THERAPEUTIC ADVANCES in
Urology

Review

A review of artificial intelligence in prostate cancer detection on imaging

Indrani Bhattacharya, Yash S. Khandwala, Sulaiman Vesal, Wei Shao, Qianye Yang, Simon J.C. Soerensen, Richard E. Fan, Pejman Ghanouni, Christian A. Kunder, James D. Brooks, Yipeng Hu, Mirabela Rusu* and Geoffrey A. Sonn*

Ther Adv Urol

2022, Vol. 14: 1–31

DOI: 10.1177/17562872221128791

***Abstract*:** A multitude of studies have explored the role of artificial intelligence (AI) in providing diagnostic support to radiologists, pathologists, and urologists in prostate cancer detection, risk-stratification, and management. This review provides a comprehensive overview of relevant literature regarding the use of AI models in (1) detecting prostate cancer on radiology images (magnetic resonance and ultrasound imaging), (2) detecting prostate cancer on histopathology images of prostate biopsy tissue, and (3) assisting in supporting tasks for prostate cancer detection (prostate gland segmentation, MRI-histopathology registration, MRI-ultrasound registration). We discuss both the potential of these AI models to assist in the clinical workflow of prostate cancer diagnosis, as well as the current limitations including variability in training data sets, algorithms, and evaluation criteria. We also discuss ongoing challenges and what is needed to bridge the gap between academic research on AI for prostate cancer and commercial solutions that improve routine clinical care.

Keywords: artificial intelligence, histopathology images, magnetic resonance imaging, prostate cancer diagnosis, registration, ultrasound images

Received: 28 January 2022; revised manuscript accepted: 30 August 2022.

Correspondence to:
Indrani Bhattacharya
Department of Radiology, Stanford University School of Medicine, 1201 Welch Road, Stanford, CA 94305, USA.

Department of Urology, Stanford University School of Medicine, Stanford, CA, USA
ibhatt@stanford.edu

Yash S. Khandwala
Sulaiman Vesal
Richard E. Fan
James D. Brooks
Department of Urology, Stanford University School of Medicine, Stanford, CA, USA

Wei Shao
Mirabela Rusu
Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA

Qianye Yang
Yipeng Hu
Centre for Medical Image Computing, University College London, London, UK

Wellcome / EPSRC Centre for Interventional and Surgical Sciences, University College London, London, UK

Simon J.C. Soerensen
Department of Urology, Stanford University School of Medicine, Stanford, CA, USA

Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, USA

Pejman Ghanouni
Geoffrey A. Sonn
Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA

Introduction

Prostate cancer screening with prostate-specific antigen (PSA) has contributed to a >50% reduction in death from prostate cancer,[1] yet it has also resulted in a major problem of overdiagnosis and overtreatment of non-aggressive prostate cancer.[2] As a result, focus has shifted to preferential diagnosis and treatment of aggressive prostate cancers. New diagnostic tests including blood- and urine-based biomarkers, genetic tests, and improved imaging modalities have great potential to save lives while at the same time reduce the problem of overdiagnosis. However, the optimal usage of the massive amount of data generated by these new tests remains a major clinical and research challenge. Artificial intelligence (AI)-based systems will play a major role in addressing this challenge.[3,4]

AI models are computational approaches that learn patterns from existing data to enable predictions in new, unseen data. Earlier AI models used, what is often referred to as, 'traditional' machine learning approaches, which were often executed in two steps. First, domain-experts (human experts in the subject matter) carefully designed features to extract quantitative variables from the data specific to the task, for example, tumor volume or shape. Second, these hand-crafted features were fed into computational models to learn which features were useful and how to combine them to maximize accuracy in classifying data in categories (e.g. benign nodules *vs* malignant tumors). Once trained, such AI models could generate predictions in new, previously unseen data. Recent advances in the computing power of graphics processing units (GPUs) have enabled the development of deep learning models. Deep learning models alleviate the need for hand-crafted features, thereby working in a completely automated manner to both identify the features and use them for the desired downstream task. Deep

Department of Urology, Stanford University School of Medicine, Stanford, CA, USA

Christian A. Kunder
Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

*Equally contributed as senior authors.

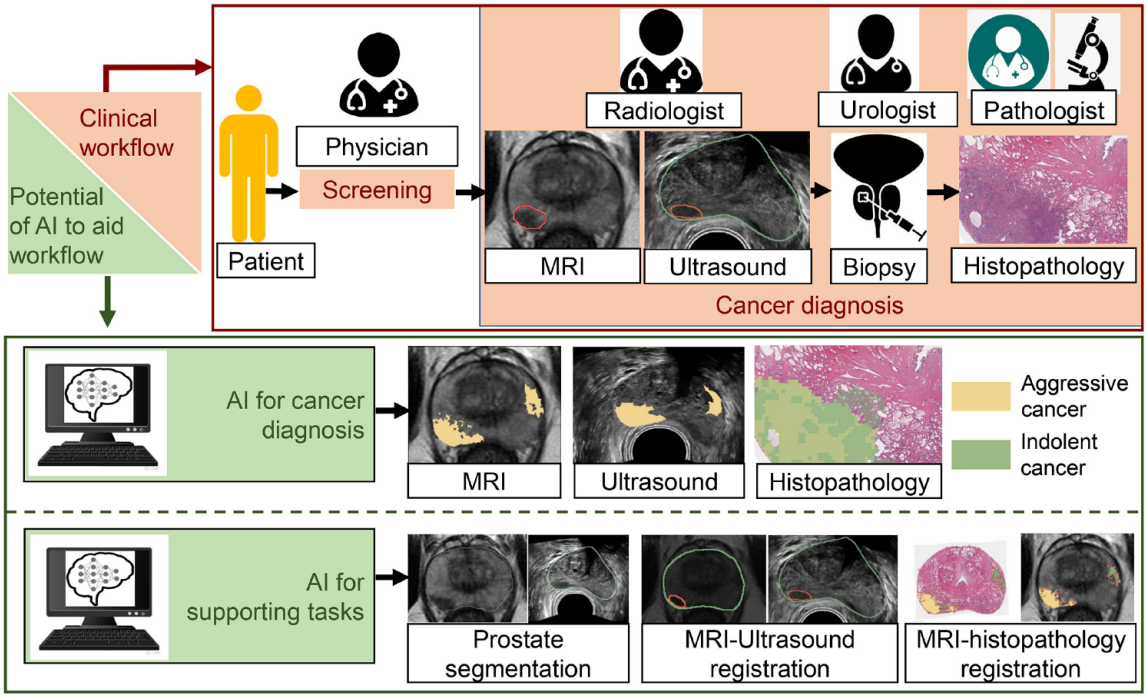

Figure 1. Potential of AI to assist prostate cancer diagnosis on imaging. AI models can help in detecting and characterizing cancer aggressiveness on non-invasive radiology images (MRI and ultrasound), as well as on histopathology images acquired through prostate biopsy. Aggressive cancer is shown in yellow, and indolent cancer in green in the 'AI for cancer diagnosis' panel. AI models can also help in supporting tasks for cancer detection, namely prostate gland segmentation, MRI-ultrasound registration, and MRI-histopathology registration.

learning models have revolutionized the field of AI through unprecedented performance that often exceeds human performance, particularly in tasks related to image analysis.

Development of medical AI models (in particular, deep learning models) that learn and predict from medical data to assist diagnosis, prognosis, and clinical decision-making for a variety of diseases is an active area of research.[5–7] Research on AI-assisted prostate cancer diagnosis is also evolving rapidly and has the potential to facilitate all aspects of the current standard diagnostic pathway (Figure 1). Although there exists a large body of research literature surrounding the use of AI in prostate cancer diagnosis, most of these methods are not yet ready for clinical deployment. Several challenges exist that impede the deployment of these widely researched AI tools for diagnostic support in the clinic.

Clinicians and AI researchers working on prostate cancer should develop a thorough understanding of this emerging interdisciplinary domain to successfully herald AI-enabled precision medicine, with a goal of revolutionizing the diagnosis and treatment of prostate cancer.

Here, we provide a systematic overview of the relevant literature involving the use of AI for prostate cancer diagnosis on medical images (Figure 1). In particular, we discuss existing AI literature for

1. Detecting prostate cancer on radiology images (magnetic resonance and ultrasound imaging).
2. Detecting prostate cancer on histopathology images.
3. Supporting tasks for prostate cancer detection. We then discuss challenges associated with implementing these AI-enabled diagnostic tools in the clinic, and possible solutions to overcome them.

Potential of AI in prostate cancer diagnosis

Medical images play an important role in prostate cancer diagnosis. For many years, this involved transrectal ultrasound alone to guide systematic

biopsy. More recently, magnetic resonance imaging (MRI) has been shown to greatly improve prostate cancer detection.[8,9] MRI-ultrasound fusion biopsies are increasingly used to target lesions outlined on MRI by radiologists. MRI-ultrasound fusion biopsies improve detection of clinically significant prostate cancer over ultrasound-guided systematic biopsies alone.[8,10–13] Finally, prostate tissue obtained through biopsy is subjected to histopathological analysis to identify the presence and grade prostate cancer. Urologists plan treatment based on the aggressiveness of prostate cancer, with a primary objective of treating aggressive cancer and reducing over-treatment of indolent cancer.

Numerous opportunities exist for optimizing this workflow (Figure 1), such as improving detection of cancer on ultrasound and MRI, reducing inter-observer variability among radiologists, and assisting pathologists in identifying and grading cancer on histopathology images. Moreover, AI models can help cancer diagnosis by facilitating supporting tasks in cancer detection that are labor- and experience-intensive, such as prostate gland segmentation, MRI-ultrasound registration, MRI-ultrasound fusion biopsies, and MRI-histopathology registration for developing cancer detection models. This review categorizes the existing studies on AI models to facilitate prostate cancer diagnosis as described below:

1. AI models for prostate cancer detection and characterization of cancer aggressiveness.
   (a) On prostate MRI.
   (b) On prostate ultrasound images.
   (c) On histopathology images collected through prostate needle biopsies.
2. AI models for supporting tasks in cancer detection.
   (a) Prostate gland segmentation on MRI and ultrasound images to facilitate MRI-ultrasound fusion biopsies.
   (b) MRI-ultrasound registration to facilitate MRI-ultrasound fusion biopsies.
   (c) MRI-histopathology registration for ground truth labeling of cancer detection models.

The following sections briefly summarize relevant AI studies in each of these areas, highlighting strengths, weaknesses, variabilities, potential, and scope for use in clinical care.

AI models for prostate cancer detection and characterizing cancer aggressiveness

AI models can help detect cancer and characterize cancer aggressiveness on three kinds of images widely used in the clinical workflow of prostate cancer diagnosis: (1) MRI, (2) ultrasound, and (3) histopathology images of prostate biopsy tissue (Figure 1).

Implications of accurate prostate cancer detection and aggressiveness characterization on imaging: accurately detecting, localizing, and characterizing lesions as aggressive or indolent using AI methods on non-invasive images (e.g. MRI and ultrasound) may significantly impact patient management and treatment planning. Non-invasive imaging can be used in conjunction with clinical variables (PSA density, race, prior biopsy history, etc) in routine clinical care to decide when biopsy is needed and how treatment is performed. For example, patients with aggressive prostate cancer accurately detected and localized on non-invasive images can be targeted with MRI-ultrasound fusion biopsies with more precision and using fewer biopsy needle samples. Patients with no cancer or with indolent cancer according to non-invasive images could safely avoid biopsy, thereby minimizing the unnecessary side-effects of invasive biopsy procedures (pain, bleeding, and infection).[8] Accurate selective identification of aggressive and indolent prostate cancer on non-invasive imaging can help prioritize and enable timely treatment planning for aggressive prostate cancer patients. Location and extent of aggressive cancer on non-invasive imaging can also help guide treatment decisions, that is, whether to perform radical prostatectomy or focal therapy or active surveillance.

Accurate automated cancer aggressiveness grading on prostate histopathology images acquired through invasive biopsy procedures can help alleviate inter- and intra-pathologist variability in Gleason grading, and also significantly reduce time required from pathologists. Such standardization of pathologist interpretations and time savings will eventually facilitate disease management.

Cancer detection on prostate MRI

MRI is increasingly used to detect prostate cancer, guide MRI-ultrasound fusion biopsies, and plan treatment.[14] Currently, it is considered to be the most sensitive non-invasive imaging modality

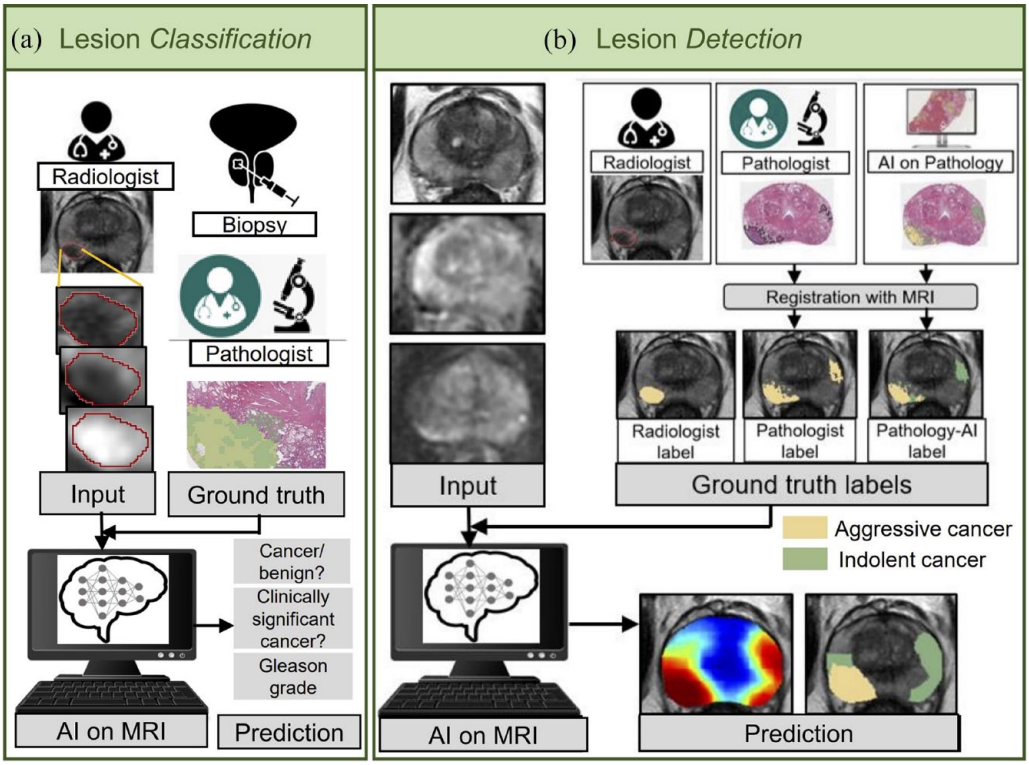

Figure 2. AI models for prostate cancer detection on MRI can be subdivided into two major tasks: lesion *classification* and lesion *detection*. Lesion classification involves classifying a radiologist-outlined lesion (region of interest) into categories (cancer *vs* benign, clinically significant cancer *vs* benign or indolent, or Gleason grade groups). Lesion *detection* involves detecting and characterizing cancer aggressiveness on the entire prostate MRI.

that enables visualization, detection, and localization of prostate cancer. However, the often subtle visual differences between benign and cancerous tissue on MRI make radiologist interpretations of MR images challenging. Despite adoption of PIRADS (Prostate Imaging-Reporting and Data System),[15] problems remain with false negatives (12% of aggressive cancers missed during screening,[8] 34% of aggressive and 81% of indolent cancers missed in men undergoing prostatectomy[10]), false positives (>35% false-positive rate[8]), and high inter-reader variability (inter-reader agreement $\kappa = 0.46-0.78$ [inter-reader agreement $\kappa = 0.46-0.78$[16,17]]. As a result, many unnecessary biopsies continue to be performed. Moreover, MRI-ultrasound fusion-targeted biopsies are usually supplemented with systematic biopsies, leading to increased risks (infection, bleeding, and pain), as well as over-detection and over-treatment of indolent cancers.

Detecting cancer and simultaneously characterizing cancers as aggressive or indolent on MRI is an unmet clinical need. Such selective identification of aggressive and indolent cancer on MRI could help identify men with aggressive prostate cancer, and reduce unnecessary biopsies in men without cancer or with indolent prostate cancer. Several studies have investigated the use of AI for prostate cancer detection on MRI with encouraging performance. A body of literature exists surrounding the use of AI to predict the likelihood of a patient having prostate cancer without explicitly detecting lesions,[18] or the likelihood of biochemical recurrence in a patient after radical prostatectomy.[19] These studies use MR images, with or without clinical variables like age, PSA-density, PIRADS scores, etc. In this review, we only focus on methods for detecting cancer in patients without known cancer and further subdivide AI models into two major tasks (Figure 2):

- Lesion *classification*: This group of AI models classify radiologist-outlined lesions (regions of interest) into categories (i.e. cancer or benign, clinically significant cancer or benign, or different Gleason grade groups) (Figure 2(a)); Table 1). AI models for lesion

Table 1. AI models for prostate lesion *classification* on MRI.

Study	Input data	Cohort size	Data type	Algorithm	Training labels	Evaluation labels	Evaluation metric	Source code availability
Algohary *et al.*[20]	T2w, ADC	231	Retrospective, 4 inst.	TML	Biopsy	Biopsy	ROC-AUC, Acc.	No
Antonelli *et al.*[21]	T2w, ADC DWI, DCE	164	Retrospective, 1 inst.	TML	Biopsy	Biopsy	ROC-AUC, Se. at 50% threshold of Sp.	No
Bleker *et al.*[22]	T2w, ADC DWI, DCE	206	Retrospective, public data set	TML	Biopsy	Biopsy	ROC-AUC, Se., Sp.	No
Bonekamp *et al.*[23]	T2w, ADC DWI	316	Retrospective, public data set	TML	Biopsy	Biopsy	ROC-AUC, Se., Sp.	No
Chen *et al.*[24]	T2w, ADC	381	Retrospective, 1 inst.	TML	Biopsy	Biopsy	ROC-AUC, Acc., Se., Sp.,	No
Akamine *et al.*[25]	DWI, DCE	52	Retrospective, 1 inst.	Hierarchical clustering	RP	RP	Acc.	No
Kwon *et al.*[26]	T2w, ADC, DWI, DCE	344	Retrospective, public data set	TML	biopsy	biopsy	ROC-AUC, Se., PPV	No
Chaddad *et al.*[27]	T2w, ADC	112	Retrospective, 1 inst., public data set	TML	Biopsy	Biopsy	ROC-AUC	No
Hectors *et al.*[28]	T2w	64	Retrospective, 1 inst.	TML	RP	RP	ROC-AUC	No
Xu *et al.*[29]	T2w	331	Retrospective, 1 inst.	TML	RP	RP	ROC-AUC, decision curve analysis	No
Viswanath *et al.*[30]	T2w	85	Retrospective, 3 inst.	TML	RP	RP	ROC-AUC	No
Transin *et al.*[31]	ADC, DCE	74	Retrospective, 1 inst.	TML	Biopsy/ RP	Biopsy/RP	ROC-AUC, Se., Sp.	No
Zhang *et al.*[32]	T2w, ADC	159	Retrospective, 2 inst.	TML	Biopsy	Biopsy	ROC-AUC	No
Deniffel *et al.*[33]	T2w, ADC, DWI	499	Retrospective, 1 inst.	DL	Biopsy	Biopsy	ROC-AUC, decision-curve analysis	No
Song *et al.*[34]	T2w, ADC, DWI	185	Retrospective, public data set	DL	Biopsy	Biopsy	ROC-AUC, Se., Sp., PPV	No
Takeuchi *et al.*[35]	T2w, ADC, DWI	334	Retrospective, 1 inst.	DL	Biopsy	Biopsy	ROC-AUC, Net-benefit curve, NPV	No
Yuan *et al.*[36]	T2w, ADC	244	Retrospective, 2 inst.	DL	Biospy	Biospy	Acc., Prec., Recall, F1-score	No
Aldoj *et al.*[37]	T2w, ADC, DWI, DCE	200	Retrospective, public data set	DL	Biopsy	Biopsy	ROC-AUC, Se., Sp.	No
Zhong *et al.*[38]	T2w, ADC	140	Retrospective, 1 inst.	DL	RP	RP	ROC-AUC, Acc., Se. Sp.	No
Abraham and Nair[39]	T2w, ADC, DWI	112	Retrospective, public data set	DL	Biopsy	Biopsy	ROC-AUC, quadratic wtd. kappa, PPV	No

Acc, accuracy; ADC, apparent diffusion coefficient; AI, artificial intelligence; DCE, dynamic contrast enhanced; DL, deep learning; DWI, diffusion weighted imaging; inst., institution; MRI, magnetic resonance imaging; NPV, negative predictive value; PPV, positive predictive value; Prec, precision; ROC-AUC, receiver operating characteristics – area under the curve; RP, radical prostatectomy; Se, sensitivity; Sp, specificity; T2w, T2-weighted MRI; TML, traditional machine learning.

classification often use traditional machine learning, which involves extracting hand-crafted features from the region of interest, and then using a classifier to attempt to classify what category that lesion falls into. Hand-crafted features assess texture, shape, volume, or image-based radiomic features. Some of the traditional machine learning classifiers include artificial neural networks, random forests, support vector machines, or logistic regression-based classifiers. With the increasing success of deep learning–based methods, several lesion *classification* methods were also developed that can classify lesions using deep neural networks, without the need to select and extract hand-crafted features. If successfully deployed in the clinic, automated lesion *classification* methods would allow a physician to select a region of interest on an MRI slice and receive AI assistance in determining whether that region is likely to be cancerous.

- Lesion *detection*: This group of AI models use all the images from a prostate MR exam as inputs, and detect, localize, and/or stratify cancer aggressiveness on the entire prostate MRI (Figure 2(b); Table 2). Often, these lesion *detection* methods provide a

Table 2. AI models for prostate lesion *detection* on MRI.

Study	Input data	Cohort size	Data type	Algorithm	Training labels	Evaluation labels	Evaluation granularity	Evaluation metric	Source code availability
Saha *et al.*[40]	T2w, ADC DWI	2732	Retrospective, 2 inst., PIRADS or biopsy	DL	Radiologist, w/o path. confirm.	Radiologist, w/o & with path. confirm. from biopsy	Lesion-level, patient-level	ROC, FROC	Yes
Yu *et al.*[41]	T2w, ADC DWI	1745	Retrospective, 4 inst., PIRADS or biopsy external validation on public data set	DL	Radiologist, w/o path. confirm.	Radiologist, w/o & with path. confirm. from biopsy	Lesion-level, patient-level	FROC, DSC, ROC-AUC	No
Schelb *et al.*[42]	T2w, DWI	312	Retrospective, 1 inst., biopsy	DL	Radiologist, path confirm. from biopsy	Radiologist, path confirm. from biopsy	Sextant-level, patient-level	Se, Sp, Prec, NPV, ROC	Yes
Sumathipala *et al.*[43]	T2w, ADC DWI	186	Retrospective, 6 inst., RP or biopsy	DL	Radiologist, path. confirm. from RP or biopsy	Radiologist, path confirm. from RP or biopsy	Patient-level	ROC-AUC	No
Bhattacharya *et al.*[44]	T2w, ADC	75	Retrospective, 1 inst., RP	DL	Pathologist, automated registration	Pathologist, automated registration	Pixel-level, lesion-level	ROC-AUC, Se, Sp	No
Sanyal *et al.*[45]	T2w, ADC DWI	77	Retrospective, 1 inst., biopsy	DL	Radiologist, path confirm. from biopsy	Radiologist, path confirm. from biopsy	Pixel-level	ROC-AUC	Yes
Jin *et al.*[46]	T2w, ADC, DWI, DCE	34	Retrospective, 1 inst.	TML	Pathologist, automated registration	Pathologist, automated registration	Pixel-level	ROC-AUC	Yes
McGarry *et al.*[47]	T2w, ADC, DWI, DCE	48	Prospecitvely recruited, 1 inst., RP	TML	Pathologist, automated registration	Pathologist, automated registration	Lesion-level	ROC-AUC	No
Cao *et al.*[48]	T2w, ADC	417	Retrospective, 1 inst., 4 scanners, RP	DL	Radiologist, path confirm., cognitive registration with RP	Radiologist, path confirm., cognitive registration with RP	Lesion-level	FROC	No
De Vente *et al.*[49]	T2w, ADC	162	Retrospective, 1 inst., public data set, biopsy	DL	Semi-automated region growing from targeted biopsy centroid	Semi-automated region growing from targeted biopsy centroid	Pixel-level, lesion-level	Quadratic weighted kappa-score	No

(Continued)

Table 2. (Continued)

Study	Input data	Cohort size	Data type	Algorithm	Training labels	Evaluation labels	Evaluation granularity	Evaluation metric	Source code availability
Seetharaman *et al.*[50]	T2w, ADC	424	Retrospective, 1 inst., biopsy & RP	DL	Automated Gleason patterns from RP, automated registration	Automated Gleason patterns from RP, radiologist labels with path confirm. from targeted biopsy	Pixel-level, lesion-level, Patient-level	ROC-AUC, Se, Sp	Yes
Bhattacharya *et al.*[51]	T2w, ADC	443	retrospective, 1 inst., biopsy & RP	DL	Pathologist & automated Gleason patterns from RP, automated registration	Automated Gleason patterns from RP, radiologist labels with path confirm. from targeted biopsy	Pixel-level, lesion-level, patient-level	ROC-AUC, PR-AUC, Se, Sp, Prec, NPV, F1-score, DSC, Acc	Soon to be released
Zhang *et al.*[52]	T2w, ADC	358	Retrospective, 1 inst., biopsy	DL	Retrospective radiologist outline from biopsy path.	Retrospective radiologist outline from biopsy path.	Pixel-level	DSC, Se, Prec, VOE, RVD	No
Alkadi *et al.*[53]	T2w	19	retrospective, 1 inst. (public), biopsy	DL	Radiologist, path confirm. from biopsy	Radiologist, path confirm. from biopsy	Pixel-level	Acc, IoU, Recall, DSC	No
Arif *et al.*[54]	T2w, DWI, ADC	292	Retrospective, 1 inst., biopsy	DL	Radiologist, path confirm. from biopsy	Radiologist, path confirm. from biopsy	Patient-level	Acc, IoU, Recall, DSC	No
Mehralivand *et al.*[55]	T2w, DWI, ADC	236	Retrospective, multi inst., biopsy	TML	Radiologist, path confirm. from biopsy or RP	Radiologist, path confirm. from biopsy or RP	Lesion-level	AUC, Se, PPV	No
Netzer *et al.*[56]	T2w, DWI	1488	Retrospective, multi inst., multi scanner biopsy	DL	Radiologist, path confirm. from biopsy or RP	Radiologist, path confirm. from biopsy or RP	Patient-level, sextant-level	ROC-AUC, Se, Se	No
Duran *et al.*[57]	T2w, ADC	318	Retrospective, 2 inst., different scanners, external validation on public data set	DL	Radiologist, cognitive alignment with RP	Radiologist, cognitive alignment with RP	Lesion-level	FROC, Cohen's Quadratic kappa	Yes (claimed)

Acc, Accuracy; ADC, apparent diffusion coefficient; AI, artificial intelligence; confim., confirmation; DCE, dynamic contrast enhanced; DL, deep learning; DSC, dice coefficient; DWI, diffusion-weighted imaging; FROC, free-response receiver operating characteristics; inst., institution; IoU, Intersection over Union; MRI, magnetic resonance imaging; NPV, negative predictive value; path., pathology; PIRADS, Prostate Imaging-Reporting and Data System; PPV, positive predictive value; PR-AUC, precision recall–area under the curve; Prec, Precision; ROC-AUC, receiver operating characteristics–area under the curve; RP, radical prostatectomy; RVD, relative volume difference; Se, Sensitivity; Sp, Specificity; T2w, T2-weighted MRI; TML, traditional machine learning; VOE, volumetric overlap error; w/o, without.

pixel-level probability of cancer distribution for the prostate, highlighting areas of the prostate which are highly suspicious for cancer. While earlier lesion *detection* methods used traditional machine learning,[58] recent studies almost always use deep learning-based models. If successfully deployed in the clinic, automated lesion *detection* methods would automatically evaluate an entire MRI exam and provide a physician with outlines of all areas that are suspicious for cancer.

Existing AI studies for prostate cancer detection on MRI (both lesion *classification* and lesion *detection*) have used a variety of traditional machine learning and deep learning approaches. These AI models also vary greatly in the following ways:

(a) Ground truth labels used for training and evaluation (biopsy or radical prostatectomy, radiologist outlines with or without pathology confirmation, pathologist outlines etc.).
(b) Evaluation criteria (patient-level, lesion-level, or pixel-level evaluation, evaluation metrics, etc.).
(c) Data set size and type (cohort size, input MRI sequences, data from single or multiple institutions, and retrospective or prospective data etc.).

Unfortunately, direct comparison across the many published AI models for prostate cancer diagnosis on MRI is not possible due to (1) the wide variability in labels, evaluation criteria, and data for trained models, (2) the lack of access to published models and source code for pre- and post-processing and training, and (3) the lack of large publicly available multi-institution MR imaging data for independent model testing.

The following sections summarize the methodologic variability that limits direct comparison across models:

(a) Ground truth labels for training and evaluation: AI models for prostate lesion *classification* mostly use radiologist outlines as inputs, and pathology confirmation from prostate biopsy or surgery as ground truth. However, AI models for prostate lesion *detection* differ widely in the ground truth labels used for training and evaluation. These approaches for ground truth labeling include:
   1. Radiologist outlines of PI-RADS 3 or above lesions with[43,45,48] or without[40,41,59] pathology confirmation.
   2. Pathologist outlines of cancer (without grade information) on whole-mount histopathology images mapped onto pre-operative MRI using MRI-histopathology registration approaches.[44,47]
   3. Automated Gleason pattern labels on whole-mount histopathology images from deep learning algorithms,[60] mapped onto MRI through automated MRI-histopathology registration.[50,51]

While the first ground truth labeling approach trains models to perform PIRADS scoring like a radiologist, the latter two approaches allow for detection of cancers that may not have been seen by a radiologist. Training a model using radiologist outlines without any pathology confirmation may lead to high rates of false-positive findings. Conversely, obtaining pathology confirmation should reduce false-positives because it enables training using only cancerous areas. Pathology-confirmation may include either targeted biopsy results from radiologist-outlined lesions,[45] or post-operative whole-mount histopathology images from radical prostatectomy patients through cognitive registration or manual matching.[43,48] For the latter two approaches of ground truth labeling, MRI-histopathology registration approaches are used to map labels from whole-mount histopathology images onto pre-operative MRI (see section "MRI-histopathology registration for ground truth labeling of cancer detection models").

All label types used to train AI models for prostate cancer detection have advantages and disadvantages. Radiologist outlines without pathology confirmation are easier to obtain in large numbers from routine clinical care (and arguably more feasible to predict), but they include many false positives,[61,62] routinely underestimate tumor extent,[63] and may miss cancers completely (up to 34% of aggressive cancers in men undergoing radical prostatectomy are missed on MRI).[8,10] Unlike radiologist annotations, pathologist outlines on whole-mount histopathology images capture the complete extent of cancer, but mapping pathologist outlines onto pre-operative MRI requires accurate MRI-histopathology registration approaches. These MRI-histopathology registrations are labor- and experience-intensive (see section "MRI-histopathology registration for ground truth labeling of cancer detection models"). Moreover, it is impossible for pathologists to annotate large data sets of whole-mount histopathology images with gland-level annotations of cancer and Gleason pattern to train machine learning models on prostate MRI. Automated Gleason pattern labels derived from deep learning on pathology images[60] have recently shown[64] to perform with similar accuracy to experienced pathologist labels, while circumventing the constraints of labor, time, and variability associated with human-annotated labels. Moreover, when cancer detection models for MRI are trained using automated pixel-level Gleason pattern labels on whole-mount histopathology images, they can selectively identify aggressive and indolent cancer components, even in mixed lesions; this is intractable with any human-annotated labels on MRI.[50,51,64]

(b) Data type and size: The input data to train AI models typically consists of one or more MR sequences [T2w-MRI, apparent diffusion coefficient (ADC) Maps, diffusion-weighted images (DWI), dynamic contrast-enhanced (DCE) sequences]. While most studies used MR images or features derived from MR images as inputs, two recent studies, CorrSigNet[44] and CorrSigNIA,[51] presented a radiology-pathology fusion approach to identify MRI features correlated to pathology features of

cancer, and used these correlated MRI features to detect and localize aggressive and indolent cancer on MRI. These studies[44,51] showed that algorithms leveraging radiology-pathology fusion to identify pathology features on non-invasive imaging performed better than algorithms that used MR-derived features alone. Recent studies[40,59,65,66] also show that adding prior knowledge about cancer distribution in the different prostate zones to the AI models improves cancer detection and localization.

The size of data sets used to train and validate AI models also varies significantly, ranging from as low as 19 patients[53] to 2732 patients[40] (Tables 1 and 2). Most studies used retrospective data, either from public data sets,[22,23,26,27,34,37,39,53] single institutions,[21,31,33,44,50,51] or multiple institutions.[20,40,43] In general, studies using large data sets tend to have lower quality labels, while those with smaller data sets tend to have high-quality labels. For example, two recent studies[40,59] used ≈2000 patients and radiologist labels without pathology confirmation to develop and validate their methods. One of these two studies[59] also showed that a large data set of radiologist labels without pathology confirmation could be used to successfully train an AI model to detect clinically significant cancer on prostate MRI.

Studies have also used patient populations with different distributions of the disease to train and validate AI models. While some studies used patients with aggressive prostate cancer that underwent radical prostatectomy,[43,48] others used patients from a population undergoing MRI-based screening who had varying distributions of cancer or no cancer.[23,40,45] To test generalizability, some studies trained AI models using one group of patients, and tested on a different group, including patients with different disease distributions[50,51] or different label types.[40,59] Due to the difficulty in acquiring large data sets of pathology-confirmed cancer labels to train AI models, a study[67] proposed a weakly supervised learning approach to alternatively learn the normal appearance of prostate MRIs using 1145 negative MRI scans, and then use this baseline model to predict pixel-wise suspicion of prostate cancer. Another recent study[68] proposed a self-supervised learning approach where the AI model first learns prostate MRI features from unlabeled data, and then further fine-tunes the learned models to detect cancer using limited labeled data.

(c) Evaluation criteria: Evaluation methods and metrics vary based on the task (lesion *classification*, lesion *detection*), as well as the granularity of the available labels. Evaluation of cancer detection models can be on a patient-level (whether the AI model correctly detects a person as having prostate cancer or not), lesion-level (whether the AI model correctly detects/classifies individual lesions, while not incorrectly predicting false positives), or pixel level (whether the AI method correctly classifies all the prostate MRI pixels into benign, cancer, or cancer aggressiveness subtypes). The definition of the evaluation metrics also differ across studies, for example, Sumathipala *et al.*[43] used a lesion-level evaluation where the negative class was defined using 3×3 voxels, while other studies[42,50,51] used different sextant-based lesion-level evaluations in line with how prostate biopsies are conducted in the clinic. Several evaluation metrics have been used in existing studies, including but not limited to, the area under the receiver operating characteristics curve (ROC-AUC), area under the precision-recall curve (PR-AUC), free-response receiver operating characteristics (FROC), sensitivity (Se), specificity (Sp), F1-score, accuracy, positive predictive value (PPV), negative predictive value (NPV), and dice coefficient. Such wide variability in evaluation methods and metrics raises the need for definition of a set of clinically relevant standardized evaluation criteria which can be used to uniformly validate and compare all AI models for cancer detection.

Summary. AI models for prostate cancer detection on MRI show great promise, but they are not ready for clinical deployment. There remains a wide variability in methods, labels, and evaluation criteria among these AI models, limiting comparison. Most of these AI models have been developed and validated on single institution, retrospective and small patient data sets, and lack tests to assess generalizability in larger, heterogeneous patient data. In order to reap the benefits of these AI models in clinical care, there remains the need for developing publicly available anonymized large patient data sets, publicly available source code and trained models, standardized evaluation criteria, external validation, multi-reader studies to assess performance of AI models, and prospective

trials (see Section "Challenges in AI for PCa" for more details).

Cancer detection on prostate ultrasound images

Prostate cancer is most commonly diagnosed using grayscale transrectal ultrasound-guided biopsy.[69,70] While grayscale ultrasound accurately identifies the prostate gland, low signal-to-noise ratio and artifacts (e.g. speckle and shadowing) prevent clinicians from reliably differentiating cancerous from non-cancerous regions. The detection rate of prostate cancer on grayscale ultrasound images is reported to be as low as 40%.[71–73] When visible on ultrasound, cancers most often appear hypoechoic[74] because they reflect significantly less sound echoes than normal tissue. To supplement grayscale ultrasound, other new ultrasound-based imaging techniques such as shear-wave elastography,[75–77] color doppler ultrasound,[77] contrast-enhanced ultrasound,[78] micro-ultrasound,[79] and their combination have been proposed. These alternative ultrasound-based imaging modalities provide enhanced image resolution and better visualization of the prostate compared to grayscale ultrasound, and enable prostate cancer detection with better sensitivity compared to grayscale ultrasound.[80] In particular, high-frequency micro-ultrasound images are showing promise in detecting clinically significant prostate cancer with similar or higher sensitivity, similar specificity, and much lower cost in comparison to MRI.[81–85] As such, research on development AI models for prostate cancer detection on micro-ultrasound images is also growing.[86,87]

Although grayscale transrectal ultrasound is widely used for prostate biopsy in clinical settings, only one AI study[88] focused on prostate cancer detection on grayscale ultrasound images. Most AI models for prostate cancer detection used new

Table 3. AI models for prostate cancer detection on ultrasound.

Study	Cohort size	Input data	Data type	Algorithm	Training labels	Evaluation granularity	Evaluation metric	Source code availability	Task
Sedghi *et al.*[89]	157	TeUS	Retrospective, 1 inst.	DL	Radiologist, path confirm. from biopsy	Lesion level	Se, Sp ACC, AUC	No	Lesion *detection*
Azizi *et al.*[90]	163	TeUS	Retrospective, 2 inst.	DL	Radiologist, path confirm. from biopsy	Lesion level	Se, Sp ACC, AUC	No	Lesion *classification*
Azizi *et al.*[72]	157	TeUS	Retrospective, 1 inst.	DL	Radiologist, path confirm. from biopsy	Lesion level	Se, Sp ACC, AUC	No	Lesion *classification*
Azizi *et al.*[91]	155	TeUS	Retrospective, 1 inst.	DL	Radiologist, path confirm. from biopsy, biopsy length	Patient level	AUC, MSE	No	Lesion *classification*
Han *et al.*[92]	51	TRUS	N/A	TML	Biopsy	Patient level, Lesion level	Se, Sp ACC, ROC-AUC	No	Lesion *detection*
Wildeboer *et al.*[80]	50	TRUS, SWE, DCE-US	Retrospective, 1 inst.	TML	RP, Biopsy	Pixel level, Lesion level	ROC-AUC	No	Lesion *detection*
Moradi *et al.*[93]	16	RF time series	Retrospective, 1 inst.	ML	RP	Patient level	Se, Sp ACC, ROC-AUC	No	Lesion *classification*
Imani *et al.*[94]	14	RF time series	Retrospective, 1 inst.	TML	RP, Biopsy	Patient level	Se, Sp ACC, ROC-AUC	No	Lesion *classification*
Hassan *et al.*[88]	1151	TRUS	Retrospective, 1 inst. public data set	TML & DL	Biopsy	Patient level	ACC	No	Lesion *classification*

Acc, accuracy; AI, artificial intelligence; AUC, area under the curve; confirm., confirmation; DCE-US, dynamic contrast-enhanced ultrasound; DL, deep learning; inst., institution; MSE, Mean Square Error; path, Pathology; RF, radio frequency; ROC-AUC, receiver operating characteristics–area under the curve; RP, radical prostatectomy; Se, sensitivity; Sp, specificity; SWE, shear-wave elastography; TeUS, temporal enhanced ultrasound; TML, traditional machine learning; TRUS, transrectal ultrasound.

ultrasound-based imaging modalities (Table 3). These models used a variety of AI methods, ranging from traditional machine learning to deep learning (Table 3). Some other studies investigated the role of radio-frequency time-series data[93,94] to detect prostate cancer using traditional machine learning. Moreover most of these studies focused on the task of lesion *classification* (classifying a physician-outlined region of interest into benign *vs* cancerous tissue)[72,90,91,93,94] while only a few focused on lesion *detection* (detecting and localizing cancer on the entire ultrasound image).[80,89,92]

Summary. Prostate cancer detection on ultrasound images using AI models remains poor, and miminal AI literature exists employing AI approaches on grayscale ultrasound. Development of AI-based approaches for prostate cancer detection on ultrasound represents a significant research opportunity. Generalizability of these methods also needs further investigation since most of these methods have been evaluated on small patient cohorts with retrospective data from single institutions. Privacy-protecting data sharing and public availability of source code and trained models are imperative to improve the performance of AI models on ultrasound (see section "Challenges in AI" for PCa for more details).

Cancer detection on prostate histopathology images

Gleason grading[95] on histopathology images is the strongest predictor of prostate cancer aggressiveness and recurrence. However, Gleason grading suffers from significant inter- and intra-pathologist variability.[96–98] While subspecialized genitourinary pathologists achieve high concordance in Gleason grading, such expertise is not universally available. The emergence of technology to digitize glass slides into whole slide images (WSIs) has revolutionized the field of computational pathology by enabling computer-assisted diagnostic support to pathologists. AI models on prostate histopathology images have been developed to distinguish cancer from non-cancer regions[99,100] and for automated Gleason grading.[60,101–108]

AI models for pathology images also suffer from challenges associated with limited labeled data sets. In addition, pathology images are extremely large, which leads to additional challenges in processing them, as well as in generating labeled data sets. To put this in perspective, a single whole-mount histopathology slice from a radical prostatectomy patient in uncompressed form is 2–4 Gigabytes (GB) in size, and a single patient data with several whole-mount slices is often more than 20 GB in size. Even in compressed form, a single patient data with several histopathology slices (biopsy or whole-mount) can occupy an average storage size of 2–3 GB.[109] When compared to natural images, around 470 whole slide images contain approximately the same number of pixels as the entire ImageNet[110] data set (the public data set of over 14 million natural images used to train AI models for classification of natural images).[102] As such, several studies considered only a subset of the complete pathology image, or tissue microarrays[111–113] for development and validation of AI models.

In this review, we only include studies that considered digital histopathology images derived from prostate needle core biopsies or radical prostatectomies. Most recent studies on cancer detection and Gleason grading on histopathology images use deep learning models. Similar to AI models for prostate cancer detection on MRI, AI models for prostate cancer detection on histopathology images also differ in (a) ground truth labels for training and validation, (b) data type and size, and (c) evaluation criteria (see Table 4).

Labels for training and evaluating AI models for Gleason grading on prostate histopathology images are derived either from pathology reports,[102,104,114] or from pixel-level annotations from experienced pathologists.[60,101,105] Digital histopathology images to train AI models may either be derived from prostate needle core biopsies[60,101–106] or radical prostatectomy[107,108] specimens. The data set size varies widely (Table 4), mostly depending on the label used for training and evaluation. AI models trained with pixel-level Gleason pattern labels from experienced pathologists typically have smaller data set sizes,[101,106,108] whereas those developed with patient-level labels from diagnostic reports have larger data set sizes.[102,104]

Evaluation of AI models are either on a pixel-level (whether the AI method correctly predicts Gleason patterns for each pixel of the image), region-level (whether the AI method assigns the correct Gleason score to a given region of the digitized histopathology image), or slide-level

Table 4. AI models for cancer detection and Gleason grading on prostate histopathology whole slide images (WSI).

Study	Input data	Cohort size	Data type	Algorithm	Training labels	Evaluation labels	Evaluation granularity	Evaluation metric	Code availability
Lucas *et al.*[101]	WSI, biopsy	38 slides	Retrospective, 1 inst.	DL	Pathologist, pixel-level	Pathologist, pixel-level	Patch-based	Se, Sp, F1-score	No
Campanella *et al.*[102]	WSI, biopsy	15,187 slides	Retrospective, multiple inst.,	DL	Reported diagnosis	Reported diagnosis	Slide-level	ROC-AUC	Yes
Bulten *et al.*[96]	WSI, biopsy	1410 slides	Retrospective, multiple inst.,	DL	Pathologists' reports	Pathologists' reports, Consensus reference standard by 3 expert urologic pathologist	slide-level	ROC-AUC, F1-score, Acc, Prec, Rec, Sp, NPV	Yes
Nagpal *et al.*[107]	WSI, RP	1557 slides	Retrospective, multiple inst.	DL	Slide-level & region-level annotations by pathologists	Slide-level & region-level annotations by pathologists	slide-level	ROC-AUC	No
Pinckaers *et al.*[114]	WSI, biopsy	5949 slides	Retrospective, multiple inst. retrospective	DL	Pathologists' reports	Pathologists' reports, Consensus reference standard by 3 expert urologic pathologist	slide-level	ROC-AUC	Yes
Ström *et al.*[103]	WSI, biopsy	1474 patients, 9001 slides	Prospectively collected, multiple inst.	DL	Annotations by single experienced urological pathologist	Annotations by 23 experienced urological pathologist	slide-level	ROC-AUC, Se, Sp, cancer length measurement, Cohen's kappa	No
Marginean *et al.*[105]	WSI, biopsy	195 patients, 735 slides	Retrospective, 1 inst., same slide different scanners	DL	Pixel-level annotations by 2 experienced pathologists	Pixel-level annotations by 2 experienced pathologists	Pixel-level slide-level	Correlation, Se, Sp	No
Kott *et al.*[106]	WSI, biopsy	80 patients, 85 slides	Retrospective, 1 inst.	DL	Pixel-level annotations by pathologists	Pixel-level annotations by pathologists	Patch-level	Acc, Se, Sp, Prec	No
Li *et al.*[108]	WSI, RP	70 patients, 543 slides	Retrospective, 1 inst.	DL & TML	Pixel-level annotations by pathologists	Pixel-level annotations by pathologists	Pixel-level	Overall Pixel Acc. IoU	No
Ryu *et al.*[60]	WSI, biopsy	1833 slides	Retrospective, 2 inst.	DL	Pixel-level annotations by 1 experienced pathologist	Slide-level annotations by 3 experienced pathologists, difficulty-level	Slide-level	Cohen's kappa, Tumor length	No

AI, artificial intelligence; DL, deep learning; inst., institution; IoU, intersection over union; NPV, negative predictive value; Prec, precision; ROC-AUC, receiver operating characteristics–area under the curve; RP, radical prostatectomy; Se, Sensitivity; Sp, specificity; TML, traditional machine learning; WSI, whole slide images.

(whether the AI method assigns the Gleason score for the entire slide). Patient-level Gleason scores are often derived from slide-level predictions. Like with MRI, evaluation metrics for histopathology images vary based on the label type and evaluation granularity. For example, pixel-level evaluation is only possible when detailed pixel-level labels are available as in Figure 3, and evaluation metrics may measure the degree of overlap or correlation between labels and predictions, sensitivity and specificity at a very fine granularity (pixel-level) (Figure 3(b)–(d)). However, in most cases, such detailed pixel-level labels are unavailable as these are impractically time-consuming for pathologists, and evaluation is performed on a coarser granularity (region-, slide-, or patient-level) using pathologist reports. Evaluation metrics for such evaluation may include ROC-AUC, Sensitivity, Specificity, Cohen's kappa

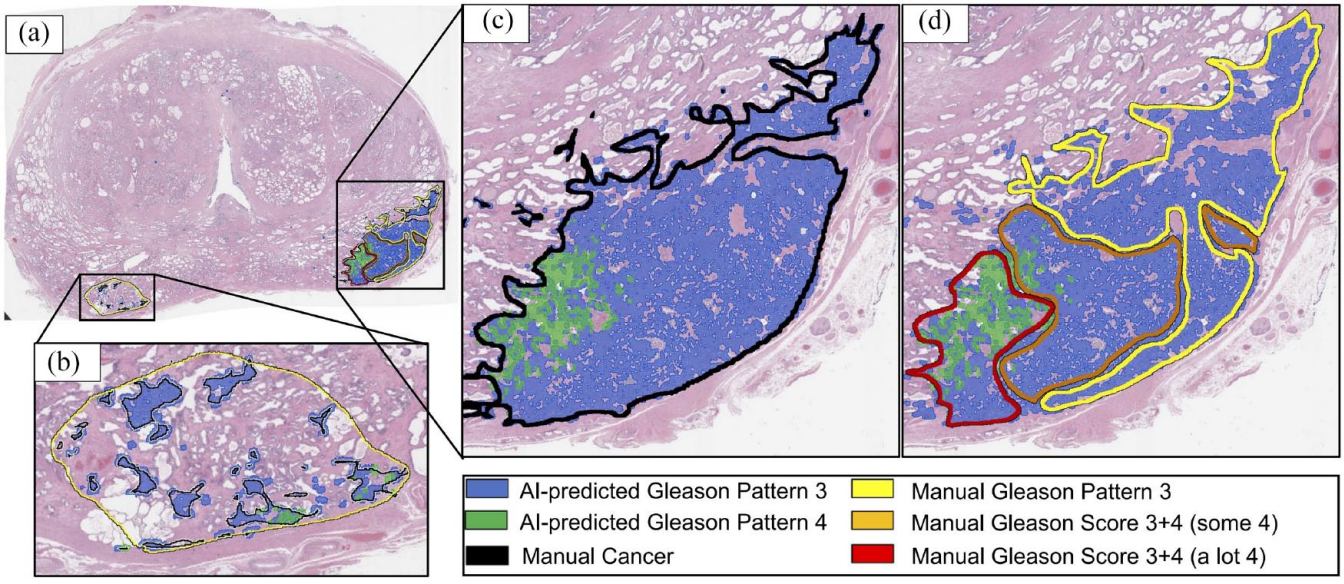

Figure 3. The AI-predicted[60] automated aggressive (Gleason pattern 4, green) and indolent (Gleason Pattern 3, blue) cancers visually match the manual cancer annotations by an expert pathologist (black, yellow, orange, red). (a) Whole mount histopathology image with (b–d) Close-up into the two cancer lesions. (c) Cancer labels manually outlined by an expert pathologist (black outline) shows high agreement with overall cancer (combined blue and green) predicted by the AI model. (b, d) It is impractically time-consuming for a human pathologist to manually assign pixel-level Gleason patterns (yellow, orange, red) to each gland in detail as done by the AI model (blue, green).

etc., but on a region-, slide-, or patient- level. A study to investigate the appropriate approach to evaluate AI classification methods on prostate histopathology images[113] found that AI models trained using data with multiple expert annotations yielded more accurate performance than models trained with single expert annotation. Moreover, patient-based cross-validation provided more realistic and unbiased evaluations of AI models than patch-based cross-validation-based evaluation methods.[113]

Several studies have compared the performance of (a) AI models *versus* pathologists, and (b) pathologists with and without AI assistance.[103,104,107,114–118] Most of these multi-reader and AI-assisted studies confirm the value of AI models in diagnostic pathology; they show increased sensitivity without statistical reduction in specificity, and reduced inter- and intra-observer variability. Recent results from the Prostate cANcer graDe Assessment (PANDA) challenge[118] show that AI models for Gleason grading are generalizable to different patient populations across the world and achieve strong concordance with expert genitourinary pathologists (see Section "Challenges in AI" for PCa for more details). Paige Prostate (Paige AI, New York, USA)[119] recently received approval from the Food and Drug Administration (FDA) as the first ever AI-based clinical pathology solution.[119] Independent studies on Paige Prostate showed generalizable performance on external test sets,[116] and significant sensitivity improvement (74%–97%)[115] of non-genitourinary specialist pathologists without prior experience in digital pathology when assisted by Paige Prostate. Although non-genitourinary pathologist sensitivity improvements were noted for cancers of all sizes and Grade Groups, the most pronounced improvements were noted for smaller and lower-grade (Grade Groups 1, 2, and 3) cancers.[115]

Gleason grading, while standardized, is constantly being tweaked. This means that AI models either need to evolve, or the data need to be linked to hard clinical endpoints, like recurrence and death. For successful clinical deployment of AI models, regulatory authorities (e.g. FDA) need to design strategies where AI models must also evolve with clinical knowledge, rather than being frozen with locked-in variables.

The accurate performance of AI models on prostate histopathology images motivated their use as labeling strategies for training prostate cancer detection methods on MRI.[50,51,66] These AI models on prostate histopathology images generate precise, gland-level annotations, that are not feasible for human pathologists (Figure 3). These automated pixel-level Gleason pattern labels, together with accurate MRI-histopathology registration (see section "MRI-histopathology registration for ground truth labeling of cancer detection models"), enable AI-based radiology-pathology fusion for selective identification of aggressive and indolent cancer on MRI.[50,51] This is not possible using human annotations on MRI.

Summary. AI models for prostate cancer detection and Gleason grading on histopathology images have demonstrated excellent performance comparable to expert genitourinary pathologists. When compared to AI models for radiology images, AI models for histopathology images have undergone more rigid experimentation and validation using larger, heterogenous patient data sets. Multi-reader studies and evaluation on external validation sets demonstrate that these AI models have generalizable performance in heterogeneous patient populations from across the globe and have the potential to help pathologists in the clinic by improving sensitivity of non-genitourinary specialist pathologists, and helping reduce inter- and intra-pathologist variability in Gleason grading.

AI models for supporting tasks in cancer detection

Supporting tasks for cancer detection include tasks that are labor-, time-, or experience-intensive, but form an integral part of the clinical workflow to detect cancer. These supporting tasks include prostate gland segmentation and MRI-ultrasound registration to guide fusion biopsy procedures. Another supporting task is MRI-histopathology registration for patients who underwent radical prostatectomy. MRI-histopathology registration is necessary to study correlations between pre-operative MRI and post-operative histopathology images of the prostate and for deriving accurate ground truth labels for training cancer detection AI models on MRI. While several AI models exist for these supporting tasks, only a few are being commercially used in the clinic for support, or as pre-processing steps for AI cancer detection models.

Prostate gland segmentation to facilitate MRI-ultrasound fusion biopsies

Targeted MRI-ultrasound fusion biopsy workflow relies on accurate prostate gland segmentations on T2-weighted MRI and ultrasound images.[120] However, manually outlining the prostate is a time-consuming and tedious task.[121] Automated methods have the potential to reduce the manual effort, time, and variability associated with prostate gland segmentations on MRI and ultrasound images during clinical biopsy procedures.

AI for prostate gland segmentation on MRI: Many studies have proposed deep learning models to segment the prostate on MRI.[121–134]

As with AI models for cancer detection tasks, AI models for prostate segmentation on MRI are mostly trained and validated on small data sets (40–250 patients),[123–131] often with retrospective, single-center data[124,131–133] without validation in external cohorts.[124,131,132] The trained models and the source code to pre-process the data and train the model are often not publicly available,[123–125,127,128,130,132,133,135] limiting the comparison between these models, as well as their usage. The better-performing models achieved Dice scores (metric of similarity between manual and AI-predicted segmentations) of at least 0.90 in internal and 0.80 in external data sets.[121–123,125–127,130] A recent study that prospectively implemented an AI model for prostate segmentation in a urology clinic found that AI performance was more accurate and 17 times faster than trained radiology technicians.[121] Finally, FDA-cleared commercial AI-based solutions for prostate gland segmentation are also available to optimize the clinical workflow.[136–139]

AI for prostate gland segmentation on ultrasound images: AI models for prostate gland segmentation on grayscale transrectal ultrasound images have used both traditional machine learning,[140–143] as well as deep learning–based approaches.[144–151] To further improve the segmentation of challenging regions (e.g. apex and base), studies have explored the use of prior shape information as statistical shape models,[152,153] and temporal information for transrectal ultrasound images.[154,155] Although these methods demonstrated good performance, most of these studies included small patient cohorts from a single institution and a single manufacturer, thus providing limited evidence

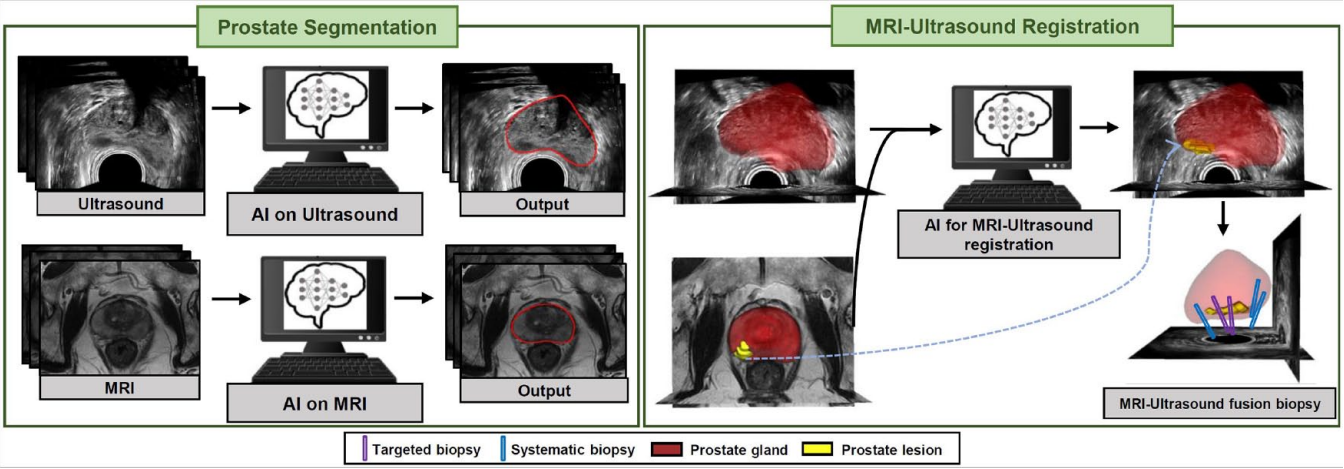

Figure 4. AI can help in supporting tasks for cancer detection like prostate gland segmentation on MRI and ultrasound (left), and MRI-ultrasound registration (right). The AI-predicted prostate segmentations on MRI and ultrasound can help in automated MRI-ultrasound registration which aligns the two modalities, mapping lesions from MRI onto ultrasound. MRI-ultrasound registration helps guide systematic and targeted fusion biopsy procedures.

about generalizability across data from other institutions and different imaging devices.

Summary. AI models for prostate gland segmentation on MRI and ultrasound have demonstrated promising results, but external validation on large patient cohorts are needed for wide clinical implementation. Moreover, source code and trained models must be shared publicly to derive the maximum benefits of the best-performing approaches.

MRI-ultrasound registration to facilitate MRI-ultrasound fusion biopsies

Registration of preoperative MRI and intra-operative ultrasound is necessary for guiding MRI-ultrasound fusion biopsy,[156,157] focal therapy,[158,159] and radiotherapy planning on MRI.[160] However, registration of the two different imaging modalities, MRI and ultrasound, is complicated due to (a) the difference in the underlying MR and ultrasound imaging processes, and (b) the deformation between the two imaging procedures. In an attempt to improve registration between the two modalities, several studies used pre-defined corresponding anatomical structures.[156–159,161,162] Some approaches used deformable transformations[161] to model patient movement, surrounding organs, for example, bladder and rectum, or interaction with surgical instrument, for example, biopsy needles and ultrasound probes. Others used AI models without constrained transformation models,[156–158] or prior knowledge in modeling soft tissue motion.[163,164]

AI models have also been proposed to learn similarity measures[165] or transformation models from either biomechanical simulations (which emphasize biologically meaningful registration)[159] or shape populations.[166] A popular class of methods utilize prostate gland segmentations on both MR and ultrasound images before registering the resulting point sets[162] (Figure 4). AI models processing point set data[167] for registration often represent corresponding structures without detailed voxel-level correspondence.[156,157,164] An advantage of using point set is its robustness to scanning protocols, largely thanks to well-established independent segmentation algorithms, and arguably even faster inference.[157] While most above methods discuss registration between 3D MR and 3D transrectal ultrasound images, there have also been advances in aligning 3D MRI to 2D ultrasound which is much easier to obtain.[168] A summary of the related references in this section are listed in Table 5.

Summary. AI methods have the potential of completely automating the MRI-ultrasound registration task. The best-performing AI-based MRI-ultrasound registration methods achieved average target registration errors of ≈2–3 mm,[161,164] although with relatively large variance. Anatomical information, such as prostate gland segmentations or surface points help improve AI performance in

Table 5. AI models for registration between MRI (T2w) and ultrasound (TRUS) images.

Study	Number of subjects	Data type	Approach	Prostate segmentation	Evaluation metric	Source code availability
Hu *et al.*[166]	143	Retrospective	DL	No	TREs, TDR, RMSE	No
Hu *et al.*[161,169]	76	Retrospective	DL	Yes; manual	TREs, DSC	Yes
Hu *et al.*[163]	76	Retrospective	DL	Yes; manual	TREs, DSC	No
Ghavami *et al.*[162]	59	Retrospective	DL + TML	Yes; DL	DSC, GVE, TREs	No
Hu *et al.*[158]	80	Retrospective	DL	Yes; manual	TREs, DSC	Yes
Haskins *et al.*[165]	679	Retrospective	DL + TML	No	TREs	No
Guo *et al.*[170]	679	Retrospective	DL	Yes; manual	TREs, SRE	Yes
Saeed *et al.*[159]	320	Retrospective	DL	Yes; manual	MAE	No
Baum *et al.*[156,157]	108	Retrospective	DL	Yes; manual	TREs, CD, HD	No
Zeng *et al.*[171]	36	Retrospective	DL	Yes; manual	TREs, DSC	No
Zeng *et al.*[160]	36	Retrospective	DL	Yes; DL	TREs, DSC	No
Song *et al.*[172]	528	Retrospective	DL	Yes; manual	SRE	Yes
Fu *et al.*[164]	50	Retrospective	DL	Yes; DL	TREs, DSC, MSD, HD	No
Guo *et al.*[168]	619	Retrospective	DL	No	TREs, NCC	Yes

AI, artificial intelligence; CD, chamfer distance; DL, deep learning; DSC, dice score coefficient; HD, Hausdorff Distance; GVE, gland volume error; MRI, magnetic resonance imaging; MAE, Mean Absolute Error; MSD, Mean Square Distance; NCC, normalized cross correlation; RMSE, Root Mean Square Error; SRE, surface registration error; T2w, T2-weighted MRI; TDR, tumor detection rate; TML, traditional machine learning; TREs, target registration Errors; TRUS, transrectal ultrasound.

MRI-ultrasound registration. However, of the studies performed to date, all have shortcomings. These shortcomings include retrospective design, lack of prospective evaluation, and development and validation using single institution data. Several MR and ultrasound manufacturers[173,174] have integrated tools to assist with MRI-ultrasound registration, although these are still semi-automated and require human input in real-time.

MRI-histopathology registration for ground truth labeling of cancer detection models

Several AI models of prostate cancer detection on MRI[43,44,48,50,51] derive ground truth labels from whole-mount prostate histopathology images through accurate registration with pre-operative MRI (Table 6). Pathologist labels mapped from histopathology images onto MRI through image registration is considered the most accurate labeling strategy[64] for training AI cancer detection models.

MRI-histopathology registrations are performed either cognitively, manually, or automatically. In cognitive methods,[43,48] researchers mentally project cancer labels from histopathology images onto the corresponding MRI slices without quantitative spatial alignment of the two modalities. Manual registration involves spatially aligning the MRI and histopathology images on a case-by-case basis by human experts. Cognitive and manual registration approaches are known to be

Table 6. MRI-histopathology registration approaches (not exhaustive) for generating ground truth cancer labels on MRI.

Study	Number of subjects	Pathology type	Registration type	Intermediate modality	Require 2D slice correspondences	Prostate sectioning	Source code availability
Chappelow *et al.*[175]	25	Whole-mount	Traditional automated	None	Yes	Manual	No
Ward *et al.*[176]	13	Whole-mount	Traditional automated	Fiducial markers	Yes	Image-guided	No
Kalavagunta *et al.*[177]	35	Pseudo-whole mount	Traditional automated	Manual landmarks	Yes	Sectioning box	No
Reynolds *et al.*[178]	6	Whole-mount	Traditional automated	Ex vivo MRI + Manual landmarks	Yes	Sectioning box	No
Li *et al.*[179]	19	Pseudo-whole mount	Traditional automated	None	Yes	Manual	No
Losnegård *et al.*[180]	12	Whole-mount	Traditional automated	None	No	Manual	No
Wu *et al.*[181]	17	Whole-mount	Traditional automated	Ex vivo MRI + fiducial markers	Yes	3D-printed mold	No
Rusu *et al.*[182]	157	Whole-mount	Traditional automated	None	Yes	3D-printed mold	Yes
Shao *et al.*[183]	152	Whole-mount	Deep learning	None	Yes	3D-printed mold	Yes
Sood *et al.*[184]	106	Whole-mount	Traditional automated	None	No	3D-printed mold	No
Shao *et al.*[185]	183	Whole-mount	Deep learning	None	Yes	3D-printed mold	No

MRI, magnetic resonance imaging.

labor-intensive, requiring highly skilled experts in both radiology and pathology. As such, these approaches can only been applied to small patient data sets. Moreover, these methods failed to map MRI-invisible or hardly visible lesions from histopathology images onto MRI. In traditional automated approaches, MRI, and histopathology images are directly registered by using customized image similarity loss functions,[175,179,182,186] fiducial markers,[176] or the use of intermediate *ex vivo* imaging modalities to facilitate the registration.[181,187] Many of these methods rely on patient-specific 3D-printed molds derived from preoperative MRI[135] to maintain slice correspondences between MRI and histopathology images, while others directly register the MRI and histopathology volumes without the need for MRI-histopathology slice correspondences.[180,188–190] However, the absence of accurate slice correspondences can lead to partial volume artifacts. A recent study used 3D super-resolution of MRI and histopathology images prior to performing a 3D registration to alleviate issues with partial volume artifacts.[184]

Although traditional automated approaches are advantageous over manual and cognitive registration, they are often time-consuming, requiring several minutes to register data from a single patient. Recent deep learning models[183] greatly speed up the registration process. Traditional automated approaches generally require manual prostate gland segmentation to facilitate the registration. To avoid this step, a recent deep learning model proposed a weakly supervised registration approach that avoids the need for prostate segmentation at inference.[185]

Summary. Registration of pre-operative MRI with post-operative histopathology images is complicated due to the inherent difference of the two modalities, and their acquisition processes. Nonetheless, it is an important task for deriving accurate ground truth labels for cancer detection models on MRI. Although several automated MRI-histopathology registration approaches have been developed, only a few studies on AI-based prostate cancer detection[44,47,50,51] use automated registration methods to derive ground truth labels. There remains the need to publicly share source code, trained models, and benchmarking data sets to compare different registration approaches, and to enable their usage in deriving accurate cancer labels.

Challenges in AI for PCa

AI models have great potential to improve diagnosis and management of prostate cancer and to bring precision medicine to patients. For example, AI models may help accurate and timely detection of aggressive cancer, reduction of cancer-related deaths, and avoiding unnecessary invasive biopsies and their associated side-effects. In addition, AI models may help streamline supporting tasks for cancer detection that are labor-, time-, and experience-intensive.

Despite promising outcomes of AI research in prostate cancer diagnosis, most methods are not ready to be used in clinical care. In the United States, only one prostate cancer detection system, the 'Paige Prostate', has received FDA approval for *in vitro* diagnostic use for detecting cancer on histopathology images of prostate biopsies.[119] Other commercially available FDA-cleared or European CE-marked applications (OnQ Prostate[137], PROView[136] Quantib Prostate,[138] qp-Prostate[139]) mostly focus on supporting tasks such as prostate segmentation, volumetry computation, or PSA density calculation.

Reducing the gap between academic research and translation of these AI models for diagnostic support in the clinic will require addressing the following challenges:

(a) Limited labeled data:

In order to be robust, generalizable, and unbiased, AI models must be trained and validated with large data sets with accurate labels, which capture variability in patient populations and imaging acquisition. For example, the AI models for natural image recognition tasks that have achieved performance exceeding humans, have been trained and validated with ≈14 million images in the publicly available ImageNet data set.[191] However, due to privacy concerns of medical data-sharing, AI models for prostate cancer diagnosis are mostly trained with small data sets, often from a single institution having patient populations with specific socio-economic or racial distributions, or images acquired with certain scanners and acquisition protocols. AI models trained on a homogeneous patient population or imaging data may not generalize to different demographics or different kinds of scanners.[192] Studies to test generalizability of AI models on prostate cancer in different racial, socio-economic, or ethnic patient populations are limited, thereby raising the question of unbiased and robust applicability of these AI-based models support in the clinic.

Development of robust, generalizable, and unbiased AI systems may require consolidated efforts from medical institutions across the globe to enable privacy-protecting medical data sharing.

Research on federated learning-[193,194] or incremental learning-based[195] AI models that enable improvement of AI models through sharing and constantly updating the model, rather than sharing of data, is another possible solution. Finally, self-supervised,[68,196] weakly supervised,[67,102] semi-supervised,[197,198] few-shot[199] learning techniques may be used to further improve robustness of AI models that are developed with large data sets, but without accurate labels (i.e. thousands of prostate MRIs in Picture Archiving and Communication System–PACS) that lack annotations about where cancer is located in the images).

(b) Limited multi-reader studies to assess AI-assisted performance of clinicians:

There are relatively few multi-reader studies that assess how AI models perform in comparison to clinicians. Moreover, only a few studies exist that analyze if the use of AI models can help standardize or improve human-reader performance.[55,104,115,117,200] This would be particularly useful in assessing the potential of AI in resource-lacking institutions, or for less-experienced radiologists or pathologists.

Multireader studies for prostate MRI interpretation[42,55,200,201] in limited patient populations demonstrated the utility of AI in providing diagnostic support to radiologists by improving sensitivity and positive predictive value of patient-level and lesion-level cancer detection. Multireader studies for prostate histopathology image interpretation[103,104,117] demonstrated that AI models performed similarly to highly experienced genitourinary pathologists. Furthermore, AI-assisted pathologists outperformed the standalone AI system and the unassisted pathologists, by reducing inter- and intra-pathologist variability in Gleason grading. These multireader studies unequivocally confirm the benefits of clinician-AI synergy, but the need remains for more extensive and carefully designed studies with larger, varied, multi-center patient populations to completely assess the clinical readiness of the AI models, particularly for radiology images. Moreover, research needs to be conducted on making the technology user-friendly and ensuring that it does not negatively affect clinical workflow.

(c) Limited prospective evaluations: Most AI models on prostate cancer have been trained and evaluated with retrospective data; only a few have focused on prospective evaluation.[103,121] To deploy AI models for diagnostic support in the clinic, models trained with retrospective data must be evaluated in a prospective setting. Moreover, clinical trials that evaluate AI models for prostate cancer detection on non-invasive imaging need to be designed and explored.

(d) Lack of standard evaluation criteria: Variability in evaluation criteria used in existing AI studies makes it difficult to compare the different automated approaches with one another. Definition of a unified, clinically relevant standard to evaluate AI models is required to assess the best-performing, robust, and unbiased AI system for clinical deployment. While patient outcome like death or recurrence can be considered as hard clinical endpoints for evaluation, such long-term outcome data are often unavailable since prostate cancer is a slow-progressing disease. Making the source code and trained models of published studies publicly available helps in testing different approaches on independent data sets, often without the need to share the data, or with different evaluation criteria.

A possible way to encourage participation of the AI community in prostate cancer research and to enable comparison, validation, and benchmarking of AI models for cancer detection could be through organization of grand challenges. Grand challenges provide large, publicly available data sets for training AI models, and also allow comparison and validation of AI models through well-curated, representative test data sets and defined clinically relevant evaluation metrics. The Prostate cANcer graDe Assessment (PANDA) Challenge[118] was organized with the aim of testing the generalizability and clinical-readiness of AI models for Gleason grading on prostate biospy histopathology images. More than 10,000 histopathology images were made publicly available through the challenge, and 1290 AI developers from 65 countries participated in this challenge. On an independent validation set of 2009 biopsies, the AI models generalized across different patient populations, imaging parameters and reference standards, achieving strong agreements with expert genitourinary pathologists. Results from the PANDA challenge suggest that AI models for histopathology images are robust enough to be implemented in clinical trials.[118]

Grand challenges for MR images have included relatively smaller patient cohorts with lesser diversity. The ProstateX and ProstateX-2 challenges[202] were organized with the aim of developing AI models that can enable cancer detection and aggressiveness characterization on prostate MR images. The ProstateX challenges included smaller patient cohort (346 patients) scanned at a single institution (Radboud University Medical Center) using two different Siemens scanners. The PROMISE-12 challenge[134] was organized with the aim of developing AI models that can segment the prostate gland. It included 100 patients from four different institutions and different scanners and scanning protocols. The NCI-ISBI 2013 challenge on automated segmentation of prostate structures[203] was organized to segment the prostate peripheral zone and central glands, in addition to the whole prostate gland. The NCI-ISBI challenge included 80 patients from two different institutions, scanned with two different scanners and scanning protocols. These studies show that AI models for prostate MR interpretation still need development and external validation. Multi-institution collaborations are needed to drive

more extensive grand challenges on prostate MRI and ultrasound images. Such grand challenges will encourage AI model-development on larger and heterogenous patient data sets. Moreover, these challenges will also allow testing generalizability and clinical readiness of AI models.

Limitations of this study

This study has several limitations. First, this is not an exhaustive review. While we curated relevant literature to cover the breadth and depth of the different applications of AI models in prostate cancer diagnosis, the multitude of research publications in the different facets of this field and the limited review space forced us to be selective. Second, we could not provide a comparative analysis of the different methods due to the variability in data sets and evaluation criteria. Third, we did not provide extensive details of the AI models. Our aim in this study was to provide a broad overview of the potential of AI models in prostate cancer care and make it generally comprehensible to both clinical and technical readers. We described algorithms as belonging to two major categories – traditional machine learning and deep learning based approaches. Fourth, we did not discuss automated systems that use clinical data, genomic data, or newer imaging modalities such as prostate-specific membrane antigen (PSMA) PET scans.

We focused on three major imaging modalities (MRI, ultrasound, and histopathology) most commonly used in routine clinical care and did not cover newer imaging modalities that are showing benefit in prostate cancer detection. One such modality is Gallium-68 PSMA-11 PET-CT scans.[204] Recent studies demonstrate that PSMA PET-CT scans can significantly improve prostate cancer detection and treatment planning, are more accurate than conventional imaging with CT and bone scanning,[205] add value to MRI for diagnosis,[206] and may enable better prediction of preoperative pathological outcomes than MRI.207–209 There is an opportunity for AI to address shortcomings of these new imaging modalities, but due to its recent FDA approval, only a few published AI models[32,210,211] exist for prostate cancer detection using PSMA PET currently.

Conclusion

AI models for prostate cancer detection on imaging are showing great promise and encouraging performance. Yet, AI models on radiology images need further development and validation on larger and diverse patient populations, standardized evaluation criteria, multi-reader studies, and prospective evaluation to make them robust, generalizable, and unbiased. AI models on histopathology images have undergone more rigid experimentation on larger data sets and multi-reader evaluations, as compared to radiology images. These studies suggest that AI models on histopathology images are generalizable to different patient populations across the globe and can assist pathologists in the clinic by reducing intra- and inter-pathologist variability in Gleason grading. AI models for supporting tasks have the potential of reducing manual labor- and time-investments, particularly as more images are acquired and processed daily in clinical care. However, research needs to be done on best practices to integrate AI predictions in the clinical workflow in a seamless way, to enable clinician–AI synergy and precision medicine. AI-enabled precision medicine may eventually help in reducing disparity and advancing health equity in prostate cancer management.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Author contributions

Indrani Bhattacharya: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing – original draft.

Yash S. Khandwala: Data curation; Investigation; Validation.

Sulaiman Vesal: Data curation; Formal analysis; Investigation.

Wei Shao: Data curation; Formal analysis; Investigation.

Qianye Yang: Data curation; Formal analysis; Investigation.

Simon J.C. Soerensen: Data curation; Investigation.

Richard E. Fan: Data curation; Project administration; Resources.

Pejman Ghanouni: Supervision; Writing – review & editing.

Christian A. Kunder: Supervision; Visualization; Writing – review & editing.

James D. Brooks: Supervision; Writing – review & editing.

Yipeng Hu: Supervision; Writing – review & editing.

Mirabela Rusu: Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing.

Geoffrey A. Sonn: Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing.

Acknowledgements
None.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (award no. R37CA260346). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors acknowledge the following funding sources: Departments of Radiology and Urology, Stanford University, GE Healthcare Blue Sky Award, National Institutes of Health, National Cancer Institute (grant no. U01CA196387, to JDB), and the generous philanthropic support of our patients (GS).

Competing interests
The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: MR has research grants from GE Healthcare and Philips Healthcare.

Availability of data and material
Not applicable.

ORCID iD
Geoffrey A. Sonn https://orcid.org/0000-0002-0850-1956

References

1. Ilic D, Djulbegovic M, Jung JH, *et al.* Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *BMJ* 2018; 362: k3519.
2. Welch HG and Albertsen PC. Reconsidering prostate cancer mortality – the future of PSA screening. *N Engl J Med* 2020; 382: 1557–1563.
3. Goldenberg SL, Nir G and Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* 2019; 16: 391–403.
4. Van Booven DJ, Kuchakulla M, Pai R, *et al.* A systematic review of artificial intelligence in prostate cancer. *Res Rep Urol* 2021; 13: 31.
5. Hamet P and Tremblay J. Artificial intelligence in medicine. *Metabolism* 2017; 69: S36–S40.
6. Kulkarni S, Seneviratne N, Baig MS, *et al.* Artificial intelligence in medicine: where are we now? *Acad Radiol* 2020; 27: 62–70.
7. Briganti G and Le Moine O. Artificial intelligence in medicine: today and tomorrow. *Front Med* 2020; 7: 27.
8. Ahmed HU, Bosaily AES, Brown LC, *et al.* Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* 2017; 389: 815–822.
9. El-Shater Bosaily A, Parker C, Brown LC, *et al.* PROMIS – prostate MR imaging study: a paired validating cohort study evaluating the role of multi-parametric MRI in men with clinical suspicion of prostate cancer. *Contemp Clin Trials* 2015; 42: 26–40.
10. Johnson DC, Raman SS, Mirak SA, *et al.* Detection of individual prostate cancer foci via multiparametric magnetic resonance imaging. *Eur Urol* 2019; 75: 712–720.
11. van der Leest M, Cornel E, Israel B, *et al.* Head-to-head comparison of transrectal ultrasound-guided prostate biopsy versus multiparametric prostate resonance imaging with subsequent magnetic resonance-guided biopsy in biopsy-naïve men with elevated prostate-specific antigen: a large prospective multicenter clinical study. *Eur Urol* 2019; 75: 570–578.
12. Sonn GA, Chang E, Natarajan S, *et al.* Value of targeted prostate biopsy using magnetic resonance–ultrasound fusion in men with prior negative biopsy and elevated prostate-specific antigen. *Eur Urol* 2014; 65: 809–815.

13. Kasivisvanathan V, Rannikko AS, Borghi M, *et al*. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med* 2018; 378: 1767–1777.

14. Liu W, Patil D, Howard DH, *et al*. Adoption of prebiopsy magnetic resonance imaging for men undergoing prostate biopsy in the united states. *Urology* 2018; 117: 57–63.

15. Turkbey B, Rosenkrantz AB, Haider MA, *et al*. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur Urol* 2019; 76: 340–351.

16. Barentsz JO, Weinreb JC, Verma S, *et al*. Synopsis of the PI-RADS v2 guidelines for multiparametric prostate magnetic resonance imaging and recommendations for use. *Eur Urol* 2016; 69: 41–49.

17. Sonn GA, Fan RE, Ghanouni P, *et al*. Prostate magnetic resonance imaging interpretation varies substantially across radiologists. *Eur Urol Focus* 2019; 5: 592–599.

18. Wang NN, Zhou SR, Chen L, *et al*. The stanford prostate cancer calculator: development and external validation of online nomograms incorporating PIRADS scores to predict clinically significant prostate cancer. *Urol Oncol* 2021; 39: 831.e19–831.e27.

19. Hiremath A, Shiradkar R, Fu P, *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: e445–e454.

20. Algohary A, Shiradkar R, Pahwa S, *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* 2020; 12: 2200.

21. Antonelli M, Johnston EW, Dikaios N, *et al*. Machine learning classifiers can predict Gleason pattern 4 prostate cancer with greater accuracy than experienced radiologists. *Eur Radiol* 2019; 29: 4754–4764.

22. Bleker J, Kwee TC, Dierckx RAJO, *et al*. Multiparametric MRI and auto-fixed volume of interest-based radiomics signature for clinically significant peripheral zone prostate cancer. *Eur Radiol* 2020; 30: 1313–1324.

23. Bonekamp D, Kohl S, Wiesenfarth M, *et al*. Radiomic machine learning for characterization of prostate lesions with MRI: comparison to ADC values. *Radiology* 2018; 289: 128–137.

24. Chen T, Li M, Gu Y, *et al*. Prostate cancer differentiation and aggressiveness: assessment with a radiomic- based model vs. PI-RADS v2. *J Magn Reson Imaging* 2019; 49: 875–884.

25. Akamine Y, Ueda Y, Ueno Y, *et al*. Application of hierarchical clustering to multi-parametric MR in prostate: differentiation of tumor and normal tissue with high accuracy. *Magn Reson Imaging* 2020; 74: 90–95.

26. Kwon D, Reis IM, Breto AL, *et al*. Classification of suspicious lesions on prostate multiparametric MRI using machine learning. *J Med Imaging* 2018; 5: 034502.

27. Chaddad A, Kucharczyk MJ and Niazi T. Multimodal radiomic features for the predicting Gleason score of prostate cancer. *Cancers* 2018; 10: 249.

28. Hectors SJ, Cherny M, Yadav KK, *et al*. Radiomics features measured with multiparametric magnetic resonance imaging predict prostate cancer aggressiveness. *J Urol* 2019; 202: 498–505.

29. Xu M, Fang M, Zou J, *et al*. Using biparametric MRI radiomics signature to differentiate between benign and malignant prostate lesions. *Eur J Radiol* 2019; 114: 38–44.

30. Viswanath SE, Chirra PV, Yim MC, *et al*. Comparing radiomic classifiers and classifier ensembles for detection of peripheral zone prostate tumors on T2-weighted MRI: a multi-site study. *BMC Med Imaging* 2019; 19: 22.

31. Transin S, Souchon R, Gonindard-Melodelima C, *et al*. Computer-aided diagnosis system for characterizing ISUP grade≥2 prostate cancers at multiparametric MRI: a cross-vendor evaluation. *Diagn Interv Imaging* 2019; 100: 801–811.

32. Zhang Y, Chen W, Yue X, *et al*. Development of a novel, multi-parametric, MRI-based radiomic nomogram for differentiating between clinically significant and insignificant prostate cancer. *Front Oncol* 2020; 10: 888.

33. Deniffel D, Abraham N, Namdar K, *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: 6867–6876.

34. Song Y, Zhang YD, Yan X, *et al*. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. *J Magn Reson Imaging* 2018; 48: 1570–1577.

35. Takeuchi T, Hattori-Kato M, Okuno Y, *et al*. Prediction of prostate cancer by deep learning

with multilayer artificial neural network. *Can Urol Assoc J* 2019; 13: E145–E150.

36. Yuan Y, Qin W, Buyyounouski M, *et al*. Prostate cancer classification with multiparametric MRI transfer learning model. *Med Phys* 2019; 46: 756–765.
37. Aldoj N, Lukas S, Dewey M, *et al*. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. *Eur Radiol* 2020; 30: 1243–1253.
38. Zhong X, Cao R, Shakeri S, *et al*. Deep transfer learning-based prostate cancer classification using 3 Tesla multi-parametric MRI. *Abdom Radiol* 2019; 44: 2030–2039.
39. Abraham B and Nair MS. Automated grading of prostate cancer using convolutional neural network and ordinal class classifier. *Inform Med Unlocked* 2019; 17: 100256.
40. Saha A, Hosseinzadeh M and Huisman H. End-to-end prostate cancer detection in bpMRI via 3D CNNs: effects of attention mechanisms, clinical priori and decoupled false positive reduction. *arXiv preprint arXiv* 2021: 210103244, https://arxiv.org/pdf/2101.03244.pdf
41. Yu X, Lou B, Zhang D, *et al*. Deep attentive panoptic model for prostate cancer detection using biparametric MRI scans. In: Martel AL, Abolmaesumi P, Stoyanov D, *et al*. (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2020, pp. 594–604.
42. Schelb P, Kohl S, Radtke JP, *et al*. Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment. *Radiology* 2019; 293: 607–617.
43. Sumathipala Y, Lay NS, Turkbey B, *et al*. Prostate cancer detection from multi-institution multiparametric MRIs using deep convolutional neural networks. *J Med Imaging* 2018; 5: 044507.
44. Bhattacharya I, Seetharaman A, Shao W, *et al*. CorrSigNet: learning correlated prostate cancer signatures from radiology and pathology images for improved computer aided diagnosis. In: Martel AL, Abolmaesumi P, Stoyanov D, *et al*. (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2020, pp. 315–325.
45. Sanyal J, Banerjee I, Hahn L, *et al*. An automated two-step pipeline for aggressive prostate lesion detection from multi-parametric MR sequence. *AMIA Jt Summits Transl Sci Proc* 2020; 2020: 552–560.
46. Jin J, Zhang L, Leng E, *et al*. Detection of prostate cancer with multiparametric MRI utilizing the anatomic structure of the prostate. *Stat Med* 2018; 37: 3214–3229.
47. McGarry SD, Bukowy JD, Iczkowski KA, *et al*. Gleason probability maps: a radiomics tool for mapping prostate cancer likelihood in MRI space. *Tomography* 2019; 5: 127–134.
48. Cao R, Mohammadian Bajgiran A, Afshari Mirak S, *et al*. Joint prostate cancer detection and Gleason score prediction in mp-MRI via FocalNet. *IEEE Trans Med Imaging* 2019; 38: 2496–2506.
49. De Vente C, Vos P, Hosseinzadeh M, *et al*. Deep learning regression for prostate cancer detection and grading in bi-parametric MRI. *IEEE Trans Biomed Eng* 2021; 68: 374–383.
50. Seetharaman A, Bhattacharya I, Chen LC, *et al*. Automated detection of aggressive and indolent prostate cancer on magnetic resonance imaging. *Med Phys* 2021; 48: 2960–2972.
51. Bhattacharya I, Seetharaman A, Kunder C, *et al*. Selective identification and localization of indolent and aggressive prostate cancers via CorrSigNIA: an MRI-pathology correlation and deep learning framework. *Med Image Anal* 2022; 75: 102288.
52. Zhang G, Shen X, Zhang Y, *et al*. Cross-modal prostate cancer segmentation via self-attention distillation. *IEEE J Biomed Health Inform*. Epub ahead of print 12 November 2021. DOI: 10.1109/JBHI.2021.3127688.
53. Alkadi R, Taher F, El-Baz A, *et al*. A deep learning-based approach for the detection and localization of prostate cancer in T2 magnetic resonance images. *J Digit Imaging* 2019; 32: 793–807.
54. Arif M, Schoots IG, Castillo Tovar J, *et al*. Clinically significant prostate cancer detection and segmentation in low- risk patients using a convolutional neural network on multi-parametric MRI. *Eur Radiol* 2020; 30: 6582–6592.
55. Mehralivand S, Harmon SA, Shih JH, *et al*. Multicenter multireader evaluation of an artificial intelligence-based attention mapping system for the detection of prostate cancer with multiparametric MRI. *Am J Roentgenol* 2020; 215: 903–912.
56. Netzer N, Weißer C, Schelb P, *et al*. Fully automatic deep learning in bi-institutional prostate magnetic resonance imaging: effects of cohort size and heterogeneity. *Invest Radiol* 2021; 56: 799–808.

57. Duran A, Dussert G, Rouvière O, *et al*. ProstAttention-Net: a deep attention model for prostate cancer segmentation by aggressiveness in MRI scans. *Med Image Anal* 2022; 77: 102347.

58. Litjens G, Debats O, Barentsz J, *et al*. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging* 2014; 33: 1083–1092.

59. Hosseinzadeh M, Saha A, Brand P, *et al*. Deep learning–assisted prostate cancer detection on bi-parametric MRI: minimum training data size requirements and effect of prior knowledge. *Eur Radiol* 2022; 32: 2224–2234.

60. Ryu HS, Jin MS, Park JH, *et al*. Automated Gleason scoring and tumor quantification in prostate core needle biopsy images using deep neural networks and its comparison with pathologist-based assessment. *Cancers* 2019; 11: 1860.

61. Stolk TT, de Jong IJ, Kwee TC, *et al*. False positives in PIRADS (v2) 3, 4, and 5 lesions: relationship with reader experience and zonal location. *Abdom Radiol* 2019; 44: 1044–1051.

62. Kasel-Seibert M, Lehmann T, Aschenbach R, *et al*. Assessment of PI-RADS v2 for the detection of prostate cancer. *Eur J Radiol* 2016; 85: 726–731.

63. Priester A, Natarajan S, Khoshnoodi P, *et al*. Magnetic resonance imaging underestimation of prostate cancer geometry: use of patient specific molds to correlate images with whole mount pathology. *J Urol* 2017; 197: 320–326.

64. Bhattacharya I, Lim DS, Aung HL, *et al*. Bridging the gap between prostate radiology and pathology through machine learning. *arXiv preprint arXiv* 2021: 211202164, https://arxiv.org/pdf/2112.02164.pdf

65. Hosseinzadeh M, Brand P and Huisman H. Effect of adding probabilistic zonal prior in deep learning-based prostate cancer detection. *arXiv preprint arXiv* 2019: 190712382, https://arxiv.org/pdf/1907.12382.pdf

66. Bhattacharya I, Shao W, Soerensen SJC, *et al*. Integrating zonal priors and pathomic MRI biomarkers for improved aggressive prostate cancer detection on MRI. In: *Medical imaging 2022: computer-aided diagnosis*. International Society for Optics and Photonics, https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12033/0000/Integrating-zonal-priors-and-pathomic-MRI-biomarkers-for-improved-aggressive/10.1117/12.2612433.short?SSO=1

67. Cao R, Zhong X, Scalzo F, *et al*. Prostate cancer inference via weakly-supervised learning using a large collection of negative MRI. In: *Proceedings of the IEEE/CVF international conference on computer vision workshops*, Seoul, Korea, 27–28 October 2019.

68. Bolous A, Seetharaman A, Bhattacharya I, *et al*. Clinically significant prostate cancer detection on MRI with self-supervised learning using image context restoration. In: Mazurowski MA and Drukker K (eds) *Medical imaging 2021: computer-aided diagnosis*, vol. 11597. Bellingham, WA: International Society for Optics and Photonics, 2021, p. 115971M.

69. Michalski JM, Pisansky TM, Lawton CA, *et al*. Chapter 53 – prostate cancer. In: Gunderson LL and Tepper JE (eds) *Clinical radiation oncology*. 4th ed. Philadelphia, PA: Elsevier, 2016, pp. 1038–1095.e18.

70. Sarkar S and Das S. A review of imaging methods for prostate cancer detection: supplementary issue: image and video acquisition and processing for clinical applications. *Biomed Eng Comput Biol* 2016; 7s1: BECB.S34255.

71. Choi YH, Kang MY, Sung HH, *et al*. Comparison of cancer detection rates between TRUS-guided biopsy and MRI-targeted biopsy according to PSA level in biopsy-naive patients: a propensity score matching analysis. *Clin Genitourin Cancer* 2019; 17: e19–e25.

72. Azizi S, Bayat S, Yan P, *et al*. Deep recurrent neural networks for prostate cancer detection: analysis of temporal enhanced ultrasound. *IEEE Trans Med Imaging* 2018; 37: 2695–2703.

73. Schimmöller L, Blondin D, Arsov C, *et al*. MRI-guided in-bore biopsy: differences between prostate cancer detection and localization in primary and secondary biopsy settings. *Am J Roentgenol* 2016; 206: 92–99.

74. Yang T, Zhang L, Chen Y, *et al*. The predictive efficacy of hypoechoic lesion in ultrasound for prostate cancer in Chinese people: five-year experience in a moderated 10-core transperineal prostate biopsy procedure. *Oncotarget* 2017; 8: 79433–79440.

75. Woo S, Suh CH, Kim SY, *et al*. Shear-wave elastography for detection of prostate cancer: a systematic review and diagnostic meta-analysis. *Am J Roentgenol* 2017; 209: 806–814.

76. Sigrist RM, Liau J, Kaffas AE, *et al*. Ultrasound elastography: review of techniques and clinical applications. *Theranostics* 2017; 7: 1303–1329.

77. Liau J, Goldberg D and Arif-Tiwari H. Prostate cancer detection and diagnosis: role of ultrasound with MRI correlates. *Curr Radiol Rep* 2019; 7: 7.

78. Ashrafi AN, Nassiri N, Gill IS, *et al*. Contrast-enhanced transrectal ultrasound in focal therapy for prostate cancer. *Curr Urol Rep* 2018; 19: 87.

79. Harland N and Stenzl A. Micro-ultrasound: a way to bring imaging for prostate cancer back to urology. *Prostate Int* 2021; 9: 61–65.

80. Wildeboer RR, Mannaerts CK, van Sloun RJG, *et al*. Automated multiparametric localization of prostate cancer based on b-mode, shear-wave elastography, and contrast-enhanced ultrasound radiomics. *Eur Radiol* 2020; 30: 806–815.

81. Cornud F, Lefevre A, Flam T, *et al*. MRI-directed high-frequency (29MhZ) TRUS-guided biopsies: initial results of a single-center study. *Eur Radiol* 2020; 30: 4838–4846.

82. Lughezzani G, Saita A, Lazzeri M, *et al*. Comparison of the diagnostic accuracy of micro-ultrasound and magnetic resonance imaging/ultrasound fusion targeted biopsies for the diagnosis of clinically significant prostate cancer. *Eur Urol Oncol* 2019; 2: 329–332.

83. Eure G, Fanney D, Lin J, *et al*. Comparison of conventional transrectal ultrasound, magnetic resonance imaging, and micro-ultrasound for visualizing prostate cancer in an active surveillance population: a feasibility study. *Can Urol Assoc J* 2019; 13: E70–E77.

84. Klotz CL. Can high resolution micro-ultrasound replace MRI in the diagnosis of prostate cancer? *Eur Urol Focus* 2020; 6: 419–423.

85. Klotz L, Lughezzani G, Maffei D, *et al*. Comparison of micro-ultrasound and multiparametric magnetic resonance imaging for prostate cancer: a multicenter, prospective analysis. *Can Urol Assoc J* 2021; 15: E11–E16.

86. Rohrbach D, Wodlinger B, Wen J, *et al*. High-frequency quantitative ultrasound for imaging prostate cancer using a novel micro-ultrasound scanner. *Ultrasound Med Biol* 2018; 44: 1341–1354.

87. Shao Y, Wang J, Wodlinger B, *et al*. Improving prostate cancer (PCa) classification performance by using three-player minimax game to reduce data source heterogeneity. *IEEE Trans Med Imaging* 2020; 39: 3148–3158.

88. Hassan MR, Islam MF, Uddin MZ, *et al*. Prostate cancer classification from ultrasound and MRI images using deep learning based explainable artificial intelligence. *Future Gener Comp Sy* 2022; 127: 462–472.

89. Sedghi A, Pesteie M, Javadi G, *et al*. Deep neural maps for unsupervised visualization of high-grade cancer in prostate biopsies. *Int J Comput Assist Radiol Surg* 2019; 14: 1009–1016.

90. Azizi S, Van Woudenberg N, Sojoudi S, *et al*. Toward a real-time system for temporal enhanced ultrasound- guided prostate biopsy. *Int J Comput Assist Radiol Surg* 2018; 13: 1201–1209.

91. Azizi S, Yan P, Tahmasebi A, *et al*. Learning from noisy label statistics: detecting high grade prostate cancer in ultrasound guided biopsy. In: Frangi AF, Schnabel JA, Davatzikos C, *et al*. (eds) *Medical image computing and computer assisted intervention – MICCAI 2018*. Cham: Springer, 2018, pp. 21–29.

92. Han SM, Lee HJ and Choi JY. Computer-aided prostate cancer detection using texture features and clinical features in ultrasound image. *J Digit Imaging* 2008; 21(Suppl. 1): S121–S133.

93. Moradi M, Abolmaesumi P, Siemens R, *et al*. P6C-7 ultrasound RF time series for detection of prostate cancer: feature selection and frame rate analysis. In: *2007 IEEE ultrasonics symposium proceedings*, New York, 28–31 October 2007, pp. 2493–2496. New York: IEEE.

94. Imani F, Abolmaesumi P, Gibson E, *et al*. Computer-aided prostate cancer detection using ultrasound RF time series: in vivo feasibility study. *IEEE Trans Med Imaging* 2015; 34: 2248–2257.

95. Epstein JI. An update of the Gleason grading system. *J Urol* 2010; 183: 433–440.

96. Bulten W, Pinckaers H, van Boven H, *et al*. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol* 2020; 21: 233–241.

97. Egevad L, Delahunt B, Berney DM, *et al*. Utility of pathology Imagebase for standardisation of prostate cancer grading. *Histopathology* 2018; 73: 8–18.

98. Ozkan TA, Eruyar AT, Cebeci OO, *et al*. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol* 2016; 50: 420–424.

99. Doyle S, Feldman M, Tomaszewski J, *et al*. A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Trans Biomed Eng* 2010; 59: 1205–1218.

100. Litjens G, Sánchez CI, Timofeeva N, *et al*. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 2016; 6: 26286.

101. Lucas M, Jansen I, Savci-Heijink CD, *et al*. Deep learning for automatic Gleason pattern classification for grade group determination

of prostate biopsies. *Virchows Arch* 2019; 475: 77–83.

102. Campanella G, Hanna MG, Geneslaw L, *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019; 25: 1301–1309.

103. Ström P, Kartasalo K, Olsson H, *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* 2020; 21: 222–232.

104. Bulten W, Balkenhol M, Belinga JJA, *et al.* Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. *arXiv preprint arXiv* 2020: 200204500, https://arxiv.org/ftp/arxiv/papers/2002/2002.04500.pdf

105. Marginean F, Arvidsson I, Simoulis A, *et al.* An artificial intelligence–based support tool for automation and standardisation of Gleason grading in prostate biopsies. *Eur Urol Focus* 2021; 7: 995–1001.

106. Kott O, Linsley D, Amin A, *et al.* Development of a deep learning algorithm for the histopathologic diagnosis and Gleason grading of prostate cancer biopsies: a pilot study. *Eur Urol Focus* 2021; 7: 347–351.

107. Nagpal K, Foote D, Tan F, *et al.* Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol* 2020; 6: 1372–1380.

108. Li J, Speier W, Ho KC, *et al.* An EM-based semi-supervised deep learning approach for semantic segmentation of histopathological images from radical prostatectomies. *Comput Med Imaging Graph* 2018; 69: 125–133.

109. Romero Lauro G, Cable W, Lesniak A, *et al.* Digital pathology consultations – a new era in digital imaging, challenges and practical applications. *J Digit Imaging* 2013; 26: 668–677.

110. Deng J, Dong W, Socher R, *et al.* ImageNet: a large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*, Miami, FL, 20–25 June 2009, pp. 248–255. New York: IEEE

111. Arvaniti E, Fricker KS, Moret M, *et al.* Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Sci Rep* 2018; 8: 12054.

112. Nir G, Hor S, Karimi D, *et al.* Automatic grading of prostate cancer in digitized histopathology images: learning from multiple experts. *Med Image Anal* 2018; 50: 167–180.

113. Nir G, Karimi D, Goldenberg SL, *et al.* Comparison of artificial intelligence techniques to evaluate performance of a classifier for automatic grading of prostate cancer from digitized histopathologic images. *JAMA Netw Open* 2019; 2: e190442

114. Pinckaers H, Bulten W, van der Laak J, *et al.* Detection of prostate cancer in whole-slide images through end-to-end training with image-level labels. *IEEE Trans Med Imaging* 2021; 40: 1817–1826.

115. Raciti P, Sue J, Ceballos R, *et al.* Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Mod Pathol* 2020; 33: 2058–2066.

116. Perincheri S, Levi AW, Celli R, *et al.* An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. *Mod Pathol* 2021; 34: 1588–1595.

117. Steiner DF, Nagpal K, Sayres R, *et al.* Evaluation of the use of combined artificial intelligence and pathologist assessment to review and grade prostate biopsies. *JAMA Netw Open* 2020; 3: e2023267

118. Bulten W, Kartasalo K, Chen PHC, *et al.* Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the panda challenge. *Nat Med* 2022; 28: 154–163.

119. Paige. Paige receives first ever FDA approval for AI product in digital pathology, 2021, https://www.businesswire.com/news/home/20210922005369/en/Paige-Receives-First-Ever-FDA-Approval-for-AI-Product-in-Digital-Pathology#:~:text=As%20a%20novel%20technology%2C%20Paige,FullFocus%E2%84%A2%20digital%20pathology%20viewer.

120. Sonn GA, Margolis DJ and Marks LS. Target detection: magnetic resonance imaging-ultrasound fusion–guided prostate biopsy. *Urol Oncol* 2014; 32: 903–911.

121. Soerensen SJC, Fan RE, Seetharaman A, *et al.* Deep learning improves speed and accuracy of prostate gland segmentations on magnetic resonance imaging for targeted biopsy. *J Urol* 2021; 206: 604–612.

122. Soerensen SJC, Fan R, Seetharaman A, *et al.* ProGNet: prostate gland segmentation on MRI with deep learning. In: Išgum I and Landman BA (eds) *Medical imaging 2021: image processing*, vol. 11596. Bellingham, WA: International Society for Optics and Photonics, 2021, p. 115962R.

123. Cheng R, Lay N, Roth HR, *et al*. Fully automated prostate whole gland and central gland segmentation on MRI using holistically nested networks with short connections. *J Med Imaging* 2019; 6: 024007.

124. Zabihollahy F, Schieda N, Krishna Jeyaraj S, *et al*. Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets. *Med Phys* 2019; 46: 3078–3090.

125. Jia H, Xia Y, Song Y, *et al*. 3D APA-Net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MR images. *IEEE Trans Med Imaging* 2019; 39: 447–457.

126. Liu Q, Dou Q, Yu L, *et al*. MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. *IEEE Trans Med Imaging* 2020; 39: 2713–2724.

127. Wang B, Lei Y, Tian S, *et al*. Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation. *Med Phys* 2019; 46: 1707–1718.

128. Jensen C, Sørensen KS, Jørgensen CK, *et al*. Prostate zonal segmentation in 1.5T and 3T T2W MRI using a convolutional neural network. *J Med Imaging* 2019; 6: 014501.

129. Tian Z, Liu L, Zhang Z, *et al*. PSNet: prostate segmentation on MRI based on a convolutional neural network. *J Med Imaging* 2018; 5: 021208.

130. Tian Z, Li X, Zheng Y, *et al*. Graph-convolutional-network-based interactive prostate segmentation in MR images. *Med Phys* 2020; 47: 4164–4176.

131. Turkbey B, Fotin SV, Huang RJ, *et al*. Fully automated prostate segmentation on MRI: comparison with manual segmentation methods and specimen volumes. *Am J Roentgenol* 2013; 201: W720–W729.

132. Lee DK, Sung DJ, Kim CS, *et al*. Three-dimensional convolutional neural network for prostate MRI segmentation and comparison of prostate volume measurements by use of artificial neural network and ellipsoid formula. *Am J Roentgenol* 2020; 214: 1229–1238.

133. Sanford TH, Zhang L, Harmon SA, *et al*. Data augmentation and transfer learning to improve generalizability of an automated prostate segmentation model. *Am J Roentgenol* 2020; 215: 1403–1410.

134. Litjens G, Toth R, van de Ven W, *et al*. Evaluation of prostate segmentation algorithms for MRI: the promise12 challenge. *Med Image Anal* 2014; 18: 359–373.

135. Turkbey B, Mani H, Shah V, *et al*. Multiparametric 3T prostate magnetic resonance imaging to detect cancer: histopathological correlation using prostatectomy specimens processed in customized magnetic resonance imaging based molds. *J Urol* 2011; 186: 1818–1824.

136. Healthcare G. PROView Body 2021, https://www.gehealthcare.com/products/magnetic-resonance-imaging/signa-works/proview-body

137. Cortechsai. OnQ Prostate, 2021, https://www.cortechs.ai/products/onq-prostate/

138. Quantib. Quantib Prostate, 2021, https://www.quantib.com/en/solutions/quantib-prostate

139. Quibim. QP-Prostate, 2021, https://qp-prostate.quibim.com/

140. Zhan Y and Shen D. Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method. *IEEE Trans Med Imaging* 2006; 25: 256–272.

141. Yang X, Rossi PJ, Jani AB, *et al*. 3D transrectal ultrasound (TRUS) prostate segmentation based on optimal feature learning framework. In: Styner MA and Angelini ED (eds) *Medical imaging 2016: image processing*, vol. 9784. Bellingham, WA: International Society for Optics and Photonics, SPIE, 2016, pp. 654–660.

142. Li X, Li C, Fedorov A, *et al*. Segmentation of prostate from ultrasound images using level sets on active band and intensity variation across edges. *Med Phys* 2016; 43: 3090–3103.

143. Ghose S, Oliver A, Martí R, *et al*. A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. *Comput Methods Programs Biomed* 2012; 108: 262–287.

144. Anas EMA, Nouranian S, Mahdavi SS, *et al*. Clinical target-volume delineation in prostate brachytherapy using residual neural networks. In: Descoteaux M, Maier-Hein L, Franz A, *et al*. (eds) *Medical image computing and computer assisted intervention MICCAI 2017*. Cham: Springer, 2017, pp. 365–373.

145. Ghavami N, Hu Y, Bonmati E, *et al*. Integration of spatial information in convolutional neural networks for automatic segmentation of intraoperative transrectal ultrasound images. *J Med Imaging* 2018; 6: 011003.

146. van Sloun RJG, Wildeboer RR, Mannaerts CK, *et al*. Deep learning for real-time, automatic, and scanner-adapted prostate (zone) segmentation of transrectal ultrasound, for example, magnetic resonance imaging;

transrectal ultrasound fusion prostate biopsy. *Eur Urol Focus* 2021; 7: 78–85.

147. Jaouen V, Bert J, Mountris KA, *et al.* Prostate volume segmentation in TRUS using hybrid edge-Bhattacharyya active surfaces. *IEEE Trans Biomed Eng* 2019; 66: 920–933.

148. Girum KB, Lalande A, Hussain R, *et al.* A deep learning method for real-time intraoperative us image segmentation in prostate brachytherapy. *Int J Comput Assist Radiol Surg* 2020; 15: 1467–1476.

149. Wang Y, Dou H, Hu X, *et al.* Deep attentive features for prostate segmentation in 3D transrectal ultrasound. *IEEE Trans Med Imaging* 2019; 38: 2768–2778.

150. Orlando N, Gillies DJ, Gyacskov I, *et al.* Deep learning-based automatic prostate segmentation in 3D transrectal ultrasound images from multiple acquisition geometries and systems. In: Fei B and Linte CA (eds) *Medical imaging 2020: image-guided procedures, robotic interventions, and modeling*, vol. 11315. Bellingham, WA: International Society for Optics and Photonics, SPIE, 2020, pp. 651–656.

151. Lei Y, Tian S, He X, *et al.* Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. *Med Phys* 2019; 46: 3194–3206.

152. Zeng Q, Samei G, Karimi D, *et al.* Prostate segmentation in transrectal ultrasound using magnetic resonance imaging priors. *Int J Comput Assist Radiol Surg* 2018; 13: 749–757.

153. Karimi D, Zeng Q, Mathur P, *et al.* Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images. *Med Image Anal* 2019; 57: 186–196.

154. Anas EMA, Mousavi P and Abolmaesumi P. A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy. *Med Image Anal* 2018; 48: 107–116.

155. Yang X, Yu L, Wu L, *et al.* Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: *Proceedings of the thirty-first AAAI conference on artificial intelligence (AAAI'17)*, San Francisco, CA, 4–9 February 2017, pp. 1633–1639. Palo Alto, CA: AAAI Press.

156. Baum ZM, Hu Y and Barratt DC. Multimodality biomedical image registration using free point transformer networks. In: Hu Y, Licandro R, Noble JA, *et al.* (eds) *Medical ultrasound, and preterm, perinatal and paediatric image analysis*. Cham: Springer, 2020, pp. 116–125.

157. Baum ZMC, Hu Y and Barratt DC. Real-time multimodal image registration with partial intraoperative point-set data. *Med Image Anal* 2021; 74: 102231.

158. Hu Y, Gibson E, Barratt DC, *et al.* Conditional segmentation in lieu of image registration. In: Shen D, Liu T, Peters TM, *et al.* (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2019, pp. 401–409.

159. Saeed SU, Taylor ZA, Pinnock MA, *et al.* Prostate motion modelling using biomechanically-trained deep neural networks on unstructured nodes. In: Martel AL, Abolmaesumi P, Stoyanov D, *et al.* (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2020, pp. 650–659.

160. Zeng Q, Fu Y, Tian Z, *et al.* Label-driven magnetic resonance imaging (MRI)-transrectal ultrasound (TRUS) registration using weakly supervised learning for MRI-guided prostate radiotherapy. *Phys Med Biol* 2020; 65: 135002.

161. Hu Y, Modat M, Gibson E, *et al.* Weakly-supervised convolutional neural networks for multimodal image registration. *Med Image Anal* 2018; 49: 1–13.

162. Ghavami N, Hu Y, Gibson E, *et al.* Automatic segmentation of prostate MRI using convolutional neural networks: investigating the impact of network architecture on the accuracy of volume measurement and MRI- ultrasound registration. *Med Image Anal* 2019; 58: 101558.

163. Hu Y, Gibson E, Ghavami N, *et al.* Adversarial deformation regularization for training image registration neural networks. In: Frangi AF, Schnabel JA, Davatzikos C, *et al.* (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2018, pp. 774–782.

164. Fu Y, Lei Y, Wang T, *et al.* Deformable MRI-US image registration using biomechanically constrained shape matching network for image-guided prostate intervention. In: Byram BC and Ruiter NV (eds) *Medical imaging 2021: ultrasonic imaging and tomography*, vol. 11602. Bellingham, WA: International Society for Optics and Photonics, 2021, p. 116020K.

165. Haskins G, Kruecker J, Kruger U, *et al.* Learning deep similarity metric for 3D MR–TRUS image registration. *Int J Comput Assist Radiol Surg* 2019; 14: 417–425.

166. Hu Y, Gibson E, Vercauteren T, *et al.* Intraoperative organ motion models with an

ensemble of conditional generative adversarial networks. In: Descoteaux M, Maier-Hein L, Franz A, *et al.* (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2017, pp. 368–376.

167. Qi CR, Su H, Mo K, *et al.* PointNet: deep learning on point sets for 3D classification and segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, 21–26 July 2017, pp. 652–660. New York: IEEE.

168. Guo H, Xu X, Xu S, *et al.* End-to-end ultrasound frame to volume registration. In: de Bruijne M, Cattin PC, Cotin S, *et al.* (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2021, pp. 56–65.

169. Hu Y, Modat M, Gibson E, *et al.* Label-driven weakly-supervised learning for multimodal deformable image registration. In: *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, Washington, DC, 4–7 April 2018, pp. 1070–1074. New York: IEEE.

170. Guo H, Kruger M, Xu S, *et al.* Deep adaptive registration of multi-modal prostate images. *Comput Med Imaging Graph* 2020; 84: 101769.

171. Zeng Q, Fu Y, Jeong J, *et al.* Weakly non-rigid MR-TRUS prostate registration using fully convolutional and recurrent neural networks. In: Išgum I and Landman BA (eds) *Medical imaging 2020: image processing*, vol. 11313. Bellingham, WA: International Society for Optics and Photonics, 2020, p. 113132Y.

172. Song X, Guo H, Xu X, *et al.* Cross-modal attention for MRI and ultrasound volume registration. In: de Bruijne M, Cattin PC, Cotin S, *et al.* (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2021, pp. 66–75.

173. Eigen. Artemis, 2018, https://www.innomedicus.com/urology/artemis/

174. Philips. UroNav, https://www.philips.co.uk/healthcare/education-resources/publications/hotspot/uronav/

175. Chappelow J, Bloch BN, Rofsky N, *et al.* Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information. *Med Phys* 2011; 38: 2005–2018.

176. Ward AD, Crukley C, McKenzie CA, *et al.* Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil. *Radiology* 2012; 263: 856–864.

177. Kalavagunta C, Zhou X, Schmechel SC, *et al.* Registration of in vivo prostate MRI and pseudo-whole mount histology using local affine transformations guided by internal structures (LATIS). *J Magn Reson Imaging* 2015; 41: 1104–1114.

178. Reynolds HM, Williams S, Zhang A, *et al.* Development of a registration framework to validate MRI with histology for prostate focal therapy. *Med Phys* 2015; 42: 7078–7089.

179. Li L, Pahwa S, Penzias G, *et al.* Co-registration of ex vivo surgical histopathology and in vivo T2 weighted MRI of the prostate via multi-scale spectral embedding representation. *Sci Rep* 2017; 7: 8717.

180. Losnegård A, Reisæter L, Halvorsen OJ, *et al.* Intensity-based volumetric registration of magnetic resonance images and whole-mount sections of the prostate. *Comput Med Imaging Graph* 2018; 63: 24–30.

181. Wu HH, Priester A, Khoshnoodi P, *et al.* A system using patient-specific 3D-printed molds to spatially align in vivo MRI with ex vivo MRI and whole-mount histopathology for prostate cancer research. *J Magn Reson Imaging* 2019; 49: 270–279.

182. Rusu M, Shao W, Kunder CA, *et al.* Registration of presurgical MRI and histopathology images from radical prostatectomy via RAPSODI. *Med Phys* 2020; 47: 4177–4188.

183. Shao W, Banh L, Kunder CA, *et al.* ProsRegNet: a deep learning framework for registration of MRI and histopathology images of the prostate. *Med Image Anal* 2021; 68: 101919.

184. Sood RR, Shao W, Kunder C, *et al.* 3D registration of pre-surgical prostate MRI and histopathology images via super-resolution volume reconstruction. *Med Image Anal* 2021; 69: 101957.

185. Shao W, Bhattacharya I, Soerensen SJC, *et al.* Weakly supervised registration of prostate MRI and histopathology images. In: de Bruijne M, Cattin PC, Cotin S, *et al.* (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2021, pp. 98–107.

186. Nir G, Sahebjavaher RS, Kozlowski P, *et al.* Registration of whole-mount histology and volumetric imaging of the prostate using particle

filtering. *IEEE Trans Med Imaging* 2014; 33: 1601–1613.

187. Zamboglou C, Kramer M, Kiefer S, *et al.* The impact of the co-registration technique and analysis methodology in comparison studies between advanced imaging modalities and whole-mount-histology reference in primary prostate cancer. *Sci Rep* 2021; 11: 5836.

188. Samavati N, McGrath DM, Lee J, *et al.* Biomechanical model-based deformable registration of MRI and histopathology for clinical prostatectomy. *J Pathol Inform* 2011; 2: S10.

189. Stille M, Smith EJ, Crum WR, *et al.* 3D reconstruction of 2D fluorescence histology images and registration with in vivo MR images: application in a rodent stroke model. *J Neurosci Methods* 2013; 219: 27–40.

190. Rusu M, Kunder C, Fan R, *et al.* Framework for the co-registration of MRI and histology images in prostate cancer patients with radical prostatectomy. In: Angelini ED and Landman BA (eds) *Medical imaging 2019: image processing*, vol. 10949. Bellingham, WA: International Society for Optics and Photonics, 2019, p. 109491P.

191. Russakovsky O, Deng J, Su H, *et al.* ImageNet large scale visual recognition challenge. *Int J Comput Vision* 2015; 115: 211–252.

192. Gibson E, Hu Y, Ghavami N, *et al.* Inter-site variability in prostate segmentation accuracy using deep learning. In: Frangi AF, Schnabel JA, Davatzikos C, *et al.* (eds) *International conference on medical image computing and computer-assisted intervention*. Cham: Springer, 2018, pp. 506–514.

193. Sheller MJ, Edwards B, Reina GA, *et al.* Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 2020; 10: 12598.

194. Rieke N, Hancox J, Li W, *et al.* The future of digital health with federated learning. *NPJ Digit Med* 2020; 3: 119–117.

195. Sheller MJ, Reina GA, Edwards B, *et al.* Multi-institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation. In: Crimi A, Bakas S, Kuijf H, *et al.* (eds) *International MICCAI Brainlesion workshop*. Cham: Springer, 2019, pp. 92–104.

196. Chen L, Bentley P, Mori K, *et al.* Self-supervised learning for medical image analysis using image context restoration. *Med Image Anal* 2019; 58: 101539.

197. Zhou Y, He X, Huang L, *et al.* Collaborative learning of semi-supervised segmentation and classification for medical images. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Long Beach, CA, 15–20 June 2019, pp. 2079–2088. New York: IEEE.

198. Wang D, Zhang Y, Zhang K, *et al.* FocalMix: semi-supervised learning for 3D medical image detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Seattle, WA, 13–19 June 2020, pp. 3951–3960. New York: IEEE.

199. Li Y, Fu Y, Yang Q, *et al.* Few-shot image segmentation for cross-institution male pelvic organs using registration-assisted prototypical learning. *arXiv preprint arXiv* 2022: 220106358, https://arxiv.org/pdf/2209.05160.pdf

200. Gaur S, Lay N, Harmon SA, *et al.* Can computer-aided diagnosis assist in the identification of prostate cancer on prostate MRI? A multi-center, multi-reader investigation. *Oncotarget* 2018; 9: 33804.

201. Schelb P, Wang X, Radtke JP, *et al.* Simulated clinical deployment of fully automatic deep learning for clinical prostate MRI assessment. *Eur Radiol* 2021; 31: 302–313.

202. Litjens G, Debats O, Barentsz J, *et al.* ProstateX challenge data. *The Cancer Imaging Archive*, 2017, http://doi.org/10.7937/K9TCIA.2017.MURS5CL

203. Bloch N, Madabhushi A, Huisman H, *et al.* NCI-ISBI 2013 challenge: automated segmentation of prostate structures. *The Cancer Imaging Archive*, 2015, http://doi.org/10.7937/K9/TCIA.2015.zF0vlOPv

204. Schmidkonz C, Cordes M, Schmidt D, *et al.* 68Ga-PSMA-11 PET/CT-derived metabolic parameters for determination of whole-body tumor burden and treatment response in prostate cancer. *Eur J Nucl Med Mol Imaging* 2018; 45: 1862–1872.

205. Hofman MS, Lawrentschuk N, Francis RJ, *et al.* Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. *Lancet* 2020; 395: 1208–1216.

206. Emmett L, Buteau J, Papa N, *et al.* The additive diagnostic value of prostate-specific membrane antigen positron emission tomography computed tomography to multiparametric magnetic resonance imaging triage in the diagnosis of prostate cancer (primary): a prospective multicentre study. *Eur Urol* 2021; 80: 682–689.

207. Roberts MJ, Morton A, Donato P, *et al.* 68Ga-PSMA PET/CT tumour intensity pre-operatively predicts adverse pathological outcomes and progression-free survival in localised prostate cancer. *Eur J Nucl Med Mol Imaging* 2021; 48: 477–482.

208. Hope TA, Eiber M, Armstrong WR, *et al.* Diagnostic accuracy of 68Ga-PSMA-11 pet for pelvic nodal metastasis detection prior to radical prostatectomy and pelvic lymph node dissection: a multicenter prospective phase 3 imaging trial. *JAMA Oncol* 2021; 7: 1635–1642.

209. Franklin A, Yaxley WJ, Raveenthiran S, *et al.* Histological comparison between predictive value of preoperative 3-T multiparametric MRI and 68Ga-PSMA PET/CT scan for pathological outcomes at radical prostatectomy and pelvic lymph node dissection for prostate cancer. *BJU Int* 2021; 127: 71–79.

210. Yi Z, Hu S, Lin X, *et al.* Machine learning-based prediction of invisible intraprostatic prostate cancer lesions on 68 Ga-PSMA-11 PET/CT in patients with primary prostate cancer. *Eur J Nucl Med Mol Imaging* 2022; 49: 1523–1534.

211. Kendrick J, Francis R, Hassan GM, *et al.* Fully automatic prognostic biomarker extraction from metastatic prostate lesion segmentations in whole-body [68Ga]Ga-PSMA-11 PET/CT images. *Eur J Nucl Med Mol Imaging*. Epub ahead of print 17 August 2022. DOI: 10.1007/s00259-022-05927-1.