

RESEARCH

Open Access



Comparison in prostate cancer diagnosis with PSA 4–10 ng/mL: radiomics-based model VS. PI-RADS v2.1

Chunxing Li^{1,2†}, Zhicheng Jin^{3†}, Chaogang Wei², Guangcheng Dai⁴, Jian Tu⁵ and Junkang Shen^{2*}

Abstract

Background To evaluate accuracy of MRI-based radiomics in diagnosing prostate cancer (PCa) in patients with PSA levels between 4 and 10 ng/mL and compare it with the latest Prostate Imaging Reporting and Data System (PI-RADS v2.1) score.

Methods 221 patients with prostate lesions and PSA levels in 4–10 ng/mL, including 154 and 67 cases in the training and validation groups. Pathological confirmation of all patients was accomplished by the use of MRI-TRUS fusion targeted biopsy or systematic transrectal ultrasound (TRUS) guided biopsy. 851 radiomic features were extracted from each lesion of ADC and T2WI images. The least absolute shrinkage and selection operator (LASSO) regression algorithm and logistic regression were employed to select features and build the ADC and T2WI model. The combined model was obtained based on the ADC and T2WI features. The clinical benefit and diagnostic accuracy of the three radiomics models and PI-RADS v2.1 score were evaluated.

Results 10 radiomic features were ultimately selected from the ADC images, 13 from the T2WI images and 7 from the combined models. The ADC, T2WI and combined models achieved satisfactory diagnostic accuracy in the training [AUC:0.945 (ADC), 0.939 (T2WI), 0.979 (combined)] and validation groups [AUC: 0.942 (ADC), 0.943 (T2WI), 0.959 (combined)], which was significantly higher than those in PI-RADS v2.1 model (0.825 for training cohort and 0.853 for validation cohort). Compared with the PI-RADS v2.1 score, the three radiomics models generated superior PCa diagnostic performance in both the training ($p=0.002$, $p=0.005$, $p<0.001$) and validation groups ($p=0.045$, $p=0.035$, $p=0.015$).

Conclusion Radiomics based on ADC and T2WI images can better identify PCa in patients with PSA 4–10 ng/mL, and MRI-based radiomics significantly outperforms the PI-RADS v2.1 score.

Clinical trial number Not applicable.

Keywords Prostate Cancer, PSA 4–10 ng/mL, Radiomics, PI-RADS v2.1, MRI

[†]Chunxing Li and Zhicheng Jin contributed equally to the work.

*Correspondence:

Junkang Shen
shenjunkang@suda.edu.cn

¹Department of MRI Room, The First People's Hospital of Yancheng, Yancheng First Hospital Affiliated Hospital of Nanjing University Medical School, Yancheng, China

²Department of Radiology, The Second Affiliated Hospital of Soochow University, Suzhou 215004, China

³Department of Nuclear Medicine, The Second Affiliated Hospital of Soochow University, Suzhou, China

⁴Department of Urology Surgery, The Second Affiliated Hospital of Soochow University, Suzhou, China

⁵Department of Pathology, The Second Affiliated Hospital of Soochow University, Suzhou, China



Introduction

Prostate cancer (PCa) is a common and significant malignancy of the genitourinary system that primarily affects older males [1]. Although serum prostate specific antigen (PSA) is widely employed as a biomarker for prostate cancer (PCa) screening, there is significant ambiguity regarding PSA levels ranging from 4 to 10 ng/mL, usually referred to as the “gray zone”. Less than 25% of men with a prostate-specific antigen (PSA) level in the gray zone were found to have prostate cancer (PCa), and nearly 70% of men with gray-zone PSA levels underwent biopsy, yielding no significant cancer [2, 3]. Enhancing diagnostic precision and minimizing unwarranted biopsy procedures would be both suitable and valuable for patients exhibiting prostate-specific antigen (PSA) values ranging from 4 to 10 ng/mL. Magnetic resonance imaging (MRI) is a non-invasive imaging tool that has had beneficial impacts on pre-operative diagnosis, clinical staging, therapy evaluation, and post-operative monitoring for prostate cancer (PCa). The use of the Prostate Imaging and Data System version 2.1 (PI-RADS v2.1) score, which is based on a multiparametric MRI (mp-MRI) protocol including T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced imaging (DCEI), has enhanced the overall accuracy for detecting prostate cancer (PCa). However, this scoring system has shown vulnerability to variations in observer experience, leading to inadequate agreement and reproducibility among different observers [4–6].

Recently, radiomics has been able to efficiently extract a significant number of imaging features from medical images, allowing for the conversion of these images into quantitative data that can be analyzed objectively. This data can provide insights into the underlying pathophysiological characteristics, particularly tumor heterogeneity [7–9]. By conducting quantitative analysis of these characteristics, it is possible to develop radiomics-based models that are associated with clinical and biological traits. These models can improve the accuracy of tumor diagnosis, identification of molecular subtypes, evaluation of treatment effectiveness, and prediction of survival. Consequently, they can provide valuable information for clinical decision-making. Several studies have indicated that the radiomics model has a beneficial impact on the diagnosis and assessment of aggressiveness in PCa [10–14]. However, previous research has not extensively examined the diagnostic value comparison between an objective radiomics-based technique and the subjective PI-RADS v2.1 score in prostate patients with prostate-specific antigen (PSA) levels in the gray zone [15, 16].

So, the purpose of this study was to investigate the impact of radiomic-based models for PCa diagnosis in the PSA level range of 4–10 ng/mL and to evaluate the

diagnostic performance by comparing it with the PI-RADS v2.1 score.

Materials and methods

Study population

This retrospective study, conducted at a single center, received approval from our institutional review board (ethics approval number: JD-HG-2024-075) and was exempt from the necessity for informed consent. It consisted of 740 consecutive patients who received prostate mp-MRI examination and were subsequently confirmed by pathology results. The data was collected from January 2021 to December 2022. Before undergoing prostate MRI imaging, all patients included in the study had elevated levels of total PSA (t-PSA) ranging from 4 to 10 ng/mL. Additionally, their levels of free PSA (f-PSA) were also assessed. However, a total of 519 patients were excluded based on the following criteria: (1) total PSA levels less than 4 ng/mL or greater than 10 ng/mL ($n=485$); (2) lesions with Pathology-MRI mismatch (pathological biopsy of lesions cannot be seen on MRI) ($n=10$); (3) too small lesion volume (maximum diameter less than 5 mm) or lesion boundary cannot be delineated ($n=6$); (4) biopsy or other therapies performed before MRI examination ($n=6$); (5) unsatisfactory MR image quality or inconsistent MRI protocol ($n=5$); (6) incomplete clinical information ($n=7$). Finally, the study population consisted of 221 patients including 60 PCa patients and 161 non-cancerous patients without any histological evidence of cancer. In these 60 patients with PCa, 16 had lesions in the transition zone (TZ) and the remaining 44 had lesions in the peripheral zone (PZ). These enrolled patients were randomly assigned to the training cohort ($n=154$) and validation cohort ($n=67$) at a ratio of 7:3. The flowchart with exclusion criteria is shown in Fig. 1.

MRI protocol and PI-RADS v2.1 assessment

The prostate mp-MRI examination was conducted using a 3.0 Tesla MRI scanner (Philips Ingenia, The Netherlands) equipped with a 32-channel body phased array coil. The scan sequences and acquisition settings, as per the PI-RADS v2.1 protocol, are presented in Table 1. The sequences mostly consisted of T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced imaging (DCEI). An axial diffusion-weighted imaging (DWI) series was performed with several b values ($b=0, 100, 1000, 2000 \text{ s/mm}^2$) to generate apparent diffusion coefficient (ADC) maps. These maps were automatically reconstructed for the purpose of visual assessment and analysis.

Two radiologists, Z.Y. and W.C., with 7 and 10 years of expertise in prostate MRI diagnosis, independently examined each prostate mp-MRI image. They were unaware of the pathological results and clinical information. The

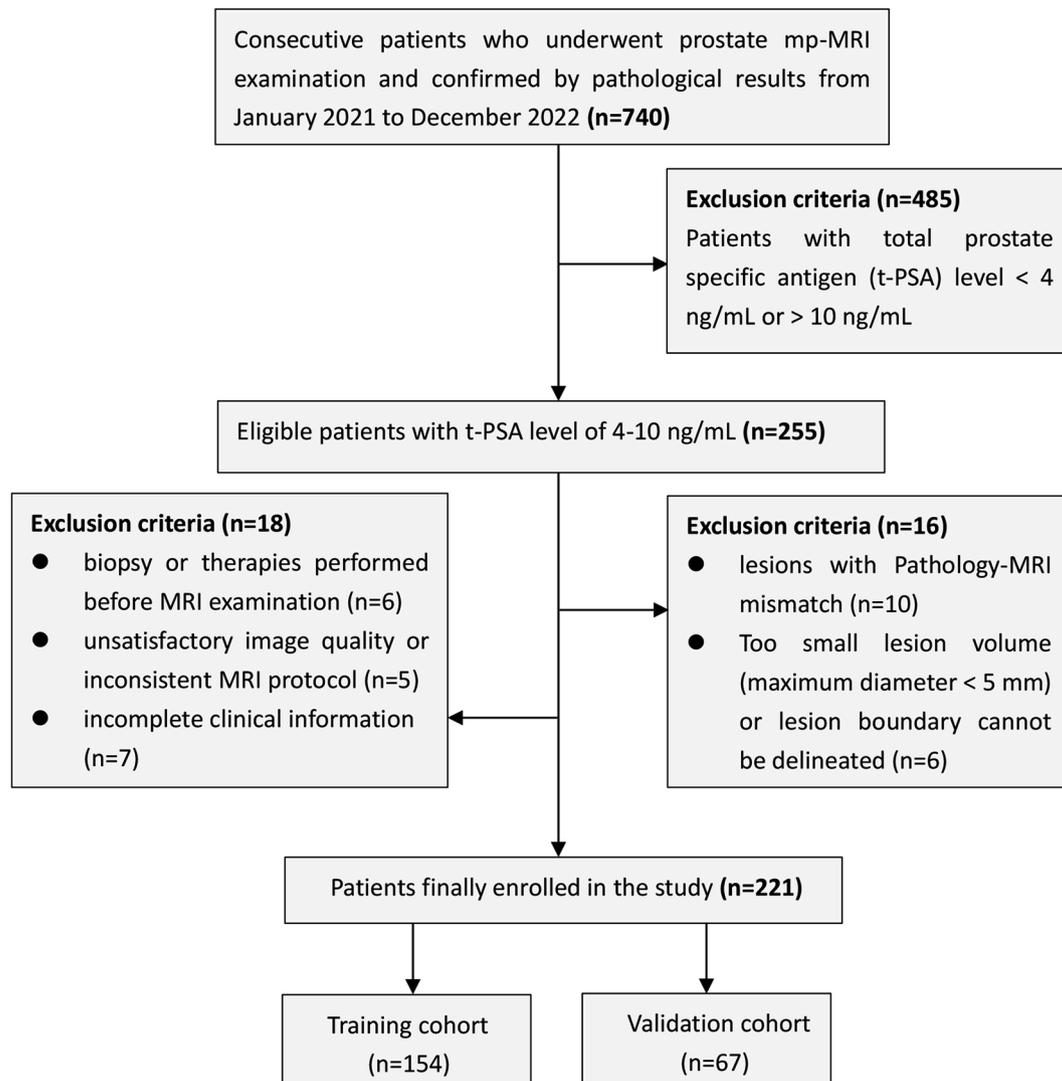


Fig. 1 Flowchart of the study population with exclusion criteria

Table 1 Prostate multi-parametric MRI acquisition protocol

Parameters	T2WI axial	T2WI sagittal	DWI	DCEI
TR (ms)	3000	4765	6000	3.2
TE (ms)	100	100	77	1.5
FOV (mm)	220×220	240×180	260×260	220×220
Slice thickness (mm)	3	4	3	3
Slice gap (mm)	0	1	0	0
Matrix	276×238	240×161	104×126	124×121
b values (s/mm ²)	-	-	0,100,1000,2000	-

Note: TR, repetition time; TE, echo time; FOV, field of view; T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging; DCEI, dynamic contrast-enhanced imaging

two radiologists reached a consensus, resolving any disagreement. A third senior radiologist specializing in the genitourinary system (S.J.) was consulted to resolve any ongoing disagreements in the PI-RADS v2.1 score and

make a final conclusion. According to PI-RADS v2.1 criteria [4], T2WI sequence plays a dominant role in the evaluation of transitional zone (TZ) lesions, while DWI performs only as a secondary role. The PI-RADS assessment category in peripheral zone (PZ) primarily depends on the DWI sequence, while DCEI sequence exerts minor effects.

Pathological analysis

All of the recipients had a 10-core systematic transrectal ultrasonography (TRUS)-guided biopsy following an MRI scan. Furthermore, MRI-TRUS fusion targeted biopsy was employed to examine worrisome PCa lesions identified on MRI with a PI-RADS v2.1 score of 3 or higher. For these lesions, an additional 2–3 targeted cores would be included. The procedure of MRI-TRUS fusion targeted biopsy was carried out utilizing the Mylab Twice Color Doppler ultrasound device, which is equipped with

the Real-time Virtual Sonography (RVS) imaging fusion system manufactured by Esaote SpA in Genova, Italy. The prostate biopsy technique was conducted by highly experienced senior urologists with more than 5 years of expertise. Experienced pathologists in our institution evaluated the specimens using the 2014 International Society of Urological Pathology (ISUP) Gleason grading system [17].

Region of interest segmentation

Due to the limited role of DCEI, we only investigated T2WI and ADC pictures for our study. The process of segmenting the region of interest (ROI) involved manually delineating each slice on both T2WI and ADC pictures. This task was performed by two radiologists (Z.R. and C.T.) using a double-blind procedure, meaning they were not provided with any clinical information or pathology data. The radiomics features were extracted using the open-source 3D Slicer software (version 5.0.3), which has been validated in previous studies for radiomics research. The detailed approach is illustrated in Fig. 2. The two radiologists meticulously delineate the boundaries of all regions of interest (ROIs) according to identical criteria.

To account for tumor heterogeneity, it is crucial to include regions of bleeding, calcification, necrosis, and cystic tissue in the specified ROIs. However, it is important to exclude normal anatomical features such as the urethra, ejaculatory duct, and seminal vesicles. Of the 221 patients enrolled, there were 60 PCa patients in this study, of whom had 50 patients with one PCa lesion, 9 patients with two PCa lesions and only 1 patient with three PCa lesions. For each patient, only the dominant lesion was ultimately selected for the delineation of ROI in our study. For PCa lesions, the dominant lesion was determined as the lesion with the highest Gleason score (GS) as validated by pathological findings. If the GSs were

the same, the dominant lesion was chosen as the one with the largest lesion volume. For non-cancerous lesions, we only delineated the dominant lesion with the largest volume [18]. Finally, we identified a total of 221 ROIs, consisting of 60 for PCa and 161 for non-cancerous tissue.

Radiomics feature extraction

The radiomics features were extracted from the regions of interest (ROIs) on the T2WI and ADC images using the 3D Slicer software, which is an open-source resource. All radiomics features were in accordance with the standards set by the Image Biomarker Standardization Initiative (IBSI). Before outlining the region of interest, we performed image registration and pre-processing of the MRI images. This included pre-processing the image with resampling normalisation and image intensity inhomogeneity (bias) correction, which was performed in the 3D-slicer software. The extraction of features from each MR image were performed based on the “Pyradiomics” package (version 3.9.1) of 3D-slicer. Then, a total of 851 features were extracted from ADC and T2WI pictures, consisting of 14 shape characteristics, 18 first-order features, 75 original texture features, and 744 wavelet features. To ensure the consistency of ROIs outlined by the two radiologists and to maintain the stability and reproducibility of the features, a total of 30 lesions from the entire dataset were randomly chosen for secondary outlining. The texture features extracted from these lesions were then tested for reliability. Features with an inter-observer correlation coefficient (ICC) greater than 0.75 were considered reliable and reproducible. These features were subsequently used for feature selection and model construction.

Radiomics feature selection and model construction

Prior to selection, all features were standardized using the Z-score approach, which enhances the

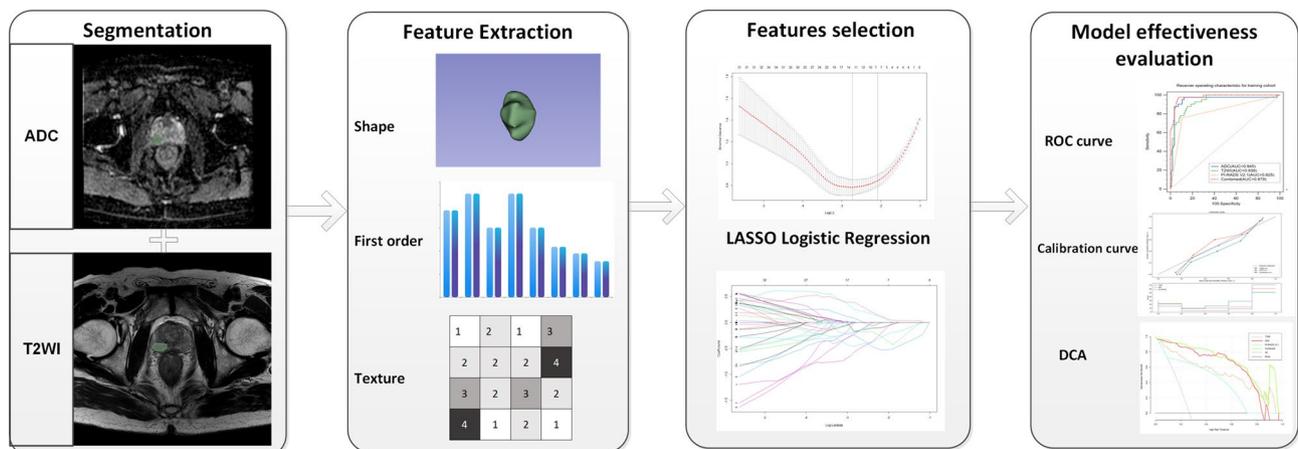


Fig. 2 Flowchart of our study

stability of the data by subtracting the mean and dividing by the standard deviation. Following the reliability test, the radiomics characteristics of the training group underwent LASSO regression for additional data selection. The LASSO regression approach utilizes a continually applied penalty term λ value to perform a continuous penalized screening of all variables. This procedure ensures that all characteristic variables continuously converge to zero. The characteristics were chosen using the ten-fold cross-validation procedure based on the least binomial deviance. The final selection of non-zero features was made using the multivariate logistic regression method to construct a classification and diagnostic model. In this study, we have initially developed two separate models based on the selected characteristics as the ADC model and the T2WI model (features were derived from ADC and T2WI images, respectively). In order to demonstrate the predictive efficacy of the radiomics model in categorizing and diagnosing PCa, we created a combined model by combining ADC with T2WI characteristics, the screening process is the same as for ADC with T2WI. The selection of texture features, creation of models, and evaluation of diagnostic efficacy for the radiomics models were conducted using R software (version 4.1.1) and Python software (version 3.9.1).

Radiomics-based models comparison

The diagnostic effectiveness of all models was assessed based on the area under the curve (AUC) of the receiver operating characteristic (ROC). The models' overall accuracy was determined using confusion matrices, as well as sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each model. The DeLong test was employed to examine the diagnostic effectiveness of the models, with a significance level of $P < 0.05$ indicating statistical significance. The calibration curve and Brier score were employed to assess the calibration performance and accuracy of the classification prediction model. A lower Brier score indicates a more optimal model fit and a closer approximation to the ideal model, resulting in improved predictive performance. Furthermore, decision curve analysis (DCA) was employed to assess the clinical efficacy of the categorical prediction model.

Statistical analysis

The statistical analyses were conducted using R software version 4.1.1, SPSS version 22.0, and MedCalc version 15.2.2. The normality of the distribution was assessed using either the Kolmogorov-Smirnov test or the Shapiro-Wilk test. The analysis involved continuous variables that followed a normal distribution. An independent samples t-test was used to compare the means, which were reported as the mean \pm standard deviation.

Non-parametric tests were used to analyze categorical variables (frequencies and percentages) or continuous variables with a non-normal distribution, represented as median (lower quartile, upper quartile). The reliability of two radiologists with differing levels of expertise was assessed using the inter-observer correlation coefficient (ICC). The ICC value, which runs from 0 to 1, is considered to have good reliability when it is more than 0.75 [19]. The inter-observer agreement for the PI-RADS v2.1 score was evaluated by the Kappa coefficients. The Kappa value ranges from 0 to 1, and was rated 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. The diagnostic performance of the radiomics-based model and PI-RADS v2.1 score was assessed by comparing the area under the curve (AUC). The diagnostic sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated using the optimal cutoff value, along with their corresponding 95% confidence intervals (CI). A two-tailed p value below 0.05 signifies statistical significance.

Results

Patient characteristics

Baseline characteristics of 154 patients (41 patients with PCa and 113 patients with non-cancerous lesions) in the training cohort and 67 patients (19 PCa patients and 48 non-cancerous patients) in the validation cohort enrolled in this study were shown in Table 2. In the training group, among the 41 patients with PCa, 11 (26.8%) had low-risk PCa (ISUP grade group 1) and 30 (73.2%) had clinically-significant PCa (ISUP grade group ≥ 2). Patients' clinical information, including age, t-PSA, f-PSA, and free/total-PSA ratio (f/t-PSA) levels were also collected. Age was significantly higher in patients with PCa than patients with non-cancerous lesions, while f/t-PSA level was significantly lower in PCa patients in the training cohort ($p < 0.05$). There was no significant difference among all parameters in the validation cohort (all $P > 0.05$).

Radiomics models building and validation

After the consistency test and excluding features with ICC coefficients less than 0.75, 735 and 698 features were extracted from the T2WI images and the ADC images, respectively. Then, based on the LASSO regression algorithm, 13 features were finally extracted from the T2WI images. After the same procedure, 10 and 7 features were finally extracted from the ADC images and combined model. All extracted features and their coefficients were detailed in Supplementary Table 1, the process was illustrated in detail in Fig. 3. The correlation heatmap of the final screening features of the ADC, T2WI and combined models were illustrated in detail in Supplementary Fig. 1. The radiomics model developed using the final

Table 2 Patients' baseline information

	Training cohort (n = 154)			Validation cohort (n = 67)		
	PCa (N = 41)	Non-cancerous lesions (N = 113)	P	PCa (N = 19)	Non-cancerous lesions (N = 48)	P
Age (year)	72 ± 7	69 ± 8	0.031*	71 (67, 76)	69 (65, 74)	0.242
t-PSA (ng/mL)	6.59 (6.18, 8.54)	6.58 (5.42, 8.22)	0.086	6.71 ± 1.29	6.24 (4.96, 8.06)	0.666
f-PSA (ng/mL)	0.83 (0.54, 1.04)	1.12 (0.89, 1.49)	0.125	1.19 ± 0.53	1.19 (0.91, 1.65)	0.444
f/t-PSA	0.13 (0.10, 0.20)	0.17 (0.14, 0.23)	0.006*	0.18 ± 0.08	0.20 ± 0.08	0.265
Location (%)			< 0.001*			< 0.001*
TZ	14 (34.1%)	99 (87.6%)		6 (31.6%)	43 (89.6%)	
PZ	27 (65.9%)	14 (12.4%)		13 (68.4%)	5 (10.4%)	
ISUP, n (%)						
grade group 1	11 (26.8%)	N.A.	N.A.	6 (31.6%)	N.A.	N.A.
grade group 2	9 (22.0%)	N.A.	N.A.	2 (10.5%)	N.A.	N.A.
grade group 3	10 (24.4%)	N.A.	N.A.	7 (36.8%)	N.A.	N.A.
grade group 4	9 (22.0%)	N.A.	N.A.	3 (15.8%)	N.A.	N.A.
grade group 5	2 (4.8%)	N.A.	N.A.	1 (5.3%)	N.A.	N.A.

Note: PCa, prostate cancer; PSA, prostate specific antigen; t-PSA, total PSA; f-PSA, free PSA; f/t-PSA, free/total-PSA ratio; TZ, Transitional zone; PZ, Peripheral zone; ISUP, International Society of Urological Pathology. N.A., not applicable. * Statistically significant

extracted features of T2WI images produced an AUC of 0.939 (95% CI: 0.888–0.971) in the training cohort. Similarly, the radiomics model based on ADC images showed a comparable diagnostic performance with an AUC of 0.945 (95% CI: 0.897–0.975). Furthermore, the combined model demonstrated a significantly enhanced diagnostic performance compared to the T2WI and ADC models. The AUC was 0.979 (95% CI: 0.943–0.996). The Delong test results indicated no statistically significant difference between the three models ($p=0.827$ between ADC model and T2WI model, $p=0.069$ between ADC and Combined model, $p=0.054$ between T2WI and combined model). The ADC model, T2WI model, and combined model in the validation cohort had comparable diagnostic performance to the models in the training group, with area under the curve (AUC) values of 0.942 (95% confidence interval [CI]: 0.856–0.984), 0.943 (95% CI: 0.858–0.985), and 0.959 (95% CI: 0.881–0.992), respectively. The Delong test indicated that there was no statistically significant difference seen among the three models. The p-values were 0.969 between the ADC model and T2WI model, 0.253 between the ADC model and combination model, and 0.359 between the T2WI model and combined model (Tables 3 and 4). The categorization diagnostic performance of all these models was reported and demonstrated. Furthermore, Fig. 4 depicts the ROC curves for all the classification models.

Diagnostic performance between the ADC model, T2WI model, combined model and the PI-RADS v2.1 score

The inter-observer agreement for the PI-RADS v2.1 score was measured with the Kappa value of 0.865 (95%CI: 0.745–0.935). This indicates that the results obtained using the PI-RADS v2.1 score were very trustworthy and reproducible. The AUC values of the PI-RADS v2.1

score in the training group and validation group were 0.825 (95% confidence interval: 0.756–0.881) and 0.853 (95% confidence interval: 0.745–0.928), respectively. The optimal threshold of the PI-RADS v2.1 score for distinguishing between PCa and non-cancerous lesions was determined to be 4. The diagnostic accuracy in the training cohort was found to be 0.857. In the training cohort, the ADC model, T2WI model, and combined model demonstrated superior diagnosis accuracy with values of 0.896, 0.877, and 0.935, respectively. Comparable outcomes were achieved in the validation group. More details about the diagnostic performance of all these models was shown in Table 3. Furthermore, the ADC model, T2WI model and combined model showed statistically significant differences compared with PI-RADS v2.1 score both the training ($p=0.002$, $p=0.005$ and $p<0.001$, respectively) and validation cohort ($p=0.046$, $p=0.035$ and $p=0.015$, respectively), as shown in Table 4.

Clinical use and calibration

The calibration curve demonstrated that the three categorization diagnostic models closely resembled the ideal curve, suggesting that these models exhibited strong fitting and predictive capabilities. The Brier score of the combined model was lower compared to the ADC and T2WI models, indicating a higher level of fitness. The Brier scores for the combined model, ADC model, and T2WI model were 0.049, 0.059, and 0.091, respectively. Figure 5 provided a comprehensive depiction of the calibration curve. The ADC model, T2WI model, and combination model demonstrated significant clinical advantages and outperformed the PI-RADS v2.1 score in terms of clinical performance gains. Figure 6 displayed the decision curve analysis. Radiomics and PI-RADS

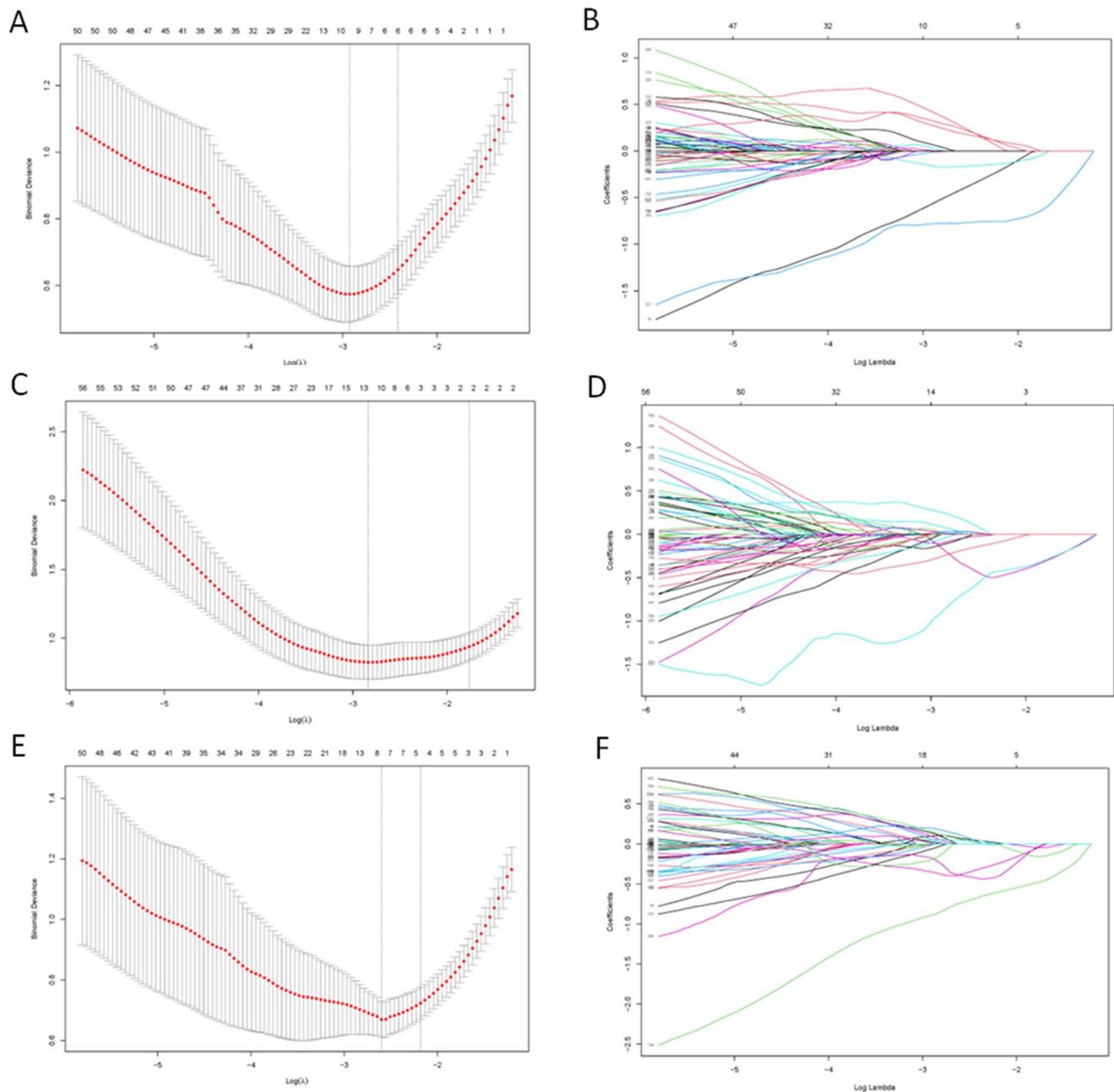


Fig. 3 (A-F) Demonstrated the specific process of least absolute shrinkage and selection operator (LASSO) regression analysis screening features for ADC, T2WI and combined models, respectively. (A, C, E) Showed process of features selection. The vertical line was plotted at the optimal γ of 0.054, 0.042 and 0.050 for ADC, T2WI and combined models, respectively. Ten, thirteen and seven features with non-zero coefficients were finally selected for ADC, T2WI and combined models, respectively. (B, D, F) Showed that features selection performed by 10-fold cross-validation with the criterion of minimum deviance

v2.1 score in the diagnosis of prostate cancer patients are demonstrated in Supplementary Figs. 2 and 3.

Discussion

In this study, we successfully constructed and internally validated radiomics models based on ADC and T2WI images. Compared to the PI-RADS v2.1 score, radiomics models could significantly improve the diagnostic accuracy of PCa in patients with PSA 4-10ng/mL. This

non-invasive method has the potential to play a crucial role in the diagnosis of PCa in patients with PSA levels of 4–10 ng/mL.

As a marker for PCa, serum PSA has been widely used in early diagnosis of PCa. PSA has high sensitivity but low specificity, which makes it elevated not only in PCa but in non-cancerous lesions (such as benign prostate hyperplasia and prostatitis), especially at the range of 4–10 ng/mL. The low specificity of PSA inevitably

Table 3 Diagnostic performance of the PI-RADS v2.1, ADC, T2WI and the combined model

	AUC (95%CI)	SEN	SPE	ACC	PPV	NPV
PI-RADS v2.1						
Training cohort	0.825(0.756–0.881)	0.756	0.894	0.857	0.756	0.894
Validation cohort	0.853(0.745–0.928)	0.789	0.917	0.806	0.737	0.833
ADC						
Training cohort	0.945(0.897–0.975)	0.976	0.876	0.896	0.902	0.894
Validation cohort	0.942(0.856–0.984)	1.000	0.792	0.867	0.737	0.896
T2WI						
Training cohort	0.939(0.888–0.971)	0.878	0.851	0.877	0.756	0.919
Validation cohort	0.943(0.858–0.985)	1.000	0.833	0.881	0.789	0.917
Combined						
Training cohort	0.979(0.943–0.996)	0.976	0.929	0.935	0.902	0.947
Validation cohort	0.959(0.881–0.992)	1.000	0.833	0.911	0.842	0.938

Note: AUC, area under the curve; SEN, sensitivity; SPE, specificity; ACC, accuracy; PPV, positive predictive value; NPV, negative predictive value; ADC, apparent diffusion coefficient; T2WI, T2-weighted imaging; Combined, the combination between T2WI and ADC

Table 4 P values for pairwise comparison of ROC curves

	Combined	PI-RADS v2.1	ADC
Training cohort			
PI-RADS v2.1	< 0.001*		
ADC	0.069	0.002*	
T2WI	0.054	0.005*	0.827
Validation cohort			
PI-RADS v2.1	0.015*		
ADC	0.253	0.046*	
T2WI	0.359	0.035*	0.969

Note: ADC, apparent diffusion coefficient; T2WI, T2-weighted imaging; Combined, the combination between T2WI and ADC; An asterisk (*) indicates a significant ($p < 0.05$) difference

leads to the possibility for overdiagnosis and unnecessary prostate biopsy. Therefore, there is an urgent need to better improve diagnostic accuracy of PCa patients with PSA levels between 4 and 10 ng/mL. PI-RADS category based on multi-parametric MRI protocol has shown great advantage in PCa diagnosis [20]. The first version was published in 2012, updated as the second version (PI-RADS v2) in 2015 and slightly revised to the latest version (PI-RADS v2.1) in 2019 [4, 21]. Compared to PI-RADS v2, it showed improved inter-observer agreement and diagnostic performance for PCa and clinically-significant PCa in PI-RADS v2.1, especially for transitional zone lesions. However, it was still not satisfactory for the diagnosis of PCa in patients with PSA ranges in 4–10 ng/

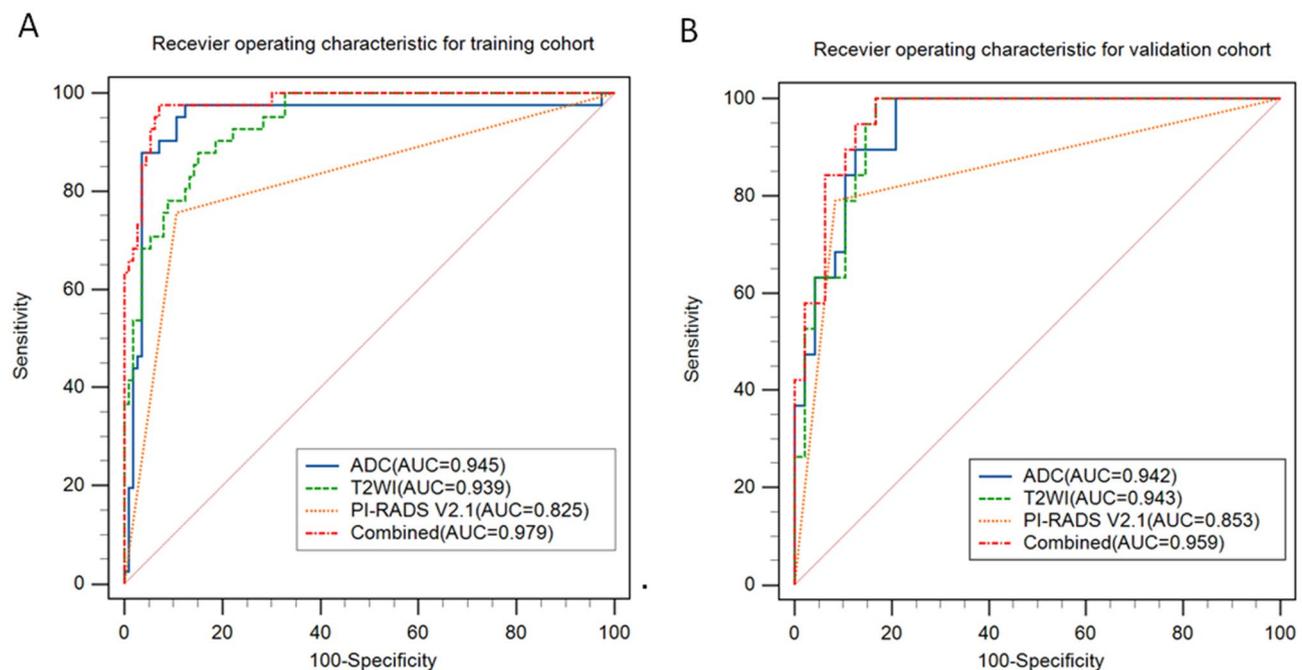


Fig. 4 Comparison of the diagnostic performance of different models. (A, B) Show the ROC curves for each model in detail while also recording the value of AUC

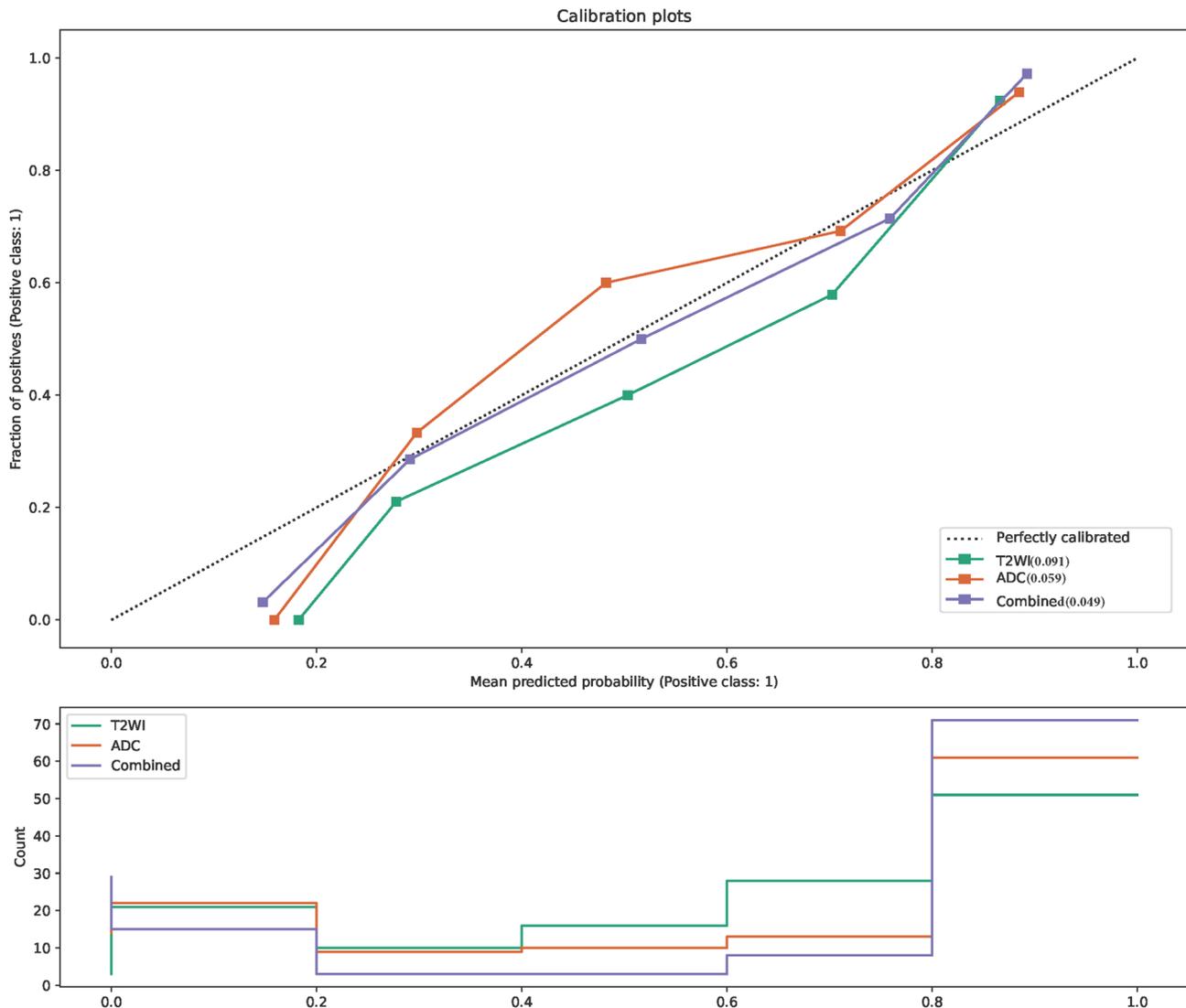


Fig. 5 The comparison of the calibration curve and Brier score across various models. All three models exhibited calibration curves that closely approximated perfect curves. In comparison to the ADC and T2WI models, the Combined model exhibited superior goodness of fit, as seen by the lesser Brier values (0.049, 0.059, and 0.091, respectively). The chart presented below illustrates the distribution of diagnostic probability across several models

mL using the PI-RADS v2.1 score. In our study, we identified a cutoff value of >3 for mpMRI-based PI-RADS v2.1 score and found an AUC of 0.882 in detecting PCa with a specificity of 88.5% in the training cohort, which had better performance than that of PI-RADS v2 in detecting PCa (AUC=0.708, Specificity 58.3%) reported by Qi et al. [22]. Similar results were reported by Han et al. [23]. They also found that the diagnostic AUC of mpMRI-based PI-RADS v2.1 score was 0.867 in detecting csPCa with a PSA level range of 4–10 ng/mL with 80.2% specificity. However, the diagnostic sensitivity of PI-RADS v2.1 in detecting PCa patients with PSA levels of 4–10 ng/mL was only 70.7% and 79.0% in the training and validation cohort. The low sensitivity made it difficult to accurately diagnose prostate lesions in patients with PSA 4–10 ng/mL. In addition, subjective PI-RADS

score was closely associated with different observers' experience [6, 24]. Observers with insufficient experience identified incorrect PI-RADS score, and further affected clinical decision-making, resulting in overdiagnosis, overtreatment and unnecessary biopsy. To reduce unnecessary biopsies, Dwivedi et al. developed a prebiopsy mpMRI-based risk score to predict the likelihood of PCa in patients with PSA 4–10 ng/mL [25].

Compared to PI-RADS score, radiomics enables medical images to be used as data for objective and in-depth analysis. It provides valuable diagnostic, prognostic and predictive information and has been used in PCa diagnosis and evaluation. In most published studies, they compared radiomics-based models to PI-RADS v2 score in the assessment of PCa and csPCa. Chen et al. [14]. revealed the superiority of radiomics to the PI-RADS v2

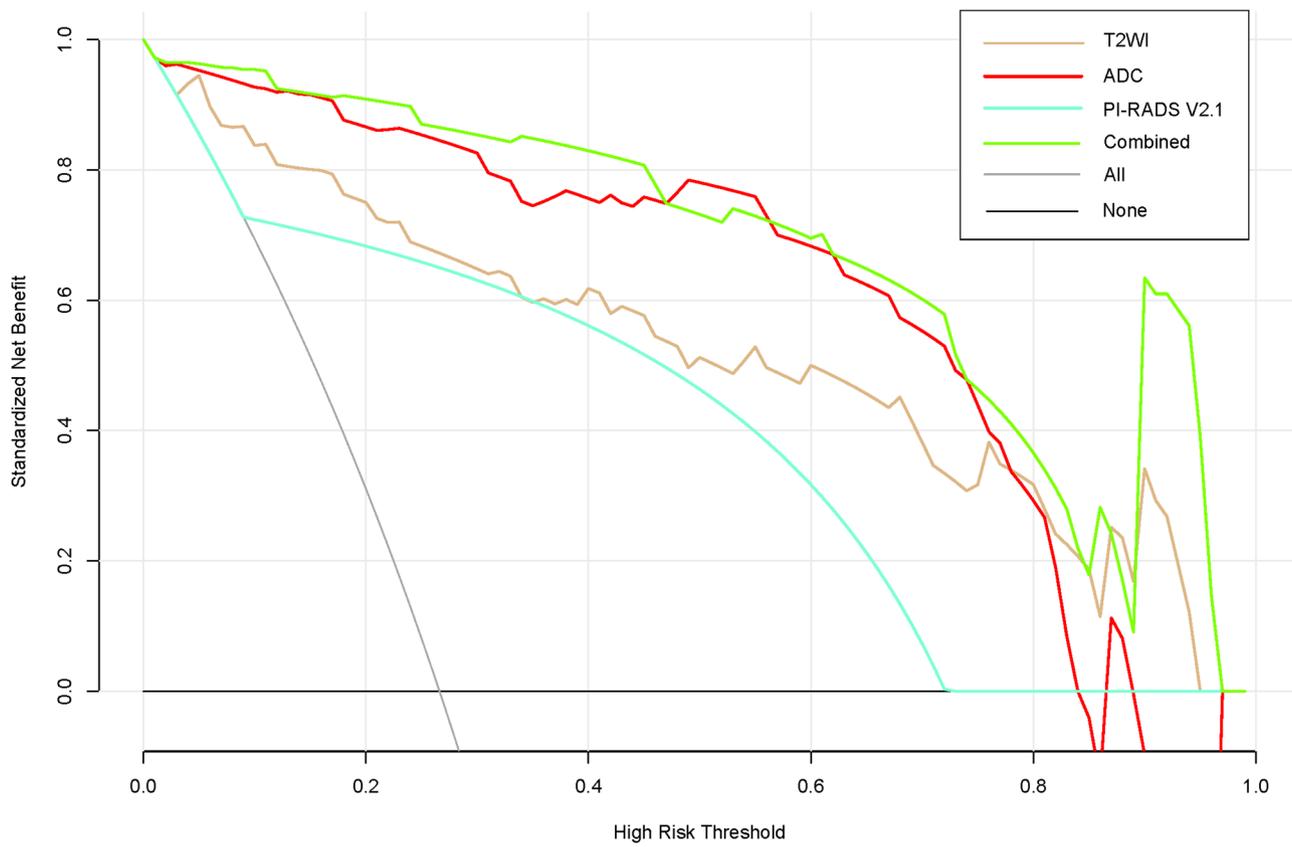


Fig. 6 Clinical benefits of these models were evaluated and compared, which indicated that the ADC, T2WI and combined models had better net clinical benefit than the PI-RADS v2.1 model

in PCa and csPCa diagnosis which was similar to that reported by Wang et al. [26]. They also concluded that machine learning-based MR radiomics contributed to improve the diagnostic value of PI-RADS v2 in PCa and csPCa. For diagnostic gray zone with PSA levels of 4–10 ng/mL, Qi et al. [22]. reported that the radiomics-based model performed better on both the training and validation cohorts compared to PI-RADS v2 score. Little prior studies focused on direct comparison between radiomic-based models and PI-RADS v2.1 score in PSA gray zone. Similar studies were only reported by Zhang et al. [15]. and Lu et al. [16]. They both established a combined model incorporating radiomic features and clinical data for the detection of PCa more accurately. Our purpose and methods differed from these two studies. In our study, head-to-head comparison between the radiomic-based model and PI-RADS v2.1 was only considered. Secondly, all lesions were pathologically evaluated only by systematic transrectal ultrasound (TRUS)-guided prostate biopsy in their studies. However, TRUS alone is unreliable for PCa detection and MRI-TRUS fusion biopsy can be suitable for enhancing PCa diagnosis [27]. In our study, MRI-TRUS fusion targeted biopsy was used for suspicious PCa lesions on MRI and more targeted

cores would be added for these lesions, which made it more accurate and reliable.

Some studies have applied radiomic features for PCa screening in patients with PSA 4–10 ng/mL. Zhong et al. concluded that MRI-based radiomics outperformed PI-RADS v2.1 for noninvasive prediction of PCa in patients with PSA levels of 4–10 ng/mL and could help improve the diagnostic performance for junior radiologists with less experience [28]. More importantly, the relationship between radiomic features and underlying biological information needs to be discussed. In our study, first-order features (mean, 10 percentile, entropy, kurtosis and skewness) differed significantly between cancerous and benign tissue on T2WI or ADC images. High values of entropy (randomness) and skewness (asymmetry) for ADC images, as well as high values of mean and kurtosis for T2WI images in cancerous lesions may reflect an increased probability of PCa due to irregular arrangement in PCa cellularity and absence of normal prostate glands. In addition, high value of entropy was related to the heterogeneity of PCa and may be useful to assess tumor aggressiveness [11]. In radiomics, second-order texture features, such as, gray level co-occurrence matrix (GLCM), gray level dependence matrix (GLDM), neighboring gray tone difference matrix (NGTDM) and gray

level size zone matrix (GLSZM) can provide a measure of intertumoral heterogeneity [9]. In our study, all four texture features exhibited a significant difference between benign and cancerous lesions. High values of dependence nonuniformity normalized (DNN) and strength, as well as low values of large dependence emphasis (LDE) and small area emphasis (SAE) on T2WI or ADC image reflected the heterogeneity of PCa, representing heterogeneous signal intensity in cancerous lesion, which was similar to that reported by Qi et al. [22]. They also proposed that high value of entropy and low value of LDE were associated with PCa heterogeneity. In addition, studies have demonstrated that higher intra-tumoral heterogeneity is closely related to a worse prognosis [29]. We further verified the diagnostic performance of the PI-RADS v2.1, ADC, T2WI and combined models for PCa. Radiomic-based models (ADC, T2WI and combined) exhibited a better performance to PI-RADS v2.1 and the combined model was optimal with the highest AUC, which was similar to previous results [15, 16, 22].

Limitations

There are several limitations to our study. First, this is a retrospective study performed in a single center with relatively small study population. A larger sample size from multi-centers will be recruited to validate our results in further studies. Second, ROIs were manually segmented, rather than semi-automatic or automatic delineation, which was time-consuming and may affect the repeatability of the segmentation. Third, transitional zone and peripheral zone lesions were not separately analyzed due to the limited number of patients in this study. In the future, we will increase the study population and evaluate the performance of radiomic-based model in each zone. Fourth, we only extracted partial data from this group as internal validation to test the model. An external validation cohort should be included to validate the robustness of our model in future.

Conclusions

In conclusion, we developed and internally validated radiomics models based on ADC and T2WI images for discriminating suspicious PCa patients with PSA 4–10 ng/mL in this study. Each radiomics model achieved satisfactory diagnostic performance, which was significantly better than the PI-RADS v2.1 score. This indicated the positive value and superiority of MRI-based radiomics in differentiating suspicious PCa in patients with PSA 4–10 ng/mL beyond routine evaluation.

Abbreviations

PCa	Prostate cancer
PSA	Transrectal ultrasound, prostate specific antigen
PI-RADS v2.1	Prostate imaging and data system version 2.1
T2WI	T2-weighted imaging

DWI	Diffusion-weighted imaging
DCEI	Dynamic contrast-enhanced imaging
f-PSA	Free PSA
t-PSA	Total PSA
ADC	Apparent diffusion coefficient
ROI	Region of interest
GS	Gleason score
IBSI	Image Biomarker Standardization Initiative
ICC	Intraclass correlation coefficient
LASSO	Least absolute shrinkage and selection operator
AUC	Area under the curve
ROC	Receiver operating characteristic
PPV	Positive predictive value
NPV	Negative predictive value
DCA	Decision curve analysis
GLCM	Gray level co-occurrence matrix
GLDM	Gray level dependence matrix
NGTDM	Neighboring gray tone difference matrix
GLSZM	Gray level size zone matrix

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12894-024-01625-2>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5

Acknowledgements

The authors would like to thank all participants who were enrolled in this study. The authors thank Yueyue Zhang, Rui Zhang and Tong Chen for their help in image post-processing.

Author contributions

The two authors (Chunxing Li and Zhicheng Jin) contributed equally to the work. CX Li*: Data curation, Design, Writing-Original draft. ZC Jin*: Writing Editing. CG Wei: Methodology. GC Dai: Visualization. J Tu: Investigation. JK Shen#: Funding acquisition.

Funding

This study was financially supported by the Suzhou Science and Technology Bureau Development Plan (Grant No. SYS2020147), the National Natural Science Foundation of China (Grant No. 81801754) and the Suzhou Science and Technology Development Plan (Grant No. SS2019012).

Data availability

The datasets analysed during the current study available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethical Committee of the Second Affiliated Hospital of Soochow University and the requirement for informed consent was waived by the Ethical Committee (ethics approval number: JD-HG-2024-075). Authors confirmed that all methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 22 January 2024 / Accepted: 10 October 2024

Published online: 23 October 2024

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(1):17–48.
2. Pienta KJ. Critical appraisal of prostate-specific antigen in prostate cancer screening: 20 years later. *Urology.* 2009;73(5 Suppl):S11–20.
3. Zhou Y, Li Y, Li X, Jiang M. Urinary Biomarker Panel to Improve Accuracy in Predicting Prostate Biopsy Result in Chinese Men with PSA 4–10 ng/mL. *Biomed Res Int.* 2017. 2017: 2512536.
4. Turkbey B, Rosenkrantz AB, Haider MA, et al. *Eur Urol.* 2019;76(3):340–51. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2.
5. Byun J, Park KJ, Kim MH, Kim JK. Direct comparison of PI-RADS version 2 and 2.1 in Transition Zone lesions for detection of prostate Cancer: preliminary experience. *J Magn Reson Imaging.* 2020;52(2):577–86.
6. Wei CG, Zhang YY, Pan P, et al. Diagnostic Accuracy and Interobserver Agreement of PI-RADS Version 2 and Version 2.1 for the detection of Transition Zone prostate cancers. *AJR Am J Roentgenol.* 2021;216(5):1247–56.
7. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical image using advanced feature analysis. *Eur J Cancer.* 2012;48(4):441–6.
8. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging.* 2012;30(9):1234–48.
9. Tomaszewski MR, Gillies RJ. The Biological meaning of Radiomic features. *Radiology.* 2021;298(3):505–16.
10. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006.
11. Gillies RJ, Kinahan PE, Hricak H, Radiomics. Images are more than pictures, they are data. *Radiology.* 2016;278(2):563–77.
12. 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(1):184–93.
13. Li T, Sun L, Li Q, et al. Development and validation of a Radiomics Nomogram for Predicting clinically significant prostate Cancer in PI-RADS 3 lesions. *Front Oncol.* 2021;11:825429.
14. Chen T, Li M, Gu Y, et al. Prostate Cancer differentiation and aggressiveness: Assessment with a Radiomic-based model vs. PI-RADS v2. *J Magn Reson Imaging.* 2019;49(3):875–84.
15. Zhang L, Zhang J, Tang M, Lei XY, Li LC. MRI-Based Radiomics Nomogram for Predicting prostate Cancer with Gray-Zone prostate-specific Antigen levels to reduce unnecessary biopsies. *Diagnostics (Basel).* 2022;12(12):3005.
16. Lu Y, Li B, Huang H, et al. Biparametric MRI-based radiomics classifiers for the detection of prostate cancer in patients with PSA serum levels of 4–10 ng/mL. *Front Oncol.* 2022;12:1020317.
17. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. 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(2):244–52.
18. Martorana E, Pirola GM, Scialpi M, et al. Lesion volume predicts prostate cancer risk and aggressiveness: validation of its value alone and matched with prostate imaging reporting and data system score. *BJU Int.* 2017;120:92–103.
19. Koo TK, Li MY. A Guideline of selecting and reporting Intraclass correlation coefficients for Reliability Research. *J Chiropr Med.* 2016;15(2):155–63.
20. Tamada T, Sone T, Higashi H, et al. Prostate cancer detection in patients with total serum prostate-specific antigen levels of 4–10 ng/mL: diagnostic efficacy of diffusion-weighted imaging, dynamic contrast-enhanced MRI, and T2-weighted imaging. *AJR Am J Roentgenol.* 2011;197:664–70.
21. Weinreb JC, Barentsz JO, Choyke PL, et al. *Eur Urol.* 2016;69(1):16–40. PI-RADS Prostate Imaging-Reporting and Data System: 2015, Version 2.
22. Qi Y, Zhang S, Wei J, et al. Multiparametric MRI-Based radiomics for prostate Cancer screening with PSA in 4–10 ng/mL to reduce unnecessary biopsies. *J Magn Reson Imaging.* 2020;51(6):1890–9.
23. Han C, Liu S, Qin XB, Ma S, Zhu LN, Wang XY. MRI combined with PSA density in detecting clinically significant prostate cancer in patients with PSA serum levels of 4–10 ng/mL: biparametric versus multiparametric MRI. *Diagn Interv Imaging.* 2020;101(4):235–44.
24. Brembilla G, Dell'Oglio P, Stabile A, et al. Interreader variability in prostate MRI reporting using prostate imaging reporting and Data System version 2.1. *Eur Radiol.* 2020;30(6):3383–92.
25. Dwivedi DK, Kumar R, Dwivedi AK, et al. Prebiopsy multiparametric MRI-based risk score for predicting prostate cancer in biopsy-naive men with prostate-specific antigen between 4–10 ng/mL. *J Magn Reson Imaging.* 2018;47(5):1227–36.
26. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. 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.* 2017;27(10):4082–90.
27. Kuru TH, Fütterer JJ, Schiffmann J, Porres D, Salomon G, Rastinehad AR. Transrectal Ultrasound (US), contrast-enhanced US, Real-time Elastography, HistoScanning, Magnetic Resonance Imaging (MRI), and MRI-US Fusion Biopsy in the diagnosis of prostate Cancer. *Eur Urol Focus.* 2015;1(2):117–26.
28. Zhong JG, Shi L, Liu J, et al. Predicting prostate cancer in men with PSA levels of 4–10 ng/mL: MRI-based radiomics can help junior radiologists improve the diagnostic performance. *Sci Rep.* 2023;13(1):4846.
29. Yadav SS, Stockert JA, Hackert V, Yadav KK, Tewari AK. Intratumor heterogeneity in prostate cancer. *Urol Oncol.* 2018;36(8):349–60.

Publisher's note

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