

<https://doi.org/10.32604/cju.2026.077632>

REVIEW

The role of artificial intelligence in urological malignancies: an overview of systematic reviews

Vasileios Sakalis,^{1,2*} Eftichia Chatzigriva,^{1,2} Karl H. Pang,³ Bhavan P. Rai,⁴ Isaak Filippidis,^{1,2} Lisa Moris,⁵ Maria Chalkidou,^{1,2} Yuhong Yuan,⁶ Michael Bussmann,⁷ Dimitrios Papanikolaou,⁸ James N'Dow,⁹ Muhammad Imran Omar⁹

¹Department of Urology, Hippokrateion Hospital of Thessaloniki, Thessaloniki, Greece

²Division of Urology, Innovative Surgical and Urological Research Hub, Thessaloniki, Greece

³Department of Urology, University College London Hospitals NHS Foundation Trust, London, UK

⁴Department of Urology, Freeman Hospital, the Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

⁵Department of Urology, University Hospitals Leuven, Leuven, Belgium

⁶Department of Medicine, McMaster University, Hamilton, ON, Canada

⁷Institute of Radiation Physics, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany

⁸Department of Urology, Papageorgiou Hospital of Thessaloniki, Thessaloniki, Greece

⁹Academic Urology Unit, University of Aberdeen, Aberdeen, UK

SAKALIS V, CHATZIGRIVA E, PANG KH, RAI BP, FILIPIDIS I, MORIS L, CHALKIDOU M, YUAN Y, BUSSMANN M, PAPANIKOLAOU D, N'DOW J, OMAR MI. The role of artificial intelligence in urological malignancies: an overview of systematic reviews. *Can J Urol* 2026;33(5):1061–1087.

Objectives: Artificial Intelligence (AI) has the potential to transform clinical medicine by enhancing diagnostic precision, prognostic accuracy, and personalized decision making. In urological oncology, AI technologies are increasingly integrated across diagnostic, prognostic, and therapeutic pathways. This overview of systematic reviews aims to synthesise the current evidence on the performance and clinical utility of AI applications in urological malignancies.

Methods: A systematic search was conducted of major bibliographic databases from 1974 to October 2024 to identify systematic reviews (SRs) evaluating AI-based models in urological cancers. We followed the Cochrane methodology of Overviews of Reviews. Outcomes of interest included diagnostic accuracy metrics (area under the receiver operating characteristic curve [AUC-ROC], sensitivity, specificity, and accuracy) and predictive performance for treatment-related outcomes.

Results: A total of 67 SRs encompassing 1139 primary studies were included. RoB was assessed as low in 21 reviews, high in 41, and unclear in 5. Forty-five SRs evaluated AI in prostate cancer (PCa). AI models demonstrated high accuracy (88–93%) in distinguishing benign from malignant lesions, and outperformed human interpretation alone in

predicting clinically significant PCa (AUC: 0.933 vs. AUC: 0.82–0.87). Performance in Gleason score classification was also very strong (AUC: 0.72–0.99, Accuracy: 70.8–93.0%). Sixteen SRs assessed AI in Renal carcinoma (RCa), reporting accurate subtype classification and tumor grade prediction, with some models exceeding 90% accuracy. Sixteen SRs focused on Bladder Cancer (BCa), showing improved detection of muscle-invasive disease and recurrence risk prediction compared with the conventional approach. Key limitations included heterogeneity in validation strategies, overlap of primary studies across reviews, and potential oversampling due to synthetic or augmented data.

Conclusions: AI-based models consistently demonstrate improved diagnostic and predictive performance across urological malignancies and hold promise for enhancing clinical decision-making and patient outcomes. At present, AI tools should be regarded as decision-support systems that complement, rather than replace, clinical expertise. Future research should emphasize prospective validation, standardized outcome definitions, and evaluation of real-world clinical impact. Robust and harmonized evaluation frameworks are also essential to address challenges related to data quality, privacy, generalizability, and ongoing regulatory and legal uncertainty, thereby enabling safe and equitable integration of AI into routine urological cancer care.

Key Words: artificial intelligence, urological malignancies, machine learning, deep Learning, prostate cancer, kidney cancer, bladder cancer

Received date 14 December 2025

Accepted for publication 02 March 2026

Published online 30 October 2026

*Corresponding Author: Vasileios Sakalis.

Email: vsakkalis@hotmail.com

Introduction

The advancement of Artificial Intelligence (AI) technologies has spurred a transformative era in healthcare. AI has the potential to revolutionize patient care by using sophisticated algorithms, natural language processing, and computer vision to accelerate diagnosis, improve accuracy, and personalize treatment.^{1,2}

AI is closely intertwined with Machine Learning (ML) and Deep Learning (DL); ML is a subset of AI that leverages algorithms and statistical models to enable computers to perform tasks without being explicitly programmed. These systems are designed to learn from data and improve over time, thereby enhancing their performance. DL is a form of ML that employs neural networks attempting to mimic the structure of and function of the human brain to identify and recognize patterns.¹ AI algorithms generally follow a structured series of steps and underlying principles.³ The process begins with data collection, during which the AI system learns from standardized and pre-processed information. The model is then trained and assessed using appropriate performance metrics, followed by validation on independent datasets to ensure generalizability. Once validated, the AI model is integrated into clinical workflows, with regular updates implemented to enhance its accuracy over time (Figure 1). Radiomics, a key component in this process, involves extracting quantitative features from medical images and combining them with clinical, laboratory, and metadata to improve diagnostic, prognostic, and predictive performance.⁴

Emerging evidence indicates that AI models demonstrate strong diagnostic and predictive capabilities in urological malignancies, achieving accuracy levels comparable to those of experienced clinicians. Computer-aided diagnosis (CAD) systems provide an objective, non-invasive tool to support decision making, improve diagnostic efficacy, and increase inter-observer agreement. This topic has been extensively investigated, with a significant increase in related literature over the last years. Therefore, through this overview of Systematic Reviews (SRs), we aim to provide a comprehensive synthesis of the current literature and to elucidate the evolving role of AI in the diagnosis and management of patients with urological cancers. An Overview of SRs was chosen due to the rapid growth and heterogeneity of the existing literature, with multiple SRs addressing overlapping research questions, enabling a high-level synthesis and critical appraisal of the available evidence.

Evidence Acquisition

Search strategy

A systematic search was performed in EMBASE, MEDLINE, Cochrane Database of Systematic Reviews, and Clinical Trial databases from 1974 up to 31 October 2024.

The study protocol is available in PROSPERO (CRD42022382416). The search strategy is described in Supplementary Fig. S1 and Supplementary Material S1, including the following search terms: Artificial intelligence, machine learning, computer-aided diagnosis, prostate cancer, kidney cancer, bladder cancer, testicular cancer, penile cancer, systematic review.

Types of study designs

We included SRs focusing on AI applications in patients with urological malignancies. Literature reviews, conference abstracts, and studies on virtual reality were excluded.

Types of participants

The study involved patients aged ≥ 18 years, diagnosed with Prostate (PCa), Renal (RCa), Upper Urinary Tract-Urothelial (UUTUCa), Bladder (BCa), Testicular (TCa), or Penile Carcinoma (PenCa) in either primary or metastatic settings.

Types of interventions

We have assessed all AI technologies used in the management of urological patients, including AI stand-alone and AI-enhanced reading applications.

Outcome measures

The primary outcome was the performance of AI in urological malignancies. We used the area under the receiver-operating characteristic curve (AUC-ROC) to evaluate AI models. Outcomes are presented as a range of AUC-ROC values. AUC-ROC summarizes the overall diagnostic accuracy of a test with values from 0 (a perfectly inaccurate test) to 1 (a perfectly accurate test).⁵ Since no clear thresholds exist for AUC, the following rule has been proposed: AUC values between 0.5 and 0.69 were considered poor, between 0.7 and 0.79 acceptable, between 0.8 and 0.89 excellent, while values above 0.9 were considered as outstanding discrimination.⁵ The secondary outcomes were accuracy (Acc), sensitivity (Sen), and specificity (Spe) of AI models. In the case of meta-analyses, the effect estimate was reported with the heterogeneity value.

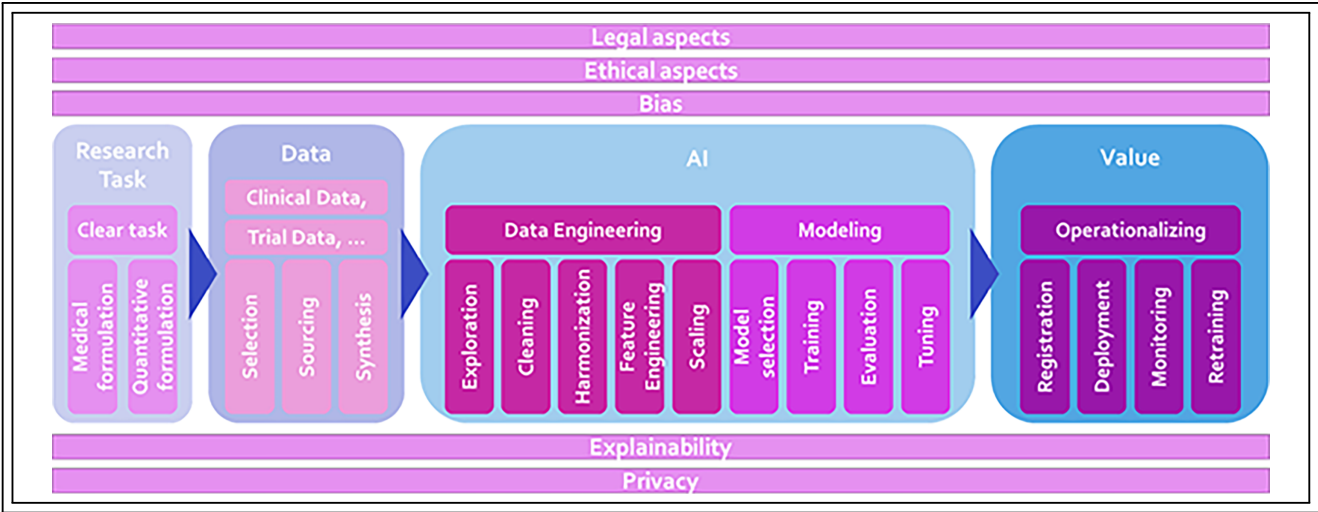


FIGURE 1. State-of-the-art workflow for creating AI models for medical applications. A specific research question or task is defined in medical terms and translated into a quantitative, mathematical formulation. Relevant data for training are selected, and data from various sources are synthesized as an input to the subsequent data engineering workflow. Data exploration includes the identification of useful data to solve the initial research task. Decluttering and harmonization of data aim to have a suitable set of raw input data that subsequent modeling can address the initial research task. Data scaling creates a single set of quantitatively comparable data objects such as a set of MRI images with same dynamic range, resolution and contrast collected from several hospitals. Before modeling, the data is divided into training data, validation data, and testing data. Then suitable AI-models are selected, trained and evaluated on the validation data set. Tuning of the model follows to optimize results until the predictive performance is reached. Before deploying, all models should be registered in a repository. The models are then deployed, meaning they are made available to work on new independent test data to identify initial biases and provide quantitative information on how good models perform. Repeated testing leads to retraining of the model to achieve better performance

Assessment of risk of bias

The methodological quality of the included SRs was assessed using the ROBIS (Risk of Bias in Systematic Reviews) tool.

Data analysis

The qualitative synthesis was considered appropriate for this overview. A subgroup analysis was planned; however, it was not possible due to the high heterogeneity of the evidence.

Evidence Synthesis

Identification of relevant systematic reviews

We have screened 599 records and identified 137 studies for full-text screening. The overlap across primary studies was evaluated by the citation matrix and the Corrected Covered Area (CCA).⁶ The CCA was calculated as the quotient of the frequency of repeated occurrences of an index publication by the product of index publications and reviews, reduced

by the number of index publications.⁷ Values up to 5 were considered to indicate a slight overlap, while values above 15 were considered as indicative of a high overlap.

Description and characteristics of the included systematic reviews

We included 67 SRs of 1139 primary studies (Supplementary Material S2). Thirty-four SRs focused on PCa,^{8–40} eleven SRs focused on RCa,^{41–51} nine SRs focused on BCa,^{52–60} and fourteen SRs evaluated studies of more than one organ.^{61–74} A meta-analysis was performed in ten SRs.^{16,20,30,39,41,43,47,49,52,59} The baseline characteristics of the included studies are presented in Table 1.

Methodological quality of included systematic reviews

According to the ROBIS criteria, twenty-one SRs^{9,11,13,20–22,27,30,32,35–38,41–43,49,50,52,57,59,65,68} were considered low-risk, five were of unclear-

TABLE 1. Baseline characteristics of included systematic reviews

Study ID	Number of included studies	Date of last search	Reported algorithms	Results from Training set/Validation set	RoB
Prostate Cancer Studies					
Sihotang et al. 2022 ⁸	6	19/11/21	DL, ANN, RNN	Training set	High risk
Sushentsev et al. 2022 ⁹	17	31/07/2021	CNN, SVM, RF	Training set/Validation set	Low risk
Michaely et al. 2022 ¹⁰	29	8/1/21	CNN	NR	High risk
Liberini et al. 2022 ¹¹	37	30/11/22	3D-CNN, Residual DL model, U-Net-CNN, Dixon-VIBE DL, Augmented GAN	Training set/Validation set	Low risk
Ghezzi et al. 2022 ¹²	76	31/12/20	NR	Training set/Validation set	High Risk
Castaldo et al. 2021 ¹³	37	24/02/2020	RF, RBM, PLS-DA, LDA, SVM, CART, ANN, DNN, CNN, DL	Training set/Validation set	Low Risk
Van Booven et al. 2021 ¹⁴	19	NR	ANN	Training set	High risk
Sugano et al. 2020 ¹⁵	40	30/04/2020	ANN, RF, XGB, RBM	Training set/Validation set	High risk
Cuocolo et al. 2020 ¹⁶	12	31/07/2019	LDA, Naive Bayesian, Linear Regression, RF, Logistic regression, SVM, CNN, DTL	Training set/Validation set	High risk
Almeida and Tavares 2020 ¹⁷	28	NR	2D-FCN, RF, 2D U-net, 3D V-net, CNN, regression, 3D ResNet encoder-decoder, pyramid convolutional blocks, 3D auto-encoder, 3D SNet, 3D V-net	Training set/Validation set	High risk
Castillo et al. 2020 ¹⁸	27	01/06/2020	CNN, LR, RF, LMM, SVM, k-nearest neighbor, Linear discriminant analysis	NR	High risk
Schröder and Kattan 2008 ¹⁹	8	30/07/2007	ANN	Training set/Validation set	High risk
Morozov et al. 2023 ²⁰	24	31/12/21	ANN, CNN, Bayesian, Adaboost	Training set/Validation set	Low risk
Quartuccio et al. 2021 ²¹	7	24/02/2021	NR	NR	Low risk
Syer et al. 2021 ²²	27	31/12/21	CNN, RF, SVM, GLMM, ANN	Training set/Validation set	Low risk
Lindgren Belal et al. 2023 ²³	26	21/03/2023	CNN, ANN, RF,	NR	High risk
Huynh et al. 2023 ²⁴	33	31/12/22	RF, LR, LASSO, SVM	Training set/Validation set	High risk
Naik et al. 2022 ²⁵	24	31/10/21	ANN, CNN, U-Net,	Training set/Validation set	High risk
Nematollahi et al. 2023 ²⁶	18	31/01/2023	SVM, KNN, LR, naive Bayes, RF, DL	NR	High risk
Rouviere et al. 2023 ²⁷	35	30/09/2022	NR	Training set/Validation set	Low risk
Spohn et al. 2021 ²⁸	67	1/1/21	NR	NR	High risk
Telecan et al. 2022 ²⁹	35	28/02/2022	NR	NR	Unclear risk
Xing et al. 2022 ³⁰	15	31/03/2019	NR	NR	Low risk

(Continued)

TABLE 1. (Continued)

Study ID	Number of included studies	Date of last search	Reported algorithms	Results from Training set/Validation set	RoB
Frewing et al. 2024 ³¹	65	31/03/2022	CNN	NR	High risk
Antolin et al. 2024 ³²	14	30/04/2024	LR, RF, SVM, CNN, DenseNet 201, Res-UNet, Attention-Gated TrumpetNet, 3D GLCM extractor with CNN, Uro-XGB, WORC Workflow, ResNet with TL	Training set/Validation set	Low risk
Calimano-Ramirez et al. 2023 ³³	11	31/08/2022	NR	Training set/Validation set	High risk
Faiella et al. 2023 ³⁴	16	31/08/2023	U-Net, SVM, CNN, LR, RF, LASSO, other Radiomics models	Training set/Validation set	High risk
Guerra et al. 2024 ³⁵	11	31/10/2023	LASSO, SVM, RF	Training set/Validation set	Low risk
Liu et al. 2024 ³⁶	11	31/07/2023	CNN, DL, ML	Training set/Validation set	Low risk
Liu et al. 2024 ³⁷	14	15/02/2024	SVM, ML, DL, RF, CNN, LR, kNN	Training set/Validation set	Low risk
Rabilloud et al. 2023 ³⁸	26	01/09/2023	SVM, CNN, ResNet, DenseNet, AlexNet, U-Net, Seg-Net, Inception V1, VGG-19	Training set/Validation set	Low risk
Wang et al. 2023 ³⁹	31	01/11/2022	SVM, DT, RF, XGB, CNN	Training set/Validation set	Unclear risk
Rabaan et al. 2023 ⁴⁰	16	NR	CNN	NR	High risk
Kirienko et al. 2021 ⁶¹	4*	24/07/2020	CNN	Training set/Validation set	High risk
Bazarkin et al. 2024 ⁶²	10*	2023	ANN, Bayesian Algorithms, RF	NR	High risk
Papalia et al. 2024 ⁶³	13*	31/12/2023	ANN, CNN, DNN, DL, LR, DT, XGB, MLP, RF, SVM, KNN, RPN, IFV, SORG, GBM, NBC, AdaBoost, Chan-Verse, Bayesian Model	NR	High risk
Schneider et al. 2022 ⁶⁴	3*	30/06/2021	CNN	Training set/Validation set	High risk
Wessels et al. 2022 ⁶⁵	16*	20/04/2021	Handcrafted feature extraction, CNN, Deep Feature Extraction	Training set/Validation set	Low risk
Prabhu et al. 2022 ⁶⁶	13*	31/05/2021	CNN, ML	NR	Unclear risk
Forsch et al. 2021 ⁶⁷	1*	31/01/2020	SVM	NR	High risk
Salem et al. 2021 ⁶⁸	61*	31/12/16	ANN, FNM, RBS, FRB, SVT, BN, DT	Training set/Validation set	Low risk
Checucci et al. 2020 ⁶⁹	26*	30/04/2020	ANN, CNN, RF,	Training set/Validation set	High risk
Chen et al. 2019 ⁷⁰	15*	NR	ANN, BN, CART, CNN, DT, LDA, LR, QDA, RF, SVM	Training set	High risk
Abbod et al. 2007 ⁷¹	15*	NR	ANN, LR, NFM, ANFIS, Bayesian Network	NR	High risk
Faiella et al. 2022 ⁷²	13*	31/12/2021	NR	Training set/Validation set	High risk

(Continued)

TABLE 1. (Continued)

Study ID	Number of included studies	Date of last search	Reported algorithms	Results from Training set/Validation set	RoB
Renal Cancer Studies					
Ursprung et al. 2020 ⁴¹	57	30/10/18	NR	NR	Low risk
Ferro et al. 2023 ⁴²	55	31/07/2022	ML, RF, SVM, CNN	NR	Low risk
Firouzabadi et al. 2023 ⁴³	6	31/12/2022	ML, SVM, RF, LR, CNN, LR, LASSO	Training set/Validation set	Low risk
Roussel et al. 2022 ⁴⁴	47	31/07/2021	LR	NR	High risk
Kocak et al. 2020 ⁴⁵	30	27/11/19	CNN, RF	Training set	Unclear risk
Pandey and Gupta 2021 ⁴⁶	28	NR	FCNN, CNN, 3D-FCNN, AlexNet, 2D U-Net, SE-Net, ResNeXTm U-Net, Hybrid V-Net, Cascade 3D U-Net, RF	NR	High risk
Muhlbauer et al. 2021 ⁴⁷	113	31/05/2020	NR	NR	Unclear risk
Bhandari et al. 2020 ⁴⁸	13	09/07/2020	LASSO, SVM	Training set/Validation set	High risk
Chandramohan et al. 2024 ⁴⁹	13	31/12/2023	CNN, MMDLM, CLAM	NR	Low risk
Khene et al. 2024 ⁵⁰	20	31/12/2022	SVM, RF, LR, LASSO, AdaBoost, CNN, ResNet, Alex-Net, DenseNet, Inception V1/V3, VGG-16/19	Training set/Validation set	Low risk
Xi et al. 2024 ⁵¹	38	01/09/2023	RF, SVM, DL, ML-based, Histogram Analysis, Inception V3	NR	High risk
Wessels et al. 2022 ⁵⁵	6*	20/04/2021	Handcrafted feature extraction, CNN, Deep Feature Extraction	Training set/Validation set	Low risk
Checucci et al. 2020 ⁶⁹	9*	30/04/2020	CNN	Training set/Validation set	High risk
Chen et al. 2019 ⁷⁰	2*	NR	ANN, CNN, CART, CNN, LDA, LR, Naïve Bayes, QDA, RF, SVM	Training set	High risk
Abbod et al. 2007 ⁷¹	1*	NR	ANN	NR	High risk
Tramanzoli et al. 2023 ⁷³	14*	09/06/2022	ML, SVM	NR	High risk
Bladder Cancer Studies					
Kozikowski et al. 2021 ⁵²	8	28/7/2020	ANN, SVM, RF, LR, LASSO	Training set/Validation set	Low risk
Cacciamani et al. 2020 ⁵³	12	31/1/2020	SVM, RF, LASSO, DL-CNN	Training set/Validation set	High risk
Huang et al. 2022 ⁵⁴	9	5/4/2022	SVM, RF, CNN, LASSO, LR,	Training set/Validation set	High risk
Liu et al. 2023 ⁵⁵	14	28/2/2022	ANN, LR	NR	High risk
Kwong et al. 2024 ⁵⁶	15	5/2/2024	SVM, ANN, RF, LASSO, DBN, NFM, Bayesian Network, VGG16, DeepSurv, RBFN	Training set/Validation set	High risk
He et al. 2024 ⁵⁷	17	30/10/2023	SVM, RF, LASSO, FGP-Net, ResNet, CNN, Naïve Bayes, Inception V3, AdaBoost, LR	Training set/Validation set	Low risk

(Continued)

TABLE 1. (Continued)

Study ID	Number of included studies	Date of last search	Reported algorithms	Results from Training set/Validation set	RoB
Khoraminia et al. 2023 ⁵⁸	33	31/3/2022	NR (DL and ML)	NR	High risk
Suartz et al. 2024 ⁵⁹	12	01/07/2023	DL, CNN, RF	NR	Low risk
Chan et al. 2021 ⁶⁰	6	12/12/2020	CNN, ResNet, U-Net, Le Net, EasyDL	Training set	High risk
Bazarkin et al. 2024 ⁶²	17*	2023	ANN, Bayesian Algorithms, RF,	NR	High risk
Checucci et al. 2020 ⁶⁹	4*	30/4/2020	ANN, Inception ResNet V2 Tool, CNN, RF, GAN,	Training set/Validation set	High risk
Salem et al. 2021 ⁶⁸	17*	31/12/2016	ANN, FNM, RBS, FRB, SVT, BN, DT	Training set/Validation set	Low risk
Chen et al. 2019 (BCa Data) ⁷⁰	7*	NR	ANN, CNN, CART, CNN, LDA, LR, Naïve Bayes, QDA, RF, SVM	Training set	High risk
Abbod et al. 2007 ⁷¹	8*	NR	ANN, LR, NFM, ANFIS, Bayesian Network	NR	High risk
Tramanzoli et al. 2023 ⁷³	4	9/6/2022	ML, SVM	NR	High risk
Drouin et al. 2013 ⁷⁴	2*	NR	ANN, LR, NFM	Training set/Validation set	High risk
Other Urological Cancer Studies					
Abbod et al. 2007 ⁷¹	1*	NR	ANN, LR, NFM, ANFIS, Bayesian Network	NR	High risk
Salem et al. 2021 ⁶⁸	2*	31/12/2016	ANN, FNM, RBS, FRB, SVT, BN, DT	Training set/Validation set	Low risk
Drouin et al. 2013 ⁷⁴	6*	NR	ANN, LR, NFM	Training set/Validation set	High risk
Wessels et al. 2022 ⁶⁵	16	20/4/2021	Handcrafted feature extraction, CNN, Deep Feature Extraction	Training set/Validation set	Low risk

Note. *Number of relevant studies. Abbreviations: 2D-FCN, 2-D Fully convolutional network; AI, Artificial Intelligence; ANFIS, Adaptive Neuro-Fuzzy Inference System; ANN, Artificial Neural Network; BN, Bayesian Network; CART, Classification and Regression Tree; CEUS, Contrast-Enhanced Ultrasound; CLAM, Clustering-constrained Attention Multiple Instance Learning; CNN, Convoluted neural networks; CRNN, convolutional recurrent neural network; CNN-HMM, convolutional neural network with hidden Markov Models; DBN, Deep Belief Network; DL, Deep Learning; DNN, Deep Neural Network; DT, Decision Trees; DTL, Deep Transfer Learning; FCNN, Fully Convolution Neural Network; FNM, Fuzzy Neural Model; FRB, Fuzzy rules based; GLMM, Generalized Linear Mixed Model; GAN, Generative Adversarial Network; kNN, k-Nearest neighbors; LASSO, Least Absolute Shrinkage and Selection Operator; LDA, Linear Discriminant Analysis; LMM, Linear Mixed model; LR, Logistic Regression; ML, Machine Learning; MMDLM, Multi Modal Deep Learning Model; NFM, Neuro-Fuzzy Modeling; NR, Not Reported; PLS-DA, Partial least Squares Discriminant Analysis; QDA, Quadratic Discriminant Analysis; RF, Random Forests; RBM, regression-based Models; RBFN, Radial Basis Function Network; RBS, Rule-based model; RNN, Recurrent Neural Network; RQS, Radiomics Quality Score; SVM, Support-vector machines; TML, Traditional Machine Learning; TL, Transfer Learning; Uro-XGBoost, uro-eXtreme Gradient Boosting.

risk,^{29,39,45,47,66} and the rest were high-risk.^{8,10,12,14–19,23–26,28,31,33,34,40,44,46,48,51,53–56,58,60–64,67,69–74}

Narrative synthesis of results

Artificial intelligence in prostate cancer

The available evidence of AI application on PCa has been assessed in forty-five SRs of 784

primary studies^{8–40,61–72} (Table 2). The overlap is considered high since CCA equals 15.78. The citation matrix is available (Supplementary Table S1). Thirty-four SRs were designed specifically for PCa.^{8–40} According to the ROBIS criteria, fifteen SRs^{9,11,13,20,21,27,30,32,35–38,65,68} were considered low-risk,

TABLE 2. Data extraction results of included systematic reviews for prostate cancer

Study ID	Objective of AI	Effect of Intervention AUC/Accuracy	Sensitivity	Specificity
Sihotang et al. 2022 ⁸	Diagnostic tool using TRUS images	0.76–0.98/-	>70%	>70%
Sushentsev et al. 2022 ⁹	Diagnostic tool using MRI images to identify csPCa (Fully automatic method (DL) vs. Semi-Automatic method (TML))	0.80–0.89 vs. 0.75–0.88/-	0.6–1.0 vs. 0.84–0.95	0.52–0.98 vs. 0.58–0.87
	Diagnostic tool using bmMRI images	0.71–0.99/72–99%	63.6–100%	65.8–99%
Michaely et al. 2022 ¹⁰	Comparison of AI vs. Radiologists in PCa diagnosis	0.78–0.999 vs. 0.816–0.867/-	57–100% vs. 53–98%	17–55% vs. 17–58%
	csPCa identification from PIRADS 3 lesions	0.89 (95% CI 0.88–0.9)/-	–	–
	ECE Prediction	0.728/-	–	–
	PSM Prediction	0.766/-	–	–
Liberini et al. 2022 ¹¹	Diagnostic tool using MRI images to identify GS7 vs. ≥GS7	0.91/-	–	–
	Diagnostic tool using MRI images to identify bone metastasis	0.76–0.95/-	73.5–78.0%	73.7–93.0%
Ghezzi et al. 2022 ¹²	MRI-Based radiomics in PCa Diagnosis	0.89/-	–	–
	MRI-Based radiomics in BCR Prediction	0.71–0.76/-	–	–
Castaldo et al. 2021 ¹³	Radiomics in PCa diagnosis	-/-	0.815	0.828
Van Booven et al. 2021 ¹⁴	Diagnostic tool using histopathology images	0.997/-	95%	68%
Sugano et al. 2020 ¹⁵	Diagnostic tool using radiomics in PCa diagnosis	0.86–0.91/-	–	–
	Diagnostic tool using radiomics in csPCa recognition	0.822–0.870/- (Combined model 0.933)	–	–
	Diagnostic tool using radiomics in ECE prediction	0.67–0.74/- (Combined model 0.80)	–	–
Cuocolo et al. 2020 ¹⁶	Diagnostic tool using MRI images for csPCa	0.86/-	–	–
Almeida and Tavares 2020 ¹⁷	Evaluation of AI methods for planning in radiation oncology (MRI)	-/64.6–95.11%	–	–
	Evaluation of AI methods for planning in radiation oncology (CT)	-/86.8–90.02%	–	–

(Continued)

TABLE 2. (Continued)

Study ID	Objective of AI	Effect of Intervention AUC/Accuracy	Sensitivity	Specificity
Castillo et al. 2020 ¹⁸	Diagnostic tool using MRI images for csPCa	0.77–0.87/-	–	–
Schröder and Kattan 2008 ¹⁹	Diagnostic tool for PCa	0.67–0.91/-	–	–
Morozov et al. 2023 ²⁰	Diagnostic tool using histopathology images	0.99–0.997/83.7–98.3%	87–100%	68–99%
	Grade prediction using histopathology images	0.8/-	77–87%	82–90%
Quartuccio et al. 2021 ²¹	Diagnostic tool using PET-radiomics in PCa diagnosis	0.8/-	94%	89%
	Diagnostic tool using PET-radiomics in csPCa recognition	0.84/-	–	–
	Diagnostic tool using PET-radiomics in BCR prediction	0.8/89%	–	–
	Diagnostic tool using MRI images for PCa per lesion (GS $\geq 3 + 3$) (Human reading alone vs. CAD system alone vs. Combination)	0.69–0.88 vs. 0.64–0.95, vs. 0.78–0.98	51–100% vs. 63–100%, vs. 68–99%	40–91% vs. 56–91% vs. 26–96%
Syer et al. 2021 ²²	Diagnostic tool using MRI images for PCa per patient (GS $\geq 3 + 3$) (Reader alone vs. CAD system alone vs. Combination)	0.82–0.88 vs. 0.83–0.95 vs. 0.78–0.85	81–100% vs. 96–100% vs. 82–91%	9–75% vs. 24–52% vs. 57–78%
	Diagnostic tool using PSMA PET/CT in PCa diagnosis	0.85–0.95/-	84–95%	80–92%
Huynh et al. 2023 ²⁴	Diagnostic tool using MRI Radiomics in GS Prediction	0.72–0.92/-	–	–
	Diagnostic tool using MRI Radiomics in ECE Prediction	0.8–0.57/-	71–94%	77–89%
Naik et al. 2022 ²⁵	Diagnostic tool using MRI images for PCa Diagnosis	0.76–0.997/-	87–98%	22–50%
	Diagnostic tool using histopathology images for PCa Diagnosis	0.83–0.94/-	–	–
Nematollahi et al. 2023 ²⁶	Diagnostic tool using MRI images for PCa Diagnosis	0.82/-	–	–
Rouviere et al. 2023 ²⁷	Diagnostic tool using MRI images for PCa Diagnosis (Human reading alone vs. CAD System)	0.44–0.874 vs. 0.55–0.945	50–97.8% vs. 30–100%	6–95.2% vs. 21–90.7%
	Diagnostic tool using MRI images for csPCa Diagnosis (Human reading alone vs. CAD System)	0.74–0.88 vs. 0.68–0.95/-	88–100% vs. 56–100%	8–40% vs. 23–95.6%

(Continued)

TABLE 2. (Continued)

Study ID	Objective of AI	Effect of Intervention AUC/Accuracy	Sensitivity	Specificity
Spohn et al. 2021 ²⁸	Diagnostic tool using MRI Radiomics in PCa Diagnosis	0.6–0.86/87%	–	–
	Diagnostic tool using MRI Radiomics in GS Prediction	0.64–0.82/-	–	–
	Diagnostic tool using PSMA-PET Radiomics in GS Prediction	0.81–0.91/-	–	–
Telecan et al. 2022 ²⁹	Diagnostic tool using MRI Images in PCa Diagnosis	-/75.0–81.8%	–	–
	Diagnostic tool using MRI Images in csPCa Diagnosis (CAD system vs. human reading)	0.75 vs. 0.5/-	90 vs. 83%	70 vs. 47%
Xing et al. 2022 ³⁰	Diagnostic tool using MRI Images in PCa Diagnosis	0.89/-	0.87	0.76
Frewing et al. 2024 ³¹	Diagnostic tool for PCa and Gleason grading (ML Models Binary to Multi-Classification)	0.85–0.99 vs. 0.78–0.9 vs. 0.71–0.98 vs. 0.59–0.97	–	–
Antolin et al. 2024 ³²	Diagnostic tool using MRI Radiomics in csPCA recognition (Handgrafted and Deep Radiomics models)	0.72–0.98 and 0.73–0.85/-	0.73–0.95 and 0.64–0.99	0.55–0.94 and 0.47–0.92
Calimano-Ramirez et al. 2023 ³³	Diagnostic tool using MRI Radiomics for ECE prediction in PCa	0.682–0.88/-	–	–
Faiella et al. 2022 ³⁴	Radiomics in Bone metastasis prediction	0.916–0.92/-	–	–
Guerra et al. 2024 ³⁵	Diagnostic tool using MRI Radiomics in ECE Prediction, Best model performance (radiomics associated with clinical/semantic features) vs. Radiomics only	0.69–0.89 vs. 0.7–0.93/-	76.8–95.2% vs. 68–90%	52.3–91.9% vs. 66.7–91.4%
Liu et al. 2024 ³⁶	Diagnostic tool using PSMA PET/CT for Evaluation mPCa.	0.76–0.98/52– 94.5%	73.5–97%	73.5–96%
Liu et al. 2024 ³⁷	Tool to identify PCa using PSMA PET/CT	-/67.1–99.2%	91%	35%
	Tool to identify ISUP GG ≥ 4 using PSMA PET/CT	-/83–88%	89%	87%
Rabilloud et al. 2023 ³⁸	Detect ECE using PSMA PET/CT	0.70–0.77/-	–	–
	Tool for PCa diagnosis using histopathology images	0.8–0.99/-	74–99%	93–96%
	Tool for PCa Gleason grading using histopathology images	0.82–0.997/70– 92.7%	77–100%	68–94%
	Tool for PCa outcome prediction using histopathology images	0.59–0.997/-	–	–

(Continued)

TABLE 2. (Continued)

Study ID	Objective of AI	Effect of Intervention AUC/Accuracy	Sensitivity	Specificity
Wang et al. 2023 ³⁹	Prediction tool for Lymph Node Mets in PCa (New ML models (training and validation set) vs. existing)	0.846 and 0.862 vs. 0.731	0.78 and 0.81 vs. 0.8	0.85 and 0.82 vs. 0.7
Rabaan et al. 2023 ⁴⁰	Diagnostic tool using MRI images for PCa Diagnosis	0.73–0.93/-	–	–
	Diagnostic tool using TRUS images for PCa Diagnosis	0.72–0.94/-	–	–
Kirienko et al. 2021 ⁶¹	Diagnostic tool using PET/CT for PCa Staging	0.95/-	–	–
	Diagnostic tool using PET/CT for PCa Re-staging	0.97/-	–	–
Bazarkin et al. 2024 ⁶²	Tool to assess AI capabilities in the genetics of PCa	0.67–0.997	–	–
Papalia et al. 2024 ⁶³	Diagnostic tool to detect bone metastatic PCa using PET	-/81.2–99.2%	72–98%	82–92%
	Diagnostic tool to detect bone metastatic PCa using CT/MRI	-/81.7–96%	80.6–89.8%	–
Schneider et al. 2022 ⁶⁴	Use of CNN fusion with omics data in PCa Diagnosis	0.74 vs. 0.71/-	–	–
Wessels et al. 2022 ⁶⁵	Diagnostic tool for PCa	0.68–0.85/-	0.53–0.71	0.61–0.70
Prabhu et al. 2022 ⁶⁶	Diagnostic tool using histopathology images	0.951/92–93%	79–90%	81–93%
	Grade prediction using histopathology images	0.81–0.87/88–91%	83–87.1%	–
Forsch et al. 2021 ⁶⁷	Grade prediction using histopathology images	–70.8%	–	–
Salem et al. 2021 ⁶⁸	Tool for PCa diagnosis	0.72–0.91/-	79–95%	23–62%
	Tool to predict PCa progression	0.71–0.74/74–80%	64–79%	61–85%
	PCa diagnosis using MRI images	0.97–0.99/-	–	–
Checucci et al. 2020 ⁶⁹	PCa diagnosis using PET images	0.81/-	–	–
	PCa diagnosis using histopathology images	0.72–0.918/-	–	–
Chen et al. 2019 ⁷⁰	Diagnostic tool for PCa	0.55–0.99/32–93%	42–75%	30–60%;
	Prediction of PCa recurrence	–/66.2–97.6%	70–94%	65–94%;
	Tool for PCa diagnosis	0.74–0.91/-	59–95%	68–92%
Abbod et al. 2007 ⁷¹	Tool for PCa Staging	0.77–0.88/74–80%	21–100%	30–82%
	Tool for PCa Prognosis	0.7–0.9/60–84%	16–85%	55–74%
Faiella et al. 2023 ⁷²	Tool for lymph node involvement in PCa Diagnosis and prediction	0.73–0.95/-	77–82%	–

Note. Abbreviations: AUC, Area Under Curve; BCR, Biochemical Recurrence; CAD, Computer-Aided Diagnosis; CEUS, Contrast-Enhanced Ultrasound; csPCa, Clinically Significant Prostate Cancer; CSS, Cancer Specific Survival; CT, Computer Tomography; ECE, Extracapsular Extension; GS, Gleason Score; MRI, Magnetic Resonance Imaging; NR, Not Reported; OS, Overall Survival; PCa, Prostate Cancer; TRUS, Transrectal Ultrasonography.

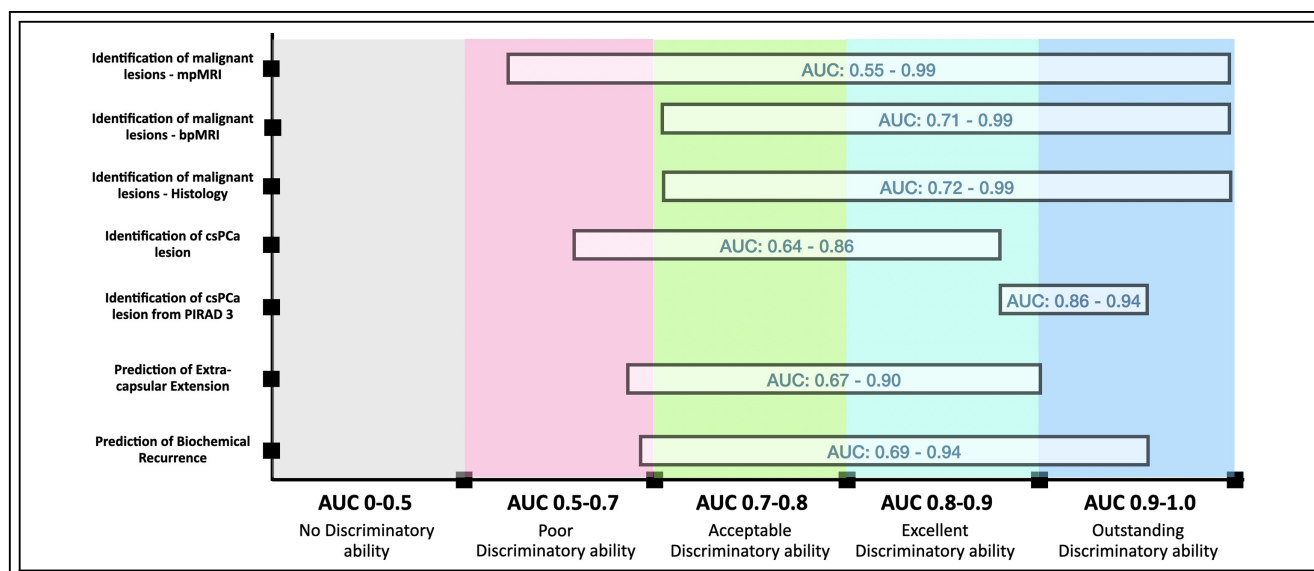


FIGURE 2. Principal findings of AI-models efficacy in Prostate cancer management

three of unclear-risk,^{29,39,66} and twenty-seven of high-risk^{8,10,12,14–19,22–26,28,31,33,34,40,61–64,67,69–72} (Figure 2).

AI Algorithms as Diagnostic Prediction Models in Prostate Cancer

Automatic diagnosis of benign from malignant prostate lesions

Ultrasonography is a traditional mode of prostate visualization with limited efficacy due to a lack of standardisation and high inter-rater variability. Four SRs reported an AUC of 0.6–0.98 with a sensitivity and specificity above 70%.^{8,40,68,70} In addition, AI expedites the process, minimizes patient discomfort, and improves prediction outcomes.^{8,40}

Multiparametric MRI (mp-MRI) is essential in PCa diagnosis.² The Prostate Imaging Reporting and Data Systems version 2 (PI-RADSv2) scoring system is characterized by good performance for PCa detection.² CAD systems improve inter-reader agreement and recognize non-visible features, enhancing diagnostic accuracy. Eighteen SRs evaluated the accuracy of AI models on PCa diagnosis from mpMRI images.^{9,12,15,19,22,24–30,40,65,68–71} Most AI models used semi-automatic or fully automatic segmentation. The AI models recognized benign from malignant lesions with an AUC of 0.55–0.99 (poor to outstanding discriminatory ability) and accuracy of 88–93%.^{12,15,19,22,24,25,29,40,65,68–71} The sensitivity and specificity ranged from 59–95% and 58–87%, respectively.^{22,25,30,70} Four SRs compared the efficacy of AI models vs. radiologists with an AUC of 0.64–0.95 vs. 0.44–0.88.^{15,22,27,29} The accuracy of AI-enhanced

reading by radiologists scored an AUC of 0.78–0.98 (acceptable to outstanding discriminatory ability).

A SR reviewed the use of ML using bi-parametric MRI (bp-MRI) images and reported an AUC of 0.71–0.99 and accuracy of 72–99%, while the comparison between radiologists-alone and AI-enhanced reading reported an AUC of 0.82–0.87 vs. 0.78–0.99 and sensitivity of 57–100% vs. 53–98%, respectively.¹⁰

Molecular imaging has become increasingly popular in PCa diagnosis and staging.² Six SRs reported on the diagnostic performance of AI models using radiomic features extracted from Positron Emission Tomography (PET) images with an AUC of 0.81–0.95 (excellent to outstanding discriminatory ability).^{15,20,21,23,61,69}

AI models detect and grade malignant cells based on their distinguishable characteristics. Ten SRs evaluated AI models for histopathological diagnosis with diagnostic accuracy of 70.8–93% and an AUC of 0.72–0.99 (acceptable to outstanding discriminatory ability).^{14,20,21,25,38,65–67,70} The use of AI-enhanced reading improved sensitivity by 16.1%.²⁰ A meta-analysis reported a pooled sensitivity of 0.96 (95%CI: 0.95–0.97) and a pooled specificity of 0.95 (95%CI: 0.95–0.96).²⁰

Identification of clinically significant prostate cancer (csPCa) lesions

Several attempts have been made to define csPCa using histopathologic and radiologic criteria, but there are insufficient data to correlate with clinical endpoints.² Seven SRs assessed AI models' efficacy to detect csPCa from mpMRIs using

Gleason score ≥ 7 or Gleason Group ≥ 2 as cut-off points.^{9,12,16,18,22,29,32} The stand-alone AI-model had an AUC of 0.64–0.86 [acceptable to outstanding discriminatory ability], with pooled sensitivity of 63–100% and specificity of 47–91%. The AI-enhanced reading improves radiologists' diagnostic accuracy compared to radiologists-alone and AI stand-alone reading (AUC: 0.78–0.98 vs. AUC: 0.69–0.88 vs. AUC: 0.64–0.95).²²

Equivocal mpMRI findings have a 23–32% risk of harboring csPCa.² AI models have shown an AUC of 0.86–0.94 [Excellent discriminatory ability] in detecting csPCa in PI-RADS 3 lesions.¹⁰ Using radiomic features, csPCa recognition appears superior to PI-RADS with an AUC of 0.85 vs. 0.73, sensitivity 90% vs. 47%, and specificity 70% vs. 47%.²⁹ The agreement between radiologists and AI-model prediction was proportionally increased from 6% in PI-RADS 2 to 80% for PI-RADS 5 lesions. The prediction of aggressiveness was comparable; however, AI minimized interobserver variability and inherent subjectivity of PIRADS score assessment.²⁹ An SR that compared the diagnostic efficacy of automated vs. semi-automated methods reported an AUC of 0.80–0.89 and 0.75–0.88, respectively.⁹

Extracted features using PET have been tested to improve accuracy and differentiate csPCa from indolent lesions. However, the results of [⁶⁸Ga] Ga-PSMA PET/CT were surprisingly low in the validation phase since diagnosis was missed in 50% of patients.¹¹ Recent evidence reported an accuracy between 0.83 and 0.88 to differentiate GG ≥ 4 lesions and an accuracy of 0.87 to identify csPCa in non-PSMA-avid lesions.³⁷

Discrimination of extracapsular extension

Extracapsular extension (ECE) is associated with adverse oncologic outcomes.² Using mpMRI images, AI models predicted ECE with an AUC of 0.67–0.90 [Poor to Excellent discriminatory ability] and accuracy of 71%–81.8%.^{10,12,15,24,28,29,33,69} Manual segmentation techniques were evaluated in most primary studies, but accuracy decreased at external validation.

A radiomic model was compared to clinical features, including the Memorial Sloan Kettering (MSK) Cancer Center nomogram for predicting ECE.^{2,12,15,35} AUC was 0.74–0.85, while the best AUC for combined models varied from 0.72–0.92 for the validation group (acceptable to outstanding discriminatory ability). However, the clinicopathological data were considered the accuracy ranged from 67.1% to 85.4%.²⁹ Using PSMA PET/CT images, AI models can detect ECE with an AUC of 0.70–0.77 (Acceptable discriminatory ability).³⁷

Recognition of metastasis (staging and restaging)

Eleven SRs assessed AI models for staging and restaging purposes.^{11,12,21,23,34,36,39,61,63,72,74} Radiomic analysis using CT images for LN involvement prediction showed an AUC of 0.95–0.97 (outstanding discriminatory ability) and accuracy of 89%.^{12,23} The predictive value of AI models in LN involvement has been addressed in an SR, with predictive performance of 0.862 (95%CI [0.819–0.905]).³⁹ Two SRs evaluated AI models using skeletal involvement using scintigraphy images with an accuracy of 90.0 to 97.4%.^{11,12} The accuracy of bone lesions using PET/CT and PET/MRI was 81.7–96% and 81.2–99.2%, respectively.⁶³ Input data from MRI images showed an AUC of 0.84–0.90 for bone metastasis identification,⁷² while data from PET/CT images reported an AUC of 0.73–0.98 and accuracy of 52–94.5%.^{11,34} Using features from the [⁶⁸Ga] Ga-PSMA-1 PET/CT, AI models identified responders from non-responders among patients with bone metastasis. The models were evaluated on bone lesions invisible on low-dose CT, with an AUC of 0.76 (acceptable discriminatory ability) and accuracy of 73.5–85.0%, while sensitivity and specificity ranged from 73.5–78% and 73.7–93%, respectively.

Prediction of Gleason score

Three SRs evaluated the role of MRI-derived models in the prediction of Gleason score (GS).^{22,24,28} Most of the included studies used models with manual segmentation and reported an AUC of 0.50–0.92. The comparison of radiologists-alone vs. AI stand-alone vs. AI-enhanced reading to identify GS $\geq 3 + 3$ and GS $\geq 3 + 4$ in a predefined MRI-lesion, reported AUC of 0.68–0.87, 0.63–0.97, and 0.78–0.98, respectively.²² For unspecified MRI-lesions, the AUCs were 0.82–0.88, 0.83–0.95, and 0.78–0.89. The prediction of GS up-scoring at prostatectomy specimens reported an AUC of 0.65, 0.87, 0.91, and an accuracy of 0.64, 0.72, and 0.79, respectively.²⁴ Two SRs estimated the performance of PSMA-PET and reported an AUC of 0.81–0.91, while the ability to discriminate GS = 7 from Gleason ≥ 8 lesions had an AUC of 0.84.^{21,28}

Two SRs summarized the available evidence on AI models for histology specimens.^{20,25} The models achieved 82–90% accuracy in recognizing high vs. low-grade tumors. Sensitivity and specificity improved with higher GSs reaching 64–80% and 96–98% for GS = 5, respectively. The AUC for automatic recognition of cribriform patterns was 0.80–0.83 (excellent discriminatory ability).

AI Models for Treatment Outcome Prediction in Prostate Cancer

The prediction of survival outcomes is based on prognostic nomograms. AI models combine data

from multiple sources and predict treatment or survival outcomes. Older studies reported an accuracy of 60–84%, an AUC of 0.70–0.90 (acceptable to outstanding discriminatory ability), sensitivity of 16–85%, and specificity of 55–74%.⁷¹ An SR evaluated AI models and compared them to image data alone and reported an AUC of 0.74 and 0.71, respectively, for predicting cancer-specific survival.⁶⁴

Three SRs evaluated the AI models to predict biochemical recurrence (BCR) with an AUC of 0.69–0.94 and accuracy of 66–97%.^{12,21,70} A SR compared the use of AI vs. standard practice (PSA, nomograms) and reported an accuracy of 89% and 70%, respectively.²¹

Two SRs have explored the use of AI models in the prediction of disease progression.^{12,68} The first evaluated the use of radiomics and reported an AUC of 0.8–0.92 (excellent to outstanding discriminatory ability).¹² The second explored the predictive ability of AI models using histopathology images from radical prostatectomy and reported an AUC of 0.71–0.74 (acceptable discriminatory ability) and accuracy of 74–80%.⁶⁸

Genomics in Prostate Cancer

Genomic or molecular biomarkers hold the potential to diagnose PCa from blood, urine, tissue, or semen samples and predict treatment response. Three SRs evaluated the available evidence and reported a sensitivity of 67–95% and a specificity of 15–97%, respectively.^{13,62,64} A meta-analysis revealed a pooled sensitivity of 0.88 (95%CI: 0.541–0.99) and a pooled specificity of 0.73 (95% CI: 0.330–0.99) (heterogeneity 73% and 92%, respectively).¹³

Theragnostic in Prostate Cancer

Radioligand therapies are recommended for metastatic disease.² However, a third of patients will progress despite PSMA positivity. Two SRs evaluated the ability of AI to identify the potential non-responders and reported a pooled AUC of 0.68–0.83.^{11,23}

Artificial intelligence in renal cancer

The available evidence of AI application on renal carcinoma (RCa) has been assessed in sixteen SRs of 209 primary studies^{41–51,63,65,69–71,73} (Table 3). The overlap was considered moderate with CCA equal to 7.81. The citation matrix is available (Supplementary Table S2). Eleven SRs were designed specifically for RCa.^{41–51} According to the ROBIS criteria for RoB assessment, six SRs^{41–43,49,50,65} were considered low risk, one was considered of unclear risk,⁴⁵ and ten were considered high risk^{44,46–48,51,63,69–71,73} (Figure 3).

AI Models as Diagnostic Prediction Models in Renal Cancer

AI models based on feature analysis from routine CT or MRI show acceptable diagnostic accuracy in differentiating benign from malignant renal lesions. Pooled evidence from eight SRs revealed an AUC of 0.52–0.97 (poor to outstanding discriminatory ability) and accuracy of 70–94.7%.^{42,44,49–51,65,70,73} The pooled sensitivity and specificity were 88–100% and 41–90.7%, respectively. Two SRs compared the AI-enhanced reading vs. radiologists alone and reported a mean diagnostic accuracy (75% vs. 65%), sensitivity (90% vs. 80%), and specificity (46.5% vs. 40%).^{42,44} AI-enhanced reading was superior to radiologists alone in differentiating angiomyolipomas from RCa with an overall accuracy of 70% vs. 60%, sensitivity of 77% vs. 69%, and specificity of 76% vs. 71%.⁴⁴

Oncocytomas are renal lesions with low metastatic potential, which can be misdiagnosed as more aggressive types of RCa. Two SRs reported that AI models using CT and/or MRI images had a pooled AUC of 0.77–0.96 (Acceptable to outstanding discriminatory ability) and an accuracy of 74.4–94% in discriminating oncocytomas from RCa.^{42,73} Another SR assessed the sensitivity and specificity of AI models, which were 81.8% and 62.8%, while for differentiating oncocytomas from chromophobe subtype were 83.0% and 92.0%, respectively.⁴³

Several studies report on the efficacy of AI in the recognition of different RCa subtypes. Three SRs evaluated AI models using CT or MRI images, showing an AUC of 0.83–0.96 (excellent to outstanding discriminatory ability) and an accuracy of 69.0–92.9%.^{42,48,73} Checcucci et al. reported that AI models recognize clear cell RCa (ccRCC) and chromophobe RCa with 94% and 88% accuracy, respectively.⁶⁹ A meta-analysis reported an accuracy of 93.5% (95%CI: 88.7–96.3) in RCC subtype detection.⁴⁹ Tramanzoli et al. reported that AI models are superior to radiologists alone in recognizing ccRCC vs. non-ccRCC with an AUC of 0.91–0.97 vs. 0.69–0.79.⁷³ In addition, the AI model predicted tumour grade with an AUC of 0.82–0.98 [Excellent to outstanding discriminatory ability] and an accuracy of 73–92.3%.^{42,44,49} A meta-analysis of radiomics in RCa diagnostics reported a log diagnostic odds ratio of 3.57 (95%CI: 2.69–4.45) in discriminating benign from malignant lesions, 2.89 (95%CI: 2.40–3.39) in discriminating angiomyolipoma from RCa, and 3.08 (95%CI: 2.09–4.06) in the oncocytoma vs. RCC discrimination.⁴⁷

AI Algorithms as Treatment Outcome Prediction Models in Renal Cancer

The treatment outcome prediction models have been assessed in three SRs.^{49,64,73} AI models could predict cancer-specific survival (CSS) and metastatic

TABLE 3. Data Extraction results of included systematic reviews for renal cancer

Study ID	Objective of AI	Effect of Intervention (AUC/Accuracy)	Sensitivity	Specificity
Ursprung et al. 2020 ⁴¹	To assess the quality of studies on the use of radiomics to AML diagnosis	RQS mean (SD): 3.41 (4.43)	–	–
Ferro et al. 2023 ⁴²	Diagnostic tool using CT images for RCa (CAD system vs. Human reading)	-/83% vs. 68%	–	–
	Diagnostic tool using MRI images for RCa (CAD system vs. Human reading)	-/70% vs. 60%	92% vs. 80%	41% vs. 35%
Firouzabadi et al. 2023 ⁴³	Diagnostic tool using radiomics images in oncocytoma recognition	-/-	34–100%	28.3–97.4%
Roussel et al. 2022 ⁴⁴	Tool for RCa diagnosis	0.54–0.87 / 70–88%	88–92%	41.0–67.0%
	Comparison of AI vs. Radiologists in RCa diagnosis	-/70% vs. 60%	–	–
	Comparison of AI vs. Radiologists in AML vs. RCa diagnosis	-/70% vs. 60%	77% vs. 69%	71% vs. 71%
Kocak et al. 2020 ⁴⁵	Diagnostic tool using radiomics to diagnose RCa	-/63%	–	–
Pandey and Gupta 2021 ⁴⁶	To analyse different kidney segmentation methods in abdominal CT scans	-/68–92%	–	–
Muhlbauer et al. 2021 ⁴⁷	Tool for RCa diagnosis	Log odds ratio (OR) 2.89 (95%CI 2.4–3.39)	–	–
Bhandari et al. 2020 ⁴⁸	Diagnostic tool using CT radiomics images in RCa	0.82–0.978/-	–	–
	Diagnostic tool using CT radiomics images in tumor differentiation	0.82–0.96/-	–	–
Chandramohan et al. 2024 ⁴⁹	Tool for RCa diagnosis	0.91/92.3%	97.5%	89.2%
Khene et al. 2024 ⁵⁰	Prediction of oncological subtyping	0.88–0.96/72–94.4%	–	–
	Diagnostic tool using histopathology	0.84–0.97/95.6–98%	–	–
	Recognition of Oncocytomas using CT Images	0.77–0.96/91.6–94.0%	84–99%	80%
Xi et al. 2024 ⁵¹	Recognition of AML using CT Images	0.706–0.95/-	98–100%	48.0–87.8%
	Recognition of benign lesions using MRI Images	0.73–0.95/-	–	–
	Recognition of benign lesions using US Images	0.84–0.96/-	76.7–100%	81.7–90.7%

TABLE 3. (Continued)

Study ID	Objective of AI	Effect of Intervention (AUC/Accuracy)	Sensitivity	Specificity
Wessels et al. 2022 ⁶⁵	Tool for RCa diagnosis	0.63–0.78/-	–	–
Checucci et al. 2020 ⁶⁹	Tool for RCa diagnosis	–/94%	–	–
Chen et al. 2019 ⁷⁰	Tool for RCa diagnosis	0.94/91.3–94.7%	95.8%	88.8%
Abbod et al. 2007 ⁷¹	Tool for RCa diagnosis	–/99%	–	–
Tramanzoli et al. 2023 ⁷³	Diagnostic tool using MRI radiomics images in RCa	0.824–0.925/-	–	–
	Prediction tool using CT radiomics images for tumor grade	0.817–0.91/-	–	–
	Prediction tool using MRI radiomics images for tumor grade	0.842–0.9/-	77.3–85%	80–84%

Note. Abbreviations: AML: Angiomyolipoma, AUC: Area Under Curve, CAD: Computer-Aided Diagnostic System, CT: Computer Tomography, MRI: Magnetic Resonance Imaging, RCa: Renal Cell Carcinoma, RQS: Radiomics Quality Score, US: Ultrasound

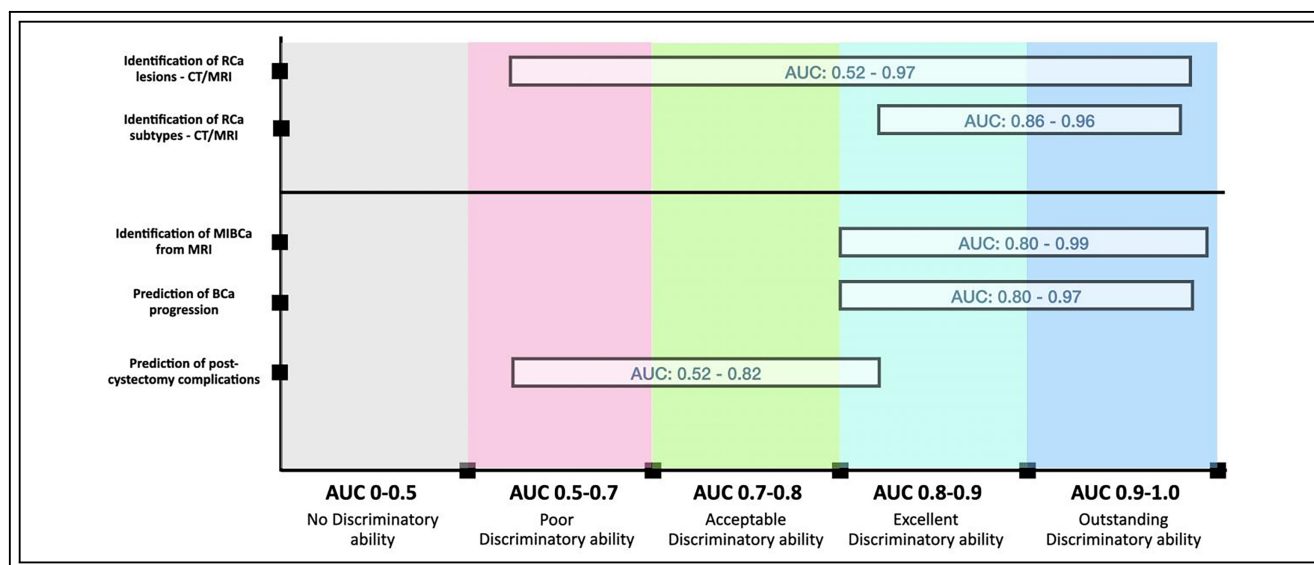


FIGURE 3. Principal findings of AI-models efficacy in kidney cancer (top) and bladder cancer (bottom) management

potential with an AUC of 0.72–0.79 and 0.82, respectively.⁷³ The accuracy of survival prediction was estimated at 81% (95% CI: 67.8–89.6%).⁴⁹

Artificial intelligence in bladder cancer

Data on the application of AI on Bladder Carcinoma (BCa) have been assessed in sixteen SRs of 121 primary studies^{52–60,62,68–71,73,74} (Table 4). The overlap is

considered slight since the CCA was 1.87 (<5). The citation matrix is available (Supplementary Table S3). Nine SRs assessed AI models specifically for BCa.^{52–60} According to the ROBIS criteria for RoB assessment, one SR was considered as low risk⁶⁸ and fifteen as high risk^{52–60,62,69–71,73,74} (Figure 3).

AI Algorithms as Diagnostic Prediction Models in Bladder Cancer

TABLE 4. Data Extraction results of included systematic reviews for BCa

Study ID	Objective of AI	Effect of Intervention (AUC/Accuracy)	Sensitivity	Specificity
Kozikowski et al. 2021 ⁵²	Tool for MIBCa prediction	0.798–0.986/-	82%	81%
Cacciamani et al. 2020 ⁵³	Prediction Tool using MRI for therapy response	0.72–0.82/66.7–75.5%	55.6–77.8%	–
	BCa grade prediction tool using MRI	0.769–0.928/76.7–83.3%	76.9–84.6%	87.1%
	BCa stage prediction tool using MRI	0.874–0.907/-	87.3%	78.1%
	Tool for MIBCa prediction CAD system vs. clinical model	0.702–0.906 vs. 0.846–0.943	–	–
Liu et al. 2023 ⁵⁵	Use of ML to predict survival outcomes	-/-	–	–
Kwong et al. 2024 ⁵⁶	Tool for NMIBC recurrence prediction	0.61–0.86/72–91%	33–100%	40–96%
	Tool for NMIBC progression prediction	0.66–1.00/75–100%	70–100%	67–100%
He et al. 2024 ⁵⁷	Diagnostic tool for MIBCa prediction (Radiomics and DL machines)	0.89 and 0.91/-	84% and 81%	82% and 87%
Khoraminia et al. 2023 ⁵⁸	Detection of benign from malignant lesions from pathology images	0.94–0.98/-	–	–
	Grading and staging tool using pathology images	-/79–96%	–	–
	Prediction of clinical outcome using pathology images	0.62–0.89/73–92%		
Suartz et al. 2024 ⁵⁹	Prediction tool for NAC response in MIBCa using CT images	–	62%	82%
Chan et al. 2021 ⁶⁰	Tool for BCa diagnosis	0.96–0.98/85.6–96.9%	89.7%	96.1%

TABLE 4. (Continued)

Study ID	Objective of AI	Effect of Intervention (AUC/Accuracy)	Sensitivity	Specificity
Bazarkin et al. 2024 ⁶²	Tool to assess AI capabilities in the genetics of BCa	0.588–0.90/-	–	–
Checcucci et al. 2020 ⁶⁹	Tool for BCa diagnosis	0.86–0.98/79–82%	82–91%	77–99%
Salem et al. 2021 ⁶⁸	Tool for Recurrence prediction	-/72–100%	76–100%	55–98%
	Prediction Tool of 5-year survival rates	-/72–80%	59–63%	77–86%
Chen et al. 2019 ⁷⁰	Tool for BCa diagnosis	0.72–0.97/68.6–87.9%	71.9–90.3%	85.5–88%
	Prediction Tool of BCa recurrence	0.69–0.92/-	68.6–88%	48–90%
	Prediction Tool for mortality after cystectomy	0.82/72.2–80%	73.5–85.6%	68.1–81.5%
Abbod et al. 2007 ⁷¹	Tool for BCa diagnosis	-/91–98%	–	52%
	Tool for BCa Prognosis	0.91–1.0/88–100%	80–100%	80–100%
Tramanzoli et al. 2023 ⁷³	Tool for MIBCa prediction using CT images (CAD vs. Radiologist)	0.979 vs. 0.856/-	79.4–90% vs. 91.2–96.7%	85.3–96.7% vs. 47.1–58.3%
	Tool for MIBCa prediction using MRI images (CAD vs. Radiologist)	-/96.3% vs. 90.12%	–	–
Drouin et al. 2013 ⁷⁴	Prediction Tool of 5-year survival rates	-/88–92%	–	–

Note. Abbreviations: AI: Artificial Intelligence, AUC: Area Under Curve, CAD: Computer-Aided Diagnostic System, CEUS: Contrast-Enhanced Ultrasound, CT: Computer Tomography, DL: Deep Learning, MIBCa: Muscle Invasive Bladder Cancer, ML: Machine Learning, MRI: Magnetic Resonance Imaging, NR: Not Reported

Multiple AI models have been developed to improve the diagnostic performance of cystoscopy from still images. In most primary studies, flat lesions, stones, or blurry images were excluded from analysis. Two SRs evaluated the outcomes and reported an AUC of 0.86–0.98 (excellent to outstanding discriminatory ability) and accuracy of 79–96.9%, to recognize BCa.^{60,69} There is limited data to evaluate

stage or grade prediction using AI-enhanced cystoscopy images; however, the prediction of large lesions is more accurate.

Four SRs evaluated the role of AI models in the diagnosis of muscle-invasive BCa (MIBCa) based on MRI features with an AUC of 0.80–0.99 (Excellent to outstanding discriminatory ability) and accuracy of 79.0–91.9%.^{52–54,73} The diagnostic efficacy of AI

models in MIBCa prediction was improved when other clinical characteristics were incorporated into the model compared to the AI model alone, with an AUC of 0.85–0.94 vs. 0.70–0.92, accuracy of 89.7% vs. 83.9–91.9%, and sensitivity of 96.4% vs. 87.3%.^{53,54} A SR reported that AI-stand-alone models outperform radiologists-alone in the diagnosis of MIBCa using contrast-enhanced CT (AUC: 0.98 vs. 0.85, sensitivity: 79.4–90.0% vs. 91.2–96.7%, and specificity: 85.3–96.7%), or MRI (accuracy: 96.3% vs. 90.1%, sensitivity: 92.6% vs. 91.1%, and specificity: 100% vs. 88.9%).⁷³

AI Algorithms as Treatment Outcome Prediction Models in Bladder Cancer

Eight SRs evaluated the role of AI as a predictive tool in BCa^{53,56,60,68–71,74} and four SRs^{47,51,54,56} evaluated the role of AI in the prediction of recurrence or progression. AI models could predict recurrence with an AUC of 0.61–0.92 (poor to outstanding discriminatory ability) and accuracy of 72–100%, disease progression with an AUC of 0.66–1.00 (excellent to outstanding discriminatory ability) and accuracy of 90%, and the 5-year overall survival with accuracy of 72–80%.^{53,68,70,74} Two SRs reported the post-radical cystectomy complications prediction with an AUC of 0.52–0.82 (poor to outstanding discriminatory ability), while the outcome of neo-adjuvant chemotherapy can be predicted with a sensitivity of 62%.^{59,69,70} The mortality associated with radical cystectomy can be predicted with an accuracy of 72–80%, a sensitivity of 73.5–85.6%, and a specificity of 68.1–81.5%.⁷⁰

Applications of artificial intelligence in other urological cancers

AI applications in other urological malignancies have limited evidence (Table 5). A SR⁷⁴ reviewed six studies on AI models' predictive efficacy for UUT-UCC patients and reported a 75.4–80%, 71%, and 70% accuracy in predicting 5-year CSS, risk of muscle-invasion, and non-organ confined disease risk. Data for TCa and AI are limited; however, three primary studies assessed and reported a 5-year recurrence prediction with an AUC of 0.87 and an accuracy of 94%, while the accuracy of disease stage prediction was 87–92%.^{68,71} To date, there is no SR on the use of AI models for penile cancer.

Discussion

In this review, we systematically summarized the available evidence of AI's role in the management of urological cancers. Most SRs assessed the AI models

in PCa, RCa, and BCa; evidence for other urological malignancies remains limited or non-existent, as our study included only SRs and these malignancies have not yet been systematically analysed.

The incorporation of AI models in urology improves diagnostic accuracy and prognostic efficiency, suggesting that AI could be integrated in various clinical workflows—radiology, pathology, risk stratification—offering support in early detection, patient selection, treatment planning, and surveillance. Data from different sources enhances diagnostic accuracy, thereby enabling personalized patient care. Given the rapid expansion of the literature on AI applications, the overview of SRs was selected as the most appropriate methodological approach to provide a higher level of evidence and critically evaluate the relevant SRs that share overlapping clinical questions and often include the same primary studies.

Across prostate, renal, and bladder malignancies, AI models consistently demonstrated high diagnostic accuracy, complementing or even outperforming human experts in imaging and histopathology interpretation. In PCa, AI models are successful in recognising benign from malignant lesions using any type of imaging modality. AI-assisted MRI interpretation improved the detection of csPCa, especially in equivocal lesions such as PI-RADS 3, and even reduced inter-reader variability. AI models have exhibited excellent to outstanding ability in predicting ECE, BCR, and survival outcomes, as indicated by their AUC range. In the realm of histopathology images, AI-enhanced reading tools improved the accuracy and sensitivity in Gleason grading, reduce inter-rater variability and improve detection of distinct entities such as the cribriform pattern PCa. In contrast, AI models for staging, metastasis detection, and treatment outcome prediction demonstrated high accuracy in selected settings, especially when multimodal data were integrated. Nevertheless, unexpected drops in performance during validation—such as with PSMA PET-based models—underscore the sensitivity of AI systems to dataset shift and highlight the need for robust, prospective, multicenter evaluation. Evidence supporting AI-driven genomics and theragnostics remains preliminary, with substantial heterogeneity and wide confidence intervals limiting clinical relevance.

In RCa, AI models demonstrated acceptable effectiveness in differentiating benign from malignant lesions, in identifying RCa subtypes and in predicting tumor grades using data either from CT and MRI images. In BCa AI models improved the accuracy

TABLE 5. Data extraction results of included systematic reviews on other urological malignancies

Study ID	Objective of AI	Effect of Intervention (AUC/Accuracy)	Sensitivity	Specificity
Abbod et al. 2007 ⁷¹	Tool for Testicular Ca diagnosis	-/92%	88%	96%
Salem et al. 2021 ⁶⁸	Prediction tool of Testicular Ca 5-year recurrence	0.87/94%	–	–
Prediction Tool of Testicular Ca stage	-/87%	–	–	–
Drouin et al. 2013 ⁷⁴	Prediction tool of UUT-UCC muscle invasive disease	-/70%	–	–
Prediction tool of UUT-UCC 5-year CSS	-/75.4–80%	–	–	–
Wessels et al. 2022 ⁶⁵	Outcome prediction in urological tumors	0.56–0.78/-	50–93%	38–100%

Note. Abbreviations: AI, Artificial Intelligence; AUC, Area under Curve-Operating Characteristic Curve; Ca, Carcinoma; CSS: Cancer Specific Survival; UUT-UCC, Upper Urinary Tract Urothelial Cell Carcinoma.

of cystoscopic image analysis and improved the predictive ability of progression to MIBC, of disease recurrence and of the response to treatment. These findings support the growing potential of AI as an adjustive tool in urological oncology, contributing to more objective, data-driven clinical decisions.

Although AI models across demonstrate encouraging diagnostic and predictive performance, the overall strength of the evidence remains limited. A large proportion of the included SRs were rated as high-risk of bias, with substantial heterogeneity in imaging protocols and outcome definitions. Additionally, most primary studies were retrospective and lacked external validation, raising concerns about generalizability and possible overestimation of performance. In particular, in PCa, there was a substantial overlap ($CCA > 15$) of primary studies, and there was a wide variability in reported accuracy, sensitivity, and specificity reflecting marked heterogeneity in AI algorithms, segmentation strategies, and reference standards.

The strength of evidence in RCa was also limited by the moderate overlap among the primary studies and the high risk of bias observed in more than half of the included SRs. There was considerable heterogeneity in imaging protocols and feature extraction methods, contributing to wide ranges in reported AUC, sensitivity, and specificity. Notably,

specificity values were frequently lower than sensitivity, highlighting the ongoing risk of false-positive diagnoses and potential overtreatment. AI models appear particularly promising for renal tumor subtype classification and tumor grading, where performance frequently exceeded that of radiologists alone. However, these results were largely derived from retrospective, single-center datasets with limited external validation. Evidence supporting AI-based treatment outcome prediction in renal cancer remains limited. Only a small number of systematic reviews addressed prognostic applications, with moderate predictive performance reported for cancer-specific survival and metastatic potential, suggesting that the current data are insufficient to support routine clinical use.

In contrast to PCa and RCa, the evidence on BCa is characterized by only slight overlap; however, most of the included studies were assessed as having a high risk of bias. More robust performance was observed in the prediction of MIBC using MRI- and CT-based AI models, which consistently achieved high diagnostic accuracy, particularly when clinical variables were integrated with imaging features. In several studies, AI models outperformed radiologists alone, highlighting their potential value in preoperative risk stratification. While AI demonstrated good accuracy in predicting recurrence, disease progression,

survival outcomes, and post-cystectomy complications, performance metrics varied widely and were often reported without consistent external validation. Moreover, the high proportion of SRs assessed as high risk of bias raises concerns regarding the robustness and reproducibility of these findings. As with diagnostic applications, many models lacked transparency and standardized reporting, limiting their immediate clinical applicability.

AI models have shown benefit in solid tumors from other organs such as breast, lung, brain, head/neck, and liver; however, the challenges remain similar to those encountered in urological malignancies.⁷⁵ In addition to the evidence provided so far, the outcomes of two major projects that haven't been included in an SR, merit further attention. The Prostate Imaging-Cancer Artificial Intelligence (PI-CAI) challenge for detecting csPCa from mpMRIs reported that AI models scored an AUC of 0.91 (95% CI 0.87–0.94) compared to 0.86 (95% CI 0.83–0.89) of radiologists using PIRADS 2.1 ($p < 0.001$).⁷⁶ The PANDA Challenge was the largest histopathology competition aiming to develop a reproducible AI algorithm for Gleason grading.⁷⁷ AI-achieved agreement of 0.868 (95% CI 0.835–0.900) with expert uropathologists with AI models to improve pathologists' diagnostic accuracy, reduce inter-observer variability, and reduce assessment time (Table 6).

From a clinical perspective, these findings highlight the potential role of AI as a supportive tool across multiple stages of urological cancer care. In PCa, AI has the potential to improve diagnosis, risk stratification, and treatment planning, particularly when integrated into clinician-led workflows. In RCa, AI models show promise as decision-support systems for lesion characterization and subtype differentiation, with potential to assist clinicians in selecting appropriate management strategies and reducing unnecessary interventions. In BCa, AI models appear most promising as adjunctive systems for image interpretation and risk stratification, particularly in the identification of MIBC and prediction of adverse outcomes following radical cystectomy.

A recurring observation throughout the included SRs is the variability in diagnostic performance metrics. This wide range is not discussed in isolation and appears consistent across cancer types and AI models. Several explanations exist for this heterogeneity: differences in the quality and size of training datasets, variability in imaging acquisition protocols, annotation standards, or even the clinical populations used.

Reproducibility of AI algorithms remains a significant challenge for every medical task, not only for urological cancers. AI models trained in a

single-center environment may not translate well to multi-center or real-world settings, leading to decreased performance due to the domain shift. Reproducibility issues are due to variations in datasets, model architectures, and evaluation metrics. These variations lead to inconsistent model performance across institutions thus limiting generalizability and impairing clinical implementation—if an AI tool shows 47% specificity in one context and 91% in another, the clinician is left uncertain as to what to expect when applying it in their own practice.⁷⁸ Ensuring generalizability requires large-scale training and rigorous external validation across heterogeneous datasets, to enable reliable AI deployment reduce concerns about reproducibility, and facilitate comparison across studies. Recognizing and addressing these limitations is essential to move AI from promise to practice.

Transparency poses a substantial risk not only in terms of traceability, that is the transparency during the design, development, evaluation and deployment of AI algorithms, but also in explainability, which refers to the transparency of the AI model's decisions. The significance of the latter lies in the fact that clinicians may face difficulties in trusting AI decisions and recommendations, despite the decision-making process not being readily comprehensible. Using standardized reporting frameworks like CLAIM (Checklist for Artificial Intelligence in Medical Imaging) and RQS (Radiomics Quality Score) ensures transparent documentation of data, models, and evaluation, while interpretable model designs enhance understanding of decisions. Federated learning offers a practical approach to train models across decentralized datasets, improving reproducibility and generalizability without compromising privacy. Together, these strategies strengthen reliability, transparency, and trust in AI systems.

Apart from the acknowledged role of AI as a diagnostic and prognostic tool, the positive impact of AI is evident in other areas of healthcare, such as virtual patient care, patient engagement and compliance, administrative applications, medical research, and drug discovery.⁷⁸ In addition, AI offers an excellent opportunity to improve medical and surgical education. Generative AI can produce customizable training material that simulates real patients and is adapted to students' needs.⁷⁹ The use of synthetic data can create virtual training cases, thus protecting patient privacy and minimizing the risk of deanonymization of personal data. The incorporation of AI in surgical training may provide solutions to training difficulties and boost surgical education.⁸⁰ A SR has reported that AI and augmented reality have

TABLE 6. Representative clinical case studies of artificial intelligence implementation in urological cancer management

Cancer type	Study/Year	Clinical task	AI approach	Input data	Key performance	Validation	Clinical relevance
Prostate cancer	PI-CAI Challenge, 2024 ⁷⁶	Detection of clinically significant PCa	Deep learning (CNN ensemble)	Multiparametric MRI	AUC 0.91 (AI) vs. 0.86 (radiologists)	External, multicenter	Supports AI as second reader in MRI workflows
Prostate cancer	PANDA Challenge, 2022 ⁷⁷	Gleason grading on histopathology	Deep learning (whole-slide image analysis)	H&E-stained biopsy slides	Cohen's κ 0.868 vs. expert pathologists	External, multi-institutional	Reduces inter-observer variability and grading time
Renal cancer	Tramanzoli et al. 2023 ⁷³	RCC subtype classification	Radiomics + machine learning	Contrast-enhanced CT	AUC 0.91–0.97 (AI) vs. 0.69–0.79 (radiologists)	External datasets	Improves non-invasive tumor characterization
Renal cancer	Ursprung et al. 2020 ⁴¹	Benign vs. malignant renal mass differentiation	Radiomics-based ML	CT and MRI	Pooled AUC ~0.90; accuracy up to 94%	Meta-analysis with external validation	Supports decision-making in renal mass management
Bladder cancer	He et al. 2024 ⁵⁷	Prediction of muscle-invasive BCa	Deep learning radiomics	Multiparametric MRI	AUC 0.80–0.99; accuracy up to 92%	External validation cohorts	Enhances preoperative risk stratification
Bladder cancer	Kwong et al. 2024 ⁵⁶	Prediction of NMIBC recurrence and progression	Machine learning prognostic models	Clinical and pathological data	AUC 0.61–0.92; accuracy up to 100%	External and temporal validation	Supports individualized surveillance strategies

Note. Performance metrics are reported as described in the original publications. Values represent ranges across validation cohorts.

the potential to improve surgical training by providing unbiased evaluations, personalized feedback, and enhanced learning processes.⁸⁰ Furthermore, AI can set up *in-silico* clinical trials using synthetic datasets and evaluate outcomes of interest for populations or conditions that would be otherwise difficult to assess in a real-world study.⁸¹ Generative AI chatbots, such as ChatGPT and other Large Language Models (LLMs), can improve doctor-patient communication, facilitate access to medical knowledge, enhance patient care, and advance research and education.⁸² There is skepticism among physicians that

LLMs pose a risk to exacerbate scientific misconduct due to a lack of accountability and transparency.⁸² A recently published survey on the potential use of LLMs in urology, reported that 47.7% of urologists use LLMs in their academic practice, while 19.8% use them in clinical practice, and 62.6% are concerned about potential ethical implications.⁸³

Although AI models exhibit promising potential in improving patient care, there are ethical issues and potential risks that must be carefully considered. The European Parliamentary Research Service recently published a study on AI in healthcare, highlighting

concerns such as patient harm from AI errors, misuse of AI tools, safeguarding patient confidentiality, and issues regarding transparency and privacy/security.⁸⁴ The use of AI in healthcare introduces new potential errors that result from misclassifications made by a computational algorithm.^{85,86} Discrepancies between data that is used to train AI algorithms and real-world scenarios are example. A recent review addressed the impact of AI errors across diagnostic specialties and especially in histopathology, proposing the different stages at which an AI-algorithm error should be considered.⁸⁵ Further research on AI should focus on key areas such as refining diagnostic algorithms, developing personalized treatment strategies, exploring digital biomarkers, implementing clinical decision support systems, addressing data interoperability challenges, assessing long-term patient outcomes, and navigating ethical and regulatory considerations. The integration of AI-based augmented reality techniques in prostate and kidney surgery is no longer a distant hope for the future.^{86–89} Furthermore, AI-based technologies to aid surgeons during robot-assisted radical prostatectomy are already available, such as the BLAIR software, which is capable of forecasting instances of intraoperative bleeding, illustrating the positive impact of AI on surgical procedures.⁹⁰ The scarcity of studies evaluating AI in other urological cancers highlights an evidence gap, indicating that this area is not yet supported by robust clinical or methodological research.

Effective collaboration between stakeholders is crucial for explainable AI development. EAU has launched two EU-funded projects (PIONEER and OPTIMA) to unlock the potential of big data for improving patient management. PIONEER is designed to standardise and analyse high-quality big data from diverse populations of PCa patients to provide evidence-based data for improving decision-making.⁹¹ Whereas, OPTIMA implements data from multiple stakeholders and, using AI-driven technology, creates dynamic, computer-interpretable guidelines to facilitate an effective shared decision-making by clinicians and patients.⁹²

The limitations of this overview are related to the weaknesses of the body of evidence. Primary studies reveal diverse methodologies, such as differences in data acquisition and validation techniques. Another limitation is the potential risk for oversampling using synthetic data, which was not easily recognized. Among the strengths of this overview is the systematic approach based on Cochrane methodology, the pre-specified protocol, and the thorough literature search, including all relevant databases.

Furthermore, this review is the result of a collaborative effort between urologists involved in patient care, a methodology expert, and an AI expert.

Conclusions

The integration of AI models in urological oncology represents a significant advancement with important implications for patient care. As demonstrated across numerous SRs, AI models can match or exceed human performance in imaging interpretation, histopathological grading, and outcome prediction. Addressing the substantial variability in AI-model performance is crucial to ensure its generalizability and reliable integration in clinical practice. However, AI's ability to analyze extensive datasets and uncover subtle patterns within complex medical information supports a move toward precision, patient-centered care. Furthermore, AI implementation could boost personalized treatment care by integrating patient-specific data, including genomics, clinicopathological, and imaging data, to optimize treatment and minimize adverse effects. However, clinical implementation remains in its infancy. Methodological weaknesses, limited external validation, and lack of standardized reporting currently preclude widespread clinical use. At present, AI tools should be viewed as decision-support systems that augment, rather than replace, clinical expertise. The high proportion of reviews rated at high risk of bias reduces the certainty of the evidence, and therefore the findings should be interpreted as hypothesis-generating rather than practice-informing. Future studies should prioritize prospective validation, standardized outcome definitions, and assessment of real-world clinical impact to ensure safe, reliable, and equitable adoption in routine cancer care. AI literacy should be integrated into medical training to prepare future healthcare professionals to work with AI, understand its potential, and appreciate its limitations. Rather than replacing clinicians, AI should be embraced as a powerful ally—enhancing accuracy, supporting decision-making, and ultimately improving patient outcomes.

Acknowledgement

Not applicable.

Funding Statement

The authors received no specific funding for this study.

Author Contributions

Study Conception and Design: Vasileios Sakalis, Muhammad Imran Omar, Yuhong Yuan, James N'Dow; Data Collection: Karl H. Pang, Bhavan P. Rai, Lisa Moris, Eftichia Chatzigriva, Isaak Filippidis, Maria Chalkidou, Dimitrios Papanikolaou; Analysis and Interpretation of Results: Vasileios Sakalis, Muhammad Imran Omar, Michael Bussmann, Eftichia Chatzigriva, Isaak Filippidis, Maria Chalkidou, Dimitrios Papanikolaou; Draft Manuscript Preparation: Vasileios Sakalis, Eftichia Chatzigriva, Muhammad Imran Omar. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials

The datasets analysed during the current study are available from the corresponding author upon reasonable request.

Ethics Approval

Not applicable.

Conflicts of Interest

The authors declare no conflicts of interest.

Supplementary Materials

The supplementary material is available online at <https://www.techscience.com/doi/10.32604/cju.2026.077632/s1>. **Supplementary Material S1:** The search strategy. **Supplementary Fig. S1.** PRISMA flow diagram. **Supplementary Material S2.** PRISMA 2020 checklist. **Supplementary Table S1.** Citation Matrix for Prostate Cancer studies. **Supplementary Table S2.** Citation Matrix for Kidney Cancer studies. **Supplementary Table S3.** Citation Matrix for Bladder Cancer studies.

References

1. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. *N Engl J Med* 2023;388(21):1981–1990. doi:10.1056/nejmra2301725.
2. Cornford P, Tilki D, van den Bergh RCN et al. EAU-EANM-ESTRO-ESUR-ISUP-SIOG guidelines on prostate cancer 2024. In: EAU Annual Meeting in Paris 2024. Arnhem, The Netherlands: EAU Guidelines Office; 2024.
3. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci* 2019;40(8):577–591. doi:10.1016/j.tips.2019.05.005.
4. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278(2):563–577.
5. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York, NY, USA: John Wiley and Sons; 2000.
6. Lunney C, Pieper D, Thabet P, Kanji S. Managing overlap of primary study results across systematic reviews: practical considerations for authors of overviews of reviews. *BMC Med Res Methodol* 2021;21(1):140. doi:10.1186/s12874-021-01269-y.
7. Pieper D, Antoine SL, Mathes T, Neugebauer EA, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. *J Clin Epidemiol* 2014;67(4):368–375. doi:10.1016/j.jclinepi.2013.11.007.
8. Sihotang RC, Agustino C, Huang F, Parikesit D, Rahman F, Hamid ARAH. Accuracy of machine learning models using ultrasound images in prostate cancer diagnosis: a systematic review. *Med J Indones* 2023;32(2):112–121. doi:10.13181/mji.oa.236765.
9. Sushentsev N, Moreira Da Silva N, Yeung M et al. Comparative performance of fully-automated and semi-automated artificial intelligence methods for the detection of clinically significant prostate cancer on MRI: a systematic review. *Insights Imaging* 2022;13(1):59. doi:10.1186/s13244-023-01594-4.
10. Michaely HJ, Aringhieri G, Cioni D, Neri E. Current value of biparametric prostate MRI with machine-learning or deep-learning in the detection, grading, and characterization of prostate cancer: a systematic review. *Diagnostics* 2022;12(4):799. doi:10.3390/diagnostics12040799.
11. Liberini V, Laudicella R, Balma M et al. Radiomics and artificial intelligence in prostate cancer: new tools for molecularhybrid imaging and theragnostics. *Eur RadiolExp* 2022;6(1):27. doi:10.1186/s41747-022-00282-0.
12. Ghezzi S, Bezzi C, Presotto L et al. State of the art of radiomic analysis in the clinical management of prostate cancer: a systematic review. *Crit Rev Oncol Hematol* 2022;169:103544. doi:10.1016/j.critrevonc.2021.103544.
13. Castaldo R, Cavaliere C, Soricelli A, Salvatore M, Pecchia L, Franzese M. Radiomic and genomic machine learning method performance for prostate cancer diagnosis: systematic literature review. *J Med Internet Res* 2021;23(4):e22394. doi:10.2196/22394.
14. Van Booven DJ, Kuchakulla M, Pai R et al. A systematic review of artificial intelligence in prostate cancer. *Res Rep Urol* 2021;13:31–39. doi:10.2147/rru.s268596.
15. Sugano D, Sanford D, Abreu A, Duddalwar V, Gill I, Cacciamani GE. Impact of radiomics on prostate cancer detection: a systematic review of clinical applications. *Curr Opin Urol* 2020;30(6):754–781. doi:10.1097/mou.0000000000000822.
16. Cuocolo R, Cipullo MB, Stanzione A et al. Machine learning for the identification of clinicallysignificant prostate cancer

- on MRI: a meta-analysis. *Eur Radiol* 2020;30(12):6877–6887. doi:10.1007/s00330-020-07027-w.
17. Almeida G, Tavares JMRS. Deep learning in radiation oncology treatment planning for prostate cancer: a systematic review. *J Med Syst* 2020;44(10):179. doi:10.1007/s10916-020-01641-3.
18. Castillo JTM, Arif M, Niessen WJ, Schoots IG, Veenland JF. Automated classification of significant prostate cancer on MRI: a systematic review on the performance of machine learning applications. *Cancers* 2020;12(6):1606. doi:10.3390/cancers12061606.
19. Schröder F, Kattan MW. The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review. *Eur Urol* 2008;54(2):274–290. doi:10.1016/j.eururo.2008.05.022.
20. Morozov A, Taratkin M, Bazarkin A et al. A systematic review and meta-analysis of artificial intelligence diagnostic accuracy in prostate cancer histology identification and grading. *Prostate Cancer Prostatic Dis* 2023;26(4):681–692. doi:10.1038/s41391-023-00673-3.
21. Quartuccio N, Marrale M, Laudicella R et al. The role of PET radiomic features in prostate cancer: a systematic review. *Clin Transl Imaging* 2021;9(6):579–588. doi:10.1007/s40336-021-00436-x.
22. Syer T, Mehta P, Antonelli M et al. Artificial intelligence compared to radiologists for the initial diagnosis of prostate cancer on magnetic resonance imaging: a systematic review and recommendations for future studies. *Cancers* 2021;13(13):3318. doi:10.3390/cancers13133318.
23. Lindgren Belal S, Frantz S, Minarik D et al. Applications of artificial intelligence in PSMA PET/CT for prostate cancer imaging. *Semin Nucl Med* 2024;54(1):141–149. doi:10.1053/j.semnucmed.2023.06.001.
24. Huynh LM, Hwang Y, Taylor O, Baine MJ. The use of MRI-derived radiomic models in prostate cancer risk stratification: a critical review of contemporary literature. *Diagnostics* 2023;13(6):1128. doi:10.3390/diagnostics13061128.
25. Naik N, Tokas T, Shetty DK et al. Role of deep learning in prostate cancer management: past, present and future based on a comprehensive literature review. *J Clin Med* 2022;11(13):3575. doi:10.3390/jcm11133575.
26. Nematollahi H, Moslehi M, Aminolroayaei F, Maleki M, Shahbazi-Gahrouei D. Diagnostic performance evaluation of multiparametric magnetic resonance imaging in the detection of prostate cancer with supervised machine learning methods. *Diagnostics* 2023;13(4):806. doi:10.3390/diagnostics13040806.
27. Rouviere O, Jaouen T, Baseilhac P et al. Artificial intelligence algorithms aimed at characterizing or detecting prostate cancer on MRI: how accurate are they when tested on independent cohorts?—A systematic review. *Diagn Interv Imaging* 2023;104(5):221–234. doi:10.1016/j.diii.2022.11.005.
28. Spohn SKB, Bettermann AS, Bamberg F et al. Radiomics in prostate cancer imaging for a personalized treatment approach-current aspects of methodology and a systematic review on validated studies. *Theranostics* 2021;11(16):8027–8042. doi:10.7150/thno.61207.
29. Telecan T, Andras I, Crisan N et al. More than meets the eye: using textual analysis and artificial intelligence as decision support tools in prostate cancer diagnosis—a systematic review. *J Pers Med* 2022;12(6):983. doi:10.3390/jpm12060983.
30. Xing X, Zhao X, Wei H, Li Y. Diagnostic accuracy of different computer-aided diagnostic systems for prostate cancer based on magnetic resonance imaging: a systematic review with diagnostic meta-analysis. *Medicine* 2021;100(3):e23817. doi:10.1097/md.00000000000023817.
31. Frewing A, Gibson AB, Robertson R, Urie PM, Corte DD. Don't fear the artificial intelligence: a systematic review of machine learning for prostate cancer detection in pathology. *Arch Pathol Lab Med* 2024;148(5):603–612. doi:10.5858/arpa.2022-0460-RA.
32. Antolin A, Roson N, Mast R et al. The role of radiomics in the prediction of clinically significant prostate cancer in the PI-RADS v2 and v2.1 era: a systematic review. *Cancers* 2024;16(17):2951. doi:10.3390/cancers16172951.
33. Calimano-Ramirez LF, Virarkar MK, Hernandez M et al. MRI-based nomograms and radiomics in presurgical prediction of extraprostatic extension in prostate cancer: a systematic review. *Abdom Radiol* 2023;48(7):2379–2400. doi:10.1007/s00261-023-03924-y.
34. Faiella E, Vaccarino F, Ragone R et al. Can machine learning models detect and predict lymph node involvement in prostate cancer? a comprehensive systematic review. *J Clin Med* 2023;12(22):7032. doi:10.3390/jcm12227032.
35. Guerra A, Wang H, Orton MR et al. Prediction of extracapsular extension of prostate cancer by MRI radiomic signature: a systematic review. *Insights Imag* 2024;15(1):217. doi:10.1186/s13244-024-01776-8.
36. Liu J, Cundy TP, Woon DTS, Lawrentschuk N. A systematic review on artificial intelligence evaluating metastatic prostatic cancer and lymph nodes on PSMA PET scans. *Cancers* 2024;16(3):486. doi:10.3390/cancers16030486.
37. Liu J, Cundy TP, Woon DTS, Desai N, Palaniswami M, Lawrentschuk N. A systematic review on artificial intelligence evaluating PSMA PET scan for intraprostatic cancer. *BJU Int* 2024;134(5):714–722. doi:10.1111/bju.16412.
38. Rabilloud N, Allaume P, Acosta O et al. Deep learning methodologies applied to digital pathology in prostate cancer: a systematic review. *Diagnostics* 2023;13(16):2676. doi:10.3390/diagnostics13162676.
39. Wang H, Xia Z, Xu Y, Sun J, Wu J. The predictive value of machine learning and nomograms for lymph node metastasis of prostate cancer: a systematic review and meta-analysis. *Prostate Cancer Prostatic Dis* 2023;26(3):602–613. doi:10.1038/s41391-023-00704-z.
40. Rabaan AA, Bakhrebah MA, AlSaihati H et al. Artificial intelligence for clinical diagnosis and treatment of prostate cancer. *Cancers* 2022;14(22):5595. doi:10.3390/cancers14225595.
41. Ursprung S, Beer L, Bruining A et al. Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma—a systematic review and meta-analysis. *Eur Radiol* 2020;30(6):3558–3566. doi:10.1007/s00330-020-06666-3.
42. Ferro M, Crocetto F, Barone B et al. Artificial intelligence and radiomics in evaluation of kidney lesions: a comprehensive literature review. *Ther Adv Urol* 2023;15:17562872231164803.
43. Dehghani Firouzabadi F, Gopal N, Homayounieh F et al. CT radiomics for differentiating oncocytoma from renal cell carcinomas: systematic review and meta-analysis. *Clin Imaging* 2023;94:9–17. doi:10.1016/j.clinimag.2022.11.007.
44. Roussel E, Capitanio U, Kutikov A et al. Novel imaging methods for renal mass characterization: a collaborative review. *Eur Urol* 2022;81(5):476–488. doi:10.1016/j.eururo.2022.01.040.
45. Kocak B, Kaya OK, Erdim C, Kus EA, Kilickesmez O. Artificial intelligence in renal mass characterization: a systematic review of methodologic items related to modeling, performance evaluation, clinical utility, and transparency. *AJR Am J Roentgenol* 2020;215(5):1113–1122. doi:10.2214/ajr.20.22847.

46. Pandey M, Gupta AK. A systematic review of the automatic kidney segmentation methods in abdominal images. *Biocybern Biomed Eng* 2021;41(4):1601–1628. doi:10.1016/j.bbe.2021.10.006.
47. Muhlbauer J, Egen L, Kowalewski K-F et al. Radiomics in renal cell carcinoma—a systematic review and meta-analysis. *Cancers* 2021;13(6):1348. doi:10.3390/cancers13061348.
48. Bhandari A, Ibrahim M, Sharma C, Liong R, Gustafson S, Prior M. CT-based radiomics for differentiating renal tumours: a systematic review. *Abdom Radiol* 2021;46(5):2052–2063. doi:10.1007/s00261-020-02832-9.
49. Chandramohan D, Garapati HN, Nangia U et al. Diagnostic accuracy of deep learning in detection and prognostication of renal cell carcinoma: a systematic review and meta-analysis. *Front Med* 2024;1:1447057. doi:10.3389/fmed.2024.1447057.
50. Khene ZE, Kammerer-Jacquet SF, Bigot P et al. Clinical application of digital and computational pathology in renal cell carcinoma: a systematic review. *Eur Urol Oncol* 2024;7(3):401–411. doi:10.1016/j.euo.2023.10.018.
51. Xi Z, Li J, Liu X, Gao X, Chen Y. The role of radiomics in computed tomography, magnetic resonance imaging, and ultrasound for renal tumors' diagnoses: a systematic review. *J Radiat Res Appl SC* 2024;17(1):100769. doi:10.1016/j.jrras.2023.100769.
52. Kozikowski M, Suarez-Ibarrola R, Osiecki R et al. Role of radiomics in the prediction of muscle-invasive bladder cancer: a systematic review and meta-analysis. *Eur Urol Focus* 2022;8(3):728–738. doi:10.1016/j.euf.2021.05.005.
53. Cacciamani GE, Nassiri N, Varghese B et al. Radiomics and bladder cancer: current status. *Bladder Cancer* 2020;6(3):343–362. doi:10.3233/blc-200293.
54. Huang X, Wang X, Lan X, Deng J, Lei Y, Lin F. The role of radiomics with machine learning in the prediction of muscle-invasive bladder cancer: a mini review. *Front Oncol* 2022;12:990176. doi:10.3389/fonc.2022.990176.
55. Liu YS, Thaliffdeen R, Han S, Park C. Use of machine learning to predict bladder cancer survival outcomes: a systematic literature review. *Expert Rev Pharmacoecon Outcomes Res* 2023;23(7):761–771. doi:10.1080/14737167.2023.2224963.
56. Kwong JCC, Wu J, Malik S et al. Predicting non-muscle invasive bladder cancer outcomes using artificial intelligence: a systematic review using APPRAISE-AI. *npj Digit Med* 2024;7(1):98. doi:10.1038/s41746-024-01088-7.
57. He C, Xu H, Yuan E et al. The accuracy and quality of image-based artificial intelligence for muscle-invasive bladder cancer prediction. *Insights Imaging* 2024;15(1):185. doi:10.1186/s13244-024-01780-y.
58. Khoraminia F, Fuster S, Kanwal N et al. Artificial intelligence in digital pathology for bladder cancer: hype or hope? A systematic review. *Cancers* 2023;15(18):4518. doi:10.3390/cancers15184518.
59. Suartz CV, Martinez LM, Cordeiro MD et al. Artificial intelligence for predicting response to neoadjuvant chemotherapy for bladder cancer: a comprehensive systematic review and meta-analysis. *Can Urol Assoc J* 2024;18(9):E276–E284. doi:10.5489/auaj.8681.
60. Chan EO, Pradere B, Teoh JY. European association of urology – young academic urologists (EAU-YAU) urothelial carcinoma working group. The use of artificial intelligence for the diagnosis of bladder cancer: a review and perspectives. *Curr Opin Urol* 2021;31(4):397–403.
61. Kirienko M, Biroli M, Gelardi F et al. Deep learning in Nuclear Medicine—focus on CNN-based approaches for PET/CT and PET/MR: where do we stand? *Clin Transl Imaging* 2021;9(1):37–55. doi:10.1007/s40336-021-00411-6.
62. Bazarkin A, Morozov A, Androsof A et al. Assessment of prostate and bladder cancer genomic biomarkers using artificial intelligence: a systematic review. *Curr Urol Rep* 2024;25(1):19–35. doi:10.1007/s11934-023-01193-2.
63. Papalia GF, Brigato P, Sisca L et al. Artificial intelligence in detection, management, and prognosis of bone metastasis: a systematic review. *Cancers* 2024;16(15):2700. doi:10.3390/cancers16152700.
64. Schneider L, Laiouar-Pedari S, Kuntz S et al. Integration of deep learning-based image analysis and genomic data in cancer pathology: a systematic review. *Eur J Cancer* 2022;160:80–91.
65. Wessels F, Kuntz S, Krieghoff-Henning E et al. Artificial intelligence to predict oncological outcome directly from hematoxylin and eosin-stained slides in urology. *Minerva Urol Nephrol* 2022;74(5):538–550. doi:10.23736/s2724-6051.22.04758-9.
66. Prabhu S, Prasad K, Robels-Kelly A, Lu X. AI-based carcinoma detection and classification using histopathological images: a systematic review. *Comput Biol Med* 2022;142(6):105209. doi:10.1016/j.combiomed.2022.105209.
67. Forsch S, Klauschen F, Hufnagl P, Roth W. Artificial intelligence in pathology. *Dtsch Arztebl Int* 2021;118(12):194–204. doi:10.3238/arztebl.m2021.0011.
68. Salem H, Soria D, Lund JN, Awwad A. A systematic review of the applications of Expert Systems (ES) and machine learning (ML) in clinical urology. *BMC Med Inform Decis Mak* 2021;21(1):223. doi:10.1186/s12911-021-01585-9.
69. Checcucci E, De Cillis S, Granato S, Chang P, Afyouni AS, Okhunov Z. Applications of neural networks in urology: a systematic review. *Curr Opin Urol* 2020;30(6):788–807. doi:10.1097/mou.0000000000000814.
70. Chen J, Remulla D, Nguyen JH et al. Current status of artificial intelligence applications in urology and their potential to influence clinical practice. *BJU Int* 2019;124(4):567–577. doi:10.1111/bju.14852.
71. Abbod MF, Catto JW, Linkens DA, Hamdy FC. Application of artificial intelligence to the management of urological cancer. *J Urol* 2007;178(4 Pt 1):1150–1156. doi:10.1016/j.juro.2007.05.122.
72. Faiella E, Santucci D, Calabrese A et al. Artificial intelligence in bone metastases: an MRI and CT imaging review. *Int J Environ Res Public Health* 2022;19(3):1880. doi:10.3390/ijerph19031880.
73. Tramanzoli P, Castellani D, De Stefano V et al. Radiomics vs. radiologist in bladder and renal cancer. Results from a systematic review. *Cent Eur J Urol* 2023;76:12–19. doi:10.5173/cej.2023.252.
74. Drouin SJ, Yates DR, Hupertan V, Cussenot O, Rouprêt M. A systematic review of the tools available for predicting survival and managing patients with urothelial carcinomas of the bladder and of the upper tract in a curative setting. *World J Urol* 2013;31(1):109–116. doi:10.1007/s00345-012-1008-9.
75. Paudyal R, Shah AD, Akin O et al. Artificial intelligence in CT and MR imaging for oncological applications. *Cancers* 2023;15(9):2573. doi:10.3390/cancers15092573.
76. Saha A, Bosma JS, Twilt JJ et al. Artificial intelligence and radiologists in prostate cancer detection on MRI (PI-CAI): an international, paired, non-inferiority, confirmatory study. *Lancet Oncol* 2024;25(7):879–887. doi:10.1016/j.eururo.2024.11.007.

77. Bulten W, Kartasalo K, Chen PC et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med* 2022;28(1):154–163. doi:10.1016/s0302-2838(22)00693-5.
78. Al Kuwaiti A, Nazer K, Al-Reedy A et al. A review of the role of artificial intelligence in healthcare. *J Pers Med* 2023;13(6):951. doi:10.3390/jpm13060951.
79. Arora A, Arora A. Generative adversarial networks and synthetic patient data: current challenges and future perspectives. *Future Healthc J* 2022;9(2):190–193. doi:10.7861/fhj.2022-0013.
80. Rodriguez Peñaranda N, Eissa A, Ferretti S et al. Artificial intelligence in surgical training for kidney cancer: a systematic review of the literature. *Diagnostics* 2023;13(19):3070. doi:10.3390/diagnostics13193070.
81. van Wijk Y, Ramaekers B, Vanneste BGL et al. Modeling-based decision support system for radical prostatectomy versus external beam radiotherapy for prostate cancer incorporating an *in silico* clinical trial and a cost-utility study. *Cancers* 2021;13(11):2687. doi:10.3390/cancers13112687.
82. Clusmann J, Kolbinger FR, Muti HS et al. The future landscape of large language models in medicine. *Commun Med* 2023;3(1):141. doi:10.1038/s43856-023-00370-1.
83. Eppler M, Ganjavi C, Ramacciotti LS et al. Awareness and use of ChatGPT and large language models: a prospective cross-sectional global survey in urology. *Eur Urol* 2024;85(2):146–153. doi:10.23736/S2724-6051.25.06133-6.
84. Scientific and technological Options Assessment (STOA). STOAwebsite. Brussels, Belgium: European Parliament; [cited 2025 Dec 10]. Available from: <http://www.europarl.europa.eu/stoa>.
85. Evans H, Snead D. Understanding the errors made by artificial intelligence algorithms in histopathology in terms of patient impact. *npj Digit Med* 2024;7(1):89. doi:10.1038/s41746-024-01093-w.
86. Challen R, Denny J, Pitt M et al. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;28(3):231–237. doi:10.1136/bmjqs-2018-008370.
87. Hameed BMZ, Dhavileswarapu AVL, Raza SZ et al. Artificial intelligence and its impact on urological diseases and management: a comprehensive review of the literature. *J Clin Med* 2021;10(9):1864. doi:10.3390/jcm10091864.
88. De Backer P, Van Praet C, Simoens J et al. Improving augmented reality through deep learning: real-time instrument delineation in robotic renal surgery. *Eur Urol* 2023;84(1):86–91. doi:10.1016/j.eururo.2023.02.024.
89. Checcucci E, Piana A, Volpi G et al. Three-dimensional automatic artificial intelligence driven augmented-reality selective biopsy during nerve-sparing robot-assisted radical prostatectomy: a feasibility and accuracy study. *Asian J Urol* 2023;10(4):407–415. doi:10.1016/j.ajur.2023.08.001.
90. Checcucci E, Piazzolla P, Marullo G et al. Development of bleeding artificial intelligence detector (BLAIR) system for robotic radical prostatectomy. *J Clin Med* 2023;12(23):7355. doi:10.3390/jcm12237355.
91. EAU UroEvidenceHub. PIONEER+. Arnhem, The Netherlands: Stichting EAU UroEvidence Hub; [cited 2025 Dec 10]. Available from: <https://prostate-pioneer.eu>.
92. OPTIMA. Tackling cancer through real world data. Munich, Germany: ARTTIC Innovation GmbH; [cited 2025 Dec 10]. Available from: <https://www.optima-oncology.eu>.