Development and Validation of a Machine Learning-Based Radiomics Model Using Ultrasound Image Features for Prostate Cancer Risk Stratification

Ann. Ital. Chir., 2026 97, 1: 63–73
https://doi.org/10.62713/aic.4250

Anli Zhao[1,†], Shunlan Du[2,†], Yanhong Du[1], Maoliang Zhang[1], Jiangfeng Wu[1], Zhengping Wang[1], Liyan Hu[1]

[1]Ultrasound Department, Dongyang People's Hospital, 322100 Dongyang, Zhejiang, China
[2]Obstetrics Department, Dongyang People's Hospital, 322100 Dongyang, Zhejiang, China

AIM: This study aimed to construct a risk stratification model for prostate cancer (PCa) ultrasound imaging data and machine learning algorithms, with the goal of providing an effective tool for early diagnosis, personalized treatment, and clinical decision-making.
METHODS: A total of 211 histopathologically confirmed PCa patients were retrospectively enrolled and categorized into low-risk (n = 65), intermediate-risk (n = 55), and high-risk (n = 91) groups based on prostate-specific antigen levels, Gleason scores, and clinical T stage. From ultrasound images, 135 quantitative radiomic features—including morphological, texture, and edge descriptors—were extracted using the PyRadiomics toolkit. Feature dimensionality was reduced using the Pearson correlation coefficient (PCC), followed by recursive feature elimination (RFE) with 10-fold nested cross-validation to select the most informative features. Three machine learning algorithms—support vector machine (SVM), random forest (RF), and logistic regression (LR)—were trained and evaluated. Model performance was assessed using accuracy, sensitivity, specificity, and area under the curve (AUC).
RESULTS: The RF model achieved the best performance in both training and test cohorts, with AUCs of 0.87 and 0.86, and accuracies of 90% and 88%, respectively. DeLong's test confirmed that RF significantly outperformed SVM (p = 0.016) and LR (p = 0.004) in AUC comparison. The RF model also demonstrated robust predictive ability across risk subgroups: in the high-risk group, it achieved an AUC of 0.89, accuracy of 89%, sensitivity of 88%, and specificity of 90%; in the intermediate- and low-risk groups, AUCs were 0.86 and 0.81, respectively. Feature importance analysis revealed that wavelet-transformed Gray Level Dependence Matrix (GLDM) texture features, particularly DependenceEntropy and DependenceVariance, were the most predictive, highlighting the value of intratumoral textural heterogeneity in risk classification.
CONCLUSIONS: The RF-based ultrasound radiomics model enables accurate stratification of PCa risk, with remarkable performance in identifying high-risk patients.

Keywords: prostate cancer; ultrasound imaging; machine learning; risk stratification; prostate-specific antigen

Introduction

Prostate cancer (PCa) is one of the most common malignancies among men, with over 1.4 million new cases reported annually worldwide [1]. Marked by ever-increasing mortality rate, PCa has become the second leading cause of cancer-related death and the fifth leading cause of overall mortality in men [2]. Early detection and accurate risk stratification remain critical challenges in clinical management for PCa, particularly in the context of improving patient outcomes. Current screening strategies primarily rely on prostate-specific antigen (PSA) testing, digital rectal examination (DRE), and imaging-based assessments. Among these, PSA detection is widely used due to its accessibility; however, its limited specificity frequently results in false positives and unnecessary biopsies [3–5]. Magnetic resonance imaging (MRI) and computed tomography (CT) have demonstrated strong diagnostic and staging performance, but high cost, procedural complexity, and dependence on advanced infrastructure significantly limit their widespread adoption, especially in low-resource settings [6].

In contrast, ultrasound imaging has become a widely utilized tool in PCa screening and surveillance due to its non-invasive nature, cost-effectiveness, and accessibility [7]. Nevertheless, conventional ultrasound interpretation is often constrained by limited spatial resolution and operator dependency, making it insufficient for capturing subtle intratumoral heterogeneity [8]. Radiomics—a computational approach that extracts a large number of quantitative features from medical images—has emerged as a powerful method for characterizing tumor biology and enhanc-

Submitted: 6 July 2025 Revised: 8 August 2025 Accepted: 21 August 2025 Published: 10 January 2026
Correspondence to: Liyan Hu, Ultrasound Department, Dongyang People's Hospital, 322100 Dongyang, Zhejiang, China (e-mail: HLY2025sci@163.com; Zhengping Wang, Ultrasound Department, Dongyang People's Hospital, 322100 Dongyang, Zhejiang, China (e-mail: WZPscizs@163.com).
†*These authors contributed equally.*
Editor: Emanuele Casciani

ing diagnostic and prognostic precision [9,10]. Recent advances have demonstrated the synergistic potential of integrating radiomics with machine learning techniques to mine high-dimensional imaging features for facilitating clinical decision-making [11,12]. These methods enable automated pattern recognition from complex datasets, offering a data-driven framework for risk prediction and disease stratification [13].

Among various imaging modalities, MRI-based radiomics has shown particular promise. For example, Li *et al*. [14] developed a combined model incorporating a radiomic signature (Rad-score) and Prostate Imaging Reporting and Data System (PI-RADS) (v2.1, https://www.acr.org/Clinical-Resources/Clinical-Tools-and-Reference/Reporting-and-Data-Systems/PI-RADS), which achieved an area under the curve (AUC) of 0.979 in the training cohort and 0.931 in the validation cohort—outperforming PI-RADS alone. A systematic review further highlighted that machine learning models based on MRI radiomics can achieve pooled AUCs of up to 0.96 in identifying prostate cancer, with corresponding sensitivities and specificities of 0.92 and 0.90, respectively [15]. In parallel, ultrasound-based radiomics is gaining momentum. Sun *et al*. [16] constructed a model based on radiomic features extracted from transrectal B-mode and contrast-enhanced ultrasound (CEUS), significantly improving detection of PCa in the peripheral zone. These studies collectively demonstrate that even within the traditionally limited framework of ultrasound, the integration of radiomics and advanced ML algorithms can markedly enhance the diagnostic and stratification capabilities for PCa.

Risk stratification is central to guiding PCa management, as outlined in the National Comprehensive Cancer Network (NCCN) guidelines that recommend distinct strategies for low-, intermediate-, and high-risk patients [17]. Accurate early diagnosis and classification are therefore critical for personalized treatment and improved clinical outcomes. In this study, we propose a machine learning–driven model for PCa risk stratification based on ultrasound-derived radiomic features. By integrating texture, morphological, and edge descriptors, the model enables automated, quantitative analysis of raw ultrasound data and delivers robust predictions of cancer risk.

Methods

Study Design and Sample Selection

This study employed a retrospective cohort design to construct and validate a machine learning model based on ultrasound imaging features for PCa risk stratification. This study was approved by the Ethics Committee of the Dongyang People's Hospital (2025-YX-003). All data collection and analysis procedures adhered to the ethical guidelines approved by the institutional review board of our institution, and all patient data were de-identified to ensure privacy protection. All procedures involving human participants were conducted in accordance with the principles of the Declaration of Helsinki. The ultrasound imaging data used in this study were sourced from the electronic health records and imaging database of PCa patients at the Dongyang People's Hospital. All imaging data were collected by experienced radiologists and ultrasound technicians using standardized equipment, ensuring the consistency of image quality.

Inclusion Criteria: This study included male patients who were diagnosed with PCa pathologically between August 2021 and September 2023 at the Dongyang People's Hospital. All included patients had complete ultrasound imaging data and relevant clinical information (e.g., PSA levels, Gleason scores, clinical staging).

Exclusion Criteria: Patients were excluded if they met any of the following criteria: (1) incomplete imaging data or clinical information; (2) poor image quality (e.g., blurry or distorted images); and (3) inability to clearly delineate the prostate region in the imaging data.

A total of 211 patients were included in the final study cohort. PCa risk stratification was performed based on PSA levels and Gleason scores:

- Low-risk group: PSA $<$10 ng/mL, Gleason $\leq$6, clinical stage T1–T2a.
- Intermediate-risk group: PSA 10–20 ng/mL, Gleason = 7, clinical stage T2b–T2c.
- High-risk group: PSA $>$20 ng/mL, Gleason $\geq$8, clinical stage $\geq$T3a.

Ultrasound Imaging Data Acquisition and Preprocessing

All ultrasound examinations were performed using the Philips iU22 G4 ultrasound system (Philips Medical Systems, Bothell, WA, USA). Patients were placed in the left lateral decubitus position. After assessing the basic anatomy of the prostate gland, contrast-enhanced ultrasound was performed. Same imaging parameters (such as gain, depth, and dynamic range) were applied across all scans. For analysis, the maximum lesion plane—defined as the largest cross-sectional area of grayscale or color Doppler abnormality—was selected for quantitative feature extraction.

Before the analysis of imaging data, all ultrasound images underwent a standardized preprocessing workflow. Gaussian filtering was applied to reduce background noise and improve image clarity. All images were resized to a consistent resolution, and intensity normalization was performed to eliminate the impact of equipment and patient-related differences on image quality. Histogram equalization was applied to enhance the image contrast, making the prostate and tumor regions more prominent. The images were segmented automatically using 3D Slicer software (version 5.1.0, Bethesda Health Institute). The region of interest (ROI) was manually delineated, and the final output included 3D images containing the segmented ROI.

Radiomics Feature Extraction

Radiomics feature extraction involves deriving high-dimensional quantitative data from ultrasound images for further analysis using machine learning. Morphological features include geometric metrics of the tumor region, such as volume, surface area, long axis, short axis, and perimeter, reflecting the size and shape characteristics of the tumor. Texture features were extracted using methods such as the Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and Local Binary Patterns (LBP), which capture the structural complexity of the tumor. Edge roughness, edge uniformity, and other features describe the geometric properties and complexity of the tumor boundary. Frequency-domain features were extracted using Fourier and wavelet transforms to capture subtle changes and complexities of the tumor. All feature extraction was performed using the open-source radiomics software package PyRadiomics (version 3.1.0; developed by the PyRadiomics open-source community) [18]. For each image, the prostate region was first extracted using a mask, and the aforementioned features were subsequently obtained from this region. All extracted features were standardized to eliminate the impact of differences in image resolution, contrast, and scanning equipment.

Machine Learning Model Development

All enrolled patients (n = 211) were retrospectively divided into a training set and a test set in a 7:3 ratio, resulting in 148 cases in the training cohort (50 positive/98 negative) and 63 cases in the test cohort (21 positive/42 negative). All feature engineering steps, including normalization, dimensionality reduction, feature selection, and hyperparameter tuning, were performed strictly within the training cohort. The test set was reserved solely for final model evaluation to prevent data leakage.

Normalization

To mitigate the influence of differing feature scales, Z-score normalization was applied to the training dataset's feature matrix. Specifically, the mean and standard deviation were calculated for each feature vector. Subsequently, each feature vector was mean-centered and scaled by its standard deviation, resulting in features with a mean of 0 and a standard deviation of 1.

Dimensionality Reduction

Given the initial high dimensionality (135 features), Pearson correlation coefficient (PCC) filtering was implemented to reduce multicollinearity. PCC values were computed for all feature pairs, and for pairs exhibiting a PCC >0.86, one feature was randomly retained while the other was discarded. The 0.86 threshold was empirically determined to balance information retention with redundancy removal.

Feature Selection

Following dimensionality reduction, recursive feature elimination (RFE), in conjunction with nested 10-fold cross-validation (CV) to prevent information leakage, was employed for further feature selection.

External CV: The training set was partitioned into 10 folds. The external validation folds were used exclusively for evaluating the model's performance.

Internal CV: Within each external training fold, RFE and hyperparameter tuning were conducted. RFE, using a linear kernel support vector machine (SVM) as the base model, iteratively removed features with the smallest absolute weight coefficients. The optimal number of features was determined based on the area under the curve in the internal validation set.

To assess feature reproducibility, bootstrap resampling (1000 iterations) was performed on the training set. For each resampled dataset, the RFE procedure was repeated, and the selection frequency of each feature was recorded. Furthermore, the intraclass correlation coefficient (2,1) [ICC (2,1)] was calculated to quantify feature consistency. Features were considered stable if they exhibited a selection frequency $\geq$80% and an ICC $\geq$0.75.

Classification Models

The performance of three classification algorithms was evaluated:

• Support vector machine: A linear kernel (kernel = 'linear') was used, with the regularization parameter C optimized via 10-fold CV over the range [0.01, 0.1, 1, 10].

• Random forest (RF): The number of trees (n_estimators) was tuned across [100, 200, 500], and the maximum number of features (max_features) was tuned across ['sqrt', 'log2'].

• Logistic regression (LR): L2 regularization was employed, with the regularization parameter C optimized via 10-fold CV over the range [0.01, 0.1, 1, 10].

Hyperparameter optimization was performed within the internal loop of the nested cross-validation, using grid search with stratified sampling to maintain class proportions. For larger parameter spaces, a random search was initially conducted for pre-selection, followed by a grid search for fine-tuning. The optimal hyperparameter set was determined by maximizing the mean AUC across the internal CV folds.

Model Evaluation and Validation

Model performance was evaluated using receiver operating characteristic (ROC) analysis. The AUC was calculated to quantify performance. Additional performance metrics, including accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), were calculated using the optimal threshold to maximize the Youden index. The bootstrap method was applied to the original dataset with 1000 iterations of resampling to construct the empirical distribution of the target statistics (e.g.,

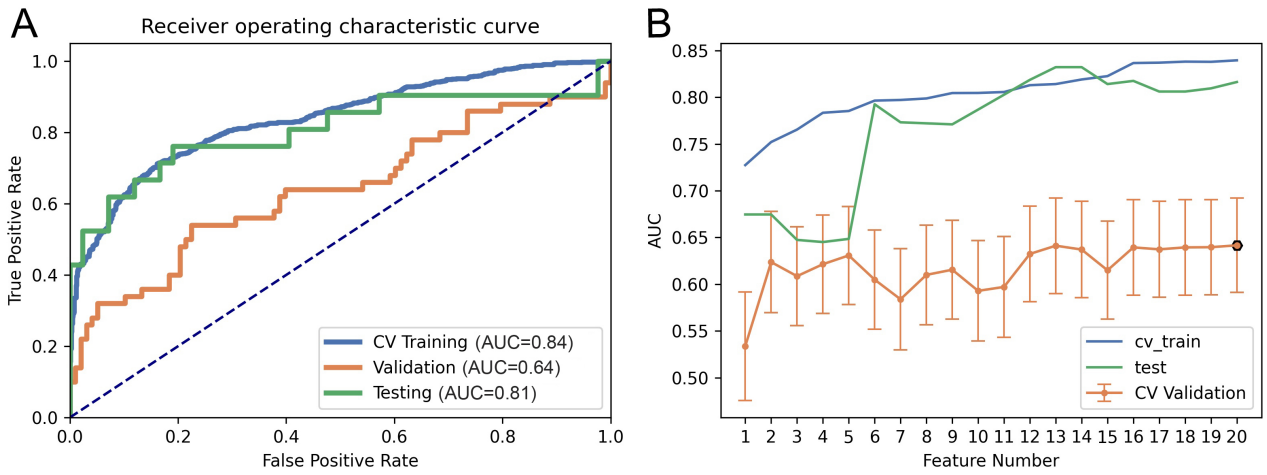

Fig. 1. Receiver operating characteristic (ROC) curve of the support vector machine (SVM) model for prostate cancer risk stratification. (A) ROC curve. (B) Relationship between hyperparameters. The blue line represents the model performance on the training set, the orange line indicates the validation set, and the green line denotes the test set's performance. Validation refers to the performance of a model on a cross-validation set. CV, cross-validation; AUC, area under the curve.

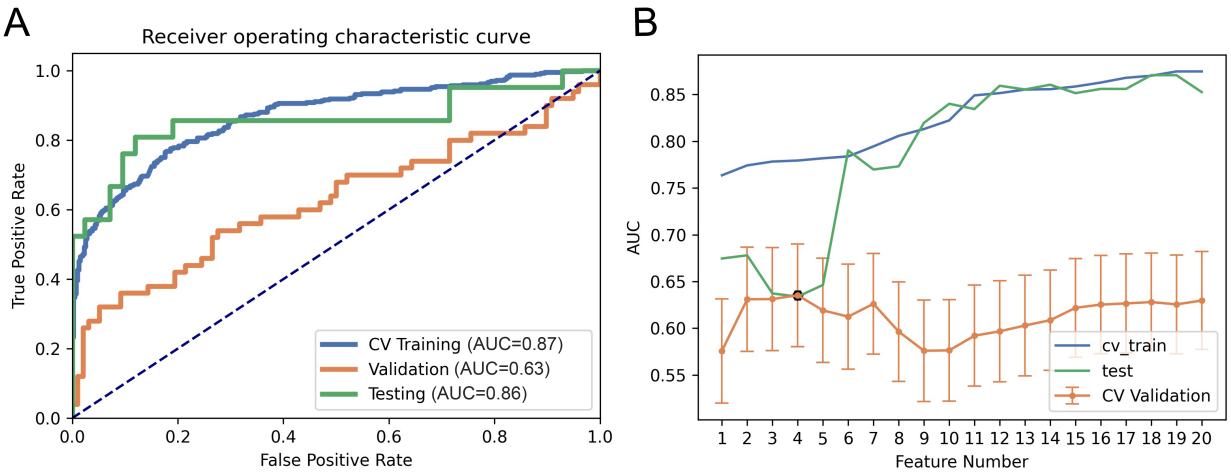

Fig. 2. ROC curve of the random forest (RF) model for prostate cancer risk stratification. (A) ROC curve. (B) Relationship between hyperparameters. The blue line represents the model performance on the training set, the orange line indicates the performance of the validation set, and the green line denotes the test set's performance. Validation refers to the performance of a model on a cross-validation set.

mean, regression coefficients), from which 95% confidence intervals were estimated.

Statistical Analysis

All data analyses were conducted using R software (version 4.0.1, R Foundation for Statistical Computing, Vienna, Austria). Manual segmentation of the ROIs was performed using 3D Slicer v5.1.0 software (developed by the Slicer Community, originally initiated by the Surgical Planning Laboratory at Brigham and Women's Hospital, Harvard Medical School). The "PyRadiomics" plugin in 3D Slicer was used to extract image features from each ROI. Forest plots for multivariable Cox regression models were visualized using the "forestplot" package in R (version 4.0.1; R Foundation for Statistical Computing, Vienna, Austria). Machine learning model development was performed using the FeAture Explorer Pro (FAEPro, v0.5.12, https://github.com/salan668/FAE) tool on Python (version 3.7.6; Python Software Foundation).

Results

Patient Characteristics

A total of 211 patients diagnosed with PCa were included in this study. Based on PSA levels, Gleason scores, and clinical staging, patients were stratified into low-risk (n = 65), intermediate-risk (n = 55), and high-risk (n = 91) groups. The mean age of the cohort was 65.39 ± 6.94 years, with no statistically significant difference in age among the three

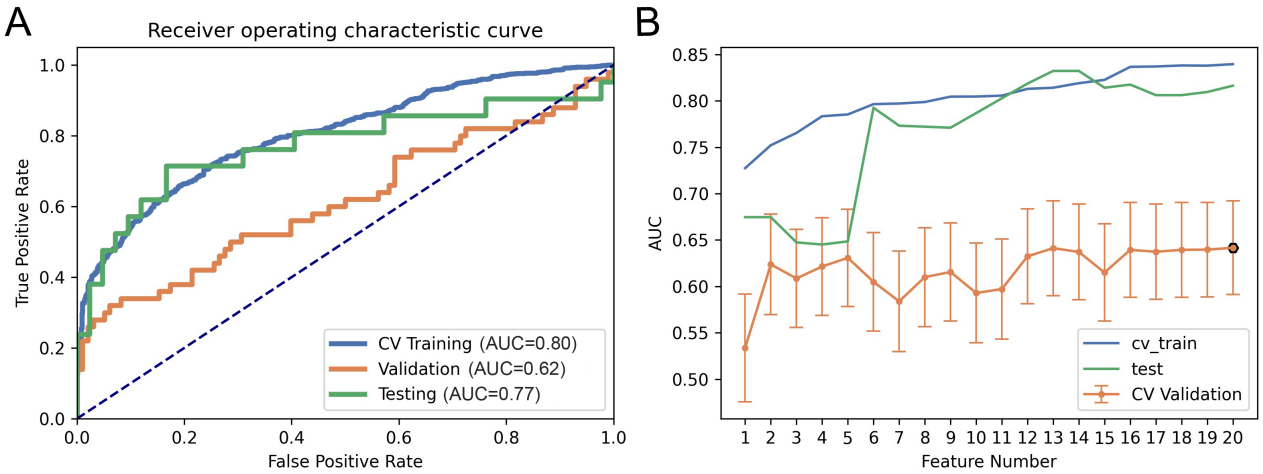

Fig. 3. Receiver operating characteristic curve of the logistic regression (LR) model for prostate cancer risk stratification. (A) ROC curve. (B) Relationship between hyperparameters. The blue line represents the model performance on the training set, the orange line indicates the performance of the validation set, and the green line denotes the test set's performance. Validation refers to the performance of a model on a cross-validation set.

Table 1. Demographic and clinical characteristics of the patients.

Variables	Total (n = 211)	Low-risk (n = 65)	Intermediate-risk (n = 55)	High-risk (n = 91)	Statistic	p
Age (years)	65.39 ± 6.94	64.58 ± 6.30	65.75 ± 6.85	65.75 ± 7.44	F = 0.65	0.525
PSA (ng/mL)	17.26 ± 9.86	6.19 ± 2.64	15.34 ± 2.20	26.32 ± 6.73	F = 340.08	<0.001
Gleason score	7.36 ± 1.39	5.84 ± 0.43	7.00 ± 0.00	8.66 ± 0.97	F = 337.99	<0.001
Clinical stage					χ^2 = 422.00	<0.001
T1	21 (9.95)	21 (32.31)	0 (0.00)	0 (0.00)		
T2a	44 (20.85)	44 (67.69)	0 (0.00)	0 (0.00)		
T2b	33 (15.64)	0 (0.00)	33 (60.00)	0 (0.00)		
T2c	22 (10.43)	0 (0.00)	22 (40.00)	0 (0.00)		
T3a	44 (20.85)	0 (0.00)	0 (0.00)	44 (48.35)		
T3b	32 (15.17)	0 (0.00)	0 (0.00)	32 (35.16)		
T4	15 (7.11)	0 (0.00)	0 (0.00)	15 (16.48)		

Notes: Categorical variables are expressed as count and percentage, whereas quantitative variables are presented as mean ± standard deviation. F, analysis of variance (ANOVA).

Abbreviation: PSA, prostate-specific antigen.

groups (p = 0.525). Serum PSA levels and Gleason scores significantly differed across the three risk groups (both $p < 0.001$). Clinical T staging also varied significantly between groups (χ^2 = 422.00, $p < 0.001$). All low-risk patients were classified as T1–T2a, intermediate-risk patients as T2b–T2c, and high-risk patients as T3–T4 (Table 1).

Machine Learning Model Training and Performance Evaluation

A total of 211 patients were divided into a training cohort (n = 148) and an independent test cohort (n = 63). Baseline characteristics between the training and test cohorts were compared using t-tests or Chi-square tests, with a significance level of $p > 0.05$ (Table 2). No statistically significant differences were observed in age, PSA level, Gleason score, and clinical stage between the training cohort and the independent test cohort (all $p > 0.05$), indicating good baseline balance between the two cohorts. A total of 135 quantitative imaging features were extracted from the ultrasound images, encompassing morphological, texture, edge, and high-order features. The stability results showed that 15 features satisfied election frequency ≥80% and ICC ≥0.75, and were included in the final model (Table 3).

To evaluate the performance of the risk stratification model, three machine learning algorithms—SVM, RF, and LR—were used. All models were trained using ten-fold cross-validation and evaluated on a test dataset. In the training set, the RF model achieved the highest performance with an accuracy of 90%, an AUC of 0.87 (95% CI: 0.83–0.91), sensitivity of 85%, and specificity of 90%. The SVM model also performed well (accuracy: 86%, AUC: 0.84, 95% CI: 0.79–0.88), while the LR model showed comparatively lower sensitivity and AUC (AUC: 0.80, 95% CI: 0.75–0.85) (Table 4, Figs. 1,2,3). In the test set, the RF model continued

Table 2. Comparison of baseline and clinical characteristics between training and test sets.

Variables	Training set (n = 148)	Test set (n = 63)	Statistic	p
Age (years)	65.58 ± 6.81	64.95 ± 7.25	t = 0.80	0.43
PSA (ng/mL)	17.67 ± 9.78	16.41 ± 10.04	t = 0.88	0.38
Gleason score	7.38 ± 1.34	7.30 ± 1.47	t = 0.69	0.49
Clinical stage			χ^2 = 3.98	0.66
T1	15 (10.1%)	6 (9.5%)		
T2a	31 (20.9%)	13 (20.6%)		
T2b	24 (16.2%)	9 (14.3%)		
T2c	16 (10.8%)	6 (9.5%)		
T3a	30 (20.3%)	14 (22.2%)		
T3b	23 (15.5%)	9 (14.3%)		
T4	9 (6.1%)	6 (9.5%)		

Notes: Categorical variables are expressed as count and percentage, whereas quantitative variables are presented as mean ± standard deviation.

Table 3. Verification results of feature stability.

Feature name	Selection frequency (%)	ICC (2,1)
wavelet-LLH_gldm_DependenceVariance	96.8	0.91
wavelet-HHL_gldm_DependenceVariance	95.4	0.90
wavelet-LHH_gldm_DependenceEntropy	94.1	0.89
wavelet-HLH_glrlm_RunLengthNonUniformityNormalized	93.5	0.88
original_shape_SurfaceVolumeRatio	93.2	0.87
original_shape_MajorAxisLength	92.7	0.88
wavelet-HLH_gldm_DependenceVariance	91.9	0.87
wavelet-LHH_gldm_DependenceNonUniformityNormalized	90.8	0.86
wavelet-LLH_glrlm_RunLengthNonUniformityNormalized	89.6	0.86
wavelet-HLH_gldm_LargeDependenceEmphasis	88.3	0.85
wavelet-LHH_gldm_SmallDependenceLowGrayLevelEmphasis	87.5	0.85
wavelet-HLH_gldm_SmallDependenceLowGrayLevelEmphasis	85.7	0.83
wavelet-HLH_glrlm_RunEntropy	83.9	0.82
wavelet-LHH_glrlm_RunLengthNonUniformityNormalized	82.1	0.81
wavelet-LHL_gldm_DependenceNonUniformityNormalized	81.2	0.77

Abbreviations: LHH, Low-pass X, High-pass Y, High-pass Z; HLH, High-pass X, Low-pass Y, High-pass Z; LLH, Low-pass X, Low-pass Y, High-pass Z; HHL, High-pass X, High-pass Y, Low-pass Z; ICC, intraclass correlation coefficient; LHL, Low-pass X, High-pass Y, Low-pass Z.

to demonstrate superior performance, with an accuracy of 88%, AUC of 0.86 (95% CI: 0.83–0.90), sensitivity of 79%, and specificity of 90%. SVM and LR models achieved AUCs of 0.81 (95% CI: 0.77–0.85) and 0.77 (95% CI: 0.72–0.82), respectively (Table 5, Figs. 1,2,3).

DeLong's test further confirmed that RF significantly outperformed both SVM (p = 0.016) and LR (p = 0.004) in AUC comparison, while the difference between SVM and LR was not statistically significant (p = 0.128) (Table 6). These results indicate that the RF model provided the most reliable and generalizable performance for PCa risk stratification.

Prediction Performance of RF Model for Different Risk Subgroups

The predictive performance of the RF model was further evaluated across different PCa risk subgroups. In the low-risk group, the model achieved an AUC of 0.81 (95% CI: 0.76–0.86), with an accuracy of 84%, sensitivity of 75%, and specificity of 88%. In the intermediate-risk group, the RF model's performance was higher, with an AUC of 0.86 (95% CI: 0.82–0.90), accuracy of 86%, sensitivity of 83%, and specificity of 88%. The high-risk group demonstrated the best predictive results, with an AUC of 0.89 (95% CI: 0.86–0.92), accuracy of 89%, sensitivity of 88%, and specificity of 90% (Table 7, Fig. 4).

Feature Importance Analysis

Feature importance analysis using the RF model revealed that texture features, such as wavelet-LHH gldm DependenceEntropy, wavelet-HHL gldm DependenceVariance, and wavelet-LLH gldm DependenceVariance, were the most important predictors for PCa risk. These features consistently ranked highly in the feature importance analysis,

Table 4. Performance comparison of machine learning models on the training set.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (95% CI)	PPV	NPV
SVM	86	80	89	0.84 (0.79–0.88)	0.76	0.74
RF	90	85	90	0.87 (0.83–0.91)	0.80	0.85
LR	85	71	86	0.80 (0.75–0.85)	0.72	0.80

Abbreviations: AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; RF, random forest; LR, logistic regression; SVM, support vector machine.

Table 5. Performance comparison of machine learning models on the test set.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (95% CI)	PPV	NPV
SVM	83	77	92	0.81 (0.77–0.85)	0.75	0.72
RF	88	79	90	0.86 (0.83–0.90)	0.76	0.77
LR	81	76	88	0.77 (0.72–0.82)	0.73	0.70

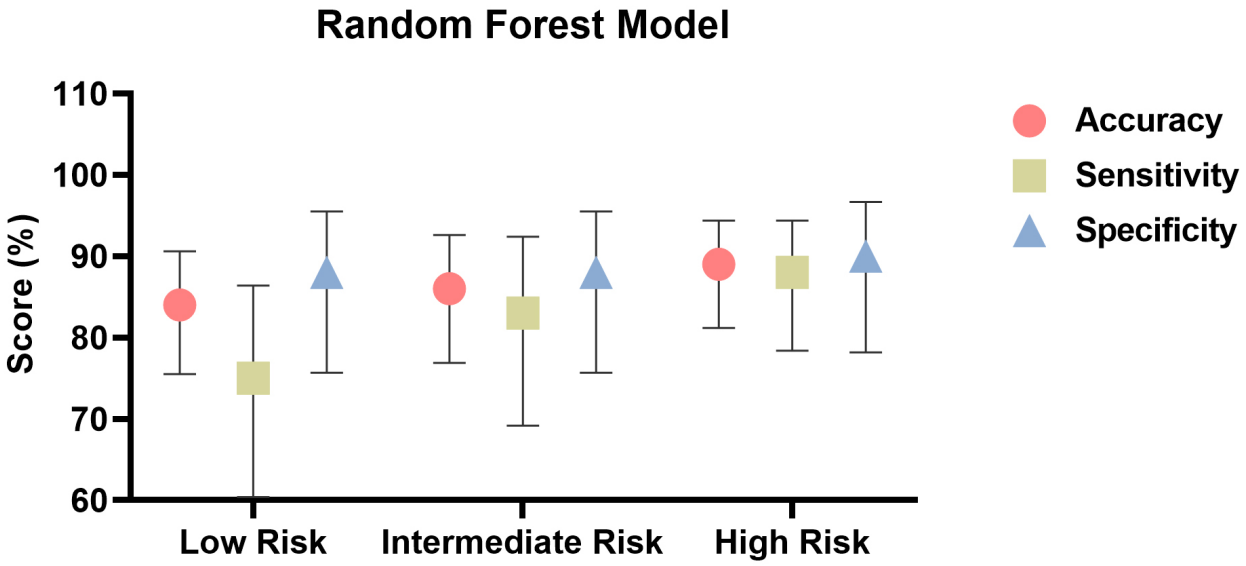

Fig. 4. Prediction performance of the RF model for different prostate cancer risk groups.

Table 6. DeLong's test for AUC comparisons.

Comparison	Difference in AUC	Z-statistic	*p*
RF vs. SVM	0.05	2.41	0.016
RF vs. LR	0.09	2.89	0.004
SVM vs. LR	0.04	1.52	0.128

indicating that texture information plays a crucial role in distinguishing high-risk from low-risk PCa cases (Fig. 5).

Discussion

This study integrated ultrasound imaging data with advanced machine learning algorithms to establish a PCa risk stratification model. By comparing this model with traditional clinical risk assessment methods, such as PSA levels and Gleason scores, we explored its potential in facilitating early diagnosis and personalized treatment decisions for PCa. The results showed that the RF model based on ultrasound imaging significantly outperformed traditional clinical methods in prediction accuracy, sensitivity, specificity, and the ability to differentiate between risk groups, especially in identifying high-risk patients.

Study has demonstrated the high value of combining imaging with machine learning algorithms for diagnosing and predicting PCa risk [19]. Our study, which utilized machine learning algorithms to mine ultrasound imaging features, confirmed the advantages of machine learning models in PCa risk stratification. For instance, Liu *et al.* [20] also achieved positive results using MRI combined with machine learning for PCa risk stratification. However, MRI is expensive and requires specialized operation, limiting its widespread application. In contrast, ultrasound is a low-cost, accessible, and non-invasive imaging modality, making it more feasible for clinical use [21]. This study indicates that ultrasound combined with machine learning offers an efficient and cost-effective method for PCa risk assessment, with good prospects for clinical implementation.

Early detection of prostate cancer is crucial for improving

Table 7. Prediction performance of the random forest model for different prostate cancer risk groups.

Risk stratification group	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (95% CI)
Low-risk	84	75	88	0.81 (0.76–0.86)
Intermediate-risk	86	83	88	0.86 (0.82–0.90)
High-risk	89	88	90	0.89 (0.86–0.92)

Fig. 5. Ranking of top imaging features contributing to prostate cancer risk classification, as determined by the random forest model. Texture features derived from wavelet-transformed images, particularly gldm DependenceEntropy and DependenceVariance from different frequency bands, were the most significant in distinguishing between risk categories, underscoring the value of radiomic texture analysis in prostate cancer assessment. L/H represents Low-pass or High-pass filtering, applied along the three directions (X, Y, Z) in wavelet decomposition. LHH, Low-pass X, High-pass Y, High-pass Z; HLH, High-pass X, Low-pass Y, High-pass Z; LLH, Low-pass X, Low-pass Y, High-pass Z; HHL, High-pass X, High-pass Y, Low-pass Z.

patient prognosis. In our study, we stratified patients based on PSA levels and Gleason scores, which are the most commonly used clinical risk indicators. Both PSA and Gleason scores play a vital role in early detection and risk prediction of PCa [22]. However, they each have limitations. PSA levels can be influenced by various physiological factors, and their performance in terms of specificity and sensitivity varies significantly across different patient populations, especially in those with low-risk or locally advanced disease [23]. Although Gleason scoring remains the standard pathological grading system for PCa, it is inherently subjective and relies on pathologists' judgment of tissue slides, which may introduce variability and potential diagnostic inconsistencies [24]. Therefore, overcoming these limitations and improving early screening and precise classification of prostate cancer remain a pressing clinical challenge.

In this study, we chose three classic machine learning algorithms—SVM, RF, and LR—to perform risk stratification based on ultrasound imaging features. These three algorithms are widely used in medical imaging, each with its advantages and disadvantages. SVM is particularly effective with high-dimensional datasets, but may face challenges with longer training times when the dataset size is large [25]. LR is computationally simpler but may not perform as well with complex data patterns [26]. RF, as an ensemble learning method that constructs multiple decision trees, has strong noise resistance and generalization capabilities, and has been successfully applied in various medical imaging analyses [27]. In our study, the RF model demonstrated strong and consistent performance in both the training and test cohorts, with AUCs of 0.87 and 0.86, respectively, underscoring the potential of RF in prostate cancer imaging analytics. Notably, in the high-risk group, the model achieved an AUC of 0.89, along with high accuracy (89%), sensitivity (88%), and specificity (90%), indicating its ability to effectively distinguish between low- and high-risk patients. A previous study using RF for PCa risk stratification based on MRI similarly reported an overall AUC of

0.87 and an AUC of 0.89 in high-risk patients [28], further supporting the robustness of our findings. RF offers intrinsic advantages in handling high-dimensional, imbalanced radiomic datasets due to its ensemble architecture (bagging) and robust assessment of feature importance, often outperforming linear models such as SVM and LR [29,30].

Feature importance analysis revealed that texture features derived from wavelet-transformed images—particularly wavelet-LHH gldm DependenceEntropy and wavelet-HHL gldm DependenceVariance—were among the most predictive variables for PCa risk stratification. These features were extracted using the Gray Level Dependence Matrix (GLDM), which quantifies the spatial relationships and intensity dependencies between neighboring voxels within tumor regions [31]. DependenceEntropy measures the randomness or complexity of gray-level dependencies, with higher values reflecting greater textural heterogeneity [32]. Elevated DependenceEntropy may indicate increased cellular heterogeneity, disrupted glandular architecture, or stromal irregularity—hallmarks of more aggressive tumor phenotypes [33]. Similarly, DependenceVariance quantifies the variability in dependence sizes, capturing fluctuations in local texture patterns that may correspond to structural disorganization within the tumor microenvironment [34]. Wavelet transformations further enhance the sensitivity of these descriptors by capturing multi-scale texture variations, which are critical for detecting subtle spatial heterogeneity that may not be apparent in raw imaging [35]. The prominence of these features in our model supports the hypothesis that radiomic texture biomarkers can serve as non-invasive surrogates for tumor heterogeneity, which is a well-recognized determinant of malignancy, treatment resistance, and clinical outcome. These findings provide mechanistic insights and emphasize the potential of radiomics in characterizing the underlying biology of PCa.

Taken together, the proposed model demonstrates strong potential for clinical translation. Ultrasound is cost-effective, radiation-free, and widely accessible—especially in primary care settings. Combined with our automated radiomics pipeline, clinicians can rapidly segment images, extract features, and generate quantitative risk scores after obtaining ultrasound scans. This workflow can be integrated into existing hospital Picture Archiving and Communication System (PACS) or Clinical Decision Support System (CDSS), enabling real-time risk assessment without additional patient burden. Furthermore, the model's interpretability through feature importance outputs may enhance clinical trust and decision-making. Future integration with multimodal imaging and biomarker data may further accelerate the implementation of precision medicine in PCa management.

Despite the promising performance of our model, several limitations warrant consideration. First, the study was conducted using data from a single institution, which may limit the generalizability of the findings. External validation on multicenter cohorts is essential to assess the model's robustness across diverse populations and clinical settings. Second, although a wide range of radiomic features (n = 135) were extracted, the relatively small sample size (n = 211) raises concerns regarding model stability and the risk of overfitting inherent in high-dimensional datasets with limited observations. This high feature-to-sample ratio can impair generalization, particularly in the absence of external data or feature reduction strategies. Third, the current model development relied on a single train-test split, which may not provide a sufficiently rigorous evaluation. Future work should incorporate repeated cross-validation or nested cross-validation schemes to ensure more reliable performance estimation. Furthermore, while conventional machine learning approaches were employed in this study, integrating deep learning methods such as convolutional neural networks (CNNs) may enhance feature representation and improve model accuracy, especially when larger annotated datasets become available. Lastly, the inclusion of multimodal imaging (e.g., MRI) and genomic data could further enrich the predictive capacity of radiomics-based models and support more comprehensive, personalized PCa risk assessment.

Conclusions

This study constructed an effective PCa risk stratification model based on ultrasound imaging radiomics features and machine learning algorithms, demonstrating the superior performance of the random forest algorithm in distinguishing high-risk and low-risk patients. The extraction and analysis of texture features offer new insights for early diagnosis, especially in resource-limited settings. Ultrasound imaging combined with radiomics methods holds broad clinical application potential. However, further research is needed to validate the stability and generalizability of the model and address its interpretability issues, with the aim of providing stronger support for precise diagnosis and personalized treatment of prostate cancer.

Availability of Data and Materials

The data analyzed are available from the corresponding authors upon reasonable request.

Author Contributions

ALZ and SLD wrote the manuscript and contributed equally to this work. ALZ and MLZ conceived and designed the study, performed formal analysis, and were responsible for the design and analysis of ultrasound data. SLD and YHD conducted clinical assessments, collected required data, and assisted in methodology implementation. JFW conceived and designed the study, assisted in data analysis, and contributed to manuscript preparation and visualization. LYH and ZPW conceived and designed the study, supervised the project and ensured methodological rigor. All authors have been involved in revising the manuscript critically for im-

portant intellectual content. All authors gave final approval of the version to be published. All authors have participated sufficiently in the work to take public responsibility for appropriate portions of the content and agreed to be accountable for all aspects of the work in ensuring that questions related to its accuracy or integrity.

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of the Dongyang People's Hospital (2025-YX-003). Given the use of fully anonymized patient information, informed consent was waived. All procedures involving human participants were conducted in accordance with the principles of the Declaration of Helsinki.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Belkahla S, Nahvi I, Biswas S, Nahvi I, Ben Amor N. Advances and development of prostate cancer, treatment, and strategies: A systemic review. Frontiers in Cell and Developmental Biology. 2022; 10: 991330. https://doi.org/10.3389/fcell.2022.991330.

[2] Chakroborty D, Singh AP. Prostate Cancer: Insights into Disease Progression and Therapeutic Challenges. International Journal of Molecular Sciences. 2024; 25: 2451. https://doi.org/10.3390/ijms25052451.

[3] Descotes JL. Diagnosis of prostate cancer. Asian Journal of Urology. 2019; 6: 129–136. https://doi.org/10.1016/j.ajur.2018.11.007.

[4] Mikulas CJ, Ali K, Onteddu N. Prostate cancer screening: Is it time for a new approach? A review article. Journal of Investigative Medicine: the Official Publication of the American Federation for Clinical Research. 2025; 73: 27–34. https://doi.org/10.1177/10815589241279414.

[5] Juliebø-Jones P, Beisland C. Prostate Specific Antigen (PSA) Testing and Cancer Detection: Guidance Notes for the Non-Urologist. British Journal of Hospital Medicine. 2025; 86: 1–5. https://doi.org/10.12968/hmed.2024.1035.

[6] Liu L, Tian Z, Zhang Z, Fei B. Computer-aided Detection of Prostate Cancer with MRI: Technology and Applications. Academic Radiology. 2016; 23: 1024–1046. https://doi.org/10.1016/j.acra.2016.03.010.

[7] Parker P, Twiddy M, Whybrow P, Rigby A, Simms M. The role of diagnostic ultrasound imaging for patients with known prostate cancer within an active surveillance pathway: A systematic review. Ultrasound (Leeds, England). 2022; 30: 4–17. https://doi.org/10.1177/1742271X21995212.

[8] Avola D, Cinque L, Fagioli A, Foresti GL, Mecca A. Ultrasound Medical Imaging Techniques: A Survey. ACM Computing Surveys (CSUR). 2021; 54: 1–38. https://doi.org/10.1145/3447243.

[9] Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, *et al*. Machine and deep learning methods for radiomics. Medical Physics. 2020; 47: e185–e202. https://doi.org/10.1002/mp.13678.

[10] Mao KZ, Ma C, Song B. Radiomics advances in the evaluation of pancreatic cystic neoplasms. Heliyon. 2024; 10: e25535. https://doi.org/10.1016/j.heliyon.2024.e25535.

[11] Liu B, Fang L, Xiong Y, Du Q, Xiang Y, Chen X, *et al*. A Machine Learning Model Based on Genetic and Traditional Cardiovascular Risk Factors to Predict Premature Coronary Artery Disease. Frontiers in Bioscience (Landmark Edition). 2022; 27: 211. https://doi.org/10.31083/j.fbl2707211.

[12] Wu S, Li H. Medical Image and Physics Analysis of Human Physiology and Dysfunction. Pakistan Journal of Medical Sciences. 2021; 37: 1550–1552. https://doi.org/10.12669/pjms.37.6-WIT.5005.

[13] Cui Y, Li Y, Xing D, Bai T, Dong J, Zhu J. Improving the Prediction of Benign or Malignant Breast Masses Using a Combination of Image Biomarkers and Clinical Parameters. Frontiers in Oncology. 2021; 11: 629321. https://doi.org/10.3389/fonc.2021.629321.

[14] Li M, Yang L, Yue Y, Xu J, Huang C, Song B. Use of Radiomics to Improve Diagnostic Performance of PI-RADS v2.1 in Prostate Cancer. Frontiers in Oncology. 2021; 10: 631831. https://doi.org/10.3389/fonc.2020.631831.

[15] Zhao Y, Zhang L, Zhang S, Li J, Shi K, Yao D, *et al*. Machine learning-based MRI imaging for prostate cancer diagnosis: systematic review and meta-analysis. Prostate Cancer and Prostatic Diseases. 2025; 10.1038/s41391-025-00997-2. https://doi.org/10.1038/s41391-025-00997-2.

[16] Sun Y, Fang J, Shi Y, Li H, Wang J, Xu J, *et al*. Machine learning based on radiomics features combing B-mode transrectal ultrasound and contrast-enhanced ultrasound to improve peripheral zone prostate cancer detection. Abdominal Radiology (New York). 2024; 49: 141–150. https://doi.org/10.1007/s00261-023-04050-5.

[17] Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, *et al*. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. European Urology. 2021; 79: 243–262. https://doi.org/10.1016/j.eururo.2020.09.042.

[18] 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. https://doi.org/10.1158/0008-5472.CAN-17-0339.

[19] Chiu PKF, 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 and Prostatic Diseases. 2022; 25: 672–676. https://doi.org/10.1038/s41391-021-00429-x.

[20] Liu G, Pan S, Zhao R, Zhou H, Chen J, Zhou X, *et al*. The added value of AI-based computer-aided diagnosis in classification of cancer at prostate MRI. European Radiology. 2023; 33: 5118–5130. https://doi.org/10.1007/s00330-023-09433-2.

[21] Abou-Elkacem L, Bachawal SV, Willmann JK. Ultrasound molecular imaging: moving toward clinical translation. European Journal of Radiology. 2015; 84: 1685–1693. https://doi.org/10.1016/j.ejrad.2015.03.016.

[22] Ilic D, Djulbegovic M, Jung JH, Hwang EC, Zhou Q, Cleves A, *et al*. Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. BMJ (Clinical Research Ed.). 2018; 362: k3519. https://doi.org/10.1136/bmj.k3519.

[23] Distler FA, Radtke JP, Bonekamp D, Kesch C, Schlemmer HP, Wieczorek K, *et al*. The Value of PSA Density in Combination with PI-RADS™ for the Accuracy of Prostate Cancer Prediction. The Journal of Urology. 2017; 198: 575–582. https://doi.org/10.1016/j.juro.2017.03.130.

[24] Xue C, Yuan J, Lo GG, Poon DMC, Chu WCW. Evaluation of the Reliability and the Performance of Magnetic Resonance Imaging Radiomics in the Presence of Randomly Generated Irrelevant Features for Prostate Cancer. Diagnostics (Basel, Switzerland). 2023; 13: 3580. https://doi.org/10.3390/diagnostics13233580.

[25] Gatou P, Tsiara X, Spitalas A, Sioutas S, Vonitsanos G. Artificial Intelligence Techniques in Grapevine Research: A Comparative

Study with an Extensive Review of Datasets, Diseases, and Techniques Evaluation. Sensors (Basel, Switzerland). 2024; 24: 6211. https://doi.org/10.3390/s24196211.

[26] Pinheiro-Guedes L, Martinho C, Martins MR. Logistic Regression: Limitations in the Estimation of Measures of Association with Binary Health Outcomes. Acta Médica Portuguesa. 2024; 37: 697–705. https://doi.org/10.20344/amp.21435.

[27] Qiu H, Luo L, Su Z, Zhou L, Wang L, Chen Y. Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. BMC Medical Informatics and Decision Making. 2020; 20: 83. https://doi.org/10.1186/s12911-020-1101-8.

[28] Shu X, Liu Y, Qiao X, Ai G, Liu L, Liao J, *et al*. Radiomic-based machine learning model for the accurate prediction of prostate cancer risk stratification. The British Journal of Radiology. 2023; 96: 20220238. https://doi.org/10.1259/bjr.20220238.

[29] Li Z, Yang L, Wang X, Xu H, Chen W, Kang S, *et al*. Intralesional and perilesional radiomics strategy based on different machine learning for the prediction of international society of urological pathology grade group in prostate cancer. BMC Medical Imaging. 2025; 25: 266. https://doi.org/10.1186/s12880-025-01812-z.

[30] Jiang M, Miao Z, Xu R, Guo M, Li X, Li G, *et al*. Clinical-radiomics hybrid modeling outperforms conventional models: machine learning enhances stratification of adverse prognostic features in prostate cancer. Frontiers in Oncology. 2025; 15: 1625158. https://doi.org/10.3389/fonc.2025.1625158.

[31] Salmanpour MR, Amiri S, Gharibi S, Shariftabrizi A, Xu Y, Weeks WB, *et al*. Radiological and Biological Dictionary of Radiomics Features: Addressing Understandable AI Issues in Personalized Prostate Cancer, Dictionary Version PM1.0. Journal of Imaging Informatics in Medicine. 2025; 10.1007/s10278-025-01585-5. https://doi.org/10.1007/s10278-025-01585-5.

[32] Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. Computer Graphics and Image Processing. 1983; 23: 341–352. https://doi.org/10.1016/0734-189X(83)90032-4.

[33] Fan M, Chen H, You C, Liu L, Gu Y, Peng W, *et al*. Radiomics of Tumor Heterogeneity in Longitudinal Dynamic Contrast-Enhanced Magnetic Resonance Imaging for Predicting Response to Neoadjuvant Chemotherapy in Breast Cancer. Frontiers in Molecular Biosciences. 2021; 8: 622219. https://doi.org/10.3389/fmolb.2021.622219.

[34] Zhang B, Zhang H, Li X, Jin S, Yang J, Pan W, *et al*. Can Radiomics Provide Additional Diagnostic Value for Identifying Adrenal Lipid-Poor Adenomas From Non-Adenomas on Unenhanced CT? Frontiers in Oncology. 2022; 12: 888778. https://doi.org/10.3389/fonc.2022.888778.

[35] Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics. 1973; SMC-3-6: 610–621. https://doi.org/10.1109/TSMC.1973.4309314.