Original Article | Open Access

Artificial Intelligence Surgery

Lin *et al. Art Int Surg.* 2026;6:268-79 **DOI:**10.20517/ais.2025.105

Application of artificial intelligence incorporating PI-RADS v2.1 to prevent unnecessary prostate biopsies

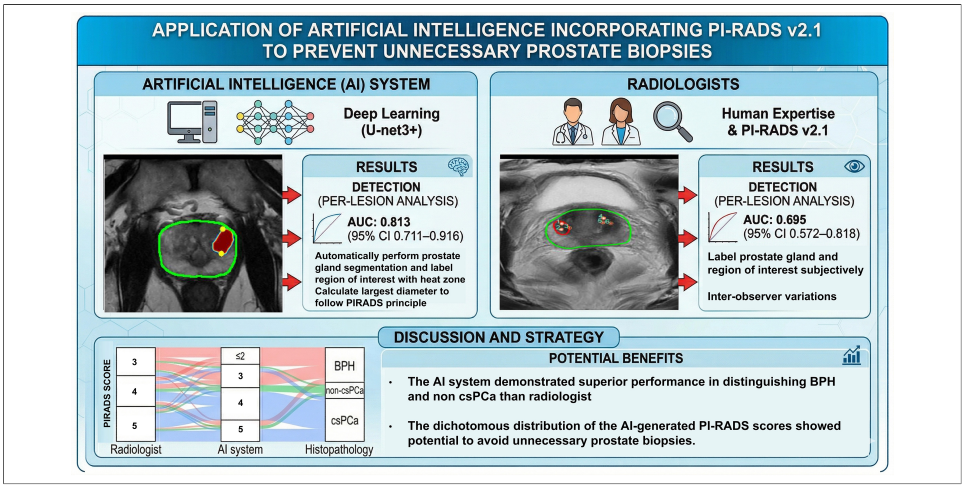
Kun-Che Lin[1], Shu-Wei Liu[2], Yung-Ming Kuo[3], Gia-Shing Shieh[4], Yuh-Shyan Tsai[1], Chien-Hui Ou[1,5], Kuan-Yu Wu[1], Chun-Hung Yang[6,†], Che-Yuan Hu[1,†]

Keywords:
Artificial intelligence, prostate cancer, multiparametric MRI, PI-RADS v2.1, risk stratification, receiver operating characteristic curve analysis

Citation: Lin KC, Liu SW, Kuo YM, Shieh GS, Tsai YS, Ou CH, Wu KY, Yang CH, Hu CY. Application of artificial intelligence incorporating PI-RADS v2.1 to prevent unnecessary prostate biopsies. *Art Int Surg.* 2026;6:268-79. https://dx.doi.org/10.20517/ais.2025.105

Received: 22 Oct 2025 **First Decision:** 25 Mar 2026 **Revised:** 24 Apr 2026 **Accepted:** 11 May 2026 **Published:** 25 May 2026

Academic Editors: Roland Croner, Eyad Elyan **Copy Editor:** Xing-Yue Zhang **Production Editor:** Xing-Yue Zhang



Abstract

Aim: Artificial intelligence (AI) systems have the potential to enhance prostate magnetic resonance imaging (MRI) interpretation by providing objective image analysis, improving lesion detection, and reducing overdiagnosis. This study aimed to develop and evaluate an AI system for analyzing prostate multiparametric MRI (mpMRI) based on Prostate Imaging Reporting and Data System version 2.1 (PI-RADS v2.1) criteria.

Methods: In this retrospective, single-center study, we developed an AI system using data from 204 patients in the open-source PROSTATEx Challenge and 30 patients from National Cheng Kung University Hospital (NCKUH). The AI algorithm was retrospectively applied to mpMRI scans of 70 patients, and AI-derived PI-RADS scores were compared to those assigned by radiologists. Histopathological results from MRI-targeted biopsies served as the reference standard. The primary endpoints included the area under the

[1]Department of Urology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704302, Taiwan.
[2]Department of Electronic Engineering, Southern Taiwan University of Science and Technology, Tainan 710301, Taiwan.
[3]Department of Electronic Engineering, National Formosa University, Yunlin County 632301, Taiwan.
[4]Department of Urology, Tainan Hospital, Ministry of Health and Welfare, Tainan 700007, Taiwan.
[5]Tainan Hospital, Ministry of Health and Welfare, Tainan 700007, Taiwan.
[6]Department of Electrical Engineering, National Formosa University, Yunlin County 632301, Taiwan.
[†]Co-corresponding author.

Correspondence to: Dr. Che-Yuan Hu, Department of Urology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704302, Taiwan. E-mail: greatoldhu@gmail.com; Prof. Chun-Hung Yang, Department of Electrical

Engineering, National Formosa University, Yunlin County 632301, Taiwan. E-mail: eli.yang@nfu.edu.tw

receiver operating characteristic curve (AUROC) for the AI system versus radiologists, and the prostate gland segmentation metrics.

Results: The AI system achieved an average F1-score of 0.896 for prostate gland segmentation, demonstrating robust performance. In the 70 NCKUH cases, the AI system outperformed radiologists in differentiating benign prostatic hyperplasia (BPH) or non-clinically significant prostate cancer (non-csPC), with an AUROC of 0.813 [95% confidence interval (CI) 0.711-0.916; $P < 0.001$], compared to 0.695 (95%CI 0.572-0.818; $P = 0.005$) for radiologists. The AI system exhibited a more dichotomous distribution of PI-RADS scores, reducing diagnostic ambiguity in PI-RADS 3 lesions.

Conclusion: The AI system demonstrated improved performance in distinguishing BPH and non-csPC compared with radiologists. The dichotomous distribution of the AI-generated PI-RADS scores showed potential to avoid unnecessary biopsies.

INTRODUCTION

Prostate cancer is one of the most prevalent cancers among men worldwide. In 2022, 1.4 million new cases of prostate cancer were diagnosed, making it the fourth most frequently diagnosed cancer globally. In the United States, approximately one in eight men will be diagnosed with prostate cancer during their lifetime[1]. Despite its high incidence, prostate cancer is associated with relatively low mortality[2]. Therefore, distinguishing clinically significant prostate cancer (csPC), which affects patient life expectancy, from non-clinically significant prostate cancer (non-csPC), and avoiding unnecessary treatment of the latter, has become an essential task in current clinical practice[3]. Prostate biopsy remains the gold standard for diagnosis. Conventional transrectal and transperineal prostate biopsies have a diagnostic limitation, with positive detection rates of only about 30% due to their random sampling approach. In this method, the prostate is divided into fixed quadrants for systematic but unguided biopsy, making it challenging to target suspected tumor lesions directly[4,5].

In addition to these examinations, the prostate health index (PHI), multiparametric magnetic resonance imaging (mpMRI), and prostate-specific membrane antigen positron emission tomography (PSMA PET) may be applied for prostate cancer diagnosis[6-8]. Besides transrectal/transperineal ultrasound prostate biopsy, mpMRI-guided prostate biopsy using cognitive guidance, ultrasound integrated with MRI fusion software, or direct in-bore guidance provides alternative approaches to improve diagnostic accuracy[9].

The Prostate Imaging-Reporting and Data System (PI-RADS) score is used to evaluate a patient's prostate gland and assess the likelihood of csPC (defined as International Society of Urological Pathology, ISUP, Grade group ≥ 2). PI-RADS score rates lesions from 1 to 5[10,11]. A recent systematic review reported that the detection rates of csPC were 6%, 12%, 48%, and 72% for PI-RADS scores 2, 3, 4, and 5, respectively[12]. Omitting biopsy in patients with PI-RADS score 2 or lower has been shown to reduce unnecessary biopsy by approximately 30%[13]. However, csPC may still be present in lesions with low PI-RADS scores, and conversely, benign pathology can occasionally be found in PI-RADS 5 lesions[14]. In addition, inter-observer variability among radiologists can lead to discrepancies in PI-RADS scoring[15]. To distinguish patients with low PI-RADS, an objective image review system is necessary for physicians in clinical practice.

In 2012, Krizhevsky et al. introduced AlexNet, a convolutional neural network (CNN) that demonstrated the potential of computer vision and achieved breakthrough performance in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)[16]. Rapid development of deep learning models followed, including VGG

(visual geometry group) and ResNet (residual neural network)[17,18]. Machine learning approaches using support vector machines on radiomic features have shown improved performance in enhancing the accuracy of PI-RADS scoring for interpreting csPC[19]. Artificial intelligence (AI) is on the verge of a revolution in prostate cancer, with clinical applications expanding across pathological diagnosis, Gleason grading, prognostic evaluation, and the determination of treatment options[20]. Detection of peripheral zone prostate cancer through radiomic classifiers on T2-Weighted MRI had achieved a cross-validated area under the curve (AUC) of 0.744 with boosted Decision Tree (DT). However, there is still limited evidence on prostate cancer lesion detection by AI using the PI-RADS v2.1 (Prostate Imaging Reporting and Data System version 2.1) principle.

In this study, we establish a prostate MRI analysis model using U-net, U-net++ and U-net3+ to perform prostate gland segmentation and prostate cancer lesion detection according to PI-RADS v2.1 principles. The pathology result of prostate gland biopsy is used as ground truth for our model training. Our goal is to establish an objective image-based diagnostic support system to aid physicians in clinical practice.

METHODS

The study protocol was approved by the Institutional Review Board of National Cheng Kung University Hospital (NCKUH) (IRB protocol number A-ER-112-005). For this retrospective analysis of clinical data and MRI scans, the requirement for informed consent was waived by the ethics committee. All data were anonymized and de-identified to protect patient privacy before development of the AI analysis system.

Patients and MRI acquisition

Open-source ProstateX dataset

The SPIE-AAPM-NCI PROSTATEx Challenges (hereafter referred to as PROSTATEx) open-source dataset was selected to supplement this study (https://www.cancerimagingarchive.net/collection/prostatex/). The PROSTATEx dataset contains 346 cases, of which 204 include ground truth annotations. It was made publicly available at the 2017 SPIE Medical Imaging conference as part of a challenge. The dataset's purpose was to promote the development and advancement of automated detection and diagnosis methods for prostate cancer. It provides various types of prostate MRI images, including T2-weighted images (T2WI), diffusion-weighted images (DWI), and dynamic contrast-enhanced (DCE) images, all from real clinical settings and annotated by professional physicians, including the location of the prostate and potential cancerous regions. Therefore, the PROSTATEx dataset is an ideal resource for developing and validating automated algorithms for prostate and lesion segmentation.

National Cheng-Kung University Hospital (NCKUH) training dataset

From January 1, 2022, to December 25, 2023, 40 patients with suspected prostate cancer who underwent MRI screening were enrolled at NCKUH. All examinations were performed using a 3-T MRI scanner (Ingenia 3.0T, Philips®) with T2WI, DWI with b-0, b-1000, and b-2000 s/mm^2, and apparent diffusion coefficient (ADC).

Prediction tool development

We used U-net, U-net++ and U-net3+ as the deep learning models, which are widely applied in medical image segmentation[21]. The major core concepts of the model employ an encoder-decoder architecture, where the encoder progressively down-samples the input image to capture contextual information, and the decoder up-samples the features to generate a precise output with high spatial resolution. These three models were implemented in Python 3.8 (https://www.python.org) and TensorFlow version 2.6.

According to the principles of PI-RADS v2.1, the peripheral zone (PZ) and transition zone should be segmented separately. The prostate and lesion segmentation models were trained independently. The prostate segmentation and PZ models were trained using T2WI. The non-PZ region was defined as the portion of the segmented prostate area after subtraction of the PZ region. The lesion segmentation model comprised three distinct input channels: T2-weighted MRI, ADC, and DWI with b-2000. We used the open-source PROSTATEx dataset and NCKUH dataset to train the models, and then evaluated their performance during the validation and testing phases. Of the 40 NCKUH cases, 10 were reserved as an independent test set (i.e., data not used in training) prior to any model training. These cases were entirely withheld from all training and cross-validation procedures to serve as unbiased external samples for the final model assessment. The remaining 30 NCKUH cases were incorporated into the training pool alongside the 204 PROSTATEx cases.

Consequently, three training dataset configurations were constructed to systematically evaluate segmentation performance across different data compositions: (1) NCKUH, comprising the 30 NCKUH training cases only; (2) PROSTATEx, comprising the 204 PROSTATEx cases only; and (3) MIX, comprising all 234 training cases from both sources combined. For each configuration in the validation phase, 5-fold cross-validation was applied to the respective training pool, with data randomly divided into five equal subsets. In each fold iteration, four subsets were used for model optimization and one was reserved for internal validation. For the model test phase, the finalized model was subsequently evaluated on the 10 held-out NCKUH patient cases. These cases remained entirely unseen by the model at any stage of training or validation, thereby constituting a truly independent test. Supplementary Table 1 shows the list of parameters explored for each model, as well as the final parameter combination chosen for the analyses.

Validation of model

Prostate segmentation was evaluated using the F1-score, which indicates the accuracy of the segmentation relative to the ground truth labels. The F1-score was defined as follows:

Precision = True positive/(True positive + False positive); Recall = True positive/(True positive + False negative); and F1-score = 2 × Precision × Recall/(Precision + Recall).

We retrospectively reviewed MRI-fusion biopsy performed in NCKUH between January 2022 and November 2024. In NCKUH, mpMRI of the prostate was performed using a 3.0-Tesla MRI scanner (Ingenia 3.0T, Philips Healthcare®, Best, The Netherlands) equipped with a phased-array surface coil. The imaging protocol included axial, sagittal, and coronal T2-weighted (T2W) sequences, DWI with corresponding ADC maps, and DCE imaging following intravenous administration of a gadolinium-based contrast agent. DWI was acquired using multiple b-values (0, 1,000, and 2,000 s/mm^2), and DCE images were obtained using a fast 3D T1-weighted spoiled gradient-echo sequence with a temporal resolution of approximately 3-5 s per phase after contrast injection. The total examination time was approximately 40 min. Approximately 70 mpMRI examinations from individual cases were processed by the AI system. Prostate gland segmentation, automated PI-RADS scores, and regions of interest (ROIs) generated by the AI system were collected and compared with radiologists' annotations. Pathological findings from targeted biopsies, particularly cases diagnosed as benign prostatic hyperplasia (BPH) or non-csPC, were used to verify whether these corresponded to cases classified by the AI system as PI-RADS ≤ 2. Therefore, the validation of ROI marked by AI system will be feasible with the ground truth of the pathological result. In this way, we could determine the successfulness of prediction by an AI system.

Statistical analysis

Statistical analyses were performed using SPSS version 22 (IBM Corp., Armonk, NY, USA) and R version 4.5.3. Continuous variables are presented as mean ± standard deviation. Receiver operating characteristic (ROC) curves were constructed to evaluate diagnostic performance, and area under the receiver operating characteristic curves (AUROCs) were compared using DeLong's test. A P-value < 0.05 was considered statistically significant.

RESULTS

For the AI system, 204 cases were selected from the PROSTATEx challenge and 30 cases were selected from the NCKUH dataset. K-fold cross-validation with 5-fold revealed the best performance that had been achieved under epoch 500 cycles of training. Three different models (U-net, U-net++ and U-net3+) were compared, and U-net3+ achieved the highest average F1-score of 0.896 across training with NCKUH alone, PROSTATEx, and MIX dataset (NCKUH + PROSTATEx datasets) [Table 1]. For prostate gland segmentation, we validated the AI system using three datasets - NCKUH, PROSTATEx, and MIX. The NCKUH dataset comprised imaging data only, acquired at NCKU hospital (30 cases). The PROSTATEx dataset included 204 cases from the PROSTATEx challenge. The MIX dataset consisted of 234 cases from both sources combined. The three network architectures (U-net, U-net++ and U-Net3+) were applied to 5-fold cross-validation. The model trained prostate gland segmentation for 500 times on each validation, generating best and final scores for the performance index - accuracy, precision and recall. Accuracy was defined as the overall rate of correct predictions. Precision was defined as the proportion of correctly predicted positive cases among all positive predictions. Recall was defined as the proportion of correctly predicted positive cases among all actual positives. Combining these parameters, the F1-score was calculated, which balances precision and recall. U-net3+ demonstrated the best average performance, 0.896, of F1-score among the mixed dataset.

The PZ of the prostate gland was segmented separately from whole prostate gland to enable analysis according to PI-RADS principles. Multiple parametric series of prostate MRI were analyzed in the same "cut" of the anatomical level. Suspicious prostate cancer lesions were visualized with heat maps, and the greatest diameter of lesion was automatically calculated. The AI system then determined whether each ROI was located in the PZ or non-PZ. Then, the algorithm interpreted the MRI findings according to the principles of PI-RADS v2.1. For example, Figure 1 shows a PI-RADS 5 case with estimated lesion measuring approximately 1.5 cm in the transition zone. The major ROI is presented in T2 phase, consistent with the principle of PI-RADS v2.1, while a milder ROI signal is also noted on DWI phase.

At our institution, we retrospectively reviewed 70 patients from the NCKUH validation dataset who underwent MRI-echo fusion biopsy. The mean prostate-specific antigen (PSA) level of patients was 14.73 ng/mL, the mean prostate volume was 50.32 mL and the mean body mass index (BMI) was 25.55. The numbers of patients in different PI-RADS categories were as follows: 26 patients with PI-RADS 3, 20 with PI-RADS 4, and 24 with PI-RADS score 5. Pathological results of magnetic resonance (MR) fusion biopsy revealed 31 patients with csPC (Gleason grade score ≥ 7) and 7 patients with ISUP grade 1 (Gleason grade score = 6), which was classified as non-csPC. The remaining 32 patients had BPH on biopsy. Basic patient characteristics of the PROSTATEx dataset, the NCKUH training dataset, and the NCKUH validation dataset are listed in Table 2. Notably, because specific clinical parameters such as BMI and PI-RADS scores were not provided in the official PROSTATEx release, these fields are indicated as 'N/A' (not available).

PI-RADS scores assigned by the radiologists and by the AI system, together with corresponding biopsy results, are listed in Table 3. For lesions labeled by radiologists, csPC detection rates for PI-RADS 5, 4, and 3 lesions were 79.16%, 45%, and 11.54%, respectively. For AI system-derived scores, csPC detection rates for

Table 1. F1 score of prostate gland segmentation under U-net, U-net++ and U-net3+Epoch 500

5-fold cross-validation for prostate segmentation									
5-fold cross validation	**Dataset**	**U-net**							
		Accuracy		**Precision**		**Recall**		**F1-score**	
		Best	**Final**	**Best**	**Final**	**Best**	**Final**	**Best**	**Final**
	NCKUH	0.979	0.979	0.904	0.860	0.742	0.784	0.815	0.820
1	PROSTATEx	0.989	0.989	0.948	0.945	0.876	0.879	0.910	0.911
	MIX	0.987	0.987	0.890	0.900	0.854	0.839	0.871	0.869
	NCKUH	0.979	0.980	0.901	0.886	0.743	0.776	0.814	0.827
2	PROSTATEx	0.984	0.984	0.970	0.968	0.841	0.845	0.901	0.902
	MIX	0.987	0.987	0.967	0.964	0.844	0.850	0.901	0.903
	NCKUH	0.992	0.991	0.913	0.921	0.849	0.825	0.880	0.870
3	PROSTATEx	0.990	0.990	0.937	0.931	0.889	0.898	0.912	0.915
	MIX	0.987	0.988	0.952	0.942	0.876	0.891	0.912	0.916
	NCKUH	0.988	0.987	0.961	0.964	0.811	0.796	0.880	0.872
4	PROSTATEx	0.988	0.988	0.866	0.878	0.940	0.930	0.901	0.903
	MIX	0.989	0.989	0.892	0.889	0.913	0.917	0.902	0.903
	NCKUH	0.993	0.993	0.934	0.936	0.861	0.859	0.896	0.896
5	PROSTATEx	0.987	0.987	0.875	0.865	0.896	0.908	0.885	0.886
	MIX	0.987	0.986	0.858	0.848	0.925	0.928	0.890	0.886
	NCKUH	0.986	0.986	0.923	0.913	0.801	0.808	0.857	0.857
Average	PROSTATEx	0.988	0.988	0.919	0.917	0.888	0.892	0.902	0.903
	MIX	0.987	0.987	0.912	0.909	0.882	0.885	0.895	0.895
5-fold cross-validation for prostate segmentation									
U-net ++									
Accuracy		**Precision**		**Recall**		**F1-score**			
Best	**Final**	**Best**	**Final**	**Best**	**Final**	**Best**		**Final**	
0.981	0.980	0.866	0.893	0.813	0.772	0.838		0.828	
0.988	0.988	0.950	0.942	0.865	0.876	0.906		0.908	
0.986	0.986	0.902	0.879	0.830	0.846	0.864		0.862	
0.985	0.984	0.947	0.945	0.798	0.797	0.866		0.865	
0.984	0.984	0.965	0.964	0.848	0.847	0.903		0.902	
0.987	0.987	0.961	0.961	0.852	0.852	0.904		0.903	
0.990	0.990	0.914	0.922	0.792	0.780	0.848		0.845	
0.989	0.989	0.925	0.924	0.892	0.893	0.908		0.908	
0.987	0.987	0.949	0.949	0.876	0.876	0.911		0.911	
0.987	0.987	0.949	0.953	0.818	0.803	0.878		0.872	
0.987	0.987	0.867	0.863	0.937	0.942	0.900		0.901	
0.988	0.988	0.877	0.881	0.923	0.918	0.899		0.899	
0.993	0.992	0.917	0.923	0.864	0.852	0.889		0.886	
0.987	0.987	0.863	0.867	0.916	0.911	0.888		0.888	
0.987	0.987	0.885	0.885	0.896	0.896	0.890		0.890	
0.987	0.987	0.919	0.927	0.817	0.801	0.864		0.859	
0.987	0.987	0.914	0.912	0.892	0.894	0.901		0.901	
0.987	0.987	0.915	0.911	0.875	0.878	0.894		0.893	
5-fold cross-validation for prostate segmentation									
U-net 3+									

Accuracy		Precision		Recall		F1-score	
Best	**Final**	**Best**	**Final**	**Best**	**Final**	**Best**	**Final**
0.981	0.981	0.868	0.855	0.813	0.820	0.840	0.837
0.989	0.989	0.948	0.951	0.873	0.871	0.909	0.909
0.986	0.986	0.867	0.861	0.861	0.870	0.864	0.865
0.985	0.984	0.885	0.886	0.867	0.863	0.876	0.874
0.984	0.983	0.962	0.964	0.851	0.836	0.903	0.895
0.987	0.987	0.954	0.959	0.865	0.859	0.907	0.906
0.991	0.992	0.863	0.901	0.884	0.869	0.873	0.885
0.990	0.990	0.930	0.928	0.900	0.902	0.915	0.915
0.987	0.987	0.948	0.942	0.884	0.891	0.915	0.916
0.990	0.990	0.957	0.958	0.866	0.858	0.909	0.905
0.988	0.988	0.866	0.866	0.945	0.945	0.904	0.904
0.988	0.988	0.870	0.868	0.930	0.933	0.899	0.899
0.993	0.993	0.918	0.933	0.882	0.869	0.900	0.900
0.987	0.987	0.868	0.880	0.897	0.888	0.882	0.884
0.987	0.987	0.867	0.858	0.927	0.932	0.896	0.893
0.988	0.988	0.898	0.907	0.862	0.856	0.880	0.880
0.988	0.987	0.915	0.918	0.893	0.888	0.903	0.901
0.987	0.987	0.901	0.898	0.893	0.897	0.896	0.896

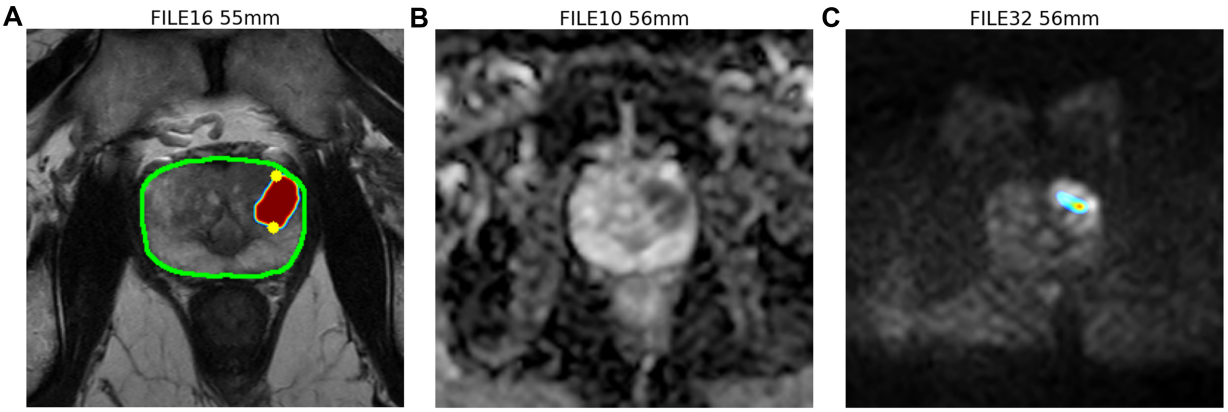


Figure 1. AI system analyzing prostate multiparametric MRI. (A) T2 weighted (T2W) phase of MRI showing whole-prostate segmentation outlined in green. The region of interest (ROI) is labeled with a heat map, with the highest possibility in red. The largest diameter of the ROI is automatically calculated and marked by two yellow dots, indicating an estimated lesion diameter of approximately 1.5 cm and a lesion proportion of 98.63% in the transition zone (TZ); (B) Apparent diffusion coefficient (ADC) phase of MRI; (C) Diffusion-weighted images (DWI) phase of MRI. The ROI is also labeled with heat map with highest possibility in red zone. The AI system will assemble every file from different phases, align each cut of prostate MRI image in order to evaluate PI-RADS and label the ROI. In this representative case, the AI system identified a 1.5 cm lesion in the TZ and assigned a PI-RADS score of 5 based on the lesion's greatest diameter. The dominant abnormality is seen on T2W imaging, while a mild corresponding abnormality is also noted on DWI, consistent with PI-RADS v2.1, in which TZ lesions are primarily assessed on T2W imaging. AI: Artificial intelligence; MRI: magnetic resonance imaging; PI-RADS: Prostate Imaging Reporting and Data System; DWI: diffusion-weighted images; PI-RADS v2.1: PI-RADS version 2.1.

PI-RADS 5, 4, and 3 were 57.14%, 55.88%, and 42.86%, respectively. Additionally, the AI system identified 16 cases with lesions classified as PI-RADS score ≤ 2. Among these cases, 15 patients had benign biopsy results and one had Gleason grade score of 6 (non-csPC). The mean analysis time for the AI system was 52.41 ± 17.64 s.

Table 2. Basic characteristics of PROSTATEx dataset, NCKUH training dataset and NCKUH validation dataset

	NCKUH validation dataset	PROSTATEx dataset	NCKUH training dataset
Patients number (*n*)	*n* = 70	*n* = 204	*n* = 30
Age	66.72 ± 8.02	63 ± 7	67 ± 8.2
BMI (kg/m^2)	25.55 ± 3.31	N/A	25.25 ± 2.61
PI-RADS			
3	26	N/A	13
4	20	N/A	9
5	24	N/A	8
PSA (ng/mL)	14.73 ± 18.73	14 ± 10	18.35 ± 27.95
Prostate volume (mL)	50.32 ± 23.20	50 ± 25	54.38 ± 26.86
PSAD	0.31 ± 0.32	0.16 ± 0.22	0.37 ± 0.40
Prostate MRI-echo fusion biopsy/Radical prostatectomy histopathology results (*n*)			
BPH	32	106	15
ISUP Grade group			
Group 1	7	29	3
Group 2	8	38	4
Group 3	13	18	4
Group 4	3	7	0
Group 5	7	6	4

NCKUH: National Cheng Kung University Hospital; PI-RADS: Prostate Imaging Reporting and Data System; MRI: magnetic resonance imaging; BMI: Body mass index; PSA: prostate-specific antigen; PSAD: prostate-specific antigen density; BPH: benign prostatic hyperplasia; ISUP: International Society of Urological Pathology; N/A: not available.

Table 3. PI-RADS interpretation by radiologists and the AI system with final biopsy pathology results

		Final pathology result					Final pathology result		
		BPH	Non-csPC	csPC			BPH	Non-csPC	csPC
Radiologists PI-RADS score	≤ 2				AI system PI-RADS score	≤ 2	15	1	
	3	21	2	3		3	3	1	3
	4	8	3	9		4	10	4	19
	5	3	2	19		5	4	2	8
					AI system analysis time (sec): 52.41 ± 17.64				

PI-RADS: Prostate Imaging Reporting and Data System; AI: artificial intelligence; BPH: Benign prostatic hyperplasia; Non-csPC: non-clinically significant prostate cancer; csPC: clinically significant prostate cancer.

For the task of distinguishing BPH and non-csPC using PI-RADS scores assigned by the AI system and by the radiologists, the AI system achieved an ROC curve of 0.813 [95% confidence interval (CI) 0.711-0.916; $P < 0.001$]. The ROC curve for the radiologists was 0.695 (95%CI 0.572-0.818; $P = 0.005$). With an 11.8% increase in the AUROC (95%CI, 1.064%-25.745%, $P = 0.033$), the AI system significantly outperformed the radiologists [Figure 2].

DISCUSSION

Deep learning-based AI systems have demonstrated non-inferior performance to radiologists for prostate segmentation and prostate cancer lesion identification in large, international multicenter studies[22]. AI assistance has also been shown to improve accuracy in the radiologic diagnosis of csPC[23]. However, many challenges remain in diagnosing prostate cancer using mpMRI, and interobserver variability in prostate MRI

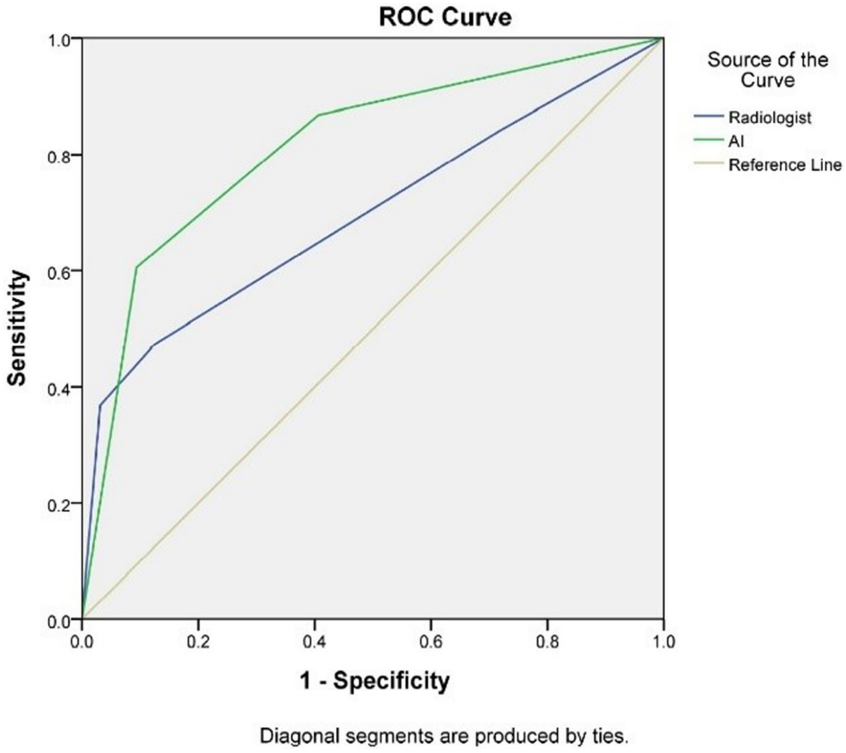


Figure 2. Receiver operating characteristic (ROC) curves for detecting benign prostatic hyperplasia (BPH) or non-clinically significant prostate cancer (non-csPC) by the radiologists and the AI system. The area under the ROC curve (AUROC) by the AI system is 0.813, and by the radiologists is 0.695. The sensitivity and specificity by the AI system are 86.8% and 59.4%, respectively, while those by the radiologists are 84.2% and 28.1%, respectively. AI: Artificial intelligence.

interpretation persists despite the principles of PI-RADS v2.1[24]. Therefore, an AI system that is widely accessible to urologists, easy to use, and capable of providing accurate PI-RADS scores is essential.

To our knowledge, this is the first AI system that detects lesions on multiple parametric MRI and provides a PI-RADS score as the interpretation output. Our system detects lesions on T2WI, DWI, and ADC, providing lesion coordinates in the axial plane according to the latest PI-RADS v2.1 guidelines. Ground truth for ROIs was established using the final pathology results from the biopsy.

The risk of csPC for PI-RADS scores 3-5 was 11%, 37%, and 70%, respectively, according to the American Urological Association guidelines[11]. We demonstrated a compatible diagnostic rate to PI-RADS scores interpreted by radiologists at our institution. PI-RADS score 3 lesions remain equivocal for biopsy; therefore, reducing the number of reporting PI-RADS score 3 is beneficial for clinical decision-making. In previous AI algorithms, the negative predictive value (NPV) has been higher than the positive predictive value[22]. Therefore, our study focuses on screening out BPH or non-csPC imaging findings on MRI to avoid unnecessary biopsies. In our AI system, PI-RADS interpretations showed a dichotomous distribution, with redistribution toward higher categories (such as PI-RADS 4 and 5) and lower categories (PI-RADS ≤ 2) [Supplementary Figure 1]. Accordingly, the proportion of PI-RADS 3 lesions was reduced, which may help decrease diagnostic ambiguity in equivocal cases. The ROC curve for detecting BPH or non-csPC demonstrated superior performance for the AI system compared with radiologists. The AI system can more effectively reclassify patients with PI-RADS 3 lesions, either upgrading them to PI-RADS 4 or 5 or downgrading them to PI-RADS 1 or 2, thereby helping to resolve the clinical dilemma of whether a biopsy is necessary.

According to PI-RADS v2.1, PI-RADS 5 is defined as lesions larger than 1.5 cm in greatest diameter. Our system defines the largest diameter as the longest distance between two points on the boundary of the lesion, which provides an accurate definition of PI-RADS 5 lesion [Figure 1]. Under these strict criteria, fewer PI-RADS 5 lesions were identified by the AI system compared to radiologists' reports, indicating heterogeneity between subjective and objective analysis.

AI and machine learning models can support physicians and patients in shared decision-making, including risk stratification, optimization of patient outcomes, and early warning of acute decompensation[25]. In a nationwide effort, high-risk patients were identified preoperatively[26]. Fusion biopsy focuses on more accurate targeting of ROIs while using a smaller number of biopsy cores. Therefore, integrating fusion biopsy with AI-based analysis may help avoid unnecessary biopsies, particularly in patients with high surgical risks but low malignancy potential.

Our study has several limitations. First, the sample size of both the training and validation cohorts was relatively small, which may limit the robustness of the model and the persuasiveness of the results. Second, this was a single-center study for institutional model development and validation, which may restrict the external validity of the findings. Third, the prevalence of prostate cancer was high across all datasets, including 48.0% in the PROSTATEx cohort, 50.0% in the NCKUH training cohort, and 54.3% in the NCKUH validation cohort. As these cohorts were derived from patients undergoing prostate MRI and MRI-targeted biopsy, they likely represent a selected higher-risk population with a more complex case mix rather than a general screening population. This may have introduced selection bias, limited the generalizability of our findings, and potentially overestimated the diagnostic performance of the AI system. In addition, the mpMRI protocol used in this study was limited to 3-T MRI with high b-value DWI ($b > 1{,}500$ s/mm^2), which may affect applicability to institutions using different imaging protocols. Therefore, further larger-scale, multicenter, and prospective studies are needed to validate the real-world clinical utility and external generalizability of this AI system.

In conclusion, the AI system demonstrated a dichotomous distribution of PI-RADS v2.1 results and outperformed radiologists in detecting BPH and non-csPC. The AI system has been shown to be a potential supportive tool for clinical decision-making and for avoiding unnecessary biopsies. Further prospective clinical trials of this system are essential.

DECLARATIONS

Authors' contributions

Concept and design of the study: Lin KC, Ou CH, Yang CH, Hu CY
Data acquisition: Lin KC, Liu SW, Shieh GS, Tsai YS
Data analysis: Lin KC, Liu SW, Kuo YM, Yang CH, Hu CY
Statistical analysis: Lin KC, Liu SW, Wu KW, Yang CH, Hu CY
Manuscript preparation: Lin KC, Yang CH, Hu CY
Manuscript editing: Yang CH, Hu CY
Manuscript review: Yang CH, Hu CY

Availability of data and materials

The open source PROSTATEx dataset was obtained from the PROSTATEx challenge and is available at https://www.cancerimagingarchive.net/collection/prostatex/ with the permission of the PROSTATEx challenge. The National Cheng Kung University Hospital (NCKUH) datasets analyzed during the current study are not publicly available due to ethical restrictions and patient confidentiality protocols mandated by the NCKUH Institutional Review Board (IRB protocol number: A-ER-112-005). However, data may be made available from the corresponding author upon reasonable request and with appropriate institutional data sharing agreements.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT, powered by GPT-5.4 Thinking (released 2026-03-05), was used solely for language editing. Graphical abstract was generated by Google Gemini 3 Flash in collaboration with Nano Banana 2. The graphic abstract is carefully reviewed and adjusted. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

Hu CY was supported by the Ministry of Science and Technology, Taiwan (NSTC 114-2314-B-006-038).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

This retrospective Institutional Review Board (IRB)-approved study was performed at a single center: NCKUH. Data collection, analysis and publication were approved under IRB protocol number A-ER-112-005. The requirement for informed consent was waived by the IRB of NCKUH due to the retrospective nature of the study and the use of de-identified imaging data.

Consent for publication

Not applicable.

Copyright

© The Authors 2026.

Supplementary Materials

Supplementary Materials

REFERENCES

1. American Cancer Society. Cancer Statistics Center. Available from http://cancerstatisticscenter.cancer.org [accessed 22 May 2026].
2. Schafer EJ, Laversanne M, Sung H, et al. Recent patterns and trends in global prostate cancer incidence and mortality: an update. *Eur Urol.* 2025;87:302-13. DOI PubMed PMC
3. Spratt DE, Srinivas S, Adra N, et al. Prostate cancer, Version 3.2026, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* 2025;23:469-93. DOI PubMed
4. Gravestock P, Shaw M, Veeratterapillay R, Heer R. Prostate cancer diagnosis: biopsy approaches. In: Barber N, Editor. Urologic cancers. Exon Publications; 2022. pp. 141-68. DOI PubMed
5. Wei JT, Barocas D, Carlsson S, et al. Early detection of prostate cancer: AUA/SUO guideline part II: considerations for a prostate biopsy. *J Urol.* 2023;210:54-63. DOI PubMed PMC
6. Stabile A, Giganti F, Rosenkrantz AB, et al. Multiparametric MRI for prostate cancer diagnosis: current status and future directions. *Nat Rev Urol.* 2020;17:41-61. DOI PubMed
7. Shen Z, Li Z, Li Y, et al. PSMA PET/CT for prostate cancer diagnosis: current applications and future directions. *J Cancer Res Clin Oncol.* 2025;151:155. DOI PubMed PMC
8. Lee IT, Hou CM, Vo TTT, et al. Optimizing prostate cancer care: clinical utility of the prostate health index. *Prostate.* 2025;85:1357-68. DOI PubMed
9. EAU Guidelines. Edn. presented at the EAU Annual Congress Amsterdam 2022. ISBN 978-94-92671-16-5. 2022. Available from https://uroweb.org/news/new-eau-guidelines-are-now-available [accessed 22 May 2026].
10. Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur Urol.* 2019;76:340-51. DOI PubMed
11. Wei JT, Barocas D, Carlsson S, et al. Early detection of prostate cancer: AUA/SUO guideline Part I: prostate cancer screening. *J Urol.* 2023;210:46-53. DOI PubMed PMC

12. Barkovich EJ, Shankar PR, Westphalen AC. A systematic review of the existing prostate imaging reporting and data system version 2 (PI-RADSv2) literature and subset meta-analysis of PI-RADSv2 categories stratified by Gleason scores. *AJR Am J Roentgenol.* 2019;212:847-54. DOI PubMed
13. Haj-Mirzaian A, Burk KS, Lacson R, et al. Magnetic resonance imaging, clinical, and biopsy findings in suspected prostate cancer: a systematic review and meta-analysis. *JAMA Netw Open.* 2024;7:e244258. DOI PubMed PMC
14. Ahmed HU, El-Shater Bosaily A, Brown LC, et al.; PROMIS study group. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet.* 2017;389:815-22. DOI PubMed
15. Sonn GA, Fan RE, Ghanouni P, et al. Prostate magnetic resonance imaging interpretation varies substantially across radiologists. *Eur Urol Focus.* 2019;5:592-9. DOI PubMed
16. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges CJ, Bottou L, Weinberger K, Editors. Advances in neural information processing systems 25. NIPS 2012; 2012 Dec 3-6; Lake Tahoe, NV, USA. New York: Curran Associates, Inc.; 2012. Available from https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf [accessed 22 May 2026].
17. Tabrizchi H, Parvizpour S, Razmara J. An improved VGG model for skin cancer detection. *Neural Process Lett.* 2023;55:3715-32. DOI
18. Zhou M, Li M, Cao Q, et al. Malignant pleural mesothelioma classification and survival prediction with CT imaging using ResNet. *Eur Radiol.* 2026;36:2603-14. DOI PubMed
19. Jin P, Shen J, Yang L, et al. Machine learning-based radiomics model to predict benign and malignant PI-RADS v2.1 category 3 lesions: a retrospective multi-center study. *BMC Med Imaging.* 2023;23:47. DOI PubMed PMC
20. Zhu M, Sali R, Baba F, et al. Artificial intelligence in pathologic diagnosis, prognosis and prediction of prostate cancer. *Am J Clin Exp Urol.* 2024;12:200-15. DOI PubMed PMC
21. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, Editors. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. MICCAI 2015; 2015 Oct 5-9; Munich, Germany. Cham: Springer International Publishing; 2015. pp. 234-41. DOI
22. Saha A, Bosma JS, Twilt JJ, et al.; PI-CAI consortium. Artificial intelligence and radiologists in prostate cancer detection on MRI (PI-CAI): an international, paired, non-inferiority, confirmatory study. *Lancet Oncol.* 2024;25:879-87. DOI PubMed PMC
23. Twilt JJ, Saha A, Bosma JS, et al.; PI-CAI Consortium. AI-assisted *vs* unassisted identification of prostate cancer in magnetic resonance images. *JAMA Netw Open.* 2025;8:e2515672. DOI PubMed PMC
24. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists. *Radiology.* 2016;280:793-804. DOI PubMed PMC
25. Giordano C, Brennan M, Mohamed B, Rashidi P, Modave F, Tighe P. Accessing artificial intelligence for clinical decision-making. *Front Digit Health.* 2021;3:645232. DOI PubMed PMC
26. Hyer JM, Ejaz A, Tsilimigras DI, Paredes AZ, Mehta R, Pawlik TM. Novel machine learning approach to identify preoperative risk factors associated with super-utilization of medicare expenditure following surgery. *JAMA Surg.* 2019;154:1014-21. DOI PubMed PMC

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.


© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.