Am J Cancer Res 2026;16(8):3461-3475
www.ajcr.us /ISSN:2156-6976/ajcr0173936

Original Article

Clinical and MRI features for predicting maximum cancer involvement of >50% in a single biopsy core among patients diagnosed with prostate cancer at initial systematic biopsy: a machine learning study

Dejun Fan[1*], Jun Li[3*], Ruirui Niu[2*], Feiyue Zhou[1], Jun Hu[1], Li Yang[1]

[1]Department of Urology, The First Affiliated Hospital of Dali University, Dali 671000, Yunnan, China; [2]Department of Pathology, The First Affiliated Hospital of Dali University, Dali 671000, Yunnan, China; [3]Department of Urology, Ankang Central Hospital, Ankang 725000, Shaanxi, China. [*]Equal contributors.

Received May 10, 2026; Accepted July 21, 2026; Epub August 15, 2026; Published August 30, 2026

Abstract: Objective: To investigate the conditional predictive value of prebiopsy clinical variables and magnetic resonance imaging (MRI) features for maximum cancer involvement of >50% in a single biopsy core among patients diagnosed with prostate cancer at initial systematic biopsy and to compare the internally validated performance of different machine learning models. Methods: This retrospective study included patients diagnosed with prostate cancer at initial systematic biopsy at the First Affiliated Hospital of Dali University between April 2023 and February 2026. Clinical, imaging, and combined models were developed to predict maximum cancer involvement of >50% in a single biopsy core using age, body mass index, total prostate-specific antigen (tPSA), prostate volume, Prostate Imaging Reporting and Data System (PI-RADS) score, and MRI features related to tumor extent. Logistic regression, elastic net, random forest, and extreme gradient boosting (XGBoost) were compared using the combined predictors. Discrimination, calibration, and overall prediction error were evaluated using five repeats of stratified five-fold cross-validation and bootstrap resampling. Results: Among 127 patients, 85 (66.9%) had maximum cancer involvement of >50% in a single biopsy core. The areas under the receiver operating characteristic curve (AUROCs) of the clinical, imaging, and combined models were 0.822, 0.773, and 0.823, respectively. The combined model showed no clear incremental discrimination over the clinical model but had a lower overall prediction error than the imaging model. Among the four algorithms using the combined predictors, elastic net achieved the highest AUROC (0.823), whereas XGBoost achieved the highest average precision (0.878), the lowest Brier score (0.151), and favorable calibration. tPSA and PI-RADS score showed relatively consistent predictive contributions across models. Sensitivity analyses yielded generally similar AUROCs, although Brier scores and calibration slopes varied. Conclusion: Prebiopsy clinical variables and MRI features enabled risk stratification for maximum cancer involvement of >50% in a single biopsy core among patients diagnosed with prostate cancer at initial systematic biopsy. Elastic net and XGBoost demonstrated favorable overall predictive performance, while tPSA and PI-RADS score showed relatively consistent contributions. External validation in independent cohorts is required.

Keywords: Prostate cancer, systematic biopsy, maximum cancer involvement in a single biopsy core, magnetic resonance imaging, machine learning, prediction model

Introduction

Prostate cancer is characterized by substantial molecular, morphological, and clinical heterogeneity. Even among patients with similar prostate-specific antigen (PSA) levels or pathological grades, tumor extent and biological behavior may differ considerably [1]. Prostate biopsy not only establishes the diagnosis and Gleason grade but also quantifies tumor burden through measures such as the number and proportion of positive cores and the extent of cancer involvement in individual biopsy cores. The extent of cancer involvement in biopsy cores is an important quantitative component of prostate biopsy pathology reporting and risk assessment [2].

The maximum percentage of cancer involvement in a single biopsy core reflects the great-

https://doi.org/10.62347/ROGZ7613

est extent of tumor involvement within locally sampled tissue and is not equivalent to the histological aggressiveness represented by pathological grade. Previous active surveillance studies have commonly used cancer involvement of no more than 50% in any individual biopsy core as a reference criterion for low tumor burden, whereas maximum cancer involvement exceeding 50% generally indicates greater tumor extent and may affect the assessment of pathological progression or risk reclassification [3]. Therefore, evaluating the ability of prebiopsy clinical variables and magnetic resonance imaging (MRI) features to predict a maximum cancer involvement of >50% in a single biopsy core may help identify patients with potentially higher tumor burden on biopsy pathology.

Total PSA (tPSA), prostate volume, and MRI are important sources of information in the prebiopsy evaluation of prostate cancer. The Prostate Imaging Reporting and Data System (PI-RADS) standardizes the acquisition, interpretation, and reporting of prostate MRI and stratifies the likelihood of clinically significant prostate cancer [4]; MRI findings such as extraprostatic extension and seminal vesicle invasion may also provide information on local tumor extent and stage [5]. However, individual clinical or imaging variables may not adequately capture the heterogeneity of prostate cancer. Integrating multidimensional clinical variables with MRI features may therefore improve risk assessment.

In recent years, machine learning has been used to integrate tPSA, prostate volume, PI-RADS score, and other MRI features to predict prostate biopsy outcomes or clinically significant prostate cancer, demonstrating potential value for risk stratification [6-8]. However, previous studies have primarily focused on the detection of prostate cancer or clinically significant prostate cancer, whereas relatively few have investigated the prebiopsy prediction of the maximum extent of cancer involvement in a single biopsy core. Therefore, this study included patients who underwent initial systematic biopsy and were diagnosed with prostate cancer on that biopsy. Prebiopsy clinical variables and MRI features were integrated to develop clinical, imaging, and combined models. The conditional predictive performance of logistic regression, elastic net, random forest, and extreme gradient boosting (XGBoost) for a maximum cancer involvement of >50% in a single biopsy core was compared, and model calibration, stability, and interpretability were evaluated to determine the potential value of prebiopsy information for stratifying tumor burden on biopsy pathology.

Materials and methods

Study design and data source

This single-center retrospective study included patients who underwent initial systematic prostate biopsy at the First Affiliated Hospital of Dali University between April 2023 and February 2026 and were diagnosed with prostate cancer on that biopsy. All candidate predictors were obtained from prebiopsy clinical and MRI examinations, whereas the study outcome was determined from the pathological findings of the index systematic biopsy. The study was restricted to patients diagnosed with prostate cancer at initial systematic biopsy, and the models were developed to evaluate the conditional predictive value of prebiopsy information for a maximum percentage of cancer involvement of >50% in a single biopsy core in this population.

The inclusion criteria were as follows: (1) initial prostate biopsy with a pathological diagnosis of prostate cancer on that biopsy; (2) transrectal or transperineal systematic biopsy without MRI-targeted biopsy; (3) completion of prostate MRI within 15 days before biopsy; (4) no endocrine therapy, radiotherapy, chemotherapy, or other prostate cancer-related treatment before biopsy; and (5) complete clinical, MRI, and biopsy pathology data.

The exclusion criteria were as follows: (1) a history of previous prostate biopsy; (2) inability to accurately determine the number of positive biopsy cores, total number of biopsy cores, or maximum percentage of cancer involvement in a single biopsy core; and (3) missing key clinical, MRI, or pathological data that could not be verified.

The study was approved by the Ethics Committee of the First Affiliated Hospital of Dali University, which waived the requirement for informed consent (approval No. DFY2026-0205001).

Collection of clinical, MRI, and pathological data

Prebiopsy data included age, body mass index (BMI), tPSA, free prostate-specific antigen (fPSA), three orthogonal prostate diameters, PI-RADS score, MRI findings of extraprostatic extension and seminal vesicle invasion, lesion location, bilateral lesions, and multifocal lesions. Biopsy route and total number of cores were also recorded. Pathological data included the number of positive cores, highest Gleason score, International Society of Urological Pathology (ISUP) grade group, and maximum cancer involvement in a single core. Prostate volume was calculated as anteroposterior diameter × transverse diameter × longitudinal diameter × 0.52.

All patients underwent MRI using a 3.0-T Toshiba system. Images were independently assessed by two radiologists, each with more than 10 years of experience in prostate MRI interpretation, who were blinded to the biopsy findings and the maximum cancer involvement in a single core. The highest PI-RADS score across all lesions was recorded, and other MRI findings were determined by consensus. Extraprostatic extension and seminal vesicle invasion were assessed and coded independently. When seminal vesicle invasion was present without definite extraprostatic extension, the former was recorded as present and the latter as absent.

Outcome definition, candidate predictors, and feature processing

The primary outcome was maximum cancer involvement of >50% in a single biopsy core among patients diagnosed with prostate cancer on the index systematic biopsy. Each core was examined separately, and cancer involvement was calculated as the percentage of evaluable tissue occupied by cancer. Fragmented cores and cores with discontinuous cancer involvement were measured according to standardized institutional procedures. The highest value across all positive cores was recorded; patients with exactly 50% involvement were assigned to the ≤50% group.

Prespecified predictors included age, BMI, tPSA, prostate volume, PI-RADS score, MRI signs of local tumor extension, and extensive lesion involvement. Local tumor extension was defined as extraprostatic extension or seminal vesicle invasion, whereas extensive lesion involvement was defined as bilateral or multifocal lesions. No predictor selection based on univariable *P* values or stepwise regression was performed. For tPSA values reported as ">100 ng/mL", the models included both log[min(tPSA, 100)] and an indicator for tPSA >100 ng/mL. Prostate volume was natural log-transformed. Continuous predictors in the logistic regression and elastic net models were standardized within the corresponding training sets, and PI-RADS score was modeled as an ordinal variable.

Because fPSA values were also subject to an upper reporting limit, fPSA was used only in the descriptive analyses. MRI lesion location was excluded from the primary models because of multiple categories, small subgroups, and inconsistent reporting. Biopsy route and total number of cores were examined only in sensitivity analyses. Postbiopsy pathological variables were used descriptively and excluded from the prediction models to prevent information leakage.

Model development

Elastic net logistic regression was used to develop the clinical, imaging, and combined models. The clinical model included age, BMI, log-transformed capped tPSA, an indicator for tPSA >100 ng/mL, and log-transformed prostate volume. The imaging model included PI-RADS score, MRI signs of local tumor extension, and extensive lesion involvement. The combined model included all clinical and MRI predictors. Its performance was compared with that of the clinical and imaging models to assess the incremental predictive value of combining both sources of information.

Using the same combined predictors, multivariable logistic regression, elastic net logistic regression, random forest, and XGBoost models were developed and compared. Logistic regression served as the reference model, elastic net provided coefficient shrinkage, and random forest and XGBoost explored potential nonlinear relationships and interactions.

Model validation, hyperparameter selection, and performance evaluation

Stratified five-fold cross-validation repeated five times was used for model development and internal validation, with identical data splits across models. Elastic net and random forest hyperparameters were selected by stratified three-fold inner cross-validation within each outer training set. All preprocessing, model fitting, and hyperparameter selection were confined to the training data, and out-of-fold predictions were generated for the corresponding validation folds. Given the limited sample size, XGBoost used prespecified shallow trees and strong regularization without extensive hyperparameter tuning. For each patient, predictions from the five repeats were averaged to obtain a patient-level mean out-of-fold probability.

For the final models fitted to the full cohort, elastic net used C=0.5 and l1_ratio=1, whereas random forest used 250 trees, a maximum depth of 2, a minimum of 5 samples per leaf, and max_features set to "sqrt". The XGBoost parameters were 80 boosting rounds, a learning rate of 0.05, a maximum depth of 2, a minimum child weight of 5, row and column subsampling ratios of 0.8, an L2 regularization coefficient of 10, and an L1 regularization coefficient of 0.5.

Discrimination was assessed using the area under the receiver operating characteristic curve (AUROC) and average precision (AP), overall prediction error using the Brier score, and calibration using calibration curves, intercepts, and slopes. Based on patient-level mean out-of-fold probabilities, 2,000 patient-level bootstrap resamples were used to estimate AUROC, AP, and Brier scores with 95% confidence intervals (CIs), and 2,000 paired bootstrap resamples were used to estimate between-model differences and their 95% CIs. Comparisons included the combined model versus the clinical and imaging models and each machine learning model versus logistic regression. Performance was evaluated by jointly considering discrimination, calibration, and overall prediction error.

Model interpretability, stability, and sensitivity analyses

The stability of the elastic net regression coefficients was evaluated using 500 patient-level bootstrap resamples. The median standardized regression coefficients, 95% empirical intervals, and directional consistency rates were reported. Permutation importance was calculated in the outer validation folds to evaluate the predictive contribution of each variable in the random forest model. SHapley Additive exPlanations (SHAP) were used to interpret predictions from the final XGBoost model fitted to the full cohort. These analyses were used solely to interpret the contributions of individual variables to model predictions and were not intended to support causal inference.

Sensitivity analyses comprised: (1) redefining the outcome as maximum cancer involvement of ≥50% or ≥70% in a single core; (2) excluding patients with tPSA >100 ng/mL; (3) restricting the analysis to transrectal systematic biopsy; and (4) adding biopsy route and total number of cores to the combined model. All analyses used the same elastic net modeling and cross-validation procedures as the primary analysis.

Statistical analysis

Continuous variables were summarized as medians and interquartile ranges (IQRs) and compared using the Mann-Whitney U test. Categorical variables were summarized as counts and percentages. Binary variables were compared using the two-sided Fisher exact test, and multicategory variables using a Monte Carlo exact test with fixed margins and the Pearson χ^2 statistic, based on 1,000,000 random samples. Baseline comparisons were descriptive and were not used for predictor selection. All tests were two-sided, with $P<0.05$ indicating statistical significance.

Analyses were performed using Python 3.13.5. Logistic regression, elastic net, random forest, and cross-validation were implemented primarily using scikit-learn 1.8.0; XGBoost models used XGBoost 3.1.3, and model interpretation used SHAP 0.50.0. NumPy 2.3.5, pandas 2.2.3, SciPy 1.17.0, statsmodels 0.14.6, and Matplotlib 3.10.8 were used for data processing, statistical analysis, and visualization.

Results

Patient characteristics

Patients who underwent initial systematic biopsy during the study period were consecutively

screened. After excluding those who did not meet the eligibility criteria or had incomplete key data, 127 patients diagnosed with prostate cancer on the index biopsy were included. Of these, 85 (66.9%) had a maximum percentage of cancer involvement of >50% in a single biopsy core, whereas 42 (33.1%) had a value of ≤50%. The median age was 72.0 years (IQR, 68.0-76.0 years). The median capped tPSA and fPSA levels were 36.8 ng/mL (IQR, 14.0-100.0 ng/mL) and 3.6 ng/mL (IQR, 1.8-12.6 ng/mL), respectively. Overall, 36 patients (28.3%) had tPSA >100 ng/mL, and 16 (12.6%) had fPSA >50 ng/mL.

Compared with the ≤50% group, the >50% group had higher tPSA and fPSA levels and higher proportions of patients with values above the corresponding upper reporting limits (all P<0.05). Age, BMI, and prostate volume did not differ significantly between the groups. Regarding MRI characteristics, the overall distribution of PI-RADS scores differed between the groups (P<0.001), with a higher proportion of PI-RADS 5 scores in the >50% group. Extraprostatic extension, seminal vesicle invasion, bilateral lesions, and multifocal lesions were also more frequent in this group (all P<0.05). The overall distribution of MRI lesion location differed between the groups (P=0.009), with concurrent peripheral-zone and transition-zone involvement being more frequent in the >50% group.

Transrectal biopsy was performed in 99 patients (78.0%), whereas transperineal biopsy was performed in 28 (22.0%). Neither biopsy route nor the total number of biopsy cores differed significantly between the groups. Regarding pathological characteristics, the number and proportion of positive biopsy cores were markedly higher in the >50% group. The overall distributions of the highest Gleason score and ISUP grade group also differed between the groups (both P<0.001), with the >50% group showing a shift toward higher pathological grades. The proportions of ISUP grade group 5 were 64.7% in the >50% group and 11.9% in the ≤50% group. Detailed baseline characteristics are presented in **Table 1**. The distribution of the maximum percentage of cancer involvement in a single biopsy core and the outcome composition across PI-RADS strata are shown in **Figure 1**.

Performance comparison among the clinical, imaging, and combined models

The AUROCs of the clinical, imaging, and combined models were 0.822 (95% CI, 0.730-0.905), 0.773 (95% CI, 0.681-0.851), and 0.823 (95% CI, 0.732-0.901), respectively. The corresponding AP values were 0.846, 0.875, and 0.867, and the corresponding Brier scores were 0.157, 0.179, and 0.156. The combined model had the numerically highest AUROC and the lowest Brier score, with a calibration intercept of -0.314 and a calibration slope of 1.580. Detailed performance estimates are presented in **Table 2**, and the receiver operating characteristic (ROC) and calibration curves are shown in **Figure 2A** and **2B**.

Paired bootstrap analysis showed that, compared with the clinical model, the combined model had an increase of 0.001 in AUROC (95% CI, -0.027 to 0.031), an increase of 0.021 in AP (95% CI, -0.045 to 0.079), and a decrease of 0.002 in Brier score (95% CI, -0.008 to 0.005); none of these differences were statistically significant. Compared with the imaging model, the combined model had an increase of 0.050 in AUROC (95% CI, -0.016 to 0.120) and a decrease of 0.008 in AP (95% CI, -0.048 to 0.033), neither of which was statistically significant. The Brier score decreased by 0.024 (95% CI, -0.045 to -0.002), with the 95% CI excluding 0, indicating that adding clinical variables reduced the overall prediction error of the imaging model. Between-model differences in performance are presented in **Table 2** and **Figure 2C**.

Performance comparison of four algorithms using the combined predictors

Using the same combined set of predictors, the AUROCs of logistic regression, elastic net, random forest, and XGBoost were 0.800 (95% CI, 0.694-0.888), 0.823 (95% CI, 0.732-0.901), 0.815 (95% CI, 0.723-0.897), and 0.822 (95% CI, 0.733-0.897), respectively. Elastic net achieved the numerically highest AUROC. XGBoost achieved the highest AP of 0.878 (95% CI, 0.790-0.948) and the lowest Brier score of 0.151 (95% CI, 0.117-0.190). Its calibration intercept and slope were 0.029 and 1.042, respectively, which were numerically closest to the ideal values. Detailed performance esti-

Table 1. Baseline clinical, MRI, procedural, and biopsy pathological characteristics according to the maximum percentage of cancer involvement in a single biopsy core

Variable	Overall (n=127)	Maximum involvement ≤50% (n=42)	Maximum involvement >50% (n=85)	*P* value
Pre-biopsy clinical characteristics				
Age, years	72.0 [68.0-76.0]	71.0 [67.0-74.0]	73.0 [68.0-77.0]	0.086
Capped tPSA, ng/mL	36.8 [14.0-100.0]	11.5 [7.3-22.0]	61.2 [26.7-100.0]	<0.001
tPSA >100 ng/mL	36 (28.3)	3 (7.1)	33 (38.8)	<0.001
Capped fPSA, ng/mL	3.6 [1.8-12.6]	1.8 [1.0-2.6]	8.2 [2.8-20.2]	<0.001
fPSA >50 ng/mL	16 (12.6)	1 (2.4)	15 (17.6)	0.02
BMI, kg/m^2	21.9 [20.1-24.4]	21.6 [20.6-24.0]	22.4 [20.0-24.8]	0.65
Prostate volume, mL	47.3 [32.3-64.0]	47.3 [30.5-64.8]	47.3 [32.5-61.3]	0.794
MRI characteristics				
PI-RADS category				<0.001
1	3 (2.4)	2 (4.8)	1 (1.2)	
2	11 (8.7)	8 (19.0)	3 (3.5)	
3	22 (17.3)	11 (26.2)	11 (12.9)	
4	30 (23.6)	12 (28.6)	18 (21.2)	
5	61 (48.0)	9 (21.4)	52 (61.2)	
MRI extraprostatic extension	49 (38.6)	6 (14.3)	43 (50.6)	<0.001
MRI seminal vesicle invasion	26 (20.5)	1 (2.4)	25 (29.4)	<0.001
MRI lesion location				0.009
No MRI-visible lesion	6 (4.7)	4 (9.5)	2 (2.4)	
Peripheral zone only	58 (45.7)	24 (57.1)	34 (40.0)	
Transition zone only	16 (12.6)	7 (16.7)	9 (10.6)	
Peripheral and transition zones	33 (26.0)	4 (9.5)	29 (34.1)	
Central-zone involvement	14 (11.0)	3 (7.1)	11 (12.9)	
Bilateral lesions	48 (37.8)	9 (21.4)	39 (45.9)	0.011
Multifocal lesions	62 (48.8)	13 (31.0)	49 (57.6)	0.005
Biopsy procedural characteristics				
Biopsy route				1.000
Transrectal	99 (78.0)	33 (78.6)	66 (77.6)	
Transperineal	28 (22.0)	9 (21.4)	19 (22.4)	
Total biopsy cores	13 [12-14]	13 [12-16]	13 [12-14]	0.106
Biopsy pathological characteristics				
Positive biopsy cores	9 [4-12]	2 [1-5]	10 [8-12]	<0.001
Positive-core rate, %	58.3 [26.8-92.0]	17.0 [9.0-33.3]	83.3 [55.6-100.0]	<0.001
Highest biopsy Gleason score				<0.001
6 (3+3)	23 (18.1)	22 (52.4)	1 (1.2)	
7 (3+4)	8 (6.3)	4 (9.5)	4 (4.7)	
7 (4+3)	6 (4.7)	2 (4.8)	4 (4.7)	
8	30 (23.6)	9 (21.4)	21 (24.7)	
9	29 (22.8)	4 (9.5)	25 (29.4)	
10	31 (24.4)	1 (2.4)	30 (35.3)	
ISUP grade group				<0.001
1	23 (18.1)	22 (52.4)	1 (1.2)	
2	8 (6.3)	4 (9.5)	4 (4.7)	
3	6 (4.7)	2 (4.8)	4 (4.7)	
4	30 (23.6)	9 (21.4)	21 (24.7)	
5	60 (47.2)	5 (11.9)	55 (64.7)	

Note: Data are presented as median [interquartile range] or number (percentage). Continuous variables were compared between groups using the Mann-Whitney U test. Binary variables were compared using the two-sided Fisher exact test, whereas multicategory variables were compared using a Monte Carlo exact test with fixed row and column margins and the Pearson χ^2 statistic as the test statistic (1,000,000 random samples). tPSA >100 ng/mL and fPSA >50 ng/mL were considered to exceed the upper reporting limits and were summarized as 100 and 50 ng/mL, respectively, in the descriptive analyses. Prostate volume was calculated using the ellipsoid formula: anteroposterior diameter × transverse diameter × longitudinal diameter × 0.52. Biopsy pathological characteristics were used for descriptive purposes only and were not included in the prebiopsy prediction models. BMI, body mass index; fPSA, free prostate-specific antigen; ISUP, International Society of Urological Pathology; MRI, magnetic resonance imaging; PI-RADS, Prostate Imaging Reporting and Data System; tPSA, total prostate-specific antigen.

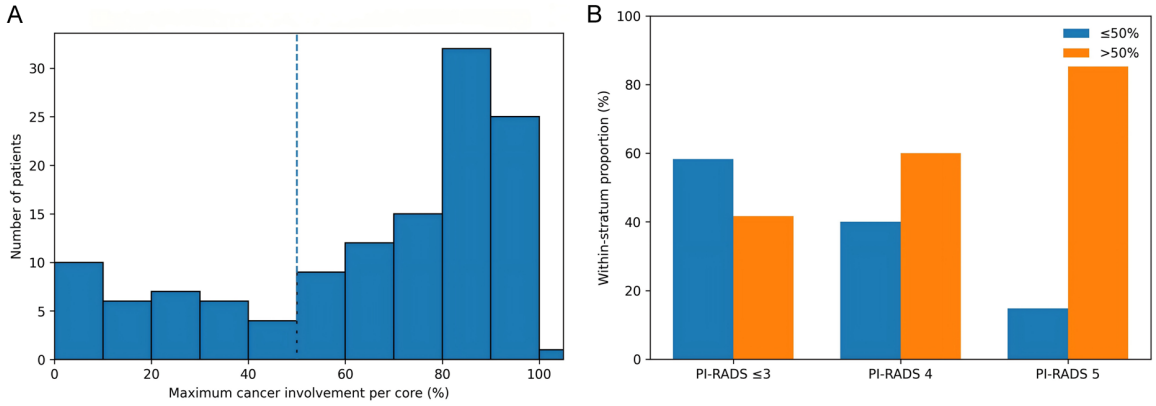

Figure 1. Distribution of the maximum percentage of cancer involvement in a single biopsy core and stratification by PI-RADS scores of ≤3, 4, and 5. Note: A. Distribution of the maximum percentage of cancer involvement in a single biopsy core; the dashed line indicates the primary outcome threshold of 50%. B. Proportions of patients in the ≤50% and >50% groups within the PI-RADS ≤3, 4, and 5 strata. PI-RADS, Prostate Imaging Reporting and Data System.

Table 2. Predictive performance of the clinical, imaging, and combined models and between-model comparisons

Item	AUROC or ΔAUROC (95% CI)	AP or ΔAP (95% CI)	Brier score or ΔBrier score (95% CI)	Calibration intercept	Calibration slope
Model performance					
Clinical model	0.822 (0.730-0.905)	0.846 (0.750-0.944)	0.157 (0.129-0.187)	-0.350	1.637
Imaging model	0.773 (0.681-0.851)	0.875 (0.796-0.937)	0.179 (0.148-0.213)	-0.096	1.185
Combined model	0.823 (0.732-0.901)	0.867 (0.777-0.947)	0.156 (0.127-0.188)	-0.314	1.580
Between-model comparisons					
Combined model vs. clinical model	0.001 (-0.027 to 0.031)	0.021 (-0.045 to 0.079)	-0.002 (-0.008 to 0.005)	-	-
Combined model vs. imaging model	0.050 (-0.016 to 0.120)	-0.008 (-0.048 to 0.033)	-0.024 (-0.045 to -0.002)	-	-

Note: Model performance was calculated using patient-level mean out-of-fold predicted probabilities obtained from five repeats of stratified five-fold cross-validation. The 95% CIs for AUROC, AP, and Brier score were estimated using 2,000 patient-level bootstrap resamples. Between-model differences and their 95% CIs were estimated using 2,000 paired patient-level bootstrap resamples. Between-model differences were defined as the combined model minus the corresponding reference model. Positive values of ΔAUROC and ΔAP indicate better performance of the combined model, whereas a negative value of ΔBrier indicates a lower overall prediction error for the combined model. All differences were calculated using the original unrounded estimates. The ideal values of the calibration intercept and calibration slope are 0 and 1, respectively. AP, average precision; AUROC, area under the receiver operating characteristic curve; CI, confidence interval.

mates are presented in **Table 3**, and the ROC, precision-recall (PR), and calibration curves are shown in **Figure 3A-C**.

Paired bootstrap analysis showed that, compared with logistic regression, elastic net had increases of 0.023 in AUROC (95% CI, 0.002-0.048) and 0.041 in AP (95% CI, 0.004-0.074), whereas XGBoost had an increase of 0.052 in AP (95% CI, 0.007-0.089). The AUROCs of random forest and XGBoost did not differ significantly from that of logistic regression, and the AP of random forest also did not differ significantly from that of logistic regression. None of the machine learning models showed a statistically significant difference in Brier score compared with logistic regression. Overall, elastic net performed favorably in terms of discrimination, whereas XGBoost showed favorable precision-recall performance, overall prediction error, and calibration. However, no single model was superior across all performance measures. Between-model differences in performance are presented in **Table 3**, and the distributions of out-of-fold predicted probabilities stratified by observed outcome are shown in **Figure 3D**.

Model interpretability and stability

Bootstrap analysis of the elastic net model showed that log-transformed capped tPSA had the largest median standardized regression coefficient (0.838; 95% empirical interval, 0.490-1.271). The median standardized coefficient for PI-RADS score was 0.128, whereas the median coefficients for most other predictors were close to or equal to 0, suggesting

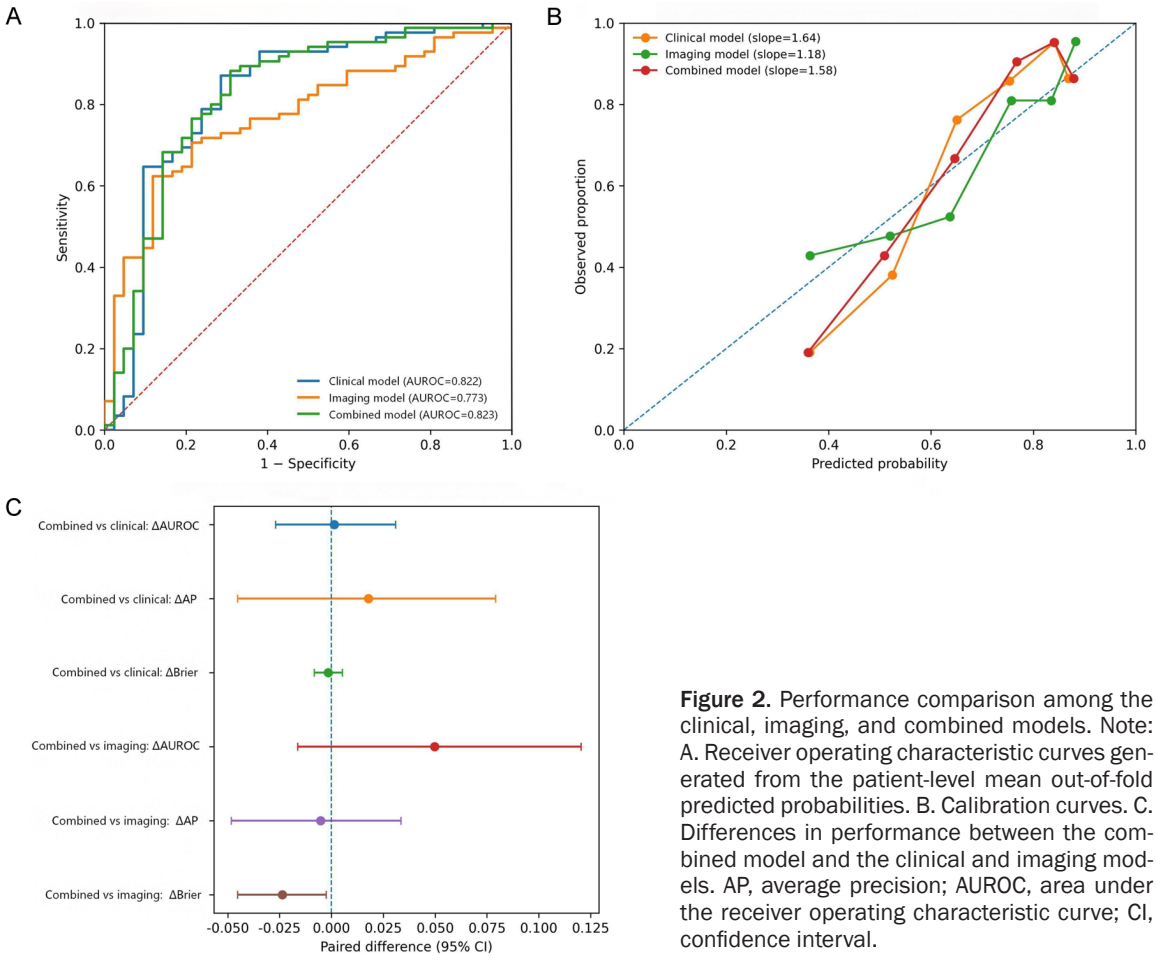

Figure 2. Performance comparison among the clinical, imaging, and combined models. Note: A. Receiver operating characteristic curves generated from the patient-level mean out-of-fold predicted probabilities. B. Calibration curves. C. Differences in performance between the combined model and the clinical and imaging models. AP, average precision; AUROC, area under the receiver operating characteristic curve; CI, confidence interval.

that their predictive contributions were relatively limited under strong penalization (**Figure 4A**). Permutation importance analysis of the random forest model identified log-transformed capped tPSA, PI-RADS score, MRI signs of local tumor extension, and the indicator variable for tPSA >100 ng/mL as relatively important predictors (**Figure 4B**). SHAP analysis identified log-transformed capped tPSA, PI-RADS score, log-transformed prostate volume, and BMI as the main contributing predictors (**Figure 4C**). The dependence plot further illustrated the relationship between log-transformed capped tPSA and its SHAP value (**Figure 4D**).

Sensitivity analyses

The primary analysis of the combined elastic net model served as the reference and yielded an AUROC of 0.823 (95% CI, 0.732-0.901). When the outcome was redefined as a maximum percentage of cancer involvement of ≥50% or ≥70% in a single biopsy core, the AUROCs were 0.818 (95% CI, 0.715-0.904) and 0.829 (95% CI, 0.749-0.900), respectively. After excluding patients with tPSA >100 ng/mL, restricting the analysis to patients who underwent transrectal systematic biopsy, or additionally including biopsy route and total number of biopsy cores in the combined model, the AUROCs were 0.803 (95% CI, 0.700-0.894), 0.836 (95% CI, 0.734-0.923), and 0.818 (95% CI, 0.718-0.903), respectively. Discrimination in the sensitivity analyses was generally similar to that in the primary analysis, although the Brier scores and calibration slopes showed some variation. These findings suggest that model discrimination was relatively robust, whereas calibration may have been affected by the study population and outcome definition. Detailed results are presented in **Table 4**.

Discussion

This study included patients who underwent initial systematic biopsy and were diagnosed with

Table 3. Predictive performance of different algorithms and comparisons with logistic regression

Item	AUROC or ΔAUROC (95% CI)	AP or ΔAP (95% CI)	Brier score or ΔBrier score (95% CI)	Calibration intercept	Calibration slope
Model performance					
Logistic regression	0.800 (0.694-0.888)	0.826 (0.728-0.935)	0.154 (0.112-0.200)	0.157	0.745
Elastic net	0.823 (0.732-0.901)	0.867 (0.777-0.947)	0.156 (0.127-0.188)	-0.314	1.580
Random forest	0.815 (0.723-0.897)	0.865 (0.770-0.946)	0.167 (0.136-0.199)	0.484	1.238
XGBoost	0.822 (0.733-0.897)	0.878 (0.790-0.948)	0.151 (0.117-0.190)	0.029	1.042
Between-model comparisons					
Elastic net vs. logistic regression	0.023 (0.002-0.048)	0.041 (0.004-0.074)	0.002 (-0.018 to 0.019)	-	-
Random forest vs. logistic regression	0.015 (-0.021 to 0.054)	0.039 (-0.013 to 0.083)	0.013 (-0.008 to 0.034)	-	-
XGBoost vs. logistic regression	0.022 (-0.010 to 0.053)	0.052 (0.007-0.089)	-0.003 (-0.015 to 0.010)	-	-

Note: All algorithms used the same combined set of predictors, and performance was evaluated using patient-level mean out-of-fold predicted probabilities obtained from five repeats of stratified five-fold cross-validation. The 95% CIs for AUROC, AP, and Brier score were estimated using 2,000 patient-level bootstrap resamples. Between-model differences and their 95% CIs were estimated using 2,000 paired patient-level bootstrap resamples. Between-model differences were defined as the corresponding machine learning model minus the logistic regression model. Positive values of ΔAUROC and ΔAP indicate better performance of the corresponding machine learning model, whereas a negative value of ΔBrier indicates a lower overall prediction error. All differences were calculated using the original unrounded estimates. The ideal values of the calibration intercept and calibration slope are 0 and 1, respectively. AP, average precision; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; XGBoost, extreme gradient boosting.

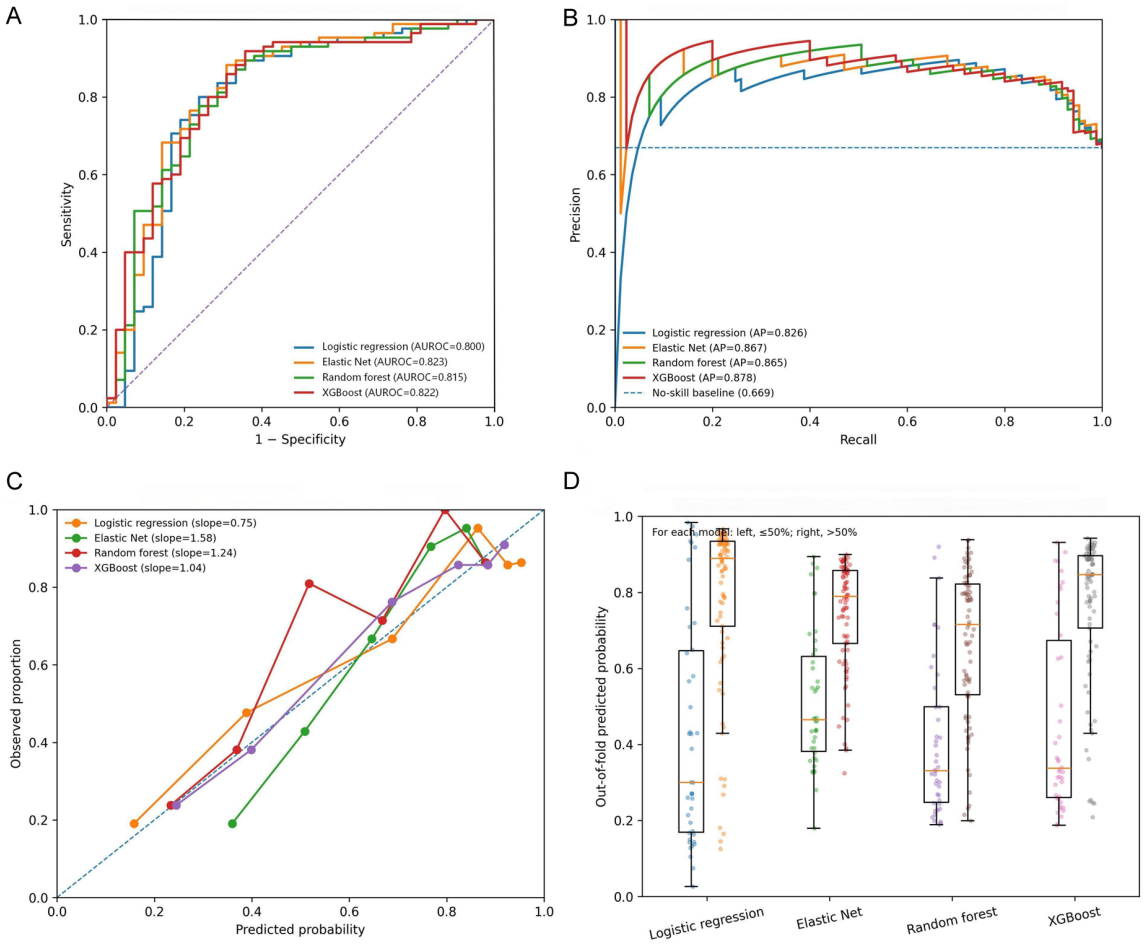

Figure 3. Performance comparison of four algorithms using the combined set of predictors. Note: A. Receiver operating characteristic curves. B. Precision-recall curves. C. Calibration curves. D. Distributions of patient-level mean out-of-fold predicted probabilities according to the observed outcome. AP, average precision; AUROC, area under the receiver operating characteristic curve; XGBoost, extreme gradient boosting.

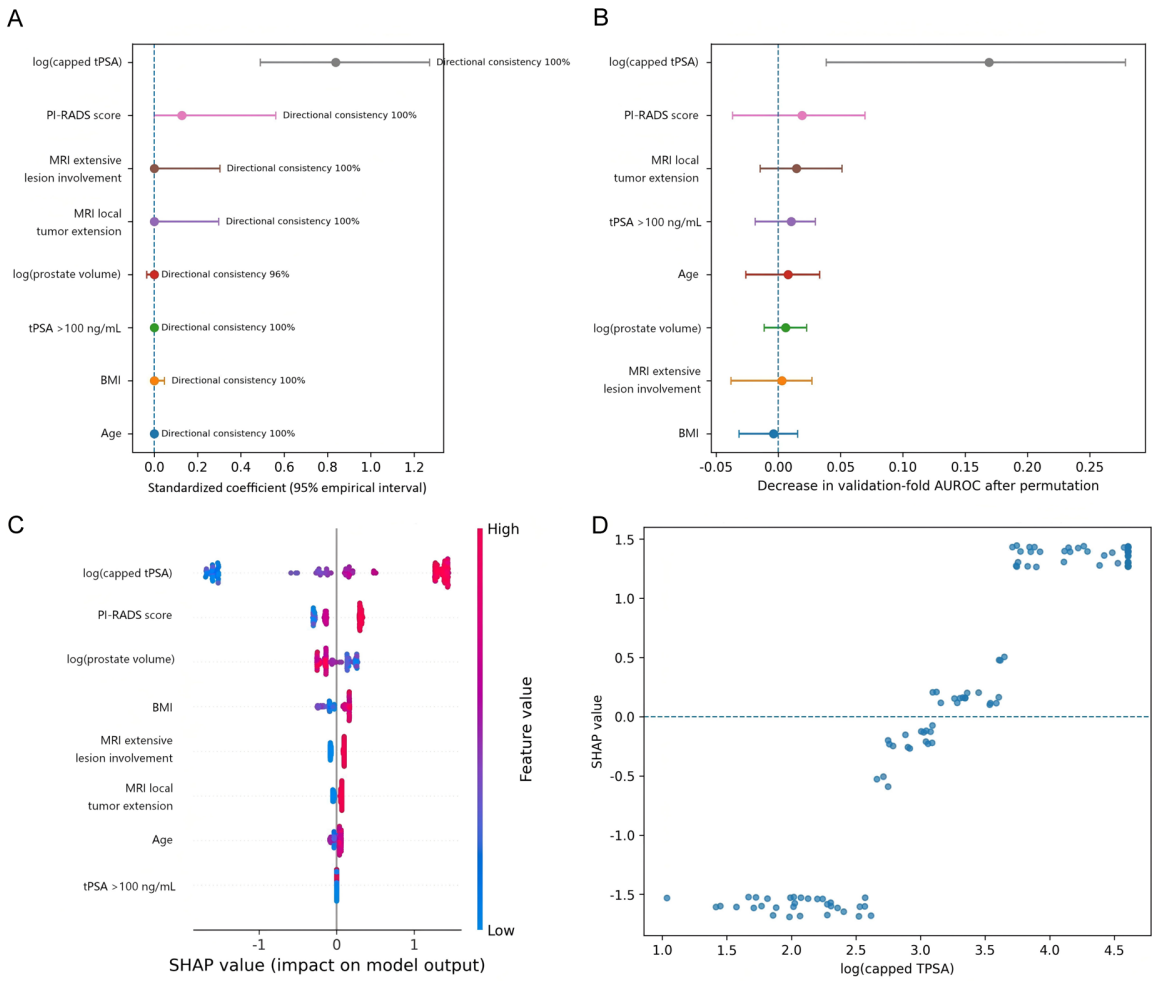

Figure 4. Model interpretability and stability analyses. Note: A. Median standardized regression coefficients, 95% empirical intervals, and directional consistency rates derived from 500 bootstrap resamples for the elastic net model. B. Permutation importance calculated in the outer validation folds for the random forest model. C. SHAP summary plot for the final XGBoost model fitted to the full cohort. D. SHAP dependence plot for log-transformed capped tPSA. BMI, body mass index; PI-RADS, Prostate Imaging Reporting and Data System; SHAP, SHapley Additive exPlanations; tPSA, total prostate-specific antigen; XGBoost, extreme gradient boosting.

prostate cancer on the index biopsy. Prebiopsy clinical variables and MRI features were integrated to develop and internally validate conditional prediction models for a maximum cancer involvement of >50% in a single biopsy core. The models were not intended to predict high tumor burden among all patients being considered for biopsy. Rather, they evaluated the predictive relationship between prebiopsy information and tumor burden on biopsy pathology in the restricted population of patients diagnosed with prostate cancer on the index biopsy. The clinical, imaging, and combined models all demonstrated predictive ability, with AUROCs of approximately 0.82 for both the clinical and combined models. The combined model showed no clear improvement in AUROC over the clinical model but had a lower Brier score than the imaging model, suggesting that adding clinical information improved overall predictive accuracy compared with imaging alone. Previous studies have shown that combining clinical variables with MRI features may facilitate prostate cancer risk stratification, although the incremental value over conventional predictors such as PI-RADS score and PSA density may depend on the study population, predictor composition, and model complexity [9-11].

The maximum percentage of cancer involvement in a single biopsy core reflects the greatest extent of local tumor involvement within an

Table 4. Sensitivity analyses of the combined elastic net model

Analysis	n	Positive outcomes, n	AUROC (95% CI)	AP (95% CI)	Brier score (95% CI)	Calibration slope
Primary analysis: maximum cancer involvement >50% in a single biopsy core	127	85	0.823 (0.732-0.901)	0.867 (0.777-0.947)	0.156 (0.127-0.188)	1.580
Outcome redefined as ≥50%	127	94	0.818 (0.715-0.904)	0.889 (0.807-0.972)	0.150 (0.119-0.183)	1.895
Outcome redefined as ≥70%	127	73	0.829 (0.749-0.900)	0.818 (0.718-0.920)	0.167 (0.141-0.195)	1.715
Patients with tPSA >100 ng/mL excluded	91	52	0.803 (0.700-0.894)	0.812 (0.694-0.927)	0.194 (0.169-0.220)	2.231
Transrectal biopsy only	99	66	0.836 (0.734-0.923)	0.864 (0.751-0.969)	0.159 (0.130-0.195)	2.051
Biopsy route and total number of cores added to the combined model	127	85	0.818 (0.718-0.903)	0.850 (0.762-0.947)	0.158 (0.130-0.189)	1.748

Note: The primary analysis defined a maximum percentage of cancer involvement of >50% in a single biopsy core as the positive outcome. Sensitivity analyses included redefining the outcome thresholds, excluding patients with tPSA >100 ng/mL, restricting the analysis to patients who underwent transrectal systematic biopsy, and adding biopsy route and total number of biopsy cores to the combined model. All analyses used the same elastic net modeling procedure and five repeats of stratified five-fold cross-validation as the primary analysis. The 95% CIs for AUROC, AP, and Brier score were estimated using 2,000 patient-level bootstrap resamples. The ideal value of the calibration slope is 1. AP, average precision; AUROC, area under the receiver operating characteristic curve; CI, confidence interval; tPSA, total prostate-specific antigen.

individual core and is an important measure of tumor burden on biopsy pathology [2]. Previous active surveillance studies and related evaluations have considered the extent of cancer involvement in individual cores, the number of positive cores, and pathological grade important for patient selection and monitoring of disease progression. Cancer involvement of no more than 50% in any individual core has commonly been used as a reference criterion for lower tumor burden [3, 12, 13]. In addition, tumor extent on MRI has been associated with adverse pathological features [14]. Therefore, using a maximum cancer involvement of >50% in a single biopsy core as the primary outcome has a clinical basis. Nevertheless, this measure reflects only local tumor involvement within a limited biopsy sample and may be affected by the biopsy scheme and sampling location. It cannot fully represent the total tumor volume of the prostate or pathological stage after radical prostatectomy [15].

tPSA showed the most consistent predictive contribution across models. Both the tPSA level and the proportion of patients with tPSA >100 ng/mL were markedly higher in the >50% group. Elastic net, random forest, and XGBoost also identified tPSA-related variables as important predictors, suggesting that higher tPSA may be associated with more extensive local tumor involvement. Wang et al. noted that tPSA is not specific to prostate cancer and may also be affected by nonmalignant factors such as prostate volume and inflammation. Nevertheless, PSA density has demonstrated favorable risk-stratification value for clinically significant prostate cancer in the MRI era, suggesting that future studies should evaluate its incremental value for predicting the maximum percentage of cancer involvement in a single biopsy core [16].

Higher PI-RADS scores and MRI findings of extraprostatic extension, seminal vesicle invasion, bilateral lesions, and multifocal lesions were more frequent in the >50% group. These findings suggest that MRI may reflect potential tumor burden through lesion suspicion, local invasion, and disease extent [4, 5]. Previous studies have shown that MRI and its quantitative features can be used to evaluate extraprostatic extension and other signs of locally advanced disease [5, 17, 18]. However, the combined model showed no clear incremental discrimination over the clinical model. This may be attributable to the strong predictive signal of tPSA, overlapping information between clinical and MRI variables, and the absence of maximum lesion diameter and quantitative apparent diffusion coefficient (ADC) measurements [9, 16]. Previous multicenter studies have found that PI-RADS score and PSA density may themselves contain substantial predictive information and that some radiomic features do not further improve model performance. Differences in MRI equipment, acquisition protocols, and readers may also affect model performance and reproducibility [10, 19-21].

In the algorithm comparison, elastic net achieved the numerically highest AUROC, whereas XGBoost showed a higher AP, a lower Brier score, and favorable calibration. However, no single model was superior across all performance measures. Artificial intelligence models have demonstrated favorable performance in prostate MRI lesion detection and risk assessment for clinically significant prostate cancer, with some studies reporting performance approaching or exceeding that of conventional radiological interpretation [11, 22]. However, multicenter and external validation studies have also shown that performance may be affected by sample size, data heterogeneity, and changes in the intended setting [10, 19, 20]. Therefore, given the limited sample size and relatively small number of predictors in the present study, the absence of a consistent advantage of complex machine learning algorithms over penalized regression is plausible. These findings indicate that model performance depends not only on algorithmic complexity but also on sample size and data heterogeneity. Model interpretation further showed that tPSA and PI-RADS score made relatively consistent predictive contributions across algorithms [23].

In the sensitivity analyses, the model AUROCs remained above 0.80 after changing the outcome thresholds, excluding patients with tPSA >100 ng/mL, restricting the analysis to patients who underwent transrectal biopsy, or additionally including biopsy route and total number of biopsy cores. These results were generally similar to those of the primary analysis, indicating that model discrimination was relatively

robust. However, calibration measures varied in some sensitivity analyses, suggesting that the accuracy of predicted probabilities may be affected by changes in the study population and outcome definition. These findings further emphasize the importance of calibration assessment and external validation [24].

This study has several limitations. First, it was a single-center retrospective study with a relatively small sample size, and only internal validation was performed. Previous studies have shown that the performance of prostate MRI and machine learning models may be lower in external data than in internal validation. The discrimination, calibration, and clinical applicability of these models therefore require further evaluation in independent multicenter cohorts [9, 17, 20]. Second, the study included only patients diagnosed with prostate cancer at initial systematic biopsy and did not include patients with negative biopsy findings. Consequently, the models cannot be extrapolated to all patients being considered for biopsy and cannot be used to determine whether a biopsy should be performed. Third, the sampling routes and numbers of biopsy cores were not identical between transrectal and transperineal biopsy. Although recent evidence suggests that both routes provide favorable cancer detection, they differ in sampling technique, safety, and lesion coverage. Despite the sensitivity analyses, their potential effects could not be completely excluded [25]. Fourth, the maximum percentage of cancer involvement in a single biopsy core may be affected by core length, specimen fragmentation, pathological measurement methods, and biopsy sampling schemes. Dichotomization of this measure may also have resulted in information loss. Fifth, no specific clinical decision context or risk threshold was established, and clinical net benefit was not evaluated using decision curve analysis. Therefore, the models have not yet been shown to improve clinical decision-making. Sixth, both tPSA and fPSA were subject to upper reporting limits. Although capped values and upper-limit indicator variables were used, the true continuous values above these limits could not be recovered. Seventh, PSA density, maximum lesion diameter, ADC values, and other potentially relevant clinical factors were not included. Previous studies suggest that PSA density and selected quantitative clinical and imaging measures may further improve risk stratification for indeterminate PI-RADS lesions. Their incremental value should be evaluated in larger cohorts [16, 26].

In conclusion, among patients diagnosed with prostate cancer at initial systematic biopsy, prebiopsy clinical variables and MRI features enabled risk stratification for maximum cancer involvement of >50% in a single biopsy core. tPSA and PI-RADS score showed relatively consistent predictive contributions across models. Elastic net achieved the numerically highest AUROC, whereas XGBoost achieved the highest AP, the lowest Brier score, and calibration closest to the ideal values. This study provides a basis for using prebiopsy information to estimate the probability of maximum cancer involvement of >50% in a single biopsy core. Further evaluation in independent cohorts is required to establish model generalizability and potential clinical applications.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Li Yang, Department of Urology, The First Affiliated Hospital of Dali University, 32 Jiasibo Road, Dali 671000, Yunnan, China. E-mail: yangliyndali@hotmail.com

References

[1] Haffner MC, Zwart W, Roudier MP, True LD, Nelson WG, Epstein JI, De Marzo AM, Nelson PS and Yegnasubramanian S. Genomic and phenotypic heterogeneity in prostate cancer. Nat Rev Urol 2021; 18: 79-92.

[2] Berney DM, Finnegan K, Chu K, Fine SW, Varma M, Cuzick J and Beltran L; Transatlantic Prostate Group. Measuring cancer burden in prostatic needle core biopsies: simplified assessments outperform complex measurements in assessing outcome: evidence to assist pathologist efficiency and minimize datasets. Histopathology 2023; 82: 1021-1028.

[3] Willemse PM, Davis NF, Grivas N, Zattoni F, Lardas M, Briers E, Cumberbatch MG, De Santis M, Dell'Oglio P, Donaldson JF, Fossati N, Gandaglia G, Gillessen S, Grummet JP, Henry AM, Liew M, MacLennan S, Mason MD, Moris L, Plass K, O'Hanlon S, Omar MI, Oprea-Lager DE, Pang KH, Paterson CC, Ploussard G, Rouvière O, Schoots IG, Tilki D, van den Bergh RCN, Van den Broeck T, van der Kwast TH, van der Poel HG, Wiegel T, Yuan CY, Cornford P,

Mottet N and Lam TBL. Systematic review of active surveillance for clinically localised prostate cancer to develop recommendations regarding inclusion of intermediate-risk disease, biopsy characteristics at inclusion and monitoring, and surveillance repeat biopsy strategy. Eur Urol 2022; 81: 337-346.

[4] Oerther B, Nedelcu A, Engel H, Schmucker C, Schwarzer G, Brugger T, Schoots IG, Eisenblaetter M, Sigle A, Gratzke C, Bamberg F and Benndorf M. Update on PI-RADS version 2.1 diagnostic performance benchmarks for prostate MRI: systematic review and meta-analysis. Radiology 2024; 312: e233337.

[5] Lin Y, Johnson LA, Fennessy FM and Turkbey B. Prostate cancer local staging with magnetic resonance imaging. Radiol Clin North Am 2024; 62: 93-108.

[6] Schrader A, Netzer N, Hielscher T, Görtz M, Zhang KS, Schütz V, Stenzinger A, Hohenfellner M, Schlemmer HP and Bonekamp D. Prostate cancer risk assessment and avoidance of prostate biopsies using fully automatic deep learning in prostate MRI: comparison to PI-RADS and integration with clinical data in nomograms. Eur Radiol 2024; 34: 7909-7920.

[7] Sungur M, Aykaç A, Aydin ME, Celik O and Kaya C. Machine learning-based prediction of prostate biopsy necessity using PSA, MRI, and hematologic parameters. J Clin Med 2024; 14: 183.

[8] Bacchetti E, De Nardin A, Giannarini G, Cereser L, Zuiani C, Crestani A, Girometti R and Foresti GL. A deep learning model integrating clinical and MRI features improves risk stratification and reduces unnecessary biopsies in men with suspected prostate cancer. Cancers (Basel) 2025; 17: 2257.

[9] Deniffel D, Healy GM, Dong X, Ghai S, Salinas-Miranda E, Fleshner N, Hamilton R, Kulkarni G, Toi A, van der Kwast T, Zlotta A, Finelli A, Perlis N and Haider MA. Avoiding unnecessary biopsy: MRI-based risk models versus a PI-RADS and PSA density strategy for clinically significant prostate cancer. Radiology 2021; 300: 369-379.

[10] Krauss W, Frey J, Heydorn Lagerlöf J, Lidén M and Thunberg P. Radiomics from multisite MRI and clinical data to predict clinically significant prostate cancer. Acta Radiol 2024; 65: 307-317.

[11] Ruan M, Liu Y, Yao K, Wang K, Fan Y, Wu S and Wang X. Development and validation of interpretable machine learning models for clinically significant prostate cancer diagnosis in patients with lesions of PI-RADS v2.1 score ≥3. J Magn Reson Imaging 2024; 60: 2130-2141.

[12] Walker CH, Marchetti KA, Singhal U and Morgan TM. Active surveillance for prostate cancer: selection criteria, guidelines, and outcomes. World J Urol 2022; 40: 35-42.

[13] Pepe P, Pepe L, Pennisi M and Fraggetta F. Which prostate biopsy in men enrolled in active surveillance? Experience in 110 men submitted to scheduled three-years transperineal saturation biopsy combined with fusion targeted cores. Clin Genitourin Cancer 2021; 19: 305-308.

[14] Ratnani P, Dovey Z, Parekh S, Sobotka S, Shukla D, Davis A, Roshandel R, Wagaskar V, Jambor I, Lundon DJ, Wiklund P, Kyprianou N, Menon M and Tewari A. Prostate MRI percentage tumor involvement or "PI-RADS percent" as a predictor of adverse surgical pathology. Prostate 2022; 82: 970-983.

[15] Burk KS, Naik S, Lacson R, Tuncali K, Lee LK, Tempany C, Cole AP, Trinh QD, Kibel AS and Khorasani R. MRI-targeted, systematic, or combined biopsy for detecting clinically significant prostate cancer. J Am Coll Radiol 2023; 20: 687-695.

[16] Wang S, Kozarek J, Russell R, Drescher M, Khan A, Kundra V, Barry KH, Naslund M and Siddiqui MM. Diagnostic performance of prostate-specific antigen density for detecting clinically significant prostate cancer in the era of magnetic resonance imaging: a systematic review and meta-analysis. Eur Urol Oncol 2024; 7: 189-203.

[17] van den Berg I, Soeterik TFW, van der Hoeven EJRJ, Claassen B, Brink WM, Baas DJH, Sedelaar JPM, Heine L, Tol J, van der Voort van Zyp JRN, van den Berg CAT, van den Bergh RCN, van Basten JA and van Melick HHE. The development and external validation of artificial intelligence-driven MRI-based models to improve prediction of lesion-specific extraprostatic extension in patients with prostate cancer. Cancers (Basel) 2023; 15: 5452.

[18] Ponsiglione A, Gambardella M, Stanzione A, Green R, Cantoni V, Nappi C, Crocetto F, Cuocolo R, Cuocolo A and Imbriaco M. Radiomics for the identification of extraprostatic extension with prostate MRI: a systematic review and meta-analysis. Eur Radiol 2024; 34: 3981-3991.

[19] Yilmaz EC, Harmon SA, Law YM, Huang EP, Belue MJ, Lin Y, Gelikman DG, Ozyoruk KB, Yang D, Xu Z, Tetreault J, Xu D, Hazen LA, Garcia C, Lay NS, Eclarinal P, Toubaji A, Merino MJ, Wood BJ, Gurram S, Choyke PL, Pinto PA and Turkbey B. External validation of a previously developed deep learning-based prostate lesion detection algorithm on paired external and in-house biparametric MRI scans. Radiol Imaging Cancer 2024; 6: e240050.

[20] Gresser E, Schachtner B, Stüber AT, Solyanik O, Schreier A, Huber T, Froelich MF, Magistro G,

Kretschmer A, Stief C, Ricke J, Ingrisch M and Nörenberg D. Performance variability of radiomics machine learning models for the detection of clinically significant prostate cancer in heterogeneous MRI datasets. Quant Imaging Med Surg 2022; 12: 4990-5003.

[21] Ren H, Peng Y, Si Y, Ye Y and Gong L. Accuracy, intra-, and inter-radiologist variability of PI-RADS v2.1 scoring for clinically significant prostate cancer detection. Quant Imaging Med Surg 2025; 15: 7080-7089.

[22] Saha A, Bosma JS, Twilt JJ, van Ginneken B, Bjartell A, Padhani AR, Bonekamp D, Villeirs G, Salomon G, Giannarini G, Kalpathy-Cramer J, Barentsz J, Maier-Hein KH, Rusu M, Rouvière O, van den Bergh R, Panebianco V, Kasivisvanathan V, Obuchowski NA, Yakar D, Elschot M, Veltman J, Fütterer JJ, de Rooij M and Huisman H; PI-CAI Consortium. Artificial intelligence and radiologists in prostate cancer detection on MRI (PI-CAI): an international, paired, non-inferiority, confirmatory study. Lancet Oncol 2024; 25: 879-887.

[23] Bhattacharya I, Khandwala YS, Vesal S, Shao W, Yang Q, Soerensen SJC, Fan RE, Ghanouni P, Kunder CA, Brooks JD, Hu Y, Rusu M and Sonn GA. A review of artificial intelligence in prostate cancer detection on imaging. Ther Adv Urol 2022; 14: 17562872221128791.

[24] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, Boulesteix AL, Camaradou JC, Celi LA, Denaxas S, Denniston AK, Glocker B, Golub RM, Harvey H, Heinze G, Hoffman MM, Kengne AP, Lam E, Lee N, Loder EW, Maier-Hein L, Mateen BA, McCradden MD, Oakden-Rayner L, Ordish J, Parnell R, Rose S, Singh K, Wynants L and Logullo P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024; 385: e078378.

[25] Zattoni F, Rajwa P, Miszczyk M, Fazekas T, Carletti F, Carrozza S, Sattin F, Reitano G, Botti S, Matsukawa A, Dal Moro F, Jeffrey Karnes R, Briganti A, Novara G, Shariat SF, Ploussard G and Gandaglia G. Transperineal versus transrectal magnetic resonance imaging-targeted prostate biopsy: a systematic review and meta-analysis of prospective studies. Eur Urol Oncol 2024; 7: 1303-1312.

[26] Ajami T, Han S, Porto JG, Kimbel I, Szczotka Z, Guerard T, VanderVeer-Harris N, Ledesma BR, Acosta PC, Kryvenko ON, Parekh DJ, Stoyanova R, Reis IM and Punnen S. Molecular and diffusion features for identification of clinically significant prostate cancer in PI-RADS 3 lesions. Urol Oncol 2024; 42: 370.e9-370.e14.