



OPEN ACCESS

EDITED BY

Anthony Chi Fai Ng,
The Chinese University of Hong Kong,
Hong Kong SAR, China

REVIEWED BY

Albino Eccher,
Integrated University Hospital Verona, Italy
Ugo Giovanni Falagario,
University of Foggia, Italy

*CORRESPONDENCE

Bin Fu

✉ urofbin@163.com

Xinxi Deng

✉ sudadengxinxi2011@163.com

Ru Chen

✉ chenru99999@126.com

†These authors have contributed equally to this work

RECEIVED 02 February 2023

ACCEPTED 24 May 2023

PUBLISHED 08 June 2023

CITATION

Liu T, Zhang X, Chen R, Deng X and Fu B (2023) Development, comparison, and validation of four intelligent, practical machine learning models for patients with prostate-specific antigen in the gray zone. *Front. Oncol.* 13:1157384. doi: 10.3389/fonc.2023.1157384

COPYRIGHT

© 2023 Liu, Zhang, Chen, Deng and Fu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development, comparison, and validation of four intelligent, practical machine learning models for patients with prostate-specific antigen in the gray zone

Taobin Liu^{1,2†}, Xiaoming Zhang^{1†}, Ru Chen^{1*}, Xinxi Deng^{3*} and Bin Fu^{1,2*}

¹Department of Urology, The First Affiliated Hospital of Nanchang University, Nanchang, China, ²Jiangxi Institute of Urology, Nanchang, Jiangxi, China, ³Department of Urology, Jiu Jiang NO.1 People's Hospital, Jiujiang, China

Purpose: Machine learning prediction models based on LogisticRegression, XGBoost, GaussianNB, and LGBMClassifier for patients in the prostate-specific antigen gray zone are to be developed and compared, identifying valuable predictors. Predictive models are to be integrated into actual clinical decisions.

Methods: Patient information was collected from December 01, 2014 to December 01, 2022 from the Department of Urology, The First Affiliated Hospital of Nanchang University. Patients with a pathological diagnosis of prostate hyperplasia or prostate cancer (any PCa) and having a prostate-specific antigen (PSA) level of 4–10 ng/mL before prostate puncture were included in the initial information collection. Eventually, 756 patients were selected. Age, total prostate-specific antigen (tPSA), free prostate-specific antigen (fPSA), fPSA/tPSA, prostate volume (PV), prostate-specific antigen density (PSAD), (fPSA/tPSA)/PSAD, and the prostate MRI results of these patients were recorded. After univariate and multivariate logistic analyses, statistically significant predictors were screened to build and compare machine learning models based on LogisticRegression, XGBoost, GaussianNB, and LGBMClassifier to determine more valuable predictors.

Results: Machine learning prediction models based on LogisticRegression, XGBoost, GaussianNB, and LGBMClassifier exhibit higher predictive power than individual metrics. The area under the curve (AUC) (95% CI), accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score of the LogisticRegression machine learning prediction model were 0.932 (0.881–0.983), 0.792, 0.824, 0.919, 0.652, 0.920, and 0.728, respectively; of the XGBoost machine learning prediction model were 0.813 (0.723–0.904), 0.771, 0.800, 0.768, 0.737, 0.793 and 0.767, respectively; of the GaussianNB machine learning prediction model were 0.902 (0.843–0.962), 0.813, 0.875, 0.819, 0.600, 0.909, and 0.712, respectively; and of the LGBMClassifier machine learning prediction model were 0.886 (0.809–0.963), 0.833, 0.882, 0.806, 0.725, 0.911, and 0.796, respectively. The LogisticRegression machine learning prediction

model has the highest AUC among all prediction models, and the difference between the AUC of the LogisticRegression prediction model and those of XGBoost, GaussianNB, and LGBMClassifier is statistically significant ($p < 0.001$).

Conclusion: Machine learning prediction models based on LogisticRegression, XGBoost, GaussianNB, and LGBMClassifier algorithms exhibit superior predictability for patients in the PSA gray area, with the LogisticRegression model yielding the best prediction. The aforementioned predictive models can be used for actual clinical decision-making.

KEYWORDS

prostate cancer, gray area PSA, machine learning, LogisticRegression, XGBoost, GaussianNB, LGBMClassifier

Introduction

Prostate cancer (PCa) is an extremely common malignancy in men, its incidence being second only to lung cancer among all malignant tumors in men (1). Approximately 1.31 million new PCa cases are diagnosed each year globally, and nearly 10 million men currently have PCa. Out of these 10 million patients, approximately 700,000 suffer from metastatic PCa, which causes nearly 400,000 deaths each year, a death toll that is expected to more than double by 2040 (2). Therefore, it is becoming increasingly important to screen more accurately for PCa at an early stage.

Prostate specific-antigen (PSA) is a specific secretion of the prostate gland, which is a key marker of the initial diagnosis of the disease (3). However, this serum PSA level is influenced by various factors, such as ejaculation, acute prostatitis, urinary retention, catheterization, digital rectal examination (DRE), cystoscopy, and additional operations, thus affecting the accuracy of PSA diagnosis for patients (4). Clinically, PSA within the range of 4–10 ng/mL is defined as being in the gray zone (5). The possibility of PCa in the Chinese population with PSA in this range is about 25%, and the same is approximately 40% with the world population (6), and these patients are more likely to be missed for PCa or to have a non-essential puncture biopsy. In a study with 2426 patients, Xu B et al. found that the positive rate of PCa biopsy was only 25.77% for patients in the PSA gray zone (7). The diagnosis of PCa in the gray zone of PSA has become a key issue in urological research.

Machine learning can independently replicates human cognition and make decisions based on its perceived environment to achieve predetermined goals (8). Zhang S et al. proposed an interpretable deep learning framework for describing and interpreting human brain states (9). Furthermore, Lee C et al. developed a novel machine learning model for predicting non-metastatic PCa in men based on the Surveillance, Epidemiology, and End Results (SEER) database (10). Madhur Nayan et al. used machine learning to predict the progress of active PCa surveillance (11). Nevertheless, to date, only a few studies have used machine

learning to forecast PCa with PSA in the gray zone. In the current study, four machine learning algorithms were compared.

Materials and methods

Data collection, the introduction of relevant variables, and handling of missing values

Information was collected from December 01, 2014 to December 01, 2022 from the Department of Urology, The First Affiliated Hospital of Nanchang University, from patients with a pathological diagnosis of prostate hyperplasia or PCa (any PCa) and prior PSA of 4–10 ng/mL. According to the exclusion and inclusion criteria followed in this study, 756 patients were finally selected (596 patients with prostate hyperplasia and 160 patients with PCa). Age, free prostate-specific antigen (fPSA), total prostate-specific antigen (tPSA), fPSA/tPSA, prostate volume (PV), prostate-specific antigen density (PSAD), (fPSA/tPSA)/PSAD, absolute neutrophil count (NEUT), lymphocyte absolute count (LYM), platelet count (PLT), neutrophil absolute count/lymphocyte absolute count (NLR), absolute platelet count/absolute lymphocyte count (PLR), alkaline phosphatase (ALP), and prostate MRI results were recorded for the samples.

The MRI findings of the prostate were abnormal, the abnormality being defined as PI-RADS score ≥ 3 . The PI-RADS score is the combined clinical presentation of mpMRI (multiparametric MRI) with DWI (diffusion-weighted imaging), T2WI (T2-weighted imaging of the prostate), and DCE (dynamic contrast enhancement). This score has considerable clinical significance in the diagnosis of PCa (12).

The cases were divided into prostate hyperplasia and PCa groups based on the pathological findings. Most of the sample data of this study were relatively complete, and the primary limitation was the presence of PSA only in some samples that did

not contain fPSA data. However, the missing data were padded using the random forest algorithm.

Inclusion and exclusion criteria

Inclusion criteria: 1, PSA in the range of 4–10 ng/mL; 2, a clear pathological diagnosis and no history of prostate surgery; 3, the relevant prediction metrics are complete. Exclusion criteria: 1, patients with acute prostatitis, acute urinary retention, within 48 h of cystoscopy and catheterization and within 1 week of DRE; 2, patients taking drugs that may affect outcomes, such as 5-alpha reductase inhibitors; 3, patients having more than two prostate punctures.

Predictive variable screening

First, the overall distribution of the 14 variables included in this study was analyzed, and then a univariate logistic analysis was performed on each variable to screen for statistically significant predictors. After the Collinearity Diagnostics of predictor variables, multivariate logistic analysis was performed after screening out variables with no covariance and multiple indicators with independent influences. Machine learning models were selected to be incorporated for modeling in conjunction with actual clinical situations.

Building machine learning models

Sample imbalance in machine learning refers to an imbalance in the number of samples per category in a dataset. When building the model, it is biased towards samples with larger proportions, resulting in a low generalization ability. The ideal ratio of positive to negative samples for building the model is 1:1. In this study, the sample data for the PCa group: and prostate hyperplasia group was 160:596, and therefore, the sample showed an imbalance. Sample imbalance commonly originates from an over-sampling of under-represented categories or under-sampling of over-represented categories. Herein, over-sampling the PCa group, which is a small percentage of the total population, resulted in less accurate true data. To address this issue, the prostate hyperplasia group was randomly under-sampled, thereby altering the PCa group: prostate hyperplasia group to 160:320 (1:2).

To reduce model overfitting and ensure its stability and generalization ability, the total number of samples is considered to be low. The model is thus constructed using five-fold cross-validation on the dataset, which is divided into five parts. Each part is uncrossed and all the parts are of the same size. In descending order, four of the five replicas are selected as the training set and one as the validation set, and five separate model training and validation runs were performed. The final results of the five validation runs

were averaged as the validation error of the model. The flow chart of this research is shown in Figure 1.

Statistical analysis

The statistical analysis of the sample data was initially performed using SPSS 26.0 software. Continuous and categorical variables were expressed as mean \pm standard deviation and percentage (%), respectively. Independent sample t-tests were performed to determine the differences between the groups. Differences in the AUC values of machine learning algorithms between groups were compared using De Lon's non-parametric approach (4). The machine learning algorithms were constructed and compared using python version 3.10.

Results

After filtering according to the inclusion and exclusion criteria, a total of 765 cases were collected, including 596 in the prostate hyperplasia group and 160 in the PCa group. The case information for both groups is shown in Table 1. The comparison yielded statistically significant differences in the 11 predictors of tPSA, fPSA, fPSA/tPSA, PV, PSAD, (fPSA/tPSA)/PSAD, ALP, NLR, PLR, NEUT, and prostate MRI abnormalities. Age, LYM, and PLT were not statistically significant between the two groups. The overall mean of tPSA, PSAD, NLR, PLR, and ALP in the PCa group was greater than that in the prostate hyperplasia group ($p < 0.05$). Conversely, the overall mean of the fPSA, fPSA/tPSA, PV, and (fPSA/tPSA)/PSAD in the PCa group were smaller than those in the prostate hyperplasia group ($p < 0.05$).

Univariate logistic analysis was performed on the aforementioned statistically significant predictors, the results of which are presented in Table 2. In addition, the ROC curves for each predictor variable are shown in Figure 2. The AUC for the predictive efficacy of prostate MRI abnormalities was 0.839, and the AUCs for (fPSA/tPSA)/PSAD and PSAD were 0.818 and 0.769, respectively.

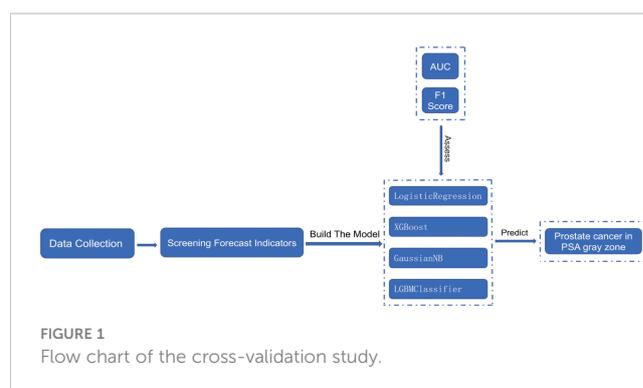


TABLE 1 Distribution of variables and comparison of differences between the two groups of cases.

Projects	Prostatic hyperplasia group (N = 596)	Prostate cancer group (N = 160)	p-value
Age (years)	74.324 ± 6.984	74.250 ± 3.162	0.846
tPSA(ng/ml)	6.569 ± 1.502	7.061 ± 1.186	<0.001
fPSA(ng/ml)	1.501 ± 0.778	1.071 ± 0.338	<0.001
fPSA/tPSA	0.226 ± 0.095	0.154 ± 0.049	<0.001
PV(cm ³)	60.089 ± 14.446	46.774 ± 12.594	<0.001
PSAD(ng/ml. cm ³)	0.116 ± 0.038	0.166 ± 0.062	<0.001
(fPSA/tPSA)/PSAD	2.162 ± 1.057	1.085 ± 0.554	<0.001
NEUT(*10 ⁹ /L)	4.730 ± 1.839	5.135 ± 0.713	<0.001
LYM(*10 ⁹ /L)	1.396 ± 0.543	1.333 ± 0.612	0.235
PLT(*10 ⁹ /L)	209.865 ± 65.421	212.410 ± 46.315	0.576
ALP(U/L)	89.432 ± 24.840	96.804 ± 19.687	<0.001
NLR	3.817 ± 2.088	4.511 ± 1.627	<0.001
PLR	166.984 ± 71.286	189.305 ± 84.711	0.003
MRI abnormalities[n(%)]			
no(0)	497(83.389%)	25(15.625%)	<0.001
yes(1)	99(16.611%)	135(84.375%)	

NEUT, absolute neutrophil count; LYM, lymphocyte absolute count; PLT, platelet count; ALP, alkaline phosphatase; NLR, neutrophil absolute count/lymphocyte absolute count; PLR, absolute platelet count/absolute lymphocyte count.

The bold numbers indicate the statistical differences.

The results of the Collinearity Diagnostics for the above variables are given in Table 3. Among these, PSAD, NLR, PLR, NEUT, MRI, and ALP are variables with covariance index VIF values less than 10, indicating no covariance.

After including variables with VIF<10 for multivariate logistic analysis, PASD, and MRI variables were found to have p-values less than 0.001, and PASD and MRI were determined to be the independent risk factors for outcome variables (Table 4).

Results of machine learning models for LogisticRegression; XGBoost; GaussianNB; LGBMClassifier

In the overall sample, N = 120 cases (25.00%) were randomly selected as the test set. The remaining samples were used as the training set for five-fold cross-validation. ROC curves for the validation and test sets of LogisticRegression, XGBoost,

TABLE 2 Results of univariate logistic analysis.

Projects	AUC	Sensitivity	Specificity	Youden's index	Optimal threshold
tPSA	0.609	0.850	0.404	0.254	6.010
fPSA5	0.664	0.396	0.925	0.321	1.570
fPSA/tPSA	0.748	0.676	0.819	0.495	0.190
PV	0.748	0.532	0.831	0.363	59.190
PSAD	0.769	0.644	0.743	0.387	0.140
NEUT	0.662	0.863	0.559	0.421	4.530
(fPSA/tPSA)/PSAD	0.818	0.624	0.925	0.549	1.700
ALP	0.632	0.581	0.721	0.303	95.650
NLR	0.650	0.594	0.716	0.310	4.160
PLR	0.574	0.588	0.638	0.225	170.520
MRI	0.839	0.844	0.834	0.678	1.000

NEUT, absolute neutrophil count; ALP, alkaline phosphatase; NLR, neutrophil absolute count/lymphocyte absolute count; PLR, absolute platelet count/absolute lymphocyte count.

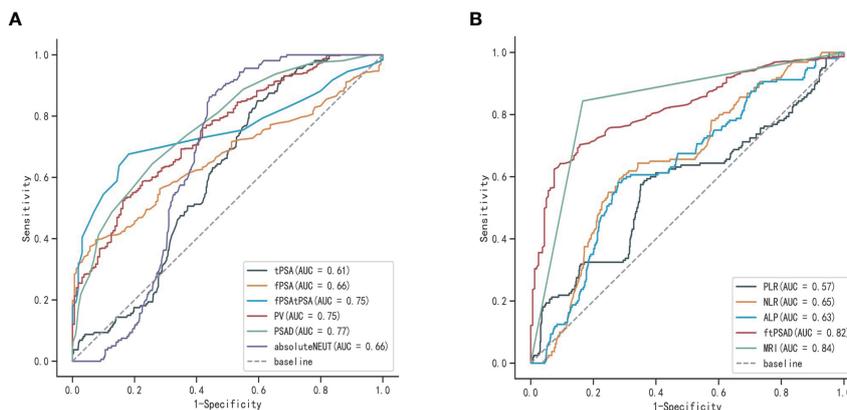


FIGURE 2 ROC curves for each predictor variable: **(A)** ROC curves of tPSA, fPSA, fPSA/tPSA, PV, PSAD, and ENEUT; **(B)** ROC curves of (fPSA/tPSA)/PSAD, ALP, NLR, PLR, and prostate MRI abnormalities.

GaussianNB, and LGBMClassifier are shown in Figure 3. The AUC values of their validation sets were 0.918 (0.855–0.979), 0.893 (0.814–0.971), 0.906 (0.839–0.974), and 0.893 (0.814–0.972), respectively. Furthermore, the AUC values of the measured sets were 0.935 (0.893–0.977), 0.933 (0.891–0.976), 0.935 (0.893–0.977), and 0.939 (0.899–0.980), respectively (Table 5). Since the AUC of the validation set did not exceed that of the test set, or the excess ratio was less than 10%, the fit was considered successful. All the aforementioned models can be used for this dataset.

Categorical multi-model construction and the comparison of respective ROC curves

The best performer in the training set was XGBoost, while that in the validation set was LogisticRegression (both ranked according to AUC) (Figure 4). A comparison of the training and validation sets

shows that the results do not match (Table 6). XGBoost most likely resulted in overfitting, while LogisticRegression demonstrated relatively excellent stability and is therefore recommended as the optimal prediction model. From the change of AUC during the training of each model, it is observed that the AUCs of the training and validation sets are finally stable and well-matched above 0.8 (Figure 5).

DeLong test for categorical multi-model AUC values

The DeLong test is used to verify whether the AUC values of the two sets of machine learning models are statistically different (Table 7). From Table 7, it is evident that the difference between LogisticRegression and the three machine learning models XGBoost, GaussianNB, and LGBMClassifier is significant ($p < 0.001$). The remaining machine learning models do not differ from each other in a statistically significant way.

TABLE 3 Covariance VIF values for each predictor variable.

Projects	VIF
fPSA/tPSA	57.783
fPSA	35.782
fPSAD	19.838
PV	11.935
tPSA	10.932
PSAD	9.452
NLR	3.454
PLR	2.265
NEUT	2.039
MRI	1.113
ALP	1.059

NEUT, absolute neutrophil count; ALP, alkaline phosphatase; NLR, neutrophil absolute count/lymphocyte absolute count; PLR, absolute platelet count/absolute lymphocyte count. Bold numbers indicate items included in subsequent studies.

Discussion

PCa is one of the most prevalent cancers in the world (13). Approximately 25% of patients with PSA between 4.0 and 10.0 ng/mL were found to have PCa, compared with approximately 40 percent in the United States (14). Currently, artificial intelligence is widely used in medicine. Machine learning algorithms accept predictor variable metrics as inputs and provide variable metrics as outcomes and build appropriate machine learning models. Machine learning can also predict the current input data based on an existing model, which considers the prediction metrics and considerably improves the prediction accuracy. This technique exhibits a better prediction performance than individual prediction metrics, so its application prospects are numerous (15, 16). Related studies have demonstrated the tremendous value of machine learning models for PCa diagnosis and the assessment of prognosis (17–20). For example, Satoshi Nitta et al. used age, PSA, PV, and urine leukocyte count as predictors to build a PCa

TABLE 4 Results of multivariate logistic analysis of predictor variables.

Projects	Estimate	SE	Z	p
NEUT	0.103	0.107	0.962	0.336
ALP	0.009	0.005	1.829	0.067
PLR	0.002	0.002	0.976	0.329
NLR	-0.056	0.116	-0.482	0.63
PSAD	25.198	3.187	7.906	<0.001
MRI	3.542	0.294	12.047	<0.001

NEUT, absolute neutrophil count; ALP, alkaline phosphatase; NLR, neutrophil absolute count/lymphocyte absolute count; PLR, absolute platelet count/absolute lymphocyte count. The bold numbers mean statistically different. Bold numbers indicate items included in subsequent studies.

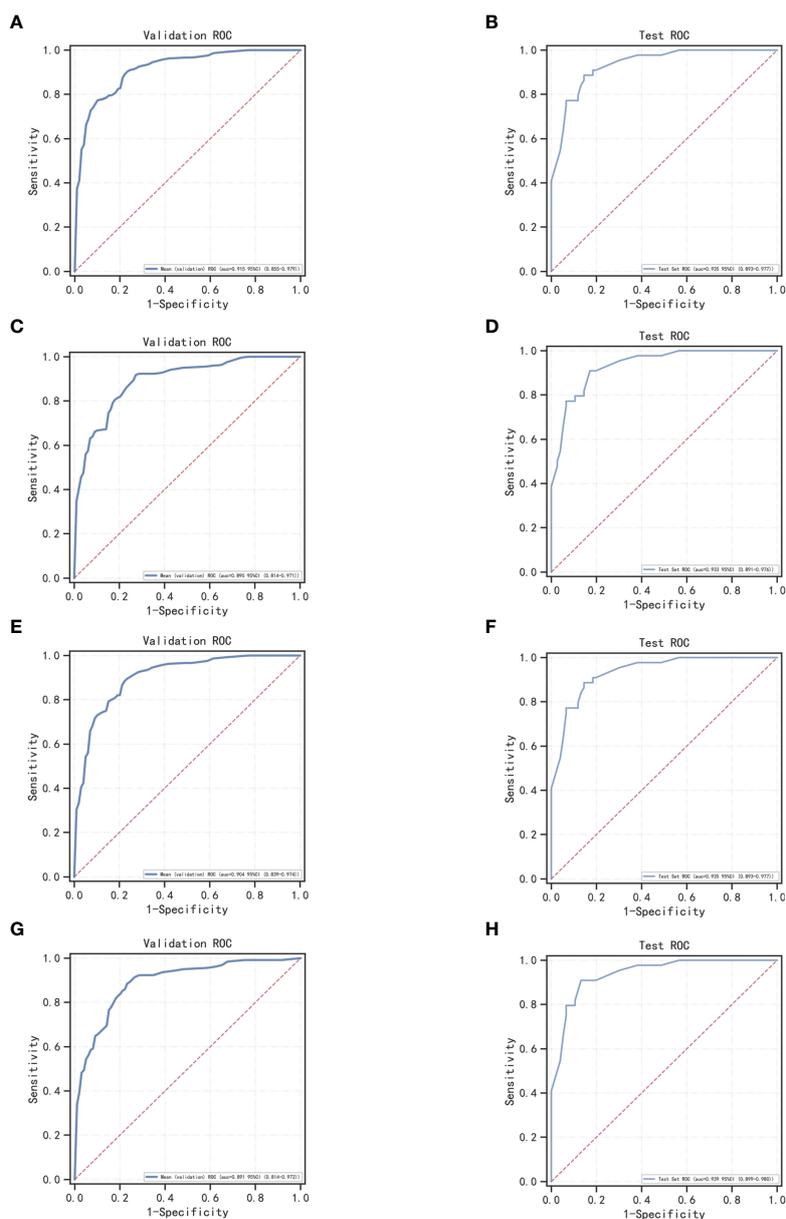


FIGURE 3 ROC curves for machine learning models for LogisticRegression, XGBoost, GaussianNB, and LGBMClassifier; (A, C, E, G) ROC curves of four machine learning methods on the validation set; (B, D, F, H) ROC curves of four machine learning methods applied on the test set.

TABLE 5 Results of the validation and test sets of four machine learning models.

	Data set	AUC (95% CI)	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1 score
LogisticRegression	Validation set	0.918 (0.855–0.979)	0.811	0.88	0.844	0.666	0.914	0.758
	Test set	0.935 (0.893–0.977)	0.875	0.886	0.855	0.872	0.877	0.879
XGBoost	Validation set	0.893 (0.814–0.971)	0.792	0.914	0.762	0.659	0.88	0.765
	Test set	0.933 (0.891–0.976)	0.85	0.909	0.829	0.81	0.872	0.856
GaussianNB	Validation set	0.906 (0.839–0.974)	0.794	0.906	0.799	0.64	0.922	0.749
	Test set	0.935 (0.893–0.977)	0.875	0.886	0.855	0.872	0.877	0.879
LGBMClassifier	Validation set	0.893 (0.814–0.972)	0.8	0.914	0.766	0.67	0.893	0.771
	Test set	0.939 (0.899–0.980)	0.883	0.909	0.868	0.875	0.888	0.892

prediction model based on Random Forest, Support Vector Machines (SVM), and artificial neural networks (ANNs) (21). Wouter Bulten et al. built a machine learning model to grade the Gleason score and compared it with professional pathologists. The results indicated that this model had diagnostic values comparable to professional pathologists (22). Simon P Hood et al. developed an ensemble machine learning predictive model to predict low and high-risk PCa (23). Further, Patrick Schelb et al. classified PCa

based on MRI by machine learning combined with PI-RADS scoring (24). Similarly, numerous studies of machine learning in the classification of PCa and prostate enlargement have been conducted to date (25–30). Fewer studies, however, have targeted patients in the gray zone of PSA. In addition, the machine learning model built for the overall range of PSA is not well suited for PCa with PSA in the gray zone. Meanwhile, in their study of the utility and limitations of cadaveric examination in organ transplantation,

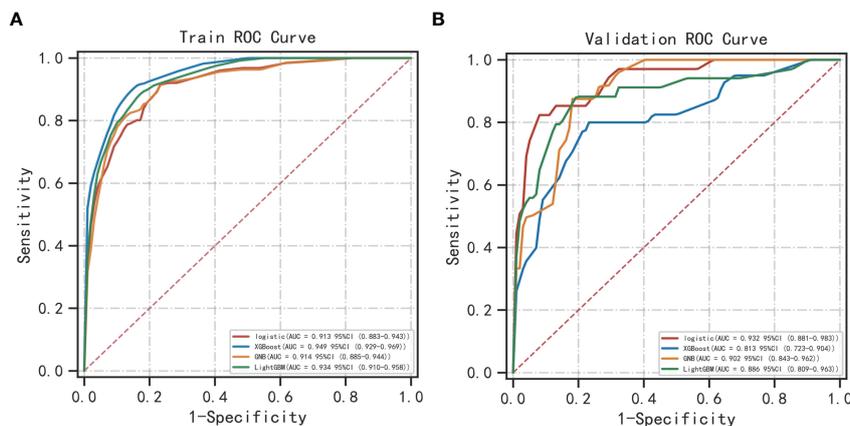


FIGURE 4 ROC curves of categorical multivariate models: (A) Training sets; (B) Validation sets.

TABLE 6 Results of the machine learning categorical multi-model.

Training sets							
Models	AUC(95% CI)	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1 score
LogisticRegression	0.913 (0.883–0.943)	0.815	0.913	0.771	0.663	0.935	0.768
XGBoost	0.949 (0.929–0.969)	0.872	0.900	0.852	0.752	0.942	0.819
GaussianNB	0.914 (0.885–0.944)	0.854	0.824	0.863	0.790	0.890	0.806
LGBMClassifier	0.934 (0.910–0.958)	0.852	0.881	0.833	0.732	0.928	0.799
Validation sets							
Models	AUC(95% CI)	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1 score
LogisticRegression	0.932 (0.881–0.983)	0.792	0.824	0.919	0.652	0.920	0.728
XGBoost	0.813 (0.723–0.904)	0.771	0.800	0.768	0.737	0.793	0.767
GaussianNB	0.902 (0.843–0.962)	0.813	0.875	0.819	0.600	0.909	0.712
LGBMClassifier	0.886 (0.809–0.963)	0.833	0.882	0.806	0.725	0.911	0.796

Desley Neil et al. (31, 32) found that although PCa is incidentally detected at the cadaveric examination of organ donors, the likelihood of this disease increases with donor age. Thus, this poses new requirements and challenges for PCA diagnosis. It is to be noted that machine learning is particularly important in the diagnosis of PCa in organ transplantation.

The differences in the 11 predictors of tPSA, fPSA, fPSA/tPSA, PV, PSAD, (fPSA/tPSA)/PSAD, ALP, NLR, PLR, NEUT#, and prostate MRI abnormalities were statistically significant as seen in the general patient data. Notably, while the mean age of the PCa group was greater than that of the prostate enlargement group, the difference was not statistically significant. The mean PV in the PCa group was also smaller

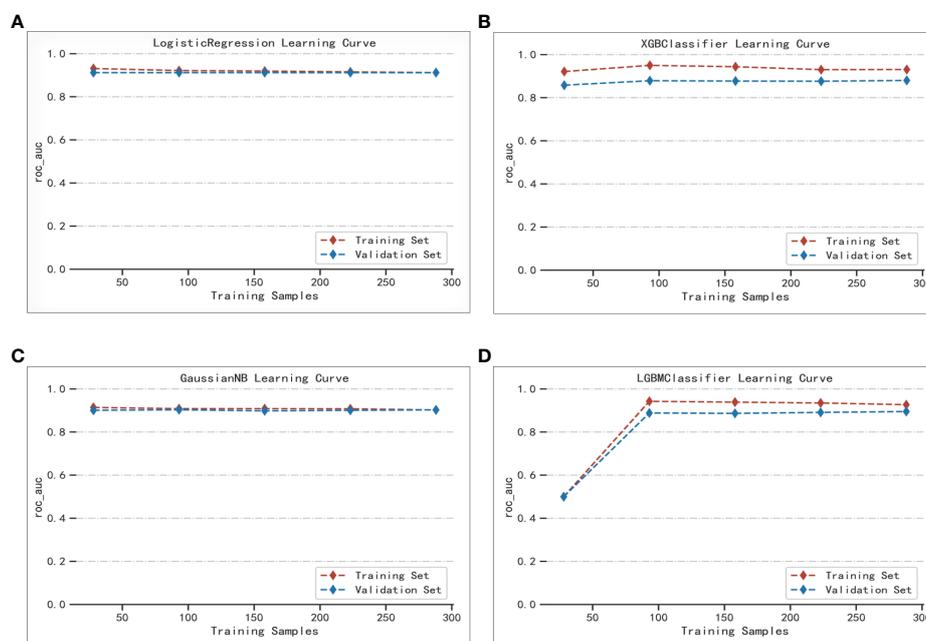


FIGURE 5 Extent of matching of the validation and test sets of the four machine learning methods: (A) LogisticRegression; (B) XGBoost; (C) GaussianNB; (D) LGBMClassifier.

TABLE 7 Results of Delong test.

LogisticRegression	XGBoost	GaussianNB	LGBMClassifier
NaN	<0.001	<0.001	<0.001
<0.001	NaN	0.667	0.769
<0.001	0.667	NaN	0.429
<0.001	0.769	0.429	NaN

NaN, Not a Number.

The bold numbers indicate statistically different.

than that in the prostate hyperplasia group, which is consistent with the overall range of the relevant literature in which PV in PCa is larger than that in prostate hyperplasia. This may be owing to the error caused by the limited sample size. In the univariate logistic analysis of predictive variables, the AUC of prostate MRI abnormalities to determine PCa was 0.84. This predictor has an elevated predictive efficacy for PCa. In addition, the prediction performances for (fPSA/tPSA)/PSAD and PSAD are also better, and this is consistent with the findings of ZhangYiyang et al. (30). Since PSA is influenced by numerous factors and the variability is not particularly significant when PSA is in the gray zone, the use of PSAD is more valuable than that of PSA in terms of predictive performance. As claimed by S. M. Bruno et al. (33) in a prospective study, PASD is a strong predictor of PCa because it can exclude the influence of elevated PSA due to prostatitis on the diagnosis. Therefore, the inclusion of PASD in this study is also essential. Since all fPSA/tPSA and (fPSA/tPSA)/PSAD are derived indicators, they exhibited covariance in the covariance analysis. After multivariate logistic analysis of tPSA, NLR, PLR, ALP, PSAD, and prostate MRI predictors, the difference in the predictive index of PSAD and prostate MRI abnormalities was significant ($p < 0.001$), and PSAD and prostate MRI abnormalities were the independent risk factors for PCa with PSA in the gray zone. Owing to the development of reflectomics, the future application of radiomics in the PCa diagnosis bears potential, as described by M. Ferro et al. (34) in their review on the application of radiomics in PCa. Although the study of radiomics is primarily focused on mpMRI, it integrates several other factors to render the PCa diagnosis more accurate and sensitive. Unfortunately, only mpMRI was used as a predictor in the current study.

Herein, the validation sets for the four machine learning models have AUC values of 0.918, 0.893, 0.906, and 0.893, accuracies of 0.811, 0.792, 0.794, and 0.800, and F1 scores of 0.758, 0.765, 0.749, and 0.771, respectively. This demonstrates the better performance of LogisticRegression over that of GaussianNB and LGBMClassifier on the validation sets, and the performance of LGBMClassifier is comparable to that of XGBoost. In the classification multi-model validation, the LogisticRegression model demonstrated the best predictive efficacy (AUC = 0.932) and the difference with XGBoost, GaussianNB, and LGBMClassifier was statistically significant ($p < 0.001$). Bermejo et al. showed that a single indicator, PSA, is not adequately effective in diagnosing PCa. Nevertheless, when logistic regression and decision tree models are built by combining age, DRE, PSA, mpMRI, and other relevant indicators, both models achieve better accuracy (35). All the aforementioned studies, although different from the subject of this study, illustrate the remarkable value of machine learning models in PCa diagnosis and prediction

capabilities. Previous studies on the diagnosis of PCa were compared using other predictive methods, such as I. M. Perez et al.'s biparametric MRI (bpMRI) as the main predictor (36). The AUC values of the resulting prediction models in the development and validation queues were 0.83 (0.77–0.89) and 0.80 (0.75–0.85), lower than those of the machine learning models in this study. In another study, S. Parekh et al. (37) used three predictive model covers (PROMOD, ERSPC RC, and ANN) with AUC values of any PCa (0.82 vs 0.70 vs 0.90) and csPCa (0.82 vs 0.78 vs 0.92). All these values are lower than the AUC values of the machine learning models in the current study, especially the LogisticRegression algorithm.

Although a total of 756 cases were included in this study, the sample size is still limited and prone to bias. Moreover, although the sample number ratio of the prostate hyperplasia group to the PCa group was maintained at 1:2 to ensure a balanced sample, the total sample size being modest, it was prone to problems such as overfitting and poor generalization ability. Therefore, an increase in sample size was needed to ensure the reliability of the results. In addition, information about prostate puncture, such as the number of needles in the puncture, transperineal, or transrectal, and whether MRI ultrasound fusion was performed or not (38), are not recorded in this study, which should be included in subsequent studies.

Conclusion

Machine learning prediction models based on LogisticRegression, XGBoost, GaussianNB, and LGBMClassifier algorithms for PCa with PSA in the gray region exhibit remarkable predictive power, and the LogisticRegression machine learning model has higher predictive performance than the rest of the models. The aforementioned machine learning models can be used for practical clinical decision-making to assist physicians in the personalized diagnosis and treatment of patients.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Ethics statement

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for

participation was not required for this study in accordance with the national legislation and institutional requirements.

Author contributions

The data was collected by XZ. TL conceived and wrote the manuscript. XZ and TL worked together on the tables and figures. RC provides guidance in statistical analysis. RC and BF provided input into the revision of the manuscript. XD made changes to the language of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of P. R. China (Grant Nos. 81560419, 81960512, and 81760457), Jiangxi Provincial “Double Thousand Plan” Fund Project (Grant No. jxsq2019201027), and Youth Project of Natural Science Foundation of Jiangxi Province (20212BAB216037).

References

- Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO classification of tumours of the urinary system and Male genital organs-part b: prostate and bladder tumours. *Eur Urol* (2016) 70(1):106–19. doi: 10.1016/j.eururo.2016.02.028
- Yamada Y, Beltran H. The treatment landscape of metastatic prostate cancer. *Cancer Lett* (2021) 519:20–9. doi: 10.1016/j.canlet.2021.06.010
- Van Poppel H, Albrecht T, Basu P, Hogenhout R, Collen S, Roobol M. Serum PSA-based early detection of prostate cancer in Europe and globally: past, present and future. *Nat Rev Urol* (2022) 19(9):562–72. doi: 10.1038/s41585-022-00638-6
- Ito K. Editorial comment: factor that influence serum prostate-specific antigen levels in terms of maintaining quality within a PSA screening system. *Eur Urol* (2007) 52(3):714.
- Liu F, Shi X, Wang F, Han S, Chen D, Gao X, et al. Evaluation and multi-institutional validation of a novel urine biomarker lncRNA546 to improve the diagnostic specificity of prostate cancer in PSA gray-zone. *Front Oncol* (2022) 12:946060. doi: 10.3389/fonc.2022.946060
- Tao T, Shen D, Yuan L, Zeng A, Xia K, Li B, et al. Establishing a novel prediction model for improving the positive rate of prostate biopsy. *Transl Androl Urol* (2020) 9(2):574–82. doi: 10.21037/tau.2019.12.42
- Xu B, Li G, Kong C, Chen M, Hu B, Jiang Q, et al. A multicenter retrospective study on evaluation of predicative factors for positive biopsy of prostate cancer in real-world setting. *Curr Med Res Opin* (2021) 37(9):1617–25. doi: 10.1080/03007995.2021.1949270
- Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* (2019) 16(7):391–403. doi: 10.1038/s41585-019-0193-3
- Zhang S, Wang J, Yu S, Wang R, Han J, Zhao S, et al. An explainable deep learning framework for characterizing and interpreting human brain states. *Med Image Anal* (2023) 83:102665. doi: 10.1016/j.media.2022.102665
- Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanapragasam VJ. Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the surveillance, epidemiology, and end results (SEER) database. *Lancet Digit Health* (2021) 3(3):e158–65. doi: 10.1016/S2589-7500(20)30314-9
- Nayan M, Salari K, Bozzo A, Ganglberger W, Lu G, Carvalho F, et al. A machine learning approach to predict progression on active surveillance for prostate cancer. *Urol Oncol* (2022) 40(4):161.e161–161.e167. doi: 10.1016/j.urolonc.2021.08.007
- Klotz L. Point: active surveillance for favorable risk prostate cancer. *J Natl Compr Canc Netw* (2007) 5(7):e93–8. doi: 10.6004/jnccn.2007.0060
- Komura K, Sweeney CJ, Inamoto T, Ibuki N, Azuma H, Kantoff PW. Current treatment strategies for advanced prostate cancer. *Int J Urol* (2018) 25(3):220–31. doi: 10.1111/iju.13512
- Chen R, Sjoberg DD, Huang Y, Xie L, Zhou L, He D, et al. Prostate biopsy collaborative G: prostate specific antigen and prostate cancer in Chinese men

Acknowledgments

The authors would like to thank all the investigators who contributed to this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

undergoing initial prostate biopsies compared with Western cohorts. *J Urol* (2017) 197(1):90–6. doi: 10.1016/j.juro.2016.08.103

15. Hou Y, Bao ML, Wu CJ, Zhang J, Zhang YD, Shi HB. A machine learning-assisted decision-support model to better identify patients with prostate cancer requiring an extended pelvic lymph node dissection. *BJU Int* (2019) 124(6):972–83. doi: 10.1111/bju.14892

16. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med* (2019) 25(1):24–9. doi: 10.1038/s41591-018-0316-z

17. Cysouw MCF, Jansen BHE, van de Brug T, Oprea-Lager DE, Pfahler E, de Vries BM, et al. Machine learning-based analysis of [(18)F]DCFPyL PET radiomics for risk stratification in primary prostate cancer. *Eur J Nucl Med Mol Imaging* (2021) 48(2):340–9. doi: 10.1007/s00259-020-04971-z

18. Seneviratne MG, Banda JM, Brooks JD, Shah NH, Hernandez-Boussard TM. Identifying cases of metastatic prostate cancer using machine learning on electronic health records. *AMIA Annu Symp Proc* (2018) 2018:1498–504.

19. Syed K, Sleeman W, Soni P, Hagan M, Palta J, Kapoor R, et al. Machine-learning models for multicenter prostate cancer treatment plans. *J Comput Biol* (2021) 28(2):166–84. doi: 10.1089/cmb.2020.0188

20. Barlow H, Mao S, Khushi M. Predicting high-risk prostate cancer using machine learning methods. *Data* (2019) 4(3):129. doi: 10.3390/data4030129

21. Nitta S, Tsutsumi M, Sakka S, Endo T, Hashimoto K, Hasegawa M, et al. Machine learning methods can more efficiently predict prostate cancer compared with prostate-specific antigen density and prostate-specific antigen velocity. *Prostate Int* (2019) 7(3):114–8. doi: 10.1016/j.pri.2019.01.001

22. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol* (2020) 21(2):233–41. doi: 10.1016/S1470-2045(19)30739-9

23. Hood SP, Cosma G, Foulds GA, Johnson C, Reeder S, McArdle SE, et al. Identifying prostate cancer and its clinical risk in asymptomatic men using machine learning of high dimensional peripheral blood flow cytometric natural killer cell subset phenotyping data. *Elife* (2020) 9:e50936. doi: 10.7554/eLife.50936

24. Schelb P, Kohl S, Radtke JP, Wiesenfarth M, Kickingereder P, Bickelhaupt S, et al. Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment. *Radiology* (2019) 293(3):607–17. doi: 10.1148/radiol.2019190938

25. Dominguez GA, Polo AT, Roop J, Campisi AJ, Somer RA, Perzin AD, et al. Detecting prostate cancer using pattern recognition neural networks with flow cytometry-based immunophenotyping in At-risk men. *biomark Insights* (2020) 15:1177271920913320. doi: 10.1177/1177271920913320

26. Wu M, Krishna S, Thornhill RE, Flood TA, McInnes MDF, Schieda N. Transition zone prostate cancer: logistic regression and machine-learning models of quantitative ADC, shape and texture features are highly accurate for diagnosis. *J Magn Reson Imaging* (2019) 50(3):940–50. doi: 10.1002/jmri.26674

27. Zheng Y, Huang Y, Cheng G, Zhang C, Wu J, Qin C, et al. Developing a new score system for patients with PSA ranging from 4 to 20 ng/mL to improve the

- accuracy of PCa detection. *Springerplus* (2016) 5(1):1484. doi: 10.1186/s40064-016-3176-3
28. Abbasi AA, Hussain L, Awan IA, Abbasi I, Majid A, Nadeem MSA, et al. Detecting prostate cancer using deep learning convolution neural network with transfer learning approach. *Cognit Neurodyn* (2020) 14(4):523–33. doi: 10.1007/s11571-020-09587-5
29. Chiu PK, Shen X, Wang G, Ho CL, Leung CH, Ng CF, et al. Enhancement of prostate cancer diagnosis by machine learning techniques: an algorithm development and validation study. *Prostate Cancer Prostatic Dis* (2022) 25(4):672–6. doi: 10.1038/s41391-021-00429-x
30. Zhang Y, Li Q, Xin Y, Lv W. Differentiating prostate cancer from benign prostatic hyperplasia using PSAD based on machine learning: single-center retrospective study in China. *IEEE/ACM Trans Comput Biol Bioinform* (2018) 16(3):936–41. doi: 10.1109/TCBB.2018.2822675
31. Girolami I, Neil D, Segev DL, Furian L, Zaza G, Boggi U, et al. Discovered cancers at postmortem donor examination: a starting point for quality improvement of donor assessment. *Transplant Rev (Orlando)* (2021) 35(2):100608. doi: 10.1016/j.tre.2021.100608
32. Eccher A, Girolami I, Marletta S, Brunelli M, Carraro A, Montin U, et al. Donor-transmitted cancers in transplanted livers: analysis of clinical outcomes. *Liver Transpl* (2021) 27(1):55–66. doi: 10.1002/lt.25858
33. Bruno SM, Falagario UG, d'Altilla N, Recchia M, Mancini V, Selvaggio O, et al. PSA density help to identify patients with elevated PSA due to prostate cancer rather than intraprostatic inflammation: a prospective single center study. *Front Oncol* (2021) 11:693684. doi: 10.3389/fonc.2021.693684
34. Ferro M, de Cobelli O, Musi G, Del Giudice F, Carrieri G, Busetto GM, et al. Radiomics in prostate cancer: an up-to-date review. *Ther Adv Urol* (2022) 14:17562872221109020. doi: 10.1177/17562872221109020
35. Bermejo P, Vivo A, Tarraga PJ, Rodriguez-Montes JA. Development of interpretable predictive models for BPH and prostate cancer. *Clin Med Insights Oncol* (2015) 9:15–24. doi: 10.4137/CMO.S19739
36. Perez IM, Jambor I, Kauko T, Verho J, Ettala O, Falagario U, et al. Qualitative and quantitative reporting of a unique biparametric MRI: towards biparametric MRI-based nomograms for prediction of prostate biopsy outcome in men with a clinical suspicion of prostate cancer (IMPROD and MULTI-IMPROD trials). *J Magn Reson Imaging* (2020) 51(5):1556–67. doi: 10.1002/jmri.26975
37. Parekh S, Ratnani P, Falagario U, Lundon D, Kewlani D, Nasri J, et al. The mount Sinai prebiopsy risk calculator for predicting any prostate cancer and clinically significant prostate cancer: development of a risk predictive tool and validation with advanced neural networking, prostate magnetic resonance imaging outcome database, and European randomized study of screening for prostate cancer risk calculator. *Eur Urol Open Sci* (2022) 41:45–54. doi: 10.1016/j.euros.2022.04.017
38. Winoker JS, Wajswol E, Falagario U, Maritini A, Moshier E, Voutsinas N, et al. Transperineal versus transrectal targeted biopsy with use of electromagnetically-tracked MR/US fusion guidance platform for the detection of clinically significant prostate cancer. *Urology* (2020) 146:278–86. doi: 10.1016/j.urology.2020.07.072