Multiparametric MRI-Based Radiomics for Prostate Cancer Screening With PSA in 4–10 ng/mL to Reduce Unnecessary Biopsies

Yafei Qi, MM,[1] Shuaitong Zhang, PhD,[2,3] Jingwei Wei, PhD,[2,3] Gumuyang Zhang, MD,[1] Jing Lei, MD,[1] Weigang Yan, MD,[4] Yu Xiao, MD,[5] Shuang Yan, MD,[1] Huadan Xue, MD,[1] Feng Feng, MD,[1] Hao Sun, MD,[1]* Jie Tian, PhD,[2,6,7]* and Zhengyu Jin, MD[1]*

Background: Whether men with a prostate-specific antigen (PSA) level of 4–10 ng/mL should be recommended for a biopsy is clinically challenging.

Purpose: To develop and validate a radiomics model based on multiparametric MRI (mp-MRI) in patients with PSA levels of 4–10 ng/mL to predict prostate cancer (PCa) preoperatively and reduce unnecessary biopsies.

Study Type: Retrospective.

Subjects: In all, 199 patients with PSA levels of 4–10 ng/mL.

Field Strength/Sequence: 3T, T_2-weighted, diffusion-weighted, and dynamic contrast-enhanced MRI.

Assessment: Lesion regions of interest (ROIs) from T_2-weighted, diffusion-weighted, and dynamic contrast-enhanced MRI were annotated by two radiologists. A total of 2104 radiomic features were extracted from the ROI of each patient. A random forest classifier was used to build the radiomics model for PCa in the primary cohort. A combined model was constructed using multivariate logistic regression by incorporating the radiomics signature and clinical-radiological risk factors.

Statistical Tests: For continuous variables, variance equality was assessed by Levene's test and Student's *t*-test, and Welch's *t*-test was used to assess between-group differences. For categorical variables, Pearson's chi-square test, Fisher's exact test, or the approximate chi-square test was used to assess between-group differences. $P < 0.05$ was considered statistically significant.

Results: The combined model incorporating the multi-imaging fusion model, age, PSA density (PSAD), and the PI-RADS v2 score yielded area under the curve (AUC) values of 0.956 and 0.933 on the primary ($n = 133$) and validation ($n = 66$) cohorts, respectively. Compared with the clinical-radiological model, the combined model performed better on both the primary and validation cohorts ($P < 0.05$). Furthermore, the use of the combined model to predict PCa could identify more negative PCa patients than the use of the clinical-radiological model by 18.4%.

Data Conclusion: The combined model was developed and validated to provide potential preoperative prediction of PCa in men with PSA levels of 4–10 ng/mL and might aid in treatment decision-making and reduce unnecessary biopsies.
Level of Evidence: 3
Technical Efficacy Stage: 3

PROSTATE CANCER (PCa) is the most common cancer and the second most common cause of cancer-related death in American men.[1] Its increasing incidence and decreasing mortality are attributed to early screening via prostate-specific antigen (PSA) testing.[2] Generally, patients with PSA levels greater than the traditional cutoff of 4 ng/mL (not including 4 ng/mL) are considered for a prostate biopsy. However, the clinical and financial costs of overdiagnosis and unnecessary biopsy procedures resulting from false positives are high, especially in men with a PSA level of 4–10 ng/mL.[3] According to the National Comprehensive Cancer Network (NCCN), only ~18% of men with a PSA level in the 4–10 ng/mL range will be diagnosed with cancer in the subsequent biopsy, which indicates that 82% of men with a PSA level in this range have undergone an unnecessary biopsy and therefore have suffered from postbiopsy complications, such as bleeding, infection, and urinary retention.[4,5] Thus, the prebiopsy prediction of PCa for individuals with PSA levels of 4–10 ng/mL is clinically challenging.

Concerning this issue, researchers have found that the PSA density (PSAD), pro-PSA isoforms,[6] PCa antigen 3,[7] DNA in the serum,[8] and microRNAs in the urine[9] are useful for PCa screening. Multiparametric magnetic resonance imaging (mpMRI) combined with Prostate Imaging Reporting and Data System Version 2 (PI-RADS v2) is considered the most promising approach for PCa screening because of its noninvasiveness and direct assessment.[10,11] Xu et al demonstrated that the optimal cutoff point for the diagnosis of PCa in individuals with PSA levels of 4–10 ng/mL was a PI-RADS v2 score of 3, which yielded a negative predictive value (NPV) of 85.1% but a low specificity of 62.1%, resulting in a high percentage of false positives.[12]

Radiomic analysis extracts high-throughput and quantitative image features from medical images and mines information related to tumor pathophysiology using machine-learning algorithms, which might contribute to medical decision-making and improve diagnostic and predictive performance.[13,14] One study demonstrated good performance of lung cancer screening with radiomic analysis based on low-dose chest computed tomography.[15] Therefore, we hypothesized that an mpMRI-based radiomics approach might improve the diagnostic performance of PCa screening, which could reduce unnecessary biopsies.

Materials and Methods

Study Design
This retrospective study was approved by the Ethics Committee of our institution, and the informed consent requirement was waived. The study design is illustrated in Fig. 1. In all, 440 patients with elevated PSA
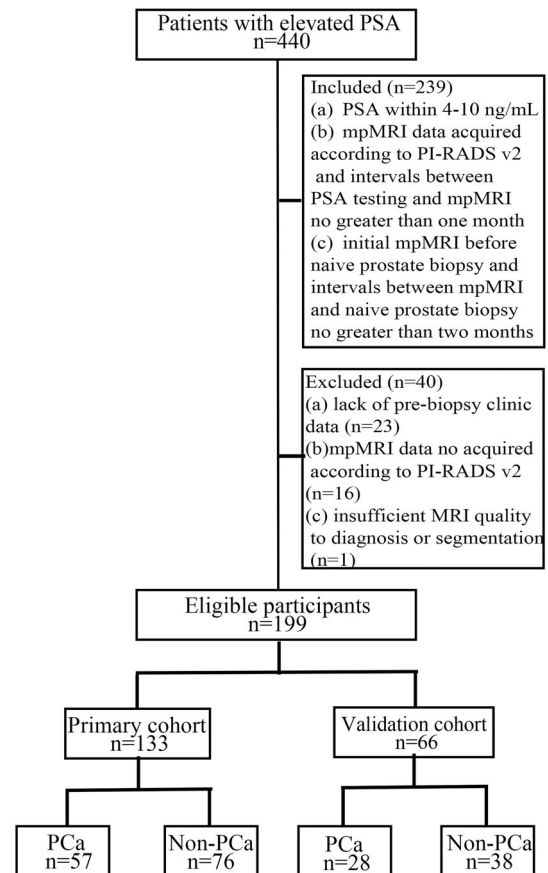
FIGURE 1: The flow diagram of this study.

PSA levels who were diagnosed between December 2015 and March 2018 were consecutively included in this study. The inclusion criteria were as follows: 1) PSA level within 4–10 ng/mL; 2) mpMRI data acquired according to the PI-RADS v2 criteria, including T_2-weighted imaging (T_2WI), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced (DCE) imaging, with an interval between PSA testing and mpMRI of no more than 1 month; and 3) initial mpMRI performed before the naïve prostate biopsy with an interval of less than 2 months. The exclusion criteria were as follows: i) lack of clinical data; ii) lack of any T_2W, DW, or DCE images; and iii) insufficient MRI quality to diagnose or segment the disease (eg, due to motion artifacts or catheter implantation). Finally, 199 patients were randomly assigned to the primary or validation cohort at a 2:1 ratio. The primary cohort was used to build the predictive model for PCa screening, and the validation cohort was completely held out; the validation cohort was not used to optimize the prognostic model and was used only to test the performance of the model built on the primary cohort.

MR Image Acquisition and Image Interpretation
MRI was performed with a 3.0T MR scanner (Discovery MR 750, GE Medical Systems, Milwaukee, WI). The mpMRI protocol

consisted of T_2W, DW, or DCE images. The details of MR image acquisition are provided in Supplement S1.

Two radiologists (Y.Q. with 1 year of experience in prostate MRI, and G.Z. with 5 years of experience in prostate MRI) assessed the images according to the PI-RADS v2 criteria but were blinded to the histopathologic interpretation. In the case of disagreement between the two radiologists, the final PI-PADS v2 score was discussed with a third radiologist (H.S. with 10 years of experience in prostate MRI) to reach a consensus (Fig. 2a). The PI-RADS v2 scores were assessed on each of the T_2WI, DWI, and DCE-MRI sequences. If there were multiple lesions, the PI-RADS v2 score of the index lesion demonstrating the largest size or the most aggressive feature was assigned to the patient.

Prostate Biopsy

Patients underwent a 12-core systematic transperineal ultrasound-guided prostate biopsy by a group of urinary specialists (one urinary specialist with 13 years of experience in prostate biopsy, another urinary specialist with 14 years of experience in prostate biopsy, and W.Y. with 20 years of experience in prostate biopsy) (Fig. 2b). The prostate gland was divided into 11 regions on ultrasound scans. At least one additional targeted biopsy was performed; the cognitive targeted biopsy using cognitive registration was based on zonal anatomy or imaging landmarks, such as remarkable nodules. The details of the cognitive targeted biopsy are as follows: first, the urologist reviewed the MRI results; and second, the urologist used the MRI information to perform the targeted biopsy for the most remarkable nodules guided by ultrasound images.

Lesion Annotation on MRI and Segmentation

To accurately annotate the prostate lesion extent on MRI, the coregistration of histopathologic specimens of the 12-core prostate biopsy and MRI was performed through a systematic review by a urinary pathologist (Y.X. with 5 years of experience in genitourinary pathology) and two radiologists (Y.Q. and G.Z.) (Fig. 2c). All of the histopathologic specimens were labeled as cancer or benign prostate lesions using a combination of primary and secondary Gleason grades by a pathologist (Y.X.) (Fig. 2d). Two radiologists (Y.Q. and G.Z.) manually delineated the maximum extent of the visible lesion using ITK-SNAP software with no prior knowledge of the histopathological results (Fig. 2e). As shown in the flowchart illustrated in Fig. 2, the PI-RADS reports, biopsy results, and radiologist's segmentations were matched for each patient.

Feature Extraction and Selection

Feature extraction was performed based on an open-source Python package (Pyradiomics 2.0.1).[16] Radiomic features indicating suspected PCa phenotypes were extracted from regions of interest (ROIs) on T_2W, apparent diffusion coefficient (ADC), and DCE map images according to 1) first-order statistics; 2) shape and size; 3) texture; 4) wavelet filter; and 5) Laplacian of Gaussian (LoG) filter features. Details of these features are shown in Supplement S2.

To select stable and reproducible features, both intra- and interobserver stability analyses were performed, and intra- and interobserver correlation coefficients greater than 0.85 were included in the subsequent analyses.[17] To select features with the greatest diagnostic value for PCa, variance filter assessment, two-sample Student's t-tests and Pearson's correlation coefficient analyses were performed sequentially on T_2W, ADC, and DCE MR images of the primary cohort. Details of feature selection are shown in Supplement S3.

Development of the Clinical-Radiological Model

The clinical-radiological risk factors for PCa included age, the PSA level, prostate volume, PSAD, location (transition zone [TZ] vs. peripheral zone [PZ]), early enhancement (yes vs. no), and the PI-RADS v2 score (1-2/3/4-5). A univariate analysis was conducted to
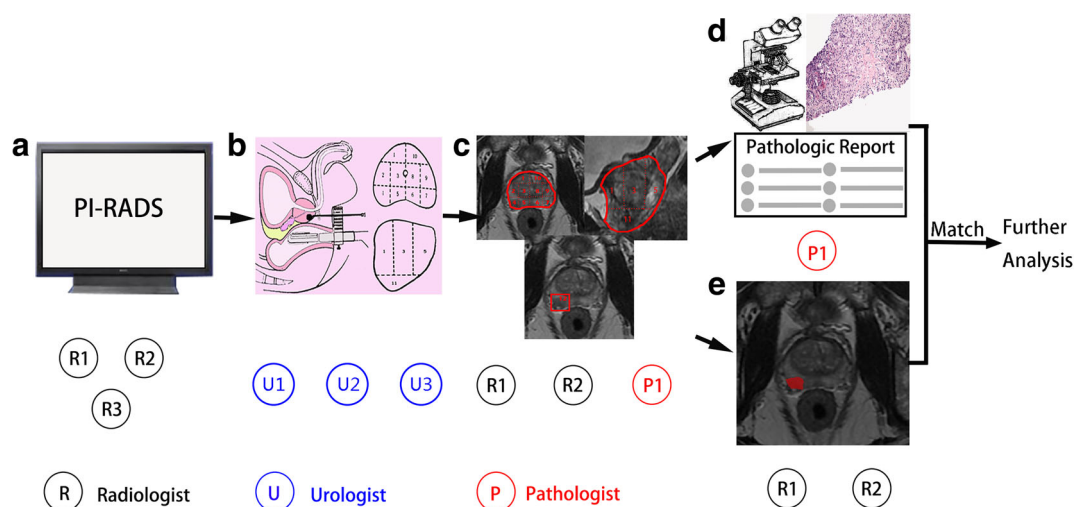
FIGURE 2: The flowchart for lesion coregistration of the histopathologic specimens and MRI. (a) Two radiologists (R1 and R2) assessed the MR images according to the PI-RADS v2 criteria. In the case of disagreement between the two radiologists, the final PI-PADS v2 score was discussed with a third radiologist (R3) to reach a consensus. (b) A group of urinary specialists (U1, U2, and U3) performed a 12-core systematic transperineal ultrasound-guided prostate biopsy for patients. (c) The coregistration of the positions of the 12-core prostate biopsy and MRI was performed through a histological-radiological correlative review by a urinary pathologist (P1) and two radiologists (R1 and R2). (d) All of the histopathologic specimens were assigned by a pathologist (P1) as a combination of the primary and secondary Gleason grades. (e) Two radiologists (R1 and R2) manually delineated the maximum extent of the visible lesion using ITK-SNAP software for further analysis.

analyze differences between patients with and without PCa for each clinical-radiological factor. In the primary cohort, factors with *P*-values less than 0.05 were used to build the clinical-radiological model by multivariate logistic regression, in which backward stepwise selection was conducted with the Akaike information criterion as the stopping rule. Because volume and PSAD are highly correlated, we could not include both in the predictive model. In this study, we included PSAD in the multivariate logistic regression analysis.

Development of the Radiomics and Combined Models

Based on the selected radiomic features, a random forest classifier was used to build the predictive models for PCa in the primary cohort using T_2W, ADC, and DCE MR images. The details of the model are shown in Supplement S4.

To explore the advantage of combining T_2W, ADC, and DCE MR images, a multi-imaging fusion model was built based on the mean of the predictive probability of the T_2W, ADC, and DCE imaging models in the primary cohort, which could reduce overfitting.

To explore the advantage of combining the radiomics and clinical-radiological models, a combined model incorporating the multi-imaging fusion model, age, PSAD, and the PI-RADS v2 score was built with multivariate logistic regression.

Validation of the Predictive Models

The predictive models were first assessed in the primary cohort (*n* = 133) and then validated in the validation cohort (*n* = 66). Receiver operation characteristic (ROC) curve, area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and NPV analyses were used to assess the performance of these models. In addition, the performance of these models in the TZ was also evaluated. Notably, the specificity and NPV were relatively important in this study. Relatively high specificity indicated a low probability of misclassifying non-prostate cancer (non-PCa) patients as positive and a low probability of overdiagnosis. In contrast, a relatively high NPV indicated a low probability of misclassifying PCa as non-PCa in negative patients and a high validity of the model for the prediction of non-PCa. Calibration curves and Hosmer–Lemeshow tests were conducted to assess the agreement between the combined model-predicted and expected probabilities. Decision curve analysis (DCA) was conducted to assess the clinical usefulness and benefits of the combined model at different threshold probabilities.

Statistical Analysis

Statistical analyses were conducted using R v. 3.3.4 and Python v. 3.5. Continuous variables are shown as the mean ± standard deviation (SD) and categorical variables are shown as percentages. For continuous variables, variance equality was assessed by Levene's test and Student's *t*-test, and Welch's *t*-test was used to assess between-group differences. For categorical variables, Pearson's chi-square test, Fisher's exact test, or the approximate chi-square test, if needed, was used to assess between-group differences. Two-sided *P*-values less than 0.05 were considered statistically significant.

Results

Study Population

No significant difference in age, the PSA level, prostate volume, PSAD, location, early enhancement, PI-RADS v2 score, digital rectal examination (DRE), positive cores, or Gleason score was observed between the primary and validation cohorts (*P* = 0.106–0.728), while a significant difference in early enhancement was found between the primary and validation cohorts (*P* < 0.001) (Supplement S5).

Development of the Clinical-Radiological Model

The univariate analysis of the clinical characteristics of PCa and non-PCa patients in the primary and validation cohorts is shown in Table 1. The results showed that age (*P* = 0.002), volume (*P* = 0.001), PSAD (*P* = 0.002), early enhancement (*P* < 0.001), PI-RADS v2 score (*P* < 0.001), the DRE (*P* = 0.004), positive cores (*P* < 0.001), and Gleason score (*P* < 0.001) were significantly different between PCa and non-PCa patients in the primary cohort. Age (*P* = 0.003), PSAD (*P* = 0.002), the DRE (*P* = 0.120), and the PI-RADS v2 score (*P* = 0.002) were selected to develop the clinical-radiological model by multivariate logistic regression.

Feature Selection

After removing features with both intra- and interobserver correlation coefficients less than 0.85, 438, 482, and 255 radiomic features remained for the ADC, T_2W, and DCE MR images, respectively. Finally, 34, 52, and 4 radiomic features were included for the ADC, T_2W, and DCE MR images, respectively (Supplement S6). In Table 2, we list the top ten features with the highest performance.

Validation of the Predictive Models

The ADC, T_2W, and DCE imaging models yielded AUC values of 0.910, 0.914, and 0.793, respectively, for the primary cohort and 0.853, 0.828, and 0.774, respectively, for the validation cohort. In addition to assessing these three models on the independent validation cohort, 10-fold cross-validation was also performed, in which the range of AUC values was 0.737–0.947 for the ADC imaging model, 0.750–0.954 for the T_2WI model, and 0.670–0.926 for the DCE imaging model. The multi-imaging fusion model performed better than the models based on each single MR image (AUC: 0.945 and 0.902 in the primary and validation cohorts, respectively). Compared to the multi-imaging fusion model, the performance of the clinical-radiological model (AUC: 0.806 and 0.858 in the primary and validation cohorts, respectively) was poorer.

The combined model incorporating the multi-imaging fusion model, age, PSAD, DRE, and PI-RADS v2 score exhibited the best performance (AUC: 0.956 and 0.933 in the primary and validation cohorts, respectively).

TABLE 1. Univariate Analysis of the Clinical Characteristics of Patients and Tumors in the Primary and Validation Cohorts

Characteristic	Primary cohort (n = 133)			Validation cohort (n = 66)		
	PCa (n = 57)	Non-PCa (n = 76)	P-value	PCa (n = 28)	Non-PCa (n = 38)	P-value
Age (yr, mean ± SD)	67.53 ± 7.28	62.51±10.05	0.002[a]	66.39±8.09	61.82±7.32	0.021[a]
PSA (ng/mL, mean ± SD)	7.04±1.67	6.97±1.61	0.811	6.96±1.49	6.89±1.66	0.862
Prostate volume (mL, mean ± SD)	40.32±20.21	53.33±24.01	0.001[a]	31.43±14.56	50.79±17.68	1.651E-05[a]
PSAD (ng/mL, mean ± SD)	0.22±0.0.11	0.16±0.09	0.002[a]	0.3±0.2	0.15±0.06	0.001[a]
Location (%)			<0.001[b]			5.514E-06[b]
TZ	27 (47.4%)	60 (78.9%)		11 (39.3%)	35 (92.1%)	
PZ	30 (52.6%)	16 (21.1%)		17 (60.7%)	3 (7.9%)	
DCE type (%)			<0.001[b]			0.745
Early enhancement	14 (24.6%)	7 (9.2%)		12 (42.9%)	16 (42.1%)	
Nonearly enhancement	43 (75.4%)	69 (90.8%)		16 (57.1%)	22 (57.9%)	
PI-RADS v2 grade (%)			1E-05[b]			0.003[b]
≤2	6 (10.5%)	27 (35.5%)		5 (17.9%)	15 (39.5%)	
3	2 (3.5%)	14 (18.4%)		0	6 (15.8%)	
≥4	49 (86%)	35 (46.1%)		23 (82.1%)	17 (44.7%)	
DRE			0.004[b]			0.011[b]
0	42 (73.7%)	70 (92.1%)		23 (82.1%)	38 (100.0%)	
1	15 (26.3%)	6 (7.9%)		5 (17.9%)	0 (0.0%)	
Positive cores			<0.001[b]			<0.001[b]
0	0 (0.0%)	76 (100.0%)		0 (0.0%)	38 (100.0%)	
1-2	26 (43.9%)	0 (0.0%)		17 (60.7%)	0 (0.0%)	
>2	31 (54.4%)	0 (0.0%)		11 (39.3%)	0 (0.0%)	
Gleason score			<0.001[b]			<0.001[b]
0	0 (0.0%)	76 (100.0%)		0 (0.0%)	38 (100.0%)	
Low (=6)	25 (43.9%)	0 (0.0%)		17 (60.7%)	0 (0.0%)	
High (>6)	32 (56.1%)	0 (0.0%)		11 (39.3%)	0 (0.0%)	

PCa, prostate cancer; Non-PCa, nonprostate cancer; PSA, prostate-specific antigen; PSAD, prostate-specific antigen density; TZ, transitional zone; PZ, peripheral zone; early enhancement, DCE time curves type 1 (progressive); nonearly enhancement, DCE time curves type 2 (plateau) or type 3 (washout); DRE, digital rectal examination; Gleason score, low (Gleason = 3 + 3) and high (Gleason >3 + 3). A P-value <0.05 was considered a significant difference.
[a]Student's t-test.
[b]Pearson's test.

TABLE 2. Top Ten Features With the Highest Performance According to the Decrease in AUC for the Radiomics Model

Feature	Interpretation	Median (IQR)		P-value	AUC
		PCA	non-PCa		
T2WI_LF_Maximum	Grayscale/intensity	0.281 (−0.138, 1.125)	−0.572 (−0.894, −0.066)	<0.001	0.781
DCE_LLF_Minimum	Grayscale/intensity	−0.525 (−1.077, 0.108)	0.323 (−0.313, 0.966)	<0.001	0.276
T2WI_WF_MAD	Grayscale/intensity	0.121 (−0.346, 0.447)	−0.447 (−0.646, −0.134)	<0.001	0.721
T2WI_LG_DE	Randomness/ variability	0.408 (−0.443, 0.799)	−0.363 (−0.878, 0.092)	<0.001	0.705
T2WI_WF_Maximum	Grayscale/intensity	0.203 (−0.666, 1.145)	−0.762 (−0.912, 0.181)	<0.001	0.703
T2WI_LG_Contrast	Intensity variation	−0.005 (−0.313, 0.767)	−0.481 (−0.813, 0.062)	<0.001	0.700
T2WI_LG_JE	Randomness/ variability	0.365 (−0.159, 0.962)	−0.278 (−0.936, 0.373)	<0.001	0.697
T2WI_LF_90Percentile	Grayscale/intensity	0.308 (−0.363, 0.9)	−0.381 (−0.829, −0.054)	<0.001	0.696
T2WI_LG_LDE	Textures	−0.506 (−1.081, 0.254)	0.214 (−0.402, 1.145)	<0.001	0.307
ADC_LG_LALGLE	Textures	−0.433 (−0.46, −0.254)	−0.266 (−0.403, 0.206)	<0.001	0.309

T2WI_LF_Maximum: T2WI_LoG-sigma-1-0-mm-3D_firstorder_Maximum; DCE_LLF_Minimum: DCE_LL_firstorder_Minimum; T2WI_WF_MAD: T2WI_wavelet-HH_firstorder_MeanAbsoluteDeviation; T2WI_LG_DE: T2WI_log-sigma-1-0-mm-3D_glcm_DifferenceEntropy; T2WI_WF_Maximum: T2WI_wavelet-HH_firstorder_Maximum; T2WI_LG_Contrast: T2WI_log-sigma-3-0-mm-3D_glcm_Contrast; T2WI_LG_JE: T2WI_log-sigma-3-0-mm-3D_glcm_JointEntropy; T2WI_LF_90Percentile: T2WI_log-sigma-1-0-mm-3D_firstorder_90Percentile; T2WI_LG_LDE: T2WI_log-sigma-3-0-3D_gldm_LargeDependeceEmphasis; ADC_LG_LALGLE: ADC_log-sigma-3-0-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis.

Furthermore, the NPV and specificity of the combined model were relatively higher than those of the clinical-radiological model in the validation cohort (0.875 vs. 0.824 and 0.921 vs. 0.737). The performance of these models is shown in Table 3, and the ROC curves of the clinical-radiological, multi-imaging fusion, and combined model are shown in Fig. 3a. The violin plots of the combined model are also shown in Fig. 3b. The calibration curve showed good agreement between the combined model predicted and expected probabilities of PCa ($P = 0.478$ and 0.837 in the primary and validation cohorts, respectively; Hosmer–Lemeshow test) (Fig. 4a,b), and DCA showed that using the combined model to predict PCa was more beneficial than using the clinical-radiological model (Fig. 4c). Furthermore, the DeLong test showed that the combined model performed better than the clinical-radiological model on both the primary and validation cohorts ($P < 0.05$, Table 4). The combined model also exhibited good performance in screening for PCa in the TZ (Table 5).

Discussion

It is important to avoid unnecessary and invasive prostate punctures and to subsequently curtail the overdiagnosis of men with PSA levels of 4–10 ng/mL because the prevalence of PCa and the positive rate of biopsies are both low. In this study, an mpMRI-based radiomics approach was conducted to develop and validate the predictive model for PCa screening. Compared to the clinical-radiological model and the PI-RADS v2 score, the combined model incorporating the clinical-radiological and radiomics models exhibited the best performance. Furthermore, using the combined model to predict PCa could identify more negative PCa than using the clinical-radiological model by 18.4%.

Several studies have investigated the diagnostic performance of clinical indicators and genes. For example, the AUC and specificity were 0.77–0.79 and 45–67.8%, respectively, for the prostate health index,[6,18] and those for DNA[19] were 0.84 and 68%, respectively. These results show that only 68% of non-PCa patients would avoid an unnecessary biopsy. Moreover, the results based on the univariate logistic regression analysis were not formally compared with PSA levels.[20] Schroder et al[21] recommended prostate mpMRI as a promising method to decrease overdiagnosis.

The PI-RADS v2 guidelines for mpMRI were published in 2015 and thereby increased the sensitivity and specificity for both small indolent tumors and large aggressive lesions.[10] Liu et al evaluated the ability of the PI-RIDS v2 score to detect PCa in patients with PSA levels of 4–10 ng/mL and

TABLE 3. Diagnostic Performance of the PI-RADS, Clinic-Radiological, ADC, T2FSE, DCE, Radiomics, and Combined Models

Model	Performance	AUC (95% CI)	ACC	SEN	SPE	PPV	NPV	*P*-value
PI-RADS ≥4	Primary cohort	0.708 (0.694–0.722)	0.677	0.837	0.583	0.583	0.837	5.82 E-6
	Validation cohort	0.726 (0.695–0.756)	0.667	0.808	0.575	0.575	0.808	6.54E-4
Clinical-radiological	Primary cohort	0.806 (0.793–0.819)	0.714	0.754	0.684	0.641	0.798	1.75E-9
	Validation cohort	0.858 (0.835–0.881)	0.758	0.786	0.737	0.688	0.824	1.42E-7
ADC	Primary cohort	0.910 (0.902–0.918)	0.812	0.754	0.855	0.796	0.823	6.88E-16
	Validation cohort	0.853 (0.829–0.878)	0.864	0.786	0.921	0.880	0.854	2.21E-7
T2WI	Primary cohort	0.914 (0.906–0.922)	0.820	0.877	0.776	0.746	0.894	3.51E-16
	Validation cohort	0.828 (0.804–0.852)	0.803	0.679	0.895	0.826	0.791	2.01E-6
DCE	Primary cohort	0.793 (0.780–0.807)	0.677	0.877	0.526	0.581	0.851	7.81E-9
	Validation cohort	0.774 (0.748–0.801)	0.682	**0.821**	0.579	0.590	0.815	1.55E-4
Radiomics	Primary cohort	0.945 (0.939–0.951)	0.842	0.965	0.750	0.743	0.966	<2.2E-16
	Validation cohort	0.902 (0.884–0.920)	0.803	0.786	0.816	0.759	0.838	1.26E-9
Combined	Primary cohort	0.956 (0.951–0.961)	0.895	0.912	0.882	0.853	0.931	<2.2E-16
	Validation cohort	**0. 933** (0.918–0.948)	**0.879**	**0.821**	**0.921**	**0.885**	**0.875**	1.92E-11

AUC = area under the curve; ACC = accuracy; SEN = sensitivity; SPE = specificity; PPV = positive predictive value; NPV = negative predictive value; ADC = apparent diffusion coefficient; T2WI = T2-weighted imaging; DCE = dynamic contrast-enhanced. The best performance in the validation cohort is indicated in bold font. The cutoff values were calculated using the xtile function in R.

found a specificity of 89.66% and an NPV of 92.86%.[22] However, this model, which was designed by a single radiologist, yielded a sensitivity of 60% and a PPV of 50%. Dwivedi et al[23] developed a combined model based on the PSA level, ADC, and non-PI-RADS-related magnetic resonance spectroscopy (MRS) to predict the risk of PCa in 137 consecutive men with PSA levels between 4 and 10 ng/mL. The authors achieved a specificity and an NPV slightly lower than our values, but the sensitivity and PPV of their model were obviously lower than those of our model. Our combined models both reduced negative cases to acceptable levels to avoid unnecessary biopsies and avoided missed diagnoses of positive cases.

Some studies have applied radiomics or texture features for PCa screening, but we were unable to find studies that examined PCa screening in patients with PSA levels of 4–10 ng/mL. Litjens et al[24] used five features to distinguish non-PCa from PCa in 70 patients, resulting in an AUC of 0.62. Sidhu et al[25] applied texture analysis based on ADC kurtosis (AUC = 0.78) and T_1 entropy (AUC = 0.66) for PCa detection in 67 cases, resulting in an AUC of 0.83. However, no independent validation cohort was included in these studies. In our study, low ADC and T_2WI values represented an increased probability of PCa, possibly because variations in cellularity in PCa result in normal gland loss,[26] representing enhanced signal reduction in the lesion. High values of entropy and size zone nonuniformity (SZN), as well as low values of large dependence emphasis (LDE), long run emphasis (LRE), and large area emphasis (LAE), for the ADC and T_2W images reflected the heterogeneity of PCa, which originates from angiogenesis and extravascular extracellularity, representing heterogeneous signal intensity in the lesion.[27] A large distribution of gray scales, such as the mean absolute deviation (MAD), maximum gray scales, and range in the ADC and/or T_2WI values indicate an increased possibility of PCa. This may be due to signaling changes caused by areas of hemorrhage and necrosis in PCa.[28] Low values of the minimum, 90th percentile, and mean in the washout phase represent the probability of PCa, possibly because of increasing tumor vascular permeability, resulting in a rapid decline in DCE sequence features.[29] PCa screening in the TZ is difficult due to confusion with benign prostatic hyperplasia (BPH); therefore, we further verified the performance of the models for PCa screening in the TZ. Ginsburg et al[30] utilized radiomic features to detect PCa in the TZ with an AUC of 0.68, which is different
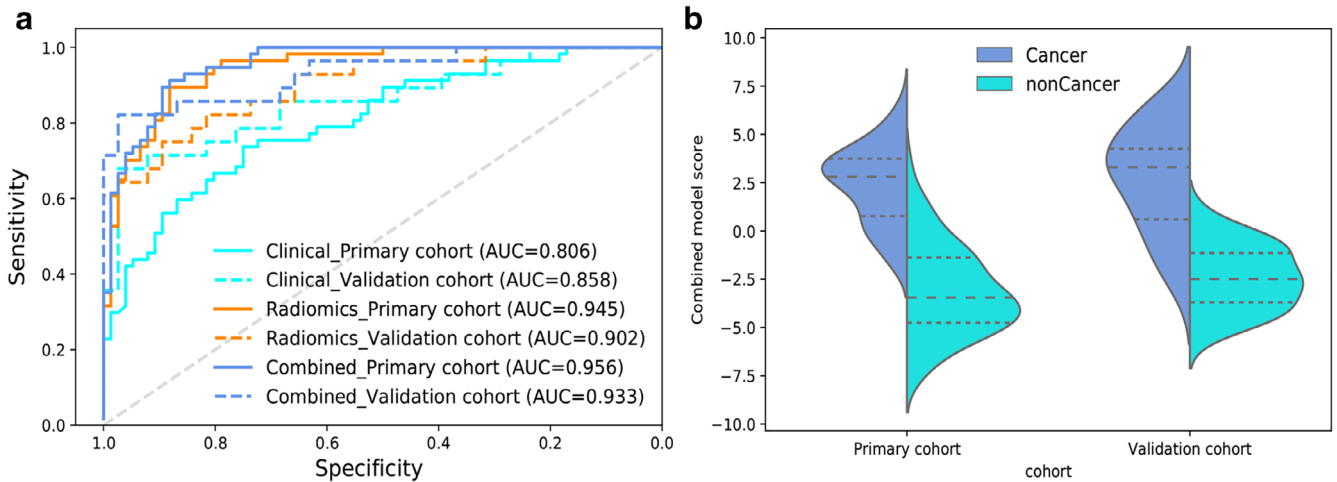
FIGURE 3: (a) ROC curves of the clinical-radiological, multi-imaging fusion, and combined models. (b) Violin plots of the combined model in the primary and validation cohorts.
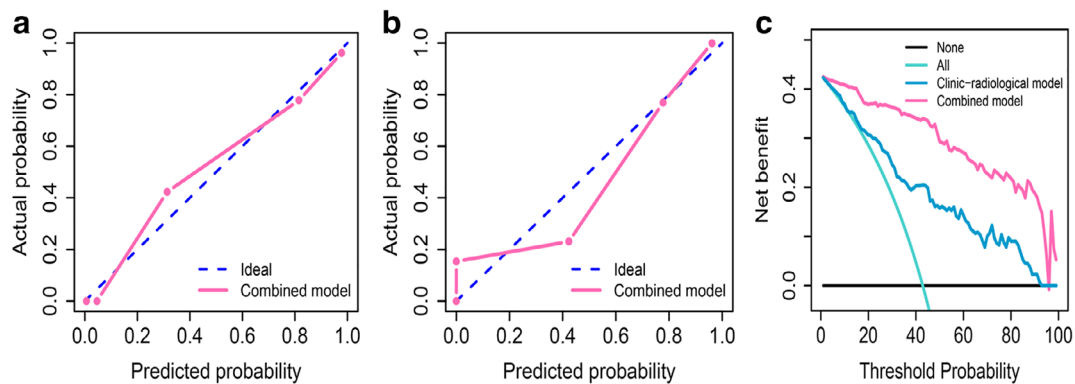

FIGURE 4: Calibration curve of the combined model in the primary (a) and validation (b) cohorts. (c) DCA of the combined (red line) and clinical-radiological (blue line) models.

TABLE 4. Pairwise Comparison Using the DeLong Test Among the Clinical-Radiological, Radiomics, and Combined Models

	DeLong test	Clinical-radiological	Radiomics	Combined
Primary cohort	Clinical-radiological	—	<0.001[a]	<0.001[a]
	Radiomics	<0.001[a]	—	0.102
	Combined	<0.001[a]	0.102	—
Validation cohort	Clinical-radiological	—	0.212	0.049[a]
	Radiomics	0.212	—	0.041[a]
	Combined	0.049[a]	0.041[a]	—

[a]A P-value <0.05 was considered a significant difference.

from the result of our combined model. This may have occurred because we obtained unified MR parameters using the same MRI equipment and coil in one institution. In our study, the radiomics model based on T_2-weighted fast spin echo (T_2FSE) sequences exhibited a similar performance to DWI sequences but a better performance than DCE sequences, which was consistent with the results of Greer et al.[31]

There were several limitations to our study. First, the sample size was limited. Although we utilized a validation cohort to verify the reproducibility of our models, the results

TABLE 5. Diagnostic Performance of the PI-RADS, Clinical-Radiological, ADC, T2WI, DCE, Radiomics, and Combined Models for PCa in the TZ

	Performance	AUC	ACC	NPV	PPV	SEN	SPE	*P*-value
PI-RADS ≥4	Primary cohort	0.706	0.678	0.833	0.487	0.704	0.667	8.11E-4
	Validation cohort	0.555	0.565	0.8	0.286	0.545	0.571	0.294
Clinical-radiological	Primary cohort	0.762	0.770	0.822	0.640	0.593	0.850	5.0E-5
	Validation cohort	0.709	0.696	0.862	0.412	0.536	0.714	0.020
ADC	Primary cohort	0.92	0.851	0.851	0.85	0.63	0.95	2.27E-10
	Validation cohort	0.813	0.848	0.889	0.7	0.636	0.914	1E-3
T2WI	Primary cohort	0.923	0.839	0.911	0.71	0.815	0.85	1.69E-10
	Validation cohort	0.839	0.87	0.914	0.727	0.727	0.914	4.07E-4
DCE	Primary cohort	0.814	0.69	0.902	0.5	0.852	0.617	1.57E-6
	Validation cohort	0.766	0.674	**0.955**	0.417	**0.909**	0.6	4.3E-3
Radiomics	Primary cohort	0.959	0.874	0.962	0.735	0.735	0.926	4.81E-12
	Validation cohort	0.878	0.826	0.909	0.615	0.727	0.857	9.42E-5
Combined	Primary cohort	0.972	0.908	0.919	0.880	0.815	0.950	1.24E-12
	Validation cohort	**0.904**	**0.891**	0.917	**0.800**	0.727	**0.943**	3.28E-5

TZ, transitional zone; AUC = area under the curve; ACC = accuracy; SEN = sensitivity; SPE = specificity; PPV = positive predictive value; NPV = negative predictive value; ADC = apparent diffusion coefficient; T2WI = T2-weighted imaging; DCE = dynamic contrast-enhanced; The best performance in the validation cohort is indicated in bold font. The cutoff values were calculated using the xtile function in R.

should be verified using a larger sample size. Second, MRI data were collected from a single center. However, we set unified MR scanning parameters according to the PI-RADS v2 criteria. Therefore, the robustness of our models should be verified by additional MRI data from multiple centers in future studies. Third, the population distribution was unbalanced in our study. PCa in the PZ is easily diagnosed and subsequently treated in local hospitals. Our hospital, which is a referral center in China, had more patients with PCa in the TZ than in the PZ. However, the stratified analysis showed that the combined model performed well among patients with lesions in the TZ. We did not verify the performance of the models to screen PCa occurring in the PZ because of the limited number of patients. In the future, we will examine the performance of our model in different zones to identify aggressive PCa. Fourth, we used a biopsy as the reference test in some patients, which resulted in the risk of mismatching some cancers. However, a systematic prostate biopsy is the recommended technique for a prostate biopsy according to the NCCN guidelines. Additionally, we identified lesions by comparing biopsy templates with mpMRI results to reduce this risk.

In conclusion, we developed and validated a radiomics model based on mpMRI, which possibly applies to predicting PCa in men with PSA levels of 4–10 ng/mL prebiopsy. Furthermore, the combined model incorporating the radiomics model and clinical-radiological risk factors might aid clinicians with better clinical treatment decision-making and reduce unnecessary prostate biopsies.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7–30.

2. Attard G, Parker C, Eeles RA, et al. Prostate cancer. Lancet 2016;387: 70–82.

3. Aminsharifi A, Howard L, Wu Y, et al. Prostate specific antigen density as a predictor of clinically significant prostate cancer when the prostate specific antigen is in the diagnostic gray zone: Defining the optimum cutoff point stratified by race and body mass index. J Urology 2018; 200:758–766.

4. Gershman B, Van Houten HK, Herrin J, et al. Impact of prostate-specific antigen (PSA) screening trials and revised PSA screening guidelines on rates of prostate biopsy and postbiopsy complications. Eur Urol 2017; 71:55–65.

5. Cicione A, De Nunzio C, Manno S, et al. An update on prostate biopsy in the era of magnetic resonance imaging. Minerva Urol Nefrol 2018; 70:264–274.

6. Tan LGL, Tan YK, Tai BC, et al. Prospective validation of %p2PSA and the Prostate Health Index, in prostate cancer detection in initial

prostate biopsies of Asian men, with total PSA 4-10 ng ml (-1). Asian J Androl 2017;19:286–290.

7. Donovan MJ, Noerholm M, Bentink S, et al. A molecular signature of PCA3 and ERG exosomal RNA from non-DRE urine is predictive of initial prostate biopsy result. Prostate Cancer Prostat Dis 2015;18: 370–375.

8. Creed J, Klotz L, Harbottle A, et al. A single mitochondrial DNA deletion accurately detects significant prostate cancer in men in the PSA 'grey zone.' World J Urol 2018;36:341–348.

9. Stuopelyte K, Daniunaite K, Bakavicius A, Lazutka JR, Jankevicius F, Jarmalaite S. The utility of urine-circulating miRNAs for detection of prostate cancer. Br J Cancer 2016;115:707–715.

10. Turkbey B, Brown AM, Sankineni S, Wood BJ, Pinto PA, Choyke PL. Multiparametric prostate magnetic resonance imaging in the evaluation of prostate cancer. CA Cancer J Clin 2016;66:326–336.

11. Manfredi M, Mele F, Garrou D, et al. Multiparametric prostate MRI: Technical conduct, standardized report and clinical use. Minerva Urol Nefrol 2018;70:9–21.

12. Xu N, Wu YP, Chen DN, et al. Can Prostate Imaging Reporting and Data System Version 2 reduce unnecessary prostate biopsies in men with PSA levels of 4-10 ng/ml? J Cancer Res Clin Oncol 2018;144: 987–995.

13. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–762.

14. Zhang S, Song G, Zang Y, et al. Non-invasive radiomics approach potentially predicts non-functioning pituitary adenomas subtypes before surgery. Eur Radiol 2018;28:3692–3701.

15. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25:954–961.

16. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017; 77:e104–e107.

17. Knight DS, Grasso AE, Quail MA, et al. Accuracy and reproducibility of right ventricular quantification in patients with pressure and volume overload using single-beat three-dimensional echocardiography. J Am Soc Echocardiogr 2015;28:363–374.

18. Ng CF, Chiu PK, Lam NY, Lam HC, Lee KW, Hou SS. The Prostate Health Index in predicting initial prostate biopsy outcomes in Asian men with prostate-specific antigen levels of 4-10 ng/mL. Int Urol Nephrol 2014;46:711–717.

19. Creed J, Klotz L, Harbottle A, et al. A single mitochondrial DNA deletion accurately detects significant prostate cancer in men in the PSA 'grey zone'. World J Urol 2018;36:341–348.

20. Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, diagnosis, and local treatment with curative intent. Eur Urol 2017;71:618–629.

21. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: Results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014; 384:2027–2035.

22. Liu C, Liu SL, Wang ZX, et al. Using the prostate imaging reporting and data system version 2 (PI-RIDS v2) to detect prostate cancer can prevent unnecessary biopsies and invasive treatment. Asian J Androl 2018;20:459–464.

23. 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;47:1227–1236.

24. Litjens GJ, Elliott R, Shih NN, et al. Computer-extracted features can distinguish noncancerous confounding disease from prostatic adenocarcinoma at multiparametric MR imaging. Radiology 2016;278: 135–145.

25. Sidhu HS, Benigno S, Ganeshan B, et al. Textural analysis of multiparametric MRI detects transition zone prostate cancer. Eur Radiol 2017;27:2348–2358.

26. Yamin G, Schenker-Ahmed NM, Shabaik A, et al. Voxel level radiologic-pathologic validation of restriction spectrum imaging cellularity index with Gleason grade in prostate cancer. Clin Cancer Res 2016;22:2668–2674.

27. Calio BP, Sidana A, Sugano D, et al. Risk of upgrading from prostate biopsy to radical prostatectomy pathology—Does saturation biopsy of index lesion during multiparametric magnetic resonance imaging-transrectal ultrasound fusion biopsy help? J Urol 2018;199:976–982.

28. Niu XK, Chen ZF, Chen L, Li J, Peng T, Li X. Clinical application of biparametric MRI texture analysis for detection and evaluation of high-grade prostate cancer in zone-specific regions. AJR Am J Roentgenol 2018;210:549–556.

29. Hansford BG, Peng Y, Jiang Y, et al. Dynamic contrast-enhanced MR imaging curve-type analysis: Is it helpful in the differentiation of prostate cancer from healthy peripheral zone? Radiology 2015;275: 448–457.

30. Ginsburg SB, Algohary A, Pahwa S, et al. Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study. J Magn Reson Imaging 2017;46:184–193.

31. Greer MD, Shih JH, Lay N, et al. Validation of the dominant sequence paradigm and role of dynamic contrast-enhanced imaging in PI-RADS version 2. Radiology 2017;285:859–869.