Submitted: 26 August, 2024 Accepted: 24 January, 2025 Published: 30 May, 2025 DOI:10.22514/jomh.2025.067

ORIGINAL RESEARCH

JOMH Journal of Men's Health

Diagnosing prostate cancer in the PSA gray zone through machine learning and transrectal ultrasound video

Qin Wu[1,†], Chengyi Wu[2,†], Maoliang Zhang[1,†], Jie Yang[1], Junxiang Zhang[1], Yun Jin[1], Yanhong Du[1], Xingbo Sun[1], Liyuan Jin[1], Kai Wang[1], Zhengbiao Hu[1], Xiaoyang Qi[1], Jincao Yao[3,4,5,*], Zhengping Wang[1,*], Dong Xu[3,4,5,*]

[1]Department of Ultrasound, The Affiliated Dongyang Hospital of Wenzhou Medical University, 322100 Dongyang, Zhejiang, China
[2]Medical College of Zhejiang University, 310033 Hangzhou, Zhejiang, China
[3]Department of Ultrasound, Zhejiang Cancer Hospital, 310022 Hangzhou, Zhejiang, China
[4]Interventional Medicine and Engineering Research Center, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, 310006 Hangzhou, Zhejiang, China
[5]Key Laboratory of Head & Neck Cancer Translational Research of Zhejiang Province, Zhejiang Provincial Research Center for Cancer Intelligent Diagnosis and Molecular Technology, 310022 Hangzhou, Zhejiang, China

***Correspondence**
yaojc@zjcc.org.cn (Jincao Yao);
wangzp0203@163.com (Zhengping Wang);
xudong@zjcc.org.cn (Dong Xu)

† These authors contributed equally.

Abstract

Background: We developed a machine learning-based predictive model for diagnosing prostate cancer within the gray zone of prostate-specific antigen (PSA) levels, leveraging transrectal prostate ultrasound video clips. **Methods**: Data were collected for patients with suspected prostate cancer, characterized by intermediate PSA levels between 4 and 10 ng/mL, who visited the Department of Urology, Dongyang People's Hospital, which is affiliated with Wenzhou Medical University, from 20 August 2021 to 30 September 2023. Among the final selection of 508 patients, a total of 851 features were extracted from the ultrasound video clips, reduced the dimensionality using least absolute shrinkage and selection operator regression, and finally selected 25 features. The selected features were employed to construct radiomics models based on four machine learning algorithms support vector machine (SVM), random forest (RF), adaptive boosting (ADB) and gradient boosting machine (GBM). The performance of the model was comprehensively assessed using receiver operating characteristic (ROC) curve analysis, with diagnostic effectiveness measured through metrics such as the area under the curve (AUC), sensitivity, specificity and overall accuracy. **Results**: The RF model demonstrated an AUC of 0.89, accuracy of 0.81, sensitivity of 0.81, specificity of 0.79, positive predictive value of 0.91 and F1 score of 0.77. As compared to the RF model, the SVM, ADB and GBM models showed similar values for AUC (range 0.80–0.86), accuracy (range 0.75–0.79), sensitivity (range 0.80–0.81), specificity (range 0.65–0.75), positive predictive value (range 0.83–0.89) and F1 score (range 0.72–0.76). In the validation set, following comprehensive evaluation, the RF model exhibited the best performance among the four models. **Conclusions**: The four machine learning models each had diagnostic value for detecting prostate cancer in patients within the PSA "gray zone", with the RF model demonstrating the highest predictive performance.

Keywords

Prostate cancer; PSA; Gray zone; Machine Learning; Ultrasound

1. Introduction

Prostate cancer (PCa) remains one of the most common malignancies affecting men worldwide, characterized by a high incidence rate and varying degrees of aggressiveness [1]. Clinically significant prostate tumors pose a significant health risk due to their potential for progression and metastasis [2]. The diagnostic approaches for PCa have traditionally relied on a combination of digital rectal examinations (DRE), prostate-specific antigen (PSA) testing [3], and advanced imaging techniques such as transrectal ultrasound (TRUS) and magnetic resonance imaging (MRI) [4].

PSA testing, in particular, is a cornerstone in the early detection of PCa. PSA is a protein produced by both normal and malignant prostate cells, with elevated levels of PSA often serving as a marker for potential malignancy [3, 5]. However, there is a so-called "PSA gray zone", typically defined as a PSA range between 4 and 10 ng/mL, presents a diagnostic challenge [6]. This range is particularly problematic due to an increased risk of false-positive results, as PSA levels within this gray zone are neither clearly indicative of benign conditions nor definitively diagnostic for PCa. As a result, additional evaluation is required to improve diagnostic accuracy.

To address the diagnostic challenges associated with the PSA gray zone, additional biomarkers and advanced imaging techniques have been explored. For instance, recent research has investigated the potential of prostate health index scores [7], 4Kscore tests, and multiparametric MRI to enhance di-

agnostic accuracy within this uncertain range [8, 9]. Despite these diagnostic advances, many studies are significantly constrained by small sample sizes, limiting their generalizability. Other studies have relied on static imaging [10], which may not capture the dynamic behavior of tumor. Moreover, some new modalities are cost-prohibitive and not universally available, further restricting their practical application in routine clinical practice [11]. These shortcomings underscore the need for innovative and accessible methods to improve the specificity and sensitivity of PCa detection in the PSA gray zone, providing an impetus for the development of novel, dynamic and cost-effective solutions.

The recent surge in machine learning (ML) techniques offers a promising avenue for addressing the challenges and complexities associated with the PSA gray zone. With the capacity to analyze vast datasets and identify subtle patterns, machine learning can potentially unveil novel imaging biomarkers and improve decision-making processes [12]. The integration of ML with dynamic imaging modalities, such as TRUS videos, could provide a more detailed and nuanced understanding of tumor characteristics, facilitating improved risk stratification for patients with PSA levels in the gray zone.

Our study harnessed the power of machine learning in conjunction with TRUS video data to extract pertinent radiomic features. Through this approach, we developed a model capable of accurately distinguishing between benign conditions and malignancies within the PSA gray zone. The rationale for combining ML with TRUS video data lies in the ability of dynamic imaging to provide a temporal dimension to tissue characterization, which static images cannot capture. By training a model with this enriched data, we can effectively capture the heterogeneity of prostate tissue and identify malignancies with greater precision.

As far as we are aware, this study represents the first attempt to combine machine learning with TRUS video data for risk assessment in the PSA gray zone of prostate tumors, and the largest-scale investigation to date. Our research not only advances the diagnostic capabilities for PCa but also introduces a scalable and cost-effective tool that can be readily implemented in diverse clinical settings. By bridging the gap between advanced ML techniques and practical clinical applications, we believe our model is a significant milestone in the management of PCa, particularly for patients navigating the diagnostic uncertainty of the PSA gray zone.

2. Methods and materials

2.1 General information

A retrospective study was performed on clinical data collected from patients presenting with suspected prostate cancer at the Department of Urology, Dongyang Hospital, affiliated with Wenzhou Medical University, between August 2021 and September 2023. Initially, 1189 cases were reviewed, of which 508 cases were eventually included, comprising 187 cases of PCa and 321 cases of benign prostatic lesions. Participant selection criteria were rigorously operationalized as follow: (1) The preoperative assessment encompassed serum biomarkers associated with prostate-specific antigen (PSA), including total PSA (tPSA), free PSA (fPSA) and the fPSA/tPSA ratio, all measured within the range of 4 to 10 ng/mL; (2) completion of TRUS examination; and (3) patients who underwent transperineal prostate biopsy and received pathological diagnostic results. The exclusion criteria were: (1) previous prostate treatment such as hormonal or radiation therapy, and (2) incomplete clinical data or imaging examinations.

Among the 508 patients, 436 cases were randomly selected for training and validating the models, whereas the remaining 72 were used for assessing the derived models. The study protocol was evaluated and approved by the ethics committee of the hospital and subsequently registered. Fig. 1 illustrates the process of participant inclusion and exclusion.

2.2 Ultrasound video clip acquisition

Ultrasound video clips were obtained using the Esaote MyLab™ Class C color Doppler ultrasound diagnostic system (TRT33 transrectal dual-plane ultrasound probe, frequency range 3–13 MHz). The procedure was performed by three highly experienced ultrasound specialists, each boasting over a decade of expertise in transrectal ultrasound diagnostics. Transrectal prostate ultrasound imaging was performed, and 10-second transverse section video clips of the prostate were recorded and preserved for subsequent data analysis.

2.3 Manual segmentation

Three physicians, each with more than five years' clinical experience in transrectal ultrasound (TRUS) diagnostics, manually annotated the images using 3D Slicer software (v.5.03). The entire prostate was outlined as the region of interest using three image sets. To ensure privacy, all identifying information, including patient names, was removed. Cases were assigned numerical identifiers and randomized to prevent sonographers from accessing any potentially relevant information prior to outlining the regions of interest.

Intraclass correlation coefficients (ICCs) were utilized to evaluate the consistency of the extracted imaging histology features, considering both within-subject and between-subject variability. Features with ICCs $\geq$0.8, indicating acceptable reproducibility, were retained for the subsequent feature selection process.

2.4 Extraction and selection of features

Imaging histology features were extracted using the Slicer Radiomics extension package (v.aa418a5) within the 3D Slicer software [13]. From each case, we extracted 851 imaging histology features, including 14 shape features, 18 first-order features, 24 gray-level co-occurrence matrix (GLCM) features, 14 gray-level dependency matrix (GLDM) features, 16 gray-level run length matrix (GLRLM) features, 16 gray-level size zone matrix (GLSZM) features, 5 neighborhood gray-tone difference matrix (NGDM) features and 744 wavelet features.

Features were selected using Python (v.3.8.8). Initially, all features underwent Levene's χ^2 test, followed by two independent sample *t*-tests (for features passing the χ^2 test) or Mann-Whitney U tests (for features failing the χ^2 test) to assess discrepancies between benign and malignant lesions,

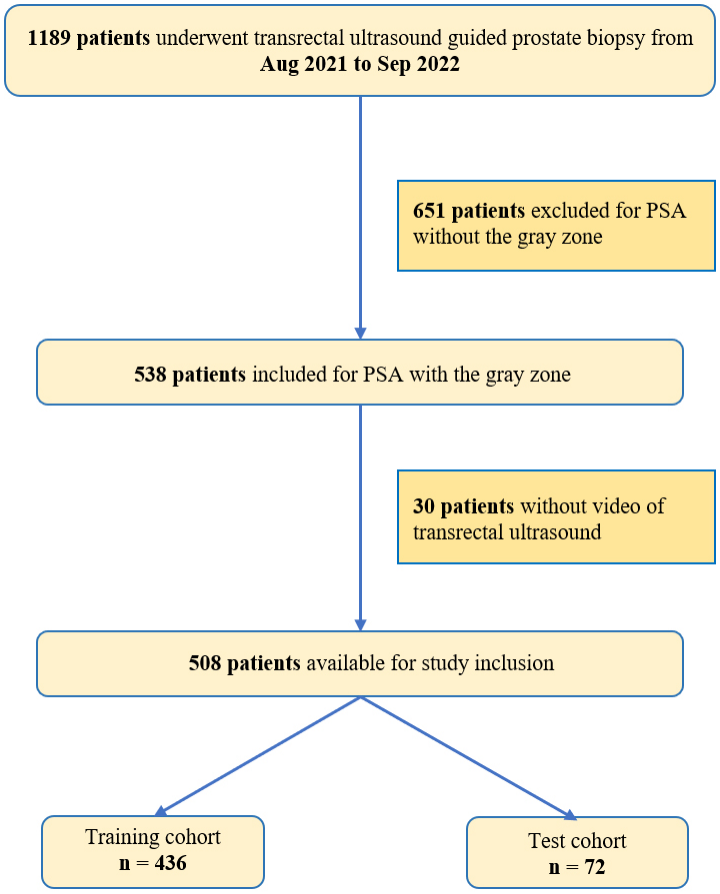

FIGURE 1. Flowchart of the study inclusion and exclusion criteria. PSA: prostate-specific antigen.

maintaining features with noteworthy differences ($p < 0.05$). Labels were assigned to each case's data based on pathological findings, where 1 labeled "malignant" and 0 labeled "benign". The residual features were then processed using least absolute shrinkage and selection operator (LASSO) regression for further screening. Within the training dataset, a rigorous tenfold cross-validation protocol was iterated 1000 cycles to calibrate the regularization parameters (λ) and isolate most informative features. Ultimately, features with non-zero coefficients were selected.

2.5 Machine Learning

Machine learning models were developed using the scikit-learn 0.24.2 library in Python [14]. The training cohort was randomly split into an 80% training subset and a 20% internal validation subset. Feature classification within the training set was conducted using four machine learning algorithms: Support Vector Machine (SVM), Random Forest (RF), Adaptive Boosting (ADB), and Gradient Boosting Machine (GBM), each applied independently. Model optimization was conducted using grid search and cross-validation (GridSearchCV) to determine the parameter configurations that achieved the highest accuracy in the validation set for each model.

2.6 Statistical analysis

Statistical computations were performed utilizing Statistical Package of Social Sciences (SPSS) version 25.0 (IBM Corp., Armonk, NY, USA). Levene's test was used to assess normality. For normally distributed continuous variables, descriptive analysis is presented as mean $\pm$ standard deviation, and intergroup differences were analyzed using independent sample *t*-tests. Non-normal distributions are described by median and interquartile range (IQR), with statistical contrasts performed through Mann-Whitney U test. Categorical variables were quantified as frequency distributions (expressed in percentages) and subjected to statistical analysis using the χ^2 test or continuity correction test. Receiver operating characteristic (ROC) curves were utilized to assess the diagnostic performance of the models for prostate cancer (PCa), including metrics such as area under the curve (AUC), sensitivity, specificity and diagnostic accuracy. A statistical significance threshold of $p < 0.05$ was applied. The overall flowchart for the study is presented in Fig. 2.

3. Results

3.1 Baseline characteristics

Following screening, a total of 508 cases were collected, comprising 321 benign prostatic lesions and 187 PCa instances. These patients were divided into a training group (436 cases) and a validation group (72 cases), with comparable baseline characteristics including age (71 years, IQR: 66–76 years *vs.* 70 years, IQR: 65–73 years), total prostate specific antigen (tPSA) concentrations (6.47 ng/mL, IQR: 5.23–7.82 ng/mL *vs.* 6.25 ng/mL, IQR: 5.06–7.85 ng/mL) and prostate volume

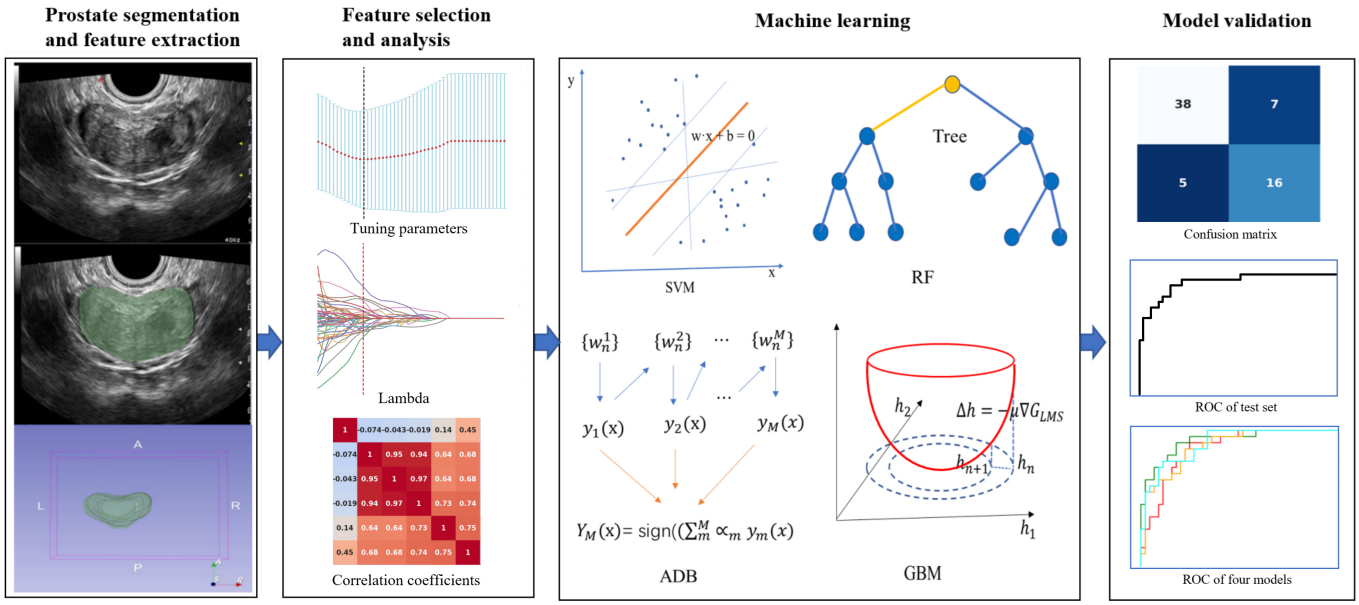

FIGURE 2. Overall study flowchart, including feature extraction, feature selection, machine learning and model validation. ADB: adaptive boosting; RF: random forest; GBM: gradient boosting machine; ROC: receiver operating characteristic; AUC: area under the curve; SVM: support vector machine; TPR: true positive rate; MSE: mean square error.

(39.2 cm^3, IQR: 31.82–50.42 cm^3 *vs.* 38.28 cm^3, IQR: 29.59–47.33 cm^3), with no significant differences observed between the training and validation groups, respectively.

Post-biopsy pathological results revealed that in the training group, 275 cases (63.1%) were diagnosed with benign prostate lesions, including benign prostatic hyperplasia (BPH) in 227 cases (52.1%), BPH with prostatitis in 45 cases (10.3%), and high-grade prostatic intraepithelial neoplasia (HGPIN) in 3 cases (0.7%). Additionally, there were 161 cases (36.9%) of PCa, with Gleason scores (GS) distributed as follows: GS 6, 96 cases (22.0%); GS 7, 54 cases (12.4%); and GS ≥8, 11 cases (2.5%). In the validation group, 47 cases (65.3%) were diagnosed with benign prostate lesions, including BPH in 38 cases (52.8%), BPH with prostatitis in 7 cases (2.6%) and HGPIN in 2 cases (2.8%). Furthermore, there were 25 cases (34.7%) of PCa, with GS disseminated as follows: GS 6, 11 cases (15.3%); GS 7, 11 cases (15.3%) and GS ≥8, 3 cases (4.1%). Clinical and pathological characteristics of the training and validation groups are comprehensively outlined in Table 1.

3.2 Feature selection

A total of 851 distinct features were derived from the ultrasound video clip datasets for each individual case, which were then subjected to feature selection using either *t*-tests or Mann-Whitney U tests based on the normality evaluation performed with Levene's test ($p > 0.05$). In LASSO regression, 10-fold cross-validation was employed to determine the optimal λ coefficient. After LASSO dimensionality reduction, 25 features were retained: 7 original features and 18 wavelet features (Table 2). Additionally, Fig. 3 demonstrates the process of identifying key parameters from the features within the training set and the subsequent development of the linear predictor. Meanwhile, Fig. 4 displays the optimal penalty coefficient (λ) determined during the model optimization process.

Among the salient features delineated in Table 2, the original shape-related features are particularly noteworthy, as they provide critical insights into the morphology and margins of tumors within the US images. These features are instrumental in facilitating accurate diagnostic assessments. Additionally, several wavelet features capture intricate internal variations of the tumors, potentially reflecting subtle pathological changes that may signify malignancy or distinct tissue characteristics, thereby enhancing the interpretative depth of the model.

3.3 Model effectiveness

The RF model exhibited an AUC of 0.89, accuracy of 0.81, sensitivity of 0.81, specificity of 0.79, positive predictive value of 0.91, and F1 score of 0.77. The SVM model demonstrated an AUC of 0.80, accuracy of 0.79, sensitivity of 0.81, specificity of 0.75, positive predictive value of 0.89, and F1 score of 0.76. The ADB model achieved an AUC of 0.85, accuracy of 0.75, sensitivity of 0.80, specificity of 0.65, positive predictive value of 0.83, and F1 score of 0.72. Lastly, the GBM model obtained an AUC of 0.86, accuracy of 0.75, sensitivity of 0.80, specificity of 0.65, positive predictive value of 0.83, and F1 score of 0.72. Upon comprehensive analysis of the evaluation metrics across the four models in the validation set, the performance of the RF model surpassed that of GBM, ADB and SVM, exhibiting superior predictive performance and the best predictive efficacy for PSA gray zone PCa when compared to the other three models. The diagnostic performances of the RF, SVM, ADB and GBM models are summarized in Table 3, whereas the ROC curves are depicted in Fig. 5.

4. Discussion

Prostate biopsy is commonly considered the definitive method for diagnosing prostate cancer (PCa). Nevertheless, this inva-

TABLE 1. Clinicopathological characteristics of the training and test cohort.

Parameters (projects)	Training cohort (n = 436)	Test cohort (n = 72)
Age (yrs, median, IQR)	71 (66, 76)	70 (65, 73)
tPSA (ng/mL, median, IQR)	6.47 (5.23, 7.82)	6.20 (5.06, 7.85)
fPSA (ng/mL, median, IQR)	0.98 (0.69, 1.30)	0.84 (0.64, 1.05)
fPSA/tPSA (median, IQR)	0.15 (0.12, 0.19)	0.14 (0.11, 0.16)
PV (cm^3, median, IQR)	39.2 (31.82, 50.45)	38.28 (29.59, 47.33)
PSAD (ng/mL, cm^3, median, IQR)	0.09 (0.07, 0.11)	0.17 (0.13, 0.21)
Pathology		
Number of Benign cases (n, %)	275 (63.1)	47 (65.3)
BPH	227 (52.1)	38 (52.8)
BPH & prostatitis	45 (10.3)	7 (9.7)
HGPIN	3 (0.7)	2 (2.8)
Number of PCa cases (n, %)	161 (36.9)	25 (34.7)
GS: 6	96 (22.0)	11 (15.3)
GS: 7	54 (12.4)	11 (15.3)
GS: 8	9 (2.1)	3 (4.1)
GS: 9	1 (0.2)	0
GS: 10	1 (0.2)	0

Abbreviations: IQR: interquartile range; tPSA: total prostate specific antigen; fPSA: free prostate specific antigen; f/tPSA: free/total prostate specific antigen ratio; PV: prostate volume; PSAD: prostate specific antigen density; BPH: benign prostatic hyperplasia; HGPIN: high-grade prostatic intraepithelial neoplasia; PCa: prostate cancer; GS: Gleason score.

TABLE 2. The subset of radiomics features ultimately selected by the LASSO algorithm.

Feature(s)	Image type	Feature class	Feature name	LASSO coefficients
1	original	shape	Least Axis Length	0.033571
2	original	shape	Maximum 2D Diameter Column	−0.093023
3	original	shape	Maximum 2D Diameter Row	0.005961
4	original	shape	Mesh Volume	0.087038
5	original	shape	Sphericity	−0.060414
6	original	shape	Surface Volume Ratio	0.024544
7	original	GLCM	Difference Average	−0.014693
8	Wavelet-LHL	First-order	Kurtosis	−0.030508
9	Wavelet-LHL	First-order	Median	0.034010
10	Wavelet-LHL	GLCM	ClusterShade	−0.021096
11	Wavelet-LHL	GLCM	Inverse Variance	0.001262
12	Wavelet-LHL	GLCM	Joint Average	0.008146
13	Wavelet-LHL	First-order	Median	−0.026531
14	Wavelet-LHL	First-order	Median	0.053700
15	Wavelet-LHL	First-order	Minimum	−0.046910
16	Wavelet-LHL	GLSZM	Gray Level Variance	0.100897
17	Wavelet-LHL	GLSZM	Small Area High Gray Level Emphasis	0.094075
18	Wavelet-LHL	First-order	Minimum	−0.059024
19	Wavelet-LHL	First-order	Minimum	0.016990
20	Wavelet-LHL	First-order	Total Energy	0.037929
21	wavelet-HHH	GLCM	Imc^2	−0.061461
22	Wavelet-HHH	GLSZM	High Gray Level Zone Emphasis	−0.068249
23	Wavelet-HHH	GLSZM	Small Area Low Gray Level Emphasis	−0.039780
24	wavelet-LLL	GLSZM	Id	0.158603
25	Wavelet-LLL	GLSZM	Idn	−0.045269

GLCM: gray-level co-occurrence matrix; GLSZM: gray-level size zone matrix; LASSO: least absolute shrinkage and selection operator. Imc^2: incremental matrix completion^2; Idn: identifier numeric; Id: identifier.

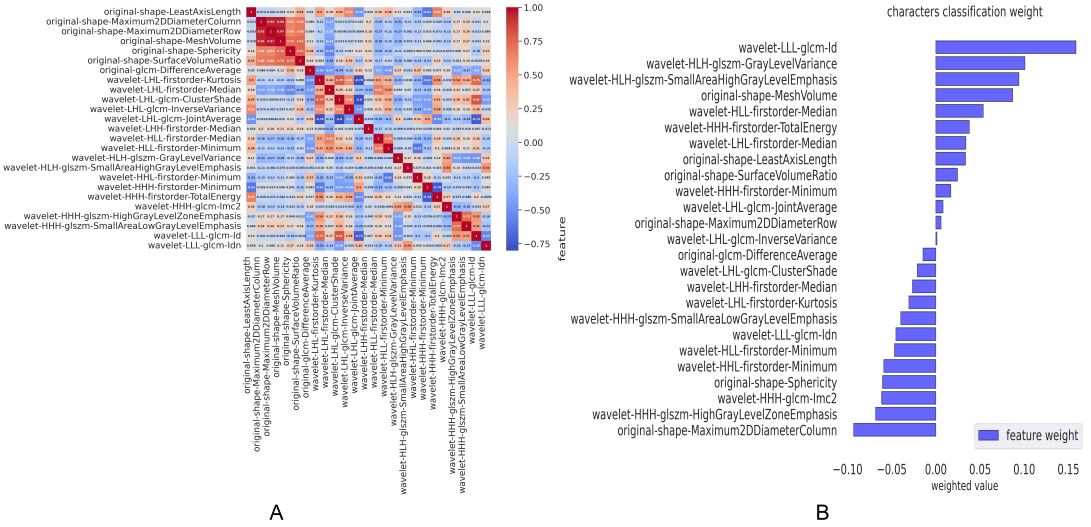

FIGURE 3. Identification of key feature parameters within the training set and development of the linear predictor. (A) Spearman's correlation coefficients were estimated for the twenty-five selected features. (B) Feature classification weights. GLCM: gray-level co-occurrence matrix; GLSZM: gray-level size zone matrix; Imc^2: incremental matrix completion^2; Idn: identifier numeric; Id: identifier.

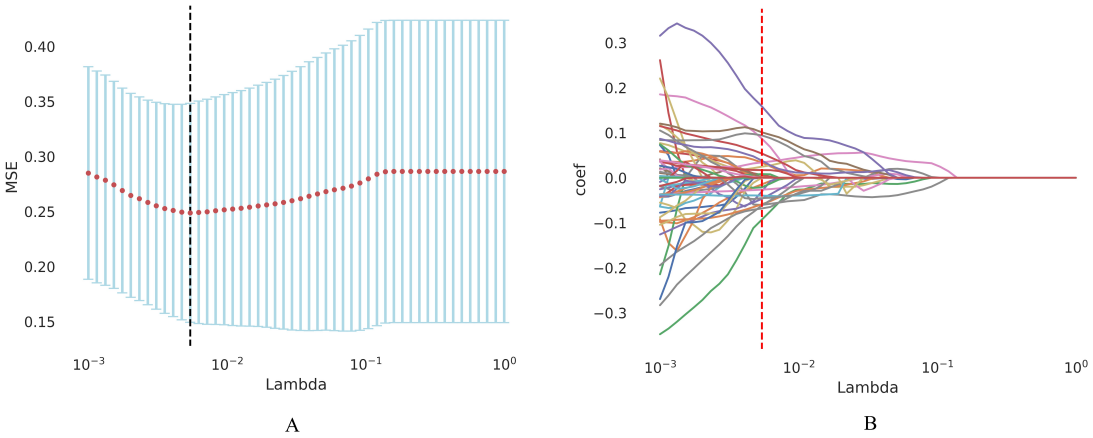

FIGURE 4. Determination of the optimal penalty coefficient, λ. (A) Ten-fold cross-validation was employed to select the tuning parameters for the LASSO model. (B) Solution path of LASSO coefficients for the 25 features. MSE: mean square error; coef: coefficients.

sive procedure may lead to considerable psychological burden in patients with PSA levels in the gray zone. Therefore, it is necessary to combine multiple techniques for diagnosing PCa in patients within the PSA gray zone. PSA has revolutionized the diagnosis of PCa, especially in the detection of early asymptomatic cases. It is noteworthy that when PSA levels are between 4–10 ng/mL, there is a significant overlap in the number of patients with BPH and PCa [15]. In recent years, the application of multiparametric MRI (mpMRI) has further improved the sensitivity and specificity of PCa diagnosis. The Prostate Imaging Reporting and Data System version 2 (PI-RADS v2) is in clinical use, but its predictive value for prostate biopsy results in patients in the PSA gray zone remains limited [16].

TRUS and mpMRI are two widely utilized imaging techniques for the evaluation of prostate diseases. TRUS is extensively used in clinical practice due to its safety, lack of radiation, low cost and simplicity [17, 18]. TRUS can usually clearly visualize the contours of the prostate, as well as indicate the transition zone with isoechoic and hypoechoic areas, with around 70% of prostate cancers are located in the peripheral zone (which generally appear as hypoechoic lesions). However, PCa situated within the transition zone can present detection challenges, as the neoplasm might be concealed by the hypoechoic background tissue. Although grayscale TRUS is helpful in diagnosing BPH, it is unreliable for detecting PCa [19, 20]. The accuracy of TRUS in detecting PCa in the general population is only moderate [21]. In studies concerning the PSA gray zone, the detection rate for PCa is 25% [22]. In this study, out of 508 patients with suspected PCa in the PSA gray zone, 187 (36%) were diagnosed with PCa by biopsy, a slightly higher fraction than in previous research results. This variation could potentially stem from a multitude of factors, including but not limited to ethnic diversity, geographical distinctions and dietary practices.

Conversely, liquid biopsies have gained significant promi-

TABLE 3. Performance outcomes for each machine learning models.

Data set	AUC	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1-score
SVM							
Validation set	0.85	0.85	0.91	0.74	0.87	0.81	0.83
Test set	0.80	0.79	0.81	0.75	0.89	0.60	0.76
RF							
Validation set	0.90	0.85	0.89	0.76	0.89	0.76	0.83
Test set	0.89	0.81	0.81	0.79	0.91	0.60	0.77
ADB							
Validation set	0.86	0.76	0.81	0.63	0.84	0.57	0.71
Test set	0.85	0.75	0.80	0.65	0.83	0.60	0.72
GBM							
Validation set	0.85	0.80	0.86	0.68	0.84	0.71	0.78
Test set	0.86	0.75	0.80	0.65	0.83	0.60	0.72

AUC: area under the curve; SVM: support vector machine; RF: random forest; ADB: adaptive boosting; GBM: gradient boosting machine.

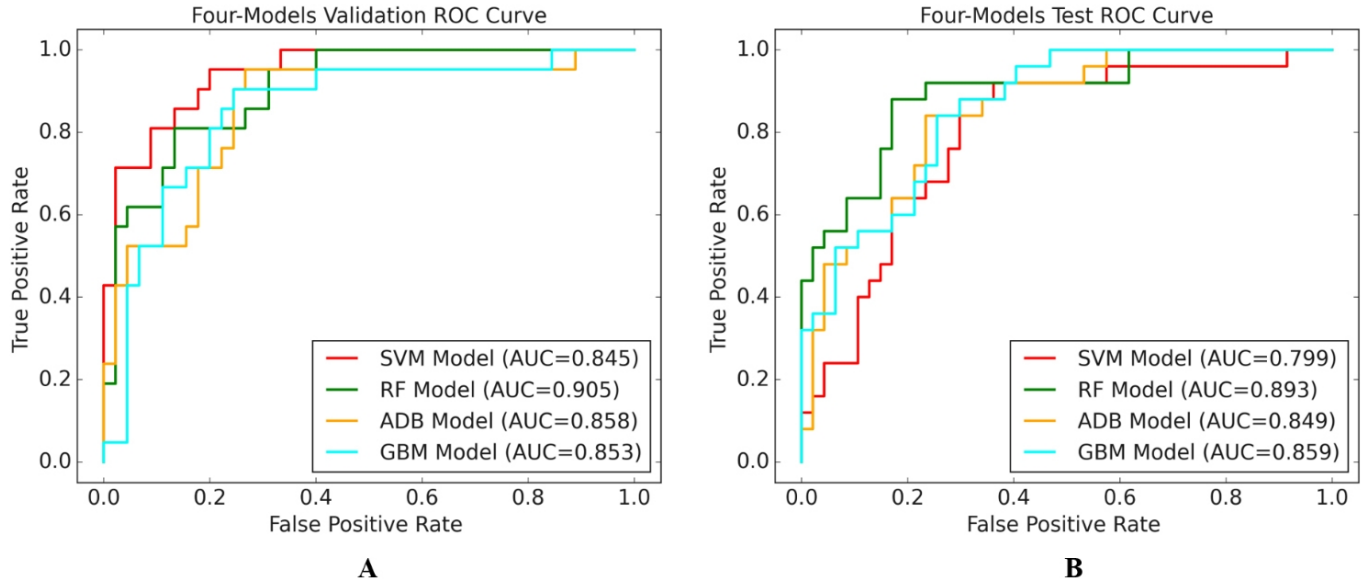

FIGURE 5. ROC curves of categorical multivariate models in the validation set (A) and test set (B). ROC: receiver operating characteristic; SVM: support vector machine; AUC: area under the curve; RF: random forest; ADB: adaptive boosting; GBM: gradient boosting machine.

nence as a valuable tool in cancer research, serving as a complement to conventional tissue biopsies [23]. This non-invasive technique enables real-time assessment of genomic, transcriptomic, and epigenomic alterations in both primary and metastatic tumors. By integrating data from multiple liquid biopsies, it becomes possible to thoroughly evaluate tumor heterogeneity and subclonal evolution, which aids in the development of personalized therapeutic strategies [24]. Various clinical studies to date have demonstrated the utility of liquid biopsy using cell-free DNA and extracellular vesicles (EVs) as tools in distinct clinical stages of PCa management. Specifically, circulating tumor DNA is increasingly used in advanced cases of metastatic castration-resistant PCa, as it targets genetic mutations within the tumor, enabling treatment choices based on these mutations. When a tumor grows, the expression of certain molecules within EVs may increase. This approach is key because, if these specific molecules can be pinpointed early on, even at the early stages of PCa, it could greatly aid in early detection [25]. Although liquid biopsy has demonstrated enormous potential to improve the current practice, additional assays for risk stratification are required to supplement the current model [26]. To accompany the PSA test for PCa screening, high-specificity tests are needed. After PSA screening, a key challenge is determining how to further select the best candidates for needle biopsy to avoid such invasiveness in a non-cancer population. Even though the future of PCa treatment using liquid biopsy holds great promise, addressing the existing challenges is imperative for its successful integration into routine clinical practice, making personalized and precise PCa management a reality for

patients.

Prostate enlargement with increasing age leads to a concomitant increase in prostate volume and cell count, which, in turn, results in elevated PSA secretion by epithelial cells. The diagnostic efficacy of total PSA (tPSA) values for PCa is notably lower within the PSA gray zone compared to cases with PSA levels >10 ng/mL [27]. Notably, Liu constructed a PCa prediction model for 197 patients in the PSA gray zone using PSA and its derivatives, with sensitivity and specificity for predicting PCa of 75.4% and 75.8%, respectively [28]. However, the aforementioned study failed to account for imaging-based assessments.

Radiomics in prostate cancer (PCa) represents a rapidly advancing field of study, holding great promise for providing non-invasive and longitudinal biomarkers to support personalized medicine [29]. This innovative approach utilizes machine learning techniques to convert visual image data into detailed quantitative metrics, extracting extensive feature information from medical images, which are then used to build predictive models [30]. Radiomics facilitates robust quantitative analysis of imaging data, enabling non-invasive evaluation of tumor biology, and has been extensively applied in the diagnosis of PCa, assessment of tumor aggressiveness, and support of clinical decision making [31]. Notably, radiomics-based methods for diagnosing PCa malignancy and GS scoring have demonstrated good predictive performance [32].

By high-throughput data extraction and building efficient and stable predictive models (such as our extraction of features from video clips), radiomics can offer auxiliary diagnostic capabilities for clinical applications. Feature selection is crucial in ML research, as the absence of high-throughput feature extraction can result in data redundancy and overfitting [33–35]. Jin predicted early cervical cancer lymph node metastasis using ML methods, extracting 106 radiomic features from lymph node ultrasound images. By integrating LASSO and ridge regression, key features were selected from high-dimensional data to mitigate the risk of overfitting. Subsequently, six features were identified for classification analysis, demonstrating a strong correlation with tumor texture complexity and heterogeneity [36]. In addition, Wang conducted ML prediction research on PCa in TRUS video clips, achieving SVM model AUCs of 0.78 and 0.75 in the validation and test sets, respectively. The diagnostic performance of the SVM model was higher than diagnostics based on MRI (AUC values of 0.78 *vs.* 0.65/0.75 and 0.75 *vs.* 0.65/0.72, respectively), indicating that it achieves greater diagnostic efficacy and consistency [37]. Furthermore, Li determined that an SVM model based on MRI achieved good diagnostic performance for the malignancy of PI-RADS category 3 prostate lesions (training set AUC = 0.93 and validation set AUC = 0.89) and invasion prediction (training set AUC = 0.92 and validation set AUC = 0.85) [38]. Qi established SVM, RF, ABD and GBM machine learning models for PCa prediction using TRUS video clips combined with MRI, with AUCs of 0.87, 0.85, 0.84 and 0.81, respectively, indicating high predictive performance [39]. Additionally, Nketiah *et al*. [40] combined texture features extracted from T2 weighted image with quantitative apparent diffusion coefficient values and dynamic contrast-enhanced pharmacokinetic parameters for predicting PCa GS grading. Their approach achieved a higher predictive performance (AUC = 0.91) compared to the RF model within the study (AUC = 0.88). This discrepancy could be explained by the limited sample size and the absence of an independent validation cohort in their analysis [40]. While these studies differ in focus, they collectively underscore the significant diagnostic and predictive value of ML models for PCa. They also highlight the versatility of ML applications across various imaging modalities and datasets.

Chen conducted a regression analysis on age, PSA-related indicators, TRUS and mpMRI to construct a binary logistic regression model to predict and evaluate the diagnostic value of PCa in patients in the PSA gray zone. Their results indicated that when selecting a critical value above 0.36, the AUC, sensitivity, and specificity of the PCa prediction model were 0.79, 73.64% and 64.23%, respectively [41]. Our predictive model demonstrated superior diagnostic performance compared to theirs, likely due to the ability of machine learning (ML) models to extract a greater amount of image feature information. As part of this work, a total of 851 feature types were extracted and subsequently filtered using the LASSO algorithm, resulting in the retention of 25 non-zero feature types that showed reliable correlations with PCa. Predictive models were then constructed using SVM, RF, ADB and GBM algorithms. In the validation set, the AUC values achieved were 0.85, 0.90, 0.86 and 0.85, respectively. Corresponding accuracy rates were 0.79, 0.81, 0.75 and 0.75; positive predictive values were 0.89, 0.91, 0.83 and 0.83; and F1 scores were 0.76, 0.77, 0.72 and 0.72, respectively. In contrast, Liu achieved AUC values of the validation set using logistic regression, Gaussian Naive Bayes, light gradient-boosting machine (LGBM) Classifier, and extreme gradient boosting (XGBoost) ML models of 0.918, 0.893, 0.906 and 0.893, respectively, with accuracies of 0.811, 0.792, 0.794 and 0.800, and F1 values of 0.758, 0.765, 0.749 and 0.771 [42]. This outcome indicates that the diagnostic performance of the four machine learning models selected in their study is higher than that of our study, which may be due to differences in samples or ML model selection.

Despite the inclusion of 508 cases in this study, the test set comprised a relatively limited sample size of 72 cases, which may introduce potential bias. Furthermore, while the ratio of sample sizes between the BPH group and the PCa group was maintained at approximately 2:1 to achieve balance, the overall sample size remained moderate. This constraint increases the risk of classifier models encountering issues such as overfitting and reduced generalizability. As a result, expanding the test set sample size is crucial to enhance the reliability and robustness of the findings.

Although the RF model demonstrated good predictive results within this unit, it has not been applied and validated in other clinical settings. Therefore, multicenter studies are needed for further testing and validation. Additionally, prospective studies are required to verify the reliability and accuracy of this model, to improve the predictive efficacy and detection rate of PCa in patients with PSA “gray zone” levels. Such efforts could potentially lower the rate of unwarranted prostate biopsies in this patient population.

This study used prostate biopsy pathology as a reference rather than prostatectomy pathology, which may have intro-

duced bias [43]. For patients diagnosed with PCa through biopsy, some patients may have missed the opportunity for surgical intervention, potentially affecting the outcomes.

5. Conclusions

Machine learning models for PCa diagnosis, utilizing algorithms such as SVM, RF, ADB and GBM, have shown significant diagnostic potential for patients with PSA "gray zone" level. Among these models, the RF algorithm demonstrated the highest predictive performance. These ML models offer promising tools to aid clinical decision-making, supporting physicians in achieving accurate diagnoses and tailoring personalized treatment plans. Additionally, they provide valuable guidance for prostate biopsy decisions and optimizing follow-up strategies for patients within the PSA "gray zone".

ABBREVIATIONS

PCa, Prostate cancer; PSA, prostate specific antigen; AUC, area under the curve; ROC, Receiver operating characteristic; SVM, support vector machine; RF, random forest; ADB, adaptive boosting; GBM, gradient boosting machine; TRUS, Transrectal ultrasound; MRI, Magnetic resonance imaging; ML, Machine learning; ICCs, Intraclass correlation coefficients; LASSO, Least absolute shrinkage and selection operator; SPSS, Statistical Package for the Social Sciences; IQR, Interquartile range; tPSA, Total prostate-specific antigen; BPH, Benign prostatic hyperplasia; HGPIN, High-grade prostatic intraepithelial neoplasia; GS, Gleason score; US, Ultrasound; mpMRI, Multiparametric magnetic resonance imaging; DRE, digital rectal examinations; fPSA, free prostate-specific antigen; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGDM, neighborhood gray-tone difference matrix; SVM, support vector machine; ADB, adaptive boosting; EVs, extracellular vesicles; PI-RADS, Prostate Imaging Reporting and Data System; LGBM, light gradient-boosting machine; XGBoost, extreme gradient boosting; CLCM, collaborative latent class model; IPR, individual prediction reliability.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are available on reasonable request from the corresponding author.

AUTHOR CONTRIBUTIONS

QW, CYW and MLZ—designed the research study and implemented the code and contributed to writing the manuscript. JY, JXZ, YJ, YHD, XBS, LYJ, KW, ZBH and XYQ—managed data collection and processing; contributed to data analysis; and assisted in writing the manuscript. JCY, ZPW and DX—as corresponding authors; supervised the research; managed the project; reviewed the manuscript; and provided guidance on the research project. All authors contributed to editorial changes in the manuscript and read and approved the final version.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The Medical Ethics Committee of Dongyang People's Hospital 2024-YX-030. Comparison of machine learning models based on transrectal ultrasound videos for the diagnosis of prostate cancer in the PSA gray zone. All participants provided their consent to participate in the study.

ACKNOWLEDGMENT

We extend our sincere gratitude to the Department of Urology at Dongyang Hospital, as affiliated with Wenzhou Medical University, for their support and cooperation. We also appreciate the technical support and equipment provided by the Esaote MyLab™ ClassC color Doppler ultrasound diagnostic system. Special thanks to the three experienced ultrasound doctors with over ten years of expertise in transrectal ultrasound diagnosis, who conducted the ultrasound scans and collected the 10-second transverse section video clips of the prostate. Additionally, we acknowledge the contributions of all medical personnel and research team members involved in the collection and processing of patient data.

FUNDING

This work was supported by the National Key Research and Development Program of China (2022YFF0608403, 2023YFC2414004) and the Pioneer and Leading Goose R&D Program of Zhejiang (2023C04039) and the Key Science and Technology Project of Jinhua, Zhejiang Province (2022-3-016, 2022-3-017 and 2023-3-012).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2021; 71: 209–249.

[2] Bergengren O, Pekala KR, Matsoukas K, Fainberg J, Mungovan SF, Bratt O, *et al*. 2022 update on prostate cancer epidemiology and risk factors—a systematic review. European Urology. 2023; 84: 191–206.

[3] Kachuri L, Hoffmann TJ, Jiang Y, Berndt SI, Shelley JP, Schaffer KR, *et al*. Genetically adjusted PSA levels for prostate cancer screening. Nature Medicine. 2023; 29: 1412–1423.

[4] Schaeffer EM, Srinivas S, Adra N, An Y, Barocas D, Bitting R, *et al*. Prostate cancer, Version 4.2023, NCCN clinical practice guidelines in oncology. Journal of the National Comprehensive Cancer Network. 2023; 21: 1067–1096.

[5] Ouyang M, Jia M, Chang Z, Wang Y, Wang K, Gao X, *et al*. Precise prostate cancer diagnosis using fluorescent nanoprobes for detecting PSA and PSMA in serum. Chemical Communications. 2024; 60: 5181–5184.

[6] Yang J, Li J, Xiao L, Zhou M, Fang Z, Cai Y, *et al*. 68Ga-PSMA PET/CT-based multivariate model for highly accurate and noninvasive diagnosis of clinically significant prostate cancer in the PSA gray zone. Cancer Imaging. 2023; 23: 81.

[7] Qian Y, Shi Y, Sheng X, Liao H, Chen H, Shi B, *et al*. Evaluating the role of morphological parameters in the prostate transition zone in PHI-

Based predictive models for detecting gray zone prostate cancer. Clinical Medicine Insights: Oncology. 2023; 17: 11795549231201122.

[8] Siddiqui MR, Li EV, Kumar SKSR, Busza A, Lin JS, Mahenthiran AK, *et al*. Optimizing detection of clinically significant prostate cancer through nomograms incorporating MRI, clinical features, and advanced serum biomarkers in biopsy naïve men. Prostate Cancer and Prostatic Diseases. 2023; 26: 588–595.

[9] Thomas J, Atluri S, Zucker I, Reis I, Kwon D, Kim E, *et al*. A multi-institutional study of 1111 men with 4K score, multiparametric magnetic resonance imaging, and prostate biopsy. Urologic Oncology. 2023; 41: 430.e9–430.e16.

[10] Sachpekidis C, Pan L, Groezinger M, Strauss DS, Dimitrakopoulou-Strauss A. Combined whole-body dynamic and static PET/CT with low-dose [18F] PSMA-1007 in prostate cancer patients. European Journal of Nuclear Medicine and Molecular Imaging. 2024; 51: 2137–2150.

[11] Hamm CA, Baumgärtner GL, Biessmann F, Beetz NL, Hartenstein A, Savic LJ, *et al*. Interactive explainable deep learning model informs prostate cancer diagnosis at MRI. Radiology. 2023; 307: e222276.

[12] Chen G, Dai X, Zhang M, Tian Z, Jin X, Mei K, *et al*. Machine learning-based prediction model and visual interpretation for prostate cancer. BMC Urology. 2023; 23: 164.

[13] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, *et al*. Computational radiomics system to decode the radiographic phenotype. Cancer Research. 2017; 77: e104–107.

[14] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, *et al*. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. 2011; 12: 2825–2830.

[15] Alonzo DG, Mure AL, Soloway MS. Prostate cancer and the increasing role of active surveillance. Postgraduate Medicine. 2013; 125: 109–116.

[16] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, *et al*. Computational radiomics system to decode the radiographic phenotype. Cancer Research. 2017; 77: e104–e107.

[17] Kaneko M, Sugano D, Lebastchi AH, Duddalwar V, Nabhani J, Haiman C, *et al*. Techniques and outcomes of MRI-TRUS fusion prostate biopsy. Current Urology Reports. 2021; 22: 27.

[18] Merriel SWD, Pocock L, Gilbert E, Creavin S, Walter FM, Spencer A, *et al*. Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (PSA) for the detection of prostate cancer in symptomatic patients. BMC Medicine. 2022; 20: 54.

[19] Van den Broeck T, van den Bergh RCN, Briers E, Cornford P, Cumberbatch M, Tilki D, *et al*. Biochemical recurrence in prostate cancer: the European association of urology prostate cancer guidelines panel recommendations. European Urology Focus. 2020; 6: 231–234.

[20] Smeenge M, Barentsz J, Cosgrove D, de la Rosette J, de Reijke T, Eggener S, *et al*. Role of transrectal ultrasonography (TRUS) in focal therapy of prostate cancer: report from a Consensus Panel. BJU International. 2012; 110: 942–948.

[21] Tomczak K, Czerwińska P, Wiznerowicz M. The cancer genome atlas (TCGA): An immeasurable source of knowledge. Contemporary Oncology. 2015; 19: A68–A77.

[22] Smith DS, Humphrey PA, Catalona WJ. The early detection of prostate carcinoma with prostate specific antigen: the Washington University experience. Cancer. 1997; 80: 1852–1856.

[23] Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic—implementation issues and future challenges. Nature Reviews Clinical Oncology. 2021; 18: 297–312.

[24] Wang Y, Wang Z, Gang X, Wang G. Liquid biopsy in prostate cancer: current status and future challenges of clinical application. The Aging Male. 2021; 24: 58–71.

[25] Urabe F, Sumiyoshi T, Tashiro K, Goto T, Kimura T, Kobayashi T. Prostate cancer and liquid biopsies: clinical applications and challenges. International Journal of Urology. 2024; 31: 617–626.

[26] Kim CJ, Dong L, Amend SR, Cho YK, Pienta KJ. The role of liquid biopsies in prostate cancer management. Lab on a Chip. 2021; 21: 3263–3288.

[27] Qin B, Liang Y, Liang Y, Zhang J, Wu Q, Ji H, *et al*. Comparative study on the diagnosis of gray zone prostate cancer by transperineal and transrectal prostate biopsy guided by transrectal ultrasound. Chinese Journal of Andrology. 2016; 22: 477–479. (In Chinese)

[28] Liu J, Wang Z, Li M, Zhou M, Yu Y, Zhan W. Establishment of two new predictive models for prostate cancer to determine whether to require prostate biopsy when the PSA level is in the diagnostic gray zone (4–10 ng ml^{-1}). Asian Journal of Andrology. 2020; 22: 213–216.

[29] Spohn SKB, Bettermann AS, Bamberg F, Benndorf M, Mix M, Nicolay NH, *et al*. Radiomics in prostate cancer imaging for a personalized treatment approach-current aspects of methodology and a systematic review on validated studies. Theranostics. 2021; 11: 8027–8042.

[30] Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, *et al*. Introduction to radiomics. The Journal of Nuclear Medicine. 2020; 61: 488–495.

[31] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, *et al*. Radiomics: extracting more information from medical images using advanced feature analysis. European Journal of Cancer. 2012; 48: 441–446.

[32] Monti S, Brancato V, Di Costanzo G, Basso L, Puglia M, Ragozzino A, *et al*. Multiparametric MRI for prostate cancer detection: new insights into the combined use of a radiomic approach with advanced acquisition protocol. Cancers. 2020; 12: 390.

[33] Abboud B, Smayra T, Jabbour H, Ghorra C, Abadjian G. Correlations of neck ultrasound and pathology in cervical lymph node of papillary thyroid carcinoma. Acta Chirurgica Belgica. 2020; 120: 238–244.

[34] Zhang H, Li X, Zhang Y, Huang C, Wang Y, Yang P, *et al*. Diagnostic nomogram based on intralesional and perilesional radiomics features and clinical factors of clinically significant prostate cancer. Journal of Magnetic Resonance Imaging. 2021; 53: 1550–1558.

[35] Xian M, Zheng X, Xu J, Li X, Chen L, Wang W. Prediction of lymph node metastasis in rectal cancer: comparison between shear-wave elastography based ultrasomics and MRI. Diagnostic and Interventional Radiology. 2021; 27: 424–431.

[36] Jin X, Ai Y, Zhang J, Zhu H, Jin J, Teng Y, *et al*. Noninvasive prediction of lymph node status for patients with early-stage cervical cancer based on radiomics features from ultrasound images. European Radiology. 2020; 30: 4117–4124.

[37] Wang K, Chen P, Feng B, Tu J, Hu Z, Zhang M, *et al*. Machine learning prediction of prostate cancer from transrectal ultrasound video clips. Frontiers in Oncology. 2022; 12: 948662.

[38] Li M, Chen T, Zhao W, Wei C, Li X, Duan S, *et al*. Radiomics prediction model for the improved diagnosis of clinically significant prostate cancer on biparametric MRI. Quantitative Imaging in Medicine and Surgery. 2020; 10: 368–379.

[39] Qi X, Wang K, Feng B, Sun X, Yang J, Hu Z, *et al*. Comparison of machine learning models based on multi-parametric magnetic resonance imaging and ultrasound videos for the prediction of prostate cancer. Frontiers in Oncology. 2023; 13: 1157949.

[40] Nketiah G, Elschot M, Kim E, Teruel JR, Scheenen TW, Bathen TF, *et al*. T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: preliminary results. European Radiology. 2017; 27: 3050–3059.

[41] Chen M, Ma T, Li J, Zhang H, Li Q, Wang J, *et al*. Diagnosis of prostate cancer in patients with prostate-specific antigen (PSA) in the gray area: construction of 2 predictive models. Medical Science Monitor. 2021; 27: e929913.

[42] Liu T, Zhang X, Chen R, Deng X, Fu B. Development, comparison, and validation of four intelligent, practical machine learning models for patients with prostate-specific antigen in the gray zone. Frontiers in Oncology. 2023; 13: 1157384.

[43] Soeterik TFW, van Melick HHE, Dijksman LM, Biesma DH, Witjes JA, van Basten JA. Active surveillance for prostate cancer in a real-life cohort: comparing outcomes for PRIAS-eligible and PRIAS-ineligible patients. European Urology Oncology. 2018; 1: 231–237.

How to cite this article: Qin Wu, Chengyi Wu, Maoliang Zhang, Jie Yang, Junxiang Zhang, Yun Jin, *et al*. Diagnosing prostate cancer in the PSA gray zone through machine learning and transrectal ultrasound video. Journal of Men's Health. 2025; 21(5): 46-55. doi: 10.22514/jomh.2025.067.