

Table S1 MRI scanning parameters

Variables	Sequence	Slice-thickness (mm)	Slice-gap (mm)	TR (ms)	TE (ms)
GE Discovery 3.0T MR	T1WI	4.0	1	620	10.2
	T2WI			3019	103
	DWI			2600	68.2
GE Signa HDxt 3.0T MR	T1WI	4.0	1	550	7.0
	T2WI			2760	68
	DWI			4150	76.4
GE Signa HDxt 1.5T MR	T1WI	4.0	1	629	10.5
	T2WI			3440	102
	DWI			4450	68.2
MAGNETOM Skyra 3.0T MR	T1WI	3.5	1	500	11
	T2WI			3800	78
	DWI			4700	83

DWI, diffusion-weighted imaging; MRI, magnetic resonance imaging; T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; TE, time to echo; TR, time of repetition.

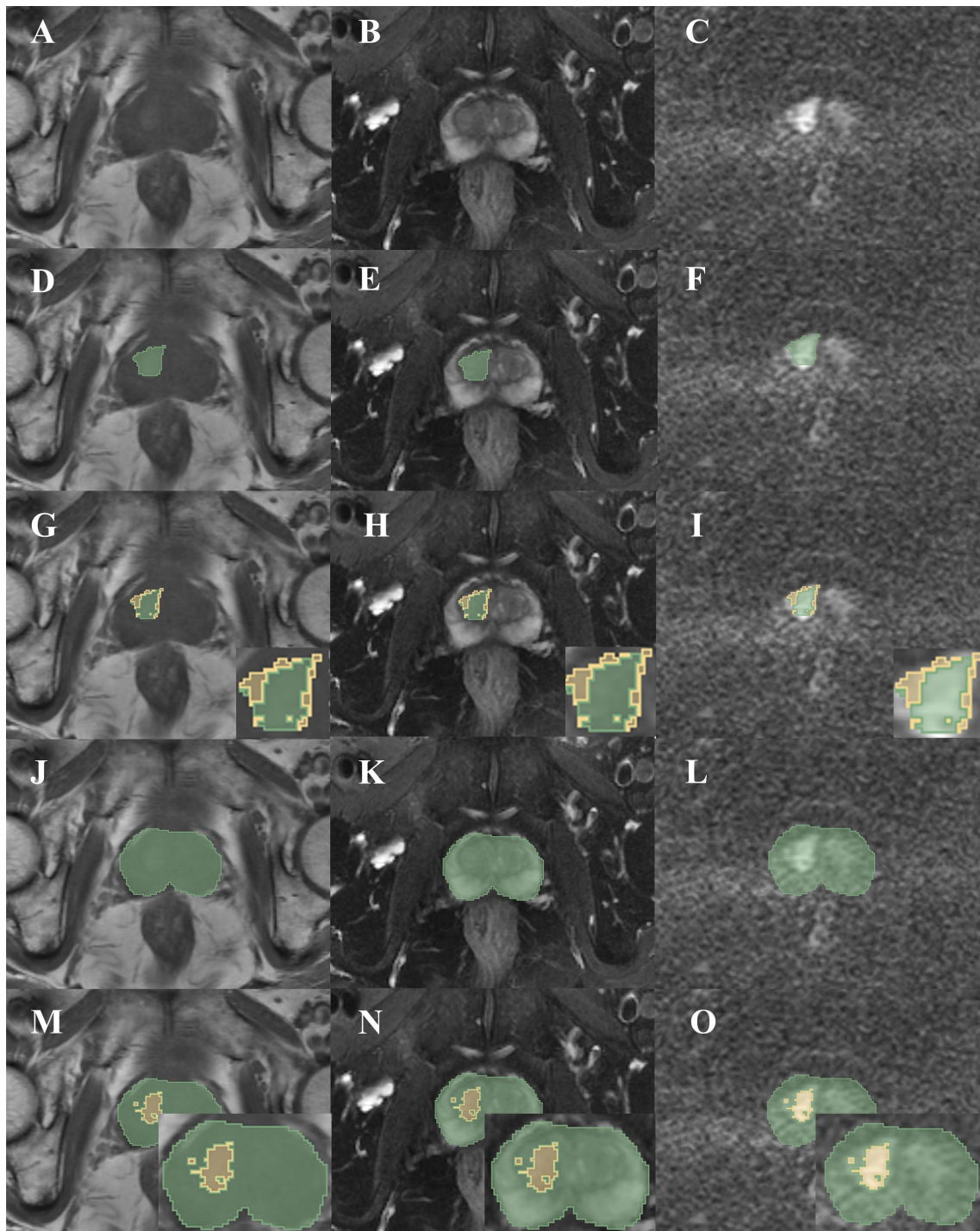


Figure S1 Representative habitat imaging. A 70-year-old man with pathologically confirmed PCa. (A-C) The images of T1WI, T2WI, and DWI sequences; (D-F) the region of interest of lesion. (G-I) the habitat subregions of lesion; (J-L) the region of interest of prostate gland; (M-O) the habitat subregions of prostate gland. The different color represented the different subregion.

Appendix 1 The process of reduction

The feature reduction process was conducted in the following manner. Initially, an intraclass correlation coefficient analysis (ICC) was employed to evaluate the consistency of the classical radiomics features between two radiologists. Features with ICC <0.75 were excluded. Secondly, the Pearson correlation coefficient was employed to evaluate the inter-feature correlation, with one of the two features exhibiting a correlation value exceeding 0.7 being excluded. Once more, the features were subjected to further screening using the least absolute shrinkage and selection operator (LASSO) 10-fold cross-validation, which not only converges on the features but also obtains the weight coefficients of the remaining features simultaneously. Ultimately, the remaining radiomics features and their corresponding coefficients is combined with constants to obtain the radiomics formula, the four kinds of Rad-score of each patient is then acquired.

Appendix 2 Feature extraction and models construction

Classical radiomics model based on prostate gland

Based on the VOI of prostate gland (PG) in each sequence image, 105 radiomics features were extracted, with a total of 315 features across three sequences. Of these, 305 features were identified as more stable through ICC analysis, 41 features were selected following Pearson correlation analysis, and 27 features were ultimately obtained through LASSO. *Figure S2A-S2C* illustrates the LASSO process and weight of remaining features. The AUC of the PCrad-score for identifying BPH and PCa in the training set, internal validation set and external validation set were 0.941 (95% CI: 0.920-0.963), 0.883 (95% CI: 0.814-0.951) and 0.865 (95% CI: 0.779-0.951), respectively.

The rad-score formula was as follows:

$$\text{PCrad-score} = 0.6369702907301562 + 0.070378 * \text{DWI_original_firstorder_Median} - 0.059828 * \text{DWI_original_firstorder_Minimum} + 0.108057 * \text{DWI_original_firstorder_TotalEnergy} + 0.071363 * \text{DWI_original_glcm_Correlation} + 0.114723 * \text{DWI_original_glcm_Idn} + 0.088840 * \text{DWI_original_glszm_SmallArea_LowGrayLevelEmphasis} + 0.034665 * \text{DWI_original_glszm_ZonePercentage} + 0.015078 * \text{DWI_original_ngtdm_Busyness} + 0.041600 * \text{DWI_original_ngtdm_Strength} - 0.152260 * \text{T1_original_firstorder_RootMeanSquared} + 0.030758 * \text{T1_original_glcm_Imc1} + 0.008431 * \text{T1_original_glszm_LowGrayLevelZoneEmphasis} + 0.087487 * \text{T1_original_glszm_SizeZoneNonUniformity} + 0.004221 * \text{T1_original_glszm_SmallAreaEmphasis} - 0.012727 * \text{T1_original_glszm_ZoneVariance} - 0.023664 * \text{T1_original_ngtdm_Contrast} + 0.047217 * \text{T1_original_ngtdm_Strength} + 0.049695 * \text{T2_original_firstorder_Kurtosis} - 0.034928 * \text{T2_original_firstorder_Skewness} + 0.072098 * \text{T2_original_glcm_Correlation} + 0.029480 * \text{T2_original_gldm_SmallDependenceLowGrayLevelEmphasis} - 0.047310 * \text{T2_original_glszm_ZoneEntropy} - 0.001576 * \text{T2_original_glszm_ZoneVariance} - 0.084681 * \text{T2_original_ngtdm_Busyness} + 0.043254 * \text{T2_original_shape_Flatness} + 0.027631 * \text{T2_original_shape_Sphericity} - 0.179976 * \text{T2_original_shape_VoxelVolume}.$$

Classical radiomics model based on lesions

Based on the VOI of lesions in each sequence image, 105 radiomics features were extracted, with a total of 315 features across three sequences. Of these, 244 features were identified as more stable through ICC analysis, 43 features were selected following Pearson correlation analysis, and 35 features were ultimately obtained through LASSO. *Figure S2D-S2F* illustrates the LASSO process and weight of remaining features. The AUC of the LCrad-score for identifying BPH and PCa in the training set, internal validation set and external validation set were 0.952 (95% CI: 0.934-0.971), 0.860 (95% CI: 0.766-0.953) and 0.854 (95% CI: 0.765-0.943), respectively.

The rad-score formula was as follows:

$$\text{LCrad-score} = 0.6441488311777703 + 0.066406 * \text{DWI_original_firstorder_10Percentile} - 0.036783 * \text{DWI_original_firstorder_Kurtosis} - 0.080558 * \text{DWI_original_firstorder_Minimum} + 0.040501 * \text{DWI_original_firstorder_Skewness} - 0.027056 * \text{DWI_original_glcm_ClusterShade} - 0.006317 * \text{DWI_original_glcm_Contrast} - 0.035593 * \text{DWI_original_glcm_Imc2} + 0.111273 * \text{DWI_original_gldm_LargeDependenceLowGrayLevelEmphasis} + 0.054521 * \text{DWI_original_glszm_ZonePercentage} - 0.033445 * \text{DWI_original_ngtdm_Complexity} + 0.131885 * \text{DWI_original_ngtdm_Strength} - 0.092696 * \text{T1_original_firstorder_RootMeanSquared} + 0.023992 * \text{T1_original_firstorder_Skewness} - 0.001385 * \text{T1_original_glcm_}$$

ClusterShade-0.035102*T1_original_glcml_Correlation+0.045272*T1_original_glcml_Idmn-0.007620*T1_original_glszm_LargeAreaEmphasis+0.017780*T1_original_glszm_LowGrayLevelZoneEmphasis-0.029562*T1_original_ngtdm_Busyness+0.008298*T1_original_ngtdm_Complexity+0.039758*T1_original_ngtdm_Contrast-0.059692*T1_original_ngtdm_Strength-0.010884*T2_original_firstorder_Minimum+0.032413*T2_original_firstorder_Skewness+0.036381*T2_original_glcml_Idmn-0.019612*T2_original_glcml_Imc2-0.026252*T2_original_glcml_InverseVariance+0.030437*T2_original_gldm_LargeDependenceHighGrayLevelEmphasis+0.154789*T2_original_gldm_RunEntropy+0.014275*T2_original_glszm_ZoneVariance-0.005360*T2_original_ngtdm_Busyness-0.067194*T2_original_ngtdm_Complexity+0.034586*T2_original_ngtdm_Strength+0.076417*T2_original_shape_Flatness-0.092250*T2_original_shape_Sphericity.

Habitat model based on lesions

Based on the VOI of each subregion in each sequence image, 105 radiomics features were extracted, with a total of 945 features across three sequences. Of these, 160 features were selected following Pearson correlation analysis, and 30 features were ultimately obtained through LASSO. *Figure S2G-S2I* illustrates the LASSO process and weight of remaining features. The AUC of the LHRad-score for identifying BPH and PCa in the training set, internal validation set and external validation set were 0.935 (95% CI: 0.913-0.957), 0.898 (95% CI: 0.841-0.955) and 0.878 (95% CI: 0.802-0.954), respectively.

The rad-score formula was as follows:

$$\begin{aligned} \text{LHRad-score} = & 0.6352845160154921 - 0.027491 * \text{DWI_original_firstorder_Skewness_h1} - 0.033663 * \text{DWI_original_} \\ & \text{glcml_ZoneVariance_h1} - 0.006145 * \text{DWI_original_ngtdm_Contrast_h1} + 0.015737 * \text{DWI_original_ngtdm_Strength_} \\ & \text{h1} - 0.037463 * \text{DWI_original_firstorder_Minimum_h2} - 0.059354 * \text{DWI_original_firstorder_RootMeanSquared_} \\ & \text{h2} + 0.003064 * \text{DWI_original_glcml_LargeAreaLowGrayLevelEmphasis_h3} - 0.006983 * \text{DWI_original_ngtdm_Strength_} \\ & \text{h3} + 0.000896 * \text{T1_original_firstorder_Kurtosis_h1} - 0.001514 * \text{T1_original_glcml_InverseVariance_h1} + 0.010685 * \text{T1_} \\ & \text{original_firstorder_Kurtosis_h2} - 0.074595 * \text{T1_original_firstorder_RootMeanSquared_h2} + 0.001627 * \text{T1_original_firstorder_} \\ & \text{Skewness_h2} + 0.003173 * \text{T1_original_glcml_Idn_h2} + 0.015271 * \text{T1_original_glcml_SmallAreaEmphasis_h2} + 0.017486 * \text{T1_} \\ & \text{original_ngtdm_Coarseness_h2} - 0.007166 * \text{T1_original_ngtdm_Complexity_h3} + 0.003291 * \text{T2_original_glcml_Imc2_} \\ & \text{h1} + 0.012030 * \text{T2_original_firstorder_Kurtosis_h2} - 0.002027 * \text{T2_original_firstorder_Minimum_h2} + 0.032984 * \text{T2_original_} \\ & \text{firstorder_RobustMeanAbsoluteDeviation_h2} + 0.007275 * \text{T2_original_firstorder_Skewness_h2} - 0.004285 * \text{T2_original_glcml_} \\ & \text{InverseVariance_h2} - 0.073183 * \text{T2_original_ngtdm_Busyness_h2} + 0.023348 * \text{T2_original_shape_Flatness_h2} - 0.237767 * \text{T2_} \\ & \text{original_shape_Sphericity_h2} + 0.000824 * \text{T2_original_firstorder_Skewness_h3} - 0.007869 * \text{T2_original_ngtdm_Complexity_} \\ & \text{h3} + 0.002066 * \text{T2_original_shape_Sphericity_h3} - 0.003685 * \text{T2_original_shape_SurfaceVolumeRatio_h3}. \end{aligned}$$

Habitat model based on PG

Based on the VOI of each habitat subregion in each sequence image, 105 radiomics features were extracted, with a total of 630 features across three sequences. Of these, 101 features were selected following Pearson correlation analysis, and 57 features were ultimately obtained through LASSO. *Figure S2J-S2L* illustrates the LASSO process and weight of remaining features. The AUC of the LHRad-score for identifying BPH and PCa in the training set, internal validation set and external validation set were 0.965 (95% CI: 0.950-0.979), 0.871 (95% CI: 0.791-0.951) and 0.773 (95% CI: 0.661-0.885), respectively.

The rad-score formula was as follows:

$$\begin{aligned} \text{PCrad_score} = & 0.6442168624847151 + 0.058497 * \text{DWI_original_glcml_ClusterShade_h1} - 0.026506 * \text{DWI_original_} \\ & \text{gldm_LargeDependenceHigh_GrayLevelEmphasis_h1} + 0.037921 * \text{DWI_original_gldm_LargeDependenceLow_} \\ & \text{GrayLevelEmphasis_h1} + 0.008339 * \text{DWI_original_glcml_SmallAreaLowGray_LevelEmphasis_h1} - 0.018552 * \text{DWI_original_} \\ & \text{glcml_ZoneEntropy_h1} + 0.030920 * \text{DWI_original_ngtdm_Contrast_h1} - 0.010342 * \text{DWI_original_firstorder_Skewness_h2} \\ & - 0.025953 * \text{DWI_original_glcml_InverseVariance_h2} - 0.042793 * \text{DWI_original_glcml_ZoneVariance_h2} - 0.063402 * \text{DWI_} \\ & \text{original_ngtdm_Busyness_h2} - 0.034716 * \text{DWI_original_ngtdm_Complexity_h2} - 0.008721 * \text{DWI_original_ngtdm_Contrast_} \\ & \text{h2} + 0.057138 * \text{DWI_original_ngtdm_Strength_h2} - 0.039679 * \text{T1_original_firstorder_Minimum_h1} + 0.025460 * \text{T1_original_} \\ & \text{firstorder_Skewness_h1} - 0.001754 * \text{T1_original_glcml_Correlation_h1} + 0.054336 * \text{T1_original_glcml_MaximumProbability_} \end{aligned}$$

$$\begin{aligned}
 &h1-0.018009*T1_original_glrlm_LongRunLow_GrayLevelEmphasis_h1+0.000103*T1_original_glszm_SmallAreaLow_ \\
 &GrayLevelEmphasis_h1+0.017317*T1_original_glszm_ZoneEntropy_h1+0.034041*T1_original_glszm_ZonePercentage_ \\
 &h1-0.035218*T1_original_glszm_ZoneVariance_h1-0.000803*T1_original_ngtdm_Contrast_h1+0.001457*T1_original_ \\
 &firstorder_Kurtosis_h2-0.047526*T1_original_firstorder_RootMean_Squared_h2+0.020848*T1_original_firstorder_ \\
 &Skewness_h2-0.002108*T1_original_glcm_Idm_h2+0.002609*T1_original_glcm_Imc2_h2-0.006617*T1_original_ \\
 &glcm_InverseVariance_h2-0.039587*T1_original_glrlm_ShortRun_Emphasis_h2+0.027699*T1_original_glrlm_ \\
 &ShortRunLowGrayLevelEmphasis_h2-0.032049*T1_original_glszm_SmallAreaEmphasis_h2-0.023046*T1_original_glszm_ \\
 &ZoneVariance_h2-0.015834*T1_original_ngtdm_Contrast_h2+0.025862*T2_original_firstorder_Kurtosis_h1+0.003799*T2_ \\
 &original_firstorder_RootMean_Squared_h1-0.031015*T2_original_firstorder_Skewness_h1+0.029199*T2_original_glcm_ \\
 &Correlation_h1+0.090207*T2_original_gldm_LargeDepen_dence_LowGrayLevelEmphasis_h1-0.040289*T2_original_ \\
 &glszm_ZoneVariance_h1-0.082215*T2_original_ngtdm_Busyness_h1+0.017568*T2_original_ngtdm_Complexity_ \\
 &h1+0.018067*T2_original_shape_Flatness_h1+0.035165*T2_original_shape_SurfaceArea_h1-0.041017*T2_original_ \\
 &firstorder_Minimum_h2-0.006014*T2_original_firstorder_Skewness_h2+0.010739*T2_original_glcm_Contrast_h2- \\
 &0.012716*T2_original_glcm_Idm_h2+0.002810*T2_original_gldm_LargeDependenceHighGrayLevelEmphasis_h2- \\
 &0.027855*T2_original_glrlm_ShortRunEmphasis_h2+0.041251*T2_original_ngtdm_Busyness_h2+0.005560*T2_original_ \\
 &ngtdm_Contrast_h2-0.016470*T2_original_ngtdm_Strength_h2-0.023913*T2_original_shape_MajorAxis_Length_h2- \\
 &0.012537*T2_original_shape_MinorAxisLength_h2-0.066864*T2_original_shape_SurfaceVolumeRatio_h2+0.045540*T2_ \\
 &original_shape_VoxelVolume_h2.
 \end{aligned}$$

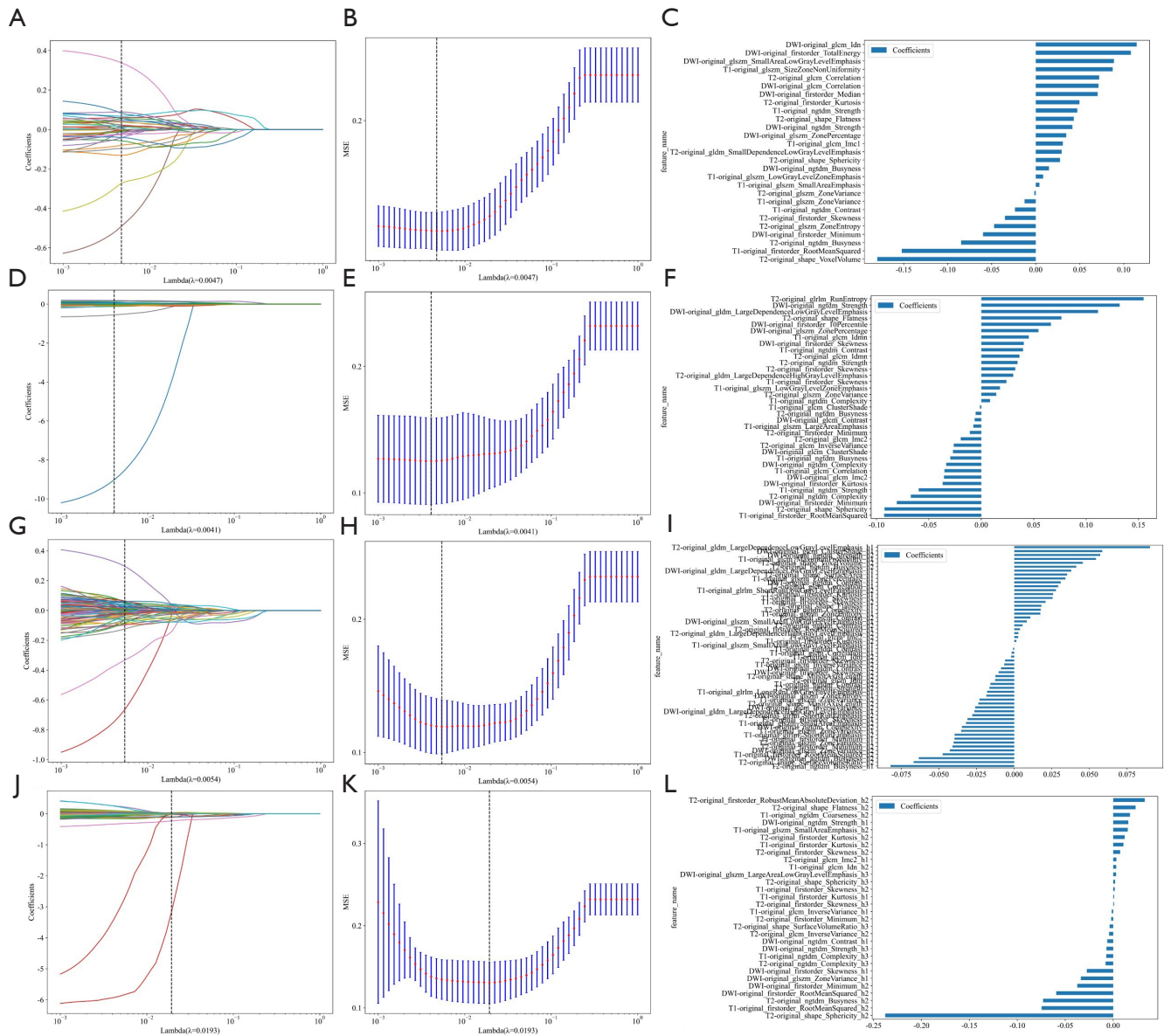


Figure S2 Feature screening process based on LASSO algorithm. (A-C) Classical radiomics features based on PG; (D-F) Classical radiomics features based on lesions; (G-I) Habitat features based on PG; (J-L) Habitat features based on lesions. From left to right is (A,D,G,J) the feature coefficient convergence process; (B,E,H,K) λ definition; and (C,F,I,L) histogram of the feature weights. PG, prostate gland.

Table S2 The contribution of each sequence in different model

Model	T1WI	T2WI	DWI
PCrad-score	0.366	0.571	0.605
PHrad-score	0.446	0.689	0.391
LCrad-score	0.362	0.599	0.625
LHrad-score	0.133	0.406	0.189

The contribution of each sequence was quantitatively illustrated using its cumulative absolute weight in the radiomics signature. LCrad-score, classical rad-score based on lesion; LHrad-score, habitat rad-score based on lesion; PCrad-score, classical rad-score based on prostate gland; PHrad-score, habitat rad-score based on prostate gland.

Appendix 3 Performance of models in non-capsular invasion subgroup

The performance of models was presented in *Table S3*. The results demonstrated that the combined model yielded the highest AUC of 0.952 (95% CI: 0.930-0.974), significantly higher than clinical model ($P<0.001$). The AUC of the combined model in the validation set exhibited a slight increase in comparison to the clinical model; however, no statistically significant difference was observed (AUC: 0.802 vs 0.764, $P=0.454$). The ROC, calibration curves, and decision curves were presented in *Figure S3*. The calibration curves demonstrated that there was no observable difference between of the combined model and the actual results (Hosmer-Lemeshow test: train set, $P=0.168$; validation set, $P=0.128$). The decision curves indicated that the clinical benefit of the combined model was superior to that of other models.

Table S3 Comparison of performance between combined model and other models

Models	AUC	95% CI	ACC	SEN	SPE	F1	P
Train set							
Clinical model	0.813	0.764-0.862	0.754	0.777	0.737	0.735	<0.001
PCrad-score	0.908	0.875-0.941	0.842	0.785	0.886	0.813	<0.001
LHrad-score	0.913	0.881-0.946	0.835	0.838	0.832	0.816	<0.001
Clinical + PCrad-score	0.922	0.892-0.951	0.862	0.777	0.928	0.831	<0.001
Clinical + LHrad-score	0.941	0.916-0.967	0.882	0.854	0.904	0.864	<0.001
Combined model	0.952	0.930-0.974	0.896	0.908	0.886	0.884	<0.001
Validation set							
Clinical model	0.764	0.676-0.852	0.730	0.721	0.734	0.626	<0.001
PCrad-score	0.762	0.67-0.854	0.766	0.651	0.819	0.636	<0.001
LHrad-score	0.751	0.651-0.85	0.752	0.698	0.777	0.638	<0.001
Clinical + PCrad-score	0.788	0.698-0.878	0.781	0.698	0.819	0.667	<0.001
Clinical + LHrad-score	0.778	0.681-0.875	0.796	0.698	0.84	0.682	<0.001
Combined model	0.802	0.710-0.895	0.825	0.674	0.894	0.707	<0.001

AUC, area under receiver operating characteristic curve; ACC, accuracy; CI, confidence interval; LHrad-score, habitat rad-score based on lesion; PCrad-score, classical rad-score based on prostate gland; SEN, sensitivity; SPE, specificity.

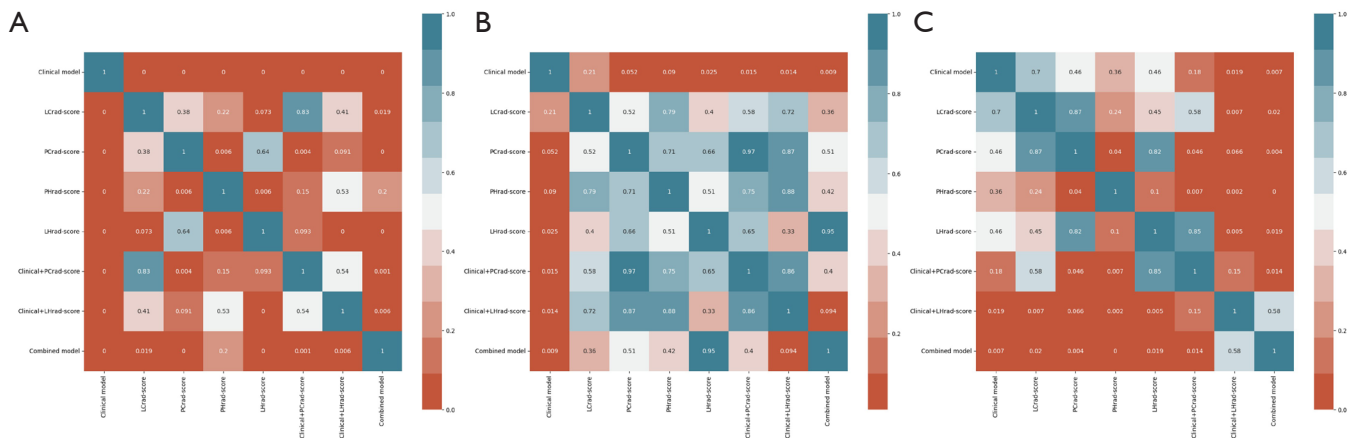


Figure S3 The results of Delong test between models in the train set (A), internal validation set (B) and external validation set (C). The closer the color is to red, the more significant the difference is between models. LCrad-score, classical rad-score based on lesion; LUradscore, habitat rad score based on lesion; PCrad-score, classical rad-score based on prostate gland; PHrad-score, habitat rad-score based on prostate gland.

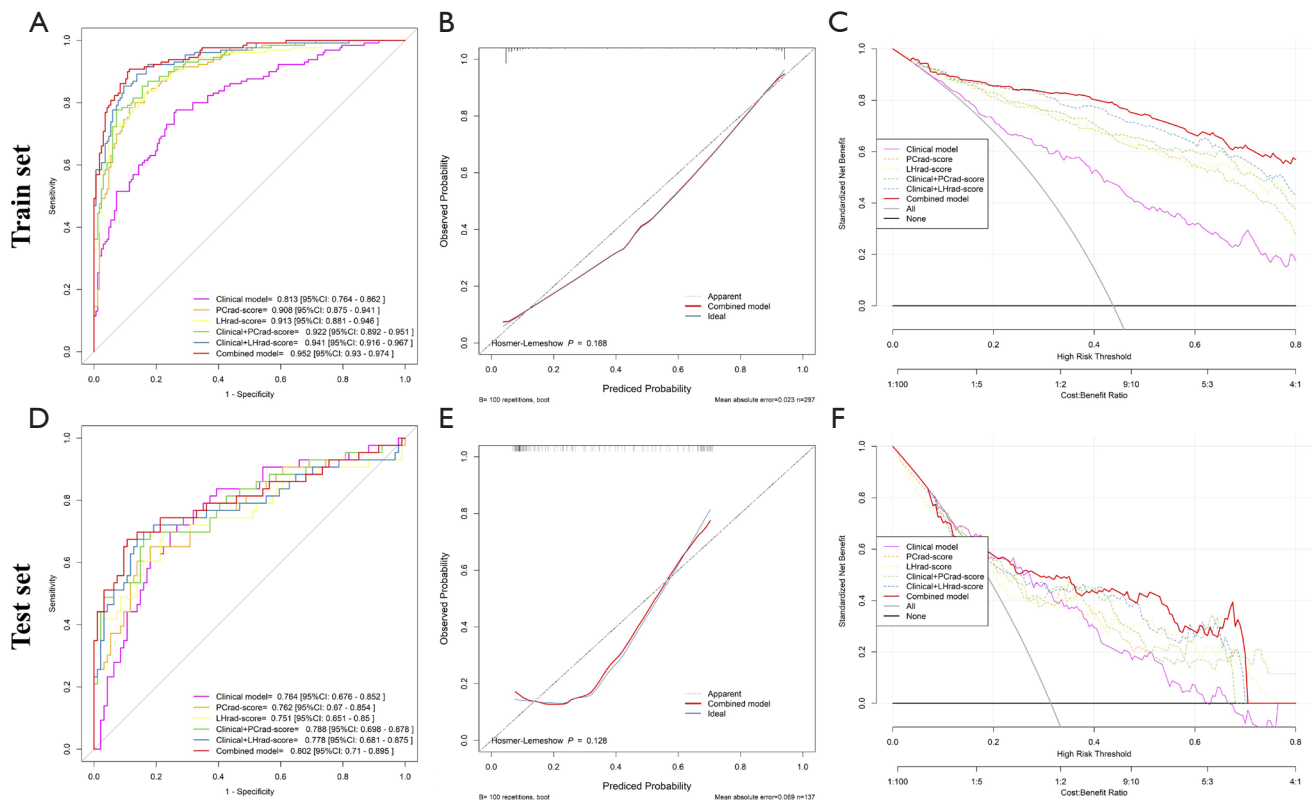


Figure S4 Performance of models in differentiating between early stage of PCa and BPH. The figures presented from left to right were (A,D) receiver operating characteristic curves, (B,E) calibration curves and (C,F) decision curves, respectively. AUC, area under receiver operating characteristic curve; BPH, benign prostatic hyperplasia; CI, confidence interval; LHrad-score, the model based on habitat rad-score based on lesion; PCrad-score, classical rad-score based on prostate gland; PCa, prostate cancer.

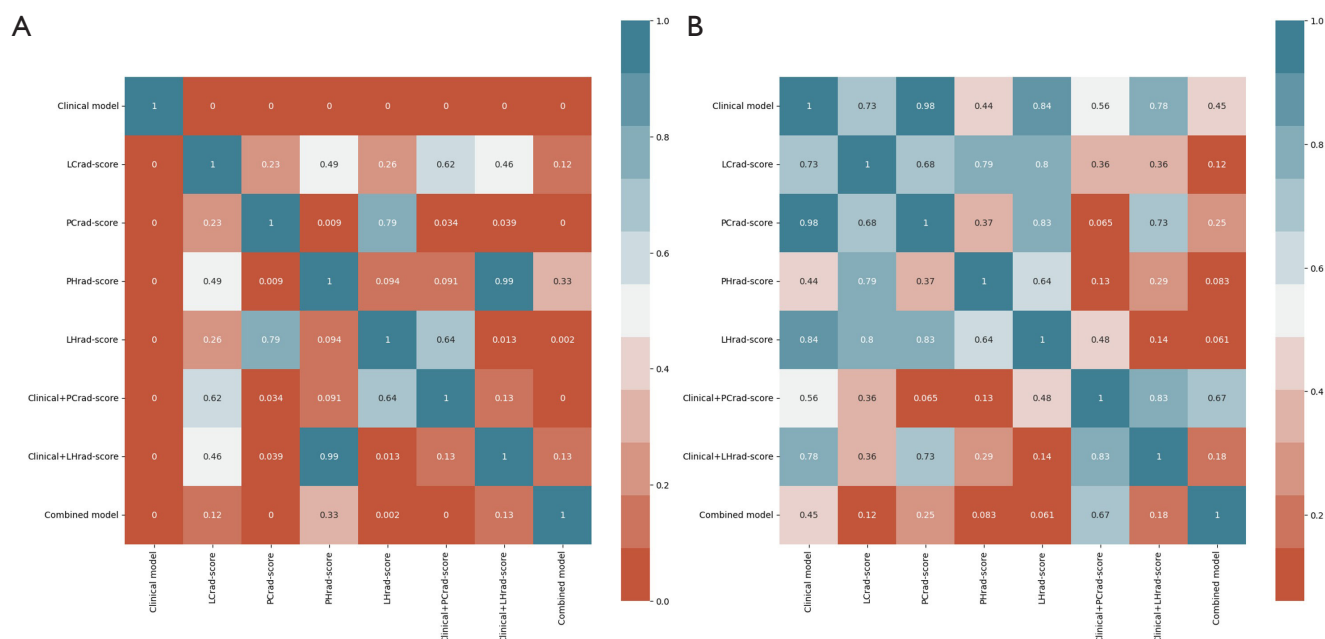


Figure S5 The results of Delong test between models in the train set (A) validation set (B). The closer the color is to red, the more significant the difference is between models. LCrad-score, classical rad-score based on lesion; LHrad-score, habitat rad score based on lesion; PCrad-score, classical rad-score based on prostate gland; PHrad-score, habitat rad-score based on prostate gland.