was used during the study period.

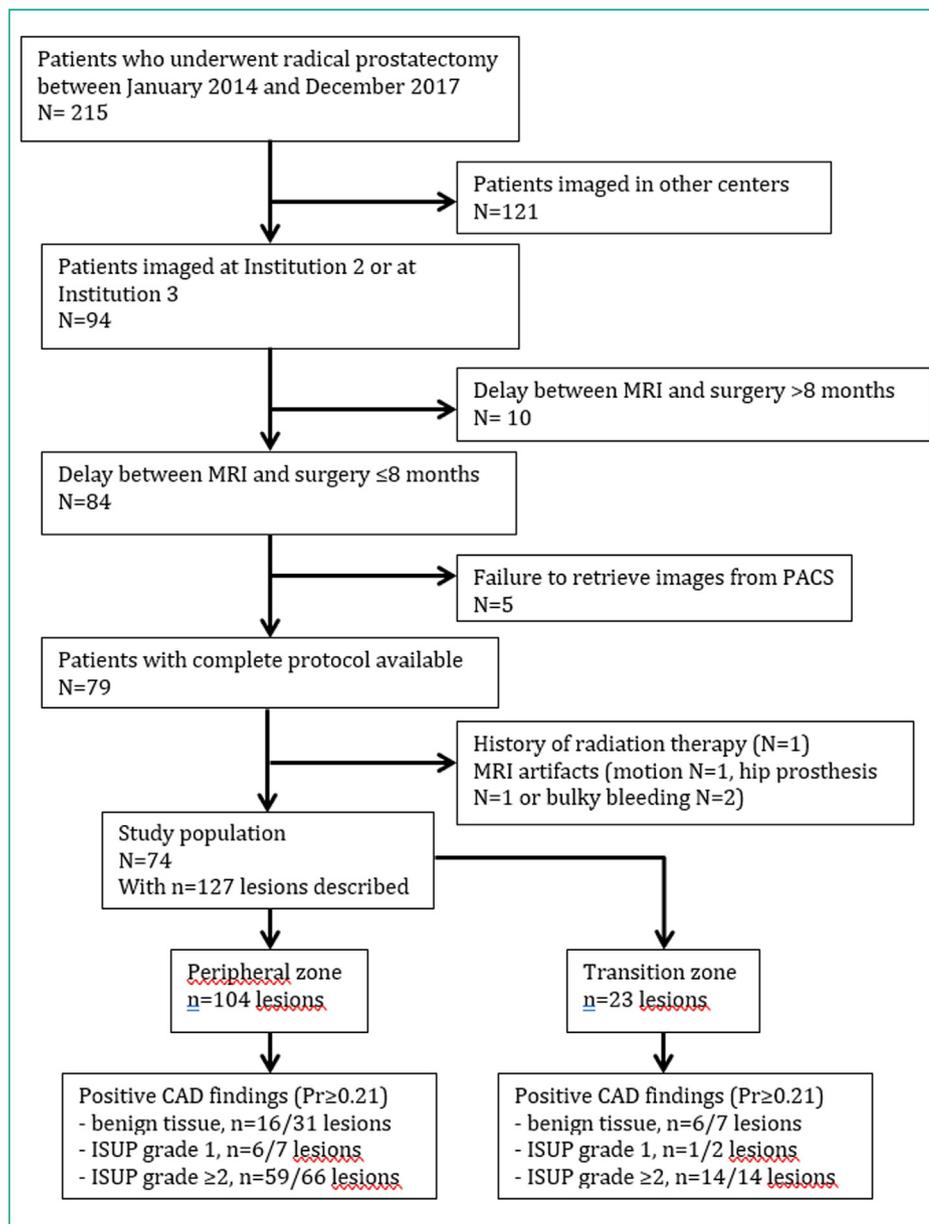


Figure 1. Standards for Reporting of Diagnostic Accuracy (STARD) flow diagram. MRI: magnetic resonance imaging; PACS: picture archiving and communication system; CAD: computer-aided diagnosis; MR: magnetic resonance; ISUP: International Society of Urological Pathology; N: number of patients; n: number of lesions.

and December 2017; all were imaged at 3T but one. The median time between MR imaging and surgery was 3.3 months (IQR, 2.1–4.8). On prostatectomy specimens, 93 cancers were found in PZ and 34 in TZ (Supplemental Tables 1 and 2).

MRI lesions

In total, the radiologist described 127 MRI lesions (PZ, $n=104$; TZ, $n=23$); 57 had bleeding artifacts (Table 2; Figs. 2 and 3). Tables 3 and 4 show the distribution of their PI-RADSv2 and CADx scores.

Performance of the CADx and PI-RADSv2 scores in PZ

The AUCs of the CADx and PI-RADSv2 scores were not significantly different, neither in the entire population of the 104 PZ lesions ($P=0.59$) nor in the subgroups of lesions with ($P=0.36$) or without ($P=0.56$) bleeding artifacts (Table 5). The AUCs of the CADx and PI-RADSv2 scores tended to be lower when bleeding artifacts were present, but the difference was not significant, neither for the CADx score ($P=0.46$) nor for the PI-RADSv2 score ($P=0.29$).

Using the ≥ 0.21 threshold, the sensitivity and specificity of the CADx score were 89% (95%CI: 82–97%)

Table 2 Focal prostatic lesions seen on multiparametric magnetic resonance imaging.

	PZ (N)	TZ (N)	Total (N)
Nature of lesions			
Benign tissue	31	7	38
ISUP grade 1 cancer	7	2	9
ISUP grade 2 cancer	36	7	43
ISUP grade 3 cancer	21	4	25
ISUP grade ≥ 4 cancer	9	3	12
Bleeding artifacts			
None	56	14	70
Mild	33	6	39
Moderate	7	1	8
Marked	8	2	10

PZ: peripheral zone; TZ: transition zone; N: number of patients; ISUP: International Society of Urological Pathology.

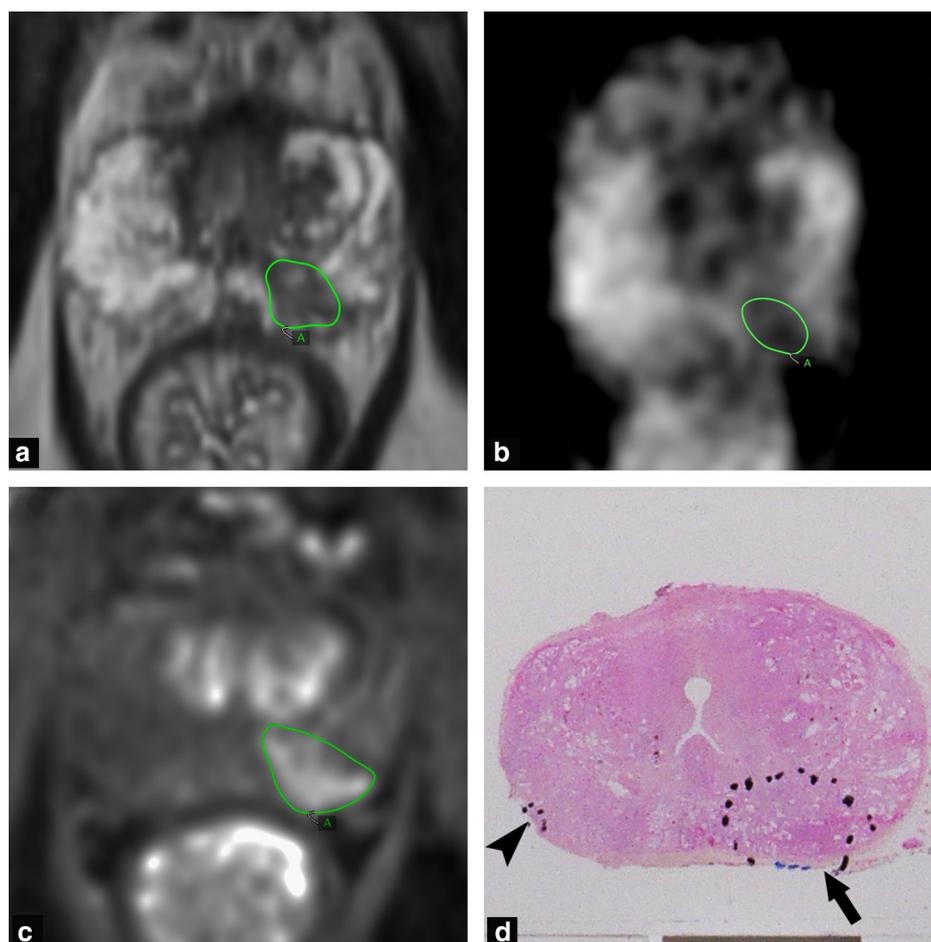


Figure 2. 66-year-old man with a prostate specific antigen (PSA) level of 3.8 ng/mL. Multiparametric magnetic resonance imaging reveals a 12-mm suspicious lesion in the peripheral zone of the left mid-gland that shows low-signal intensity on T2-weighted imaging (A, green outline), restriction of diffusion on the apparent diffusion coefficient map (B, green outline) and early focal enhancement on dynamic contrast-enhanced imaging (C, green outline). The radiologist assigned to that lesion a PI-RADSv2 score of 4. The CADx score was 0.60. On prostatectomy whole-mount, the lesion corresponded to an ISUP grade group 2 cancer (D, arrow). The black outline in the right lobe visible on Fig. D (arrowhead) corresponds to a microscopic focus of ISUP grade group 1 cancer.

and 42% (95%CI: 26–58%), respectively. These sensitivity and specificity values tended to be lower than those obtained using the same threshold in the training database, but the differences were not statistically

significant ($P=0.06$ for both). Using the ≥ 3 -threshold, the sensitivity and specificity of the PI-RADSv2 score were 97% (95%CI: 93–100%) and 37% (95%CI: 22–52%) respectively. The sensitivities of the CADx and PI-RADSv2 scores were

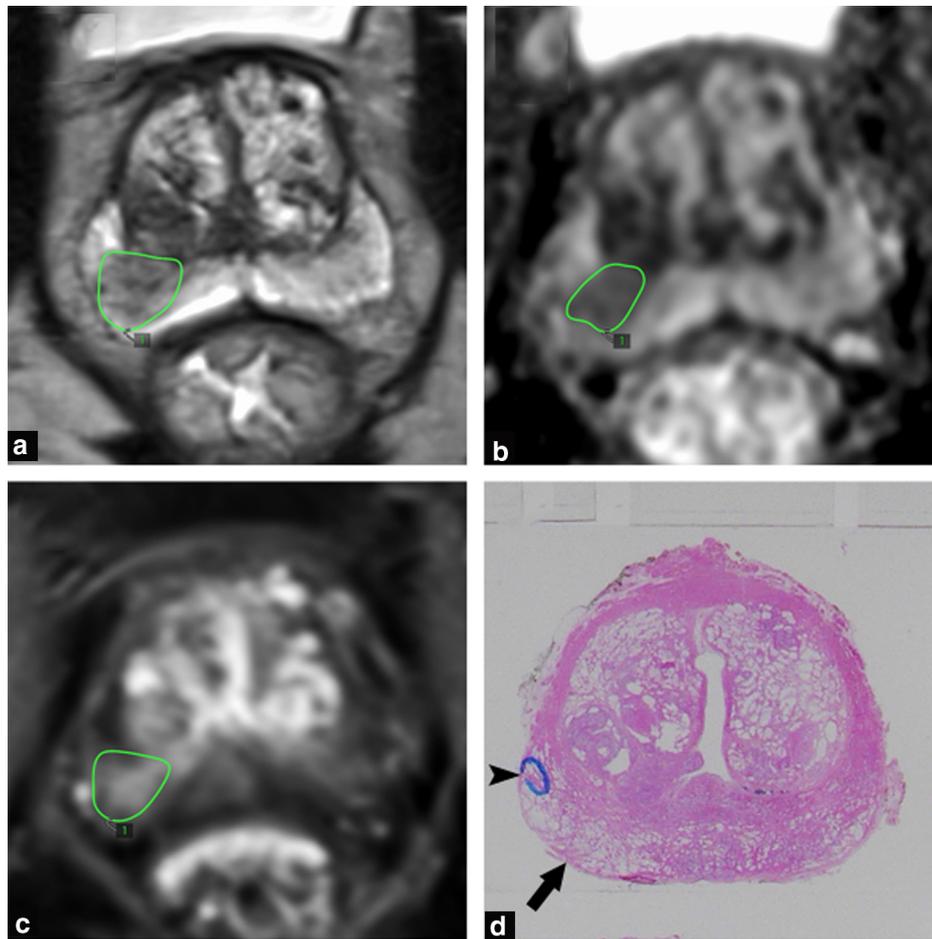


Figure 3. 62-year-old man with a prostate specific antigen (PSA) level of 4.2 ng/mL. Multiparametric magnetic resonance imaging showed a 13-mm suspicious lesion in the peripheral zone of the right mid-gland that showed low-signal intensity at T2-weighted imaging (A, green outline), restriction of diffusion on the apparent diffusion coefficient map (B, green outline) and early focal enhancement on dynamic contrast-enhanced imaging (C, green outline). The radiologist assigned to that lesion a PI-RADSv2 score of 4. The CADx score was 0.02. On prostatectomy whole-mount, there was no matching histological cancer (D, arrow). The blue outline on Fig D (arrowhead) shows a microscopic focus of ISUP grade group 1 cancer.

significantly different ($P=0.04$) but not their specificities ($P=0.83$).

Performance of the CADx and PI-RADSv2 scores in TZ

Of the 23 TZ lesions, only two had a CAD score < 0.21 ; one was benign and the other was an ISUP grade 1 cancer. All TZ lesions had a PI-RADS score ≥ 3 , irrespective of their nature.

Missed cancers

Forty-nine histological cancers (26 ISUP grade 1, 23 ISUP grade ≥ 2) were missed by the radiologist. Sixteen of these (3 ISUP grade 1, 13 ISUP grade ≥ 2) were retrospectively visible on MR images and could be delineated. Thirteen of these 16 cancers (2 ISUP grade 1, 11 ISUP grade ≥ 2) had a CADx score ≥ 0.21 .

Discussion

The purpose of this study was to evaluate the performances of a CADx system combining ADC_{10th} and TTP in a test population of patients from a different institution and imaged on MR scanners from a different manufacturer than those included in the training database. This 'external' test population was composed of patients treated by prostatectomy, although the main target for our CADx system would be patients referred to prostate biopsy. We made this methodological choice because comparison with prostatectomy specimens allows precise radiologic-pathologic correlations. Using targeted biopsy findings as ground truth is debatable since targeted biopsy may miss some prostate cancer foci or underestimate the aggressiveness of the detected cancers [35]. Had the CADx system showed poor results, we wanted to make sure that these poor results were due to the poor performance of the algorithm and not to imprecise histological reference standard.

The patients of the test population were imaged on three different MRI scanners, and imaging parameters varied substantially during the study period. DCE imaging temporal

Table 3 Distribution of the PI-RADSv2 and CADx scores in prostatic lesions seen on multiparametric magnetic resonance imaging.

PI-RADSv2 score	Benign tissue	ISUP grade 1 cancer		ISUP grade 2 cancer		ISUP grade \geq 3 cancer
		\leq 0.5 cc	> 0.5 cc	\leq 0.5 cc	> 0.5 cc	
PZ						
1						
CAD –	5 (2)	0	0	1 (1)	0	1 (1)
CAD +	0	0	0	0	0	0
2						
CAD –	4 (1)	0	0	0	0	0
CAD +	4 (2)	1 (1)	0	0	0	0
3						
CAD –	1	0	0	0	0	0
CAD +	2 (1)	0	0	0	1 (1)	0
4						
CAD –	5 (2)	0	0	1	3 (2)	1 (1)
CAD +	8 (2)	1 (1)	2 (2)	0	25 (12)	11 (7)
5						
CAD –	0	1 (1)	0	0	0	0
CAD +	2	0	2 (1)	5 (2)	0	17 (5)
TZ						
3						
CAD –	1	0	0	0	0	0
CAD +	2 (1)	0	1	0	1 (1)	0
4						
CAD –	0	1	0	0	0	0
CAD +	4 (2)	0	0	0	3	2 (1)
5						
CAD –	0	0	0	0	0	0
CAD +	0	0	0	0	3 (2)	5 (2)

PZ: peripheral zone; TZ: transition zone; PI-RADSv2: Prostate Imaging-Reporting and Data System version 2; ISUP: International Society of Urological Pathology; cc: cubic centimetre; CADx: computer-aided diagnosis system. The table shows numbers of lesions. The numbers in parentheses are numbers of lesions with bleeding artefacts among the lesions of the cell. For example, a total of five lesions in PZ had a PI-RADS score of 1, negative CADx findings and corresponded to benign tissue on prostatectomy specimens. Among these five lesions, two had bleeding artefacts on magnetic resonance imaging. CADx score was dichotomized using a \geq 0.21 threshold.

Table 4 CADx scores of prostatic lesions seen on multiparametric magnetic resonance imaging.

	PZ		TZ	
	Median (IQR)	<i>n</i>	Median (IQR)	<i>n</i>
Benign lesions and ISUP grade 1 cancers	0.30 (0.04–0.67)	38	0.85 (0.33–0.93)	9
ISUP grade 2 cancers	0.85 (0.46–0.97)	36	0.98 (0.93–0.99)	7
ISUP grade 3 cancers	0.78 (0.76–0.97)	21	0.74 (0.66–0.86)	4
ISUP grade \geq 4 cancers	0.97 (0.92–0.97)	9	0.93 (0.86–0.95)	3

PZ: peripheral zone; TZ: transition zone; IQR: interquartile range; *n*: number of lesions; ISUP: International Society of Urological Pathology; CADx: computer-aided diagnosis system.

resolution, that is essential for accurate estimation of TTP, was highly variable (6.8–17.6 s). The acquired *b*-values were also variable. Maximum *b*-values were lower than those recommended by the PI-RADSv2 guidelines and different from the one used in the training database (2000 s/mm²). This is explained by the fact that most MRI examinations were performed before the PI-RADSv2 guidelines were published, and

by the absence of selection criteria based on image parameters or image quality; instead, we included all consecutive MRI examinations in order to obtain a 'real-life' database. As showed by a recent survey, the level of adherence to PI-RADSv2 technical standards is highly variable across imaging facilities [36]. In the only study that evaluated a CAD system for prostate mpMRI using a truly 'external' database,

Table 5 Areas under the receiver operating characteristic curves of the CADx and PI-RADSv2 scores in the peripheral zone.

	No bleeding artifacts ($n = 56$)	Bleeding artifacts ($n = 48$)	Overall population ($n = 104$)
CADx	0.81 (0.68–0.92)	0.74 (0.59–0.87)	0.78 (0.69–0.87)
PI-RADSv2	0.78 (0.66–0.89)	0.67 (0.51–0.83)	0.74 (0.62–0.86)

n: number of lesions. Data in parentheses are 95% confidence intervals.

there was also high variability in imaging parameters and all protocols were not entirely compliant with PI-RADSv2 specifications, even if the participating centers were academic institutions [30]. In the same study, 24% of MRI examinations showed poor image quality based on rectal distension and prostate motion [30]. CADs are not supposed to replace non-valid MRI protocols and the uro-radiological community must strive to facilitate the diffusion of PI-RADSv2 standards across institutions. However, CAD systems will not be useful in daily practice if they are not robust to variations in imaging parameters. The fact that mpMRIs of the test population do not entirely comply with up-to-date standards impairs the generalization of our results. It can also be viewed, on the other hand, as a challenging test for our CAD system.

Unsurprisingly, the CADx sensitivity and specificity tended to be lower in this difficult 'external' test population than in the training population and the differences almost reached statistical significance ($P = 0.06$). However, the difference between the observed sensitivity (89%; 95%CI: 82–97%) and the expected 95% sensitivity remained small, while the difference in specificity (42% [95%CI: 26–58%] vs. 57% [95%CI: 24–69%]) was substantial. The lower specificity obtained in the test population could be explained, at least partially, by the large proportion of lesions with bleeding artifacts, that are known to create false positive-findings at mpMRI [37,38] and that showed substantial impact on the CADx AUC in this series. Because mpMRI is increasingly used before biopsy, these bleeding artefacts will be less problematic in the future.

In this series, the CADx AUC for characterizing ISUP grade ≥ 2 cancers (0.78; 95%CI: 0.69–0.87) was similar to that of the PI-RADSv2 score assigned by an experienced radiologist (0.74; 95%CI: 0.62–0.86). However, when both tests were dichotomized, the sensitivity of the CADx score (threshold ≥ 0.21) was significantly lower than that of the PI-RADSv2 score (threshold ≥ 3). This latter result must be interpreted with care. Indeed, a CADx only provides a score for lesions that have been detected by the human reader. Its expected added value lies in characterizing these lesions, i.e. in improving the specificity of reading. Only a CADE system can be expected to improve the sensitivity of human interpretation. In this study, 16 of the 49 histological cancers missed by the radiologist were retrospectively visible on MR images. The CADx score was ≥ 0.21 in 13 of these 16 cancers, including 11 ISUP grade ≥ 2 cancers. This strongly suggests that a CADE approach has the potential to improve the sensitivity of cancer detection, even for experienced readers. However, as discussed above, this necessitates to deal with more noisy parameters, since quantification is made at the pixel level rather than in large ROIs, and to address segmentation and co-registration issues that are not trivial [30].

CAD systems are not designed to replace radiologists but to help them. They may be mostly needed for the PI-RADSv2 3–4 lesions, which are the most difficult to characterize. Our results suggest the CADx could help in this setting, at least in PZ. There were only 4 PI-RADSv2 3 lesions in PZ; the CADx was positive in three including the only ISUP ≥ 2 cancer. Similarly, ISUP ≥ 2 cancers corresponded to 50% (5/10) of PI-RADSv2 4 lesions with negative CADx findings and to 77% (36/47) of PI-RADSv2 4 lesions with positive CADx findings.

The current version of the CADx was trained only on PZ lesions and showed poor specificity on suspicious TZ lesions. This was also observed in the 'internal' validation cohort [31]. Diagnosis of TZ cancers relies on other criteria than signal changes (e.g., absence of capsule, lenticular shape, anterior-apical position) [33] that are not yet taken into consideration by our CADx system.

This study has some limitations. First, PI-RADSv2 assessment for diffusion-weighted imaging was based only on ADC maps. However, PI-RADSv2 criteria do not account for large discrepancies between ADC maps and high *b*-value images (e.g., marked hypointense on ADC and mildly hyperintense on high *b*-value images; or, mildly hypointense on ADC and markedly hyperintense on high *b*-value images), probably because they are rare. Thus, we believe that using only ADC maps for scoring had little impact on the final score. In our experience, high *b*-value images are mostly useful for tumor detection because they provide good contrast between cancers and background prostate tissue. Thus, the proportion of cancers missed by human lecture may have been overestimated in the present study. Second, only one radiologist read the MR examinations. However, our purpose was only to provide an initial assessment of the CADx in an 'external' test cohort. Third, because we used patients treated by prostatectomy, the results reported herein may not be fully reproducible on patients with suspicion of prostate cancer. Nonetheless, in an independent 'internal' test population of patients referred for prostate biopsy, using the ≥ 0.21 threshold, the CADx sensitivity (96%; 95%CI: 92–100%) and specificity (44%; 95%CI: 36–52%) for ISUP grade ≥ 2 cancers were close to those observed in the present cohort [31].

In conclusion, our study is one of the first evaluations of a CAD system designed for prostate mpMRI in a truly 'external' test population. In this 'external' heterogeneous validation cohort of mpMRIs performed during routine practice according to variable protocols, the CADx system tended to show lower sensitivity and specificity than in its training database, but yielded an overall performance similar to the PI-RADSv2 score assigned by an experienced radiologist. In the future, CAD systems may therefore help readers with less experience standardize routine interpretation of prostate mpMRI.

Human and animal rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patient(s).

Funding

This work was supported by the RHU PERFUSE (ANR-17-RHUS-0006) of the Université Claude-Bernard Lyon 1 (UCBL), within the program “Investissements d’Avenir” operated by the French National Research Agency (ANR).

Authors’ contribution

Sarah Transin: Data curation; Investigation; Writing—original draft.

Rémi Souchon: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Writing—Review & editing.

Christelle Gonindard-Melodelima: Formal analysis; Methodology; Writing—Review & editing.

Romane de Rozario: Data curation; Software; Writing—Review & editing.

Paul Walker: Investigation; Methodology; Writing—Review & editing.

Mathilde Funes de la Vega: Investigation; Writing—Review & editing.

Romarc Loffroy: Investigation; Methodology; Writing—Review & editing.

Luc Cormier: Investigation; Writing—Review & editing.

Olivier Rouvière: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Supervision; Validation; Writing—original draft.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.diii.2019.06.012>.

Disclosure of interest

The authors declare that they have no competing interest.

References

- [1] Bratan F, Niaf E, Melodelima C, et al. Influence of imaging and histological factors on prostate cancer detection and localisation on multiparametric MRI: a prospective study. *Eur Radiol* 2013;23:2019–29.
- [2] Le JD, Tan N, Shkoliar E, et al. Multifocality and prostate cancer detection by multiparametric magnetic resonance imaging: correlation with whole-mount histopathology. *Eur Urol* 2015;67:569–76.
- [3] Borofsky S, George AK, Gaur S, et al. What are we missing? False-negative cancers at multiparametric MR imaging of the prostate. *Radiology* 2018;286:186–95.
- [4] Drost FH, Osses DF, Nieboer D, et al. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. *Cochrane Database Syst Rev* 2019;4: CD012663.
- [5] Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med* 2018;378:1767–77.
- [6] Rouvière O, Puech P, Renard-Penna R, et al. Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. *Lancet Oncol* 2019;20: 100–9.
- [7] van der Leest M, Cornel E, Israel 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-naive men with elevated prostate-specific antigen: a large prospective multicenter clinical study. *Eur Urol* 2019;75:570–8.
- [8] Renard-Penna R, Rouvière O, Puech P, et al. Current practice and access to prostate MR imaging in France. *Diagn Interv Imaging* 2016;97:1125–9.
- [9] Muthigi A, Sidana A, George AK, et al. Current beliefs and practice patterns among urologists regarding prostate magnetic resonance imaging and magnetic resonance-targeted biopsy. *Urol Oncol* 2017;35:32e1–1e7.
- [10] Kasel-Seibert M, Lehmann T, Aschenbach R, et al. Assessment of PI-RADS v2 for the detection of prostate cancer. *Eur J Radiol* 2016;85:726–31.
- [11] 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;280:793–804.
- [12] Tewes S, Mokov N, Hartung D, et al. Standardized reporting of prostate MRI: comparison of the Prostate Imaging Reporting and Data System (PI-RADS) Version 1 and Version 2. *PLoS One* 2016;11:e0162879.
- [13] Lemaitre G, Marti R, Freixenet J, Vilanova JC, Walker PM, Meriaudeau F. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Comput Biol Med* 2015;60:8–31.
- [14] Fei B. Computer-aided diagnosis of prostate cancer with MRI. *Curr Opin Biomed Eng* 2017;3:20–7.
- [15] Liu L, Tian Z, Zhang Z, Fei B. Computer-aided detection of prostate cancer with MRI: technology and applications. *Acad Radiol* 2016;23:1024–46.
- [16] Niaf E, Lartizien C, Bratan F, et al. Prostate focal peripheral zone lesions: characterization at multiparametric MR imaging: influence of a computer-aided diagnosis system. *Radiology* 2014;271:761–9.
- [17] Hambrock T, Vos PC, Hulsbergen-van de Kaa CA, Barentsz JO, Huisman HJ. Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging: effect on observer performance. *Radiology* 2013;266:521–30.
- [18] Greer MD, Lay N, Shih JH, et al. Computer-aided diagnosis prior to conventional interpretation of prostate mpMRI: an international multi-reader study. *Eur Radiol* 2018;28:4407–17.
- [19] Petrick N, Sahiner B, Armato 3rd SG, et al. Evaluation of computer-aided detection and diagnosis systems. *Med Phys* 2013;40:087001.

- [20] Savadjiev P, Chong J, Dohan A, et al. Demystification of AI-driven medical image interpretation: past, present and future. *Eur Radiol* 2019;29:1616–24.
- [21] Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. *Radiology* 2015;277:813–25.
- [22] Malyarenko DI, Newitt DLJW, et al. Demonstration of nonlinearity bias in the measurement of the apparent diffusion coefficient in multicenter trials. *Magn Reson Med* 2016;75:1312–23.
- [23] Fedeli L, Belli G, Ciccarone A, et al. Dependence of apparent diffusion coefficient measurement on diffusion gradient direction and spatial position: a quality assurance inter-comparison study of forty-four scanners for quantitative diffusion-weighted imaging. *Phys Med* 2018;55:135–41.
- [24] Kim H. Variability in quantitative DCE-MRI: sources and solutions. *J Nat Sci* 2018;4.
- [25] Schlett CL, Hendel T, Hirsch J, et al. Quantitative, organ-specific interscanner and intrascanner variability for 3T whole-body magnetic resonance imaging in a multicenter, multivendor study. *Invest Radiol* 2016;51:255–65.
- [26] Shukla-Dave A, Obuchowski NA, Chenevert TL, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging* 2019;49:e101–21.
- [27] Peng Y, Jiang Y, Antic T, Giger ML, Eggener SE, Oto A. Validation of quantitative analysis of multiparametric prostate MR images for prostate cancer detection and aggressiveness assessment: a cross-imager study. *Radiology* 2014;271:461–71.
- [28] Hoang Dinh A, Souchon R, Melodelima C, et al. Characterization of prostate cancer using T2 mapping at 3T: a multi-scanner study. *Diagn Interv Imaging* 2015;96:365–72.
- [29] Hoang Dinh A, Melodelima C, Souchon R, et al. Quantitative analysis of prostate multiparametric MR images for detection of aggressive prostate cancer in the peripheral zone: a multiple imager study. *Radiology* 2016;280:117–27.
- [30] Gaur S, Lay N, Harmon SA, et al. Can computer-aided diagnosis assist in the identification of prostate cancer on prostate MRI? A multi-center, multi-reader investigation. *Oncotarget* 2018;9:33804–17.
- [31] Dinh AH, Melodelima C, Souchon R, et al. Characterization of prostate cancer with Gleason score of at least 7 by using quantitative multiparametric MR imaging: validation of a computer-aided diagnosis system in patients referred for prostate biopsy. *Radiology* 2018;287:525–33.
- [32] Brunelle S, Zemmour C, Bratan F, et al. Variability induced by the MR imager in dynamic contrast-enhanced imaging of the prostate. *Diagn Interv Imaging* 2018;99:255–64.
- [33] Chesnais AL, Niaf E, Bratan F, et al. Differentiation of transitional zone prostate cancer from benign hyperplasia nodules: evaluation of discriminant criteria at multiparametric MRI. *Clin Radiol* 2013;68:e323–30.
- [34] Samaratunga H, Montironi R, True L, et al. International Society of Urological Pathology (ISUP) Consensus Conference on handling and staging of radical prostatectomy specimens. Working group 1: specimen handling. *Mod Pathol* 2011;24:6–15.
- [35] Mortezaei A, Marzendorfer O, Donati OF, et al. Diagnostic accuracy of multiparametric magnetic resonance imaging and fusion guided targeted biopsy evaluated by transperineal template saturation prostate biopsy for the detection and characterization of prostate cancer. *J Urol* 2018;200:309–18.
- [36] Esses SJ, Taneja SS, Rosenkrantz AB. Imaging facilities' adherence to PI-RADS v2 minimum technical standards for the performance of prostate MRI. *Acad Radiol* 2018;25:188–95.
- [37] Ikonen S, Kivisaari L, Vehmas T, et al. Optimal timing of post-biopsy MR imaging of the prostate. *Acta Radiol* 2001;42:70–3.
- [38] Tamada T, Sone T, Jo Y, et al. Prostate cancer: relationships between postbiopsy hemorrhage and tumor detectability at MR diagnosis. *Radiology* 2008;248:531–9.