

REVIEW

Open Access

# What benefit can be obtained from magnetic resonance imaging diagnosis with artificial intelligence in prostate cancer compared with clinical assessments?

Li-Tao Zhao<sup>1,2†</sup>, Zhen-Yu Liu<sup>3,4†</sup>, Wan-Fang Xie<sup>1,2</sup>, Li-Zhi Shao<sup>3</sup>, Jian Lu<sup>5\*</sup>, Jie Tian<sup>1,3,6\*</sup> and Jian-Gang Liu<sup>1,6,7\*</sup>

## Abstract

The present study aimed to explore the potential of artificial intelligence (AI) methodology based on magnetic resonance (MR) images to aid in the management of prostate cancer (PCa). To this end, we reviewed and summarized the studies comparing the diagnostic and predictive performance for PCa between AI and common clinical assessment methods based on MR images and/or clinical characteristics, thereby investigating whether AI methods are generally superior to common clinical assessment methods for the diagnosis and prediction fields of PCa. First, we found that, in the included studies of the present study, AI methods were generally equal to or better than the clinical assessment methods for the risk assessment of PCa, such as risk stratification of prostate lesions and the prediction of therapeutic outcomes or PCa progression. In particular, for the diagnosis of clinically significant PCa, the AI methods achieved a higher summary receiver operator characteristic curve (SROC-AUC) than that of the clinical assessment methods (0.87 vs. 0.82). For the prediction of adverse pathology, the AI methods also achieved a higher SROC-AUC than that of the clinical assessment methods (0.86 vs. 0.75). Second, as revealed by the radiomics quality score (RQS), the studies included in the present study presented a relatively high total average RQS of 15.2 (11.0–20.0). Further, the scores of the individual RQS elements implied that the AI models in these studies were constructed with relatively perfect and standard radiomics processes, but the exact generalizability and clinical practicality of the AI models should be further validated using higher levels of evidence, such as prospective studies and open-testing datasets.

**Key words** Clinically significant prostate cancer, Adverse pathology, Radiomics quality score, Artificial intelligence, Magnetic resonance imaging

## Background

Prostate cancer (PCa) is one of the most prevalent cancers among men, especially in the United States, with the highest incidence and second highest mortality rate[1-3]. In China, PCa has three epidemiological characteristics. First, it ranks highest in the annual increase in both morbidity and mortality in men[4]. Second, the ratio of mortality to morbidity of PCa is higher than that in some Western countries[2-4]. Third, the proportion of patients with high-risk advanced PCa is high due to limited prostate specific antigen (PSA) screening[1,5]. Two challenges in PCa diagnosis and treatment are the precise diagnosis of PCa and the prediction of the therapeutic outcomes or PCa progression, which have attracted extensive interest from researchers[6-12].

Invasive biopsy is a common method used in clinical practice to monitor PCa[13-16]. However, the randomness of the needle position for biopsy sampling limits the ability of the biopsy to capture the spatial state of the lesions and therefore, leads to the omission of true tumors. Additionally, patients who undergo biopsies may have some reactions, such as bleeding, pain, infection, and even life-threatening sepsis in severe cases[16,17]. On the other hand, medical imaging can provide a comprehensive macroscopic description of the tumor phenotype and peritumoral context, which can be a compensatory and noninvasive approach to provide information by quantifying tumor progression before, during, and after treatment[18,19]. Therefore, characterization based on medical imaging is a practical method for quantifying the heterogeneity of PCa and potentially facilitating the development of precision medicine.

Magnetic resonance imaging (MRI) is a common medical imaging methodology with high spatial resolution and can also describe different physiological and anatomical characteristics

<sup>†</sup>Li-Tao Zhao and Zhen-Yu Liu contributed equally to this work

\*Correspondence: [lujian@bjmu.edu.cn](mailto:lujian@bjmu.edu.cn), [jie.tian@ia.ac.cn](mailto:jie.tian@ia.ac.cn), [jgliu@buaa.edu.cn](mailto:jgliu@buaa.edu.cn)

<sup>1</sup>School of Engineering Medicine, Beihang University, Beijing 100191, China

<sup>3</sup>Department of Urology, Peking University Third Hospital, Peking University, Beijing 100191, China

Full list of author information is available at the end of the article

based on various sequences. For example, T<sub>2</sub>-weighted imaging (T<sub>2</sub>WI) can describe the anatomical structures of tumors and is, therefore, useful for delineating the profiles and appearances of suspicious lesions. Diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) maps derived from DWI can reflect the degree of random motion of water molecules related to the tumor's aggressiveness. Additionally, the dynamic contrast-enhanced sequence can be used for the functional and physiological assessment of tumors with the guide of a gadolinium contrast agent. Further, compared with computed tomography (CT) and positron emission tomography (PET), MRI has no radiation risk and has thus been widely used for tumor diagnosis and monitoring[15,18,20-24].

In clinical diagnosis, magnetic resonance (MR) images of prostate lesions are visually assessed according to the prostate imaging reporting and data system (PI-RADS) guidelines based on some visually quantitative features of lesions (e.g., location, shape, size, and intensity)[25,26]. The visual assessment of MR images plays an important role in the diagnosis but has several limitations[6,27-30]. First, the visual assessment of MR images is greatly dependent on the high-level expertise of radiologists, leading to discrepancies in the assessment results. Second, some features of MR images reflecting tumor heterogeneities cannot be observed on the visual assessment. Third, the visual assessment is qualitative or semi-quantitative[26]. These limitations may lead to a decrease in the accuracy and robustness of PCa diagnoses.

Artificial intelligence (AI) is a data-driven method. In other words, after being trained with many samples, an AI model can automatically select the optimal feature pattern to accurately predict the novel samples. Thus, when AI methodology is used to analyze medical images of PCa, it can mine fine and deep information that may reflect relatively complete heterogeneities of the suspicious lesions, regardless of whether this information is visually representable. Due to its advantages in analyzing medical images, AI methodology has been widely applied to aid in the diagnosis and treatment of PCa[12,31-35] and other malignancies[13,36-46]. Increasing evidence supports the ability of AI methods to facilitate precise diagnosis and treatment of tumors. In fact, some AI software that can help identify PCa has been approved by the Food and Drug Administration (FDA). For example, ProstatID software aims to interpret prostate MRI and assist radiologists in identifying suspicious PCa regions and analysing their likelihood of malignancy[47]. AI-Rad Companion Prostate MR software aims to assist the radiologists in automatically segmenting prostate, estimating volume and manually delineating the location of lesions with MR images, which can

be used to support the planning of biopsies[48]. Additionally, one AI software, i.e., Paige Prostate software, is developed with pathological images instead of MRI. This software is designed to aid pathologists in detecting suspicious areas on prostate biopsy images and further assessing the likelihood of malignancies[49].

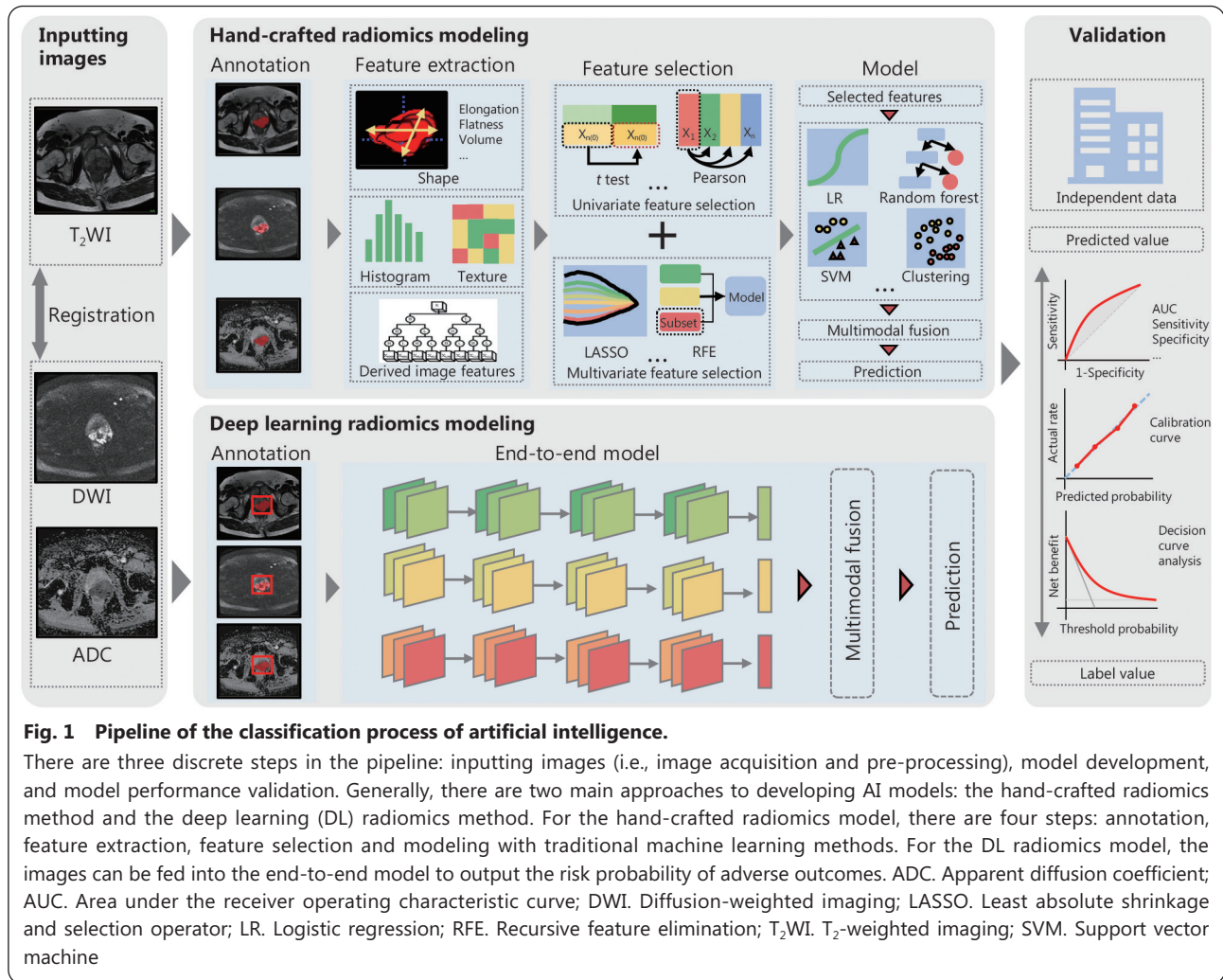
The National Comprehensive Cancer Network (NCCN) reported that MRI could generally guide PCa monitoring[15]. In clinical practice, patients with PI-RADS scores 3–5 are recommended to undergo biopsies for further pathological confirmation[17]. However, patients with a PI-RADS score 3 are equivocal in detecting clinically significant prostate cancer (csPCa). As a result, it may lead to low specificity and, therefore, overdiagnosis[6,28,30]. Additionally, according to the NCCN, some patients are recommended to undergo radical prostatectomy (RP) or other therapies. However, some of these patients may have high risks of the presence of adverse pathology (AP), disease recurrence, and subsequent metastasis[10,50-53]. Advance identification of these patients before treatment may be beneficial to their prognoses. To address these problems, many studies have constructed a variety of AI models for the diagnoses and treatments of PCa, such as the diagnosis of csPCa[54,55], prediction of Gleason grade[56], prediction of biochemical recurrence (BCR)[57], and extracapsular extension (ECE)[58]. They have compared the performances of these models with those of visual assessments based on PI-RADS or other clinical assessments. Currently, most of the published reviews focus on the analysis of the modeling processes and tasks of AI methods[59-61]. However, reviews comparing the performance between AI and clinical assessment methods are limited, though they can highlight the clinical value and potential of AI methodology to aid clinicians in precisely diagnosing PCa and predicting therapeutic outcomes or progression of PCa.

To bridge this gap, in the present study, we focused on studies that reported results from both AI and clinical assessment methods. Then, we analyzed and summarized these studies, comparing the diagnostic and predictive performance for PCa between AI and common clinical assessment methods based on MR images and/or clinical characteristics, thereby exploring the potential of AI in the diagnosis and treatment of PCa. Specifically, we compared the performance between AI and clinical assessment methods for the diagnosis and prediction fields of PCa. In particular, we quantitatively compared the abilities of these two methodologies to diagnose csPCa and predict AP. Additionally, the quality of the studies was assessed based on the radiomics quality score (RQS).

### AI pipeline on the diagnosis and prediction fields of PCa

This study focused on two fields of AI application to PCa: the diagnosis field, which refers to the identification of malignant lesions and stratification of PCa risk; and the prediction field, which refers to the prediction of therapeutic outcomes or progression of PCa.

The pipeline of the process of the AI methodology includes several discrete steps: image acquisition and preprocessing, model development, and model performance validation. Generally, there are two main approaches to developing AI models for medical imaging analysis: the hand-crafted radiomics method and the deep learning (DL) radiomics method (Fig. 1).



**Fig. 1 Pipeline of the classification process of artificial intelligence.**

There are three discrete steps in the pipeline: inputting images (i.e., image acquisition and pre-processing), model development, and model performance validation. Generally, there are two main approaches to developing AI models: the hand-crafted radiomics method and the deep learning (DL) radiomics method. For the hand-crafted radiomics model, there are four steps: annotation, feature extraction, feature selection and modeling with traditional machine learning methods. For the DL radiomics model, the images can be fed into the end-to-end model to output the risk probability of adverse outcomes. ADC. Apparent diffusion coefficient; AUC. Area under the receiver operating characteristic curve; DWI. Diffusion-weighted imaging; LASSO. Least absolute shrinkage and selection operator; LR. Logistic regression; RFE. Recursive feature elimination; T<sub>2</sub>WI. T<sub>2</sub>-weighted imaging; SVM. Support vector machine

The hand-crafted radiomics method can provide a set of high-throughput features. Specifically, prostate lesions are first manually delineated in MR images. Then the features including shape, histogram, and textural features are extracted from the delineated lesions in the original MR images and their derived images (e.g., wavelet). The shape and histogram features refer to the metrics characterizing the shape (e.g., size, volume and flatness) and histogram (e.g., mean, entropy and skewness), respectively. The textural features were extracted by the calculation matrixes reflecting the distribution of gray intensity, such as Gray Level Cooccurrence Matrix, Gray Level Size Zone Matrix, and Gray Level Run Length Matrix. These extracted

features are fed into traditional machine learning models such as logistic regression, support vector machine (SVM), and random forests (RF) after a feature selection step, which finally outputs a quantitative score with a value ranging from zero to one, indicating the risk probability of adverse outcomes such as cSPCa and BCR. Compared with the DL radiomics method, the hand-crafted radiomics method is simple owing to fewer parameters in traditional machine learning models and is, therefore, easier to achieve. Additionally, specific features have relatively evident semantic information, thereby increasing the interpretability of the models[39]. However, it requires precise manual delineation of the tumor slice by slice, which is time-

consuming, laborious, and can lead to subjective disagreements among different radiologists.

In contrast, the DL radiomics method can automatically extract the features of medical images without requiring the precisely manual delineation of lesions. Specifically, DL radiomics models have been constructed using various networks, such as ResNet[62], MobileNet[63], and ShuffleNet[64]. DL radiomics models use original images or rectangular volumes of interest containing lesions as inputs, from which image features can be extracted directly through convolution operation of networks. Therefore, precise manual segmentation of lesions can be avoided. Similar to the hand-crafted radiomics method, the DL radiomics models output a value indicating the risk probability of adverse outcomes. Thus, the DL radiomics method is particularly suitable for the analysis of a large number of samples. Additionally, owing to multi-layer construction, DL radiomics models can mine deep and subtle image information that can accurately characterize the heterogeneity of PCa, thereby showing excellent performance for predicting adverse outcomes. Both hand-crafted and DL radiomics methods can be used for the diagnosis and prediction fields of PCa. They can be employed alone or in combination *via* the fusion of features or integration of models. In this study, AI models constructed based on the hand-crafted and DL radiomics methods are referred to as hand-crafted (HC) and DL models, respectively.

### **Comparing the performances of AI and clinical assessment methods in the diagnosis field of PCa**

The application of AI in diagnosis fields of PCa based on MR images has attracted extensive interest. In clinical practice, diagnostic tasks for PCa mainly include the risk stratification of prostate lesions, such as PCa detection (i.e., the discrimination between benign and malignant lesions) and csPCa detection (i.e., the discrimination between non-csPCa and csPCa). According to the International Society of Urological Pathology Gleason grade group (GGG), patients with GGG <1 and GGG ≥1 were defined as having benign and malignant lesions, respectively. Patients with GGG <3 and GGG ≥3[18] (or GGG <2 and GGG ≥2[54]) were defined as having noncsPCa and csPCa, respectively. Table 1[31,54,65-84] listed the studies for the diagnosis field of PCa, which were included in the present study.

### **Comparing the performances of AI and clinical assessment methods for csPCa diagnosis**

In clinical practice, patients with a high risk of csPCa assessed using MRI are recommended to undergo biopsies for further pathological confirmation. Therefore, accurate identification

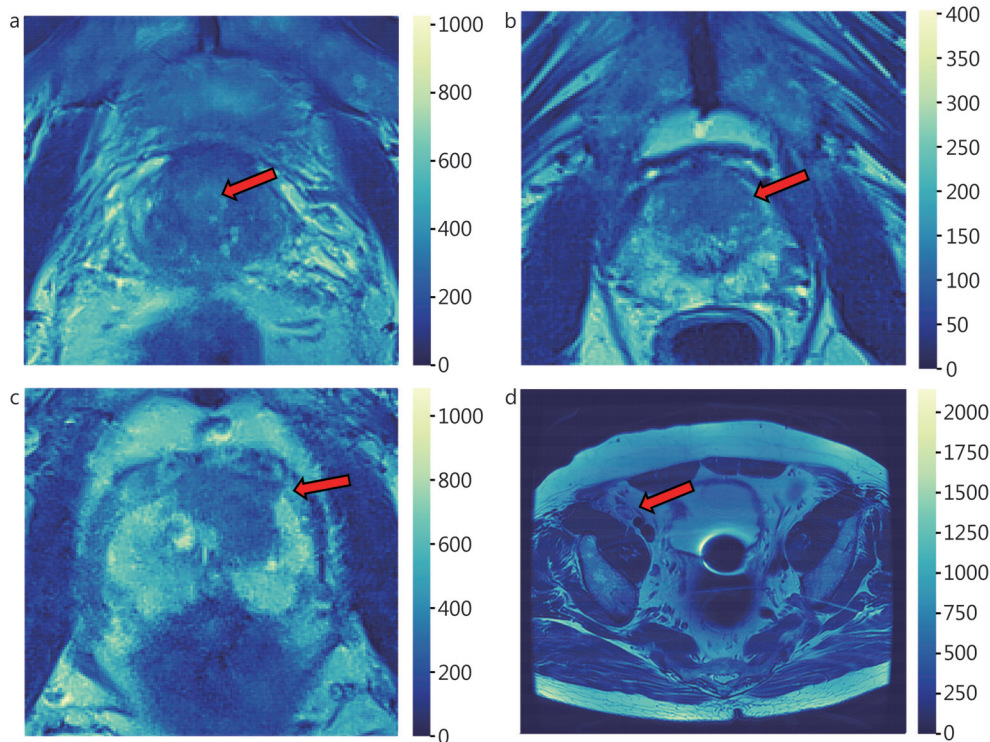
of csPCa candidates can help to reduce unnecessary biopsies. Figure 2a and b shows representative examples of patients with non-csPCa and csPCa, respectively. MRI interpretation following the PI-RADS guidelines has been used as a common clinical assessment method for csPCa diagnosis. Thus, most studies have compared the performance of AI methods with that of radiologists' interpretations of PI-RADS.

Some studies reported that AI models achieved performances similar to or better than those of clinical PI-RADS assessments (Table 1). Winkel *et al.*[65] proposed several HC models based on quantitative radiomics features. All models outperformed the PI-RADS assessment for the detection of csPCa in the peripheral zone (PZ). Dinh *et al.*[66] trained a computer-aided diagnosis system to identify csPCa in the PZ based on handcrafted radiomics features, which achieved comparable performance to those of experienced radiologists and higher performance than those of less-experienced radiologists. Schelb *et al.*[54] used a DL model [i.e., a twodimensional (2D) U-Net] trained with bi-parametric MRI (bpMRI) to achieve automatic segmentation and csPCa detection. The U-Net showed good agreement with the PI-RADS assessment by comparing the sensitivities and specificities. Similarly, Netzer *et al.*[67] confirmed that 2D U-Net could achieve a performance similar to that of PI-RADS assessment for the identification of csPCa. Zhong *et al.*[68] indicated that a DL model (i.e., a 2D ResNet) that used a transfer-learning method could distinguish indolent from csPCa lesions based on bpMRI, achieving a comparable performance to that of PI-RADS assessment. Deniffel *et al.*[69] used a DL model [i.e., a shallow 3D convolutional neural network (CNN)] based on bpMRI to diagnose csPCa. They revealed that the model resulted in fewer unnecessary biopsies compared to clinical assessment methods [e.g., PI-RADS-only assessment or the combination of PI-RADS assessment and PSA density (PSAD)]. Hiremath *et al.*[31] developed an integrated nomogram combining 2D DL, PI-RADS, and clinical characteristics, using a logistic regression method to identify csPCa. They found that the nomogram outperformed the PI-RADS assessment and a diagnostic model based on the combination of PI-RADS and clinical characteristics for diagnosing csPCa. Liu *et al.*[70] combined hand-crafted radiomics features and a DL method to diagnose csPCa by integrating a 3D gray-level co-occurrence matrix extractor into a DL network. The model was superior to PI-RADS assessment for detecting csPCa. Zhao *et al.*[71] developed 3D DL models based on multi-center bpMRI for diagnosing csPCa, which showed comparable performance to PI-RADS assessments of expert-level radiologists. Further, the integrated

**Table 1. Baseline characteristics of studies in the diagnosis field of prostate cancer**

No.	Study ID	Years	Country	Diagnosis task	Number of centers	Number of patients	AI methods	Clinical assessment methods	Comparison <sup>#</sup>
1	Hiremath et al.[31]	2021	USA	csPca	5	592	DL(2D)	Expert reader	>
2	Schelb et al.[54]	2019	Germany	csPca	1	312	DL(2D)	Expert reader and clinical characteristics <sup>†</sup>	>
3	Winkel et al.[65]	2020	Switzerland	csPca	1	201	HC	Expert reader	> (PZ)
4	Dinh et al.[66]	2018	France	csPca	2	235	HC	Expert reader	≈ (PZ)
5	Netzer et al.[67]	2021	Germany	csPca	2	1488	DL(2D)	Less-experience	> (PZ)
6	Zhong et al.[68]	2019	USA	csPca	1	140	DL(2D)	Expert reader	≈
7	Deniffel et al.[69]	2020	Canada	csPca	1	499	DL(3D)	Expert reader	>
8	Liu et al.[70]	2021	USA	csPca	1	402	HC, DL(3D)	Expert reader and PSA density <sup>†</sup>	>
9	Zhao et al.[71]	2023	China	csPca, PCa	7	1861	DL(3D)	Expert reader	>
10	Youn et al.[72]	2021	Korea	csPca	1	121	DL(3D)	Expert reader	<
11	Yu et al.[73]	2023	China	csPca	4	1540	DL(2D)	Less-experience	≈
12	Hectors et al.[74]	2021	USA	csPca	1	240	HC	Residents	>
13	Hou et al.[75]	2020	China	csPca	1	263	HC	General radiologist	> (internal)
14	Wang et al.[76]	2017	China	Pca	1	54	HC	General radiologist	≈ (external)
15	Li et al.[77]	2021	China	Pca	1	203	HC	Expert reader	<
16	Song et al.[78]	2018	China	Pca	1	195	DL(2D)	PSA density <sup>†</sup>	>
17	Aussavavirojekul et al.[79]	2022	Thailand	Pca	1	101	HC	Prostate volume <sup>†</sup>	>
18	Kan et al.[80]	2020	China	Pca	2	346	HC	Expert reader	>
19	Antonelli et al.[81]	2019	UK	Gleason pattern 4	1	164	HC	Reader	>
20	Niu et al.[82]	2018	China	High grade	1	184	HC	Expert reader	>
21	Algohary et al.[83]	2020	USA	D'Amico	4	231	HC	Reader	>
22	Zhang et al.[84]	2019	China	Risk	1	59	HC	Reader	>
							Clinical characteristics <sup>†</sup>		>

<sup>#</sup>Comparison between AI and clinical assessment methods. <sup>†</sup>Diagnostic models based on clinical characteristics. AI. Artificial intelligence; csPca. Clinically significant prostate cancer; DL(2D). Deep learning model based on two-dimensional networks; DL(3D). Deep learning model based on three-dimensional networks; HC. Hand-crafted; PCa. Prostate cancer; PZ. Peripheral zone; PSA. Prostate specific antigen; TZ. Transition zone



**Fig. 2 Representative examples of patients with prostate cancer.**

a. Patient with non-cSPCa; b. Patient with csPCa; c. Patient with ECE; d. Patient with LNI. The colorbar indicates the gray intensity of corresponding voxel in each sub-figure. The hot and cold colors indicate the strong and weak gray intensity, respectively. The red arrow points to the location of the lesion region. csPCa. Clinically significant prostate cancer; ECE. Extracapsular extension; LNI. Lymph node involvement

model combining the DL signature and PI-RADS assessment score achieved higher or equal area under the receiver operating characteristic curve (AUC) and greatly increased the specificity compared to PI-RADS assessment in the diagnosis of csPCa.

However, two recent studies reported decreased performance of a DL model for diagnosing csPCa compared to radiologists' assessments. Specifically, Youn *et al.*[72] found that, for the diagnosis of csPCa, the AUC of the DL method was similar to that of less-experienced radiologists but lower than that of experts. However, the sensitivity and specificity of the DL model were comparable to those of experts at a threshold PI-RADS score  $\geq 4$ . Yu *et al.*[73] developed a DL model that segmented prostate lesions automatically and diagnosed caPCa, whose performance was comparable or superior to general radiologists, but inferior to expert-level radiologists in diagnosing csPCa. Additionally, some other studies also developed the two-stage DL models including automatic lesion segmentation and diagnosis, though they did not compare their respective models with clinical assessment[85-87]. These studies suggested that automatic delineation of prostate lesions is of vital importance to reduce

the burden on radiologists and improve diagnostic accuracy.

Some recent studies using AI methods focused on the diagnosis of csPCa in lesions with a PI-RADS score 3 because they were equivocal for detecting csPCa[25,26]. For example, Hectors *et al.*[74] constructed an HC model to identify csPCa in lesions with a PI-RADS score 3. The model achieved a higher AUC than the diagnostic model based on clinical characteristics (e.g., PSA density or prostate volume). Hou *et al.*[75] developed an HC model to diagnose csPCa from lesions with a PI-RADS score 3. The model showed better performance than the reassessment results of expert radiologists.

The diagnosis of csPCa is a typical task in the diagnosis field of PCa, accounting for the largest proportion of the included studies (Table 1). For further comparison of the performance between AI and clinical assessment methods in csPCa diagnosis, we calculated the area under the summary receiver operator characteristic curve (SROC-AUC), pooled sensitivity, and pooled specificity of these two methods for csPCa diagnosis among the above-mentioned studies. The pooled sensitivity and specificity of each method (i.e., AI and clinical assessment) were calculated based on the summation

of the true positive (TP), false positive (FP), false negative (FN), and true negative (TN) across all included studies about csPCa diagnosis (Table 1). As summarized in Table 2, the AI methods of the studies on csPCa diagnosis presented an SROC-AUC of 0.87, a pooled sensitivity of 0.90, and a pooled specificity of 0.60. In contrast, the clinical assessment methods

of all studies of csPCa diagnosis presented an SROC-AUC of 0.82, a pooled sensitivity of 0.93, and a pooled specificity of 0.46. Compared with clinical assessment methods, AI methods achieved higher specificity with slight decrease in sensitivity. Additionally, in terms of SROC-AUC, the AI methods were superior to the clinical assessment methods.

**Table 2 Comparison of AI and clinical assessment methods in the diagnosis field of PCa**

Comparison	AI methods		Clinical assessment methods
	HC model	DL model	
Overall performance	Relatively high	-	Relatively poor
SROC-AUC*	0.87	-	0.82
Pooled sensitivity†	0.90	-	0.93
Pooled specificity‡	0.60	-	0.46
Qualitative or quantitative	Quantitative	-	Semi-quantitative
Expert dependence	Moderate	Low	High
Consistency	High	-	Low
Manual delineation	Yes	No	No
Features	High-throughput features extracted using specific algorithms (e.g., shape, histogram, and textural features)	Automatic extraction of deep and subtle image features using networks with substantial parameters	Features for visual assessments (e.g., location, shape, size, and intensity) and some clinical characteristics

\*Performance indexes pooled across the studies on csPCa diagnoses. "-": No data. AI. Artificial intelligence; csPCa. Clinically significant prostate cancer; DL. Deep learning; HC. Hand-crafted; PCa. Prostate cancer; PI-RADS. Prostate imaging reporting and data system; SROC-AUC. Area under the summary receiver operating characteristic curves

In general, the total performance of AI methods was better than that of clinical assessments for the studies on csPCa diagnosis. Specifically, for the studies on csPCa diagnosis, AI methods showed higher specificity but slightly less sensitivity compared to clinical assessment methods. In these studies, the clinical assessment methods showed a ceiling-approximate pooled sensitivity but a very low pooled specificity. Thus, the advantages of AI methods are mostly observed in the improved specificity of csPCa diagnosis. Notably, in both studies and clinical practice, PI-RADS is a common guideline for csPCa diagnosis based on medical images[26], which has been reported to have a low specificity[6]. Thus, the comparison results of these studies revealed that AI methods could improve the specificity of csPCa diagnosis, thereby reducing unnecessary confirmatory biopsies.

**Comparing the performances of AI and clinical assessment methods for PCa diagnosis**

In clinical practice, mis-identifying benign as malignant lesions in patients will cause anxiety and unnecessary trauma. Thus, accurately discriminating between benign (e.g., prostatic hyperplasia and inflammation) and malignant prostate lesions is crucial. MRI interpretation following the PI-RADS guidelines has been used as a common clinical assessment method for PCa diagnosis[25,26]. Thus, most studies

have compared the performance between AI methods and radiologists' interpretation of PI-RADS.

Several studies have reported that AI models can aid radiologists in PCa diagnosis by improving their visual assessments (Table 1). Wang *et al.*[76] reported that an HC model could improve the performance of PI-RADS assessment in diagnosing PCa, especially for lesions in the transitional zone (TZ). Li *et al.*[77] found that the difference of AUC between HC model and PI-RADS was insignificant. However, the integration of HC model and PI-RADS better diagnosed PCa than the PI-RADS assessment. Song *et al.*[78] reported that a joint model of DL and PI-RADS assessment outperformed either the DL model or PI-RADS assessment in diagnosing PCa. Zhao *et al.*[71] developed 3D DL models based on multi-center bpMRI for diagnosing PCa, which showed comparable performance to PI-RADS assessments of expert-level radiologists. Additionally, both Aussavavirojekul *et al.*[79] and Kan *et al.*[80] constructed HC models to detect PCa from lesions with a PI-RADS score 3, achieving specificities of 72% and 50%, respectively. It should be noted that in these two studies (i.e., [79] and [80]), equivocal lesions with PI-RADS score 3 were confirmed by biopsies. Thus, the AI methods in these studies can aid decision-making regarding whether a lesion with a PI-RADS score 3 should undergo

biopsy confirmation. Further, Luo *et al.*[88] reported that an AI-based image reconstruction algorithm might increase the MRI resolution and thereby improve the display effect, aiding in a better identification of PCa from benign prostatic hyperplasia.

#### **Comparing the performances of AI and clinical assessment methods for other diagnostic tasks for risk stratification of PCa**

In addition to diagnosing PCa and csPCa, AI methods have been used to aid other diagnostic tasks for risk stratification of PCa. Several studies have reported that AI models can aid radiologists in improving their visual assessments of other diagnostic tasks (Table 1). For example, Antonelli *et al.*[81] found that an HC method showed better performance in recognizing lesions with Gleason pattern 4 than three board-certified radiologists' assessments. Niu *et al.*[82] reported that an HC model could detect high-grade PCa and performs better than PI-RADS assessment. Algohary *et al.*[83] developed an HC model combining peritumoral and intratumoral radiomics features to accurately stratify PCa risk that was defined by the D'Amico Risk Classification System, resulting in higher accuracy compared to the PI-RADS assessment. Zhang *et al.*[84] used the logistic regression method combining hand-crafted radiomics features and clinical characteristics to differentiate between high- and low-grade PCa, which outperformed the diagnosis model based on clinical characteristics.

#### **Summary of the comparison between AI and clinical assessment methods in the diagnosis field of PCa**

Table 2 summarized the differences between AI and clinical assessment methods for the diagnosis field of PCa. First, AI methods achieved a better overall performance than that of clinical assessments of radiologists for the diagnosis of PCa. In particular, for the detection of csPCa, the SROC-AUC and pooled specificity of AI methods were both higher than those of the clinical assessment methods. Further, different from the clinical assessment methods, the AI methods provided a quantitative result, which relied much less on the individual expertise of radiologists and thereby could achieve consistent diagnosis results. The better performance of AI compared with clinical assessment methods was because the former can mine subtle and deep information of MR images of PCa, which was not accessible by common clinical assessment methods. Though the HC methods required precisely manual delineations of prostate lesions, the DL methods achieved an automatic image-to-decision diagnosis.

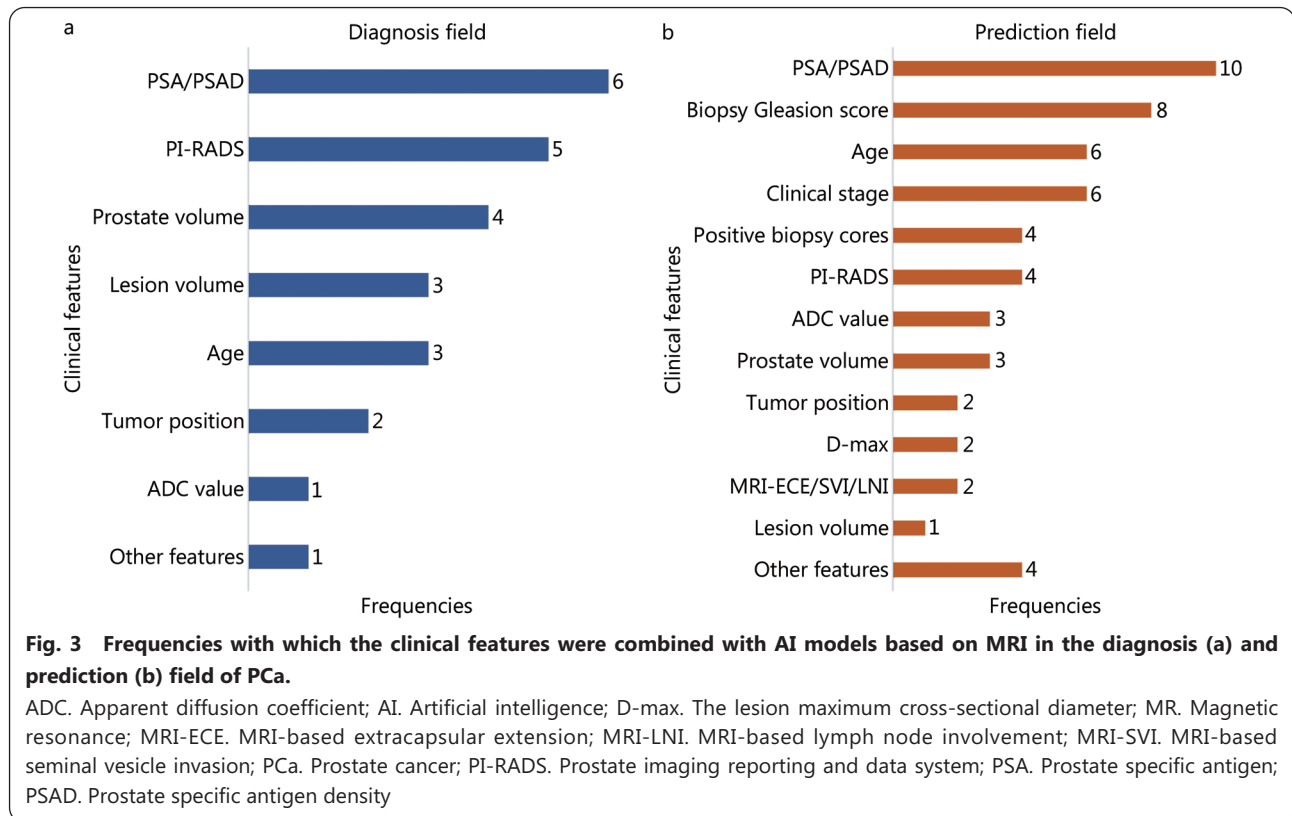
Because the image features and clinical features include different information for characterizing prostate lesions,

the combination of the clinical characteristics and the AI models based on MR images improves the performance of diagnosing PCa (e.g.,[31,71]). As shown in Fig. 3a, for the integrated AI models included in the present study, the clinical characteristics of PSA/PSAD, PI-RADS, and prostate volume were frequently combined with the AI models.

Additionally, in some of the included studies, AI models were trained and tested using a large number of samples from two- or multi-centers[31,66,67,71,73,80,83], demonstrating, to some degree, the potential robustness and generalization of these models in the diagnosis field of PCa. These findings suggest that AI models may effectively aid radiologists in improving the diagnosis field of PCa. Furthermore, as in most studies, PI-RADS is a common clinical assessment method for the diagnosis field of PCa in clinical practice. However, as a semiquantitative scoring system, PI-RADS is associated with low specificity in the diagnosis field of PCa[6], leading to unnecessary confirmation using biopsies[18]. Thus, the combination of AI models and PI-RADS assessment potentially reduces over-biopsies by improving specificity in the diagnosis field of PCa.

However, these studies had several limitations regarding the AI-assisted diagnosis of PCa. First, although two or multi-center samples were used in a few studies, most employed small and monocenter patient cohorts. Thus, the generalizability of the proposed AI models for the diagnosis field of PCa requires further validation. Second, most DL radiomics models for PCa diagnosis were constructed based on a 2D CNN. However, for models with 2D CNN, the final patient-level output results are usually the average of the predicted values of multiple slice images without considering the spatial relationship between slices[89]. 3D CNN can fully utilize spatial information and achieve accurate patient-level prediction. Thus, 3D DL radiomics models should be used in future research. Third, some studies used independent external validation cohorts to evaluate model performance; however, they were generally retrospective. Therefore, these models should be validated using prospective data. Finally, HC models still require precise manual annotation, which is time-consuming and laborious, and requires automatic annotation for PCa lesions.

In general, when applied to the analysis of medical images, AI methods present advantages over clinical assessments in various aspects, such as the high-throughput extraction of image information, overall characterization of lesion heterogeneity, and multi-variable analysis of image features. Thus, it was not unexpected that the AI methods outperformed the clinical assessment methods in the diagnosis field of PCa. The advantages of AI methods over clinical assessment methods are



highly consistent with studies on other tumors, such as breast cancer[90,91], brain tumor[92], renal cancer[93], and cervical cancer[94], suggesting that AI methods are potential tools to aid in the precise diagnosis of PCa.

**Comparing the performances of AI and clinical assessment methods in the prediction field of PCa**

AI methodology has been widely utilized to aid in the prediction fields of PCa. In clinical practice, the prediction tasks mainly include predicting lymph node involvement (LNI), ECE, postoperative BCR, and other events. Table 3[57,58,95-108] listed the studies for the prediction field of PCa, which were included in the present study.

**Comparing the performances of AI and clinical assessment methods for AP prediction**

AP features (e.g., ECE and LNI) are known to be important predictors of tumor metastasis, and therefore, accurate prediction of the presence of AP features can significantly aid in treatment decisions (e.g., planning of personalized surgical treatment)[10,15]. Figure 2c and d show representative examples of patients with ECE and LNI, respectively. In clinical practice, radiologists' interpretations and nomograms based on clinical characteristics (e.g., PSA level, Gleason grade, and positive biopsy cores) are commonly used as clinical assessment methods to predict AP. Thus, most studies

have compared the performance between AI methods and radiologists' interpretations or nomograms based on clinical characteristics.

Several studies have reported that the performance of AI models is equal to or better than that of clinical assessment methods (Table 3). Hou *et al.*[58] developed a DL model that contained an attention map of experts' prior knowledge to detect ECE. It showed better performance than radiologists' interpretations. The study also reported that the performance of radiologists' interpretations in ECE prediction was improved with the assistance of the DL model. Xu *et al.*[95] built an HC model to predict ECE that outperformed a prediction model combining clinical and pathological characteristics. Ma *et al.*[96] also constructed an HC model to predict the presence of ECE that outperformed radiologists' interpretations. Additionally, some studies have reported that prediction models combining radiomics features and clinical characteristics achieved excellent performance. For example, Bai *et al.*[97] constructed a logistic regression model combining peritumoral hand-crafted radiomics features and clinical characteristics to predict ECE. The model achieved comparable or better performance than a prediction model based on clinical characteristics. Bourbonne *et al.*[98] proposed a DL model combining hand-crafted radiomics features and clinical characteristics to predict LNI in PCa

**Table 3 Baseline characteristics of studies in the prediction field of prostate cancer**

No.	Study ID	Years	Country	Prediction task	Number of centers	Number of patients	AI methods	Clinical assessment methods	Comparison <sup>#</sup>
1	Yan et al.[57]	2021	China	BCR	3	485	HC, DL(2D)	CAPRA-S/NCCN/GGG	>
2	Hou et al.[58]	2021	China	ECE	2	849	DL(2D)	Expert reader	>
3	Xu et al.[95]	2020	China	ECE	1	95	HC	Clinical and pathological characteristics <sup>†</sup>	>
4	Ma et al.[96]	2019	China	ECE	1	210	HC	Reader	>
5	Bai et al.[97]	2021	China	ECE	2	284	HC	Clinical characteristics <sup>†</sup>	< (internal) ≈ (external)
6	Bourbonne et al.[98]	2021	France	LNI	1	280	HC, DL(2D)	Partin/Roach/Yale/MSKCC/...	>
7	Hou et al.[99]	2019	China	LNI	1	248	HC	MSKCC	>
8	Hou et al.[100]	2021	China	LNI	2	401	HC, DL(2D)	MSKCC/Briganti/...	>
9	Li et al.[101]	2021	USA	AP, BCR	4	198	HC	CAPRA/Decipher/CAPRA-S	>/≈
10	Bourbonne et al.[102]	2020	France	BCR	2	195	HC	Clinical characteristics <sup>†</sup>	>
11	Sushentsev et al.[103]	2022	UK	Progressing	1	64	HC	PRECISE	≈
12	Xie et al.[104]	2021	China	Upgrade	1	59	HC	Biopsy	>
13	Zhang et al.[105]	2020	China	Upgrade	1	166	HC	Clinical characteristics <sup>†</sup>	>
14	Wu et al.[106]	2017	China	Upgrade	1	46	HC	Clinical characteristics <sup>†</sup>	>
15	Zheng et al.[107]	2021	USA	Biopsy results	1	330	HC	PSA density <sup>†</sup>	>
16	Wang et al.[108]	2018	China	Organ-confined Pca	1	541	HC	Partin	>

<sup>#</sup>Comparison between AI methods and clinical assessment methods. <sup>†</sup>Predictive models based on clinical characteristics. AI. Artificial intelligence; AP. Adverse pathology; BCR. Biochemical recurrence; CAPRA. Cancer of the Prostate Risk Assessment; DL(2D). Deep learning model based on two-dimensional networks; ECE. Extracapsular extension; GGG. Gleason grade group; HC. Hand-crafted; LNI. Lymph node involvement; MSKCC. Memorial Sloan Kettering Cancer Center; NCCN. National Comprehensive Cancer Network; PSA. Prostate specific antigen; PRECISE. Prostate Cancer Radiological Estimation of Change in Sequential Evaluation

patients. The model provided a higher C-index than other clinical nomograms [i.e., Partin, Roach, Yale, and Memorial Sloan Kettering Cancer Center (MSKCC)]. Hou et al.[99] selected 18 features, including hand-crafted radiomics features and clinical characteristics, and developed several models (i.e., logistic regression, SVM, and RF) to predict LNI. The predictive performances of these models were superior to those of the MSKCC nomogram. Hou et al.[100] used an RF model combining clinicopathological factors, radiologists' interpretations, hand-crafted radiomics features, and DL radiomics features to predict LNI. The performance of the model was superior to those of the MSKCC, Briganti and any other model based on a single type of characteristic or a combination of the two types of characteristics in the internal and external testing cohorts. Li et al.[101] developed a nomogram combining hand-crafted radiomics and clinicopathologic features to predict the presence of AP of PCa. The nomogram outperformed the Cancer of the Prostate Risk Assessment (CAPRA) and the Decipher test for predicting the presence of AP.

AP prediction is a common task in the prediction field of PCa accounting for the largest proportion of the included

studies (Table 3). For further comparison of the performance between AI and clinical assessment methods in AP prediction, we calculated the SROC-AUC, pooled sensitivity, and pooled specificity of these two methods among the above-mentioned studies. As summarized in Table 4, the AI methods for AP prediction presented an SROC-AUC of 0.86, a pooled sensitivity of 0.75, and a pooled specificity of 0.84. In contrast, the clinical assessment methods of AP feature prediction presented an SROC-AUC of 0.75, a pooled sensitivity of 0.68, and a pooled specificity of 0.79. In terms of the above performance indexes, AI methods were superior to clinical assessment methods.

#### Comparing the performances of AI and clinical assessment methods for BCR prediction

In clinical practice, patients with BCR after RP or other therapies may present more advanced disease, distant metastasis, and even death[52]. Thus, early identification of BCR can help to make treatment decisions.

Recently, AI has been widely used for BCR prediction[57,101,102,109,110]. Several studies have compared AI and clinical assessment methods and demonstrated that the performance of AI models is better than that of clinical

**Table 4 Comparison of AI and clinical assessment methods in the prediction field of PCa**

Comparison	AI methods		Clinical assessment methods
	HC model	DL model	
Overall performance	Relatively high	-	Relatively poor
SROC-AUC*	0.86	-	0.75
Pooled sensitivity†	0.75	-	0.68
Pooled specificity†	0.84	-	0.79
Qualitative or quantitative	Quantitative	-	Quantitative/qualitative
Expert dependence	Moderate	Low	High
Consistency	High	-	Moderate
Manual delineation	Yes	No	No
Features	High-throughput features extracted using specific algorithms (e.g., shape, histogram and textural features)	Automatic extraction of deep and subtle image features using networks with substantial parameters	Clinical characteristics (e.g., PSA, Gleason grade and positive biopsy cores) or features for visual assessments (e.g., location, shape, size, and intensity)

\*Performance indexes pooled across the studies of AP prediction. "-". No data. AI. Artificial intelligence; AP. Adverse pathology; DL. Deep learning; HC. Hand-crafted; PCa. Prostate cancer; PSA. Prostate specific antigen; SROC-AUC. Area under the summary receiver operator characteristic curves

assessment methods for BCR prediction (Table 3). For example, Yan *et al.*[57] extracted hand-crafted radiomics features and developed a DL model to predict BCR after RP with MR images, which outperformed other clinical assessment methods (e.g., CAPRA-S score, NCCN model, and Gleason grade group systems). Li *et al.*[101] developed a nomogram combining hand-crafted radiomics features and clinicopathologic features to predict the post-surgical BCR of PCa. The nomogram yielded a higher C-index than CAPRA and Decipher and was equal to CAPRA-S for the prediction of BCR. Bourbonne *et al.*[102] trained an HC model to predict the BCR for high-risk PCa, which outperformed other prediction models based on clinical characteristics.

**Comparing the performances of AI and clinical assessment methods for other predictive tasks for PCa prognosis**

In addition to predicting AP and BCR, AI methods have been used to aid other predictive tasks for PCa. Several studies have reported that the performance of AI models is equal to or better than that of clinical assessment methods for other predictive tasks (Table 3). For example, Sushentsev *et al.*[103] developed an HC model for predicting PCa progression in patients undergoing active surveillance. The model's performance was comparable to that of the clinical assessment method [i.e., Prostate Cancer Radiological Estimation of Change in Sequential Evaluation (PRECISE)]. Xie *et al.*[104] extracted textural features from ADC maps and developed HC models to predict pathological upgrading from biopsy to RP. These models showed excellent performance, suggesting that they can improve the diagnostic accuracy of biopsy and avoid missed detection of high-grade PCa. Zhang *et al.*[105] built a logistic regression model combining hand-crafted radiomics

features and clinical characteristics to predict upgrading from biopsy to RP. It outperformed the prediction model based on clinical characteristics including clinical stage and time from biopsy to RP. Wu *et al.*[106] developed a logistic regression model combining diffusion kurtosis imaging and PSA to predict upgrading after RP. It outperformed other prediction models based on clinical characteristics. Zheng *et al.*[107] trained an SVM model combining hand-crafted radiomics features and clinical characteristics to predict biopsy results for patients with negative MRI findings. The proposed model was superior to PSA density-based risk assessment. Wang *et al.*[108] developed an SVM model combining clinical characteristics (i.e., age, PSA, clinical stage, and biopsy Gleason score) and MRI findings (i.e., tumor location, PI-RADS scores, diameter, and 6-point MRI stage) for the prediction of organ-confined PCa, which outperformed the clinical assessment method (i.e., Partin table).

**Summary of the comparisons between AI and clinical assessment methods in the prediction field of PCa**

Table 4 summarized the differences between AI and clinical assessment methods for the prediction field of PCa. First, AI methods achieved a better overall performance than that of clinical assessments of radiologists for the prediction of PCa. In particular, for the prediction of AP presence, all the SROC-AUC, pooled sensitivity and pooled specificity of AI methods were higher than those of the clinical assessment methods. Although both AI and clinical assessment methods can provide a quantitative result, the former rely much less on the individual expertise of radiologists and thereby could achieve consistent prediction results. Additionally, the AI methods can extract high-throughput features and subtle information,

the majority of which were not accessible by the clinical assessment methods.

Like those in the diagnosis field of PCa, the integrated AI models combining the AI models based on MR images and clinical characteristics achieved an increased performance for the prediction of PCa (e.g., [100,101]). As shown in Fig. 3b, for the integrated AI models included in the prediction field, the clinical characteristics of PSA/PSAD, biopsy Gleason score, age, clinical stage (C-stage), positive biopsy cores, and PI-RADS were frequently combined with the AI models.

Overall, studies in the prediction field of PCa mostly focused on the prediction of LNI, ECE, BCR, and pathological upgrading from biopsy to RP. All abovementioned studies reported better or comparable performances of AI models compared to those of clinical assessment methods. In particular, in some of these studies, the AI models were tested using an external testing cohort [57,58,97,100-102], demonstrating, to some degree, the potential robustness and generalization of these models in clinical application to the prediction field of PCa. These findings suggest that AI models may effectively improve preoperative prediction performance and assist clinicians in making treatment decisions. Furthermore, as in some studies, some risk assessment tools, such as Partin tables, MSKCC nomogram, and CAPRA score, showed moderate predictive performance on validation [57,98,100,101]. Additionally, most of studies employed HC models as prediction methods, the number of which was much larger than that of studies using DL models. Furthermore, all studies using DL models employed 2D CNN without considering the 3D spatial information of tumors.

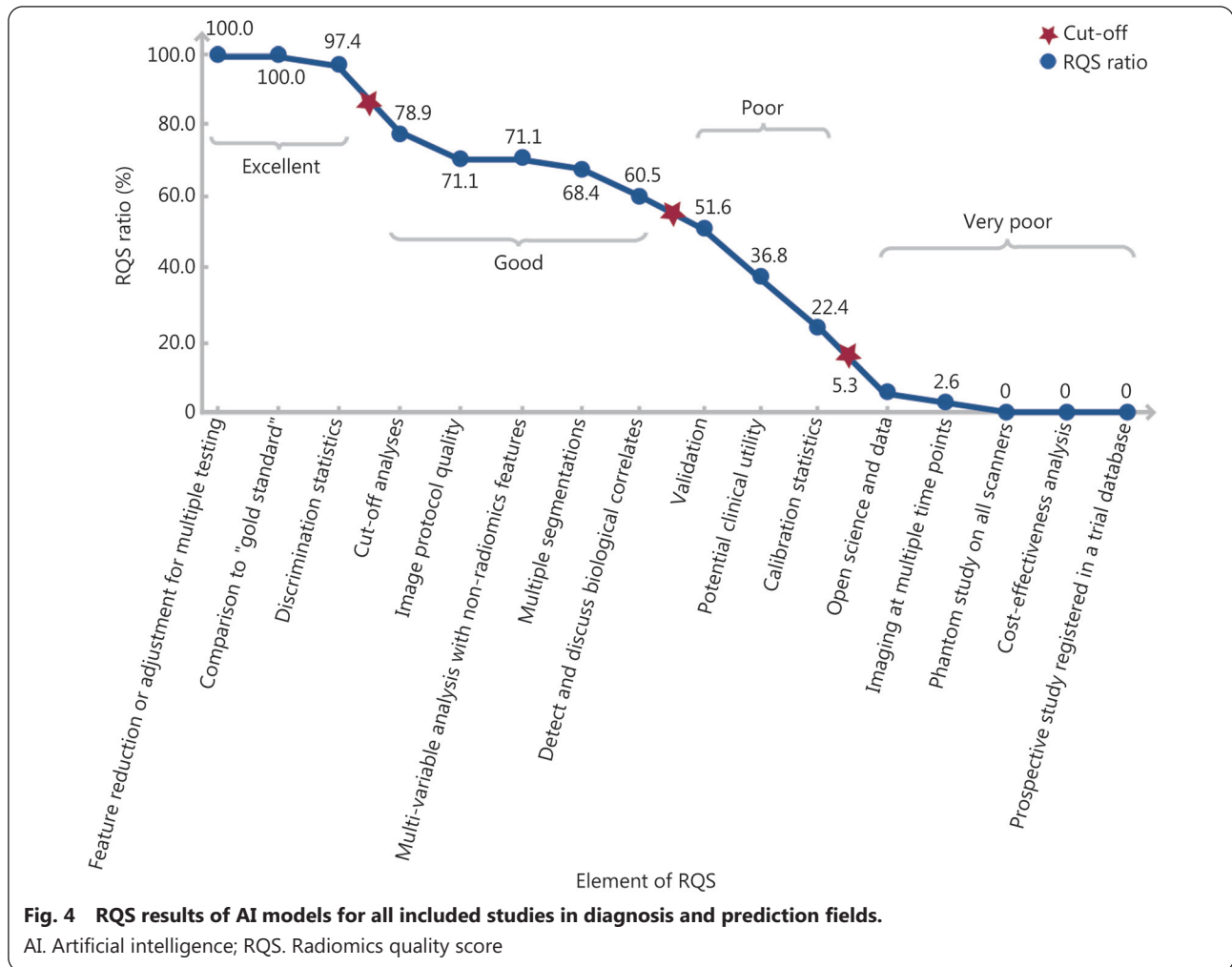
In clinical practice, radiologists' interpretations and nomograms, such as CAPRA scores and MSKCC nomograms, are commonly used to predict therapeutic outcomes. These nomograms integrated multiple clinicopathological features, such as PSA, Gleason grade, and positive biopsy cores, but failed to account for tumor heterogeneity, resulting in relatively poor performance. MRI can visually and comprehensively describe the characteristics and morphology of tumors associated with tumor aggressiveness and progression [15]. However, MRI interpretation based on some evidently visual features of lesions (e.g., size, location, and intensity) requires a high level of expertise by radiologists, leading to inter-observer variability. Furthermore, lesions with low volumes may be missed in visual assessments, and various invisible features (e.g., subtly textural and advanced features) have also been associated with PCa aggressiveness and progression. In contrast, AI methods can automatically extract features from images, reducing the dependence on the high-level expertise

of radiologists. Additionally, AI methods can extract visible features and mine invisible high-throughput information, thus overcoming the limitations of radiologists' interpretations. The advantages of AI methods over clinical assessment methods were highly consistent with those of studies on other tumors, such as breast cancer [36,90], brain tumor [92,111,112], rectal cancer [37,113,114], gastric cancer [41,114,115], colon cancer [116], lung cancer [117-119], and cervical cancer [120,121], suggesting that AI methods are potential tools to aid the precise prediction of PCa.

### RQS assessment

The RQS assessment [122] was performed for all included studies in both diagnosis (Table 1) and prediction (Table 3) fields of PCa of the present study. The RQS assessments were conducted independently by two reviewers. The disagreement between the reviewers was resolved by discussion until achieving an agreement. These studies presented a total average RQS of 15.2 (11.0–20.0) with a total average RQS ratio of 42.2%, which was defined as the ratio of the total average RQS to the full points (i.e., 15.2/36). This total average RQS ratio is higher than those of some recent radiomics studies [123-125] (average RQS ratio: 14.3%–29.6%). The RQS ratio of each RQS element was defined as the ratio of the average RQS across all included studies to the full points for the corresponding element, which reflected the degree to which all included studies met the requirements of the corresponding RQS element. When the RQS elements were sorted in descending order of their RQS ratios, they were divided into four levels by three evident cutoffs: excellent, good, poor, and very poor (Fig. 4).

As shown in Fig. 4, the excellent elements (i.e., RQS ratio: 97.4%–100.0%) included “feature reduction or adjustment for multiple testing”, “comparison to gold standard”, and “discrimination statistics”. The good elements (i.e., RQS ratio: 60.5%–78.9%) included “cut-off analyses”, “image protocol quality”, “multi-variable analysis with non-radiomics features”, “multiple segmentations”, and “detect and discuss biological correlates”. Some of the excellent and good elements (e.g., “feature reduction or adjustment for multiple testing”, “comparison to gold standard”, “discrimination statistics”, “cut-off analyses”, and “multiple segmentations”) are closely related to several processes of AI model construction, such as the extraction and selection of radiomics features, as well as the assessment and comparison of model performance, suggesting the relatively perfect and standard construction processes of AI models for the included studies. Additionally, the RQS elements of “detect and discuss biological correlates”



and “multi-variable analysis with non-radiomics features” also presented a relatively high RQS ratio, highlighting some clinical significance of the included studies.

Poor elements (i.e., RQS ratio: 22.4%–51.6%) included “validation”, “potential clinical utility”, and “calibration statistics”. The very poor elements (i.e., RQS ratio: 0%–5.3%) included “open science and data”, “imaging at multiple time points”, “phantom study on all scanners”, “cost-effectiveness analysis”, and “prospective study registered in a trial database”. It is noted that, among these elements, three have the top three full points, namely “prospective study registered in a trial database” (full points of 7), “validation” (full points of 5), and “open science and data” (full points of 4). However, they presented relatively low RQS, particularly the element of “prospective study registered in a trial database” even presenting a zero average RQS ratio, suggesting that none of the included studies used prospective data to test the AI models. These three RQS elements are related to the assessment of the generalizability and replicability of the AI models. Therefore, these low RQS ratios suggest that the robustness of AI models

in most of the included studies is unclear. Additionally, the elements of “phantom study on all scanners”, “imaging at multiple time points”, and “cost-effectiveness analysis” also had very low average RQS ratios. This may be because all included studies were retrospective. Considering that “phantom study on all scanners” and “imaging at multiple time points” facilitates the examination of feature robustness to inter-scanner differences and temporal variabilities [122], this should be considered in future prospective studies.

According to above description, the RQS element can be mostly categorised into two groups. One is related to the performance improvement of AI models, such as “feature reduction or adjustment for multiple testing”, “multivariable analysis with non-radiomics features”, “multiple segmentations” and “phantom study on all scanners”. Specifically, for the element of “feature reduction or adjustment for multiple testing”, feature selection or dimensionality reduction for the extracted features with high redundancy and/or strong collinearity can optimize feature space, thereby improving the performance of the model [74-76,97,99]. For the element

of “multi-variable analysis with non-radiomics features”, AI can mine subtle information that may reflect heterogeneities of the lesions, but the commonly used clinical characteristics (e.g., PSA, age, family history, and routine habits) also contain information relevant to the diagnosis and prognosis. Thus the clinical characteristics were complementary to the radiomics features of MR images to improve the performance of model [31,71,101]. For the element of “multiple segmentations”, the delineation of lesions with different methods (e.g., automatic and manual), by different radiologists and software, and in different stages of the breathing cycles is helpful to reduce the discrepancy between the delineated region of interest (ROI) and the actual lesion. This may heighten the robustness and accuracy of the extracted features, based on which an AI model is constructed [31,54,71]. For the element of “phantom study on all scanners”, when radiomics features come from images scanned by multiple scanners, it is important to consider feature variabilities between scanners. The phantom study is an appropriate way to measure the uncertainties of different scanners.

The other group of elements is related to the performance evaluation of the AI model, such as “validation”, “prospective study registered in a trial database”, “cut-off analyses” and “potential clinical utility”. Specifically, for “validation” and “prospective study registered in a trial database”, the testing of an AI model using independent external cohort, particularly the prospective samples, can fully evaluate the robustness and generalisation of the model, thereby reducing the overfitting of AI models. For the element of “cut-off analyses”, some performance indexes, such as sensitivity and specificity, are calculated dependently on the risk threshold. The traditional default value of 0.5 doesn't exactly reflect the clinical problem. Therefore, the appropriate risk threshold should be selected in conjunction with clinical decisions [54,67,71]. For the element of “potential clinical utility”, analyzing potential applications of the model in a clinical practice is of vital importance to make clinical promotion [31,67,75].

In addition to the RQS elements, several factors may affect AI performance, such as the development of new AI models, large data queues and interaction between AI methods and clinical problems. First, new breakthroughs of AI technology can provide new networks model with a powerful ability to extract the deep features of medical imaging and combine multiple imaging modalities and multiple time points. This can make the model more fully and accurately characterize the tumor heterogeneity. Second, an AI model trained using a larger data queue based on different institutions and different regions (e.g., cities or countries) can show stronger robustness

and generalization. Finally, the development of an AI model that is orientated to a specific clinical problem may make the model have more clinical applicability. Like the RQS elements, these three factors can also effectively improve the performance of AI modes for aiding in the precise diagnoses and treatments of prostate tumors.

Although many proposed AI models have been demonstrated better performance than clinical assessment methods for the diagnosis and prediction of PCa, they are not yet extensively used in the clinical practice. The causes are various and difficult to clearly list. One possible cause may be that the AI models were usually trained using a relatively limited number of samples. Compared to the expertise of clinicians that is accumulated based on tens of years of experience, the limited-trained AI models provide the clinicians with less confidence in the management of PCa. Additionally, the weak interpretability of the AI model may be another cause. Thus, it is very difficult for the clinicians to combine the predicting results of the model with their expertise for the diagnoses and treatments of PCa.

## Conclusions and prospect

In this review, we summarized the studies including the performance comparisons between AI and clinical assessment methods applied to PCa. Several findings were obtained: First, the performance of AI methods was generally better than clinical assessment methods for the diagnosis and prediction fields of PCa, particularly for the detection of csPCa and prediction of some AP features, indicating that AI can aid clinicians in making accurate decisions (e.g., reducing the frequency of unnecessary biopsies and making personalized treatment plans). Second, the AI models were constructed with relatively perfect and standard radiomics processes. However, due to inadequate multi-center validation, prospective data testing, or the opening of the research material, the generalizability and clinical practicality of AI models should be further validated.

In the future, AI models can be improved in the following potential aspects. First, the AI models will be validated using a high level of evidence, such as different race data and prospective data. Second, the combination of radiomics models based on MR images and Natural Language Processing based on medical records will provide more comprehensive information and reduce the burden on radiologists. Third, more state-of-the-art and complex AI methods, such as ones integrating the expertise of radiologists into the latest network architectures, will be developed to further improve the diagnosis and prediction of PCa. Additionally, for the

applications of AI methods in the management of PCa, AI methodology should further extend to the fields beyond the above-mentioned tasks in the present study, such as the identification of the patients qualified for active surveillance, the prediction of local recurrence, survival analysis, and comparison of the prognosis of different treatment plans.

### Abbreviations

ADC: Apparent diffusion coefficient; AI: Artificial intelligence; AP: Adverse pathology; AUC: Area under the receiver operating characteristic curve; BCR: Biochemical recurrence; bpMRI: Bi-parametric MRI; CAPRA: Cancer of the Prostate Risk Assessment; CNN: Convolutional neural network; csPCa: Clinically significant prostate cancer; CT: Computed tomography; DL: Deep learning; DL(2D): Deep learning model based on two-dimensional networks; DL(3D): Deep learning model based on three-dimensional networks; DWI: Diffusion-weighted imaging; D-max: The lesion maximum cross-sectional diameter; ECE: Extracapsular extension; GGG: Gleason grade group; HC: Hand-crafted; LASSO: Least absolute shrinkage and selection operator; LNI: Lymph node involvement; LR: Logistic regression; MRI: Magnetic resonance imaging; MRI-ECE: MRI-based extracapsular extension; MRI-LNI: MRI-based lymph node involvement; MRI-SVI: MRI-based seminal vesicle invasion; MSKCC: Memorial Sloan Kettering Cancer Center; NCCN: National Comprehensive Cancer Network; PCa: Prostate cancer; PI-RADS: Prostate imaging reporting and data system; PRECISE: Prostate Cancer Radiological Estimation of Change in Sequential Evaluation; PSA: Prostate specific antigen; PSAD: Prostate specific antigen density; PZ: Peripheral zone; RFE: Recursive feature elimination; RP: Radical prostatectomy; RQS: Radiomics quality score; SROC: Summary receiver operator characteristic curve; SROC-AUC: Area under the summary receiver operating characteristic curve; SVM: Support vector machine; T<sub>2</sub>WI: T<sub>2</sub>-weighted imaging; TZ: Transition zone.

### Acknowledgements

Not applicable.

### Author contributions

JGL, ZYL, JT and JL conceived and designed the project; LTZ and WFX performed radiomics quality score assessment; LTZ, JGL and ZYL drafted the manuscript; LZS performed data review; all authors provide review, editing, and final approval of the manuscript.

### Funding

This work was supported by the Natural Science Foundation of Beijing (Z200027), the National Natural Science Foundation of China (62027901, 81930053), and the Key-Area Research and Development Program of Guangdong Province (2021B0101420005).

### Availability of data and materials

Not applicable.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

### Competing interests

The authors declare that there is no competing interests.

### Author details

<sup>1</sup>School of Engineering Medicine, Beihang University, Beijing 100191, China. <sup>2</sup>School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China. <sup>3</sup>CAS Key Laboratory of Molecular Imaging, Institute of Automation, Beijing 100190, China. <sup>4</sup>University of Chinese Academy of Sciences, Beijing 100080, China. <sup>5</sup>Department of Urology, Peking University Third Hospital, Peking University, Beijing 100191, China. <sup>6</sup>Key Laboratory of Big Data-Based Precision Medicine (Beihang University), Ministry of Industry and Information Technology of the People's Republic of China, Beijing 100191, China. <sup>7</sup>Beijing Engineering Research Center of Cardiovascular Wisdom Diagnosis and Treatment, Beijing 100029, China.

### 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 Cancer J Clin.* 2021;71(3):209–49.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin.* 2021;71(1):7–33.
3. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72(1):7–33.
4. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115–32.
5. Cao W, Chen HD, Yu YW, Li N, Chen WQ. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin Med J (Engl).* 2021;134(7):783–91.
6. Woo S, Suh CH, Kim SY, Cho JY, Kim SH. Diagnostic performance of prostate imaging reporting and data system version 2 for detection of prostate cancer: a systematic review and diagnostic meta-analysis. *Eur Urol.* 2017;72(2):177–88.
7. Woo S, Suh CH, Kim SY, Cho JY, Kim SH. Diagnostic performance of magnetic resonance imaging for the detection of bone metastasis in prostate cancer: a systematic review and meta-analysis. *Eur Urol.* 2018;73(1):81–91.
8. Nagpal K, Foote D, Tan F, Liu Y, Chen PHC, Steiner DF, et al. Development and validation of a deep learning algorithm for gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol.* 2020;6(9):1372–80.
9. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol.* 2020;21(2):233–41.
10. Hwang WL, Tendulkar RD, Niemierko A, Agrawal S, Stephans KL, Spratt DE, et al. Comparison between adjuvant and early-salvage postprostatectomy radiotherapy for prostate cancer with adverse pathological features. *JAMA Oncol.* 2018;4(5):e175230.
11. Kornberg Z, Cooperberg MR, Cowan JE, Chan JM, Shinohara K, Simko JP, et al. A 17-gene genomic prostate score as a predictor of adverse pathology in men on active surveillance. *J Urol.* 2019;202(4):702–9.
12. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol.* 2019;16(7):391–403.
13. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol.* 2016;2(12):1636–42.

14. NICE Guidance. Prostate cancer: diagnosis and management. *BJU Int.* 2019;124(1):9–26.
15. Schaeffer EM, Srinivas S, Antonarakis ES, Armstrong AJ, Cheng HH, D'Amico AV, et al. NCCN clinical practice guidelines in oncology, Prostate Cancer version 1. 2022. NCCN. 2022. <https://www.isotopia-global.com/wp-content/uploads/2022/04/NCCN-guidelines-prostate-cancer-2022.pdf>.
16. Loeb S, Vellekoop A, Ahmed HU, Catto J, Emberton M, Nam R, et al. Systematic review of complications of prostate biopsy. *Eur Urol.* 2013;64(6):876–92.
17. Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Myrders LA, Vaarala MH, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med.* 2018;378(19):1767–77.
18. Ahmed HU, El-Shater Bosaily A, Brown LC, Gabe R, Kaplan R, Parmar MK, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet.* 2017;389(10071):815–22.
19. Renard-Penna R, Mozer P, Cornud F, Barry-Delongchamps N, Bruguière E, Portalez D, et al. Prostate imaging reporting and data system and likert scoring system: multiparametric MR imaging validation study to screen patients for initial biopsy. *Radiology.* 2015;275(2):458–68.
20. Kuhl CK. Abbreviated magnetic resonance imaging (MRI) for breast cancer screening: rationale, concept, and transfer to clinical practice. *Annu Rev Med.* 2019;70(1):501–19.
21. Meier-Schroers M, Homsí R, Gieseke J, Schild HH, Thomas D. Lung cancer screening with MRI: evaluation of MRI for lung cancer screening by comparison of LDCT- and MRI-derived lung-RADS categories in the first two screening rounds. *Eur Radiol.* 2019;29(2):898–905.
22. Kim SY, Cho N, Hong H, Lee Y, Yoen H, Kim YS, et al. Abbreviated screening MRI for women with a history of breast cancer: comparison with full-protocol breast MRI. *Radiology.* 2022;305(1):36–45.
23. Rosén R, Nilsson E, Rahman M, Rönnow CF. Accuracy of MRI in early rectal cancer: national cohort study. *Br J Surg.* 2022;109(7):570–2.
24. Tatsubayashi T, Tanizawa Y, Miki Y, Tokunaga M, Bando E, Kawamura T, et al. Treatment outcomes of hepatectomy for liver metastases of gastric cancer diagnosed using contrast-enhanced magnetic resonance imaging. *Gastric Cancer.* 2017;20(2):387–93.
25. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS prostate imaging–reporting and data system: 2015, version 2. *Eur Urol.* 2016;69(1):16–40.
26. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, 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(3):340–51.
27. Le JD, Tan N, Shkolyar E, Lu DY, Kwan L, Marks LS, et al. Multifocality and prostate cancer detection by multiparametric magnetic resonance imaging: correlation with whole-mount histopathology. *Eur Urol.* 2015;67(3):569–76.
28. Radtke JP, Schwab C, Wolf MB, Freitag MT, Alt CD, Kesch C, et al. Multiparametric magnetic resonance imaging (MRI) and MRI-transrectal ultrasound fusion biopsy for index tumor detection: correlation with radical prostatectomy specimen. *Eur Urol.* 2016;70(5):846–53.
29. Rosenkrantz AB, Ginocchio LA, Cornfeld D, Froemming AT, Gupta RT, Turkbey B, et al. Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists. *Radiology.* 2016;280(3):793–804.
30. Moldovan PC, Van den Broeck T, Sylvester R, Marconi L, Bellmunt J, van den Bergh RCN, et al. What is the negative predictive value of multiparametric magnetic resonance imaging in excluding prostate cancer at biopsy? A systematic review and meta-analysis from the European Association of urology prostate Cancer Guidelines Panel. *Eur Urol.* 2017;72(2):250–66.
31. Hiremath A, Shiradkar R, Fu P, Mahran A, Rastinehad AR, Tewari A, et al. An integrated nomogram combining deep learning, prostate imaging-reporting and data system (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: a retrospective multicentre study. *Lancet Digit Health.* 2021;3(7):e445–54.
32. Shao L, Liu Z, Yan Y, Liu J, Ye X, Xia H, et al. Patient-level prediction of multi-classification task at prostate MRI based on end-to-end framework learning from diagnostic logic of radiologists. *IEEE Trans Biomed Eng.* 2021;68(12):3690–700.
33. Shao L, Yan Y, Liu Z, Ye X, Xia H, Zhu X, et al. Radiologist-like artificial intelligence for grade group prediction of radical prostatectomy for reducing upgrading and downgrading from biopsy. *Theranostics.* 2020;10(22):10200–12.
34. Gong L, Xu M, Fang M, He B, Li H, Fang X, et al. The potential of prostate gland radiomic features in identifying the gleason score. *Comput Biol Med.* 2022;144:105318.
35. Gong L, Xu M, Fang M, Zou J, Yang S, Yu X, et al. Noninvasive prediction of high-grade prostate cancer via biparametric MRI radiomics. *J Magn Reson Imaging.* 2020;52(4):1102–9.
36. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. *Clin Cancer Res.* 2019;25(12):3538–47.
37. Liu Z, Meng X, Zhang H, Li Z, Liu J, Sun K, et al. Predicting distant metastasis and chemotherapy benefit in locally advanced rectal cancer. *Nat Commun.* 2020;11(1):4308.
38. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics.* 2019;9(5):1303–22.
39. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol.* 2022;19(2):132–46.
40. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18(8):500–10.
41. Li W, Zhang L, Tian C, Song H, Fang M, Hu C, et al. Prognostic value of computed tomography radiomics features in patients with gastric cancer following curative resection. *Eur Radiol.* 2019;29(6):3079–89.
42. Wu Q, Wang S, Zhang S, Wang M, Ding Y, Fang J, et al. Development of a deep learning model to identify lymph node metastasis on magnetic resonance imaging in patients with cervical cancer. *JAMA Netw Open.* 2020;3(7):e2011625.
43. Zhang L, Dong D, Zhong L, Li C, Hu C, Yang X, et al. Multi-focus network to decode imaging phenotype for overall survival prediction of gastric cancer patients. *IEEE J Biomed Health Inform.* 2021;25(10):3933–42.
44. Chen C, Cao Y, Li W, Liu Z, Liu P, Tian X, et al. The pathological risk score: a new deep learning-based signature for predicting survival in cervical cancer. *Cancer Med.* 2022;12(2):1051–63.

45. Zhou X, Liu Z, Du Y, Xiong Q, Wang K, Tian J. Abstract P1-10-29: radiomics improved pre-therapeutic prediction of breast cancers insensitive to neoadjuvant chemotherapy. *Cancer Res.* 2020;80(4Supplement):P1–10.
46. Sun C, Tian X, Liu Z, Li W, Li P, Chen J, et al. Radiomic analysis for pretreatment prediction of response to neoadjuvant chemotherapy in locally advanced cervical cancer: a multicentre study. *EBioMedicine.* 2019;46:160–9.
47. ProstatID. Food and Drug Administration. 2021; K212783. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K212783>.
48. AI-Rad Companion Prostate MR, Food. and Drug Administration. 2020; K193283. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K193283>.
49. Paige Prostate. Food and Drug Administration. 2021; DEN200080. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN200080>.
50. Thompson IM, Tangen CM, Paradelo J, Lucia MS, Miller G, Troyer D, et al. Adjuvant radiotherapy for pathologically advanced prostate cancer: a randomized clinical trial. *JAMA.* 2006;296(19):2329–35.
51. Bolla M, van Poppel H, Tombal B, Vekemans K, Da Pozzo L, de Reijke TM, et al. Postoperative radiotherapy after radical prostatectomy for highrisk prostate cancer: long-term results of a randomised controlled trial (EORTC trial 22911). *Lancet.* 2012;380(9858):2018–27.
52. Van den Broeck T, van den Bergh RCN, Arfi N, Gross T, Moris L, Briers E, et al. Prognostic value of biochemical recurrence following treatment with curative intent for prostate cancer: a systematic review. *Eur Urol.* 2019;75(6):967–87.
53. Panebianco V, Villeirs G, Weinreb JC, Turkbey BI, Margolis DJ, Richenberg J, et al. Prostate magnetic resonance imaging for local recurrence reporting (PI-RR): international consensus-based guidelines on multiparametric magnetic resonance imaging for prostate cancer recurrence after radiation therapy and radical prostatectomy. *Eur Urol Oncol.* 2021;4(6):868–76.
54. Schelb P, Kohl S, Radtke JP, Wiesenfarth M, Kickingereder P, Bickelhaupt S, et al. Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment. *Radiology.* 2019;293(3):607–17.
55. Gutiérrez Y, Arevalo J, Martínez F. An inception-based deep multiparametric net to classify clinical significance MRI regions of prostate cancer. *Phys Med Biol.* 2022;67(22):225004.
56. Bao J, Hou Y, Qin L, Zhi R, Wang XM, Shi HB, et al. High-throughput precision MRI assessment with integrated stack-ensemble deep learning can enhance the preoperative prediction of prostate cancer gleason grade. *Br J Cancer.* 2023;128(7):1267–77.
57. Yan Y, Shao L, Liu Z, He W, Yang G, Liu J, et al. Deep learning with quantitative features of magnetic resonance images to predict biochemical recurrence of radical prostatectomy: a multi-center study. *Cancers (Basel).* 2021;13(12):3098.
58. Hou Y, Zhang YH, Bao J, Bao ML, Yang G, Shi HB, et al. Artificial intelligence is a promising prospect for the detection of prostate cancer extracapsular extension with mpMRI: a two-center comparative study. *Eur J Nucl Med Mol Imaging.* 2021;48(12):3805–16.
59. Penzkofer T, Padhani AR, Turkbey B, Haider MA, Huisman H, Walz J, et al. ESUR/ESUI position paper: developing artificial intelligence for precision diagnosis of prostate cancer using magnetic resonance imaging. *Eur Radiol.* 2021;31(12):9567–78.
60. Suarez-Ibarrola R, Sigle A, Eklund M, Eberli D, Miernik A, Benndorf M, et al. Artificial intelligence in magnetic resonance imaging-based prostate cancer diagnosis: where do we stand in 2021? *Eur Urol Focus.* 2022;8(2):409–17.
61. Chaddad A, Kucharczyk MJ, Cheddad A, Clarke SE, Hassan L, Ding S, et al. Magnetic resonance imaging based radiomic models of prostate cancer: a narrative review. *Cancers (Basel).* 2021;13(3):552.
62. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770–8.
63. Phan H, Liu Z, Huynh D, Savvides M, Cheng KT, Shen Z. Binarizing MobileNet via evolution-based searching. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020. p. 13417–26.
64. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018. p. 6848–56.
65. Winkel DJ, Breit HC, Shi B, Boll DT, Seifert HH, Wetterauer C. Predicting clinically significant prostate cancer from quantitative image features including compressed sensing radial MRI of prostate perfusion using machine learning: comparison with PI-RADS v2 assessment scores. *Quant Imaging Med Surg.* 2020;10(4):808–23.
66. Dinh AH, Melodelima C, Souchon R, Moldovan PC, Bratan F, Pagnoux G, et al. Characterization of prostate cancer with gleason score of at least 7 by using quantitative multiparametric MR imaging: validation of a computer-aided diagnosis system in patients referred for prostate biopsy. *Radiology.* 2018;287(2):525–33.
67. Netzer N, Weißer C, Schelb P, Wang X, Qin X, Gortz M, et al. Fully automatic deep learning in bi-institutional prostate magnetic resonance imaging: effects of cohort size and heterogeneity. *Invest Radiol.* 2021;56(12):799–808.
68. Zhong X, Cao R, Shakeri S, Scalzo F, Lee Y, Enzmann DR, et al. Deep transfer learning-based prostate cancer classification using 3 Tesla multi-parametric MRI. *Abdom Radiol (NY).* 2019;44(6):2030–9.
69. Deniffel D, Abraham N, Namdar K, Dong X, Salinas E, Milot L, et al. Using decision curve analysis to benchmark performance of a magnetic resonance imaging-based deep learning model for prostate cancer risk assessment. *Eur Radiol.* 2020;30(12):6867–76.
70. Liu Y, Zheng H, Liang Z, Miao Q, Brisbane WG, Marks LS, et al. Textured-based deep learning in prostate cancer classification with 3T multiparametric MRI: comparison with PI-RADS-based classification. *Diagnostics (Basel).* 2021;11(10):1175.
71. Zhao L, Bao J, Qiao X, Jin P, Ji Y, Li Z, et al. Predicting clinically significant prostate cancer with a deep learning approach: a multicentre retrospective study. *Eur J Nucl Med Mol Imaging.* 2023;50(3):727–41.
72. Youn SY, Choi MH, Kim DH, Lee YJ, Huisman H, Johnson E, et al. Detection and PI-RADS classification of focal lesions in prostate MRI: performance comparison between a deep learning-based algorithm (DLA) and radiologists with various levels of experience. *Eur J Radiol.* 2021;142:109894.
73. Yu R, Jiang KW, Bao J, Hou Y, Yi Y, Wu D, et al. PI-RADS(AI): introducing a new human-in-the-loop AI model for prostate cancer diagnosis based on MRI. *Br J Cancer.* 2023;128(6):1019–29.
74. Hectors SJ, Chen C, Chen J, Wang J, Gordon S, Yu M, et al. Magnetic resonance imaging radiomics-based machine learning

- prediction of clinically significant prostate cancer in equivocal PI-RADS 3 lesions. *J Magn Reson Imaging*. 2021;54(5):1466–73.
75. Hou Y, Bao ML, Wu CJ, Zhang J, Zhang YD, Shi HB. A radiomics machine learning-based redefining score robustly identifies clinically significant prostate cancer in equivocal PI-RADS score 3 lesions. *Abdom Radiol (NY)*. 2020;45(12):4223–34.
  76. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol*. 2017;27(10):4082–90.
  77. 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. *Front Oncol*. 2021;10:631831.
  78. Song Y, Zhang YD, Yan X, Liu H, Zhou M, Hu B, et al. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. *J Magn Reson Imaging*. 2018;48(6):1570–7.
  79. Aussavavirojekul P, Hoonlor A, Srinualnad S. Optimization of clinical risk factor interpretation and radiological findings with machine learning for PIRADS category 3 patients. *Prostate*. 2022;82(2):235–44.
  80. Kan Y, Zhang Q, Hao J, Wang W, Zhuang J, Gao J, et al. Clinicoradiological characteristic-based machine learning in reducing unnecessary prostate biopsies of PI-RADS 3 lesions with dual validation. *Eur Radiol*. 2020;30(11):6274–84.
  81. Antonelli M, Johnston EW, Dikaios N, Cheung KK, Sidhu HS, Appayya MB, et al. Machine learning classifiers can predict gleason pattern 4 prostate cancer with greater accuracy than experienced radiologists. *Eur Radiol*. 2019;29(9):4754–64.
  82. Niu XK, Chen ZF, Chen L, Li J, Peng T, Li X. Clinical application of biparametric MRI texture analysis for detection and evaluation of high grade prostate cancer in zone-specific regions. *AJR Am J Roentgenol*. 2018;210(3):549–56.
  83. Algohary A, Shiradkar R, Pahwa S, Puryrsko A, Verma S, Moses D, et al. Combination of peri-tumoral and intra-tumoral radiomic features on bi-parametric MRI accurately stratifies prostate cancer risk: a multi-site study. *Cancers (Basel)*. 2020;12(8):2200.
  84. Zhang Z, Xu H, Xue Y, Li J, Ye Q. Risk stratification of prostate cancer using the combination of histogram analysis of apparent diffusion coefficient across tumor diffusion volume and clinical information: a pilot study. *J Magn Reson Imaging*. 2019;49(2):556–64.
  85. Turkbey B, Haider MA. Artificial intelligence for automated cancer detection on prostate MRI: opportunities and ongoing challenges, from the AJR special series on AI applications. *AJR Am J Roentgenol*. 2022;219(2):188–94.
  86. Mehralivand S, Yang D, Harmon SA, Xu D, Xu Z, Roth H, et al. Deep learning-based artificial intelligence for prostate cancer detection at biparametric MRI. *Abdom Radiol (NY)*. 2022;47(4):1425–34.
  87. Duran A, Dussert G, Rouviere O, Jaouen T, Jodoin PM, Lartizen C. ProstAttention-Net: a deep attention model for prostate cancer segmentation by aggressiveness in MRI scans. *Med Image Anal*. 2022;77:102347.
  88. Luo R, Zeng Q, Chen H. Artificial intelligence algorithm-based MRI for differentiation diagnosis of prostate cancer. *Comput Math Methods Med*. 2022;2022:8123643.
  89. Saha A, Hosseinzadeh M, Huisman H. End-to-end prostate cancer detection in bpMRI via 3D CNNs: effects of attention mechanisms, clinical priori and decoupled false positive reduction. *Med Image Anal*. 2021;73:102155.
  90. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. *Breast Cancer Res Treat*. 2018;169(2):217–29.
  91. Romeo V, Cuocolo R, Apolito R, Stanzione A, Ventimiglia A, Vitale A, et al. Clinical value of radiomics and machine learning in breast ultrasound: a multicenter study for differential diagnosis of benign and malignant lesions. *Eur Radiol*. 2021;31(12):9511–9.
  92. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, et al. Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. *AJNR Am J Neuroradiol*. 2018;39(2):208–16.
  93. Suarez-Ibarrola R, Basulto-Martinez M, Heinze A, Gratzke C, Miernik A. Radiomics applications in renal tumor assessment: a comprehensive review of the literature. *Cancers (Basel)*. 2020;12(6):1387.
  94. Li Y, Ren J, Yang JJ, Cao Y, Xia C, Lee EYP, et al. MRI-derived radiomics analysis improves the noninvasive pretreatment identification of multimodality therapy candidates with early-stage cervical cancer. *Eur Radiol*. 2022;32(6):3985–95.
  95. Xu L, Zhang G, Zhao L, Mao L, Li X, Yan W, et al. Radiomics based on multiparametric magnetic resonance imaging to predict extraprostatic extension of prostate cancer. *Front Oncol*. 2020;10:940.
  96. Ma S, Xie H, Wang H, Han C, Yang J, Lin Z, et al. MRI-based radiomics signature for the preoperative prediction of extracapsular extension of prostate cancer. *J Magn Reson Imaging*. 2019;50(6):1914–25.
  97. Bai H, Xia W, Ji X, He D, Zhao X, Bao J, et al. Multiparametric magnetic resonance imaging-based peritumoral radiomics for preoperative prediction of the presence of extracapsular extension with prostate cancer. *J Magn Reson Imaging*. 2021;54(4):1222–30.
  98. Bourbonne V, Jaouen V, Nguyen TA, Tissot V, Doucet L, Hatt M, et al. Development of a radiomic-based model predicting lymph node involvement in prostate cancer patients. *Cancers (Basel)*. 2021;13(22):5672.
  99. Hou Y, Bao ML, Wu CJ, Zhang J, Zhang YD, Shi HB. A machine learning-assisted decision-support model to better identify patients with prostate cancer requiring an extended pelvic lymph node dissection. *BJU Int*. 2019;124(6):972–83.
  100. Hou Y, Bao J, Song Y, Bao ML, Jiang KW, Zhang J, et al. Integration of clinicopathologic identification and deep transferrable image feature representation improves predictions of lymph node metastasis in prostate cancer. *EBioMedicine*. 2021;68:103395.
  101. Li L, Shiradkar R, Leo P, Algohary A, Fu P, Tirumani SH, et al. A novel imaging based nomogram for predicting post-surgical biochemical recurrence and adverse pathology of prostate cancer from pre-operative bi-parametric MRI. *EBioMedicine*. 2021;63:103163.
  102. Bourbonne V, Fournier G, Vallières M, Lucia F, Doucet L, Tissot V, et al. External validation of an MRI-derived radiomics model to predict biochemical recurrence after surgery for high-risk prostate cancer. *Cancers (Basel)*. 2020;12(4):814.
  103. Sushentsev N, Rundo L, Blyuss O, Nazarenko T, Suvorov A, Gnanapragasam VJ, et al. Comparative performance of MRI-derived PRECISE scores and delta-radiomics models for the prediction of prostate cancer progression in patients on active surveillance. *Eur Radiol*. 2022;32(1):680–9.
  104. Xie J, Li B, Min X, Zhang P, Fan C, Li Q, et al. Prediction of pathological upgrading at radical prostatectomy in prostate

- cancer eligible for active surveillance: a texture features and machine learning-based analysis of apparent diffusion coefficient maps. *Front Oncol.* 2021;10:604266.
105. Zhang GMY, Han YQ, Wei JW, Qi YF, Gu DS, Lei J, *et al.* Radiomics based on MRI as a biomarker to guide therapy by predicting upgrading of prostate cancer from biopsy to radical prostatectomy. *J Magn Reson Imaging.* 2020;52(4):1239–48.
  106. Wu CJ, Zhang YD, Bao ML, Li H, Wang XN, Liu XS, *et al.* Diffusion kurtosis imaging helps to predict upgrading in biopsy-proven prostate cancer with a gleason score of 6. *AJR Am J Roentgenol.* 2017;209(5):1081–7.
  107. Zheng H, Miao Q, Liu Y, Raman SS, Scalzo F, Sung K. Integrative machine learning prediction of prostate biopsy results from negative multiparametric MRI. *J Magn Reson Imaging.* 2021;55(1):100–10.
  108. Wang J, Wu CJ, Bao ML, Zhang J, Shi HB, Zhang YD. Using support vector machine analysis to assess PartinMR: a new prediction model for organ-confined prostate cancer. *J Magn Reson Imaging.* 2018;48(2):499–506.
  109. Zhong QZ, Long LH, Liu A, Li CM, Xiu X, Hou XY, *et al.* Radiomics of multiparametric MRI to predict biochemical recurrence of localized prostate cancer after radiation therapy. *Front Oncol.* 2020;10:731.
  110. Bourbonne V, Vallieres M, Lucia F, Doucet L, Visvikis D, Tissot V, *et al.* MRI-derived radiomics to guide post-operative management for high-risk prostate cancer. *Front Oncol.* 2019;9:807.
  111. Wu C, Zheng H, Li J, Zhang Y, Duan S, Li Y, *et al.* MRI-based radiomics signature and clinical factor for predicting H3K27M mutation in pediatric high-grade gliomas located in the midline of the brain. *Eur Radiol.* 2022;32(3):1813–22.
  112. Bao D, Zhao Y, Li L, Lin M, Zhu Z, Yuan M, *et al.* A MRI-based radiomics model predicting radiation-induced temporal lobe injury in nasopharyngeal carcinoma. *Eur Radiol.* 2022;32(10):6910–21.
  113. Liu X, Zhang D, Liu Z, Li Z, Xie P, Sun K, *et al.* Deep learning radiomicsbased prediction of distant metastasis in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy: a multicentre study. *EBioMedicine.* 2021;69:103442.
  114. Bedrikovetski S, Dudi-Venkata NN, Maicas G, Kroon HM, Seow W, Carneiro G, *et al.* Artificial intelligence for the diagnosis of lymph node metastases in patients with abdominopelvic malignancy: a systematic review and meta-analysis. *Artif Intell Med.* 2021;113:102022.
  115. Cui Y, Zhang J, Li Z, Wei K, Lei Y, Ren J, *et al.* A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: a multicenter cohort study. *EClinicalMedicine.* 2022;46:101348.
  116. Yao X, Sun C, Xiong F, Zhang X, Cheng J, Wang C, *et al.* Radiomic signature-based nomogram to predict disease-free survival in stage II and III colon cancer. *Eur J Radiol.* 2020;131:109205.
  117. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, *et al.* Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J.* 2019;53(3):1800986.
  118. Yang L, Yang J, Zhou X, Huang L, Zhao W, Wang T, *et al.* Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur Radiol.* 2019;29(5):2196–206.
  119. Hu T, Wang S, Huang L, Wang J, Shi D, Li Y, *et al.* A clinical-radiomics nomogram for the preoperative prediction of lung metastasis in colorectal cancer patients with indeterminate pulmonary nodules. *Eur Radiol.* 2019;29(1):439–49.
  120. Song J, Hu Q, Ma Z, Zhao M, Chen T, Shi H. Feasibility of T2WI-MRI-based radiomics nomogram for predicting normal-sized pelvic lymph node metastasis in cervical cancer patients. *Eur Radiol.* 2021;31(9):6938–48.
  121. Zhou Y, Gu HL, Zhang XL, Tian ZF, Xu XQ, Tang WW. Multiparametric magnetic resonance imaging-derived radiomics for the prediction of disease-free survival in early-stage squamous cervical cancer. *Eur Radiol.* 2022;32(4):2540–51.
  122. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749–62.
  123. Chen Q, Zhang L, Mo X, You J, Chen L, Fang J, *et al.* Current status and quality of radiomic studies for predicting immunotherapy response and outcome in patients with non-small cell lung cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* 2021;49(1):345–60.
  124. Zhong J, Hu Y, Si L, Jia G, Xing Y, Zhang H, *et al.* A systematic review of radiomics in osteosarcoma: utilizing radiomics quality score as a tool promoting clinical translation. *Eur Radiol.* 2021;31(3):1526–35.
  125. Chang S, Han K, Suh YJ, Choi BW. Quality of science and reporting for radiomics in cardiac magnetic resonance imaging studies: a systematic review. *Eur Radiol.* 2022;32(7):4361–73.

<https://doi.org/10.1186/s40779-023-00464-w>

**Cite this article as:** Zhao LT, Liu ZY, Xie WF, Shao LZ, Lu J, Tian J, *et al.* What benefit can be obtained from magnetic resonance imaging diagnosis with artificial intelligence in prostate cancer compared with clinical assessments?. *Mil Med Res.* 2023;10(1):29.