

Original Article

Diagnostic utility of serum prostate-specific antigen and circulating inflammatory markers for differentiating prostate cancer from benign prostatic hyperplasia

Dawei Luo, Hongjun Xie

Department of Urology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China

Received July 13, 2025; Accepted November 6, 2025; Epub November 15, 2025; Published November 30, 2025

Abstract: Objective: To assess the diagnostic performance of serum prostate-specific antigen (PSA), the Prostate Health Index (PHI), and peripheral blood inflammatory markers (neutrophil-lymphocyte ratio (NLR), lymphocyte-monocyte ratio (LMR), neutrophil-apolipoprotein A1 ratio (NAR) apolipoprotein A1 (ApoA1)) in differentiating prostate cancer (PCa) from biopsy-negative benign prostatic hyperplasia (BPH), and to construct an optimized machine learning diagnostic model. Methods: A retrospective analysis was conducted on 701 patients referred for prostate biopsy between March 2018 and January 2024, including 421 PCa and 280 BPH cases. Patients were divided into training (60%; n=421), validation (20%; n=140), and test (20%; n=140) cohorts. LASSO regression identified key predictors, which were used to develop five machine learning models-logistic regression, decision tree, random forest, support vector machine, and XGBoost. model performance was evaluated using ROC and precision-recall curves, calibration plots, Brier Scores, and decision curve analysis (DCA). AUCs were compared using the DeLong test. Results: PCa patients exhibited higher PSA, Neu, MONO, NLR, NAR, and PHI but lower ApoA1 and LMR than BPH patients (all $P<0.05$). XGBoost achieved the best performance (AUC: training 0.994; validation 0.953; test 0.979), significantly surpassing PSA (AUC difference: 0.055-0.118, $P<0.001$) and PHI (AUC difference: 0.077-0.084, $P<0.007$). Calibration curves indicated low Brier Scores (0.0326-0.0751) and excellent model fit. DCA confirmed superior clinical benefit. NLR and NAR were major contributors to PCa risk prediction. Conclusions: The XGBoost model integrating NLR, LMR, and NAR demonstrates superior diagnostic accuracy and clinical utility compared with PSA and PHI, potentially improving pre-biopsy risk stratification and reducing unnecessary invasive procedures.

Keywords: Prostate cancer, benign prostatic hyperplasia, PSA, circulating inflammatory markers, nomogram

Introduction

Prostate cancer (PCa) is one of the most common malignancies in the male genitourinary system, particularly in the United States, where it ranks as the second most fatal malignancy for men, posing a serious health threat to older males [1]. In China, the detection rate of PCa continues to rise due to improvements in quality of life and advancements in detection technologies [2]. However, many patients are diagnosed at an advanced stage or with distant metastases, leading to poor prognosis and increasing mortality each year [3]. Therefore, early detection and accurate diagnosis of PCa are critical.

Serum prostate-specific antigen (PSA) is a core biomarker for early screening and monitoring

therapeutic effects in PCa, playing a vital role in clinical practice [4]. However, PSA's diagnostic specificity is limited, as its levels can also rise in benign prostatic hyperplasia (BPH), prostatitis, and other non-malignant conditions, affecting its accuracy in differential diagnosis [5]. Consequently, there is a pressing need to explore new biomarkers that, in combination with PSA, can improve the accuracy of PCa diagnosis.

Recent studies have highlighted the role of inflammation in tumorigenesis. Chronic inflammation and excessive inflammatory responses may trigger histiocytic cancers [6]. Approximately 20% of cancers are linked to chronic inflammatory stimuli, such as gastric cancer from gastritis and liver cancer from hepatitis

[7]. Neutrophils, through the secretion of inflammatory factors like vascular endothelial growth factor (VEGF), interleukins (IL-6/10), and prostaglandins, significantly influence the tumor microenvironment (TME) in PCa [8, 9]. Monocytes and their derivatives, such as macrophages, are crucial in tumor progression, spread, and therapeutic responses [10]. Emerging evidence suggests that circulating inflammatory markers (CIMs) like neutrophil-lymphocyte ratio (NLR), lymphocyte-monocyte ratio (LMR), neutrophil-apolipoprotein A1 ratio (NAR), and Prostate Health Index (PHI) have potential diagnostic and prognostic applications in oncology. Song et al. [11] demonstrated associations between NLR, LMR, and platelet-to-lymphocyte ratios with breast cancer prognosis, while Shi et al. [12] showed that NLR, LMR, and nutritional risk index predict outcomes in non-small cell lung cancer. However, systematic research on the role of NLR, LMR, NAR, and PHI in the differential diagnosis of PCa is lacking.

In this study, we systematically evaluate the clinical application of serum PSA, circulating inflammatory markers (NLR, LMR, NAR, apolipoprotein A1 (ApoA1)), and PHI in distinguishing PCa from BPH. By comparing baseline characteristics and inflammatory factor levels in training, validation, and test groups, key diagnostic variables (NLR, LMR, NAR) are selected using Lasso regression. Various machine learning models are then constructed and evaluated to assess model performance comprehensively. This study innovates by combining CIMs with machine learning techniques to optimize the PCa diagnostic model. Its goal is to provide clinically applicable, accurate, and reliable diagnostic tools that reduce unnecessary invasive examinations and improve patient prognoses.

Methods and materials

Sample source

We retrospectively evaluated 701 consecutive patients (421 PCa and 280 BPH cases.) treated between March 2018 and January 2024. The PCa cohort comprised patients with histologically confirmed PCa, while the control cohort consisted of men with benign biopsy results (primarily BPH). Samples were randomly allocated to training (60%), validation (20%), and test sets (20%) with matched case distribu-

tions (training: 289 PCa/132 BPH; validation: 100 PCa/40 BPH; test: 95 PCa/45 BPH) for model development, validation, and evaluation. The study was approved by the Ethics Committee of The First Affiliated Hospital of Xi'an Jiaotong University (**Figure 1**).

Inclusion and exclusion criteria (PCa patients)

Inclusion criteria: PCa confirmed by prostate biopsy or post-surgery pathology; no recent (≤ 3 months) infections; no blood/immune system abnormalities; complete clinical and lab data.

Exclusion other malignancies; critical organ (cardiac, hepatic, or renal) impairment; severe infections or chronic inflammatory conditions (e.g., prostatitis); autoimmune or hypersensitivity disorders.

Inclusion and diagnostic criteria (BPH cases)

Inclusion criteria: Controls were biopsy-negative men with benign disease (predominantly BPH) referred for prostate biopsy due to elevated PSA and/or suspicious DRE/imaging findings. Histopathology revealed no malignancy, and clinical evaluation was consistent with BPH. Men were followed for $\geq X$ months to exclude missed cancers (if available). Normal bloodwork, biochemistry, immunology, and microbiology assessments excluding acute illness; negative imaging findings for PCa and other malignancies (e.g., prostate ultrasound, CT, MRI); complete clinical records; no hematologic/immune pathologies.

Clinical data collection

Patient clinical and laboratory data were obtained through the electronic medical record system. Clinical data included age, BMI, and histories of smoking, alcohol consumption, hypertension, hyperlipidemia, and diabetes. Laboratory parameters included PSA, neutrophil count (Neu), lymphocyte count (Lym), monocyte count (MONO), ApoA1, and calculated ratios (NLR, LMR, NAR, PHI).

Model construction

Five machine learning models were constructed: Logistic Regression, Decision Tree (DT), Random Forest (RF), Support Vector Machine

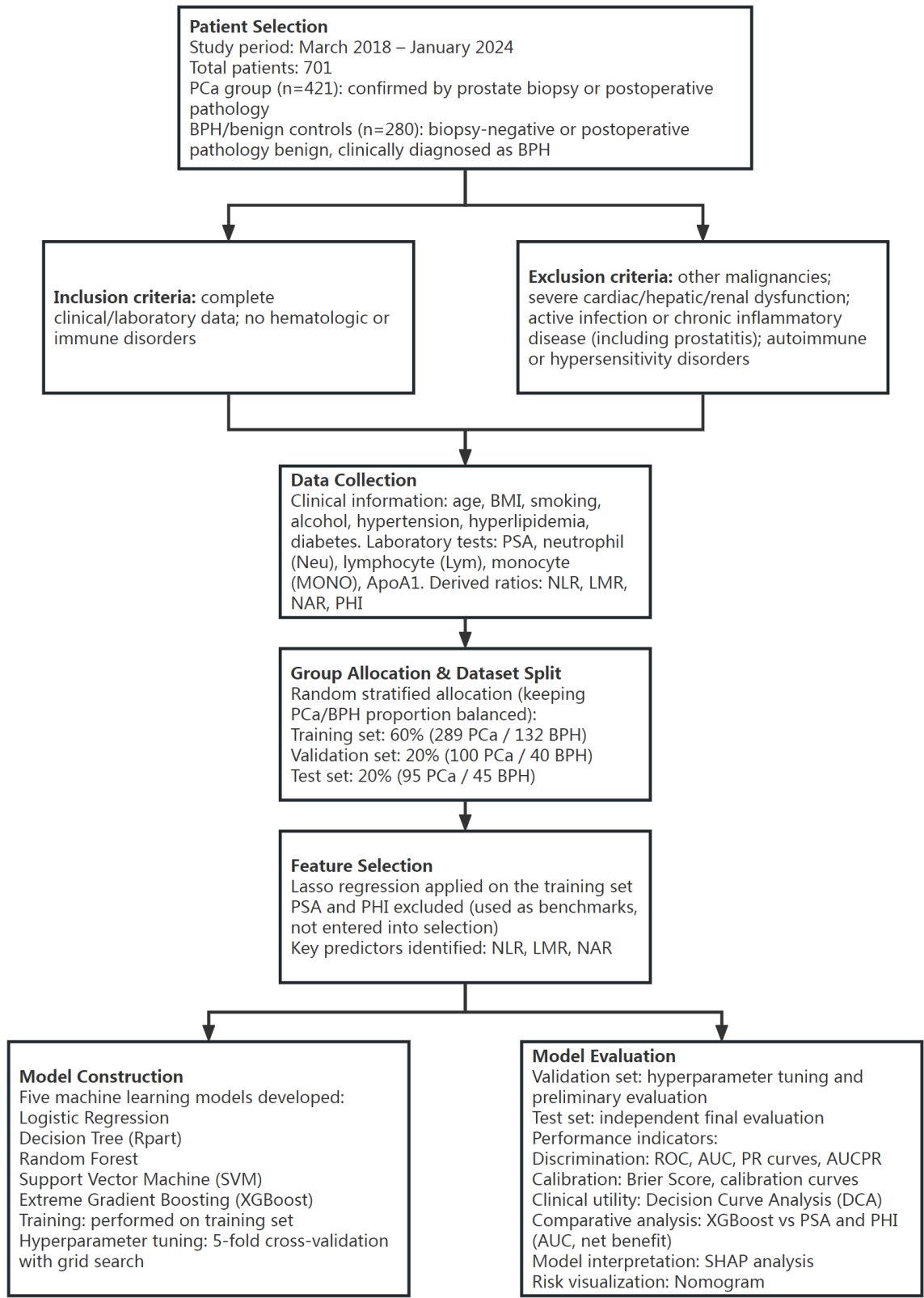


Figure 1. Research flow chart.

(SVM), and Extreme Gradient Boosting (XGBoost). All models were implemented using mlr3, and predictions were set to return probabilities (predict type="prob"). Model optimization was performed via 5-fold cross-validation using the training set. Key parameters (e.g., DT's cp and minbucket, RF's mtry and nodesize, SVM's cost and gamma, XGBoost's eta, nrounds, and maxdepth) were adjusted using grid search. Logistic Regression used default configurations without extensive tuning. The training set was used for model training, the validation set for hyperparameter selection and preliminary checks, and the test set for final performance assessment. Model performance was evaluated using accuracy (classif. acc), Brier Score (classif.brier), AUC (classif. auc), sensitivity (classif.sensitivity), specificity (classif.specificity), area under the precision-recall (PR) curve (classif.prauc), precision (classif.precision), recall (classif.recall), and F1 score (classif.fbeta). Model calibration and clinical net benefit were assessed via ROC curves, PR curves, calibration curves, and decision curve analysis (DCA). SHAP analysis was employed to identify the most important features in the top-performing model. To compare calibration performance, Brier Scores were decomposed into per-subject squared errors, and paired Wilcoxon signed-rank tests were conducted, with Holm correction for multiple comparisons. Robustness was confirmed using 2,000 rounds of bootstrap resampling and bias-corrected accelerated (BCa) 95% confidence intervals.

Outcome measurements

Baseline characteristics of the training, validation, and test groups were compared. The differences in clinical data and CIMs between PCa and BPH groups were analyzed. The diagnostic ability of serum PSA was evaluated against CIMs (Neu, MONO, ApoA1, NLR, LMR, NAR) and PHI. Key variables for PCa diagnosis were identified via Lasso regression, and the diagnostic value of NLR, LMR, and NAR was analyzed. Five machine learning models (Logistic Regression, DT, RF, XGBoost, SVM) were constructed based on the selected variables, and their performance was compared across training, validation, and test groups. A nomogram was used to analyze the contribution of NLR, LMR, and NAR in PCa risk prediction.

Statistical methods

Statistical analyses were performed using SPSS 26.0 and R software. Qualitative data were presented as numbers and percentages, and inter-group comparisons were made using the χ^2 test. Quantitative data were tested for normal distribution. Normally distributed data were expressed as mean $\pm$ standard deviation (SD) and compared using t-tests or ANOVA; non-normally distributed data were expressed as median and interquartile range, and analyzed using the Mann-Whitney U test or Kruskal-Wallis test. Lasso regression was used to select diagnostic variables, and five machine learning models (Logistic, DT, RF, XGBoost, SVM) were constructed. Model performance was evaluated using ROC, PR, and calibration curves, with AUC and Brier Score calculated. The DeLong test was used to compare AUC differences between models. A *P* value<0.05 was considered statistically significant.

Results

Comparison of baseline data among training, validation, and test groups

No significant differences were found in age, BMI, the prevalences of smoking, alcoholism, hypertension, hyperlipidemia, or diabetes among the groups (all *P*>0.05). PSA, Neu, Lym, MONO, ApoA1, NLR, LMR, NAR, and PHI also showed no significant differences (all *P*>0.05, **Table 1**).

Comparison of baseline characteristics and CIMs between PCa and BPH groups

No significant intergroup differences were found in age, BMI, smoking status, alcohol abuse, or comorbidities such as hypertension, hyperlipidemia, and diabetes (all *P*>0.05). Lym levels were comparable between groups (*P*=0.935). However, PCa patients exhibited significantly higher PSA (*P*<0.001), neutrophil (*P*<0.001), and monocyte (*P*=0.022) levels than BPH patients. In contrast, ApoA1 levels were markedly higher in the BPH group (*P*<0.001).

Regarding inflammatory markers, the PCa group demonstrated significantly elevated NLR and NAR values (both *P*<0.001) and a reduced LMR (*P*<0.001). Similarly, PHI levels were sub-

Table 1. Baseline feature comparisons across the training, validation, and test groups

Variable	Training group (n=421)	Validation group (n=140)	Test group (n=140)	Statistic	P
Age				0.368	0.832
≥70	252 (59.86%)	87 (62.14%)	87 (62.14%)		
<70	169 (40.14%)	53 (37.86%)	53 (37.86%)		
Body mass index				0.511	0.774
≥25 kg/m ²	58 (13.78%)	22 (15.71%)	18 (12.86%)		
<25 kg/m ²	363 (86.22%)	118 (84.29%)	122 (87.14%)		
Smoking history				3.265	0.195
With	352 (83.61%)	110 (78.57%)	109 (77.86%)		
Without	69 (16.39%)	30 (21.43%)	31 (22.14%)		
Alcoholism history				1.112	0.574
With	37 (8.79%)	10 (7.14%)	15 (10.71%)		
Without	384 (91.21%)	130 (92.86%)	125 (89.29%)		
Hypertension history				1.720	0.423
With	106 (25.18%)	42 (30.00%)	41 (29.29%)		
Without	315 (74.82%)	98 (70.00%)	99 (70.71%)		
Hyperlipidemia history				0.784	0.676
With	57 (13.54%)	16 (11.43%)	21 (15.00%)		
Without	364 (86.46%)	124 (88.57%)	119 (85.00%)		
Diabetes history				1.649	0.438
With	72 (17.10%)	25 (17.86%)	18 (12.86%)		
Without	349 (82.90%)	115 (82.14%)	122 (87.14%)		
PSA (ng/mL)	27.25 [20.16, 36.39]	29.94 [21.58, 38.67]	28.22 [19.69, 35.00]	3.191	0.203
Neu (×10 ⁹ /L)	4.86±0.65	4.89±0.70	4.93±0.64	0.565	0.569
Lym (×10 ⁹ /L)	1.20±0.32	1.22±0.33	1.25±0.31	0.926	0.397
MONO (×10 ⁹ /L)	0.38 [0.31, 0.45]	0.39 [0.33, 0.46]	0.39 [0.32, 0.45]	1.813	0.404
ApoA1 (g/L)	1.04 [0.99, 1.10]	1.02 [0.98, 1.08]	1.04 [0.98, 1.12]	2.218	0.330
NLR	4.01 [3.71, 4.53]	3.99 [3.72, 4.43]	3.95 [3.70, 4.32]	1.945	0.378
LMR	3.13 [3.02, 3.34]	3.12 [3.00, 3.31]	3.16 [3.04, 3.36]	2.471	0.291
NAR	4.68 [4.30, 5.12]	4.77 [4.35, 5.15]	4.67 [4.36, 5.15]	1.264	0.531
PHI	38.31±7.88	38.76±6.93	38.34±7.86	0.184	0.832

Note: PSA, prostate-specific antigen; Neu, neutrophil count; Lym, lymphocyte count; MONO, monocyte count; ApoA1, apolipoprotein A1; NLR, neutrophil-lymphocyte ratio; LMR, lymphocyte-monocyte ratio; NAR, neutrophil-apolipoprotein A1 ratio; PHI, prostate health index.

stantially higher in PCa compared with BPH (P<0.001, **Table 2**).

PCa diagnostic variables using Lasso regression

Lasso regression was used to identify key diagnostic variables in the training set. PSA and PHI, as established reference indicators, were excluded from the regression analysis. Instead, Lasso was applied to the remaining inflammatory markers (Neu, MONO, ApoA1, NLR, LMR, NAR) to find additional predictors. NLR, LMR,

and NAR were identified as the most relevant predictors (**Figure 2**).

Comparison of model performance

Five models (Logistic Regression, DT, RF, XGBoost, SVM) were tested on the training, validation, and test data. Model performance varied across datasets. All models showed performance metrics in the training, validation, and test sets, which were presented in a heatmap reflecting each model's predictive capabilities (**Figure 3**).

Table 2. Comparison of baseline characteristics and circulating inflammatory makers between PCa and BPH groups in training group

Variable	PCa (n=289)	BPH (n=132)	Statistic	P
Age			0.000	0.998
≥70	173 (59.86%)	79 (59.85%)		
<70	116 (40.14%)	53 (40.15%)		
Body mass index			0.736	0.391
≥25 kg/m ²	37 (12.80%)	21 (15.91%)		
<25 kg/m ²	252 (87.20%)	111 (84.09%)		
Smoking history			1.730	0.188
With	237 (82.01%)	115 (87.12%)		
Without	52 (17.99%)	17 (12.88%)		
Alcoholism history			0.022	0.882
With	25 (8.65%)	12 (9.09%)		
Without	264 (91.35%)	120 (90.91%)		
History of hypertension			0.089	0.765
With	74 (25.61%)	32 (24.24%)		
Without	215 (74.39%)	100 (75.76%)		
Hyperlipidemia history			0.777	0.378
With	42 (14.53%)	15 (11.36%)		
Without	247 (85.47%)	117 (88.64%)		
Diabetes history			0.995	0.319
With	53 (18.34%)	19 (14.39%)		
Without	236 (81.66%)	113 (85.61%)		
PSA (ng/mL)	32.94±10.37	19.41±5.49	-14.099	<0.001
Neu (×10 ⁹ /L)	4.97±0.65	4.62±0.58	-5.308	<0.001
Lym (×10 ⁹ /L)	1.20±0.31	1.21±0.32	0.081	0.935
MONO (×10 ⁹ /L)	0.39±0.10	0.36±0.09	-2.297	0.022
ApoA1 (g/L)	1.02 [0.97, 1.08]	1.08 [1.03, 1.12]	5.831	<0.001
NLR	4.11 [3.76, 4.54]	3.81 [3.49, 4.39]	4.515	<0.001
LMR	3.07 [2.97, 3.24]	3.31 [3.17, 3.40]	7.007	<0.001
NAR	4.88 [4.53, 5.24]	4.30 [4.04, 4.60]	10.257	<0.001
PHI	41.79±6.08	30.69±5.71	-17.712	<0.001

Note: PCa, prostate cancer; BPH, benign prostatic hyperplasia; PSA, prostate-specific antigen; Neu, neutrophil count; Lym, lymphocyte count; MONO, monocyte count; ApoA1, apolipoprotein A1; NLR, neutrophil-lymphocyte ratio; LMR, lymphocyte-monocyte ratio; NAR, neutrophil-apolipoprotein A1 ratio; PHI, prostate health index.

ROC curve evaluation of model performance

ROC curve analysis was performed to evaluate the classification performance of the five machine-learning models across the training, validation, and test datasets. In the training set (**Figure 4A**), XGBoost demonstrated the highest discriminative ability with an AUC of 0.9942, followed by RF (0.987) and SVM (0.9708), whereas logistic regression and DT yielded AUCs of 0.9618 and 0.9515, respectively. In the validation set (**Figure 4B**), XGBoost again achieved the highest AUC (0.9527), with RF (0.948) and SVM (0.944) showing comparable

performance, while logistic regression and DT obtained AUCs of 0.9052 and 0.9289. In the test set (**Figure 4C**), XGBoost maintained the best performance with an AUC of 0.9788, followed by RF (0.9724) and SVM (0.9682), and logistic regression and DT reached AUCs of 0.9401 and 0.923, respectively. Overall, XGBoost consistently showed the strongest discriminative power across all datasets.

PR curve evaluation of model performance

PR curve analysis was used to evaluate the classification performance of the five models

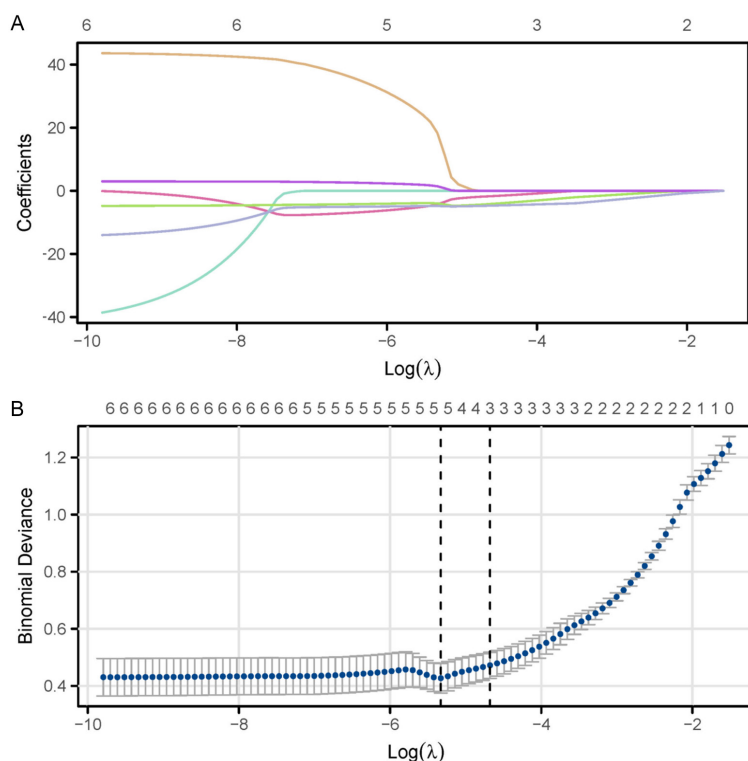


Figure 2. Lasso regression-based variable screening results. A: Coefficient trajectory plot for each variable, displaying how coefficients of variables change during Lasso regression. B: Bias curve, showing the trend in which bias varies with $\text{Log}(\lambda)$ in the model's cross-validation process.

in the training, validation, and test sets. In the training set, XGBoost showed the highest AUC (0.9974), indicating excellent performance in predicting positive cases ($P < 0.001$). RF and SVM had AUCs of 0.9941 and 0.9856, respectively, with good performance ($P < 0.001$). Logistic (AUC=0.9709) and DT (AUC=0.9723) showed some discriminative ability ($P < 0.001$). In the validation set, XGBoost, SVM, and RF achieved AUCs of 0.979 ($P < 0.001$), 0.974, and 0.9783, respectively, showing robustness. In the test set, XGBoost had the highest AUC (0.9907, $P < 0.001$), with RF and SVM achieving AUCs of 0.9877 and 0.985, respectively, demonstrating excellent classification performance (Figure 5).

Calibration curve evaluation of model performance

Calibration curves and Brier Scores were used to evaluate the prediction probabilities of the five models across the training, validation, and test sets. In the training set, XGBoost had the lowest Brier Score (0.0326), indicating the most accurate predictions ($P < 0.001$),

followed by RF (0.0435). In the validation group, SVM had the lowest Brier Score (0.0751, $P < 0.001$), with XGBoost (0.0742) showing favorable calibration. In the test set, SVM again had the lowest Brier Score (0.0640, $P < 0.001$), followed by RF (0.0642) and XGBoost (0.0561), all showing strong calibration. XGBoost and SVM consistently outperformed other models across all datasets (Figure 6). Pairwise Wilcoxon tests demonstrated that in the training set, XGBoost achieved significantly lower Brier Scores than all other models (all adjusted $P < 0.001$). In the validation and test sets, XGBoost and SVM outperformed Logistic Regression and DT, as well as Random Forest ($P < 0.05$). Differences between XGBoost and SVM were not significant (validation $P = 0.36$; test $P = 0.08$), indicating that both models had the best calibration (Table S1).

DCA curve evaluation of models' clinical decision-making value

DCA results showed that the XGBoost model outperformed other models across all datasets. Using dcurves: dca in R, thresholds were set from 0.00 to 1.00 at 0.01 increments with smoothing enabled (span=0.5). XGBoost consistently demonstrated the highest net benefit across the clinically relevant threshold range (0.10-0.70) in the training, validation, and test cohorts, outperforming both RF and SVM, as well as the treat-all and treat-none strategies. RF and SVM showed competitive, but slightly lower net benefit, while Logistic Regression and DT underperformed. These findings confirm the clinical utility of the XGBoost model (Figure 7A-C).

AUC comparison of XGBoost with PSA and PHI across different datasets

The AUC of XGBoost was consistently superior to that of PSA and PHI across all datasets. In the training set, the AUC differences between XGBoost and PSA (0.118) and between XGBoost

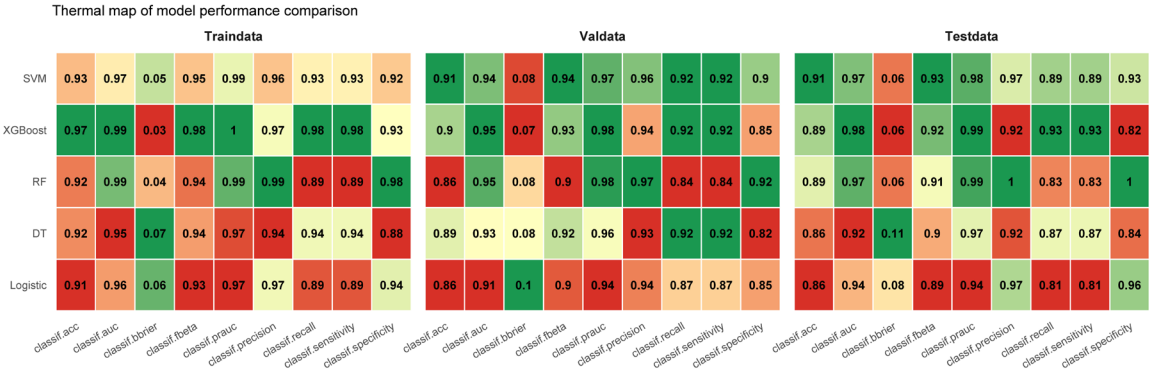


Figure 3. Heat map of model performance. The heatmap presents a performance comparison of Logistic, DT, RF, XGBoost, and SVM across training, validation, and test sets. The color gradient indicates the level of performance indicators. Red represents higher performance, while green represents lower performance. Note: DT, decision tree; RF, random forest; XGBoost, extreme gradient boosting; SVM, support vector machine.

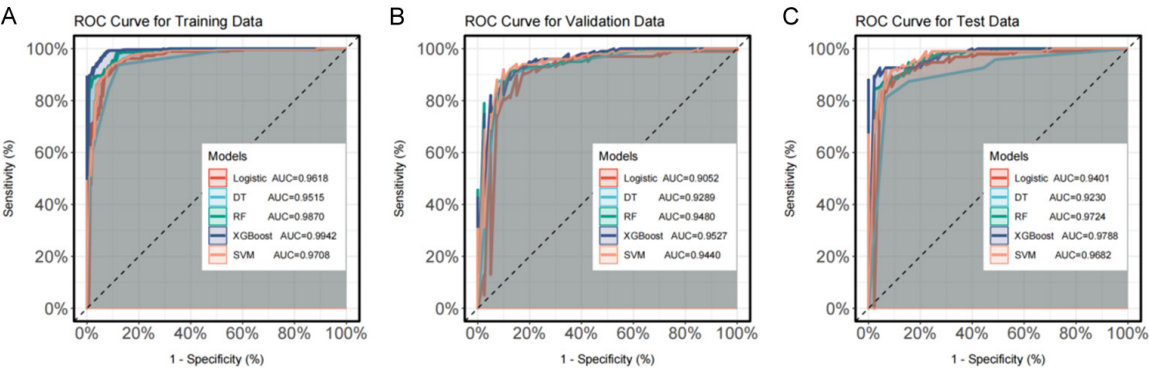


Figure 4. ROC curve analysis. A: This ROC curve is for training data and shows the sensitivity and specificity of Logistic, DT, RF, XGBoost, and SVM; B: This ROC curve is for validation data, showing the sensitivity and specificity of the five models; C: Test data ROC curve, displaying the five models' sensitivity and specificity. Note: ROC, receiver operating characteristic; DT, decision tree; RF, random forest; XGBoost, extreme gradient boosting; SVM, support vector machine; AUC, area under the curve.

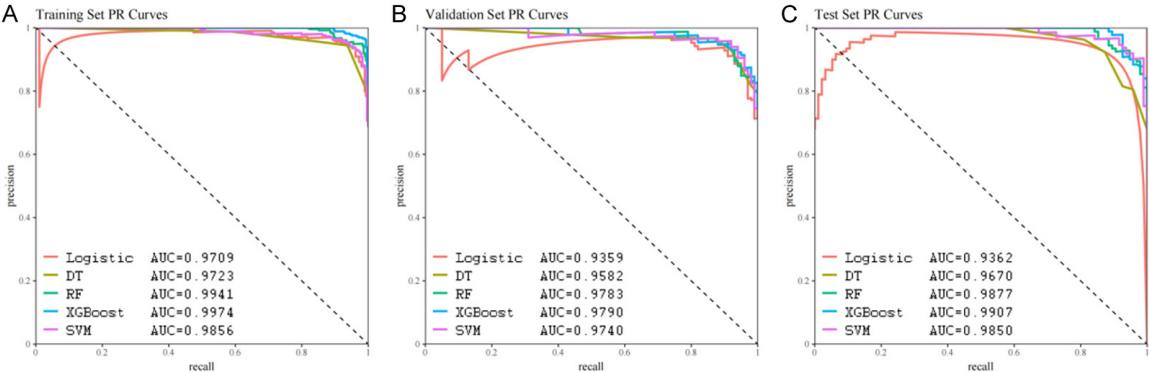


Figure 5. PR curves. A: PR relationships for Logistic, DT, RF, SVM, and XGBoost by PR curves (training data); B: The PR curve for the validation dataset, showing the precision and recall rates of these models; C: The PR curve for the test group data, presenting the precision and recall rates of these five models. Note: PR, precision-recall; DT, decision tree; RF, random forest; XGBoost, extreme gradient boosting; SVM, support vector machine; AUC, area under the curve.

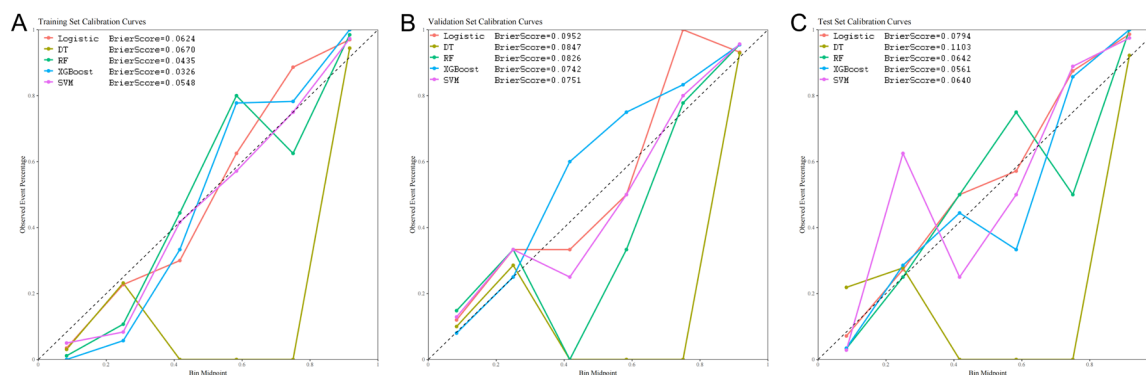


Figure 6. Calibration curves. A: DCA curves of the training data, showing the clinical net benefit of Logistic, Rpart, RF, SVM, and XGBoost models; B: The five models' clinical net benefit rates by DCA curves (validation data); C: Test data DCA curves showing the five models' clinical net benefits. Note: DT, decision tree; RF, random forest; XGBoost, extreme gradient boosting; SVM, support vector machine.

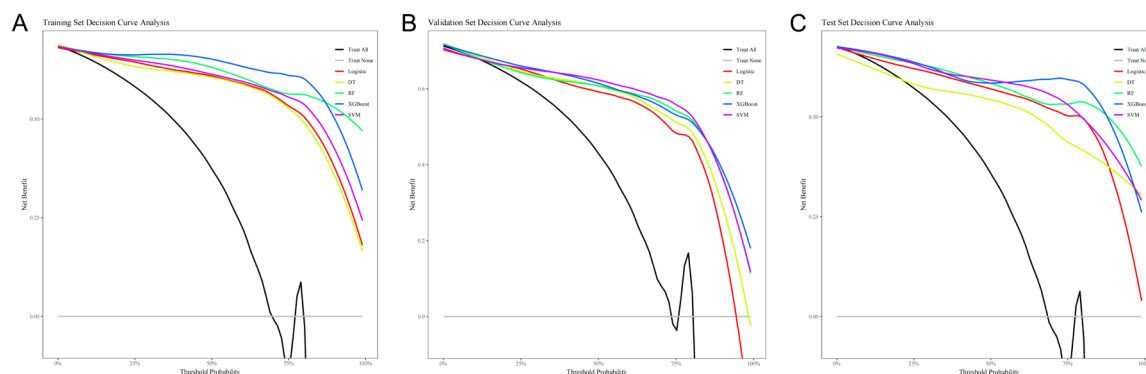


Figure 7. DCA. A: DCA of net benefits of Logistic Regression, DT, RF, SVM, and XGBoost in the training set; B: DCA-derived net benefit rates for all models in the validation set; C: DCA-derived net benefit rates for all models in the independent test set. Note: DCA, decision curve analysis; DT, decision tree; RF, random forest; XGBoost, extreme gradient boosting; SVM, support vector machine.

and PHI (0.084) were statistically significant ($P < 0.001$). In the validation set, the AUC differences were 0.055 when comparing XGBoost with PSA and 0.083 when comparing XGBoost with PHI, both reaching statistical significance ($P < 0.001$). In the test set, XGBoost again outperformed PSA and PHI, with AUC differences of 0.109 ($P < 0.001$) and 0.077 ($P = 0.007$), respectively. These findings further highlight the strong diagnostic potential of the XGBoost model, particularly in comparison with PSA and PHI (Figure 8).

Nomogram based on NLR, LMR, and NAR and its application value

A significant positive correlation was found between NLR and NAR, as evidenced by their similar trends in score changes in the logistic

regression model. This suggests that NLR and NAR play a crucial collaborative role in risk prediction. In contrast, LMR showed a weaker relationship with the risk, with its influence on the model being less pronounced. Although all three variables contribute to risk prediction, NLR and NAR appear to be more significant predictors, while the impact of LMR is relatively smaller (Figure 9).

Discussion

This study retrospectively analyzed 701 patients. With population aging and dietary changes in China, the incidence and mortality of PCa continue to rise annually [13], underscoring the need for more accurate diagnostic strategies. In our study, the XGBoost model based on NLR, LMR, and NAR achieved AUCs of

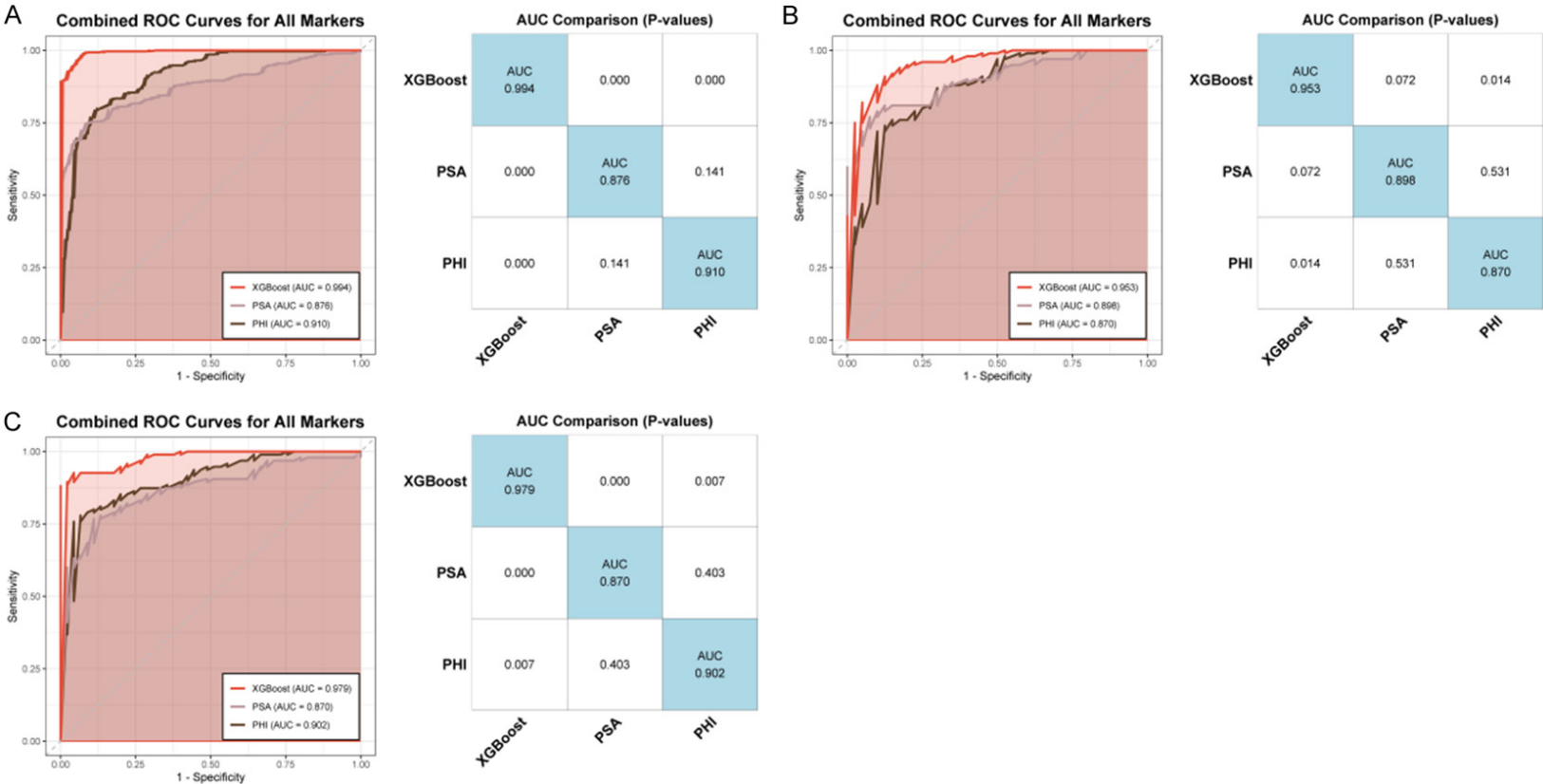


Figure 8. AUC comparison of XGBoost with PSA and PHI. A: XGBoost-PSA and XGBoost-PHI AUC comparisons (training set); B: XGBoost-PSA and XGBoost-PHI AUC comparisons (validation set); C: XGBoost-PSA and XGBoost-PHI AUC comparisons (test set). Note: ROC, receiver operating characteristic; AUC, area under the curve; XGBoost, extreme gradient boosting; PSA, prostate-specific antigen; PHI, prostate health index.

Serum PSA and inflammatory biomarkers in prostate cancer differential diagnosis

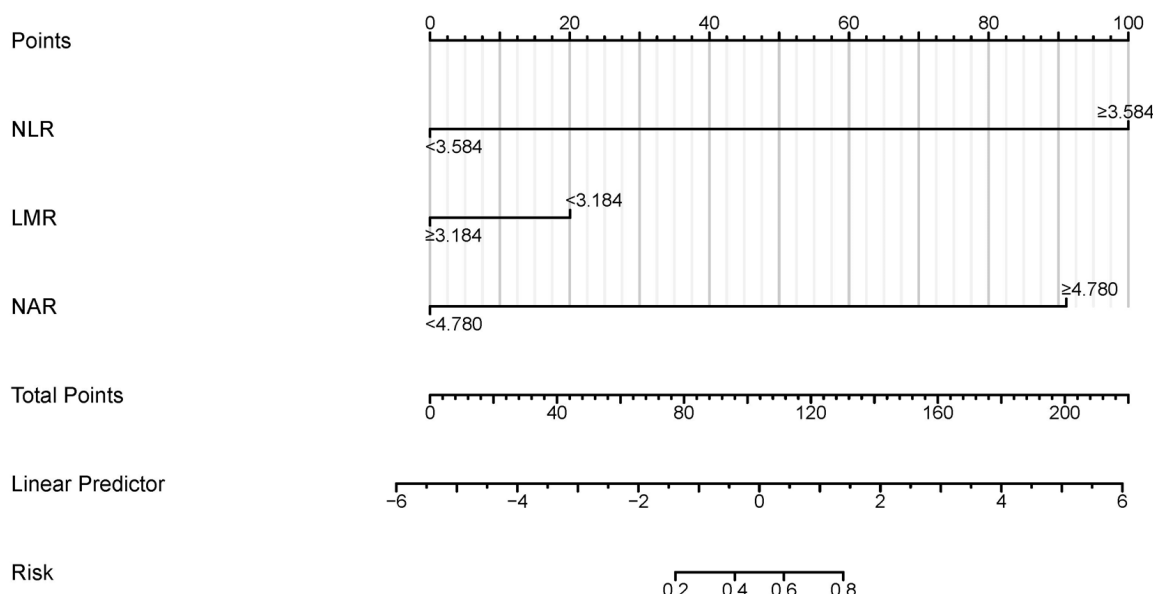


Figure 9. Nomogram and risk prediction based on NLR, LMR and NAR. Note: NLR, neutrophil-lymphocyte ratio; LMR, lymphocyte-monocyte ratio.

0.994, 0.953, and 0.979 in the training, validation, and test sets, respectively, consistently surpassing the diagnostic performance of PSA and PHI. The AUC differences between XGBoost and PSA were 0.118, 0.055, and 0.109 across the three datasets, while the differences between XGBoost and PHI were 0.084, 0.083, and 0.077, respectively. Calibration curves and DCA analysis further confirmed the excellent calibration and the highest net clinical benefit of the XGBoost model. These findings suggest that CIM-based machine learning models hold substantial clinical potential for improving diagnostic accuracy and reducing unnecessary invasive examinations.

PSA and PHI were excluded from the Lasso regression, as they were established benchmarks for comparison. Our approach specifically aimed to assess whether adding simple inflammatory markers could provide incremental diagnostic value over these traditional indicators.

The study found that NLR and NAR were significantly higher in PCa patients than in BPH cases, while LMR and ApoA1 levels were lower. These findings align with the known pathophysiological mechanisms of tumor-related inflammation. Inflammation is closely associated with cancer, particularly within the TME [16]. Tumor cells release cytokines and chemokines that trigger

systemic inflammation, altering immune cell function and promoting processes like angiogenesis and metastasis [17]. Neutrophils, in particular, can reshape the TME, promote cancer cell spread, and form neutrophil extracellular traps (NETs), enhancing cancer progression [18].

The advantage of NAR over NLR lies in its combination of inflammatory response and lipid metabolism disorders. NLR reflects neutrophil-mediated inflammation, while NAR integrates ApoA1, which is critical for antioxidant and anti-inflammatory functions in PCa [19]. ApoA1 is a key protein component of high-density lipoprotein (HDL) and plays a crucial role in cholesterol efflux. In PCa, this pathway is often silenced through hypermethylation of the ABCA1 promoter [20]. Experimental studies further show that SR-B1-mediated uptake of HDL/ApoA1 stimulates PCa cell proliferation, while genetic ablation of SR-B1 abrogates this effect [21]. At the population level, low circulating ApoA1/HDL levels in AMORIS cohort are associated with an increased risk of PCa [22]. Beyond its effects on PCa, ApoA1 also exerts systemic anti-inflammatory and antioxidant effects through the SR-B1/PDZK1/PI3K-Akt axis [24]. ApoA1's depletion has been linked to impaired immune function and altered lipid metabolism, making it an essential component of NAR in PCa diagnosis. Additionally, the decrease in LMR reflects

the weakening of immune surveillance due to the dysfunction of T lymphocytes and the conversion of monocytes into tumor-associated macrophages [24]. During the progression of PCa, tumor cells secrete immunosuppressive factors such as TGF- β and IL-10, which induce the proliferation of regulatory T cells and promote the differentiation of effector T cells into an exhausted phenotype-characterized by high expression of PD-1, TIM-3, and LAG-3 [25].

ApoA1 not only mediates lipid transport but also exerts important anti-inflammatory, antioxidant, and immunomodulatory effects. Mechanistically, ApoA1 exerts direct anti-tumor effects through multiple pathways. It disrupts lipid raft structures in PCa cell membranes by activating ABCA1-mediated cholesterol efflux, thereby interfering with key signaling cascades such as PI3K/AKT [26-28]. Decreased ApoA1 levels in PCa patients reflect impaired anti-inflammatory capacity, while elevated NAR levels indicate enhanced inflammatory responses and weakened anti-inflammatory ability-providing a theoretical basis for the application of CIMs in PCa.

Previous studies have shown the diagnostic and prognostic significance of CIMs in various cancers. For example, Song et al. [11] found that higher NLR and lower LMR are associated with poorer prognosis in breast cancer. Similarly, Shi et al. reported that NLR, LMR, and nutritional risk index function as independent prognostic factors in non-small cell lung cancer [12]. Investigations into cervical cancer have shown that serum inflammatory factors are highly expressed, with an AUC of up to 0.846 [29]. Similarly, Deepthi et al. reported that salivary gland inflammatory cytokines show promise as biomarkers for oral leukoplakia and oral squamous cell carcinoma [30]. These studies indicate that CIMs are closely associated with tumor initiation and progression. However, relatively few systematic studies have explored the combined use of multiple inflammatory factors in the differential diagnosis of PCa.

In this study, the combination of NLR, LMR, and NAR was applied to the differential diagnosis of PCa and BPH for the first time, with Lasso regression identifying the key diagnostic variables. The combined use of multiple markers provides a more comprehensive reflection of

the body's inflammatory state and immune function, improving diagnostic accuracy.

XGBoost outperformed other machine learning models in this study. As an ensemble learning algorithm, XGBoost excels in managing complex, non-linear relationships between features, which is critical for medical data analysis. Unlike traditional logistic regression, XGBoost can automatically detect interactions between variables without assuming linearity. Its superior performance was evident across various dimensions: highest AUC and Youden index in ROC curve analysis, excellent positive prediction in PR curve analysis, and strong calibration with the lowest Brier Score. DCA confirmed the model's highest clinical net benefit, highlighting its potential as a diagnostic tool. In medical diagnosis, combined prediction enhances diagnostic accuracy and comprehensiveness by integrating multiple biomarkers [31], making it particularly suitable for the diagnosis of complex diseases [32]. A key advantage of machine learning algorithms lies in their ability to process large volumes of complex data and identify patterns and associations inaccessible to humans, thereby enabling more precise diagnosis and prediction.

The XGBoost model's diagnostic tool has practical advantages. It is cost-effective, easily accessible, and based on routine blood tests, making it suitable for widespread use in medical institutions. Unlike ultrasound and MRI, which have limitations in early PCa detection [14, 15], this model offers a non-invasive and accurate alternative that can significantly reduce unnecessary prostate biopsies. Furthermore, it enhances diagnostic performance by combining multiple biomarkers, which improves risk stratification and reduces misdiagnosis rates.

A combined prediction strategy significantly enhances diagnostic performance. Guo et al. reported that combined detection of TGF- β 1, p2PSA, and PSA yielded an AUC of 0.932 for PCa diagnosis [33]. Similarly, another study demonstrated that the AUC for PCa assessment with combined multiple indicators was 0.900 [34]. These studies indicate that combined detection provides more comprehensive disease status information, reduces the risks of misdiagnosis and missed diagnosis, and

improves patient risk stratification capacity [35]. Furthermore, the model facilitates early detection and risk stratification of PCa, allowing clinicians to identify high-risk patients earlier and promptly formulate individualized diagnosis and treatment plans. Nomogram analysis demonstrated that NLR and NAR made substantial contributions to risk prediction, suggesting that clinicians should monitor changes in these two indicators. Moreover, combined prediction plays a crucial role in disease monitoring and prognostic assessment [36].

The model could also support telemedicine and AI-assisted diagnosis, facilitating collaborative diagnosis between primary and specialized hospitals.

However, this study has limitations. As a single-center retrospective study, it may have selection and information biases, and multi-center prospective studies are needed for external validation. The sample was primarily from the Chinese population, and its applicability to other ethnicities and regions remains unverified. Long-term follow-up data is also lacking, and the absence of biomarkers such as imaging features and genomic data limits the model's predictive power. Additionally, the "black box" nature of machine learning models poses challenges to clinical interpretation, and future work should focus on improving model transparency. The effects of complications, medication history, and lifestyle factors on outcomes were not fully considered. Future research should expand the sample size, include diverse populations, integrate imaging and genomic data, and enhance the model's accuracy and predictive capabilities.

Conclusion

The XGBoost machine learning model, based on NLR, LMR, and NAR, outperforms traditional PSA and PHI tests in the differential diagnosis of PCa and BPH. This model offers a cost-effective, accurate, and clinically applicable non-invasive diagnostic tool that is expected to reduce unnecessary invasive procedures and improve diagnostic precision.

Disclosure of conflict of interest

None.

Address correspondence to: Hongjun Xie, Department of Urology, The First Affiliated Hospital of Xi'an Jiaotong University, No. 277 Yanta West Road, Xi'an 710061, Shaanxi, China. Tel: +86-4000003222; E-mail: apolodite@163.com

References

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394-424.
- [2] Fu ZT, Guo XL, Zhang SW, Zheng RS, Zeng HM, Chen R, Wang SM, Sun KX, Wei WW and He J. Statistical analysis of incidence and mortality of prostate cancer in China, 2015. *Zhonghua Zhong Liu Za Zhi* 2020; 42: 718-722.
- [3] He H, Liang L, Han D, Xu F and Lyu J. Different trends in the incidence and mortality rates of prostate cancer between china and the USA: a joinpoint and age-period-cohort analysis. *Front Med (Lausanne)* 2022; 9: 824464.
- [4] Hugosson J, Månsson M, Wallström J, Axcrone U, Carlsson SV, Egevad L, Geterud K, Khatami A, Kohestani K, Pihl CG, Socratous A, Stranne J, Godtman RA and Hellström M. Prostate cancer screening with PSA and MRI followed by targeted biopsy only. *N Engl J Med* 2022; 387: 2126-2137.
- [5] Kachuri L, Hoffmann TJ, Jiang Y, Berndt SI, Shelley JP, Schaffer KR, Machiela MJ, Freedman ND, Huang WY, Li SA, Easterlin R, Goodman PJ, Till C, Thompson I, Lilja H, Van Den Eeden SK, Chanock SJ, Haiman CA, Conti DV, Klein RJ, Mosley JD, Graff RE and Witte JS. Genetically adjusted PSA levels for prostate cancer screening. *Nat Med* 2023; 29: 1412-1423.
- [6] Weinlich R, Oberst A, Beere HM and Green DR. Necroptosis in development, inflammation and disease. *Nat Rev Mol Cell Biol* 2017; 18: 127-136.
- [7] Schottenfeld D and Beebe-Dimmer J. Chronic inflammation: a common and important factor in the pathogenesis of neoplasia. *CA Cancer J Clin* 2006; 56: 69-83.
- [8] DeBerardinis RJ. Tumor microenvironment, metabolism, and immunotherapy. *N Engl J Med* 2020; 382: 869-871.
- [9] Powell DR and Huttenlocher A. Neutrophils in the tumor microenvironment. *Trends Immunol* 2016; 37: 41-52.
- [10] Ugel S, Canè S, De Sanctis F and Bronte V. Monocytes in the tumor microenvironment. *Annu Rev Pathol* 2021; 16: 93-122.
- [11] Song D, Li X and Zhang X. Expression and prognostic value of ratios of platelet lymphocyte,

- neutrophil lymphocyte and lymphocyte monocyte in breast cancer patients. *Am J Transl Res* 2022; 14: 3233-3239.
- [12] Shi Z, Zheng D, Tang X and Du Y. Correlation of immune inflammatory indices and nutritional risk index with prognosis in patients with non-small cell lung cancer. *Am J Transl Res* 2023; 15: 4100-4109.
- [13] Parker C, Castro E, Fizazi K, Heidenreich A, Ost P, Procopio G, Tombal B and Gillessen S. Prostate cancer: esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2020; 31: 1119-1134.
- [14] Ahdoot M, Wilbur AR, Reese SE, Lebastchi AH, Mehralivand S, Gomella PT, Bloom J, Gurram S, Siddiqui M, Pinsky P, Parnes H, Linehan WM, Merino M, Choyke PL, Shih JH, Turkbey B, Wood BJ and Pinto PA. MRI-targeted, systematic, and combined biopsy for prostate cancer diagnosis. *N Engl J Med* 2020; 382: 917-928.
- [15] Bulten W, Kartasalo K, Chen PC, Ström P, Pinckaers H, Nagpal K, Cai Y, Steiner DF, van Boven H, Vink R, Hulsbergen-van de Kaa C, van der Laak J, Amin MB, Evans AJ, van der Kwast T, Allan R, Humphrey PA, Grönberg H, Samaratunga H, Delahunt B, Tsuzuki T, Häkkinen T, Egevad L, Demkin M, Dane S, Tan F, Valkonen M, Corrado GS, Peng L, Mermel CH, Ruusuvaari P, Litjens G and Eklund M; Panda challenge consortium. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the panda challenge. *Nat Med* 2022; 28: 154-163.
- [16] Chan JM, Zaidi S, Love JR, Zhao JL, Setty M, Wadosky KM, Gopalan A, Choo ZN, Persad S, Choi J, LaClair J, Lawrence KE, Chaudhary O, Xu T, Masilionis I, Linkov I, Wang S, Lee C, Barlas A, Morris MJ, Mazutis L, Chaligne R, Chen Y, Goodrich DW, Karthaus WR, Pe'er D and Sawyers CL. Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science* 2022; 377: 1180-1191.
- [17] Herre M, Cedervall J, Mackman N and Olsson AK. Neutrophil extracellular traps in the pathology of cancer and other inflammatory diseases. *Physiol Rev* 2023; 103: 277-312.
- [18] Propper DJ and Balkwill FR. Harnessing cytokines and chemokines for cancer therapy. *Nat Rev Clin Oncol* 2022; 19: 237-253.
- [19] Thang NVV, Luyen LT, Vi NTT and Hai PD. Neutrophil-to-lymphocyte-to-albumin ratio as a prognostic marker for mortality in sepsis and septic shock in Vietnam. *Acute Crit Care* 2025; 40: 244-251.
- [20] Lee BH, Taylor MG, Robinet P, Smith JD, Schweitzer J, Sehayek E, Falzarano SM, Magi-Galluzzi C, Klein EA and Ting AH. Dysregulation of cholesterol homeostasis in human prostate cancer through loss of ABCA1. *Cancer Res* 2013; 73: 1211-1218.
- [21] Xiong T, Xu G, Huang XL, Lu KQ, Xie WQ, Yin K and Tu J. ATP-binding cassette transporter A1: a promising therapy target for prostate cancer. *Mol Clin Oncol* 2018; 8: 9-14.
- [22] Van Hemelrijck M, Walldius G, Jungner I, Hammar N, Garmo H, Binda E, Hayday A, Lambe M and Holmberg L. Low levels of apolipoprotein A-I and HDL are associated with risk of prostate cancer in the Swedish AMORIS study. *Cancer Causes Control* 2011; 22: 1011-1019.
- [23] Kluck GEG, Qian AS, Sakarya EH, Quach H, Deng YD and Trigatti BL. Apolipoprotein A1 protects against necrotic core development in atherosclerotic plaques: PDZK1-dependent high-density lipoprotein suppression of necroptosis in macrophages. *Arterioscler Thromb Vasc Biol* 2023; 43: 45-63.
- [24] Heidegger I, Fotakis G, Offermann A, Goveia J, Daum S, Salcher S, Noureen A, Timmer-Boscha H, Schäfer G, Walenkamp A, Perner S, Beatovic A, Moisse M, Plattner C, Krogsdam A, Haybaeck J, Sopper S, Thaler S, Keller MA, Klocker H, Trajanoski Z, Wolf D and Pircher A. Comprehensive characterization of the prostate tumor microenvironment identifies CXCR₄/CXCL₁₂ crosstalk as a novel antiangiogenic therapeutic target in prostate cancer. *Mol Cancer* 2022; 21: 132.
- [25] Joller N, Anderson AC and Kuchroo VK. LAG-3, TIM-3, and TIGIT: distinct functions in immune regulation. *Immunity* 2024; 57: 206-222.
- [26] Li Z, Li JN, Li Q, Liu C, Zhou LH, Zhang Q and Xu Y. Cholesterol efflux regulator ABCA1 exerts protective role against high shear stress-induced injury of HBMECs via regulating PI3K/Akt/eNOS signaling. *BMC Neurosci* 2022; 23: 61.
- [27] Kiebish MA, Cullen J, Mishra P, Ali A, Milliman E, Rodrigues LO, Chen EY, Tolstikov V, Zhang L, Panagopoulos K, Shah P, Chen Y, Petrovics G, Rosner IL, Sesterhenn IA, McLeod DG, Granger E, Sarangarajan R, Akmaev V, Srinivasan A, Srivastava S, Narain NR and Dobi A. Multi-omic serum biomarkers for prognosis of disease progression in prostate cancer. *J Transl Med* 2020; 18: 10.
- [28] Kluck GEG, Qian AS, Sakarya EH, Quach H, Deng YD and Trigatti BL. Apolipoprotein A1 protects against necrotic core development in atherosclerotic plaques: PDZK1-dependent high-density lipoprotein suppression of necroptosis in macrophages. *Arterioscler Thromb Vasc Biol* 2023; 43: 45-63.
- [29] Sha J, Du J, Yang J, Hu X and Li L. Changes of serum levels of tumor necrosis factor (TNF- α) and soluble interleukin-2 receptor (SIL 2R) in patients with cervical cancer and their clinical significance. *Am J Transl Res* 2021; 13: 6599-6604.

- [30] G D, Nandan SRK and Kulkarni PG. Salivary tumour necrosis factor- α as a biomarker in oral leukoplakia and oral squamous cell carcinoma. *Asian Pac J Cancer Prev* 2019; 20: 2087-2093.
- [31] Li L, Zhang F, Zhang J, Shi X, Wu H, Chao X, Ma S, Lang J, Wu M, Zhang D and Liang Z. Identifying serum small extracellular vesicle microRNA as a noninvasive diagnostic and prognostic biomarker for ovarian cancer. *ACS Nano* 2023; 17: 19197-19210.
- [32] Qiu C, Duan Y, Wang B, Shi J, Wang P, Ye H, Dai L, Zhang J and Wang X. Serum anti-pdlim1 autoantibody as diagnostic marker in ovarian cancer. *Front Immunol* 2021; 12: 698312.
- [33] Guo X, Sun X and Ye P. The diagnostic value of serum TGF- β 1, p2PSA combined with PSA in prostate cancer. *Altern Ther Health Med* 2024; 30: 184-191.
- [34] Guo T, Wang XX, Fu H, Tang YC, Meng BQ and Chen CH. Early diagnostic role of PSA combined miR-155 detection in prostate cancer. *Eur Rev Med Pharmacol Sci* 2018; 22: 1615-1621.
- [35] Boehm BE, York ME, Petrovics G, Kohaar I and Chesnut GT. Biomarkers of aggressive prostate cancer at diagnosis. *Int J Mol Sci* 2023; 24: 2185.
- [36] Kaneko M, Lenon MSL, Storino Ramacciotti L, Medina LG, Sayegh AS, La Riva A, Perez LC, Ghoreifi A, Lizana M, Jadvar DS, Lebastchi AH, Cacciamani GE and Abreu AL. Multiparametric ultrasound of prostate: role in prostate cancer diagnosis. *Ther Adv Urol* 2022; 14: 17562872221145625.

Serum PSA and inflammatory biomarkers in prostate cancer differential diagnosis

Table S1. Brier Scores of five models across training, validation, and test sets with pairwise wilcoxon comparisons

Dataset	Model	Mean Brier Score	Lowest Model(s)	Significant comparisons (Wilcoxon, Holm-adjusted)
Training	Logistic	0.0624		XGB<Logistic (P<0.001), XGB<Rpart (P<0.001), XGB<RF (P<0.001), XGB<SVM (P<0.001)
	Rpart (DT)	0.067		
	RF	0.0435		
	SVM	0.0548		
	XGBoost	0.0326	√	
Validation	Logistic	0.0952		XGB<Logistic (P<0.001), XGB<Rpart (P<0.001), XGB<RF (P=0.002); XGB vs SVM (P=0.36, n.s.)
	Rpart (DT)	0.0847		
	RF	0.0826		
	SVM	0.0751	√ (with XGB)	
	XGBoost	0.0742	√ (with SVM)	
Test	Logistic	0.0794		XGB<Logistic (P<0.001), XGB<Rpart (P<0.001), XGB<RF (P=0.04); XGB vs SVM (P=0.08, n.s.)
	Rpart (DT)	0.1103		
	RF	0.0642		
	SVM	0.064	√ (with XGB)	
	XGBoost	0.0561	√ (with SVM)	

DT: Decision Tree, RF: Random Forest, SVM: Support Vector Machine, and XGBoost: Extreme Gradient Boosting.