Article

Discovery Medicine 2025; 37(203): 3003–3012
https://doi.org/10.24976/Discov.Med.202537203.251

CNN-Based Deep Learning for MRI-Assisted Classification of Benign and Malignant Prostate Nodules: A Diagnostic Performance Study

Tianchang Guo[1], Xiaomei Zeng[1], Caiyun Wu[1], Chongyue Cai[2], Qiwei Xu[1], Ruihua Yao[1], Yuzhi Liang[1,*]

[1]Department of Medical Imaging, The Affiliated Dongguan Songshan Lake Central Hospital, Guangdong Medical University, 523320 Dongguan, Guangdong, China
[2]Department of Urology, The Affiliated Dongguan Songshan Lake Central Hospital, Guangdong Medical University, 523320 Dongguan, Guangdong, China
*Correspondence: 13416243629@163.com (Yuzhi Liang)

Submitted: 18 September 2025 Revised: 4 November 2025 Accepted: 14 November 2025 Published: 20 December 2025

Background: Prostate cancer (PCa) is one of the most common malignancies among men worldwide, and accurate differentiation between benign and malignant nodules remains challenging. Magnetic resonance imaging (MRI) provides valuable soft-tissue contrast but still suffers from interpretive variability. Deep learning–based computer-aided diagnostic (CAD) systems may help improve diagnostic accuracy and consistency. Therefore, this study aimed to evaluate a convolutional neural network (CNN)–based deep learning system for computer-aided classification of benign versus malignant prostate nodules using multiparametric MRI (mpMRI).

Methods: In this retrospective study, 80 patients with histopathologically confirmed prostate nodules (39 malignant, 41 benign) were enrolled between January 2021 and June 2024. Univariate and multivariate analyses were conducted to identify key imaging risk factors associated with malignancy. The diagnostic performance of conventional MRI was compared with that of five CNN-assisted models (CNN-1 to CNN-5) based on sensitivity, specificity, and overall accuracy.

Results: Significant differences ($p < 0.001$) were observed between malignant and benign groups in imaging features, including signal distribution ($\chi^2 = 31.473$), lesion margins ($\chi^2 = 19.776$), lesion volume ($t = 19.421$), short-axis diameter ($t = 10.337$), long-axis diameter ($t = 9.071$), and the product of diameters ($t = 6.548$). Multivariate logistic regression identified these parameters as independent malignancy predictors, with odds ratios (ORs) ranging from 2.818 to 3.277 across training and validation cohorts. These variables were incorporated into a malignancy risk score model. Among all CNN models, CNN-3, characterized by three max-pooling layers, achieved the highest diagnostic sensitivity and accuracy. Receiver operating characteristic (ROC) analysis further confirmed its superior performance, demonstrating the largest area under the curve (AUC) and outperforming both conventional MRI and other CNN variants.

Conclusion: The CNN-based deep learning diagnostic system significantly enhances the classification accuracy of prostate nodules on mpMRI. The CNN-3 model enables automated lesion detection and feature extraction, improving early diagnosis and risk stratification. It shows strong potential to support clinical decision-making in prostate cancer management.

Keywords: convolutional neural network; deep learning; multiparametric MRI; prostate cancer; prostate hyperplasia; computer-aided diagnosis

Introduction

Prostate cancer (PCa) is one of the most common malignant tumors affecting men worldwide [1–3]. With the ongoing global trend of population aging, its incidence continues to rise. Early diagnosis plays a pivotal role in improving treatment outcomes and prognosis in patients with PCa. Among available diagnostic modalities, magnetic resonance imaging (MRI) has emerged as a valuable imaging tool due to its non-invasive nature and superior soft-tissue contrast resolution [4,5]. MRI provides detailed anatomical and tissue-specific information of the prostate and facilitates differentiation between benign and malignant lesions, assessment of tumor extent, and guidance of therapeutic decisions [6,7].

In particular, T2-weighted imaging (T2WI) provides high-resolution images that enable clear visualization of prostate architecture and lesion morphology. PCa typically appears as a hypointense focus on T2WI, in contrast to the hyperintense signals of benign prostatic hyperplasia (BPH), thus enhancing diagnostic discrimination [8]. MRI has demonstrated marked advantages in differentiating PCa from benign conditions, particularly in the peripheral and transition zones [9].

Currently, multiparametric MRI (mpMRI) protocols are widely employed in the clinical assessment of PCa and include diffusion-weighted imaging (DWI), dynamic contrast-enhanced MRI (DCE-MRI), and magnetic resonance spectroscopy (MRS). DWI evaluates tissue cellularity by measuring water molecule diffusion, with malignancies typically exhibiting hyperintensity due to restricted diffusion resulting from increased cell density. DCE-MRI reflects tumor perfusion and vascular permeability, with prostate tumors often showing early contrast uptake and washout patterns consistent with neoangiogenesis [10]. The integration of these imaging sequences provides complementary biological information, thereby enhancing the diagnostic accuracy of PCa [11].

Despite these advances, MRI still faces notable limitations, including false-negative and false-positive findings. Lesions located in peripheral or anatomically complex zones may present ambiguous imaging features, compromising sensitivity and increasing the risk of misdiagnosis. Additionally, small or low-grade tumors may lack classical imaging signatures, further complicating detection. Accurate interpretation of prostate MRI requires substantial radiological expertise, especially when synthesizing multi-sequence data. Consequently, reader variability and subjectivity remain significant challenges. In recent years, computer-aided diagnostic (CAD) systems powered by convolutional neural networks (CNNs), a form of deep learning architecture, have shown great promise in prostate MRI analysis [12]. CNNs emulate the structure of biological neural networks and can automatically learn hierarchical and abstract features from large volumes of medical imaging data. This facilitates objective, efficient, and reproducible classification of prostate lesions, potentially reducing interobserver variability and enhancing diagnostic performance [13].

Therefore, this study aims to investigate the efficacy of a CNN-based deep learning approach for computer-aided diagnosis and malignancy classification of prostate nodules using MRI. The objective is to provide a scientific basis for improving diagnostic precision and supporting clinical decision-making in PCa management.

Methods

Patient Population

This retrospective study enrolled 80 patients with pathologically confirmed prostatic nodules who underwent MRI examination at our institution between January 2021 and June 2024. All patients met the clinical indications for prostate surgery and were subsequently confirmed by postoperative histopathology. The inclusion criteria were as follows: (1) histopathologically confirmed prostate nodules, with all malignant lesions meeting diagnostic criteria for PCa; (2) complete clinical and imaging data; and (3) eligibility for surgical intervention. The exclusion criteria were as follows: (1) presence of other malignancies; (2) digital rectal examination within 7 days prior to MRI; and (3) suboptimal or incomplete MRI imaging.

MRI Acquisition

Prostate MRI was performed at The Affiliated Dongguan Songshan Lake Central Hospital using a 1.5 Tesla Siemens MAGNETOM Avanto scanner (Siemens Healthineers, Erlangen, Germany). Imaging sequences included axial T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), and sagittal and coronal T2_tirm for structural assessment. Dynamic contrast-enhanced (DCE) imaging was conducted after intravenous injection of gadopentetate dimeglumine (Gd-DTPA, Magnevist®, Bayer Healthcare, Berlin, Germany) at a dose of 0.1 mmol/kg and a rate of 2.5 mL/s, followed by a 25 mL saline flush. Scanning commenced 25 seconds post-injection with the following parameters: repetition time (TR) = 4.89 ms, echo time (TE) = 2.38 ms, flip angle = 30°, and slice thickness = 3.5 mm across 6 dynamic phases. Diffusion-weighted imaging (DWI) was performed with TR = 4800 ms, TE = 89 ms, slice thickness = 3.5 mm, matrix = 272 × 320, and b-values of 800 and 1500 s/mm^2. Apparent diffusion coefficient (ADC) maps were generated for further lesion characterization.

CNN Model Development and Radiomics Analysis

Patients were randomly divided into a training set (n = 55) and a validation set (n = 25) using a 7:3 ratio. A ResNet-34 CNN served as the backbone for deep feature extraction. ResNet-34, pretrained on ImageNet, was used as a 2D backbone to extract slice-level deep features, which were then integrated into a custom 3D CNN for volumetric classification. Five CNN variants (CNN-1 to CNN-5) were constructed by varying the network depth to evaluate performance differences. Regions of interest (ROIs) were manually annotated on MRI slices using ITK-SNAP software and processed with AK software for standardization and texture feature extraction. ROI labeling was independently performed by two radiologists with 3 and 5 years of prostate MRI experience, who were blinded to the histopathological results. Any discrepancies between the two radiologists were resolved by consensus.

Prior to feature extraction and CNN training, all MRI images were preprocessed to ensure input consistency. Each scan was resampled to an isotropic voxel size of 1 × 1 × 1 mm^3 and intensity normalized to zero mean and unit variance. Rigid registration was applied to align sequences to the T2-weighted reference space using ANTs software, followed by cropping to a fixed field of view centered on the prostate. These preprocessing steps minimized variability caused by acquisition differences and ensured reproducibility of the model input.

Both hand-crafted radiomic features and CNN-derived deep features were obtained. Feature selection was performed using the minimum redundancy maximum relevance (mRMR) and least absolute shrinkage and selection operator (LASSO) algorithms to construct optimized radiomic signatures. These features were incorporated into a logistic regression model to calculate a radiomic score (Rad-score). Model robustness was evaluated using 100-fold leave-p-out cross-validation.

A transfer learning approach was used to train the CNN, whose outputs were integrated with radiomic features to build a hybrid diagnostic model. The input to the 3D CNN consisted of five consecutive 47 × 47-pixel image slices centered on the lesion. The network included one initial convolutional layer (32 feature maps), followed by three sequential max-pooling layers, which progressively reduced the feature map size ($47 \times 47 \rightarrow 23 \times 23 \rightarrow 11 \times 11$). Dropout layers (rate = 0.5) were applied to prevent overfitting. To further enhance model robustness, data augmentation was performed on the training dataset using random rotation (±10°), horizontal flipping, and intensity normalization. Hyperparameters, including learning rate, batch size, and dropout rate, were optimized through five-fold cross-validation within the training set. The model was trained using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) for up to 100 epochs with early stopping (patience = 10, Δval-loss $<1 \times 10^{-4}$) and learning-rate reduction on plateau (factor = 0.1). The final configuration (learning rate = 1×10^{-4}, batch size = 16, dropout rate = 0.5) ensured stable convergence and minimized validation loss. Five CNN variants (CNN-1 to CNN-5) were constructed with increasing network depth. CNN-3 contained three convolutional-pooling blocks (32, 64, 128 filters) and achieved the best performance, while CNN-1/2 used fewer layers, and CNN-4/5 added an extra convolutional and fully connected layer. Finally, extracted features were flattened and passed through two fully connected layers (128 neurons each), and classification was performed using a Softmax output layer.

Evaluation Metrics

The study employed a multifaceted evaluation framework to comprehensively assess model performance and clinical relevance. Baseline demographic and clinical variables—including age, body mass index (BMI), comorbidities (hypertension, diabetes mellitus, and hyperlipidemia), signal distribution, lesion margin characteristics, lesion volume, short diameter, long diameter, and the product of diameters—were compared between the benign and malignant groups using independent-samples *t*-tests and chi-square tests, as appropriate. Subsequently, multivariate logistic regression analysis was conducted to identify independent risk factors associated with malignant prostate nodules and to construct a predictive model. Additionally, the training and validation sets were statistically evaluated to ensure comparability across key clinical and imaging parameters.

Diagnostic efficacy was assessed by comparing conventional MRI interpretations with five CNN-assisted diagnostic models (CNN-1 to CNN-5). Metrics included sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV). Receiver operating characteristic (ROC) curves were generated, and the area under the curve (AUC) was calculated for each method to evaluate discrimination ability.

Statistical Analysis

All statistical analyses were conducted using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). Data normality was tested using the Kolmogorov–Smirnov test. Continuous variables following a normal distribution were presented as mean ± standard deviation (SD) and compared using independent-samples *t*-tests. Categorical variables were assessed using Chi-square (χ^2) or Fisher's exact tests. A two-sided p-value <0.05 was considered statistically significant. To construct the malignancy prediction model, variables with $p < 0.05$ in the univariate analysis were entered into a multivariate logistic regression using the maximum likelihood estimation method. The regression coefficients (β) obtained from this model were used to generate the logistic regression formula presented in the Results section.

Results

Comparison of Baseline Characteristics Between the Malignant and Benign Groups

A total of 80 patients with histologically confirmed prostatic nodules were enrolled in this study, including 39 with malignant and 41 with benign lesions. The ages of participants ranged from 52 to 89 years, with a mean ± SD of 70.13 ± 7.80 years. The average BMI was 24.61 ± 1.71 kg/m^2. Among all patients, 77.5% (62/80) had hypertension, 56.3% (45/80) had diabetes mellitus, and 47.5% (38/80) had hyperlipidemia. No statistically significant differences were observed between the malignant and benign groups with respect to age ($p = 0.229$), BMI ($p = 0.375$), hypertension ($p = 0.904$), diabetes mellitus ($p = 0.382$), or hyperlipidemia ($p = 0.814$).

However, significant differences were identified in multiple imaging and lesion-related parameters. As shown in Table 1, compared with the benign group, the malignant cohort showed a markedly higher incidence of irregular lesion margins ($p < 0.001$) and heterogeneous signal distribution ($p < 0.001$). Quantitative measurements also revealed substantial differences, with malignant nodules exhibiting greater lesion volume (1.57 ± 0.35 cm^3 *vs* 0.42 ± 0.11 cm^3; $p < 0.001$), short-axis diameter (11.97 ± 1.22 mm *vs* 8.85 ± 1.46 mm; $p < 0.001$), long-axis diameter (19.00 ± 1.54 mm *vs* 16.10 ± 1.32 mm; $p < 0.001$), and product of diameters (0.55 ± 0.08 *vs* 0.42 ± 0.10; $p < 0.001$).

Table 1. **Comparison of baseline characteristics between the malignant and benign groups.**

Variable	Malignant group (n = 39)	Benign group (n = 41)	χ^2/t	p
Age (years)	71.21 ± 6.79	69.10 ± 8.61	1.212	0.229
BMI (kg/m^2)	24.78 ± 1.23	24.44 ± 2.07	0.893	0.375
Hypertension, n (%)	30 (76.92)	32 (78.05)	0.015	0.904
Diabetes mellitus, n (%)	20 (51.28)	25 (60.98)	0.763	0.382
Hyperlipidemia, n (%)	18 (46.15)	20 (48.78)	0.055	0.814
Lesion margin (smooth/ irregular)	5/34	25/16	19.776	<0.001
Signal distribution (homogeneous/heterogeneous)	6/33	32/9	31.473	<0.001
Lesion volume (cm^3)	1.57 ± 0.35	0.42 ± 0.11	19.421	<0.001
Lesion short diameter (mm)	11.97 ± 1.22	8.85 ± 1.46	10.337	<0.001
Lesion long diameter (mm)	19.00 ± 1.54	16.10 ± 1.32	9.071	<0.001
Product of short and long diameters (mm^2)	0.55 ± 0.08	0.42 ± 0.10	6.548	<0.001

BMI, body mass index.

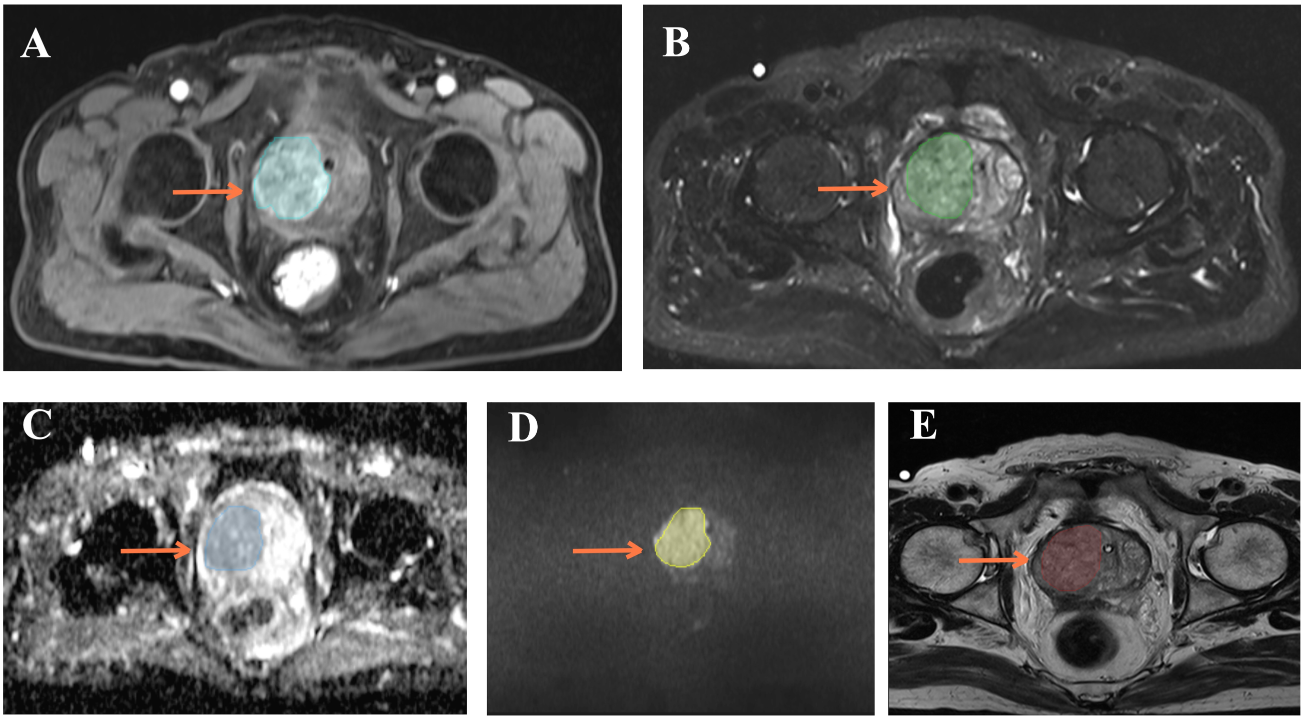

Fig. 1. Multi-parametric magnetic resonance imaging (MRI) images of a prostate lesion confirmed as malignant. (A) T1-weighted contrast-enhanced image (T1+C) showing moderate lesion enhancement with surrounding gland hyperplasia. (B) T2-weighted image with fat suppression (T2WIFS) indicating heterogeneous low signal intensity, highlighting lesion boundaries. (C) Apparent diffusion coefficient (ADC) map showing markedly low signal intensity in the lesion region. (D) Diffusion-weighted imaging (DWI, b = 1500 s/mm^2) reveals high signal intensity within the lesion. (E) T2-weighted image (T2WI) demonstrating an irregular hypointense lesion. Arrows indicate the lesion areas.

These imaging findings are illustrated in Fig. 1, which demonstrates typical malignant features, including heterogeneous low signal on T2WI, restricted diffusion on ADC maps, and hyperintensity on high b-value DWI.

Multivariate Logistic Regression Analysis of Predictors for Malignant Prostate Nodules

As shown in Table 2, lesion margin, signal distribution, lesion volume, short diameter, long diameter, and the product of diameters differed significantly ($p < 0.05$) between the malignant and benign subgroups. These parameters were subsequently included in a multivariate logistic regression analysis (Table 3) to identify independent predictors of malignancy.

In the training cohort, six features were independently associated with malignancy: irregular lesion margin (odds ratio (OR) = 2.818), heterogeneous signal distribution (OR = 3.121), lesion volume (OR = 3.056), short diameter (OR = 3.074), long diameter (OR = 2.886), and the product of diameters (OR = 2.912), all with $p < 0.001$. The validation

Table 2. Comparison of general characteristics between the training and validation sets.

Variable	Training set				Validation set			
Group	Malignant (n = 27)	Benign (n = 28)	χ^2/t	p	Malignant (n = 12)	Benign (n = 13)	χ^2/t	p
Age (years)	71.78 ± 6.03	68.57 ± 9.07	1.549	0.128	69.92 ± 8.40	70.23 ± 7.75	0.097	0.923
BMI (kg/m^2)	24.75 ± 1.22	24.41 ± 1.55	0.900	0.372	24.85 ± 1.29	24.50 ± 2.98	0.386	0.704
Hypertension, n (%)	21 (77.78)	21 (77.78)	0.059	0.808	9 (75.00)	11 (84.62)	0.361	0.548
Diabetes mellitus, n (%)	14 (51.85)	16 (57.14)	0.155	0.694	6 (50.00)	9 (69.23)	0.962	0.327
Hyperlipidemia, n (%)	12 (44.44)	14 (50.00)	0.170	0.680	6 (50.00)	6 (46.15)	0.037	0.848
Lesion margin (smooth/irregular)	4/23	18/10	14.017	<0.001	1/11	7/6	5.940	0.015
Signal distribution (homogeneous/ heterogeneous)	5/22	24/4	24.900	<0.001	1/11	8/5	7.667	0.006
Lesion volume (cm^3)	1.54 ± 0.33	0.39 ± 0.11	17.166	<0.001	1.64 ± 0.41	0.48 ± 0.08	9.617	<0.001
Lesion short diameter (mm)	12.04 ± 1.26	8.82 ± 1.44	8.809	<0.001	11.83 ± 1.19	8.92 ± 1.55	5.221	<0.001
Lesion long diameter (mm)	18.96 ± 1.56	16.07 ± 1.33	7.414	<0.001	19.08 ± 1.56	16.15 ± 1.34	5.034	<0.001
Product of diameters (mm^2)	0.53 ± 0.09	0.39 ± 0.08	6.164	<0.001	0.60 ± 0.03	0.48 ± 0.11	3.653	0.001

Table 3. Multivariate logistic regression analysis of malignant prostate nodules.

Dataset	Variable	β	Standard error	χ^2	p	OR	95% CI
Training set	Lesion margin	1.036	0.178	33.875	<0.001	2.818	1.988
	Signal distribution	1.138	0.139	67.028	<0.001	3.121	2.376
	Lesion volume (cm^3)	1.117	0.130	73.828	<0.001	3.056	2.368
	Lesion short diameter (mm)	1.123	0.108	108.121	<0.001	3.074	2.488
	Lesion long diameter (mm)	1.060	0.127	69.663	<0.001	2.886	2.250
	Product of diameters (mm^2)	1.069	0.155	47.565	<0.001	2.912	2.149
Validation set	Lesion margin	1.187	0.113	110.343	<0.001	3.277	2.626
	Signal distribution	1.156	0.158	53.531	<0.001	3.177	2.331
	Lesion volume (cm^3)	1.059	0.130	66.360	<0.001	2.883	2.235
	Lesion short diameter (mm)	1.052	0.149	49.849	<0.001	2.863	2.138
	Lesion long diameter (mm)	1.047	0.186	31.686	<0.001	2.849	1.979
	Product of diameters (mm^2)	1.125	0.168	44.842	<0.001	3.080	2.216

OR, odds ratio.

cohort yielded consistent findings, with corresponding ORs ranging from 2.849 to 3.277 and all p-values also below 0.001, confirming the robustness of these predictors. Based on these predictors, a malignancy risk prediction model was developed using the following logistic regression formula: Risk score = 12.369 + 1.036 × (Lesion margin) + 1.174 × (Signal distribution) + 1.216 × (Lesion volume) + 1.160 × (Short diameter) + 1.019 × (Long diameter) + 1.128 × (Product of short and long diameters). Complete regression results are detailed in Table 3, and variable encoding and feature definitions are summarized in Table 4.

Diagnostic Performance Comparison Between Conventional MRI and CNN-Assisted Models

To evaluate the performance of deep learning models in PCa classification, conventional MRI interpretations were compared with five CNN-based models (CNN-1 to CNN-5). Among them, CNN-3—featuring three max-pooling layers—achieved the best overall performance. Its architectural design likely enhanced its ability to extract high-level semantic features while suppressing irrelevant noise.

As shown in Table 5, CNN-3 yielded the highest diagnostic metrics in the training set, with a sensitivity of 85.19%, a specificity of 96.43%, and an overall accuracy of 90.91%. It also demonstrated the best positive and negative predictive values (PPV = 95.83%, NPV = 87.10%). In the validation set, although CNN-3's sensitivity decreased to 50.00%, it maintained favorable specificity (76.92%) and accuracy (69.57%), remaining the most effective model among the CNN variants. These results underscore the potential of CNN-based models, particularly CNN-3, for enhancing diagnostic precision in PCa MRI interpretation.

Table 4. Assignment table for logistic regression variables.

Variable	Value definition
Lesion margin	0 = Smooth, 1 = Irregular
Signal distribution	0 = Homogeneous, 1 = Heterogeneous
Lesion volume	Continuous value (cm^3)
Lesion short diameter	Continuous value (mm)
Lesion long diameter	Continuous value (mm)
Product of diameters	Continuous value (mm^2)

Table 5. Comparison of diagnostic performance between CNN-assisted and conventional MRI interpretation.

Dataset	Method	TP	FP	TN	FN	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Training set	Conventional MRI	11	11	17	16	50.91	40.74	60.71	50.00	51.52
	CNN-1	13	12	16	14	52.73	48.15	57.14	52.00	53.33
	CNN-2	10	2	26	17	65.45	37.04	92.86	83.33	60.47
	CNN-3	23	1	27	4	90.91	85.19	96.43	95.83	87.10
	CNN-4	12	1	27	15	70.91	44.44	96.43	92.31	64.29
	CNN-5	13	5	23	14	65.45	48.15	82.14	72.22	62.16
Validation set	Conventional MRI	8	2	11	4	76.00	66.67	84.62	66.67	84.62
	CNN-1	7	6	7	5	60.87	58.33	53.85	53.85	58.33
	CNN-2	8	6	7	4	65.22	66.67	53.85	57.14	63.64
	CNN-3	6	3	10	6	69.57	50.00	76.92	66.67	62.50
	CNN-4	5	3	10	7	65.22	41.67	76.92	62.50	58.82
	CNN-5	5	4	9	7	60.87	41.67	69.23	55.56	56.25

Abbreviations: TP, true positive; FP, false positive; TN, true negative; FN, false negative; PPV, positive predictive value; NPV, negative predictive value; CNN, convolutional neural network.

Table 6. ROC curve analysis of conventional MRI and CNN-based models.

Dataset	Method	Standard error	AUC	95% CI	*p*
Training set	Conventional MRI	20.573	0.721	0.155–0.880	<0.001
	CNN-1	15.483	0.778	0.216–0.876	<0.001
	CNN-2	13.505	0.759	0.239–0.900	<0.001
	CNN-3	17.03	0.864	0.192–0.850	<0.001
	CNN-4	19.732	0.708	0.213–0.918	<0.001
	CNN-5	17.171	0.775	0.174–0.951	<0.001
Validation set	Conventional MRI	19.294	0.785	0.264–0.901	<0.001
	CNN-1	19.303	0.655	0.285–0.897	<0.001
	CNN-2	18.098	0.745	0.167–0.988	<0.001
	CNN-3	17.387	0.838	0.243–0.930	<0.001
	CNN-4	18.234	0.656	0.255–0.929	<0.001
	CNN-5	16.135	0.719	0.220–0.986	<0.001

ROC Curve Analysis of Model Performance

ROC curve analysis was performed to evaluate the overall diagnostic discrimination ability of conventional MRI and CNN-based models. In both the training and validation datasets, the CNN-3 model exhibited the highest AUC, indicating superior classification performance compared to the other models. As illustrated in Table 6 and Fig. 2, the CNN-3 model consistently demonstrated the strongest discriminatory power in both the training and validation cohorts. In the training set, CNN-3 achieved the highest AUC (0.864), outperforming conventional MRI (0.721) and all other CNN variants, including CNN-1 (0.778) and CNN-5 (0.775). In the validation set, CNN-3 maintained its superior performance with an AUC of 0.838, exceeding that of conventional MRI (0.785) and the next-best model, CNN-2 (0.745). These findings indicate that CNN-3 exhibits strong generalization ability and robust classification performance across datasets. The ROC curves for the training and validation sets are shown in Fig. 2A,B, respectively.

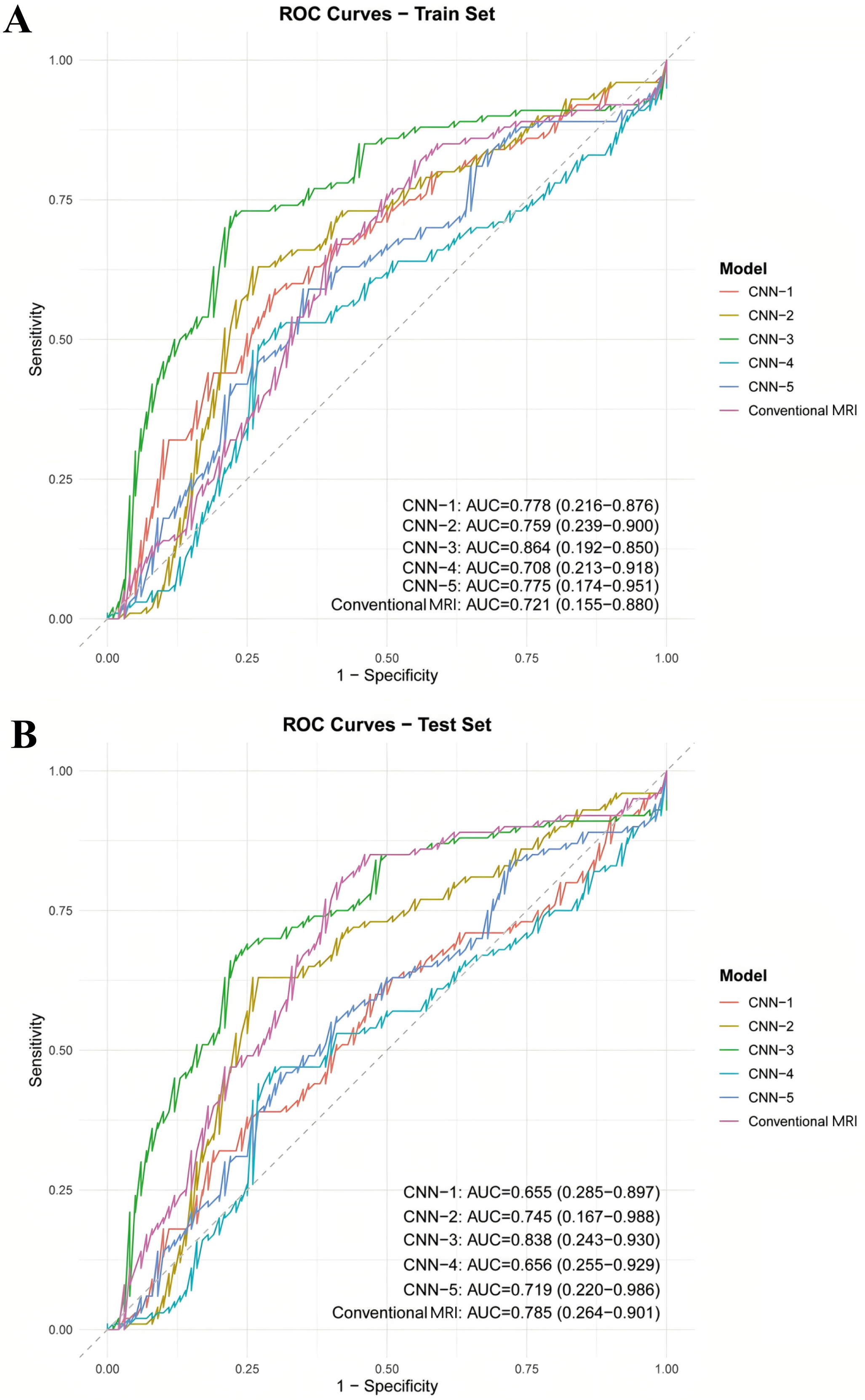

Fig. 2. ROC curve analysis of model performance. (A) ROC curves for the training set comparing conventional MRI and five CNN-assisted diagnostic models. CNN-3 achieved the highest AUC (0.864). (B) ROC curves for the validation set showing consistent superiority of CNN-3 with an AUC of 0.838. ROC, receiver operating characteristic; AUC, area under the curve.

Discussion

PCa remains one of the most prevalent malignancies among men, with its incidence steadily increasing alongside global population aging [1]. MRI, as a non-invasive modality, plays a crucial role in the diagnosis, staging, and treatment planning of PCa. The advent of mpMRI has further enhanced diagnostic performance by providing both functional and anatomical insights through T2WI, DWI, DCE-MRI, and MRS [14,15]. In particular, T2WI facilitates anatomical localization, while DWI and DCE-MRI contribute to biological characterization based on cellularity and vascularity, respectively [16]. Despite these advancements, mpMRI interpretation remains subject to variability due to heterogeneous lesion presentation and anatomical complexity [17,18]. This necessitates robust computer-aided systems to improve consistency and diagnostic precision.

In this study, we demonstrated that a CNN-based deep learning model—specifically, CNN-3—outperformed both conventional MRI interpretation and other CNN architectures. CNNs emulate neural information processing through layered convolution and pooling operations, enabling automated extraction of multi-level imaging features [19,20]. Traditional interpretation depends heavily on the radiologist's expertise. In contrast, CNNs reduce observer variability by learning discriminative patterns directly from imaging data [21]. Prior studies have confirmed that CNNs outperform handcrafted radiomics models by autonomously identifying features such as lesion irregularity, heterogeneity, and morphology, which are critical for distinguishing malignancy [22,23]. Our findings reinforce this result: CNN-3 showed the highest sensitivity and specificity across datasets, and key imaging predictors—lesion margin, signal heterogeneity, volume, and diameters—were consistently selected in both training and validation cohorts. These features align with known risk markers for PCa, and their integration into deep models enhances diagnostic robustness. The superior performance of CNN-3 may be attributed to its triple max-pooling structure, which facilitates deeper abstraction and reduces overfitting by compressing irrelevant spatial noise [24,25]. This architectural advantage likely contributed to improved generalizability, as evidenced by consistent AUC performance across cohorts. Importantly, early-stage PCa often presents as small, low-contrast lesions—especially in the peripheral zone—making detection with conventional MRI challenging [26,27]. CNNs trained on large-scale annotated datasets can capture subtle imaging cues, such as textural and perfusion-related variations, improving early detection sensitivity [28–30]. This is critical for timely intervention and a better prognosis. Although CNN-3 achieved the highest overall accuracy, its sensitivity in the validation set decreased compared with the training set, which may reflect limited sample size and distributional heterogeneity between datasets. The observed difference in AUCs suggests mild overfitting, which could be mitigated by regularization and external multicenter validation.

Despite the promising results, several limitations should be acknowledged. The sample size was relatively small and derived from a single institution, which may constrain generalizability. Future work should validate the CNN model in larger, multicenter cohorts and across diverse imaging protocols. The current study also lacked external validation, which restricts the assessment of the model's generalization on unseen datasets. Data sourced from a single institution may introduce center-specific biases related to imaging protocols or scanner settings. Therefore, future multicenter studies with external validation cohorts are essential to confirm the robustness and clinical applicability of the proposed CNN model. This limitation may reduce the reliability of model predictions when applied to images acquired from different scanners, institutions, or patient populations. Additionally, integrating clinical, biochemical, and genomic data could further enhance diagnostic performance and enable personalized risk assessment. Prospective studies exploring workflow integration and radiologist-AI collaboration are also warranted to assess real-world feasibility.

CNN-based computer-aided diagnostic systems offer substantial advantages in prostate MRI interpretation by providing automated, objective, and accurate lesion classification. Our CNN-3 model demonstrates strong potential to support clinical decision-making and reduce diagnostic uncertainty in PCa management.

Conclusion

In summary, this study demonstrates that a CNN-based deep learning model, particularly CNN-3, significantly outperforms conventional MRI in classifying prostate nodules on multiparametric MRI. By effectively leveraging key imaging features and deep feature abstraction, the model improves diagnostic accuracy and reliability. These findings highlight the potential of integrating AI-driven tools into prostate cancer diagnosis and provide a solid foundation for future clinical implementation.

Availability of Data and Materials

The data that support the findings of this study are not publicly available due to privacy reasons, but are available from the corresponding author upon request.

Author Contributions

TCG: Conceptualization, Methodology, Data curation, Investigation, Writing – original draft. XMZ: Data curation, Formal analysis, Visualization, Validation. CYW: Software, Resources, Investigation. CYC: Data curation, Resources, Supervision. QWX: Formal analysis, Visualization, Writing – review & editing. RHY: Investigation, Resources, Project administration. YZL: Supervision, Funding acquisition, Conceptualization, Writing – review & editing, Project administration. All authors contributed to the critical revision of the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of The Affiliated Dongguan Songshan Lake Central Hospital (approval: Dong Song Medical Research Ethics Review 2025 No. 30), and written informed consent was obtained from all participants prior to data collection. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Acknowledgment

Not applicable.

Funding

The Scientific Research Funding Project of the Nanfang Medical Imaging Alliance (FS2023010023).

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. https://doi.org/10.3322/caac.21660.

[2] Graham LS, Lin JK, Lage DE, Kessler ER, Parikh RB, Morgans AK. Management of Prostate Cancer in Older Adults. American Society of Clinical Oncology Educational Book. American Society of Clinical Oncology. Annual Meeting. 2023; 43: e390396. https://doi.org/10.1200/EDBK_390396.

[3] Tossetta G, Fantone S, Marzioni D, Mazzucchelli R. Cellular Modulators of the NRF2/KEAP1 Signaling Pathway in Prostate Cancer. Frontiers in Bioscience (Landmark Edition). 2023; 28: 143. https://doi.org/10.31083/j.fbl2807143.

[4] Yoo S, Kim JK, Jeong IG. Multiparametric magnetic resonance imaging for prostate cancer: A review and update for urologists. Korean Journal of Urology. 2015; 56: 487–497. https://doi.org/10.4111/kju.2015.56.7.487.

[5] Stabile A, Giganti F, Rosenkrantz AB, Taneja SS, Villeirs G, Gill IS, *et al*. Multiparametric MRI for prostate cancer diagnosis: current status and future directions. Nature Reviews. Urology. 2020; 17: 41–61. https://doi.org/10.1038/s41585-019-0212-4.

[6] Mir-Bashiri S, Yaqubi K, Woźnicki P, Westhoff N, von Hardenberg J, Huber T, *et al*. Multiparametric prostate MRI and structured reporting: benefits and challenges in the PI-RADS era. Chinese Journal of Academic Radiology. 2021; 4: 21–40.

[7] Wu Y, Tian J, Ma F, Wang C. Can dynamic contrast-enhanced MR imaging based on radiomics improve the diagnostic efficiency of clinically significant prostate cancer? British Journal of Hospital Medicine. 202; 85: 1–13. https://doi.org/10.12968/hmed.2024.0131.

[8] Xing P, Chen L, Yang Q, Song T, Ma C, Grimm R, *et al*. Differentiating prostate cancer from benign prostatic hyperplasia using whole-lesion histogram and texture analysis of diffusion- and T2-weighted imaging. Cancer Imaging: the Official Publication of the International Cancer Imaging Society. 2021; 21: 54. https://doi.org/10.1186/s40644-021-00423-5.

[9] Fuschi A, Suraci PP, Pastore AL, Al Salhi Y, Capodiferro P, Scalzo S, *et al*. Multiparametric Prostate MRI Accuracy of Prostate Imaging Reporting and Data System (v2.1) Scores 4 and 5: The Influence of Image Quality According to the Prostate Imaging Quality Score. Journal of Clinical Medicine. 2024; 13: 3785. https://doi.org/10.3390/jcm13133785.

[10] Chatterjee A, Pineda F, Karczmar GS, Oto A. Dynamic Contrast-Enhanced Imaging. Prostate MRI Essentials: A Practical Guide for Radiologists (pp. 75–87). Springer International Publishing: Cham. 2020.

[11] Yilmaz EC, Harmon SA, Law YM, Huang EP, Belue MJ, Lin Y, *et al*. External Validation of a Previously Developed Deep Learning-based Prostate Lesion Detection Algorithm on Paired External and In-House Biparametric MRI Scans. Radiology. Imaging Cancer. 2024; 6: e240050. https://doi.org/10.1148/rycan.240050.

[12] Li H, Liu H, von Busch H, Grimm R, Huisman H, Tong A, *et al*. Deep Learning-based Unsupervised Domain Adaptation via a Unified Model for Prostate Lesion Detection Using Multisite Biparametric MRI Datasets. Radiology. Artificial Intelligence. 2024; 6: e230521. https://doi.org/10.1148/ryai.230521.

[13] Twilt JJ, van Leeuwen KG, Huisman HJ, Fütterer JJ, de Rooij M. Artificial Intelligence-Based Algorithms for Prostate Cancer Classification and Detection on Magnetic Resonance Imaging: A Narrative Review. Diagnostics (Basel, Switzerland). 2021; 11: 959. https://doi.org/10.3390/diagnostics11060959.

[14] Feng X, Chen X, Peng P, Zhou H, Hong Y, Zhu C, *et al*. Values of multiparametric and biparametric MRI in diagnosing clinically significant prostate cancer: a multivariate analysis. BMC Urology. 2024; 24: 40. https://doi.org/10.1186/s12894-024-01411-0.

[15] Giganti F, Moreira da Silva N, Yeung M, Davies L, Frary A, Ferrer Rodriguez M, *et al*. AI-powered prostate cancer detection: a multi-centre, multi-scanner validation study. European Radiology. 2025; 35: 4915–4924. https://doi.org/10.1007/s00330-024-11323-0.

[16] Reynolds HM, Tadimalla S, Wang YF, Montazerolghaem M, Sun Y, Williams S, *et al*. Semi-quantitative and quantitative dynamic contrast-enhanced (DCE) MRI parameters as prostate cancer imaging biomarkers for biologically targeted radiation therapy. Cancer Imaging: the Official Publication of the International Cancer Imaging Society. 2022; 22: 71. https://doi.org/10.1186/s40644-022-00508-9.

[17] Simon BD, Merriman KM, Harmon SA, Tetreault J, Yilmaz EC, Blake Z, *et al*. Automated Detection and Grading of Extraprostatic Extension of Prostate Cancer at MRI via Cascaded Deep Learning and Random Forest Classification. Academic Radiology. 2024; 31: 4096–4106. https://doi.org/10.1016/j.acra.2024.04.011.

[18] Walker SM, Choyke PL, Turkbey B. What You Need to Know Before Reading Multiparametric MRI for Prostate Cancer. AJR. American Journal of Roentgenology. 2020; 214: 1211–1219. https://doi.org/10.2214/AJR.19.22751.

[19] Lin A, Su B, Ning Y, Zhang L, He Y. Convolutional Neural Networks in Medical Imaging: A Review. In International Con-

ference on Swarm Intelligence. Singapore: Springer Nature Singapore. 2024.

[20] Reddy SK, Kathirvelu K. Hybrid Optimization Enabled Deep-Learning for Prostate Cancer Detection. Sensing and Imaging. 2024; 25: 52.

[21] Alqarni M, Jones E, Mullassery V, Morris S, Verma H, Cooper S, *et al*. 1986: External validation of a deep learning prostate MR auto-contouring model. Radiotherapy and Oncology. 2024; 194: S3081–S3084.

[22] Castillo T JM, Arif M, Starmans MPA, Niessen WJ, Bangma CH, Schoots IG, *et al*. Classification of Clinically Significant Prostate Cancer on Multi-Parametric MRI: A Validation Study Comparing Deep Learning and Radiomics. Cancers. 2021; 14: 12. https://doi.org/10.3390/cancers14010012.

[23] Zhang Y, Li W, Zhang Z, Xue Y, Liu YL, Nie K, *et al*. Differential diagnosis of prostate cancer and benign prostatic hyperplasia based on DCE-MRI using bi-directional CLSTM deep learning and radiomics. Medical & Biological Engineering & Computing. 2023; 61: 757–771. https://doi.org/10.1007/s11517-022-02759-x.

[24] Zafar A, Aamir M, Mohd Nawi N, Arshad A, Riaz S, Alruban A, *et al*. A comparison of pooling methods for convolutional neural networks. Applied Sciences. 2022; 12: 8643.

[25] Zhao L, Zhang Z. A improved pooling method for convolutional neural networks. Scientific Reports. 2024; 14: 1589. https://doi.org/10.1038/s41598-024-51258-6.

[26] Jurka M, Macova I, Wagnerova M, Capoun O, Jakubicek R, Ourednicek P, *et al*. Deep-learning-based reconstruction of T2-weighted magnetic resonance imaging of the prostate accelerated by compressed sensing provides improved image quality at half the acquisition time. Quantitative Imaging in Medicine and Surgery. 2024; 14: 3534–3543. https://doi.org/10.21037/qims-23-1488.

[27] Wong PF, McNeil C, Wang Y, Paparian J, Santori C, Gutierrez M, *et al*. Clinical-Grade Validation of an Autofluorescence Virtual Staining System With Human Experts and a Deep Learning System for Prostate Cancer. Modern Pathology: an Official Journal of the United States and Canadian Academy of Pathology, Inc. 2024; 37: 100573. https://doi.org/10.1016/j.modpat.2024.100573.

[28] Laudicella R, Comelli A, Schwyzer M, Stefano A, Konukoglu E, Messerli M, *et al*. PSMA-positive prostatic volume prediction with deep learning based on T2-weighted MRI. La Radiologia Medica. 2024; 129: 901–911. https://doi.org/10.1007/s11547-024-01820-z.

[29] Schrader A, Netzer N, Hielscher T, Görtz M, Zhang KS, Schütz V, *et al*. Prostate cancer risk assessment and avoidance of prostate biopsies using fully automatic deep learning in prostate MRI: comparison to PI-RADS and integration with clinical data in nomograms. European Radiology. 2024; 34: 7909–7920. https://doi.org/10.1007/s00330-024-10818-0.

[30] 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. https://doi.org/10.1038/s41391-025-00997-2. (online ahead of print)