Prostate cancer heterogeneity: texture analysis score based on multiple magnetic resonance imaging sequences for detection, stratification and selection of lesions at time of biopsy

Clement Orczyk*[†‡§] [iD], Arnauld Villers[¶], Henry Rusinek**, Vincent Lepennec[††],
Céline Bazille[‡‡], Francesco Giganti*[§§], Artem Mikheev**, Myriam Bernaudin[‡],
Mark Emberton*[†] [iD], Audrey Fohlen[‡††] and Samuel Valable[‡]

*Division of Surgery and Interventional Sciences, University College London, [†]Department of Urology, University College London Hospitals, London, UK, [‡]ISTCT/CERVOxy Group, GIP CYCERON, Normandie Université, UNICAEN, CEA, CNRS, [§]Department of Urology, University Hospital of Caen, Caen, [¶]Department of Urology, University Hospital of Lille, Nord de France, Lille, France, **Department of Radiology, New York University Medical Center, New York, NY, USA, [††]Department of Radiology, [‡‡]Department of Pathology, University Hospital of Caen, Caen, France, and [§§]Department of Radiology, University College London Hospitals, London, UK

Objective

To undertake an early proof-of-concept study on a novel, semi-automated texture-based scoring system in order to enhance the association between magnetic resonance imaging (MRI) lesions and clinically significant prostate cancer (SPCa).

Patients and Methods

With ethics approval, 536 imaging volumes were generated from 20 consecutive patients who underwent multiparametric MRI (mpMRI) at time of biopsy. Volumes of interest (VOIs) included zonal anatomy segmentation and suspicious MRI lesions for cancer (Likert Scale score >2). Entropy (E), measuring heterogeneity, was computed from VOIs and plotted as a multiparametric score defined as the entropy score (ES) = E ADC + E K^{trans} + E Ve + E T2WI. The reference test that was used to define the ground truth comprised systematic saturation biopsies coupled with MRI-targeted sampling. This generated 422 cores in all that were individually labelled and oriented in three-dimensions. Diagnostic accuracy for detection of SPCa, defined as Gleason score ≥3 + 4 or >3 mm of any grade of cancer on a single core, was assessed using receiver operating characteristics, correlation, and descriptive statistics. The proportion of cancerous lesions detected by ES and visual scoring (VS) were statistically compared using the paired McNemar test.

Results

Any cancer (Gleason score 6–8) was found in 12 of the 20 (60%) patients, with a median PSA level of 8.22 ng/mL.

SPCa (mean [95% confidence interval, CI] ES = 17.96 [0.72] NATural information unit [NAT]) had a significantly higher ES than non-SPCa (mean [95% CI] ES = 15.33 [0.76] NAT). The ES correlated with Gleason score (r_s = 0.568, P = 0.033) and maximum cancer core length (ρ = 0.781; P < 0.001). The area under the curve for the ES (0.89) and VS (0.91) were not significantly different (P = 0.75) for the detection of SPCa amongst MRI lesions. Best ES estimated numerical threshold of 16.61 NAT led to a sensitivity of 100% and negative predictive value of 100%. The proportion of MRI lesions that were found to be positive for SPCa using this ES threshold (54%) was significantly higher (P < 0.001) than using the VS (24% of score 3, 4, 5) in a paired analysis using the McNemar test. In all, 53% of MRI lesions would have avoided biopsy sampling without missing significant disease.

Conclusion

Capturing heterogeneity of prostate cancer across multiple MRI sequences with the ES yielded high performances for the detection and stratification of SPCa. The ES outperformed the VS in predicting positivity of lesions, holding promise in the selection of targets for biopsy and calling for further understanding of this association.

Keywords

detection, stratification, MRI, image processing, biopsy, radiomics, #ProstateCancer, #PCSM

Introduction

Current diagnostic strategies for prostate cancer detection have relied upon PSA and random sampling of the gland using TRUS-guided biopsy. We now know that this approach is associated with an over-representation of clinically insignificant prostate cancer, as well as an under-representation of clinically significant prostate cancer (SPCa) [1].

Multiparametric (mp)MRI has been introduced into the pathway in order to mitigate some of the deficiencies associated with the standard of care. Its introduction has been associated with an increased detection of SPCa and, in certain circumstances, it has provided a strategy to address the problem of over-diagnosis [1–3] raised in major screening trials [4–6].

In current practice, prostate mpMRI visual reporting requires considerable expertise and is prone to inter-observer variability, even when applying standardised guidelines such as the Prostate Imaging-Reporting And Data System (PI-RADS) version 2.0 [7]. It also takes a considerable amount of time to do given the number of sequences that need to be looked at and compared with each other, and possibly with previous imaging. Quantitative parameters derived from functional sequences do not offer a reliable alternative as they are prone to substantial variabilities [8].

Some recent developments in image analysis (by means of neural networks) are now making the possibility of automatic MRI-segmentation of both the prostate itself and of the MRI-lesions within it [9,10]. Automated assessment of the lesions, if it were indeed possible, opens up the possibility of fully automated MRI reporting in the future, and access to and computation of information not perceived by visual assessment in the context of radiomic analysis of standard-of-care images [11,12].

Tumour heterogeneity is considered to be one of the main drivers of progression and resistance to treatment [13]. The degree of histological prostate cancer heterogeneity (i.e., the gradual loss of a structured cellular architecture) is reflected at the tissue level by the Gleason grade [14,15] and was found at an intratumoral level as opposed to benign tissue [15–17] by whole genomic sequencing.

Heterogeneity of prostate cancer may also be non-invasively assessed using parameters derived from imaging. Entropy computes heterogeneity of a given volume [12,18,19], measuring spatial randomness of image intensities. Higher entropy represents a more disordered distribution of values within the tissue [20], potentially reflecting the biological feature of heterogeneity of prostate cancer, as in other carcinomas using MRI [20].

There are evidences suggesting that lesion entropy could be derived from specific mpMRI sequences for the detection,

stratification, and prognosis of prostate cancer [21–24]. Combining entropy from different MRI sequences might improve its ability for cancer detection and stratification, as the image generated by each sequence reflects different aspects of the underlying lesion's histopathology and does not necessary detect a cancer focus due to histology variation in tissue composition [25].

Our aim in the present study was to undertake an early proof-of-concept study on a novel texture-based scoring system, relying on entropy, in order to enhance the association of MRI-derived target generation with SPCa.

Patients and Methods

Study Population and Image Acquisition (Standard Test)

Under Local Ethics Committee approval, which waives written informed consent, we analysed the data of 21 consecutive patients referred for suspicion of prostate cancer without contraindications to MRI. Patients underwent mpMRI before biopsy, except for those with suspicion of high-risk disease in whom biopsies were carried out first. All patients were scanned on a 1.5 T MRI system with a pelvic phased-array (Siemens Magnetom Avanto®; Siemens Healthcare Limited, Camberley, Surrey, UK), using the same mpMRI protocol (T2-weighted imaging [T2WI], diffusion-weighted imaging [DWI] with generation of apparent diffusion coefficient [ADC] maps and dynamic contrast-enhanced [DCE] imaging) in 2013 and 2014. The details of the protocol are listed in Table S1 and are consistent with the European Society of Urogenital Radiology (ESUR) 2012 guidelines that were available at the start of the study [26].

Two radiologists (with 6 and 3 years experiences in prostate mpMRI interpretation, respectively) reported the scans according to a Likert Scale ranging from 1 to 5 [27]. All sequences were used to allocate the score. Lesions with scores of 3–5 were considered to be suspicious for cancer. After consensus review of the images with the urologist performing the biopsy, a map of 36 regions of interest was generated before the procedure [28]. Visual scoring (VS) of MRI was the standard test.

Reference Standard: Biopsy Procedure and Correlation

All patients underwent subsequent TRUS-guided biopsy as per local protocol, which implied a saturation TRUS-guided biopsy for PSA levels <15 ng/mL with a 22-core template. Each core, either systematic or targeted, was labelled according to its location and sent in a separate jar for analysis. In cases of suspicious lesion at mpMRI before biopsy, targeted biopsies were performed using cognitive

registration by a single urologist trained in this technique, before systematic sampling [29]. A dedicated uro-pathologist reported the biopsy. A drawing reporting location and numbering of the cores was issued for each patient and sent to pathology. Systematic TRUS-guided biopsy combined with targeted biopsy corresponds to the reference test. Significant disease was defined as the presence of Gleason pattern 4 (primary or secondary) and/or >3 mm cancer in a single core [30].

Biopsy and MRI maps were compared in consensus to establish concordance for cancer leading to true-positive, false-positive, false-negative, and true-negative MRI regions.

Index Test: Image Processing

The image processing protocol is shown in Fig. S1.

Quantitative maps from DCE imaging were computed with correction of motion artefacts. The two parameters were the volume transfer constant (K^{trans}, in mL/min) and fraction of extracellular extravascular space (Ve, no unit) (see Appendix S1 for processing pharmacokinetic model).

All quantitative maps, ADC map and three-dimensional (3D) DCE-derived maps (K^{trans} and Ve) were aligned in the anatomical reference space (T2WI) using automatic rigid registration. Therefore, the same volume was analysed across the difference sequences.

Masks to generate volumes of interest (VOIs) were manually segmented in T2WI within a dedicated platform by an individual trained across datasets of prostate mpMRI-histology correlation and performing an image-guided procedure based on mpMRI. The mask segmentation was then propagated across sequences. The different VOIs according to zonal anatomy were: prostate (excluding seminal vesicles), peripheral zone (PZ), and transition zone (TZ). Patient-related VOIs were the MRI suspicious lesions at VS, normal peripheral zone (nPZ) and BPH nodules (BPHn). Those are referred to as discrete VOIs.

As results, a set of VOIs at MRI was generated for: prostate, PZ, TZ, nPZ, BPHn, non-cancerous and cancerous area, including lesion at VS for four imaging parameters.

Texture analysis was computed from the VOIs as a 3D assessment using a statistical analysis of the histogram, known as a first-order analysis [11,18,22]. Entropy was directly computed for the VOI, corresponding to the whole lesion in 3D [19] (Appendix S3).

The computation entropy was calibrated to sample the relevant information from a 0.2 mL lesion, as this is considered as the minimum volume of significant disease when Gleason pattern 4 is identified [11]. Methods of

calibration are detailed in Appendix S3. Consequently, all sequences were sampled with the same precision for the relevant information.

The entropy unit is the NATural unit of information (NAT) for each parameter.

To assess the entropy derived from different images, we introduced an entropy score (ES) representing the combined entropy within each VOI to characterise quantitatively a volume in the same manner mpMRI is visually reported with a multiparametric approach.

ES for a volume (*v*) is defined in the present work as

$$ES(v) = E^{ADC}(v) + E^{Ktrans}(v) + E^{Ve}(v) + E^{T2WI}(v)$$

Statistical Analysis

Power calculation was performed to demonstrate significant differences in the mean value of K^{trans} between cancer foci and benign tissue ($\alpha = 0.05$; $\beta = 1-0.9$, two-tailed hypothesis).

Statistical analysis was performed by using R software (The R Project for Statistical Computing, Vienna, Austria; http://www.R-project.org). Figure 1 describes the different level of analysis and study workflow.

The quantitative parameters and derivative metrics were tested as paired data to compare the values of cancer VOIs to the matching normal tissue (either nPZ or BPHn depending of the zone of origin of the tumour).
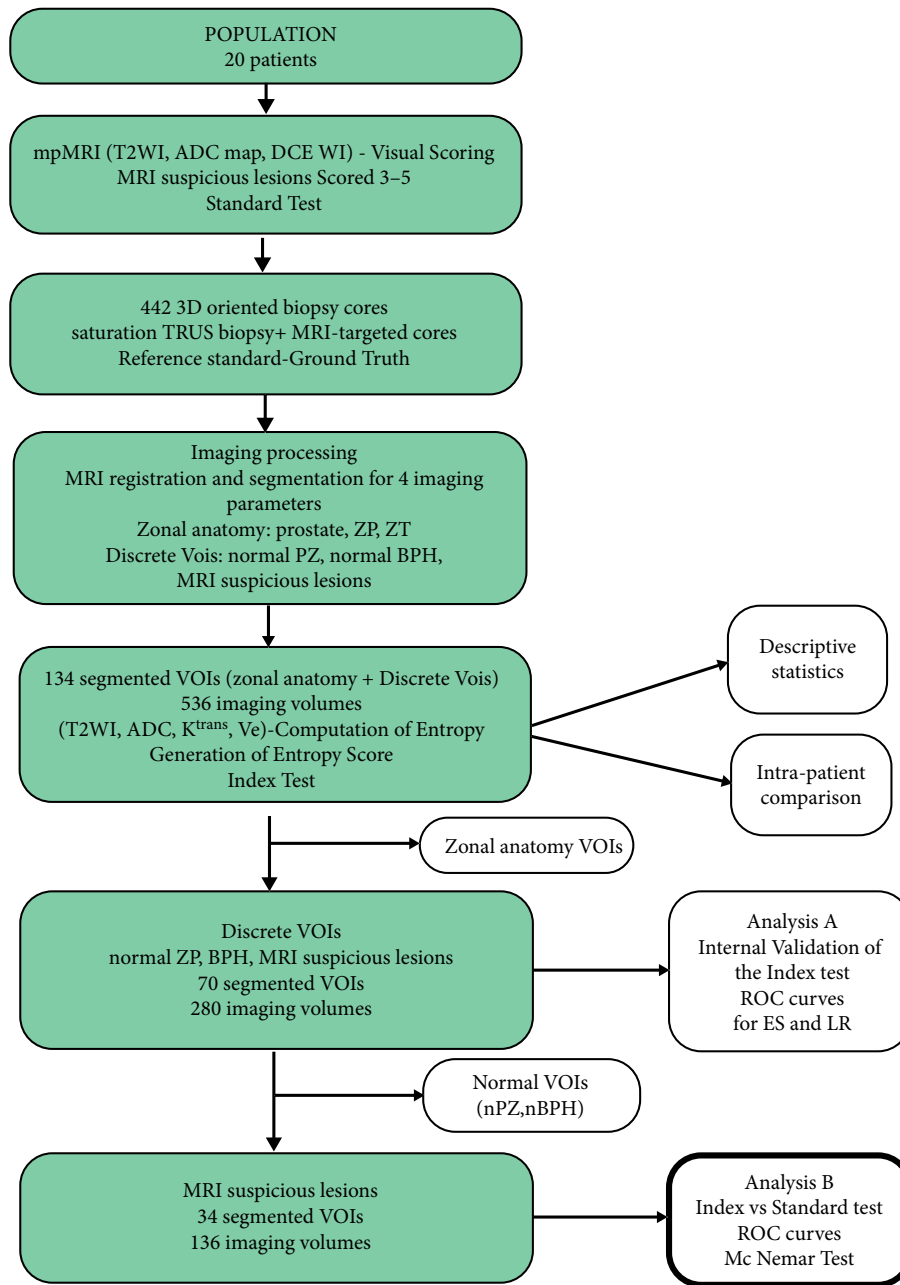
We used a variance analysis test of Friedman for paired data to assess if the tested parameters were independent for a given type of VOI.

Performances for detection were assessed at two levels of analysis (Fig. 1). Analysis 'A' considered all VOIs related to the patient. Analysis 'B' considered only VS lesions (either cancerous or not).

For continuous data and ES at the two aforementioned levels of analysis, receiver-operating characteristic (ROC) curves were generated and area under the curve (AUC) calculated (*pROC* package) to perform binary classification between SPCa and non-SPCa (either insignificant or non-cancerous). Comparison of the AUC (significant if $P < 0.05$) of the ES was carried out against each ES component taken individually. To fulfil the primary objective we compared the AUC of the VS to the ES.

Logistic regression (LR) modelling was performed for both the A and B analyses, including each single entropy parameter and promising raw quantitative value for K^{trans}, Ve

Fig. 1 Study profile and different levels of statistical analysis with matching investigated population.

```
┌─────────────────────────────────┐
│           POPULATION            │
│           20 patients           │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│ mpMRI (T2WI, ADC map, DCE WI) - Visual Scoring │
│  MRI suspicious lesions Scored 3–5  │
│           Standard Test          │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│     442 3D oriented biopsy cores │
│ saturation TRUS biopsy+ MRI-targeted cores │
│  Reference standard-Ground Truth │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│        Imaging processing        │
│ MRI registration and segmentation for 4 imaging │
│             parameters           │
│  Zonal anatomy: prostate, ZP, ZT │
│ Discrete Vois: normal PZ, normal BPH, │
│       MRI suspicious lesions     │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐        ┌──────────────────┐
│ 134 segmented VOIs (zonal anatomy + Discrete Vois) │──────▶ │    Descriptive   │
│         536 imaging volumes      │        │     statistics   │
│ (T2WI, ADC, Ktrans, Ve)-Computation of Entropy │  └──────────────────┘
│     Generation of Entropy Score  │        ┌──────────────────┐
│             Index Test           │──────▶ │   Intra-patient  │
└─────────────────────────────────┘        │    comparison    │
                │    ┌──────────────────┐   └──────────────────┘
                ├───▶│ Zonal anatomy VOIs │
                ▼    └──────────────────┘
┌─────────────────────────────────┐        ┌──────────────────┐
│           Discrete VOIs          │        │    Analysis A    │
│  normal ZP, BPH, MRI suspicious lesions │─▶│ Internal Validation of │
│        70 segmented VOIs         │        │   the Index test │
│         280 imaging volumes      │        │    ROC curves    │
└─────────────────────────────────┘        │    for ES and LR │
                │    ┌──────────────────┐   └──────────────────┘
                ├───▶│   Normal VOIs     │
                ▼    │   (nPZ,nBPH)      │
┌─────────────────────────────────┐  └───┘ ┌──────────────────┐
│      MRI suspicious lesions      │        │    Analysis B    │
│        34 segmented VOIs         │──────▶ │ Index vs Standard test │
│        136 imaging volumes       │        │    ROC curves    │
└─────────────────────────────────┘        │    Mc Nemar Test │
                                            └──────────────────┘
```

and ADC, and clinical data. Performances of the model were assessed using ROC curves for both the A and B analyses. The AUC of the LR models, the VS and the ES were compared.

Sensitivity, specificity, positive (PPVs) and negative predictive values (NPVs) were also calculated for the LR models and the ES for both levels of analysis using the best calculated threshold from the ROC. The best estimated threshold was computed as maximising the sum of test sensitivity (Se) and test specificity (Sp).

Applying this best estimated threshold, we compared proportions of positive lesions for SPCa estimated by the VS and ES in the B analysis population with a paired McNemar test.

The correlation of the ES with the Gleason score was performed using the Spearman coefficient. The correlation of the ES with the maximum core length was assessed using the Pearson coefficient.

All statistical tests were conducted at the two-sided 5% significance level.

Results

Population and Descriptive Results

From the 21 patients enrolled, one patient was found to be ineligible with inadequate DCE imaging.

Table 1 Included patients' characteristics.

Variable	Value
Number patients	20
Age, years, median (range)	65 (55–74)
PSA level, ng/mL, median (range)	8.22 (4.54–52.72)
N	
Patients with positive biopsy	12
Patients under active surveillance	2
Biopsy naïve patients	14
Patients with previous negative TRUS-guided biopsy	4
High-risk patients	2
mpMRI naïve patients	20
Maximum Gleason score per patient	
Gleason 6 (3 + 3)	8
Gleason 7 (3 + 4)	3
Gleason 8 (4 + 4)	1
Number of target per patient, median (range)	1.5 (0–3)
Maximum target score per patient, median (range)	3 (3–5)
n/N of positive targets for any cancer (SPCa, *n*)	
Score 3/5	2/19 (0)
Score 4/5	2/5 (2)
Score 5/5	3/3 (3)
Maximum core length, mm, mean (range)	5.5 (1–17)

Table 2 Values of ES for different VOIs.

| VOI (*n*) | ES, NAT | | | |
	Mean	Median	SD	95% CI
Prostate (20)	18.31	18.78	1.17	0.51
Whole PZ (20)	17.96	18.56	1.30	0.57
Whole TZ (20)	17.96	18.20	1.26	0.55
nPZ (19)***	**15.58**	15.47	3.84	1.72
BPHn (18)***	**16.26**	16.38	5.30	2.45
Target (28)	16.31	16.31	1.22	0.45
Positive target (5)*	**17.21**	16.87	1.00	0.42
Negative target (23)*	**16.11**	15.88	1.19	0.50
Prostate cancer (12)	16.93	16.81	1.59	0.90
SPCa (8)**	**17.73**	17.40	1.23	0.85
Non-SPCa (4)**	**15.33**	15.12	0.78	0.76

In bold values significantly different. *P = 0.035; **P = 0.003; ***P = 0.04.

From the 20 remaining patients, prostate cancer was detected in 12 (60%) patients following biopsies either from systematic or targeted cores. Patient characteristics are shown in Table 1. The reference test represents a data set of 442 biopsy cores independently labelled and oriented in space. The median time between mpMRI and biopsy was 35 days. The total number of targets was 28 of which seven were positive for cancer. Targets were scored 3, 4 and 5 in 20, five and three cases, respectively. The detection rate of any cancer was two of 20 (10%), two of five (40%), and three of three (100%), for the same range of score. All targets from MRI were segmented in the post-processing software.

Full imaging workflow was successfully carried out in all cases, leading to a data set of 536 independent volumes. After correlation with histology findings, 60 MRI sequences specific volumes were generated from 15 cancer foci after fusion of the sequences and maps in the T2WI space. Histogram generation and texture analysis were available for all the 536 VOIs.

The descriptive results of quantitative parameters, derived metrics and ES are listed in Table 2 and Table S2. Table S3 shows the mean values of quantitative parameters and entropy in paired match analysis between cancer VOIs and normal paired tissue.

There was a consistent trend for the entropy of each parameter for positive targets to be higher than negative, as shown in the box plot in Fig. S2.

The Friedman test showed independence ($P < 0.001$) of each individual component of the ES enabling a meaningful score.

For stratification of the disease, SPCa (mean [SD] ES = 17.73 [1.23] NAT) showed a significant higher ES than non-SPCa (mean [SD] ES = 15.33 [0.78] NAT; $P = 0.003$).

The Pearson's correlation coefficient of the ES of cancerous VOIs with the maximum cancer core length was found to be positive with $\rho = 0.781$ and $P < 0.001$.

The Spearman's correlation coefficient of the ES of cancerous VOIs with matching biopsy Gleason score was also positive with $r_s = 0.568$ and $P = 0.033$.

Table 3 Comparison of performances of the ES and multi LR models.

| | A analysis – All VOIs | | B analysis – MRI lesions | |
	ES	Regression model	ES	Regression model
Sensitivity, % (95% CI)	100 (63.0–100)	87.5 (47.3–99.6)	100 (63.0–100)	87.5 (47.3–99.6)
Specificity, % (95% CI)	69.3 (56.3–80.4)	88.71 (78.1–95.3)	72 (50.6–87.9)	84.3 (63.9–95.4)
PPV, % (95% CI)	29.6 (13.7–50.1)	50 (95% CI 23.0–76.)	53.3 (26.5–78.7)	63.6 (30.7–89.0)
NPV, % (95% CI)	100 (91.7–100)	98.2 (90.4–99.9)	100 (81.4–100)	95.4 (77.1–99.8)
Diagnostic accuracy, % (95% CI)	72.8 (60.9–82.8)	88.5 (78.7–94.9)	78.7 (61.0–91.0)	84.8 (68.1–94.8)
Youden index (95% CI)	0.69 (0.19–0.8)	0.76 (0.25–0.95)	0.72 (0.14–0.88)	0.72 (0.11–0.95)
AUC (95% CI)	0.88 (0.76–0.97)	0.93 (0.84–0.99)	0.89 (0.76–0.99)	0.93 (0.82–1.00)
P value for AUC comparison	0.16		0.32	

Performances of the Index Test (ES) for Discrimination of MRI Lesions

Table 3 summarises diagnostic performances for the ES and regression model for both the A and B analyses.

For analysis A, there was a significant difference ($P < 0.001$) in the ES between positive VOIs for SPCa (mean [SD] 17.73 [1.23] NAT) and those negative (mean [SD] 15.33 [0.78] NAT).

ROC curves were built for the ES, entropy and mean value of each individual MRI parameter for detection of SPCa. ROC curves were generated and AUCs calculated from a total of 240 imaging volumes representing 70 VOIs from MRI suspicious lesions and normal tissue (Fig. 2). The AUC results and comparison are summarised in Table S4.

The ES achieved an AUC of 0.88 (95% CI 0.76, 0.97) and was higher than any other single parameter. The best threshold for ES was 16.69 NAT, with a sensitivity of 100% (95% CI 63.06, 100) and specificity of 69.35% (95% CI 56.35, 80.44). The diagnostic accuracy was 72.86 (95% CI 60.9, 82.8). The Youden's index was 0.69 (95% CI 0.19, 0.80).

For analysis B, there was a significant difference ($P = 0.001$) in the ES between VOIs positive for SPCa (mean [SD] 17.21 [1.00] NAT) and those that were negative (mean [SD] 16.11 [1.19] NAT).

The same ROC curves (Fig. 2) were built for the 132 imaging volumes representing the 33 MRI lesions (targets of score >2 and cancer). The AUC for the ES for detection of SPCa was 0.89 (95% CI 0.76, 0.99), with a trend to be higher than other entropy or quantitative value for any other parameter (Table S4).

The best ES threshold was found to be 16.61 NAT, with a sensitivity of 100% and specificity of 72%.

Using this threshold of 16.61 NAT for the detection of SPCa, we calculated a NPV of 100% and PPV of 53.3%. The diagnostic accuracy was 78.79% (95% CI 61.09, 91.02). The Youden index of the ES was 0.72 (95% CI 0.14, 0.88).

The AUC of the LR model using entropy K^{trans}, entropy VE, entropy T2WI, mean ADC did not show significant differences to the AUC of the ES with a value of 0.93 for both the A and B analyses.

Performances of Index Against Standard Test: ES vs VS

The ROC curve was plotted for VS in the B analysis. The AUC reached 0.91, which was not significant either with the ES ($P = 0.75$) or the LR model ($P = 0.74$), as shown in Fig. 2.

In all, 18 MRI lesions over the 33 were under the ES threshold of 16.61 NAT from analysis B. None of them harboured SPCa (Table 4).

Fig. 2 ROC curves: **(A)** for detection of SPCa using the ES and each of its components within the MRI lesions and **(B)** for the ES (AUC 0.89) and VS (0.91) plotted for the B analysis ($P = 0.75$)
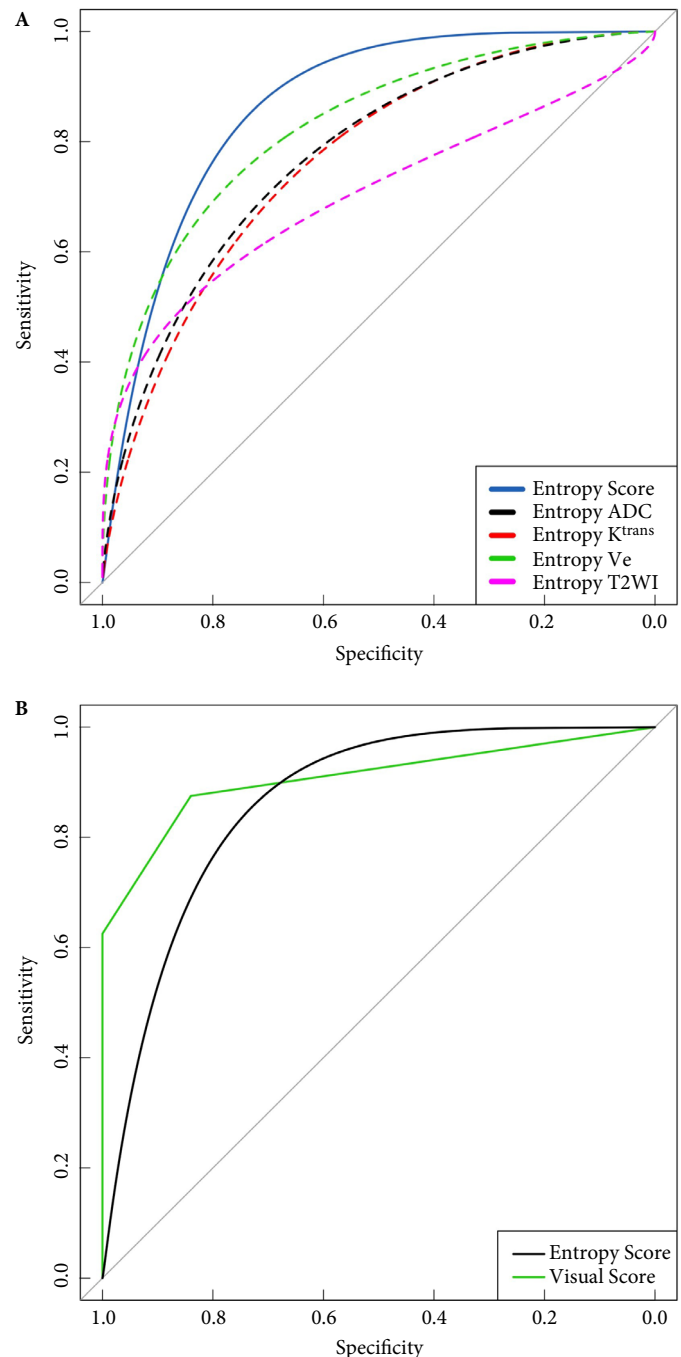
The proportion of lesions found to be positive using this ES threshold (53%) was significantly higher ($P < 0.001$) than positive lesions with VS (24% of score 3, 4, 5) in a paired analysis using the McNemar test.

An estimated detection rate for score 3, 4 and 5 MRI lesions would have been 52%, 100% and 100%, respectively,

Fig. 3 Two examples of high and low ES of cancerous lesions of Gleason (3 + 3) in A and (3 + 4) in B with an ES of 14.71 NAT and 18.75 NAT, respectively.
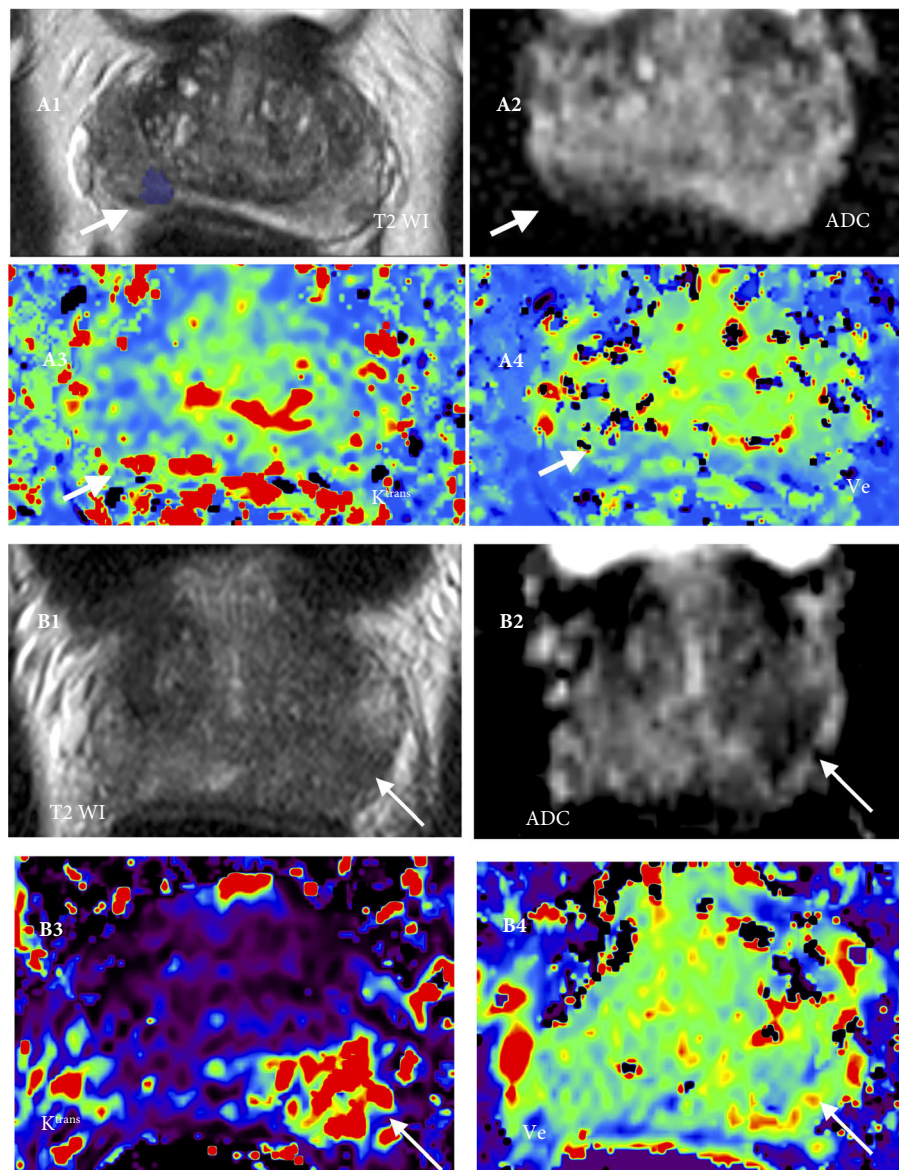
Table 4 Contingency table for lesion detection by VS and ES.

	VS	ES	
	3, 4 and 5	**<16.61 NAT**	**>16.61 NAT**
Positive for SPCa	8	0	8
Negative for SPCa	25	18	7
Number of lesions	33	18	15
Detection rate	0.24*	0	0.53*

Significant difference P < 0.001.

and would result in avoiding sampling in 54% of the MRI lesions if the estimated ES threshold was applied in this data set. Demonstrative exemples are shown in Fig. 3.

Discussion

In the present report, we focus on MRI-generated lesions for the detection and stratification of SPCa. We used the concept of radiomics and texture analysis to enhance the association of imaging findings with SPCa. A particularity of this preliminary report relies on the choice of a feature, heterogeneity, which can be both identified in the biology of cancer and imaging features. We showed an association of a texture feature, heterogeneity, with the presence of SPCa within an MRI lesion.

Our present report illustrates the growing interest in the use of radiomics to mine the amount of information that medical imaging presents to increase diagnostic accuracy [12,31,32].

Also, recent developments in the field of image analysis permit us to envisage automatic segmentation of prostate MRI [9,33,34]. It is therefore critical to link imaging findings to biology and cancer outcomes [35].

In oncology, intratumour heterogeneity drives neoplastic progression [13]. Regarding prostate cancer, recent reports further explored the heterogeneity of prostate cancer at molecular and genetic analyses [17], even when considering the intratumoral level [15,16]. It has been previously reported that heterogeneity of prostate cancer is linked to prognosis [15].

The MRI signature of prostate cancer is complex, heterogeneous, and still not fully decrypted. Neovascularisation, cellularity, proportion of histological components are reported to quantitatively and qualitatively impact overall detection of prostate cancer across different sequences [36,37]. Even the subtype of Gleason influences detectability of cancer foci by mpMRI [38].

We questioned in the present experiment the association between this biological feature of heterogeneity and the MRI signature of SPCa using texture analysis.

The present study presents original aspects in the unified analysis of texture of three different sequences, incorporating T2WI, DWI, DCE imaging, within one quantitative score capturing heterogeneity of distinct tissue component of prostate cancer for the purpose of validation of MRI lesions for biopsy. Capturing radiomics features across multiple MRI sequences has already been described for breast, liver and nasopharyngeal carcinomas and even linked to stratification and prognosis of the disease [20,39,40]. To our knowledge no unified score based on radiomics features extracted from MRI findings has been described for this purpose. The present work adds to the hypothesis of texture analysis as a link between the biology of carcinomas and their visualisation using MRI.

There was a significant difference in the ES between SPCa and benign tissue, and also with insignificant cancer foci. The clinical potential relevance for stratification of the disease is further illustrated by the strong positive correlation with the Gleason score, an important driver of oncological outcome [41], and maximum cancer core length. The ES had a stronger correlation ($r_s = 0.568$; $P = 0.03$) with Gleason score than previous reports of inverse correlation of ADC of $r = -0.376$ by Oto et al. [41] or $r = -0.39$ by Verma et al. [42]. Entropy of ADC for a whole lesion was found to be an independent predictor of biochemical failure by Rosenkrantz et al. [24].

Supporting the potential value of radiomics, Vignati et al. [43] found an excellent AUC of 0.96 for a texture parameter based on T2WI alone for differentiating Gleason 6 and >6 in a radical prostatectomy cohort, limiting translation in a diagnostic setting.

In a recent report, Ginsburg et al. [44] trained a model based on numerous radiomics features in multi-institutional data-set for detection of cancer on a voxel basis. Their maximal AUC of 0.71 was lower than our present results and based on a single sequence analysis. Rosenkrantz et al. [45] applied different features of first order to discriminate lesions with Gleason >6 to others amongst VS lesions. Harbouring similar range of values to our present results, entropy of ADC alone failed to discriminate those even in their larger cohort.

In our present study, the lack of significance of the entropy of each sequence taken individually (Fig. 2), underlies that heterogeneity assessment by MRI should be performed through a multiparametric approach, as diagnosis is performed with VS [46,47].

As a multi-sequence strategy, the ES achieved similar and very good performance to a LR model integrating imaging features (including mono-sequential entropy), by comparison of AUC (0.93 vs 0 .89, $P = 0.32$). The ES outperformed the LR model when considering NPV (100%) and Se (100%). This is an important finding to envisage generalisation of the ES. A LR model has to be calibrated for a given population when the ES needs calibration of the sequences acquisition, which can be achieved with phantoms [32]. Pending validation studies, we could envisage development of pre-calibrated phantom to setup MRI scanners of different magnetic field strengths associated with automated reporting of MRI.

The VS has already been reported to be of utility for detection of prostate cancer using a targeted approach in the PRostate Evaluation for Clinically Important Disease: Sampling Using Image-guidance Or Not? (PRECISION) trial, a multicentre randomised control trial against random TRUS-guided biopsy [2]. Detection of significant disease was 12% higher with targeted biopsy compared to standard TRUS-guided biopsy. The proportion of insignificant disease was also lower using the MRI-guided strategy.

The PRECISION trial also successfully tested the strategy to avoid biopsy in 28% of patients with negative imaging findings [2]. However, VS of mpMRI was described with variability, especially in NPV ranging from 63% to 95% [48]. A high and replicable NPV would avoid the unnecessary sampling of some lesion without missing SPCa, whilst high sensitivity is conserved.

The process leading to visualisation of MRI lesions and radiologists' general impression appear to be difficult to systematise as shown by the non-negligible inter-observer variability of the VS with either the Likert or PI-RADS version 2 scoring systems [1,2,7]. The ES, as a quantitative score, offers to overcome this limitation. The ES is, by definition, little sensitive to volume segmentation.

Performance of the ES was comparable to the VS with AUCs of 0.89 and 0.91, respectively ($P = 0.75$). With an optimised

threshold, the ES significantly showed a higher proportion (54%) of positive lesion for SPCa than the VS (24%), without missing SPCa. Depending of the PI-RADS version 2 score, the rate of false positives generated by the VS has been reported to range from 17% to 88% in a population undergoing targeted biopsy. The positive association of the ES from MRI lesions to SPCa offers the possibility to select the MRI-generated lesions to target with biopsy.

Automated texture analysis attempts to replicate a deep learning process empowered by computer calculation capabilities, but is highly dependent on the quality of input data and their clinical relevance [10]. The current workflow leading to computation of the ES can be fully automated.

As a potential tool, longitudinal analysis of the ES in an MRI-guided active surveillance population might be of interest to detect progression of cancer leading to change in management [49]. Even entropy from other imaging modalities, e.g. positron emission tomography, could possibly be added and extend performance.

Our choice of a biopsy population permits avoiding the bias of selection of patients undergoing radical prostatectomy, with inclusion of patients potentially eligible for active surveillance and without the diagnosis of prostate cancer. The reference standard, providing ground truth, meets the criteria of standard of care by sampling systematically the whole gland with addition of targeting MRI lesions. In a recent meta-analysis, no difference was shown between cognitive and fusion biopsy for detection rates [50].

Limitations

Regardless the high number of generated imaging volumes, the cohort is relatively small even if some statistical differences have been found either for quantitative values or the ES. This early proof-of-concept study was powered to detect a significant difference in quantitative parameters, not specifically the ES. Technically, mpMRI did not include high b value sequences and was acquired at 1.5 T. Some refinements might increase the accuracy of the ES, e.g., using a 3-T scanner to increase the signal-to-noise ratio. The ES for selection of targets was tested for lesions generated by VS, the standard test for reporting mpMRI. A fully automated workflow to compute the ES can be implemented and would currently rely on generation of MRI lesions with human assistance, by its segmentation on a single sequence, awaiting validation of automation of this step by the means of artificial neural networks.

Also, the computed threshold in the present study depends on the chosen conservative definition of SPCa, including Gleason 6. Variation of this debated definition will probably impact performance [51].

Research Implications

This early proof-of-concept study has multiple research implications. Validation of imaging biomarkers has to follow an established process [35]. Principally, the association of the ES, heterogeneity at imaging, has to be confirmed as a meaningful signature of SPCa based on biological features and ultimately patient outcomes [52]. Testing this score in a larger, prospectively acquired biopsy naïve population with a robust reference test is a mandatory. We plan to confront those preliminary findings for detection in the PROstate MRI Imaging Study (PROMIS) trial cohort that fulfils those criteria in a retrospective analysis [1]. This is a compulsory step before undertaking a prospective validation study in a new population. For longitudinal analysis, quantitative assessment of progression using the ES of MRI lesions in a surveillance setting would first need a longitudinal descriptive study linked to clinical outcomes. This would require a robust histological 'gold standard', where an MRI lesion is systematically re-sampled at different time points. A deep understanding of the heterogeneity of prostate tissue, either benign or cancerous, is needed at both the histology and biomolecular level of analysis. There is a need for high-quality data to input in texture analysis platforms [10]. Other and more complex texture analysis features might outperform the entropy and need to be tested [31].

Conclusion

We report results of a proof-of-concept study, which shows association between a quantitative texture-based ES, linked to heterogeneity across MRI sequences, and SPCa. For similar overall performances to the VS, the ES presented a higher proportion of positive lesions for SPCa without compromising NPV and Se. This permits us to envisage selection of MRI-generated targets for biopsy. Multiple steps are needed to validate these initial findings as an imaging biomarker.

Conflict of Interests

References

1 **Ahmed HU**, **El-Shater Bosaily A**, **Brown LC et al.** Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* 2017; 389: 815–22.

2 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

3 Moldovan PC, Van den Broeck T, Sylvester R et al. What is the negative predictive value of multiparametric magnetic resonance imaging in excluding prostate cancer at biopsy? A systematic review and meta-analysis from the European association of urology prostate cancer guidelines panel. *Eur Urol* 2017; 72: 250–66

4 Martin RM, Donovan JL, Turner EL et al. Effect of a low-intensity PSA-based screening intervention on prostate cancer mortality: the CAP randomized clinical trial. *JAMA* 2018; 319: 883–95

5 Andriole GL, Crawford ED, Grubb RL et al. Prostate cancer screening in the randomized prostate, lung, colorectal, and ovarian cancer screening trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst* 2012; 104: 125–32

6 Schröder FH, Hugosson J, Roobol MJ et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* 2014; 384: 2027–35

7 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

8 Azahaf M, Haberley M, Betrouni N et al. Impact of arterial input function selection on the accuracy of dynamic contrast-enhanced MRI quantitative analysis for the diagnosis of clinically significant prostate cancer. *J Magn Reson Imaging* 2016; 43: 737–49

9 Drozdzal M, Chartrand G, Vorontsov E et al. Learning normalized inputs for iterative estimation in medical image segmentation. *Med Image Anal* 2018; 44: 1–13

10 Litjens G, Kooi T, Bejnordi BE et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017; 42: 60–88

11 Depeursinge A, Foncubierta-Rodriguez A, Van De Ville D, Müller H. Three-dimensional solid texture analysis in biomedical imaging: review and opportunities. *Med Image Anal* 2014; 18: 176–96

12 Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278: 563–77

13 Andor N, Graham TA, Jansen M et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* 2016; 22: 105–13

14 Epstein JI, Egevad L, Amin MB et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol* 2016; 40: 244–52

15 Cyll K, Ersvær E, Vlatkovic L et al. Tumour heterogeneity poses a significant challenge to cancer biomarker research. *Br J Cancer* 2017; 117: 367–75

16 Wei L, Wang J, Lampert E et al. Intratumoral and intertumoral genomic heterogeneity of multifocal localized prostate cancer impacts molecular classifications and genomic prognosticators. *Eur Urol* 2017; 71: 183–92

17 Abeshouse A, Ahn J, Akbani R et al. The molecular taxonomy of primary prostate cancer. *Cell* 2015; 163: 1011–25

18 Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol* 2004; 59: 1061–9

19 Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948; 27: 379–423

20 Kim JH, Ko ES, Lim Y et al. Breast cancer heterogeneity: MR imaging texture analysis and survival outcomes. *Radiology* 2016; 282: 665–75

21 Wibmer A, Hricak H, Gondo T et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur Radiol* 2015; 25: 2840–50

22 Rosenkrantz AB, Triolo MJ, Melamed J, Rusinek H, Taneja SS, Deng FM. Whole-lesion apparent diffusion coefficient metrics as a marker of

percentage Gleason 4 component within Gleason 7 prostate cancer at radical prostatectomy. *J Magn Reson Imaging* 2014; 41: 708–14

23 Sidhu HS, Benigno S, Ganeshan B et al. Textural analysis of multiparametric MRI detects transition zone prostate cancer. *Eur Radiol* 2016; 27: 2348–58

24 Rosenkrantz AB, Ream JM, Nolan P, Rusinek H, Deng FM, Taneja SS. Prostate cancer: utility of whole-lesion apparent diffusion coefficient metrics for prediction of biochemical recurrence after radical prostatectomy. *AJR Am J Roentgenol* 2015; 205: 1208–14

25 Rosenkrantz AB, Mendrinos S, Babb JS, Taneja SS. Prostate cancer foci detected on multiparametric magnetic resonance imaging are histologically distinct from those not detected. *J Urol* 2012; 187: 2032–8

26 Barentsz JO, Richenberg J, Clements R et al. ESUR prostate MR guidelines 2012. *Eur Radiol* 2012; 22: 746–57

27 Rosenkrantz AB, Kim S, Lim RP et al. Prostate cancer localization using multiparametric MR imaging: comparison of Prostate Imaging Reporting and Data System (PI-RADS) and Likert Scales. *Radiology* 2013; 269: 482–92

28 Barentsz J, Villers A, Schouten M. Reply to Letter to the Editor re: ESUR prostate MR guidelines. *Eur Radiol* 2013; 23: 2322–3

29 Haffner J, Lemaitre L, Puech P et al. Role of magnetic resonance imaging before initial biopsy: comparison of magnetic resonance imaging-targeted and systematic biopsy for significant prostate cancer detection. *BJU Int* 2011; 108: E171–8

30 Epstein JI, Walsh PC, Carmichael M, Brendler CB. Pathologic and clinical findings to predict tumor extent of nonpalpable (Stage T1 c) prostate cancer. *JAMA* 1994; 271: 368–74.

31 Lambin P, Leijenaar RT, Deist TM et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; 14: 749–62

32 Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In IEEE, 2016: 565–71. Available at: http://ieeexplore.ieee.org/document/7785132/. Accessed Jan 2018

33 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* [Internet]. Cham: Springer International Publishing, 2015: 234–41. Available at: http://link.springer.com/10.1007/978-3-319-24574-4_28. Accessed Jan 2018

34 O'Connor JP, Aboagye EO, Adams JE et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 2016; 14: 169–86

35 van Niekerk CG, van der Laak JA, Börger ME et al. Computerized whole slide quantification shows increased microvascular density in pT2 prostate cancer as compared to normal prostate tissue. *Prostate* 2009; 69: 62–9

36 van Niekerk CG, van der Laak JA, Hambrock T et al. Correlation between dynamic contrast-enhanced MRI and quantitative histopathologic microvascular parameters in organ-confined prostate cancer. *Eur Radiol* 2014; 24: 2597–605

37 Truong M, Hollenberg G, Weinberg E, Messing EM, Miyamoto H, Frye TP. Impact of Gleason subtype on prostate cancer detection using multiparametric magnetic resonance imaging: correlation with final histopathology. *J Urol* 2017; 198: 316–21

38 Hectors SJ, Wagner M, Bane O et al. Quantification of hepatocellular carcinoma heterogeneity with multiparametric magnetic resonance imaging. *Sci Rep* 2017; 7: 2452. https://doi.org/10.1038/s41598-017-02706-z

39 Zhang B, Tian J, Dong D et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res* 2017; 23: 4259–69

40 Choy B, Pearce SM, Anderson BB et al. Prognostic significance of percentage and architectural types of contemporary Gleason pattern 4

prostate cancer in radical prostatectomy. *Am J Surg Pathol* 2016; 40: 1400–6.

41 Oto A, Yang C, Kayhan A et al. Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: correlation of quantitative MR parameters with Gleason score and tumor angiogenesis. *AJR Am J Roentgenol* 2011; 197: 1382–90

42 Verma S, Rajesh A, Morales H et al. Assessment of aggressiveness of prostate cancer: correlation of apparent diffusion coefficient with histologic grade after radical prostatectomy. *AJR Am J Roentgenol* 2011; 196: 374–81

43 Vignati A, Mazzetti S, Giannini V et al. Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness. *Phys Med Biol* 2015; 60: 2685–701

44 Ginsburg SB, Algohary A, Pahwa S et al. Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: preliminary findings from a multi-institutional study. *J Magn Reson Imaging* 2017; 46: 184–93

45 Rosenkrantz AB, Meng X, Ream JM et al. Likert score 3 prostate lesions: association between whole-lesion ADC metrics and pathologic findings at MRI/ultrasound fusion targeted biopsy. *J Magn Reson Imaging* 2016; 43: 325–32

46 Dickinson L, Ahmed HU, Allen C et al. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. *Eur Urol* 2011; 59: 477–94

47 Weinreb JC, Barentsz JO, Choyke PL et al. PI-RADS prostate imaging-Reporting and data system: 2015, version 2. *Eur Urol* 2016; 69: 16–40

48 Fütterer JJ, Briganti A, De Visschere P et al. Can clinically significant prostate cancer be detected with multiparametric magnetic resonance imaging? A systematic review of the literature. *Eur Urol* 2015; 68: 1045–53

49 Moore CM, Giganti F, Albertsen P et al. Reporting magnetic resonance imaging in men on active surveillance for prostate cancer: the PRECISE Recommendations-A report of a European School of Oncology Task Force. *Eur Urol* 2017; 71: 648–55

50 Wegelin O, van Melick HH, Hooft L et al. Comparing three different techniques for magnetic resonance imaging-targeted prostate biopsies: a systematic review of in-bore versus magnetic resonance imaging-transrectal ultrasound fusion versus cognitive registration. Is there a preferred technique? *Eur Urol* 2017; 71: 517–31.

51 Netto GJ. Tumor volume threshold of insignificant prostate cancer – was Dr. Stamey right all along? *J Urol* 2011; 185: 10–1.

52 D'Amico AV. Active surveillance versus treatment of prostate cancer: should metastasis be the primary end point? *J Clin Oncol* 2017; 35: 1638–40

Correspondence: Clement Orczyk, Associate Professor of Urology, Honorary Consultant Urological Surgeon, University College London, Charles Bell House, 43-45 Foley St, London W1W 7TS, UK.

e-mails: clementorczyk@yahoo.fr and c.orczyk@ucl.ac.uk

Abbreviations: (N)(P)PV, (negative) (positive) predictive value; (n)PZ, (normal) peripheral zone; 3D, three-dimensional; ADC, apparent diffusion coefficient; AUC, area under the curve; BPHn, BPH nodules; DCE, dynamic contrast-enhanced; DWI, diffusion-weighted imaging; ES, entropy score; K^{trans}, volume transfer constant; LR, logistic regression; mpMRI, multiparametric MRI; NAT, NATural unit of information; PI-RADS, Prostate Imaging - Reporting And Data System; PRECISION, PRostate Evaluation for Clinically Important Disease: Sampling Using Image-guidance Or Not? (trial); ROC, receiver-operating characteristic; Se, test sensitivity; SPCa, clinically significant prostate cancer; Sp, test specificity; T2WI, T2-weighted imaging; TZ, transition zone; Ve, extracellular extravascular space; VOI, volume of interest; VS, visual scoring.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article:

Fig. S1. Image processing workflow. **Fig. S2.** Comparison using boxplot of entropy for the different MRI features between significant cancerous and non-cancerous targets.
Table S1. Parameters of mpMRI sequences.
Table S2. Descriptive results of MRI quantitative parameters and texture metrics by MRI feature and type of VOI.
Table S3. Raw values and entropy in paired analysis for cancerous and matching benign tissue.
Appendix S1. Computing quantitative maps using the pharmacokinetic (PK) Toft model.
Appendix S2. Assessment of consistency of histograms.
Appendix S3. Entropy formula and calibration.

86 BJU International © 2018 BJU International