

# Clinical evaluation of a computer-aided diagnosis system for determining cancer aggressiveness in prostate MRI

Geert J. S. Litjens<sup>1</sup> · Jelle O. Barentsz<sup>1</sup> · Nico Karssemeijer<sup>1</sup> · Henkjan J. Huisman<sup>1</sup>

Received: 20 November 2014 / Revised: 13 March 2015 / Accepted: 25 March 2015 / Published online: 10 June 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

## Abstract

**Objectives** To investigate the added value of computer-aided diagnosis (CAD) on the diagnostic accuracy of PIRADS reporting and the assessment of cancer aggressiveness.

**Methods** Multi-parametric MRI and histopathological outcome of MR-guided biopsies of a consecutive set of 130 patients were included. All cases were prospectively PIRADS reported and the reported lesions underwent CAD analysis. Logistic regression combined the CAD prediction and radiologist PIRADS score into a combination score. Receiver-operating characteristic (ROC) analysis and Spearman's correlation coefficient were used to assess the diagnostic accuracy and correlation to cancer grade. Evaluation was performed for discriminating benign lesions from cancer and for discriminating indolent from aggressive lesions.

**Results** In total 141 lesions (107 patients) were included for final analysis. The area-under-the-ROC-curve of the combination score was higher than for the PIRADS score of the radiologist (benign vs. cancer, 0.88 vs. 0.81,  $p=0.013$  and indolent vs. aggressive, 0.88 vs. 0.78,  $p<0.01$ ). The combination score correlated significantly stronger with cancer grade (0.69,  $p=0.0014$ ) than the individual CAD system or radiologist (0.54 and 0.58).

**Conclusions** Combining CAD prediction and PIRADS into a combination score has the potential to improve diagnostic accuracy. Furthermore, such a combination score has a strong correlation with cancer grade.

## Key Points

- Computer-aided diagnosis helps radiologists discriminate benign findings from cancer in prostate MRI.
- Combining PIRADS and computer-aided diagnosis improves differentiation between indolent and aggressive cancer.
- Adding computer-aided diagnosis to PIRADS increases the correlation coefficient with respect to cancer grade.

**Keywords** Magnetic resonance imaging · Prostate cancer · Computer-aided diagnosis · Observer study · Diagnostic performance

## Introduction

Multi-parametric magnetic resonance imaging (mpMRI) is emerging as an important modality in prostate cancer diagnosis [1–3]. Several studies have shown that in patients with initial negative trans-rectal ultrasound-guided biopsies (TRUSGB) and persistently elevated prostate-specific antigen (PSA) expert readers using mpMRI find cancer in 38–59 % of the cases [4, 5]. Furthermore, it has been shown that mpMRI correctly upgrades TRUSGB-detected cancers in up to 30 % of cases [6]. Several other studies found that the negative predictive value of mpMRI is high enough to avoid TRUSGB in 30–50 % of men with persistently elevated PSA [7, 8]. However, one of the main limitations for broader acceptance of mpMRI is the lack of required expertise, especially in the acquisition and interpretation of the MR images [1, 9, 10].

In order to improve the acquisition and interpretation of mpMRI, the European Society for Urogenital Radiology (ESUR) established initial guidelines for acquisition and standardized interpretation of mpMRI (PIRADS) [1]. These guidelines have been evaluated by several groups, for

✉ Geert J. S. Litjens  
Geert.Litjens@radboudumc.nl

<sup>1</sup> Department of Radiology, Radboud University Medical Center, 766, Postbus 9101, 6500HB Nijmegen, The Netherlands

detection of cancer both prior to biopsy [8, 11] and after initial negative TRUSGB [12–15]. There are, however, still two major issues in current prostate MRI: determining which cancers need treatment (assessment of aggression) and the large amount of false positives resulting in unnecessary biopsies. Computer-aided diagnosis (CAD) might be able to address these problems.

The purpose of this study is twofold. One is to provide a clinical investigation of the effect of CAD [16] on the diagnostic accuracy of prostate mpMRI reporting via independent combination of PIRADS scores and CAD prediction into a radiologist/CAD combination score. The performance of this combination score is evaluated in a comparatively large cohort of patients with MR-guided MR-biopsy histopathological outcome as reference standard. Secondly, we investigate the ability of CAD to estimate prostate cancer aggressiveness.

## Materials and methods

### Patient data

An institutional review board (IRB) waiver applies to this study as it uses anonymized imaging data and MR-guided biopsy results obtained through regular clinical care. In total 130 consecutive patients from 1 January to 1 September 2013 who received both an mpMRI and a subsequent MR-guided MR-biopsy at our institution were included. The inclusion criteria for the detection mpMRI were an initial negative TRUSGB and persistently elevated PSA (consistently above 4 ng/mL).

Multi-parametric MRIs were acquired according to the ESUR guidelines and included T2-weighted imaging in three orthogonal directions, diffusion-weighted imaging and dynamic contrast-enhanced imaging. All MRIs were performed at a Siemens 3 T MRI scanner (TRIOTIM or Skyra) without an endo-rectal coil. Full acquisition details are presented in Table 1.

Each mpMRI was regularly, prospectively read by one radiologist out of a group of seven radiologists who reported prostate MRI in our clinic. Experience levels of the reporting radiologist ranged from moderately experienced (2 years) to very experienced (J.B., 20 years). Details on the amount of cases read by each radiologist are presented in Table 2. The ESUR prostate imaging reporting and data system (PIRADS) classification was used to assign a five-point PIRADS score to one or more lesions.

Each MR study was reported using a dedicated prostate MR workstation that allowed radiologists to indicate one or more areas of suspicion with a sphere enclosing the lesions. If no suspicious areas could be identified, a location was marked which was deemed normal/benign and assigned a PIRADS score of 1 or 2. This is performed in routine clinical care for accountability, to ensure that each case has been read and

**Table 1** MRI sequence details for the different types of acquisitions

SN	SR	ST	AM	FOV	ET	RT	FA	SS
T2W/Turbo spin-echo	0.28–0.6 mm	3.0–3.2 mm	320×320–384×384	108×108–192×192 mm	101–104 ms	4480–6840 ms	120–160°	Acquired in three orthogonal directions: transversal, sagittal and coronal
DWIEcho planar	2 mm	3 mm	128×128	256×256 mm	63–81 ms	2800–3600 ms	90°	3 b-values: 50, 400–500, 800 averaged over three directions. Apparent diffusion coefficient map calculated by the scanner software
DCEFast low-angle shot spoiled gradient recalled echo	1.5–1.8 mm	3.2–5 mm	128×128	192×192–230×230 mm	1.41 ms	36 ms	10–14°	Temporal resolution of 3.38–4.65 seconds, 36–50 timepoints. 15 mL contrast agent used (Dotarem, Guerbet, France)

SN=sequence name, SR=slice thickness, ST=slice thickness, AM=acquisition matrix, FOV=Field of View, ET=echo time, RT=repetition time, FA=flip angle, SS=sequence specific details

**Table 2** Overview of radiologist reading cases in the study cohort, including amount of cases read (out of 107 included studies) and years of experience

Reader	Years of experience	Cases read
<i>J.B.</i>	20	25
<i>J.F.</i>	12	14
<i>P.Z.</i>	8	16
<i>S.J.</i>	3	20
<i>M. vd. L.</i>	2	22
<i>R.M.</i>	2	5
<i>J.H.</i>	2	5

reported. Typically, PIRADS 1 or 2 lesions do not get biopsied. Sometimes a PIRADS 2 lesion was also biopsied, when a PIRADS 3 or higher lesion was also identified and the patient was thus already scheduled for MR-guided MR biopsy. The locations and scores were automatically recorded in a database.

MR-guided biopsies were performed by medical experts with multiple years of experience in MR-guided prostate biopsies. At the start of the biopsy procedure a T2-weighted volume and an ADC map were acquired according to the prostate cancer detection protocol (Table 1). These sequences were used to relocate the lesions in the prior detection MRI. After lesions have been identified, a needle guide is inserted transrectally. Consecutive sagittal and transversal MRIs are made during repositioning of the needle guide to assess whether the correct position has been reached. Once the correct position has been reached a biopsy needle is inserted and a biopsy taken. To verify the biopsy location, sagittal and transversal images were made with the needle in situ. Subsequently, biopsies were histopathologically processed, inspected and graded by an experienced uropathologist (17 years of experience in prostate pathology).

### Computer-aided diagnosis system

The computer-aided diagnosis system evaluated in this paper was previously presented in [16]. First, the system computes quantitative voxel features, which were designed to capture characteristics described by the PIRADS guidelines. A full feature listing can be found in Table 3. These voxel features are then fed to a random forest classifier trained to determine a continuous likelihood score for each voxel to identify cancer, resulting in a likelihood image. Subsequently, in a second stage, the system used the centre of the sphere indicated by the radiologist as a starting point for lesion segmentation, which is performed on the pre-computed likelihood image. After lesion segmentation histogram statistics are calculated on the voxel features within the lesion (e.g., percentiles, mean, standard deviation). In addition, local contrast is calculated by

comparing the voxel feature values within the lesion to values outside the lesion. Symmetry is calculated by comparing the feature values within the lesion to the feature values at the same relative position on the contra-lateral side of the prostate. The statistical, local contrast and symmetry features are then combined using a second random forest classifier trained to predict cancer likelihood per lesion. The system is able to take into account the zonal location of the lesion via the use of a probabilistic segmentation of the prostate zones as one of its features (Table 3). The construction of this probabilistic segmentation is detailed in [17].

The CAD system was trained with independent, retrospective patient data (237 patients), which had no overlap with the data set used in this study. The retrospective data was acquired in a similar manner (same MRI protocol) as the evaluation data and also had MR-guided biopsy as the reference standard.

### Combination of PIRADS score and computer-aided diagnosis (CAD) likelihood

The use of the system as proposed in this paper is presented schematically in Fig. 1. The initial identification of potential suspicious regions was performed by the radiologist, after which the radiologist and the CAD system gave independent scores on whether clinically significant cancer was present [1]. The radiologist did this by assigning a five-point PIRADS score, while the CAD system assigned a continuous likelihood score between 0 and 1.

The reported scores of the radiologists (PIRADS) and CAD (likelihoods) were combined into a combination score via logistic regression, which is a technique to map multiple variables to one single, continuous outcome variable (between 0.0 and 1.0) in an independent manner. The regression model was created with SPSS (version 20.0.01, Chicago, IL, USA). The logistic regression was performed on the retrospective data that was also used to train the CAD system to ensure no bias would occur by training and testing on the same data. It was thus independent of the evaluation data used in this paper. Alternative methods of incorporating CAD results may be used in clinical workflow, such as asking a second radiologist to make a final decision based on the two scores, but these were not investigated in this paper.

### Statistical evaluation

Radiologist-identified lesions were categorized into either benign or cancer based on the MR-guided MR biopsy outcome. Cancerous lesions were further subdivided into low-grade, intermediate-grade or high-grade cancer based on the MR-guided biopsy Gleason scores, similar to Vos et al. [18] and Hambroek et al. [19]. Our high-sensitivity MR-guided biopsy strategy has been shown to have a concordance of 95 % with prostatectomy Gleason grade [20].

**Table 3** Descriptions of the voxel features used in the computer-aided diagnosis system

Name	Type	Description
T2W	Intensity	T2-weighted voxel grey value, related to voxel T2
ADC	Intensity	Apparent diffusion coefficient, measure for cellular density
b800	Intensity	High b-value image, areas with low diffusivity appear bright
T2-map	Intensity	Calculated T2-map based on proton density and transversal T2W image [17]
x-pos	Anatomical	Relative cumulative position within the prostate mask between 0 and 1 in the x-direction
y-pos	Anatomical	Relative cumulative position within the prostate mask between 0 and 1 in the y-direction
z-pos	Anatomical	Relative cumulative position within the prostate mask between 0 and 1 in the z-direction
Distance	Anatomical	Relative distance to the prostate boundary between 0 and 1
PZ	Likelihood	Anatomical likelihood of being a peripheral zone voxel between 0 and 1 [17]
Ktrans	Pharmacokinetic	Pharmacokinetic parameter, related to vessel permeability
kep	Pharmacokinetic	Pharmacokinetic parameter, related to permeability and extracellular volume
tau	Pharmacokinetic	Dynamic parameter, related to the time-to-peak of contrast agent concentration
LateWash	Pharmacokinetic	Dynamic parameter, related to the washout of contrast agent
Gaussian texture bank	Texture	Calculate multi-scale Gaussian derivatives on the T2W image
ADC	Spatial filter	Multi-scale focal lesion detection using the Li spatial filter [27] on ADC map
Ktrans	Spatial filter	Multi-scale focal lesion detection using the Li spatial filter [27] on Ktrans map
LateWash	Spatial filter	Multi-scale focal lesion detection using the Li spatial filter [27] on LateWash map
tau	Spatial filter	Multi-scale focal lesion detection using the Li spatial filter [27] on tau map

We used two different settings for evaluation in this study: either benign versus cancerous or indolent versus aggressive lesions. In the latter case the benign and low-grade lesions are considered indolent and intermediate- and high-grade lesions are considered aggressive. These settings are summarized in Table 4. The CAD system and logistic regression model were constructed separately for each setting using the retrospective data.

The statistical evaluation consisted of three parts. First, we investigated the hypothesized increase in predictive power of the combination score over the radiologist PIRADS score alone using the likelihood ratio test on the logistic regression models.

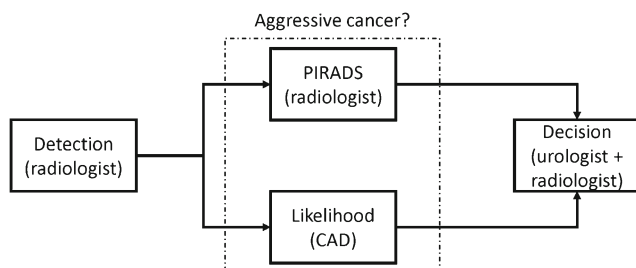
Second, the diagnostic performance of the CAD system, the radiologist PIRADS score and the combination score was evaluated using receiver-operating characteristic (ROC) analysis for both evaluation strategies. The significance of improvement for area under the ROC curve, and different

sensitivity-specificity pairs at the different PIRADS thresholds, was tested using bootstrapping. A total of 10,000 bootstrap samples was used to obtain the 95 % confidence intervals (CIs). Bootstrapping was stratified according to patient to circumvent bias introduced by multiple lesions per patient.

To assess the effect of the zonal location of the lesions on the performance of the radiologist, the CAD system and the combination score the dataset was split into two sets, one set containing only central gland lesions and one set only containing peripheral zone lesions.

The effect of observer experience on the performance of the combination score was also assessed. The dataset was split into two groups, one containing the cases reported by the experienced radiologists (more than 5 years) and one containing the cases reported by the less experienced radiologists (less than 5 years, but more than 2 years).

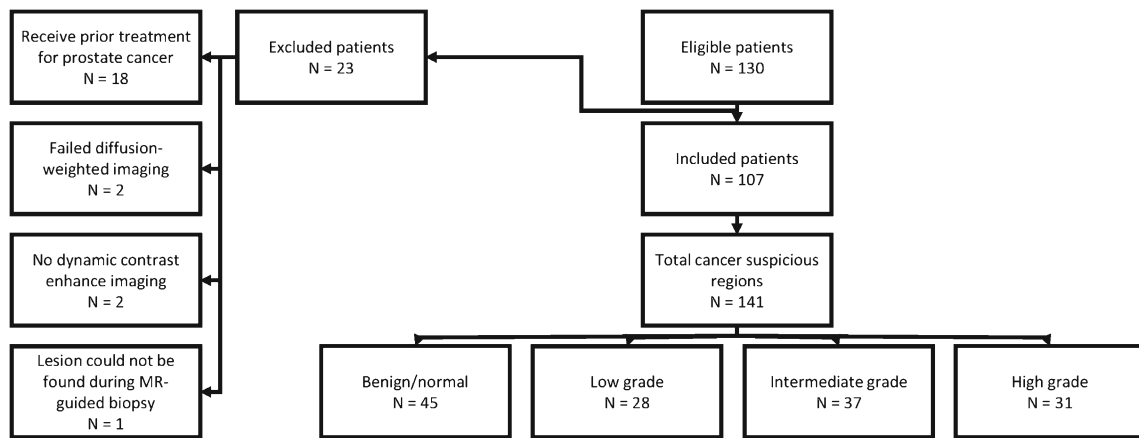
Third, we correlated radiologist PIRADS, CAD score and the combined score to cancer grade. As cancer grade is an ordinal variable, Spearman's rank correlation coefficient was used. The significance of differences in correlation



**Fig. 1** Suggested workflow for the proposed computer-aided diagnosis (CAD) system. The biopsy decision can be made by the radiologist, another attending clinician or by using the combination score to independently combine the PIRADS score and the CAD likelihood

**Table 4** Mapping of Gleason scores to cancer grade

Gleason scores	Grade	Category
None	Benign	Indolent
3+3 or lower, no 4 or 5 component	Low-grade	Indolent
2+4, 3+4, 2+5	Intermediate-grade	Aggressive
3+5, any cancer with a major 4 or 5 component	High-grade	Aggressive



**Fig. 2** STARD diagram of inclusion and exclusion criteria of the prospective patient cohort

coefficients was tested using Steiger’s z-test for dependent correlation coefficients [21]. For all significance tests a p-value threshold of 0.05 was chosen. SPSS (SPSS, version 20.0.01) and in-house developed tools for bootstrapping were used for all statistical analysis.

**Results**

**Patient data**

Of the initially included 130 patients, 23 were excluded, 18 due to previous treatment for prostate cancer, two for failed diffusion-weighted imaging, two because they did not undergo dynamic contrast-enhanced imaging and one patient because no biopsy was taken during the biopsy session. The 107 included patients had a median age and PSA level of 66 years (range 48–83) and 13 ng/mL (range 1–56),

respectively, which is similar to other studies using patient data with the same inclusion criteria (initial negative TRUS GB and persistently elevated PSA) [12–15]. Further details are summarized in Fig. 2 and Table 5.

In total 141 suspicious regions were identified in these patients. All these regions were biopsied under MR-guidance. Of these regions, 68 % were positive and 32 % were negative for prostate cancer. The zonal distribution of the lesions was almost equal, with 69 regions being located in the peripheral zone and 72 in the central gland. Gleason grades were 2+5 (1 %), 3+2 (2 %), 3+3 (27 %), 3+4 (38 %), 4+3 (13 %), 4+4 (5 %), 4+5 (10 %), 5+4 (3 %) and 5+5 (1 %). Further details about the distribution of the lesion grades can be found in Table 5.

**The effect of CAD on radiologist performance**

First, the logistic regression procedure showed that including the CAD system likelihood in addition to the radiologist PIRA

**Table 5** Characteristics of patients and biopsy specimens for the prospective cohort used to evaluate the potential added value of a computer-aided diagnosis system for the assessment of prostate cancer.

Number of patients	107			
PSA level, ng/ml, median (range)	13 (1–56)			
Age, y, median (range)	66 (48–83)			
Percentage of cancer per core, median (range)	50 (7–100)			
Gleason score	Grade	No. of lesions	Totals	141 (69/72)
Normal/Benign	Normal/Benign	45 (28/17)	No cancer	45 (28/17)
2+5	Intermediate	1 (0/1)	Low	28 (10/18)
3+2	Low	2 (0/2)	Intermediate	37 (16/21)
3+3	Low	26 (10/16)	High	31 (15/16)
3+4	Intermediate	36 (16/20)		
4+3	High	12 (7/5)		
4+4	High	5 (2/3)		
4+5	High	10 (4/6)		
5+4	High	3 (2/1)		
5+5	High	1 (0/1)		

DS score resulted in a model with significantly improved predictive power ( $p < 0.001$ , likelihood ratio test) for both evaluation settings (benign vs. cancer and indolent vs. aggressive). Using the obtained regression coefficients we created a weighted combination score:

$$\text{Combination score} = 4.739C + 1.06P$$

for the benign versus cancer setting and

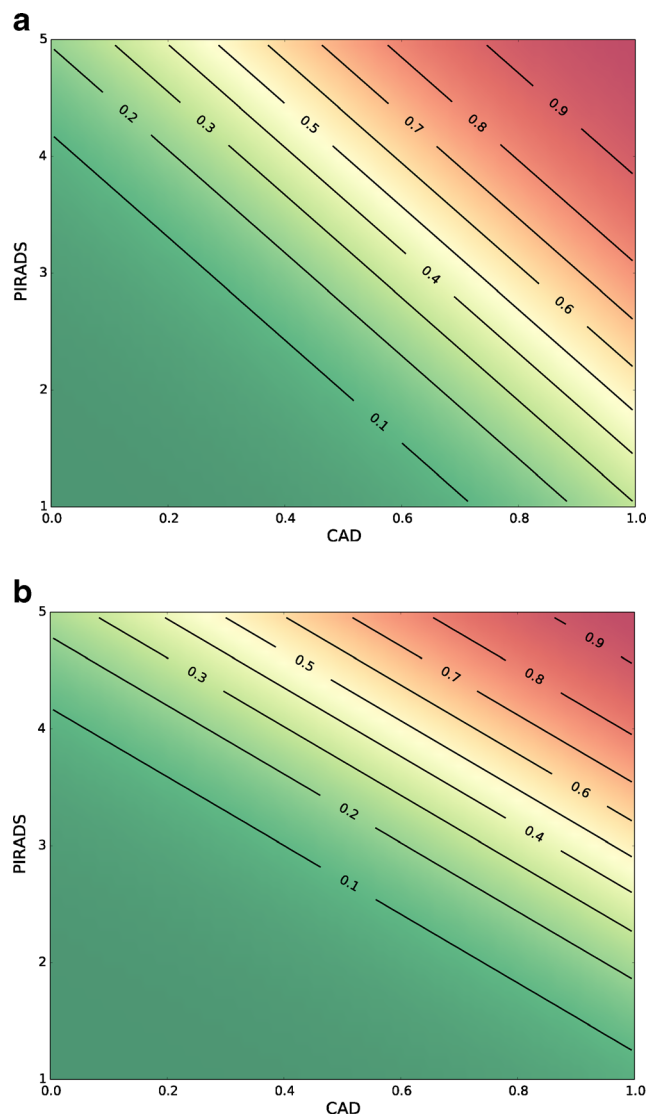
$$\text{Combination score} = 3.886C + 1.295P$$

for the indolent versus aggressive setting.  $C$  is the CAD system likelihood (ranging from 0 to 1) and  $P$  is the radiologist PIRADS score (ranging from 1 to 5). The regression models are visually represented in Fig. 3.

The ROC analyses showed a significant increase ( $p < 0.05$ ) in area under the ROC curve from 0.81 to 0.88 in the benign versus cancer setting and from 0.78 to 0.87 in the indolent versus aggressive setting when using the combination score versus only PIRADS (Fig. 4a, b and Table 6). Furthermore, this increase is not affected by the zonal location of the lesion under investigation (Fig. 4c, d, e and f and Table 7). The ROC analysis also shows that radiologists and CAD have a comparable diagnostic accuracy. Lastly, Table 6 also includes the increases in sensitivity and specificity obtained at specific PIRADS scores when using the combination score. For example, at PIRADS 4 without using CAD a sensitivity of 0.93 is attainable at a specificity of 0.37 in the indolent versus aggressive setting. However, when combining the PIRADS score with the CAD score we obtain a significantly increased sensitivity of 0.98 at a significantly increased specificity of 0.59 ( $p < 0.05$ ). Last, we show that both less experienced and experienced readers can improve their performance by using CAD. In the indolent versus aggressive evaluation setting both groups improve significantly when using CAD with an increase in AUC from 0.76 to 0.85 for less experienced and from 0.78 to 0.87 for experienced readers (Fig. 5, Table 8). In the benign versus cancer settings both groups also improve, but only the less experienced readers significantly (AUC from 0.79 to 0.89).

### Correlation of likelihood and cancer grade

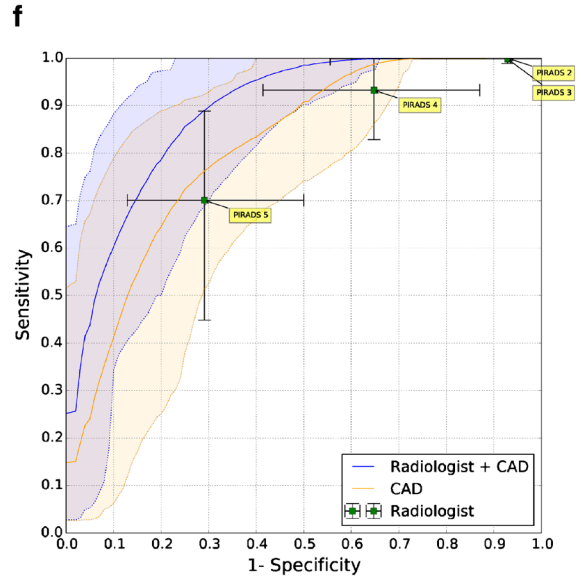
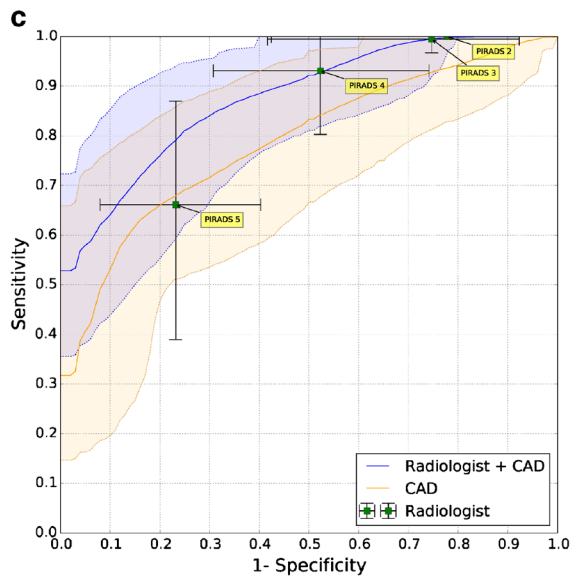
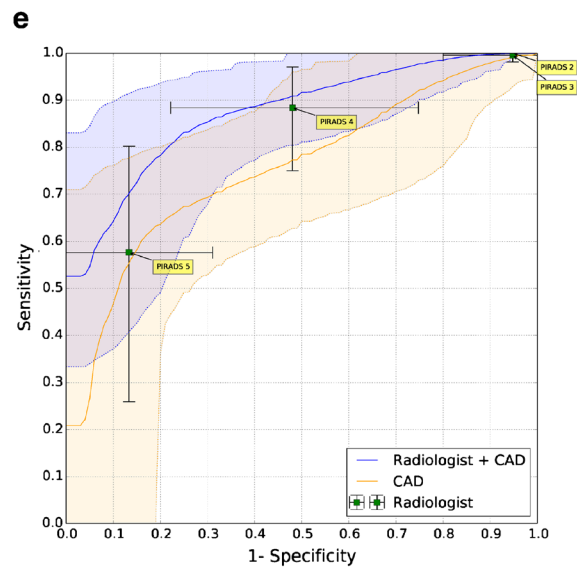
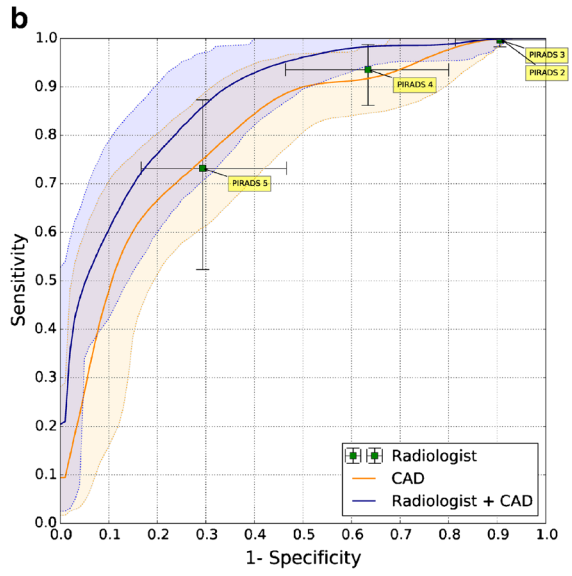
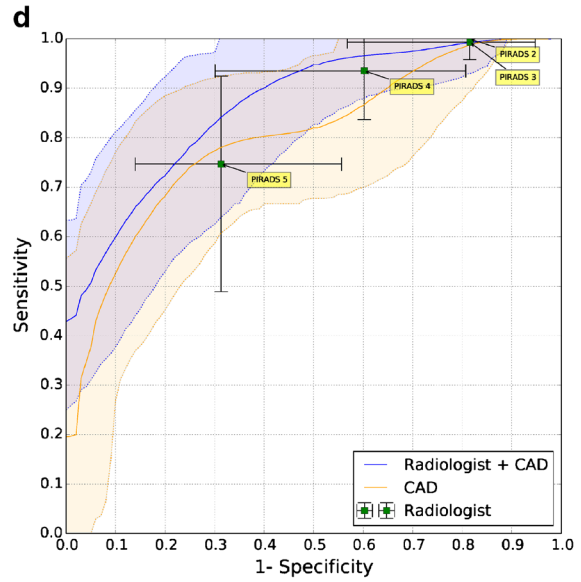
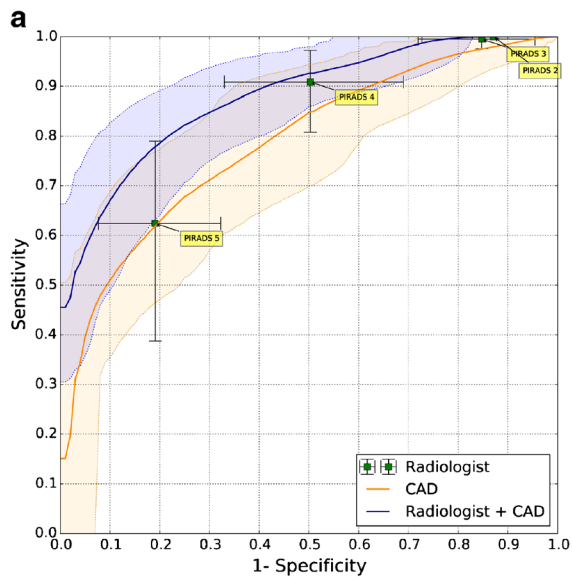
Both CAD likelihood and PIRADS score correlate significantly with cancer grade, but the combination score shows the strongest correlation. This is confirmed when assessing the correlation coefficients. In the benign versus cancer setting, correlation was 0.534, 0.582 and 0.684 for CAD, radiologist and combination, respectively. In the indolent versus aggressive setting the correlation coefficient was 0.536, 0.582 and 0.694 for CAD, radiologist and combination, respectively. The increase in correlation when using the combination score instead of just PIRADS or CAD was significant ( $p < 0.01$ ). The



**Fig. 3** Visual depictions of the regression models to generate the combination score of the radiologist and the computer-aided diagnosis (CAD) system: **(a)** shows the model for the benign vs. cancer setting, **(b)** for the indolent vs. aggressive setting. The likelihood of cancer is indicating by the colour coding and the contour labels and ranges from 0 to 1. Green indicates low likelihood and red indicates high likelihood

ability of each of the three scores to predict aggressiveness is visualized in Fig. 6 for both evaluation settings.

**Fig. 4** Receiver-operating characteristic (ROC) curve showing the performance of the computer-aided diagnosis (CAD) system (orange) and the radiologist/CAD-system combination (blue). The shaded areas indicated the 95 % confidence intervals (CIs) as calculated using bootstrapping. The radiologist performance is indicated with point for the different PIRADS thresholds. The vertical error bars indicate the 95 % CI on the sensitivity and horizontal error bars indicated the 95 % CI on the specificity as estimated by bootstrapping. **a, c and e** are the results of the benign versus cancer evaluation settings, **b, d and f** are the results of the indolent versus aggressive settings. **a and b** show the results over all lesions, **c and d** only the peripheral zone lesions and **e and f** only the central gland lesions



**Table 6** Sensitivity-specificity pairs and area under the receiver operating characteristic (ROC) curve for the radiologist and the computer-aided diagnosis (CAD)-radiologist combination including 95 % confidence intervals (CIs) and p-values determined bybootstrapping. Significant p-values ( $p < 0.05$ ) are indicated in bold. The first part of the table contains the results for the benign vs. cancer evaluation setting whereas the second part of the table contains the results for the indolent vs. aggressive evaluation setting

CAD+radiologist				Radiologist		
Benign vs. cancer	Sensitivities, mean (95 % CI)	p-value	Specificities, mean (95 % CI)	p-value	Sensitivities, mean (95 % CI)	Specificities, mean (95 % CI)
PIRADS 2	1.0 (1.0–1.0)	1	0.25 (0.0–0.44)	<b>0.008</b>	1.0 (1.0–1.0)	0.12 (0.0–0.27)
PIRADS 3	0.99 (0.99–1.0)	0.49	0.30 (0.17–0.45)	<b>0.02</b>	0.99 (0.98–1.0)	0.15 (0.04–0.28)
PIRADS 4	0.92 (0.84–0.98)	0.44	0.50 (0.26–0.72)	0.48	0.91 (0.81–0.97)	0.49 (0.30–0.67)
PIRADS 5	0.76 (0.61–0.88)	0.08	0.90 (0.77–0.98)	0.098	0.62 (0.38–0.79)	0.81 (0.67–0.92)
AUC	0.878 (0.824–0.928)	<b>0.013</b>			0.808 (0.728–0.880)	
CAD+radiologist				Radiologist		
Indolent vs. aggressive	Sensitivities, mean (95 % CI)	p-value	Specificities, mean (95 % CI)	p-value	Sensitivities, mean (95 % CI)	Specificities, mean (95 % CI)
PIRADS 2	1 (1–1)	1	0.259 (0.00–0.604)	<b>0.023</b>	1 (1–1)	0.094 (0.0–0.185)
PIRADS 3	0.99 (0.98–1.0)	0.51	0.259 (0.00–0.604)	<b>0.023</b>	0.997 (0.983–1.0)	0.094 (0.0–0.185)
PIRADS 4	0.98 (0.94–1.0)	<b>0.029</b>	0.585 (0.379–0.763)	<b>0.013</b>	0.934 (0.861–0.98)	0.366 (0.200–0.536)
PIRADS 5	0.82 (0.68–0.96)	0.09	0.78 (0.64–0.90)	0.105	0.731 (0.523–0.873)	0.707 (0.534–0.833)
AUC	0.874 (0.813–0.927)	<b>0.001</b>			0.779 (0.701–0.848)	

## Discussion

A recently developed CAD system for the computerized analysis of prostate MR was shown to have a similar diagnostic accuracy to well trained prostate MR radiologists. When combined with the PIRADS score into a combination score, diagnostic performance improved significantly. It is important to note that the CAD system was used in a regular clinical practice setting on a large cohort. This provides further evidence of the benefit of CAD in helping improve diagnostic accuracy.

The CAD system score showed a significant correlation (0.54) with cancer grade similar to the PIRADS score (0.58). A significantly higher correlation (0.69) was obtained by using the combination score. This correlation coefficient is the highest currently reported in literature [18, 22]. A noticeable difference with other multivariate aggressiveness correlation studies is that this study attains the correlation in a setting with radiologist-indicated regions instead of pathology pre-defined regions of interest, which is more similar to regular clinical practice.

Although the performance of both the radiologist and the CAD system were similar, they provide complementary information, as the combination score results in an improved ROC curve (blue curve, Fig. 4a, b). In both evaluation settings (benign vs. cancer and indolent vs. aggressive) the area under the ROC curve increased significantly (0.81 to 0.88,  $p=0.013$  and 0.78 to 0.87,  $p=0.001$ , respectively). If we compare these results to those found in the literature, we observe similar values for performance of readers with CAD; however, the PIRADS performance is somewhat lower in our study (0.84 to 0.87 in [23], 0.85 to 0.91 in [24]). We believe these

differences to be caused by the difference in reading setting and the fact that we used a substantially larger and more difficult cohort. Our study used prospective clinical reading, whereas these previous studies used a retrospective batch reading setting. Furthermore, previous studies were limited to patients scheduled for radical prostatectomy, which differs from the regular clinical population used in this study.

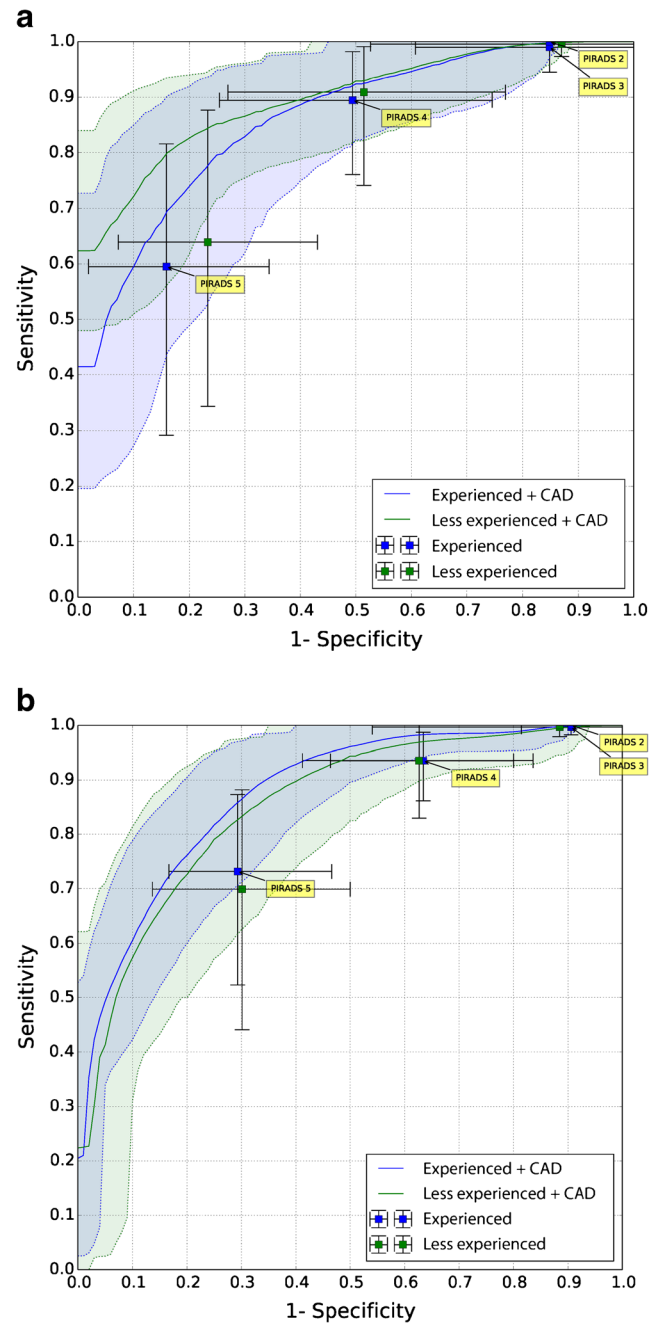
Furthermore, we showed that identifying clinically significant disease using CAD can improve both experienced and less experienced readers (increase of 0.09 in AUC for both groups). However, when discriminating any cancer from benign lesions only less experienced readers improved significantly, indicating that CAD might especially be helpful in identifying clinically significant disease. Note that direct comparison between the performance of less experienced and experienced readers cannot be performed in this study as both groups reported on different patients. However, we do not expect results to change as both groups evaluated a sufficiently large and similar subset of cases.

The added value of CAD did not seem to depend on the zonal location of the lesions. Increase in performance when using CAD was similar for both peripheral zone and central gland lesions. The CAD system performed equally well on peripheral zone and central gland lesions, indicating that it has successfully learned how to take into account the zonal characteristics of the lesions. CAD could possibly help shift the biopsy threshold from PIRADS 3 to PIRADS 4, which would lead to a significant reduction of MR-guided biopsies. At a PIRADS threshold of 3, CAD significantly increased the specificity in both evaluation settings (Table 1; 0.15 to 0.30,  $p=0.020$  and 0.09 to 0.26  $p=0.023$ ). At the PIRADS 4



**Table 7** Receiver operating characteristic (ROC) analysis differentiated with respect to the zonal location of the lesions. Area under the ROC curve is reported for both the benign vs. cancer and indolent vs. aggressive evaluation settings. P-values measuring whether the increase in area under the ROC curve is significant when using computer-aided diagnosis (CAD) were calculated using bootstrapping. Significant p-values are indicated in bold

Area under the ROC curve	Radiologist (PZ)	Radiologist (CG)	CAD (PZ)	CAD (CG)	Combined score (PZ)	Combined score (CG)	p-value	p-value
Benign vs. cancer	0.81 (0.70–0.90)	0.83 (0.72–0.92)	0.79 (0.67–0.89)	0.76 (0.62–0.87)	0.88 (0.80–0.94)	0.87 (0.78–0.95)	<b>0.04</b>	0.15
Indolent vs. aggressive	0.79 (0.69–0.88)	0.77 (0.66–0.87)	0.80 (0.70–0.89)	0.81 (0.69–0.91)	0.87 (0.79–0.93)	0.89 (0.80–0.95)	<b>0.002</b>	<b>0.002</b>



**Fig. 5** Receiver-operating characteristic (ROC) curve showing the performance of the combined score vs. the radiologist alone with respect to the level of experience. The shaded areas indicated the 95 % confidence intervals (CIs) as calculated using bootstrapping. The radiologist performance is indicated with point for the different PIRADS thresholds. The vertical error bars indicate the 95 % CIs on the sensitivity and horizontal error bars indicated the 95 % CI on the specificity as estimated by bootstrapping. **a** is the result of the benign versus cancer evaluation setting, **b** is the result of the indolent versus aggressive settings

threshold a significant increase in sensitivity and specificity was found (0.93 to 0.98,  $p=0.029$ , 0.37 to 0.59,  $p=0.013$ ) in the indolent versus aggressive setting. The latter indicates that by using CAD and increasing the biopsy threshold from PIRA

**Table 8** Receiver operating characteristic (ROC) analysis comparing inexperienced and experienced readers (less or more than 8 years of experience with prostate MRI) when using computer-aided diagnosis (CAD). Area under the ROC curve including 95 % confidence intervals (CIs) are reported for both the benign vs. cancer and indolent vs. aggressive evaluation settings. P-values measuring whether the increase in area under the ROC curve is significant when using CAD were calculated using bootstrapping. Significant p-values are indicated in bold

Area under the ROC curve	Less experienced readers	Experienced readers	Combined score of inexperienced readers+CAD	p-value	Combined score of experienced readers+CAD	p-value
Benign vs. cancer	0.79 (0.69–0.90)	0.82 (0.71–0.92)	0.89 (0.82–0.96)	<b>0.004</b>	0.86 (0.77–0.94)	0.25
Indolent vs. aggressive	0.76–(0.66–0.86)	0.78 (0.70–0.85)	0.85 (0.76–0.93)	<b>0.006</b>	0.87 (0.81–0.93)	<b>0.001</b>

DS 3 to 4 almost no loss in detection of aggressive lesions (sensitivity 0.99 vs. 0.98) occurs, while a specificity improvement is obtained (0.59 vs. 0.26), reducing unnecessary biopsies. Of course this has to be investigated further in future clinical studies. One important aspect relevant to patient prognosis was not assessed by the CAD system; the presence or absence of extracapsular extension. Being able to identify this aspect would further enhance the applicability of the CAD system.

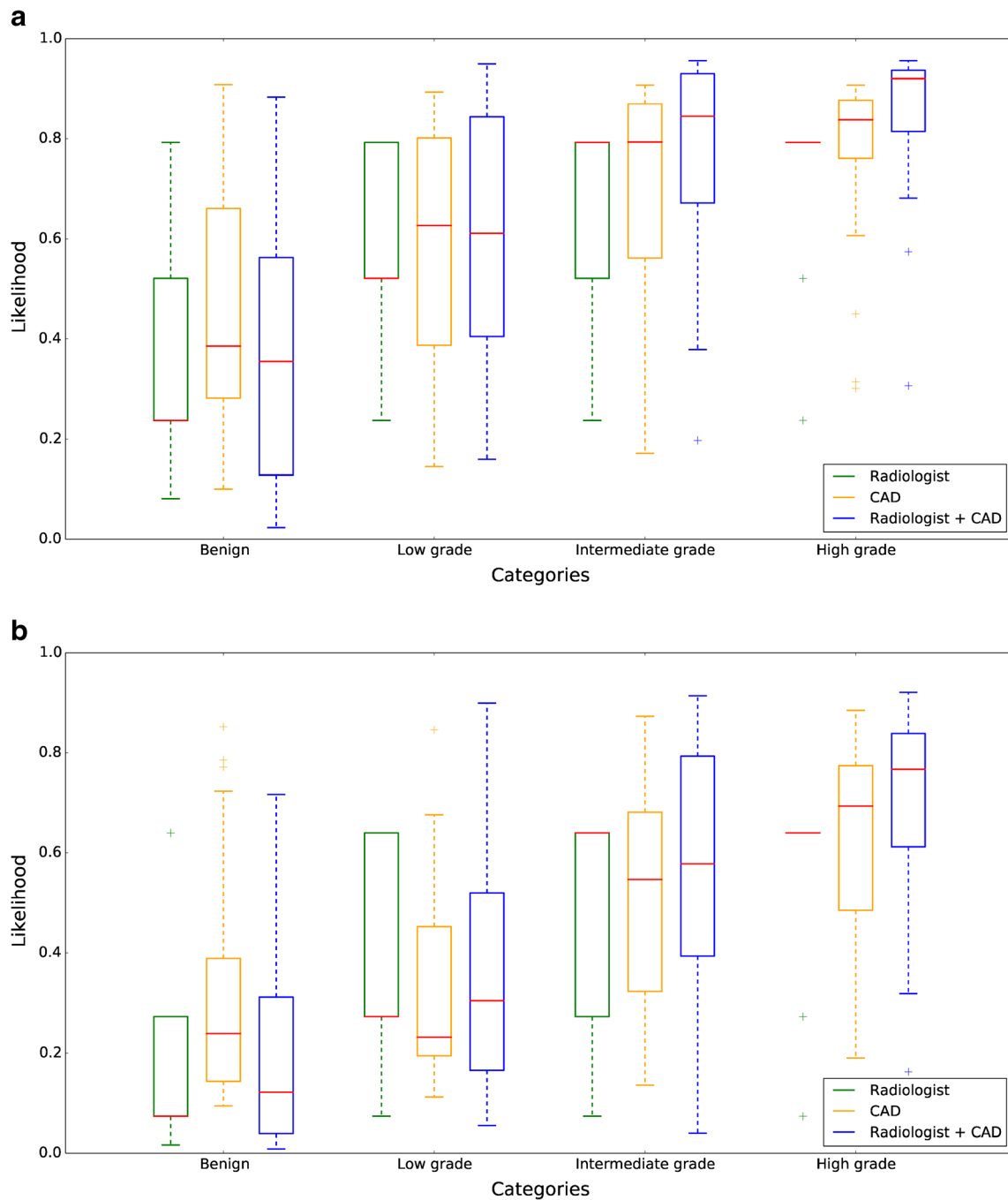
This study has some limitations. Firstly, MR-guided biopsy has a 95 % concordance with prostatectomy Gleason grade for Gleason 4 and 5 components [20, 25] and has been shown to be able to detect clinically significant cancer in men with previous negative TRUS biopsies [26]. Although in general concordance rates with prostatectomy Gleason grade are high, they are not perfect. Thus, some of the cancers in our study may be under- or overgraded. We expect the effect on our results to be minimal, as this only effects the indolent versus aggressive setting.

A second limitation is that each case in this study was read by one of seven radiologists. We know diagnostic accuracy is dependent on reader experience and thus our results depend on the average reader experience of the group. All our readers have had reasonable training and experience of at least 2 years [8, 11].

Third, due to the single reader per case and the prospective reading setting inter- and intraobserver variability and false-negative rates could not be assessed. As such, this study does not vacate the need for retrospective observer studies, in which these aspects could be assessed, but provides a different, more clinically realistic view on the added value of CAD. Furthermore, due to our comparatively large number of cases we were still able to show significantly improved diagnostic performance when using CAD.

The proposed method of implementing CAD in clinical practice (independent combination of PIRADS score and CAD likelihood into a combination score) might not be feasible, as radiologists or urologists will always have the final say. Nevertheless, we choose to perform independent combination to assess the potential observer-independent effect of CAD. In future work, one could assess the most optimal way for radiologists to incorporate CAD results in their reports.

PIRADS 1 and 2 lesions were generally not biopsied and therefore are only partially included in this study. This precludes assessment of the effect of the CAD system in those lesions. However, this has little impact on the results of this study. Of all seven biopsied PIRADS 2 lesions, none were categorized as cancer. The negative predictive value of PIRADS 1 and 2 scores is already so high that radiologist do not need computer aid for these PIRADS scores. The literature also confirms this assessment, with the studies by Thompson



**Fig. 6** Relationship between computer-aided diagnosis (CAD) system likelihood and cancer grade presented by box-plots. One can observe a positive correlation between cancer grade and CAD system likelihood. **a**

is the result of the benign vs. cancer evaluation setting, **b** is the result of the indolent vs. aggressive settings

et al. [8] and Pokorny et al. [11] reporting MRI sensitivities and negative predictive values of 97 and 96.9 %, respectively. The CAD system has most potential in more accurately discriminating which PIRADS 3, 4 or 5 lesions require biopsy and the results at these scores are not affected by the lack of PIRADS 1 or 2 biopsies.

Last, the patient population in this study contained only patients for whom initial PSA tests and TRUS biopsies were

inconclusive. As such, the results of this study cannot be directly translated to other patient groups (e.g., staging). However, due to the similar protocols for detection and staging MRI we expect results to be comparable. Furthermore, with prostate cancer guidelines in many countries now recommending MRI if PSA/TRUS results are inconclusive, we expect that the majority of prostate MRIs will be done for detection purposes.

## Conclusions

In this paper the use of a CAD system in conjunction with the radiologist to accurately characterize prostate lesions was investigated. Results showed that a significant increase in diagnostic performance can be achieved when combining the radiologist PIRADS score and CAD system likelihood into a combination score. Furthermore, a significant correlation between CAD likelihood and cancer grade exists; this increases further when using the combination score.

**Acknowledgments** The scientific guarantor of this publication is Henkjan Huisman. The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article. This study has received funding by the Dutch Cancer Society, grant KUN2007-3971. One of the authors has significant statistical expertise. Institutional Review Board approval was not required because using imaging data from regular care for research is allowed under a waiver of the Institutional Review Board. Written informed consent was waived by the Institutional Review Board. Methodology: retrospective with prospective reading by radiologists, experimental, performed at one institution.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G et al (2012) ESUR prostate MR guidelines 2012. *Eur Radiol* 22:746–757
- Axel H, Joaquim B, Michel B, Steven J, Malcolm M, Vsevolod M et al (2011) EAU guidelines on prostate cancer. part 1: screening, diagnosis, and treatment of clinically localised disease. *Eur Urol* 59: 61–71
- Thompson J, Lawrentschuk N, Frydenberg M, Thompson L, Stricker P (2013) The role of magnetic resonance imaging in the diagnosis and management of prostate cancer. *BJU Int* 112:6–20
- Hoeks CMA, Barentsz JO, Thomas H, Derya Y, Somford DM, Heijmink SWTPJ et al (2011) Prostate cancer: Multiparametric MR imaging for detection, localization, and staging. *Radiology* 261:46–66
- Lawrentschuk N, Fleshner N (2009) The role of magnetic resonance imaging in targeting prostate cancer in patients with previous negative biopsies and elevated prostate-specific antigen levels. *BJU Int* 103:730–733
- Siddiqui MM, Rais-Bahrami S, Truong H, Stamatakis L, Vourganti S, Nix J et al (2013) Magnetic resonance imaging/ultrasound-fusion biopsy significantly upgrades prostate cancer versus systematic 12-core transrectal ultrasound biopsy. *Eur Urol* 64:713–719
- Noboru N, Soichiro Y, Chikako I, Yoshinobu K, Toshiki K, Minato Y et al (2013) Potential of prebiopsy multiparametric magnetic resonance imaging to reduce initial biopsies in men with suspected clinically localized prostate cancer. *J Urol* 189:e602, ISSN 0022–5347. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022534713032096?showall=true>
- James T, Phillip S, Phillip B, Daniel M, Ron S, Warick D et al (2013) Magnetic resonance imaging detects significant prostate cancer and could be used to reduce unnecessary biopsies: initial results from a prospective trial. *J Urol* 189:e910–e911, ISSN 0022–5347. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022534713024051?showall=true>
- Kirkham APS, Emberton M, Allen C (2006) How good is MRI at detecting and characterising cancer within the prostate? *Eur Urol* 50:1163–1175
- Dickinson L, Hashim U, Ahmed CA, Jelle O, Barentsz BC, Jurgen J et al (2011) Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. *Eur Urol* 59:477–494
- Pokorny MR, de Rooij M, Duncan E, Schröder FH, Parkinson R, Barentsz JO et al (2014) Prospective study of diagnostic accuracy comparing prostate cancer detection by transrectal ultrasound-guided biopsy versus magnetic resonance (MR) imaging with subsequent MR-guided biopsy in men without previous prostate biopsies. *Eur Urol* 66:22–29
- Portalez D, Mozer P, Cornud F, Renard-Penna R, Misrai V, Thoulouzan M et al (2012) Validation of the European Society of Urogenital Radiology scoring system for prostate cancer diagnosis on multiparametric magnetic resonance imaging in a cohort of repeat biopsy patients. *Eur Urol* 62:986–996
- Rosenkrantz AB, Kim S, Lim RP, Hindman N, Deng F-M, Babb JS et al (2013) Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert scales. *Radiology* 269:482–492
- Rosenkrantz AB, Lim RP, Haghighi M, Somberg MB, Babb JS, Taneja SS (2013) Comparison of interreader reproducibility of the prostate imaging reporting and data system and likert scales for evaluation of multiparametric prostate MRI. *AJR Am J Roentgenol* 201:W612–W618
- Schimmöller L, Quentin M, Arsov C, Lanzman RS, Hiester A, Rabenalt R et al (2013) Inter-reader agreement of the ESUR score for prostate MRI using in-bore MRI-guided biopsies as the reference standard. *Eur Radiol* 23:3185–3190
- Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H (2014) Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging* 33:1083–1092
- Litjens G, Debats OA, van de Ven WJM, Karssemeijer N, Huisman HJ (2012) A pattern recognition approach to zonal segmentation of the prostate on MRI. *Med Image Comput Comput Assist Interv* 7511:413–420, of *Lect Notes Comput Sci*
- Vos EK, Litjens G, Thiele K, Thomas H, Kaa CA, Barentsz JO et al (2013) Assessment of prostate cancer aggressiveness using dynamic contrast-enhanced magnetic resonance imaging at 3 t. *Eur Urol* 64:448–455
- Hambrock T, Somford DM, Huisman HJ, van Oort IM, Witjes JA, Hulsbergen-van de Kaa CA et al (2011) Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. *Radiology* 259:453–461
- Hambrock T, Hoeks C, Hulsbergen-van de Kaa C, Scheenen T, Fütterer J, Bouwense S et al (2012) Prospective assessment of prostate cancer aggressiveness using 3-T diffusion-weighted magnetic resonance imaging-guided biopsies versus a systematic 10-core transrectal ultrasound prostate biopsy cohort. *Eur Urol* 61:177–184
- Steiger JH (1980) Tests for comparing elements of a correlation matrix. *Psychol Bull* 87:245–251
- Thiele K, Vos PC, Thomas H, Maarten De R, de Kaa CAH-V, Barentsz JO et al (2012) Prostate cancer aggressiveness: In vivo assessment of MR spectroscopy and diffusion-weighted imaging at 3 t. *Radiology* 265:457–467
- Niaf E, Lartizien C, Bratan F, Roche L, Rabilloud M, MÃ“ge-Lechevallier F, RouviÃ“re O (2014) Prostate focal peripheral zone

- lesions: Characterization at multiparametric MR imaging—influence of a computer-aided diagnosis system. *Radiology* :130448. doi: [10.1148/radiol.14130448](https://doi.org/10.1148/radiol.14130448)
24. Thomas H, Vos PC, de Kaa CAH-v, Barentsz JO, Huisman HJ (2013) Prostate cancer: Computer-aided diagnosis with multiparametric 3-t MR imaging—effect on observer performance. *Radiology* 266:521–530
  25. Moore CM, Robertson NL, Nasr A, Thomas M, Arnauld V, Laurence K et al (2013) Image-guided prostate biopsy using magnetic resonance imaging-derived targets: a systematic review. *Eur Urol* 63:125–140
  26. Roethke M, Anastasiadis AG, Lichy M, Werner M, Wagner P, Kruck S et al (2012) Mri-guided prostate biopsy detects clinically significant cancer: analysis of a cohort of 100 patients after previous negative trus biopsy. *World J Urol* 30(2):213–218
  27. Li Q, Sone S, Doi K (2003) Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. *Med Phys* 30:2040–2051