



OPEN

Evaluation of a multiparametric MRI radiomic-based approach for stratification of equivocal PI-RADS 3 and upgraded PI-RADS 4 prostatic lesions

Valentina Brancato^{1✉}, Marco Aiello¹, Luca Basso¹, Serena Monti³, Luigi Palumbo², Giuseppe Di Costanzo², Marco Salvatore¹, Alfonso Ragozzino² & Carlo Cavaliere¹

Despite the key-role of the Prostate Imaging and Reporting and Data System (PI-RADS) in the diagnosis and characterization of prostate cancer (PCa), this system remains to be affected by several limitations, primarily associated with the interpretation of equivocal PI-RADS 3 lesions and with the debated role of Dynamic Contrast Enhanced-Magnetic Resonance Imaging (DCE-MRI), which is only used to upgrade peripheral PI-RADS category 3 lesions to PI-RADS category 4 if enhancement is focal. We aimed at investigating the usefulness of radiomics for detection of PCa lesions (Gleason Score ≥ 6) in PI-RADS 3 lesions and in peripheral PI-RADS 3 upgraded to PI-RADS 4 lesions (upPI-RADS 4). Multiparametric MRI (mpMRI) data of patients who underwent prostatic mpMRI between April 2013 and September 2018 were retrospectively evaluated. Biopsy results were used as gold standard. PI-RADS 3 and PI-RADS 4 lesions were re-scored according to the PI-RADS v2.1 before and after DCE-MRI evaluation. Radiomic features were extracted from T2-weighted MRI (T2), Apparent diffusion Coefficient (ADC) map and DCE-MRI subtracted images using PyRadiomics. Feature selection was performed using Wilcoxon-ranksum test and Minimum Redundancy Maximum Relevance (mRMR). Predictive models were constructed for PCa detection in PI-RADS 3 and upPI-RADS 4 lesions using at each step an imbalance-adjusted bootstrap resampling (IABR) on 1000 samples. 41 PI-RADS 3 and 32 upPI-RADS 4 lesions were analyzed. Among 293 radiomic features, the top selected features derived from T2 and ADC. For PI-RADS 3 stratification, second order model showed higher performances (Area Under the Receiver Operating Characteristic Curve—AUC—= 80%), while for upPI-RADS 4 stratification, first order model showed higher performances respect to superior order models (AUC = 89%). Our results support the significant role of T2 and ADC radiomic features for PCa detection in lesions scored as PI-RADS 3 and upPI-RADS 4. Radiomics models showed high diagnostic efficacy in classify PI-RADS 3 and upPI-RADS 4 lesions, outperforming PI-RADS v2.1 performance.

Prostate cancer (PCa) is one of the most commonly diagnosed malignant neoplasms among men¹. Multiparametric Magnetic Resonance Imaging (mpMRI) has gradually gained in importance for both a timely diagnosis and an accurate characterization of PCa lesions, which play a key-role in all PCa patient management steps^{2,3}.

With the goal of standardizing the acquisition and reporting of prostatic mpMRI imaging examinations, the European Society of Urogenital Radiology (ESUR) developed the Prostate Imaging-Reporting and Data System (PI-RADS) in 2013 and then updated it in 2015 (PI-RADS v2) and 2019 (PI-RADS v2.1)⁴.

PI-RADS evaluation is based on a 5-point scale associated with the probability that a combination of findings on mpMRI modalities (namely T2-weighted—T2, Diffusion Weighted MRI—DWI and Dynamic Contrast-Enhanced MRI—DCE-MRI, abbreviated) correlates with the presence of a clinically significant cancer for detected prostatic lesion. PI-RADS score ranges between 1 and 5, respectively indicating a very low and a very high likelihood that a lesion is malignant. The PI-RADS classification had a crucial role in PCa management since

¹IRCCS SDN, Naples, Italy. ²Department of Radiology, S. Maria Delle Grazie Hospital, Pozzuoli, Italy. ³Institute of Biostructures and Bioimaging, National Research Council, Naples, Italy. ✉email: valentina.brancato@synlab.it

its development, and has proven to be a powerful tool for the identification and aggressiveness characterization of prostatic lesions^{5–8}.

Although encouraging results have been reported in the literature on the role of the PI-RADS in the diagnosis and characterization of PCa, this system remains to be affected by several limitations, primarily associated with the interpretation of PI-RADS category 3 lesions, namely those lesions on prostate MRI that are termed as ‘intermediate’ or ‘equivocal on the presence of clinically significant cancer’⁴.

Despite the increase in diagnostic accuracy, reproducibility and score assignment easing of PI-RADS v2 and v2.1 compared to the first version, limitations related to risk stratification of PI-RADS 3 lesions still remain, since guidelines do not say how to deal with such imaging findings that are indeterminate^{4,9,10}. PI-RADS 3 lesions present a challenge to treating urologists who must decide the optimal management option, namely whether to monitor with follow-up prostate-specific antigen (PSA) testing and imaging, or schedule immediate biopsy. An accurate and calibrated patient selection for prostate biopsy is essential in order to avoid unnecessary biopsies⁹. This equivocal imaging characterization may lead to high variability in practice patterns, costs, and potentially clinical outcomes across different institutions. Studies aimed at stratifying PI-RADS 3 lesions are limited^{9,11–15}. Another limitation directly affecting PI-RADS 3 lesion assignment concerns the role of DCE-MRI, which is still regarded as very controversial and debated, and its added value in combination with T2 and DWI was still not clearly assessed. Currently, according to the new PI-RADS v2.1, DCE-MRI is only used to upgrade PI-RADS category 3 lesions to PI-RADS category 4, but only for lesions located in peripheral zone (PZ)⁴. However, the diagnostic value of DCE-MRI in prostate MRI remains subject of the current debate and it must be shown how many additional cases of PCa are found in PI-RADS 3 upgraded to PI-RADS 4 lesions. Radiomics, the extraction of multiple quantitative imaging features from medical images, represents an attractive tool which could overcome the clinical challenge associated with radiologist uncertainties related to PI-RADS 3 lesions and PI-RADS 3 upgraded to PI-RADS 4 lesions. Radiomic tool has been widely explored in the field of PCa and led to promising results, but especially in studies aiming at differentiating between normal and cancerous prostatic tissue, characterizing PCa lesions in terms of aggressiveness according to Gleason Score (GS), and also comparing diagnostic power of radiomic features with that of PI-RADS scoring^{16–22}. However, to our knowledge, only Giambelluca et al.²³ applied radiomic approach to stratify PI-RADS 3 lesions, and there are any studies aiming at investigating the power of radiomics in stratify PI-RADS 3 upgraded to PI-RADS 4.

In this context, we aimed at investigating the usefulness of radiomics for detection of PCa (GS ≥ 6) in PI-RADS category 3 lesions and in PI-RADS 3 upgraded to PI-RADS 4 lesions (upPI-RADS 4) in PZ.

Methods and materials

Patient population. We performed retrospective analysis of all mpMRI data of patients who underwent mpMRI of the prostate between April 2013 and September 2018 due to elevated PSA level and/or clinical suspicion of PCa and subsequently biopsy. mpMRI images and histopathology records were collected at H.S. Maria delle Grazie, Italy and informed consent was given before Magnetic Resonance (MR) examination²⁴. The study was conducted in accordance with the Declaration of Helsinki, and the study protocol was approved by the Ethics Committee of the Istituto Nazionale Tumori “Fondazione G. Pascale (protocol number 1/20). PI-RADS 1, 2 and 5 lesions were excluded. Examinations where a PI-RADS 3 lesion was present together with PI-RADS 4 or 5 lesion were excluded. Biopsy results were used as gold standard.

MRI protocol. Routine clinical mpMRI acquisition includes T2, DCE-MRI, and DWI. The DWI includes an apparent diffusion coefficient (ADC) map generated at the time of acquisition. Patients were injected with contrast agent Gadoteridol (Gd-HP-DO3A; Pro Hance, Bracco Diagnostics, Princeton, NJ, USA) with a dose of 0.1 mL/kg before DCE-MRI acquisition. All patients were imaged using MAGNETOM-Avanto scanner (Siemens Healthcare, Erlangen, Germany) at 1.5 T with both endorectal coil and phase-array pelvic coil²⁴. More details on the technical parameters of the MRI sequences are shown in Supplementary Table S1.

PI-RADS assignment. After exclusion of PI-RADS 1, 2 and 5, imaging findings of the remaining patients were re-scored according the PI-RADS v2.1 by two radiologists, respectively with 10 and 8 years of experience. They independently reviewed MR images blinded to the results of biopsy and any clinical information. Evaluation of MR images was performed in two reading-sessions, following a similar procedure performed in our previous study²⁴: the first reading-session was performed considering a biparametric MRI (bpMRI) protocol consisting of axial, sagittal and coronal T2 images and axial DWI images with their corresponding ADC maps and a b-computed image with $b = 1400 \text{ s/mm}^2$; the second reading-session was performed after 2 weeks considering the entire mpMRI protocol, and rearranging patient IDs in a different order, to reduce memory bias. Discrepancies were resolved by consensus. For both reading sessions, we assessed the inter-reader agreement between the two radiologists based on the ratings assigned before the consensus using the weighted Cohen’s kappa (κ) with linear weights^{25–27}. The strength of agreement was evaluated as excellent if $\kappa = 0.81–1.00$, good if $\kappa = 0.61–0.80$, moderate if $\kappa = 0.41–0.60$, fair if $\kappa = 0.21–0.40$, and poor if $\kappa = 0–0.20$. After reaching consensus, PI-RADS 3 scored lesions after mpMRI evaluation and PI-RADS 3 lesions upgraded to PI-RADS 4 after DCE-MRI evaluation were selected.

Biopsy protocol. All prostatic biopsies were TRUS-guided and performed using an 18-gauge tru-cut needle, under anesthesia. Each patient underwent both systematic biopsies, with an average of 12 random samples of the entire prostate gland, and target biopsy, with at least three samples taken from each lesion identified by MRI. The number of randomly-taken samples could vary depending on dimensions of prostate gland, as well as number of target samples could do depending on dimension of each lesion. Target sampling was performed with

an MRI/TRUS fusion, using alternately the cognitive technique or dedicated software, coupled with ultrasound platforms from various companies²⁴. The gross description included number and core lengths of needle biopsies. The specimens were fixed in buffered 10% formalin, and routinely processed. Thin sections of four microns were cut and stained with hematoxylin and eosin stain (H&E). Supplementary sections were performed for possible immunohistochemical stains to prove the loss of basal cells in small focus of cancer (p63, and high molecular weights keratin) combined with other antibodies overexpress in prostatic cancer (anti-AMACR/p504S). One senior pathologist (with more than 10 years of experience in prostate specimen interpretation) who was blinded to the MRI reports, reviewed pathological slices and classified tumors according to the 4th WHO classification, further grading them by Gleason scores and the group grade cancer^{28–30}. The final report also included tumor extent in each needle biopsies and the percent core involvement by tumor.

Image preprocessing and 3D ROI segmentation. ADC images were non-rigidly coregistered on T2 image using Elastix software (v. 4.9.0) in order to correct for typical spatial distortion arising from DWI acquisition. Subtraction DCE-MRI images were all resliced on T2 images. Two experienced radiologists were asked to consensually draw 3D regions of interest (ROIs) in the biopsied lesions with PI-RADS3 and upPI-RADS4, while also looking at the $b = 1000$ coregistered volume. Lesion segmentation was performed on T2 images using an in-house developed software for region labeling. During the segmentation procedure, the radiologists were blinded to both the histological results and all clinical information relative to the retrospective prostate mpMR images. Prior to radiomic features extraction, normalization was applied on T2 images intensities. Specifically, intensities were normalized by centering them at their respective mean value with standard deviation of all gray values in the original image^{31–34}.

Radiomic analysis. Feature extraction. Extraction of radiomic features from 3D ROIs on T2, registered ADC and resliced subtraction DCE-MRI images was performed using PyRadiomics, an open-source package for standardizing the extraction of radiomics data³². The extracted features can be classified into three classes: shape ($n = 14$), first-order statistics ($n = 18$), and second order textural statistics including grey level cooccurrence matrix—glcm—($n = 24$), grey level run length matrix—glrlm—($n = 16$), grey level size zone matrix—glszm—($n = 16$), neighbouring gray tone difference matrix—ngtdm—($n = 5$) and gray level dependence matrix—gldm—($n = 14$). Detailed description and computing algorithms of the radiomic features were available at <https://pyradiomics.readthedocs.io>. The first and second order textural features were computed for T2, ADC and the DCE-MRI subtraction series with the highest mean signal intensity within the ROI^{35,36}. Finally, a total of 293 features were extracted for each lesion, namely 14 shape features and, for each of the three MRI images (T2, ADC, DCE-MRI), 93 features including first- and second-order features. The full list of extracted radiomic features is provided in Supplementary Table S2. Two classification tasks were investigated: PCa/non-PCa PI-RADS3 and PCa/non-PCa upPI-RADS 4. So, procedures described in the following paragraphs are to be considered per classification task.

Feature selection. Feature normalization was performed before feature selection by using z normalization. Specifically, each feature was normalized as $z = (x - \bar{x})/s$, where x , \bar{x} , and s are the feature, the mean, and the standard deviation, respectively^{37,38}. Due to the relatively small patient sample size and high-dimensional feature size, we then performed feature selection process to select features most related to biopsy outcome, in order to construct prediction models. Feature selection was performed including two steps. In the first step the feature set was restricted through a univariate analysis by using nonparametric Wilcoxon rank-sum test performed to investigate their statistical significance with respect to the outcome (PCa vs non-PCa). The significantly different features ($p < 0.05$) were then selected and further reduced in the second step using Minimum Redundancy Maximum Relevance (mRMR) algorithm. mRMR algorithm selects an optimal set of features considering both the relevance for outcome prediction and the redundancy between features, using mutual information (MI) to measure both the relevance and the redundancy. At each step of mRMR feature selection process, the feature with the highest predictor importance score (defined as the difference between MI between outcome and the considered feature and the average MI of previously selected feature and the considered feature) will be added to the selected feature set^{39,40}. The top five features with highest predictor importance score were finally used to construct radiomics prediction models. Feature selection procedures were implemented in MATLAB R2019b (The MathWorks Inc., Natick, MA, USA).

Multivariable model building and analysis. For each classification task, the reduced feature set was used to build logistic regression models of order from 1 to 5 that would best predict the presence of PCa using an imbalance-adjusted bootstrap resampling (IABR) approach on 1000 bootstrap samples⁴¹. Specifically, 1000 bootstrap samples were randomly drawn with replacement from the available dataset and used as training set, while instances that do not appear in the bootstrap sample are the testing set⁴². Then, the imbalance-adjustment step was applied duplicating the number of positive instances by a factor equal to the number of negative instances, and the number of negative instances by a factor equal to the number of positive instances. This operation made the probability of picking a positive and a negative instance in the bootstrap sample the same⁴¹.

For each model order, the combination of features maximizing the 0.632+ area under the receiver operating characteristic curve (AUC) within 1000 bootstrap training and testing samples was identified^{43,44}. Once optimal combination of features was identified for model orders 1–5, IABR on 1000 samples was performed again for all models in order to evaluate prediction performances.

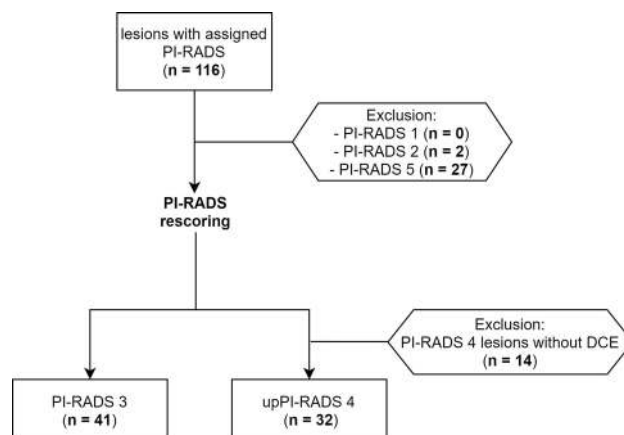


Figure 1. Flowchart showing how final cohort was selected. *PI-RADS* prostate imaging reporting and data system, *upPI-RADS 4* lesions scored as *PI-RADS 4* after mpMRI *PI-RADS* assignment but evaluated as *PI-RADS 3* after biparametric MRI reading session.

Variable	PI-RADS 3	upPI-RADS 4
Clinical variables		
No. lesions [n]	41	32
Median age [years (range)]	66 (54 – 78)	68 (54 – 80)
Mean PSA level (ng/mL)	10.36	8.56
Prostate volume (cm ³)	55.76	55.6
Prostatic zone		
PZ [n (%)]	35 (85.4)	32 (100)
TZ [n (%)]	2 (4.9)	0 (0)
CG [n (%)]	4 (9.7)	0 (0)
Biopsy results		
PCa negative [n (%)]	15 (36.58)	8 (25)
PCa positive [n (%)]	26 (63.42)	24 (75)
Gleason scores for PCa lesions [n (%)]		
3 + 3	15 (57.7)	13 (52)
3 + 4	3 (11.5)	2 (8)
4 + 3	5 (19.3)	6 (24)
4 + 4	3 (11.5)	4 (16)

Table 1. Characteristics of included patients. *PSA* prostate-specific antigene, *PCa* prostate cancer, *PZ* peripheral zone, *TZ* transition zone, *CG* central gland, *PI-RADS* prostate imaging-reporting and data system.

Results

Patients characteristics. 116 PCa lesions were finally identified. Among patients with double lesions, no one presented *PI-RADS 3* lesion together with *PI-RADS 4* or *5* lesion. 41 lesions were scored as *PI-RADS 3* and 46 as *PI-RADS 4*. For lesions evaluated as *PI-RADS 3*, *PI-RADS* reevaluation confirmed a *PI-RADS 3* score in both bpMRI and mpMRI reading session. For lesions evaluated as *PI-RADS 4*, *PI-RADS* reevaluation confirmed a *PI-RADS 4* score in mpMRI reading session. Among these, 32 lesions were scored as *PI-RADS 3* in bpMRI reading session, but were upgraded to 4 due to focal, positive DCE-MRI. The remaining 14 *PI-RADS 4* lesions were excluded from our analysis. Finally, our study results were based on 41 *PI-RADS 3* and 32 up*PI-RADS 4* lesions. See Fig. 1 for the patient selection flow chart and Table 1 for characteristics of included patients. Inter-reader agreement based on the ratings assigned before the consensus was excellent ($\kappa > 0.8$) for both sessions (See Supplementary Section S3). Of selected *PI-RADS 3* and up*PI-RADS 4* lesions, 26 and 24 were positive for PCa, respectively. Examples of *PI-RADS 3* and up*PI-RADS 4* PCa and non-PCa lesions are shown in Figs. 2 and 3.

Radiomic analysis. Univariate analysis revealed 36 and 43 statistically significant features, respectively for *PI-RADS 3* and up*PI-RADS 4* classification tasks. Statistically significant features are reported in Supplementary Tables S4 and S5. By using the mRMR method on these features, the five highest mRMR-ranked features were

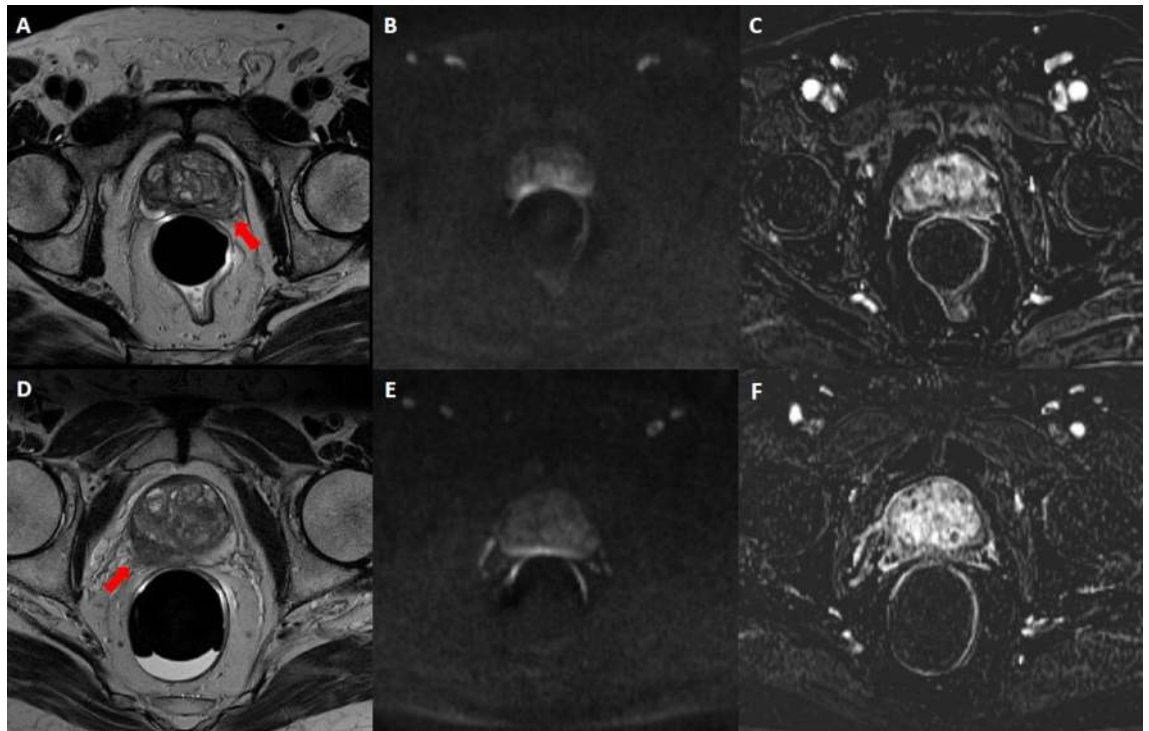


Figure 2. PI-RADS 3 scored alterations. In the first row, an example of PI-RADS 3 lesion confirmed at biopsy as GS 3 + 3: mpMRI showed reduced T2 signal intensity in the left PZ (A), altered diffusivity at b1000 DWI (B) without contrast enhancement (C). In the second row, an example of PI-RADS 3 lesion not confirmed at biopsy: mpMRI showed reduced T2 signal intensity in the right PZ (D), altered diffusivity at b1000 DWI (E) without contrast enhancement (F). PI-RADS prostate imaging-reporting and data system, mpMRI multiparametric MRI, DWI diffusion weighted MRI, PZ peripheral zone.

selected to build the prediction models. Bar plots of predictor importance score for the top five features selected by mRMR for each classification task are shown in Fig. 4.

Multivariable logistic regression models of order from 1 to 5 were obtained, and their prediction performance for the two classification tasks were reported in Supplementary Table S6 and showed in Fig. 5. By inspecting curves in Fig. 5 and values in Supplementary Table S6, we determined that the simplest multivariable model with the best prediction performances were reached by second order model for PI-RADS 3 and for first and third upPI-RADS4 classification task. For PI-RADS 3 lesion detection, second order model was chosen due to a slightly higher mean sensitivity, specificity and accuracy (80%, 51%, 71%, respectively) respect to first order model (76%, 42%, 65%, respectively), which showed comparable AUC (AUC = 74% for first order model, AUC = 0.76 for second order model). For models of order from 3 to 5, prediction performances get worse. For upPI-RADS 4 classification task, first order model showed higher performances (AUC = 89%, sensitivity = 87%, specificity = 62%, accuracy = 82%) respect to higher order models. However, promising results were also obtained from third order model performance metrics.

Discussions

The aim of our study was to investigate whether an mpMRI-based radiomic approach was able to outperform PI-RADS v2.1 performance in stratifying PI-RADS 3 and PI-RADS 3 upPI-RADS4 lesions, which are directly related to still opening challenges and controversies for PI-RADS score assignment. In particular, concerning PI-RADS3 lesions, indicating an equivocal likelihood of significant PCa, the major clinical challenge is to avoid excessive biopsies and simultaneously improve PCa detection^{10,14}. On the other hand, the upgrade of upPI-RADS 4 lesions depends on DCE-MRI, its role in prostate mpMRI remains debated for PCa detection^{45,46}. On the basis of promising results offered by radiomic approach for PCa detection, characterization and assessment in comparison with PI-RADS assignment, we investigated the power of radiomics for PI-RADS 3 and upPI-RADS 4 stratification, which, despite its relevancy for the previous mentioned reasons, still represents a little-explored field. Therefore, using biopsy as reference standard, we tested two classification tasks aiming at detecting PCa (namely lesions with GS ≥ 6) in PI-RADS 3 and upPI-RADS 4 scored lesions, respectively. Prediction models performances varied depending on the classification task and the model order. The obtained results for PCa positive PI-RADS 3 detection showed that the most relevant features for this classification task were texture features arising from T2 and ADC. However, the best multivariable prediction model was those built using T2 glrlm SRLGLE and T2 glrlm LGLRE texture features. The obtained results for PCa positive upPI-RADS 4 detection showed that the most relevant features for this classification task were texture features arising from T2 and ADC and interquartile range of ADC value. However, the best multivariable prediction model was that

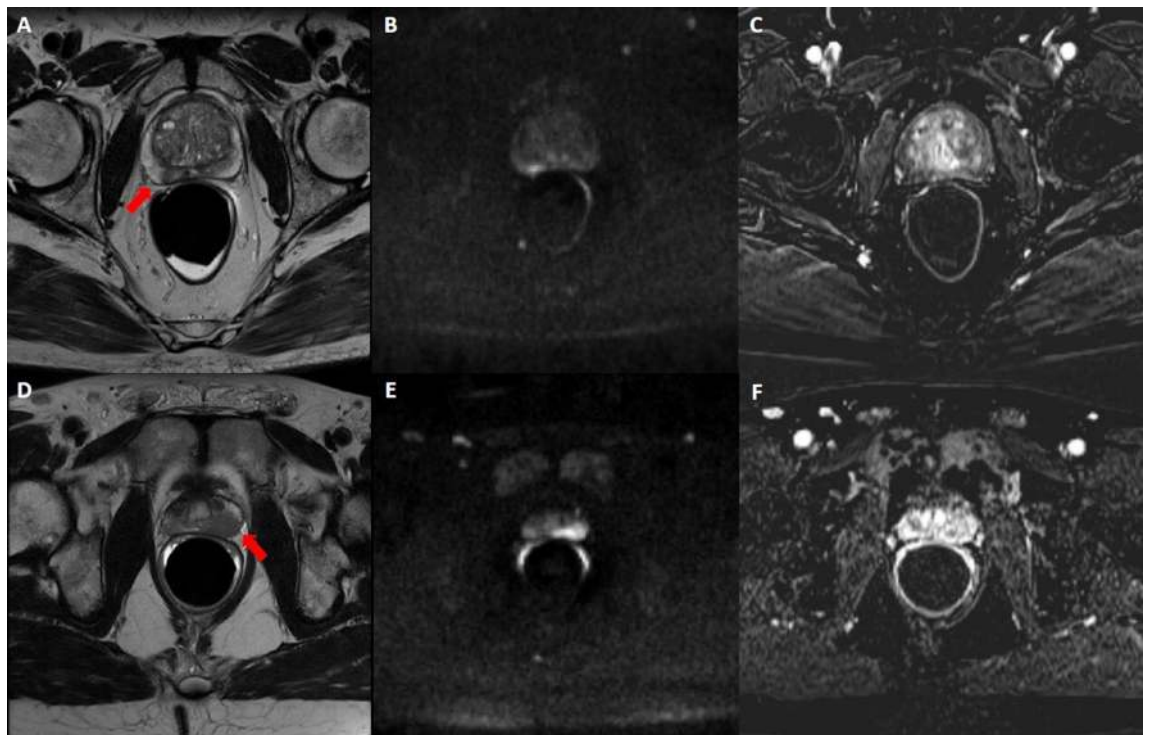


Figure 3. upPI-RADS 4 scored alterations. In the first row, an example of upPI-RADS 4 lesion confirmed at biopsy as GS 4+4: mpMRI showed reduced T2 signal intensity in the right PZ (A), altered diffusivity at b1000 DWI (B) with focal contrast enhancement (C). In the second row, an example of upPI-RADS 4 lesion classified as inflammatory at biopsy: mpMRI showed reduced T2 signal intensity in the right PZ (D), altered diffusivity at b1000 DWI (E) with contrast enhancement (F). PI-RADS prostate imaging-reporting and data system, mpMRI multiparametric MRI, DWI diffusion weighted MRI, PZ peripheral zone.

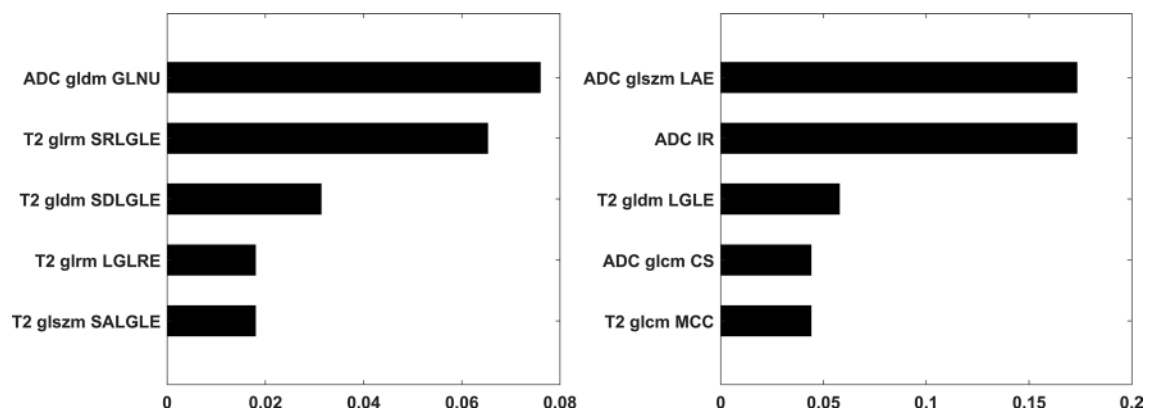


Figure 4. Bar plots of radiomic features ranked according to their relevance-redundancy predictor importance score for PI-RADS 3 (left side) upPI-RADS4 (right side) lesion detection tasks. ADC apparent diffusion coefficient, glcm gray level co-occurrence matrix, glrm grey level run length matrix, gldm gray level dependence matrix, glszm grey level size zone matrix, SDLGLE Small Dependence Low Gray Level Emphasis, SALGLE Small Area Low Gray Level Emphasis; SRLGLE Short Run Low Gray Level Emphasis; LAE Large Area Emphasis; IR Interquartile Range; LGLE Low Gray Level Emphasis; CS Cluster Shade; MCC Maximal Correlation Coefficient.

built using ADC IR. Promising results were also obtained using third order model built using ADC IR together with glszm LAE and T2 glcm MCC texture features. Interestingly, for both classification tasks, specificity was found to be lower than the other performance parameters across each model order, highlighting that our models may be prone to false positive errors. Previous studies aimed at stratifying PI-RADS 3 lesions^{9,11,13-15}, but only one investigated the power of radiomic features for this purpose²³. Our results on PI-RADS 3 lesions are in line with those obtained in this work, in which authors also reported that T2 and ADC texture features could help in stratification of PI-RADS 3 lesions. However, several differences respect to our work should be highlighted. First, in the above-mentioned work, authors just analyzed texture features arising from T2 and ADC maps

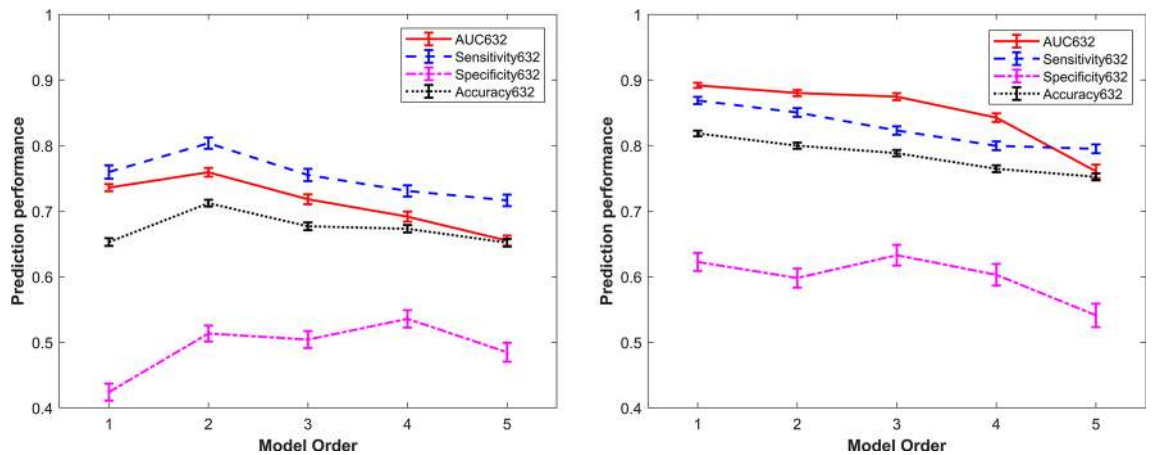


Figure 5. Prediction performances of models from order 1 to 5 for detection of PCa in PI-RADS 3 lesions (on the left) and in upPI-RADS 4 lesions (on the right).

images, omitting analysis of shape and first order features. In our work we also analyzed these features, even if none of them were useful for building predictive models. Second, we also evaluated features associated with DCE-MRI. However, none of them were useful for building predictive models. These results suggest that the inclusion of radiomic features derived from DCE-MRI does not provide a clear added value for PCa detection of PI-RADS 3 lesions. On one hand, this would justify the choice to exclude DCE-MRI from dominant sequences for PI-RADSv2 and PI-RADS v2.1 score assessment^{17,24,45}. Similar considerations also apply to upPI-RADS 4 classification task, for which DCE-MRI features did not even survive in the univariate analysis preceding the mRMR step. However, a direct comparison with studies applying radiomic analysis restricted to this kind of lesions could not be applied, since literature is lacking in this field to our knowledge. Ullrich et al.⁴⁶ discussed about the added value of DCE-MRI in a large cohort of patients with assigned PI-RADS 4 in a recent study, finding DCE-MRI useful to avoid underestimation and misclassification of clinically significant PCa and improve detection rates in PI-RADS 4 patients.

Strengths of our study were that we use PI-RADS v2.1 assessment, which should not affect the overall diagnostic accuracy when compared to PI-RADS v2, but should improve inter-reader variability and simplify score assignment^{4,24,45}. Moreover, the use of 3D ROI should reduce inter-reader variability by avoiding the need to select a single-slice corresponding to a portion of a lesion and allow for a more complete description of the lesion given by the increasing of the number of points considered for features computation, which should improve the accuracy of characterization of heterogeneous lesions and reduce sampling errors^{47,48}. Then, as recommended by the IBSI (Image Biomarker Standardization Initiative) guidelines, we normalized raw images to account for the variable intensity ranges of MRI data (acquired in arbitrary units) and improve robustness of radiomics features^{34,49–52}. Lastly, a special attention should be given to the potential clinical impact of our work that lies in the fact that we attempted to overcome the clinical challenge associated with radiologist uncertainties related to PI-RADS 3 lesions, as well as those related to upPI-RADS 4 lesions due to the previously mentioned reasons^{10,45,53}.

Despite the above, this study suffers from some limitations. First, PI-RADS 3 and upPI-RADS 4 datasets were too small and imbalanced to generalize results. A larger and more balanced study group is needed to better conduct a radiomic analysis and build more robust prediction models. In particular, although the IABR strategy we used for model building is a common reliable approach in case of small and imbalanced datasets, a larger sample size would allow to use part of the dataset for the training, and part for testing and validating the performance of the classifier with external datasets^{44,54,55}. Given the small sample size for both PI-RADS 3 and upPI-RADS 4 groups, we did not perform a separate analysis aiming at detecting clinically significant PCa lesions (namely GS ≥ 7). Moreover, in PI-RADS 3 analysis, the low number of lesions prevent us to perform a sub-analysis on the basis of prostatic zone, similar to analysis performed by Ginsburg et al. for PCa detection in histopathological proven lesion⁵⁶. In addition, although features arising from DCE-MRI parameters proved not to be useful for building prediction models. Further analysis on original DCE-MRI images⁵⁷ and/or maps of pharmacokinetic parameters⁵⁸ may be investigated, helping to overcome controversies related to DCE-MRI and clarify its role in PCa management. It could be also interesting to investigate if performances of prediction models for PCa detection in PI-RADS 3 and upPI-RADS 4 lesions could improve adding features arising from advanced diffusion models prediction model, which were found to be promising for detection and characterization of PCa, even if their role is not clearly affirmed due to the lack of a standardized diffusion MRI protocol^{17,59}. Another aspect to highlight is that also delineations of prostatic lesions are prone to inter-observer variability. In fact, it should be considered that we delineated only the prostatic lesion for feature extraction, and it is well-recognized that small volumes might lead to uncertainties in feature extraction^{60–63}.

In conclusion, our preliminary results support the significant role of T2 and ADC radiomic features, mainly second-order texture features, for PCa detection in lesions scored as PI-RADS 3 and upPI-RADS 4.

Received: 24 August 2020; Accepted: 24 December 2020

Published online: 12 January 2021

References

- Rawla, P. Epidemiology of prostate cancer. *World J. Oncol.* **10**, 63–89. <https://doi.org/10.14740/wjon1191> (2019).
- Sun, Y. *et al.* Multiparametric MRI and radiomics in prostate cancer: A review. *Aust. Phys. Eng. Sci. Med.* **42**, 3–25. <https://doi.org/10.1007/s13246-019-00730-z> (2019).
- Hegde, J. V. *et al.* Multiparametric MRI of prostate cancer: An update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. *J. Magn. Reson. Imaging* **37**, 1035–1054. <https://doi.org/10.1002/jmri.23860> (2013).
- Turkbey, B. *et al.* Prostate Imaging Reporting and Data System Version 2.1: 2019 update of prostate imaging reporting and data system Version 2. *Eur. Urol.* **2019**(76), 340–351. <https://doi.org/10.1016/j.eururo.2019.02.033> (2019).
- Zhao, C. *et al.* The efficiency of multiparametric magnetic resonance imaging (mpMRI) using PI-RADS Version 2 in the diagnosis of clinically significant prostate cancer. *Clin. Imaging* **40**, 885–888. <https://doi.org/10.1016/j.clinimag.2016.04.010> (2016).
- Cash, H. *et al.* The detection of significant prostate cancer is correlated with the Prostate Imaging Reporting and Data System (PI-RADS) in MRI/transrectal ultrasound fusion biopsy. *World J. Urol.* **34**, 525–532. <https://doi.org/10.1007/s00345-015-1671-8> (2016).
- Baldissarotto, M. *et al.* Validation of PI-RADS v2 for prostate cancer diagnosis with MRI at 3T using an external phased-array coil: 3T MRI With PI-RADS v.2 in Prostate Cancer. *J. Magn. Reson. Imaging* **44**, 1354–1359. <https://doi.org/10.1002/jmri.25284> (2016).
- Kasel-Seibert, M. *et al.* Assessment of PI-RADS v2 for the detection of prostate cancer. *Eur. J. Radiol.* **85**, 726–731. <https://doi.org/10.1016/j.ejrad.2016.01.011> (2016).
- Awamleh, B. A. H. A. *et al.* Multicenter analysis of clinical and MRI characteristics associated with detecting clinically significant prostate cancer in PI-RADS (v2.0) category 3 lesions. *Urol. Oncol. Semin. Original Investig.* **38**, 637.e9–637.e15. <https://doi.org/10.1016/j.urolonc.2020.03.019> (2020).
- Schoots, I. G. MRI in early prostate cancer detection: How to manage indeterminate or equivocal PI-RADS 3 lesions?. *Transl. Androl. Urol.* **7**, 70–82. <https://doi.org/10.21037/tau.2017.12.31> (2018).
- Hermie, I., Van Besien, J., De Visschere, P., Lumen, N. & Decaestecker, K. Which clinical and radiological characteristics can predict clinically significant prostate cancer in PI-RADS 3 lesions? A retrospective study in a high-volume academic center. *Eur. J. Radiol.* **114**, 92–98. <https://doi.org/10.1016/j.ejrad.2019.02.031> (2019).
- Felker, E. R. *et al.* Risk stratification among men with prostate imaging reporting and data system version 2 category 3 transition zone lesions: Is biopsy always necessary?. *Am. J. Roentgenol.* **209**, 1272–1277. <https://doi.org/10.2214/AJR.17.18008> (2017).
- Görtz, M. *et al.* The value of prostate-specific antigen density for prostate imaging-reporting and data system 3 lesions on multiparametric magnetic resonance imaging: A strategy to avoid unnecessary prostate biopsies. *Eur. Urol. Focus* <https://doi.org/10.1016/j.euf.2019.11.012> (2019).
- Yang, S. *et al.* Combining clinical and MRI data to manage PI-RADS 3 lesions and reduce excessive biopsy. *Transl. Androl. Urol.* **9**, 1252–1261. <https://doi.org/10.21037/tau-19-755> (2020).
- Kan, Y. *et al.* Clinico-radiological characteristic-based machine learning in reducing unnecessary prostate biopsies of PI-RADS 3 lesions with dual validation. *Eur. Radiol.* <https://doi.org/10.1007/s00330-020-06958-8> (2020).
- Smith, C. P. *et al.* Radiomics and radiogenomics of prostate cancer. *Abdom. Radiol.* **44**, 2021–2029. <https://doi.org/10.1007/s00261-018-1660-7> (2019).
- Monti, S. *et al.* Multiparametric MRI for prostate cancer detection: New insights into the combined use of a radiomic approach with advanced acquisition protocol. *Cancers* **12**, 390. <https://doi.org/10.3390/cancers12020390> (2020).
- Wang, J. *et al.* Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur. Radiol.* **27**, 4082–4090. <https://doi.org/10.1007/s00330-017-4800-5> (2017).
- Chaddad, A. *et al.* Predicting gleason score of prostate cancer patients using radiomic analysis. *Front. Oncol.* **8**, 630. <https://doi.org/10.3389/fonc.2018.00630> (2018).
- Chaddad, A., Kucharczyk, M. & Niazi, T. Multimodal radiomic features for the predicting gleason score of prostate cancer. *Cancers* **10**, 249. <https://doi.org/10.3390/cancers10080249> (2018).
- Gong, L. *et al.* Noninvasive prediction of high-grade prostate cancer via biparametric MRI radiomics: Noninvasive prediction of high-grade PCA via biparametric MRI radiomics. *J. Magn. Reson. Imaging* <https://doi.org/10.1002/jmri.27132> (2020).
- Woźnicki, P. *et al.* Multiparametric MRI for prostate cancer characterization: Combined use of radiomics model with PI-RADS and clinical parameters. *Cancers* **12**, 1767. <https://doi.org/10.3390/cancers12071767> (2020).
- Giambelluca, D. *et al.* PI-RADS 3 lesions: Role of prostate MRI texture analysis in the identification of prostate cancer. *Curr. Problems Diagnostic Radiol.* <https://doi.org/10.1067/j.cpradiol.2019.10.009> (2019).
- Brancato, V. *et al.* Assessment of DCE utility for PCA diagnosis using PI-RADS v2.1: Effects on diagnostic accuracy and reproducibility. *Diagnostics* <https://doi.org/10.3390/diagnostics10030164> (2020).
- Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
- Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* **70**, 213–220. <https://doi.org/10.1037/h0026256> (1968).
- Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46. <https://doi.org/10.1177/001316446002000104> (1960).
- Gordetsky, J. & Epstein, J. Grading of prostatic adenocarcinoma: Current state and prognostic implications. *Diagn. Pathol.* **11**, 25. <https://doi.org/10.1186/s13000-016-0478-2> (2016).
- Barakzai, M. A. Prostatic adenocarcinoma: A grading from the old to the new grade-group system: A historical and critical review. *Asian Pac. J. Cancer Prev.* **20**, 661–666. <https://doi.org/10.31557/APJCP.2019.20.3.661> (2019).
- Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E. & Ulbright, T. M. The 2016 WHO classification of tumours of the urinary system and male genital organs—Part A: Renal, penile, and testicular tumours. *Eur. Urol.* **70**, 93–105. <https://doi.org/10.1016/j.eururo.2016.02.029> (2016).
- Scalco, E. *et al.* T2w-MRI signal normalization affects radiomics features reproducibility. *Med. Phys.* **47**, 1680–1691. <https://doi.org/10.1002/mp.14038> (2020).
- van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339> (2017).
- Collewet, G., Strzelecki, M. & Mariette, F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn. Reson. Imaging* **22**, 81–91. <https://doi.org/10.1016/j.mri.2003.09.001> (2004).
- Duron, L. *et al.* Gray-level discretization impacts reproducible MRI radiomics texture features. *PLoS ONE* **14**, e0213459. <https://doi.org/10.1371/journal.pone.0213459> (2019).
- Fedorov, A., Vangel, M. G., Tempny, C. M. & Fennessy, F. M. Multiparametric magnetic resonance imaging of the prostate: Repeatability of volume and apparent diffusion coefficient quantification. *Invest. Radiol.* **52**, 538–546. <https://doi.org/10.1097/RLI.0000000000000382> (2017).

36. Yu, J.-S. *et al.* Prostate cancer: Added value of subtraction dynamic imaging in 3T magnetic resonance imaging with a phased-array body coil. *Yonsei Med. J.* **49**, 765–774. <https://doi.org/10.3349/ymj.2008.49.5.765> (2008).
37. Haga, A. *et al.* Standardization of imaging features for radiomics analysis. *J. Med. Invest.* **66**, 35–37. <https://doi.org/10.2152/jmi.66.35> (2019).
38. Salkind, N. J. & Rasmussen, K. (eds) *Encyclopedia of Measurement and Statistics* (SAGE Publications, Thousand Oaks, 2007) (ISBN 978-1-4129-1611-0).
39. Ding, C., Peng, H. Minimum redundancy feature selection from microarray gene expression data. in *Proceedings of the Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003* 523–528 (IEEE Comput. Soc: Stanford, CA, USA, 2003).
40. Peng, H., Long, F. & Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1226–1238. <https://doi.org/10.1109/TPAMI.2005.159> (2005).
41. Vallières, M., Freeman, C. R., Skamene, S. R. & El Naqa, I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys. Med. Biol.* **60**, 5471–5496. <https://doi.org/10.1088/0031-9155/60/14/5471> (2015).
42. Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1–26. <https://doi.org/10.1214/aos/1176344552> (1979).
43. Efron, B. & Tibshirani, R. Improvements on cross-validation: The 0.632+ bootstrap method. *J. Am. Stat. Assoc.* **92**, 548. <https://doi.org/10.2307/2965703> (1997).
44. Sahiner, B., Chan, H.-P. & Hadjiiski, L. Classifier performance prediction for computer-aided diagnosis using a limited dataset: Classifier performance prediction using a limited data set. *Med. Phys.* **35**, 1559–1570. <https://doi.org/10.1118/1.2868757> (2008).
45. Ullrich, T. & Schimmöller, L. Perspective: A critical assessment of PI-RADS 2.1. *Abdom. Radiol.* <https://doi.org/10.1007/s00261-020-02424-7> (2020).
46. Ullrich, T. *et al.* Value of dynamic contrast-enhanced (DCE) MR imaging in peripheral lesions in PI-RADS-4 patients. *Fortschr. Röntgenstr.* **192**, 441–447. <https://doi.org/10.1055/a-1020-4026> (2020).
47. Parmar, C. *et al.* Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS ONE* **9**, e102107. <https://doi.org/10.1371/journal.pone.0102107> (2014).
48. Rosenkrantz, A. B. *et al.* Whole-lesion apparent diffusion coefficient metrics as a marker of percentage Gleason 4 component within Gleason 7 prostate cancer at radical prostatectomy: ADC Entropy in Gleason 7 Prostate Cancer. *J. Magn. Reson. Imaging* **41**, 708–714. <https://doi.org/10.1002/jmri.24598> (2015).
49. Zwanenburg, A., Leger, S., Vallières, M., Löck, S. Initiative, for the I.B.S. Image biomarker standardisation initiative. *arXiv:1612.07003 [cs]* (2016).
50. Rizzo, S. *et al.* Radiomics: The facts and the challenges of image analysis. *Eur. Radiol. Exp.* **2**, 36. <https://doi.org/10.1186/s41747-018-0068-z> (2018).
51. Cattell, R., Chen, S. & Huang, C. Robustness of radiomic features in magnetic resonance imaging: Review and a phantom study. *Vis. Comput. Ind. Biomed. Art* **2**, 19. <https://doi.org/10.1186/s42492-019-0025-6> (2019).
52. Schwier, M. *et al.* Repeatability of multiparametric prostate MRI radiomics features. *Sci. Rep.* **9**, 9441. <https://doi.org/10.1038/s41598-019-45766-z> (2019).
53. Scialpi, M. *et al.* Round table: Arguments in supporting abbreviated or biparametric MRI of the prostate protocol. *Abdom. Radiol.* <https://doi.org/10.1007/s00261-020-02510-w> (2020).
54. Forghani, R. *et al.* Radiomics and artificial intelligence for biomarker and prediction model development in oncology. *Comput. Struct. Biotechnol. J.* **17**, 995–1008. <https://doi.org/10.1016/j.csbj.2019.07.001> (2019).
55. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **14**, e0224365. <https://doi.org/10.1371/journal.pone.0224365> (2019).
56. Ginsburg, S. B. *et al.* Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study: Radiomic Features for Prostate Cancer Detection on MRI. *J. Magn. Reson. Imaging* **46**, 184–193. <https://doi.org/10.1002/jmri.25562> (2017).
57. Liu, B. *et al.* Prediction of prostate cancer aggressiveness with a combination of radiomics and machine learning-based analysis of dynamic contrast-enhanced MRI. *Clin. Radiol.* **74**(896), e1-896.e8. <https://doi.org/10.1016/j.crad.2019.07.011> (2019).
58. Franiel, T., Hamm, B. & Hricak, H. Dynamic contrast-enhanced magnetic resonance imaging and pharmacokinetic models in prostate cancer. *Eur. Radiol.* **21**, 616–626. <https://doi.org/10.1007/s00330-010-2037-7> (2011).
59. Brancato, V., Cavaliere, C., Salvatore, M. & Monti, S. Non-Gaussian models of diffusion weighted imaging for detection and characterization of prostate cancer: A systematic review and meta-analysis. *Sci. Rep.* **9**, 16837. <https://doi.org/10.1038/s41598-019-53350-8> (2019).
60. Pavic, M. *et al.* Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol.* **57**, 1070–1074. <https://doi.org/10.1080/0284186X.2018.1445283> (2018).
61. Steenbergen, P. *et al.* Prostate tumor delineation using multiparametric magnetic resonance imaging: Inter-observer variability and pathology validation. *Radiother. Oncol.* **115**, 186–190. <https://doi.org/10.1016/j.radonc.2015.04.012> (2015).
62. Fernandes, C. D. *et al.* Biochemical recurrence prediction after radiotherapy for prostate cancer with T2w magnetic resonance imaging radiomic features. *Phys. Imaging Radiat. Oncol.* **7**, 9–15. <https://doi.org/10.1016/j.phro.2018.06.005> (2018).
63. Zhang, X. *et al.* The effects of volume of interest delineation on MRI-based radiomics analysis: Evaluation with two disease groups. *Cancer Imaging* **19**, 89. <https://doi.org/10.1186/s40644-019-0276-7> (2019).

Acknowledgements

This work was supported by “Ricerca Corrente” Grant from Italian Ministry of Health (IRCCS SDN).

Author contributions

C.C. and V.B. conceived the study; C.C., L.B., A.R., G.D.C., and L.P. enrolled the patients and acquired data; V.B., S.M and L.B. performed the data analysis; V.B. wrote the original draft; C.C., M.A. and A.R. reviewed and edited the manuscript. M.S. and C.C. supervised the project. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-80749-5>.

Correspondence and requests for materials should be addressed to V.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021