MAGNETIC RESONANCE

Logistic regression model for diagnosis of transition zone prostate cancer on multi-parametric MRI

Nikolaos Dikaios · Jokha Alkalbani · Harbir Singh Sidhu · Taiki Fujiwara · Mohamed Abd-Alazeez · Alex Kirkham · Clare Allen · Hashim Ahmed · Mark Emberton · Alex Freeman · Steve Halligan · Stuart Taylor · David Atkinson · Shonit Punwani

Abstract

Objectives We aimed to develop logistic regression (LR) models for classifying prostate cancer within the transition zone on multi-parametric magnetic resonance imaging (mp-MRI).

Methods One hundred and fifty-five patients (training cohort, 70 patients; temporal validation cohort, 85 patients) underwent mp-MRI and transperineal-template-prostate-mapping (TPM) biopsy. Positive cores were classified by cancer definitions: (1) any-cancer; (2) definition-1 [≥Gleason 4+3 or≥6 mm cancer core length (CCL)] [high risk significant]; and (3) definition-2 (≥Gleason 3+4 or≥4 mm CCL) cancer [intermediate–high risk significant]. For each, logistic-regression mp-MRI models were derived from the training cohort and validated internally and with the temporal cohort. Sensitivity/specificity and the area under the receiver operating characteristic (ROC-AUC) curve were calculated. LR model performance was compared to radiologists' performance.

Results Twenty-eight of 70 patients from the training cohort, and 25/85 patients from the temporal validation cohort had significant cancer on TPM. The ROC-AUC of the LR model for classification of cancer was 0.73/0.67 at internal/temporal validation. The radiologist A/B ROC-AUC was 0.65/0.74 (temporal cohort). For patients scored by radiologists as Prostate Imaging Reporting and Data System (Pi-RADS) score 3, sensitivity/specificity of radiologist A 'best guess' and LR model was 0.14/0.54 and 0.71/0.61, respectively; and radiologist B 'best guess' and LR model was 0.40/0.34 and 0.50/0.76, respectively.

Conclusions LR models can improve classification of Pi-RADS score 3 lesions similar to experienced radiologists.

Key Points
- *MRI helps find prostate cancer in the anterior of the gland*
- *Logistic regression models based on mp-MRI can classify prostate cancer*
- *Computers can help confirm cancer in areas doctors are uncertain about*

Nikolaos Dikaios and Shonit Punwani are contributed equally towards the manuscript

N. Dikaios · J. Alkalbani · H. S. Sidhu · T. Fujiwara · S. Halligan · S. Taylor · D. Atkinson · S. Punwani (✉)
Centre for Medical Imaging, University College London, 3rd Floor East Wing, 250 Euston Road, London, UK NW1 2PG
e-mail: shonit.punwani@gmail.com

N. Dikaios · A. Kirkham · C. Allen · S. Halligan · S. Taylor · D. Atkinson · S. Punwani
Departments of Radiology, University College London Hospital, 235 Euston Road, London, UK NW1 2BU

M. Abd-Alazeez · H. Ahmed · M. Emberton
Research Department of Urology, University College London, London, UK NW1 2PG

A. Freeman
Department of Histopathology, University College London Hospital, London, UK NW1 2PG

Abbreviations

PSA	Prostate specific antigen
TRUS	Transrectal ultrasound guided biopsy
CAD	Computer assisted diagnosis
LR	Logistic regression
TPM	Template mapping biopsy
DWI	Diffusion weighted imaging
ADC	Apparent diffusion coefficient
DCE	Dynamic contrast enhanced
mp	Multi-parametric
SI	Signal intensity
T2-nSI	Normalized T2 signal intensity
DCE-nSI	Early contrast enhanced T1 signal intensity

SoE	Slope of enhancement
ME	Maximum enhancement
Etype	Curve type
AUCtot	Total area under the dynamic contrast enhanced signal intensity time curve
ROC	Receiver operator characteristic
ROI	Region of interest
AUC	Area under curve

Introduction

Increasing recognition of anteriorly located prostate cancers has resulted from the gradual transition of a diagnostic pathway reliant primarily on prostate specific antigen (PSA) and transrectal ultrasound (TRUS) guided biopsy—a strategy that selectively identifies peripheral zone cancers—to one that incorporates imaging, notably multi-parametric magnetic resonance imaging (mp-MRI) [1, 2].

Although mp-MRI facilitates identification of more anterior cancers, its overall diagnostic accuracy is likely lower in this part of the prostate compared to the posteriorly located peripheral zone [3, 4]. The reason for this relates to benign hyperplasia nodules. Against the heterogeneous signal of an expanded transition zone, tumours are much more difficult to differentiate [5]. In other cancer types (colorectal [6], lung [7] and breast [8]), discriminating tumours has been in part aided by computer-assisted diagnosis (CAD). This term refers to software that prompts observers towards potential abnormalities. However, most developed prostate CAD relies upon the evaluation of a single MRI parameter informed by data exclusively from the peripheral zone [3, 9–12]. It is possible that when applied to the transition zone, the accuracy of these diagnostic models is compromised as a result [13, 14]. Indeed, recent work suggests that a multi-parametric CAD model may provide improved results [15].

Our aim is to derive and validate logistic regression (diagnostic) models for classifying transition zone prostate cancer on multi-parametric MRI (mp-MRI).

Material and methods

Our local institutional review board approved the study and waived the requirement for individual consent for retrospective analysis of patient data collected as part of clinical trials/ routine care (R&D No: 12/0195 date: 16 July 2012).

Patient cohort

The cohort comprised men with clinically suspected prostate cancer (elevated PSA ± abnormal digital rectal examination ± family history of prostate cancer, etc.) undergoing prostatic mp-MRI prior to template-prostate-mapping (TPM) biopsies. Two temporally separated patient cohorts were identified: February 2007 to September 2010 (training cohort) and October 2010 to January 2013 (temporal validation cohort).

Patients were identified from a total of 509 TPM clinical records. Men who had undergone TPM biopsy within 6 months prior to mp-MRI, or previous treatment for prostate cancer, were excluded to avoid biopsy-associated MRI signal changes. Men with incomplete mp-MRI data sets were also excluded.

In total, 155 men were eligible. Seventy men (mean age 61 years, range 42–78) with a mean PSA of 8.39 ng/ml (range 1.2–40 ng/ml) and mean prostate gland volume of 44.1 ml (range 18.9–101 ml) comprised the training cohort. Eighty-five men (mean age 63 years, range 45–77) with a mean PSA of 8.66 ng/ml (range 0.2–39 ng/ml) and mean prostate gland volume of 44.1 ml (range 20–133 ml) comprised the temporal validation cohort.

Multi-parametric magnetic resonance imaging

Subjects underwent mp-MRI using a 1.5 T static magnet (Avanto, Siemens, Erlangen, Germany) and pelvic-phased array coil; 0.2 mg/kg (maximum 20 mg) of spasmolytic (Buscopan; Boehringer Ingelheim, Ingelheim, Germany) was administered intravenously to reduce peristalsis. The mp-MRI included axial and coronal small field of view T2-weighted imaging; and was supplemented by axial diffusion weighted imaging (DWI) and dynamic contrast enhanced (DCE) imaging (Table 1).

Transperineal template-prostate-mapping biopsy

TPM was conducted following mp-MRI over a median time of 127 days (range 17–332 days) for the training cohort, and over a mean time of 75 days (range 13–384 days) for the temporal validation cohort. The biopsy was performed as previously reported [16, 17]. Briefly, systematic biopsy of the whole gland was conducted through a brachytherapy template-grid and 5 mm sampling frame, conforming to a sampling density of approximately 1 core/cc of tissue uniformly over the prostate. Biopsies were grouped and potted into 20 zones, modified from a technique reported by Barzell et al. [18]; each zone, with the exception of the lateral zone, had both apical and basal samples.

Histopathology review

Three differing definitions of cancer significance were used for the study, reflecting the lack of current consensus regarding clinically important disease (i.e., disease that warrants treatment and therefore detection). The first was 'any-cancer', with the second and third definitions using different thresholds

Table 1 Multi-parametric MRI sequence parameters

	T2w TSE axial / coronal	EPI-DWI	T1 3D FLASH*
Repetition time (ms)	5170 / 5240	2100	5.61
Echo time (ms)	92 / 104	98	2.5
Flip angle (degrees)	180 / 150	90	15
Echo train length	22 / 24	172	n/a
Bandwidth (Hz/Px)	190 / 190	968	300
Field of view (mm)	180 / 180	260	260
Phase FoV %	100 / 100	100	100
Slice thickness (mm)	3 / 3	5	3
Slice gap (mm)	0.3 / 0.3	0	0.6
Averages	2 / 2	16	1
Phase encoding direction	A>P / R>L	A>P	A>P
Fat saturation	No / No	Yes	Yes
Base matrix	256 / 256	172	192
Matrix phase %	95 / 95	100	100
b-values (s.mm^{-2})	n/a	0, 300, 500, 1000	n/a
No. of acquisitions	1 / 1	1	35
Temporal resolution (seconds)	n/a	n/a	16
Total acquisition time (mins)	3 m 54 s / 4 m 18 s	3 m 39 s	10 m

*dynamic contrast enhanced MRI – 0.2 ml/kg intravenous gadolinium contrast agent injected at the beginning of acquisition 6 at 3 ml/s followed by a saline flush of 20 ml

T2w TSE – T2 weighted turbo spin echo; EPI-DWI – Echo planar imaging - diffusion weighted imaging; FLASH – Fast low angle shot

Fig. 1 Standard modified Barzell zone schematic used for reporting of template biopsy results. Green disease (any-cancer definition) defined as ≤ Gleason 3+3 and a cancer core length of <4 mm; yellow disease (definition-2 cancer definition) as ≥ Gleason 3+4 or a cancer core length of ≥4 mm but <6 mm; and red (definition-1 cancer definition) as ≥ Gleason 4+3 or a cancer core length of ≥6 mm. Barzell zones 9 (right-medial-anterior-apical) and 3 (right-parasagittal-anterior-apical) were grouped into a *right anterior apical*; 7 (left-medial-anterior-apical) and 1 (left-parasagittal-anterior-apical) into *left anterior apical*; 10 (right-medial-anterior-basal) and 4 (right-parasagittal-anterior-basal) into *right anterior basal*; and 8 (left-medial-anterior-basal) and 2 (left-parasagittal-anterior-basal) into *left anterior basal* quadrants

for significant cancer based on widely used volume assessments (0.5 ml and 0.2 ml), combined with dominant/non-dominant Gleason grade 4. TPM cancer core length (CCL) values that best predict volumes were used to infer volume as described previously, where a TPM CCL of ≥4 mm signifies a tumour volume of ≥0.2 ml and a TPM CCL of ≥6 mm signifies a tumour volume of ≥0.5 ml [19]. For clinically significant disease; definition-1 [high-risk] cancer was defined as ≥ Gleason 4+3 or ≥6 mm CCL; and definition-2 [intermediate and high-risk] as ≥ Gleason 3+4 or ≥4 mm CCL.

For each patient, the modified Barzell zones were grouped into four anterior sectors (Fig. 1). Histopathology from the posterior and lateral Barzell zones was disregarded, as this predominantly represented the peripheral zone. Anterior sectors were classified as either positive or negative for each of the three cancer definitions of any-cancer, definition-1 cancer and definition-2 cancer [19].

Training cohort review

Images were viewed using Osirix (version. 3.5.1; Geneva, Switzerland). Two radiologists with 5 and 7 years of mp-
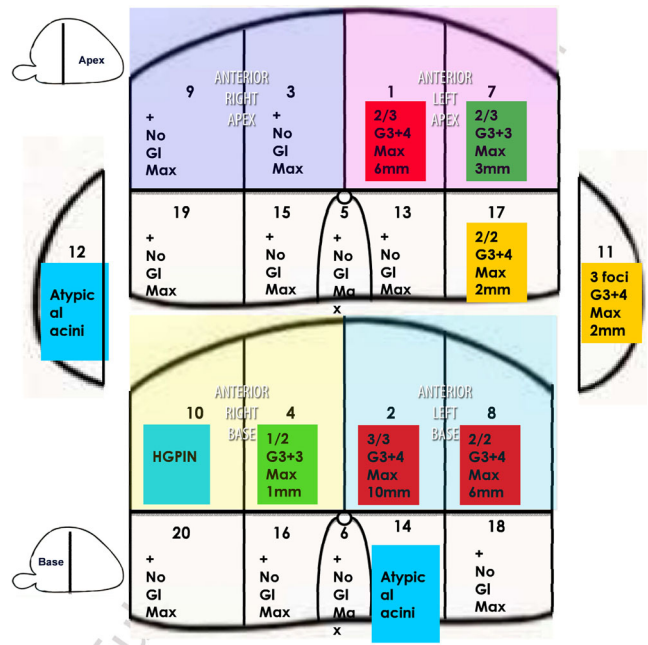
MRI experience (300–400 mp-MRI prostate scans/year each) reviewed each image in consensus, while being aware of pathological status of each patient and having access to histopathology reports.

Observers visually matched mp-MRI findings with the histological reference. For the purposes of the study, assessment was limited to the transition zone left/right apical and basal sectors only. Apical/basal sectors were split at half the cranio-caudal height of the prostate; anterior/posterior sectors at the level of the urethra; right/left sides by the midline; and lateral sectors were defined as extending 2 cm from the midline or in smaller glands up to the medial highest extent of the anterior horn of the peripheral zone.

T2-weighted, DWI and DCE images were reviewed simultaneously, and the transition zone focus most suspicious for disease (as confirmed on the histopathology report) was identified as per Prostate Imaging Reporting and Data System (Pi-RADS) criteria [20] and contoured. Where no transition zone tumour was identified on mp-MRI despite positive histopathology, a 1-cm^2 region of interest (ROI) was located within the sector pathologically reported as containing the most significant tumour definition. Where no tumour was present

within anterior sectors on histopathology, a 1-cm^2 ROI was located within any one of the four sectors to encompass a representative area of non-malignant tissue.

Consequently, a single ROI confined to an individual sector was contoured within the transition zone for each patient and a record was made of the positive/negative status of the ROI for each cancer definition.

Calculation of mp-MRI quantitative parameters of the training cohort

The mean signal intensity (SI) of the ROI on T2-weighted, early arterial contrast enhanced images and corresponding apparent diffusion coefficient (ADC) images were recorded for each patient. Additionally, the mean SI of the obturator internus muscle was measured on T2-weighted and early arterial contrast-enhanced images for normalization of T2 and contrast-enhanced transition zone SIs between patients. Normalized T2 (T2-nSI) and early contrast enhanced T1 SI (DCE-nSI) were derived by taking the ratio of prostate/muscle ROI SI. Quantitative parameters were derived from the DCE MRI signal enhancement time curve. Initial slope of enhancement (SoE), maximum enhancement (ME), curve type (Etype), and total area under the time-intensity curve (AUCtot) were defined as previously reported [21]. Briefly, ME was defined as (peak enhancement SI – baseline SI)/ baseline SI; and SoE as the gradient of the SI time curve between the last point of baseline SI and the inflexion point following initial rapid enhancement. Finally, Etype was categorized as type 1 if the SI increased with no subsequent decrease or levelling (threshold as late enhancement>10 % of ME); and type 3 if SI increased rapidly and then decreased after a peak (threshold as late enhancement<10 % of ME); and type 2 when the SI levelled following maximum enhancement (threshold as late enhancement within ±10 % of ME).

DCE analysis based on the extended Toft pharmacokinetic [22] model was also performed, deriving blood plasma volume v_p; transfer rates K_{trans}, k_{ep} between plasma and extracellular extravascular space.

Development of LR models

Individual logistic regression models were derived for each of the three cancer definitions from the training cohort. The score test was used to select the mp-MRI parameters most likely to contribute significantly ($p<0.05$) for inclusion in each model [23].

Logistic regression was preferred over linear discriminant analysis (LDA), because it does not require the independent variables to be normally distributed, linearly related, nor of equal variance within each group [24]. Moreover, similar to LDA, logistic regression provides a deterministic model yielding weighting factors for each contributing variable, easing

interpretation and clinical application of results compared with neural network models (which can also more easily over-fit the data).

Validation of LR models

Data were analysed with the Statistical Package for Social Sciences, version 19.0 (SPPS Inc., Chicago, IL, USA).

Internal validation of each of the three LR models was performed separately using the training cohort ($n=70$). The receiver operator characteristic area under the curve (ROC AUC) was calculated for each model prior to and following leave-one-out analysis [25]. For leave-one-out analysis, one patient's data was excluded, and a model was generated from the remaining patients. The model was then tested on the excluded data to calculate a predictive probability. The process was repeated for all patients to calculate 70 predictive probabilities. An ROC curve was then constructed from the predictive probabilities. Temporal validation of the LR model was then performed by its application to the temporal validation cohort ($n=85$).

Performance of a single parameter T2 alone LR model was assessed against the multi-parametric LR model for prediction of definition-2 cancer using ROC analysis.

Comparison of LR models against radiologists' performance in the temporal validation cohort

Two radiologists (A and B) unaware of the histopathology independently reviewed and contoured suspected transition zone lesions on temporal cohort mp-MRI examinations. Where no suspect lesion was identified, the radiologists placed a 1-cm ROI in any of the four transition zone sectors. The sector containing each ROI was recorded.

The radiologists scored all ROIs for the presence/absence of significant (a minimum of definition-2 [17]) cancer using a 1 to 5 scale (1=highly likely benign, 2=likely benign, 3= equivocal, 4=likely significant cancer, 5=highly likely significant cancer) as per Pi-RADS mp-MRI scoring criteria [20]. Figures 2, 3 and 4 demonstrate radiologists' Pi-RADS scores of 2, 3 and 5 respectively. Whenever a score of 3 was ascribed, the radiologists further dichotomized the score to a 3+ or a 3- based on the 'best guess' of whether significant cancer was more or less likely than 50 %.

The sector containing the radiologists' scored ROI was then matched against histopathology and classified as positive/negative for each of the three cancer definitions.

Separate ROC curves were derived and sensitivity/ specificity analysis was performed using Pi-RADS scores of≥3 and a≥4 threshold score as positive for significant cancer on mp-MRI.

Quantitative parameters derived from ROIs were (as described for the training cohort) used to test the LR models.

Fig. 2 Axial multi-parametric MR images [(*a*) T2 weighted, (*b*) Apparent diffusion coefficient map, (*c*) pre-contrast T1, and (*d*) early post contrast T1] of a right anterior basal region (*yellow arrows*) Pi-RADS score 2/5. Template mapping biopsy revealed benign change only within the corresponding Barzell zones (10 and 4)
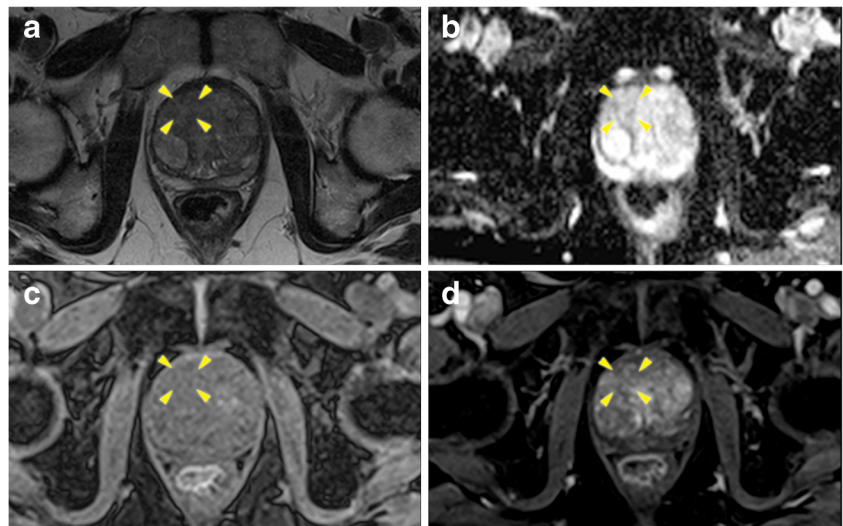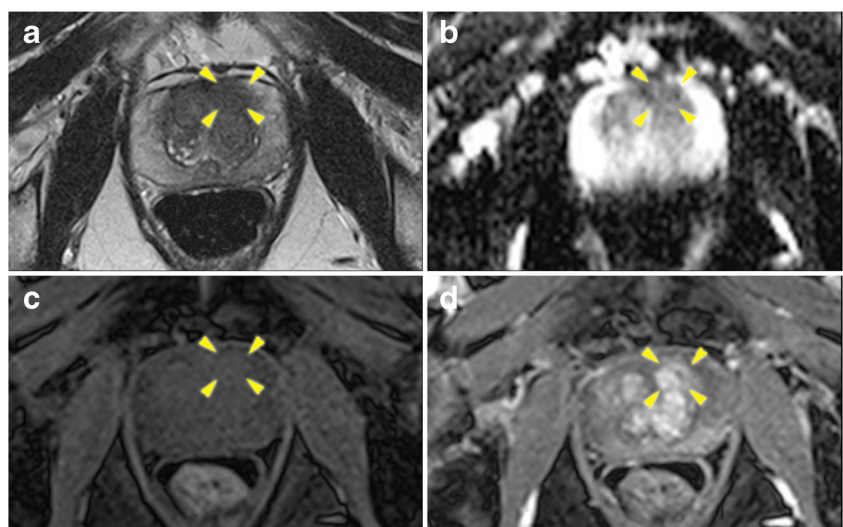

ROC and sensitivity/specificity analysis at a probability threshold of >0.5 was performed.

Results

For the training cohort, 37 of the 70 (53 %) localized ROIs were classified by histopathology as positive for any-cancer; 28 (40 %) of these reached the threshold for definition-2, and 25 (36 %) reached definition-1 cancer. For the temporal validation patient cohort, 46 of the 85 (54 %) localized ROIs were classified by histopathology as positive for any-cancer; 25 (29 %) of these reached the threshold for definition-2, and 14 (16 %) reached definition-1 cancer. The median tumour volume of histologically confirmed definition-2 and definition 1 lesions scored as Pi-RADS scores 3 to 5, and was localizable on mp-MRI at 0.8 ml (range, 0.2–1.2 ml).

Derivation of LR models

The score test confirmed that T2-nSI, ADC, and ME were most contributory to ROI classification for all cancer definitions (Table 2). The logistic regression equations for the diagnostic multi-parametric model for any-cancer, definition-2 and definition-1 cancer (Eqs. 1–3) are given below:

Any-cancer:

$$\ln(Odds) = 5.952 - 1.343 \cdot ADC - 0.227 \cdot T2nSI - 2.379 \cdot ME \quad (1)$$

Definition-2 cancer:

$$\ln(Odds) = 5.347 + 0.332 \cdot ADC - 0.974 \cdot T2nSI - 1.730 \cdot ME \quad (2)$$

Definition-1 cancer:

$$\ln(Odds) = 4.844 + 0.194 \cdot ADC - 0.895 \cdot T2nSI - 1.659 \cdot ME \quad (3)$$

Fig. 3 Axial multi-parametric MR images [(*a*) T2 weighted, (*b*) Apparent diffusion coefficient map, (*c*) pre-contrast T1, and (*d*) early post contrast T1] of a left anterior apical region (*yellow arrows*) Pi-RADS score 3/5. Template mapping biopsy revealed definition one tumour within the corresponding Barzell zones (1 and 7)

Fig. 4 Axial multi-parametric MR images [(*a*) T2 weighted, (*b*) Apparent diffusion coefficient map, (*c*) pre-contrast T1, and (*d*) early post contrast T1] of a right anterior basal region (*yellow arrows*) Pi-RADS score 5/5. Template biopsies revealed definition one tumour within the corresponding Barzell zones (10 and 4)
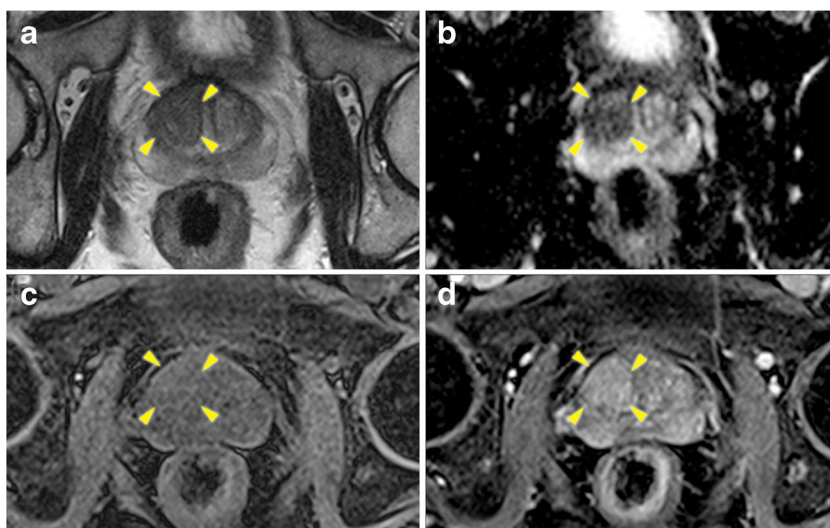
Validation of LR model performance

The ROC-AUC of the diagnostic model for a positive diagnosis for any-cancer, definition-2 cancer and definition-1 cancer was 0.78 (95 % CI 0.67–0.89), 0.8 (95 % CI 0.70–0.90), and 0.79 (95 % CI 0.69–0.89), respectively, prior to leave-one-out analysis; and 0.73 (95 % CI 0.61–0.85), 0.73 (95 % CI 0.61–0.85), and 0.71 (95 % CI 0.58–0.84), respectively, following leave-one-out analysis.

The ROC curve for each diagnostic model applied to the temporal validation patient group is given in Fig. 5a. The ROC-AUC for the any-cancer, definition-2 and definition-1 cancer model was 0.76 (95 % CI 0.66–0.87), 0.67 (95 % CI

Table 2 Score test results for quantitative mp-MRI parameters

MR Parameter	Score test (p value) for any-cancer model	Score test (p value) for definition-2 cancer model	Score test (p value) for definition-1 cancer model
[†]ADC (x10^{-3} mm^{2}s^{-1})	7.95 (0.005*)	7.08 (0.008*)	6.74 (0.009*)
[†]T2-nSI	7.34 (0.007*)	12.3 (<0.001*)	11.0 (0.001*)
[†]DCE-nSI	1.32 (0.250)	2.47 (0.116)	1.65 (0.199)
SoE	0.66 (0.417)	0.17 (0.678)	0.55 (0.459)
ME	12.9 (<0.001*)	9.46 (0.002*)	8.79 (0.003)
*E$_{type}$	0.38 (0.144)	0.39 (0.532)	0.14 (0.711)
AUC$_{tot}$	4.77 (0.029*)	2.51 (0.113)	3.23 (0.720)

Definition-2 cancer defined as ≥Gleason 3+4 or a template biopsy cancer core length of ≥4 mm. Definition-1 cancer defined as ≥Gleason 4+3 or a template biopsy cancer core length of ≥6 mm

ADC – apparent diffusion coefficient; T2-nSI – normalized T2 signal intensity, DCE-nSI – normalized early arterial enhanced T1 signal intensity, SoE – slope of enhancement, ME – maximum enhancement, Etype – curve shape type, AUC$_{tot}$ – area under enhancement curve, ROC – receiver operating characteristic, AUC – area under curve

* Significant contribution to model

0.55–0.79) and 0.70 (95 % CI 0.55–0.85). For the T2 alone LR model, the ROC-AUC was 0.62 (95 % CI 0.50–0.76), suggesting little additional benefit of mp-MRI.

Applying a probability threshold of >0.5 as positive for cancer, the sensitivity/specificity of LR for any-cancer, definition-2 and definition-1 was 0.91/0.26; 0.60/0.73; 0.71/0.70; respectively.

Specifically for equivocally (Pi-RADS 3; 20/85 for radiologist A, 39/85 for radiologist B) scored MRI lesions within the temporal validation group: for radiologist A, the mp-MRI LR model had an ROC-AUC of 0.65 (95 % CI 0.40–0.89) and the T2 alone LR model had an ROC-AUC of 0.49 (95 % CI 0.24-0.75), suggesting additional benefit of mp-MRI in this patient subset. For radiologist B, the mp-MRI LR model had an ROC-AUC of 0.59 (95 % CI 0.39–0.79) and the T2 LR model had an ROC-AUC of 0.50 (95 % CI 0.28–0.71), again suggesting additional benefits of an mp-MRI model. The sensitivity/ specificity of the definition-2 mp-MRI LR model was 0.71/0.61 for radiologist A and 0.50/076 for radiologist B for the Pi-RADS 3 group.

Comparison with radiologist performance

The radiologists' Pi-RADS score for the temporal validation patient cohort are given in Table 3.

Figure 5b illustrates the ROC curve of overall performance of radiologists A and B (and comparative LR model) for prediction of clinically significant (definition-2) cancer for the temporal validation group.

The ROC-AUC for radiologist A was 0.65 (95 % CI 0.51–0.79). The sensitivity/specificity of radiologist A for cancer was 0.45/1.0, 0.52/0.94 and 0.64/0.93 using a score ≥3 as positive; and 0.20/1.0, 0.24/0.95 and 0.35/0.94 using ≥4 as positive for any-cancer, definition-2 and definition-1 cancer, respectively.
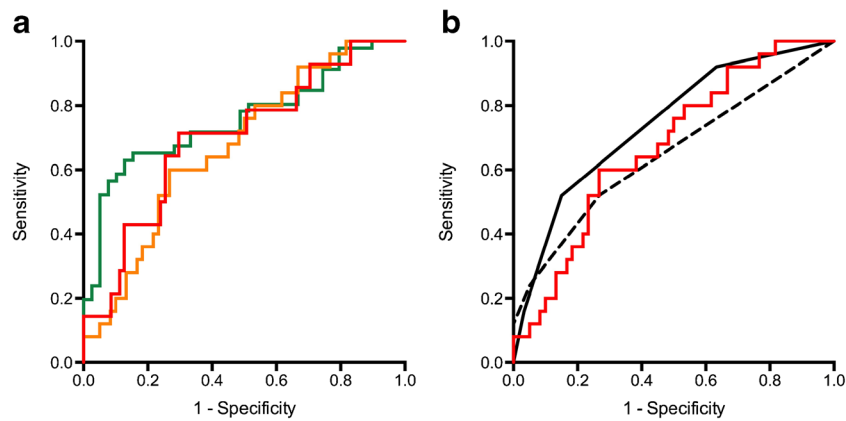
Fig. 5 (*a*) Receiver operating characteristic curves for temporal validation of CAD models to classify the presence of any-cancer [*green*], definition-2 cancer (≥4 mm with ≥ Gleason 3+4) [*yellow*], and definition-1 cancer (≥6 mm with ≥ Gleason 4+3) [*red*]. Corresponding area-under-curve of 0.76 (95 % CI 0.66–0.87), 0.67 (95 % CI 0.55–0.79) and 0.70 (95 % CI 0.55–0.85) (*b*) Receiver operating characteristic curve of radiologist A (*black line*) and radiologist B (*dashed black line*) Pi-RADS score for classification of clinically significant transition zone cancer; area under curve of 0.65 (95 % CI 0.51–0.79) and 0.74 (95 % CI 0.63–0.86), respectively. The *red line* demonstrates the ROC curve for the logistic regression model

Twenty of 85 patients within the temporal validation group were classified as equivocal for significant tumours by radiologist A (Pi-RADS score 3); 8/20 were subclassified as 3+ and 12/20 as 3-. Within this group, there were seven (35 %) definition-2 cancers and 13 (65 %) benign/insignificant cancers. The sensitivity and specificity of the radiologist A subclassification for significant cancer were 0.14 and 0.54, respectively (c.f. LR model sensitivity of 0.71 and specificity of 0.61).

The ROC-AUC for radiologist B was 0.74 (95 % CI 0.63–0.86). The sensitivity/specificity of radiologist B for cancer was 0.88/0.47, 0.95/0.37 and 0.92/0.32 using a score≥3 as positive; and 0.42/0.92, 0.54/0.85 and 0.64/0.81 using≥4 as positive for any-cancer, definition-2 and definition-1 cancer, respectively.

Thirty-nine of 85 patients within the temporal validation group were classified as equivocal for significant tumour by radiologist B (Pi-RADS score 3); 14/39 were sub-classified as 3+ and 25/39 as 3-. Within this group, there were ten (26 %) definition-2 cancers and 29 (74 %) benign/insignificant cancers. The sensitivity and specificity of radiologist B sub-classification for significant cancer were 0.40 and 0.34, respectively (c.f. LR model sensitivity of 0.50 and specificity of 0.76).

Discussion

Summary of results

There are two reasons why a radiologist may miss a cancer: (1) the radiologist saw the cancer, but dismissed it as benign change; or (2) the radiologist did not see the cancer. In this

Table 3 Radiologists Pi-RADS score for the temporal validation cohort (*n*=85)

Pi-RADS score for the presence of significant disease within the selected ROI (1=significant tumour highly unlikely, 2=significant tumour unlikely, 3=equivocal, 4=significant tumour likely, 5=significant tumour highly likely).

ROI – region of interest

*≥Gleason 3+4 or a template biopsy CCL of≥4 mm

† ≥Gleason 4+3 or a template biopsy CCL of≥6 mm

Radiologist Pi-RADS Score	Total number of ROIs	ROIs positive for *definition-2 cancer	ROIs negative for *definition-2 cancer	ROIs positive for any-cancer	ROIs positive for †definition-1 cancer
Radiologist A					
1	0	0	0	0	0
2	56	12	44	25	5
3	20	7	13	12	4
4	6	3	3	6	3
5	3	3	0	3	2
Radiologist B					
1	0	0	0	0	0
2	24	2	22	6	1
3	39	10	29	21	4
4	16	9	7	13	5
5	6	4	2	6	4

study, we assessed whether a mp-MRI derived LR model could address the former problem; i.e., whether it has the potential to aid the radiologist in making classification decisions. Whilst overall we found that our LR model was as accurate as an experienced radiologist at determining whether an area was clinically significant prostate cancer or a benign change, ROC-AUC 0.65 (radiologist A), 0.74 (radiologist B) versus 0.67 (LR model); for Pi-RADS score=3 (equivocal) areas, our LR model outperformed radiologists' "best guess".

Comparison with existing work

The three variables that proved to be the best classifiers in our LR models were: ADC, T2nSI and ME. In turn, these relate to: tissue cellularity; water content and tissue vascularity [26]. Others have used similar inputs into their models, but mainly where applied to peripheral zone tumours. For example, Langer et al. [3] found ADC, T2 and K_{trans} were the best classifiers within the peripheral zone, whilst Shah et al. [27] report performance of ADC>DCE> normalized T2 weighted imaging for classification of peripheral zone tumour. In line with peripheral zone classification, we confirm that ADC is also of primary importance for classifying transition zone tumours.

Our results show that the principal problem faced by the radiologist when assessing the transition zone is a reduction in sensitivity for diagnosis of clinically significant disease. This concurs with the findings of others [28, 29].

In agreement with other findings [29, 15], we found no overall benefit of using mp-MRI LR versus T2-only LR models when applied to all patients. However, when specifically applied to lesions classified by radiologists as Pi-RADS 3 (equivocal), the mp-MRI LR model was a better classifier than the T2-only LR model.

Similar to our work, a recent publication also tested a mp-MRI TZ-specific diagnostic model [15]. Their results show a higher ROC-AUC of 0.87 (95 % CI 0.78, 0.96) for the depiction of TZ prostate cancer when compared with our work. This, however, is likely to be explained by differences in study population and histological reference standard, where we use a lower threshold for significant tumour and smaller tumours within our patient cohort. By way of explanation, the median volume of TZ prostate cancer in the study by Hambrock et al. [15] was 2.5 ml (range, 0.5–12.48 ml). In addition, Hambrock et al. defined significant tumour based on a threshold volume of 0.5 ml (approximately equivalent to a 1-cm diameter lesion), irrespective of Gleason grade. In comparison, our median volume of TZ prostate cancer was significantly lower at 0.8 ml (range, 0.2–1.2 ml) and significant tumour was defined either by volume (>0.2 ml; equivalent to a CCL of 4 mm [19]), or *any* size tumour that includes Gleason primary or secondary pattern 4.

Methodological limitations

Our study has some potential limitations. Firstly, we did not have access to radical prostatectomy specimens as ground-truth. However, we believe that our employment of a TPM-based reference standard [30] helps to avoid spectrum bias that can occur with studies based on prostatectomy specimens. Secondly, TPM is a significant improvement over a TRUS biopsy reference standard [19, 31] used by others. We acknowledge that there may still be some error in matching TPM histopathology with mp-MRI. However, it should be noted that prostatectomy specimens themselves are not free from artefacts due to processing (shrinkage, distortion, tissue-loss), which also impair accurate co-registration with imaging [32]. Furthermore, we aimed to reduce misregistration error by limiting the number of divisions of the transition zone (four sectors), basing divisions on easily identifiable landmarks and employing consensus for experienced radiologists' opinions on ascribing the location of ROIs. Indeed, we believe that misregistration error has not impeded our LR model performance.

For our MRI protocol, we followed the recommendations from the European Consensus meeting [1], which suggests that imaging could be adequately performed at a 1.5 T (such as the one used in this work), but we acknowledge that a 3 T MRI protocol may have improved performance of both our radiologists and the derived LR models.

Whilst recommended by other studies [33], we did not use endorectal coils, and we note that the European Consensus Meeting failed to reach agreement on their necessity even at 1.5 T [1]. In our experience, the use of endorectal coils has not proven to be a significant detriment to overall performance of mp-MRI at 1.5 T [34]. Our previous work and that of others who have not employed endorectal coils and imaged at 1.5 T has shown comparable performance of mp-MRI with those studies that image with endorectal coils [17, 35–37]. However, the performance of our LR model in data sets generated at 3 T and with use of endorectal coils remains to be established.

Finally, whilst we have derived and tested LR models, the next step remains to integrate these models into a CAD package and test radiologist performance with and without CAD.

Clinical implications

Radiologists tend to use an ordinary scale of 1 to 5 when attributing the likelihood of clinically significant cancer to the whole prostate or to a given sector within [1]. These scores only have clinical utility if they assist in a management decision. When a radiologist scores a 4 or 5, he/she expects that a biopsy will be triggered for that sector. In contrast, a patient who is attributed a score of 1 to 2 may indeed avoid a biopsy [10, 38–40]. However, patients scored as 3 remain a

heterogeneous group, with some undergoing biopsy whilst others often opt for active surveillance. We therefore assessed if the LR model could help specifically classify patients classified as Pi-RADS 3.

Indeed, the sensitivity/specificity of the LR model within this group was higher than both of our radiologists' "best-guesses"; suggesting that the LR model could have utility as a 'second opinion' when radiologists are unsure about a transition zone region.

The next step is the integration of the LR model into a CAD package that generates probability maps. Future work will assess whether the CAD package indeed improves a radiologist's performance, as suggested by our results.

Conclusions

We demonstrate that overall performance of transition zone specific mp-MRI LR model is similar to an experienced radiologist, and moreover that the LR model improves classification of Pi-RADS score 3 lesions located within the transition zone.

References

1. Dickinson L, Ahmed HU, Allen C 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
2. Emberton M (2013) Has magnetic resonance-guided biopsy of the prostate become the standard of care? Eur Urol 64:720–721
3. Langer DL, van der Kwast TH, Evans AJ, Trachtenberg J, Wilson BC, Haider MA (2009) Prostate cancer detection with multi-parametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. J Magn Reson Imaging 30:327–334
4. Zakian KL, Eberhardt S, Hricak H et al (2003) Transition zone prostate cancer: metabolic characteristics at 1H MR spectroscopic imaging–initial results. Radiology 229:241–247
5. Chesnais AL, Niaf E, Bratan F et al (2013) Differentiation of transitional zone prostate cancer from benign hyperplasia nodules: evaluation of discriminant criteria at multiparametric MRI. Clin Radiol 68:e323–e330
6. Robinson C, Halligan S, Iinuma G et al (2011) CT colonography: computer-assisted detection of colorectal cancer. Br J Radiol 84:435–440
7. Sprindzuk MV, Kovalev VA, Snezhko EV, Kharuzhyk SA (2010) Lung cancer differential diagnosis based on the computer assisted radiology: The state of the art. Pol J Radiol 75:67–80
8. Yuan Y, Giger ML, Li H, Bhooshan N, Sennett CA (2010) Multimodality computer-aided breast cancer diagnosis with FFDM and DCE-MRI. Acad Radiol 17:1158–1167
9. Engelbrecht MR, Huisman HJ, Laheij RJ et al (2003) Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using dynamic contrast-enhanced MR imaging. Radiology 229:248–254
10. Nogueira L, Wang L, Fine SW et al (2010) Focal treatment or observation of prostate cancer: pretreatment accuracy of transrectal ultrasound biopsy and T2-weighted MRI. Urology 75:472–477
11. Sung YS, Kwon HJ, Park BW, Cho G, Lee CK, Cho KS, Kim JK (2011) Prostate cancer detection on dynamic contrast-enhanced MRI: computer-aided diagnosis versus single perfusion parameter maps. Am J Roentgenol 197:1122–1129
12. Niaf E, Rouvière O, Mège-Lechevallier F, Bratan F, Lartizien C (2012) Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Phys Med Biol 57:3833–3851
13. Dikaios NAM, Emberton M, Fujiwara T, Atkinson D, Punwani S (2013) Derivation and comparison of site specific peripheral and transition zone quantitative DCE MRI logistic regression models for prostate cancer detection: does cancer location matter? International Society of Magnetic Resonance in Medicine, Salt Lake City
14. Puech P, Betrouni N, Makni N, Dewalle AS, Villers A, Lemaitre L (2009) Computer-assisted diagnosis of prostate cancer using DCE-MRI data: design, implementation and preliminary results. Int J Comput Assist Radiol Surg 4:1–10
15. Hambrock T, Vos PC, Hulsbergen-van de Kaa CA, Barentsz JO, Huisman HJ (2013) Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging–effect on observer performance. Radiology 266:521–530
16. Onik G, Barzell W (2008) Transperineal 3D mapping biopsy of the prostate: an essential tool in selecting patients for focal prostate cancer therapy. Urol Oncol 26:506–510
17. Taira AV, Merrick GS, Galbreath RW et al (2010) Performance of transperineal template-guided mapping biopsy in detecting prostate cancer in the initial and repeat biopsy setting. Prostate Cancer Prostatic Dis 13:71–77
18. Barzell WE, Melamed MR (2007) Appropriate patient selection in the focal treatment of prostate cancer: the role of transperineal 3-dimensional pathologic mapping of the prostate—a 4-year experience. Urology 70:27–35
19. Ahmed HU, Hu Y, Carter T et al (2011) Characterizing clinically significant prostate cancer using template prostate mapping biopsy. J Urol 186:458–464
20. Barentsz JO, Richenberg J, Clements R et al (2012) ESUR prostate MR guidelines 2012. Eur Radiol 22:746–757
21. Zelhof B, Lowry M, Rodrigues G, Kraus S, Turnbull L (2009) Description of magnetic resonance imaging-derived enhancement variables in pathologically confirmed prostate cancer and normal peripheral zone regions. BJU Int 104:621–627

22. Tofts PS (1197) Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. J Magn Reson Imaging 7:91–101

23. SM (1995) Applied logistic regression analysisSage University Paper Series on Quantitative Applications in the Social Sciences. SAGE pp 7-106

24. Tabachnick BG, Fidell LS (1996) Using Multivariate Statistics. HarperCollins, NY

25. Fukunaga KHD (1989) Leave-one-out procedures for nonparametric error estimates. IEEE Trans Pattern Anal Mach Intell 11:421–423

26. Choi YJ, Kim JK, Kim N, Kim KW, Choi EK, Cho KS (2007) Functional MR imaging of prostate cancer. Radiographics 27:63–75, discussion 75-67

27. Shah V, Turkbey B, Mani H et al (2012) Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. Med Phys 39:4093–4103

28. Akin O, Sala E, Moskowitz CS et al (2006) Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging. Radiology 239:784–792

29. Hoeks CM, Hambrock T, Yakar D et al (2013) Transition zone prostate cancer: detection and localization with 3-T multiparametric MR imaging. Radiology 266:207–217

30. Ahmed HU, Emberton M, Kepner G, Kepner J (2012) A biomedical engineering approach to mitigate the errors of prostate biopsy. Nature Reviews. Urology 9:227–23131

31. Hoeks CM, Barentsz JO, Hambrock T et al (2011) Prostate cancer: multiparametric MR imaging for detection, localization, and staging. Radiology 261:46–66

32. Xu S, Kruecker J, Turkbey B et al (2008) Real-time MRI-TRUS fusion for guidance of targeted prostate biopsies. Comput Aided Surg 13:255–264

33. Bonekamp D, Jacobs MA, El-Khouli R, Stoianovici D, Macura KJ (2011) Advancements in MR imaging of the prostate: from diagnosis to interventions. Radiographics 31:677–703

34. Kirkham AP, Haslam P, Keanie JY, McCafferty I, Padhani AR, Punwani S, Richenberg J, Rottenberg G, Sohaib A, Thompson P, Turnbull LW, Kurban L, Sahdev A, Clements R, Carey BM, Allen C (2013) Prostate MRI: who, when, and how? Report from a UK consensus meeting. Clin Radiol 68:1016–1023

35. Kirkham AP, Emberton M, Allen C (2006) How good is MRI at detecting and characterising cancer within the prostate? Eur Urol 50:1163–1174

36. Bratan F, Niaf E, Melodelima C, Chesnais AL, Souchon R, Mège-Lechevallier F, Colombel M, Rouvière O (2013) Influence of imaging and histological factors on prostate cancer detection and localisation on multiparametric MRI: a prospective study. Eur Radiol 23:2019–2029

37. Lee SH, Park KK, Choi KH, Lim BJ, Kim JH, Lee SW, Chung BH (2010) Is endorectal coil necessary for the staging of clinically localized prostate cancer? Comparison of non-endorectal versus endorectal MR imaging. World J Urol 28:667–672

38. Arumainayagam N, Kumaar S, Ahmed HU et al (2010) Accuracy of multiparametric magnetic resonance imaging in detecting recurrent prostate cancer after radiotherapy. BJU Int 106:991–997

39. Mueller-Lisse U, Scheidler J, Klein G, Reiser M (2005) Reproducibility of image interpretation in MRI of the prostate: application of the sextant framework by two different radiologists. Eur Radiol 15:1826–1833

40. Villers A, Puech P, Mouton D, Leroy X, Ballereau C, Lemaitre L (2006) Dynamic contrast enhanced, pelvic phased array magnetic resonance imaging of localized prostate cancer for predicting tumor volume: correlation with radical prostatectomy findings. J Urol 176:2432–2437