Review

Multimodal AI for Prostate Cancer Diagnosis and Risk Stratification: Scoping Review

Jesus Alejandro Alzate-Grisales[1,2,3,4], MSc; Felipe Ocampo Osorio[5,6], MSc; Jose Daniel Restrepo Orozco[7]; Juan Diego Rendon Tabares[7]; Miguel Perán-Teruel[8], MD, PhD; Azahar Navarro Beltrán[8], MD; Clara Ruiz Torres[8], MD; Francisco García-García[4], PhD; Reinel Tabares-Soto[7,9,10,11], PhD; Juan M García-Gómez[12], PhD; Maria de la Iglesia-Vayá[1,13], PhD

[1]Unidad Mixta de Imagen Biomédica e Inteligencia Artificial FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, Valencia, Spain
[2]Sistemas Informáticos y de Comunicaciones y Economía de Tecnologías Sanitarias Research Team, Universitat Politècnica de València, Valencia, Spain
[3]Doctoral Program in Health and Wellbeing Technologies, Universitat Politècnica de València, Valencia, Spain
[4]Computational Biomedicine Laboratory, ríncipe Felipe Research Center, Valencia, Spain
[5]Centro de Investigaciones Clínicas, Fundación Valle del Lili, Cali, Valle del Cauca Department, Colombia
[6]Departamento de Salud Pública y Medicina Comunitaria, Icesi University, Cali, Valle del Cauca Department, Colombia
[7]Department of Systems and Informatics, University of Caldas, Manizales, Caldas Department, Colombia
[8]Department of Urology, Hospital Arnau de Vilanova, Valencia, Spain
[9]Department of Electronics and Automation, Universidad Autónoma de Manizales, Manizales, Caldas Department, Colombia
[10]Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Santiago, Chile
[11]GobLab School of Government, Universidad Adolfo Ibáñez, Santiago, Chile
[12]BDSLab, Instituto Universitario de Tecnologías de la Información y Comunicaciones, Universitat Politècnica de València, Valencia, Spain
[13]Conselleria de Sanidad, Generalitat Valenciana, Valencia, Spain

Corresponding Author:

Jesus Alejandro Alzate-Grisales, MSc
Unidad Mixta de Imagen Biomédica e Inteligencia Artificial FISABIO-CIPF
Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana
21 Catalunya Avenue, Benimaclet
Valencia 46020
Spain
Phone: 34 961925700
Email: jesus.alzate@fisabio.es

Abstract

Background: Multimodal AI may improve clinically significant prostate cancer (csPCa) detection and risk stratification by integrating imaging, histopathology, molecular, radiomic, and structured clinical data.

Objective: This scoping review aimed to map recent evidence on multimodal AI for prostate cancer diagnosis, staging, prognosis, and treatment personalization, with emphasis on data modalities, fusion strategies, validation, reproducibility, and clinical translation. Its added value is a decision-centered evidence map linking fusion strategy and input burden to clinical task, validation strength, reproducibility, and translational readiness across the prostate cancer pathway.

Methods: A PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews)–guided search of IEEE Xplore, Scopus, Web of Science, PubMed, and SpringerLink was conducted to identify English-language peer-reviewed studies published from January 2021 to December 2025. Eligible studies applied AI to at least 2 data sources for prostate cancer detection, grading, staging, prognosis, recurrence prediction, or treatment selection. Two reviewers screened records and charted clinical tasks, modalities, fusion strategy, validation design, performance, and reproducibility indicators.

Results: Twenty-six studies were included. Clinical variables were incorporated in 24 studies, multiparametric magnetic resonance imaging (MRI) in 17 studies, and whole-slide histopathology in 6 studies. Multimodal models usually outperformed unimodal baselines in paired comparisons. including prostate imaging: cancer AI (PI-CAI) performance above the median radiologist area under the receiver operating characteristic curve (AUROC; 0.91 vs 0.86) and trimodal positron emission

tomography (PET)/MRI/clinical models reporting area under the curve (AUC) values of up to 0.955. The evidence was limited by retrospective designs, small cohorts, geographic concentration, scarce public code, inconsistent calibration reporting, and a lack of prospective workflow validation.

Conclusions: Multimodal AI is promising for biopsy triage, grading, staging, prognostication, and treatment selection, but current evidence supports research prioritization rather than routine deployment. Prospective, diverse, transparent, and clinically embedded validation is required before multimodal AI can guide routine prostate cancer decisions.

J Med Internet Res 2026;28:e105912; doi: 10.2196/105912

Keywords: prostate cancer; multimodal AI; machine learning; magnetic resonance imaging; data fusion; scoping review

Introduction

Prostate cancer remains a major global health problem, with approximately 1.47 million new cases and almost 400 000 deaths estimated worldwide in 2022 [1,2]. The absolute burden is expected to rise as populations age, placing increasing pressure on diagnostic pathways that already depend on careful risk stratification [3]. Over the past decade, these pathways have improved through prebiopsy magnetic resonance imaging (MRI), targeted biopsy, standardized reporting, and risk-adapted treatment strategies [4]. Randomized evidence indicates that MRI-directed pathways increase the detection of clinically significant prostate cancer (csPCa), reduce the diagnosis of indolent disease, and spare some men from biopsy [5,6]. Multiparametric MRI (mpMRI) is now embedded in international guidelines for the diagnostic workup [4,7]. Nevertheless, prostate-specific antigen (PSA) testing has limited specificity, systematic biopsy is invasive and vulnerable to sampling error, and MRI interpretation varies across readers, scanners, protocols, and patient populations [8,9]. These limitations matter clinically because missed csPCa, unnecessary biopsy, and inappropriate risk stratification can each produce substantial patient harm.

AI has been investigated as a way to reduce this uncertainty, reflecting the broader expansion of deep learning across medical imaging and clinical decision support [10]. Most early work focused on single-modality imaging, particularly mpMRI and radiomics, in which quantitative image features are linked to diagnosis, grading, and prognosis [11]. More recent models combine multiple MRI sequences, positron emission tomography (PET)/computed tomography (CT), including prostate-specific membrane antigen (PSMA) PET, which improves the accuracy of nodal and metastatic staging in high-risk disease [12], digital pathology, genomics, proteomics, and structured clinical variables such as PSA, age, prostate volume, the Prostate Imaging Reporting and Data System (PI-RADS) score, and tumor stage. This direction mirrors a wider movement in medicine toward integrating imaging, text, and structured records within a single model [13,14]. Although multimodal AI commonly refers to the integration of heterogeneous data sources, terminology in medical imaging is not fully uniform. Multisequence MRI approaches may also be considered multimodal when sequences are processed as distinct information streams and explicitly fused. Accordingly, this review adopts a broad operational definition encompassing both cross-domain integration and explicit multisequence MRI fusion [13]. The premise is clinically plausible. Different data streams describe complementary aspects of tumor biology, anatomy, and patient risk. However, multimodal integration also introduces new methodological risks, including missing input modalities, overfitting in small cohorts, domain shift between institutions, limited interpretability, and unclear regulatory pathways.

The clinical appeal of multimodal AI is therefore different from the appeal of a purely image-based detector. A biopsy-triage tool must decide which men can safely avoid biopsy; a staging model must support surgical planning; a recurrence model must identify patients whose follow-up or treatment intensity should change; and a treatment-personalization model must improve decision-making beyond guideline-based risk groups. These tasks require evidence of validation, calibration, workflow integration, and patient safety, not only high areas under the curve (AUCs). For this reason, a scoping review is useful at the present stage: it can map which clinical questions have been studied, which data sources drive performance gains, and where methodological gaps prevent clinical translation.

Prior syntheses have provided broad narrative overviews of AI across prostate imaging, pathology, prognostication, and treatment prediction, or surveys of computer-aided diagnosis using multiple imaging modalities [15,16]. Neither was designed as a PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews)–guided evidence map of recent explicit fusion studies across the prostate cancer pathway, with joint charting of validation, reproducibility, missing-modality handling, and translational readiness. The added value of the present review is therefore a decision-centered map linking each integration strategy and its input burden to the clinical task, validation strength, and barriers to implementation, rather than providing an aggregate ranking of reported accuracy.

This scoping review maps the extent, design, and translational maturity of recent multimodal AI studies in prostate cancer. The primary objective was to characterize how multimodal AI has been applied to csPCa detection, grading, staging, prognosis, and therapy personalization. Secondary objectives were to summarize data fusion strategies, validation approaches, reported performance relative to unimodal baselines, reproducibility practices, and research priorities needed before clinical adoption.

Methods

Ethical Considerations

This scoping review synthesized previously published aggregate evidence and did not involve new recruitment, intervention, patient-level data extraction, or human tissue analysis.

Study Design

This scoping review was reported in accordance with the PRISMA-ScR guidelines [17]. The completed checklist is provided in Checklist 1. A scoping review design was selected because the field is heterogeneous in clinical tasks, end points, model architectures, data modalities, and validation strategies, making a single pooled meta-analysis inappropriate at this stage.

The research questions were structured around the population, concept, and context framework. The population was men with suspected or confirmed prostate cancer. The concept was multimodal AI, defined as a model integrating at least 2 distinct data sources, including imaging, histopathology, molecular data, radiomics, or structured clinical variables. Studies integrating multiple MRI sequences through explicit fusion architectures were retained when the model treated the sequences as independent information channels; because this definition may be broader than some readers expect, it was recorded during charting and considered in interpretation. The context included screening, biopsy triage, grading, staging, prognosis, treatment selection, and follow-up. The 5 prespecified review questions that guided data charting and synthesis are listed in Multimedia Appendix 1.

Eligibility Criteria

The eligibility criteria required peer-reviewed original studies or conference proceedings that applied AI or machine learning to at least 2 data sources for a prostate cancer–related task and reported sufficient methodological detail and quantitative performance metrics for evidence mapping. Studies were excluded if they were reviews, editorials, letters, case reports, pre-2021 publications, non-English reports, unimodal AI studies without cross-source integration, or studies that did not address csPCa-related diagnosis, grading, staging, prognosis, or treatment personalization as a primary clinical objective.

This boundary was necessary because several recent adjacent studies used related terminology but did not meet the final evidence-map definition. Examples included full-text-unavailable multimodal nomograms, radiology-clinical models without eligible fusion, MRI-only AI, lymph-node metastasis models, and bone metastasis prediction studies outside the csPCa-focused review question [18-22]. These reports were retained in the screening audit and exclusion table rather than synthesized as eligible evidence.

Information Sources and Search Strategy

Five databases were searched: IEEE Xplore, Scopus, Web of Science, PubMed, and SpringerLink. The searches covered English-language, peer-reviewed publications from January 2021 to December 2025 and were executed on January 12, 2026. The search combined terms for AI, prostate cancer, imaging, and nonimaging modalities, and multimodal integration. The complete Boolean search strings are provided in Multimedia Appendix 1. Records were exported in RIS format, imported into Mendeley (Elsevier), and deduplicated based on normalized title matching. Review articles and other contextual reports were excluded from the primary corpus but were used for contextual interpretation when relevant [15,16,23].

The search strategy deliberately included broad modality terms because multimodal AI in prostate cancer is not restricted to radiology. Candidate studies could integrate mpMRI, biparametric MRI, PET/CT, ultrasound, whole-slide pathology, genomic expression, proteomics, radiomics, structured clinical variables, or combinations of these sources. During screening, studies were retained only if the integration was central to the model rather than being a post hoc descriptive comparison. This distinction was important for studies using several MRI sequences. These were included only when the architecture explicitly fused sequence-specific inputs or features. This broader interpretation of multimodality was considered during synthesis. Detailed inclusion and exclusion criteria are provided in Multimedia Appendix 1.

The search strategy was structured into 4 Boolean blocks addressing the target population, AI methodology, imaging, and nonimaging modalities, and multimodal integration. The complete search combined terms for deep learning, AI, machine learning, neural networks, prostate cancer, csPCa, mpMRI, CT, PET, PSMA, ultrasound, histopathology, digital pathology, clinical reports, PSA, radiomics, multimodal modeling, data fusion, and hybrid or integrated models. SpringerLink contributed 997 of the 1211 initial records, reflecting its broader full-text indexing scope relative to the other databases. The subsequent deterministic title filter was applied to the pooled postdeduplication, nonreview set from all 5 databases and was not a SpringerLink-specific step. The full search strings for each of the 4 Boolean blocks, including database-specific syntax, are provided in Multimedia Appendix 1.

Study Selection

After removal of 53 duplicate records and 221 records identified as reviews or systematic reviews, a deterministic Python (Python Software Foundation) filter was applied to the pooled set of 937 records from all 5 databases. The script evaluated only the document title field. Titles were converted to lowercase, and literal, case-insensitive Python substring matching was used to create one binary indicator for each of 12 prespecified categories: deep learning, machine learning, AI, multimodality, mpMRI, CT, PET, ultrasound, histopathology, clinical data, radiomics, and prostate cancer. Each category was counted at most once per title. A record was

retained when at least 3 categories matched and the prostate-cancer category was one of them. The filter retained 153 records for manual title and abstract screening and excluded 784. During revision, rerunning the original implementation reproduced the retained set exactly. Two reviewers (JAA-G and FOO) then assessed a randomly selected sample of 200 of the 784 excluded records against the prespecified eligibility criteria, with final decisions resolved by consensus. None met the inclusion criteria (0/200; 0%; exact 95% CI 0%-1.8%). In retrospective known-positive checks, all 24 included studies that could be reliably linked to the historical input and all 52 records in the public full-text assessment file had passed the filter. Two included studies could not be reliably linked to the historical input and were therefore excluded from the known-positive denominator. These analyses provide retrospective evidence of reproducibility and potential false exclusions but do not constitute prospective validation of the title-only filter.

The literal title substrings were as follows: deep learning (deep learning, cnn, convolutional neural network, rnn, lstm, transformer, resnet, vgg, densenet); machine learning (machine learning, support vector machine, svm, random forest, decision tree, xgboost, and logistic regression); AI (AI, ai, neural network, and neural networks); multimodality (multimodal, multimodal, multimodality, data fusion, data integration, integrated model, hybrid model, and ensemble); mpMRI (multiparametric mri, mpmri, t2-weighted, diffusion-weighted, dwi, adc, and contrast-enhanced mri); CT (computed tomography, ct); PET (positron emission tomography, pet, and psma); ultrasound (transrectal ultrasound, trus, and micro-ultrasound); histopathology (whole-mount pathology, histopathology, and digital pathology); clinical data (psa, psa density, clinical data, and patient demographics); radiomics (radiomic features and radiomics); and prostate cancer (prostate, prostatic, prostate cancer, prostatic cancer, and pca). Matching used literal substrings rather than whole-word boundaries. The complete reproducibility details are provided in Multimedia Appendix 1 and the public repository.

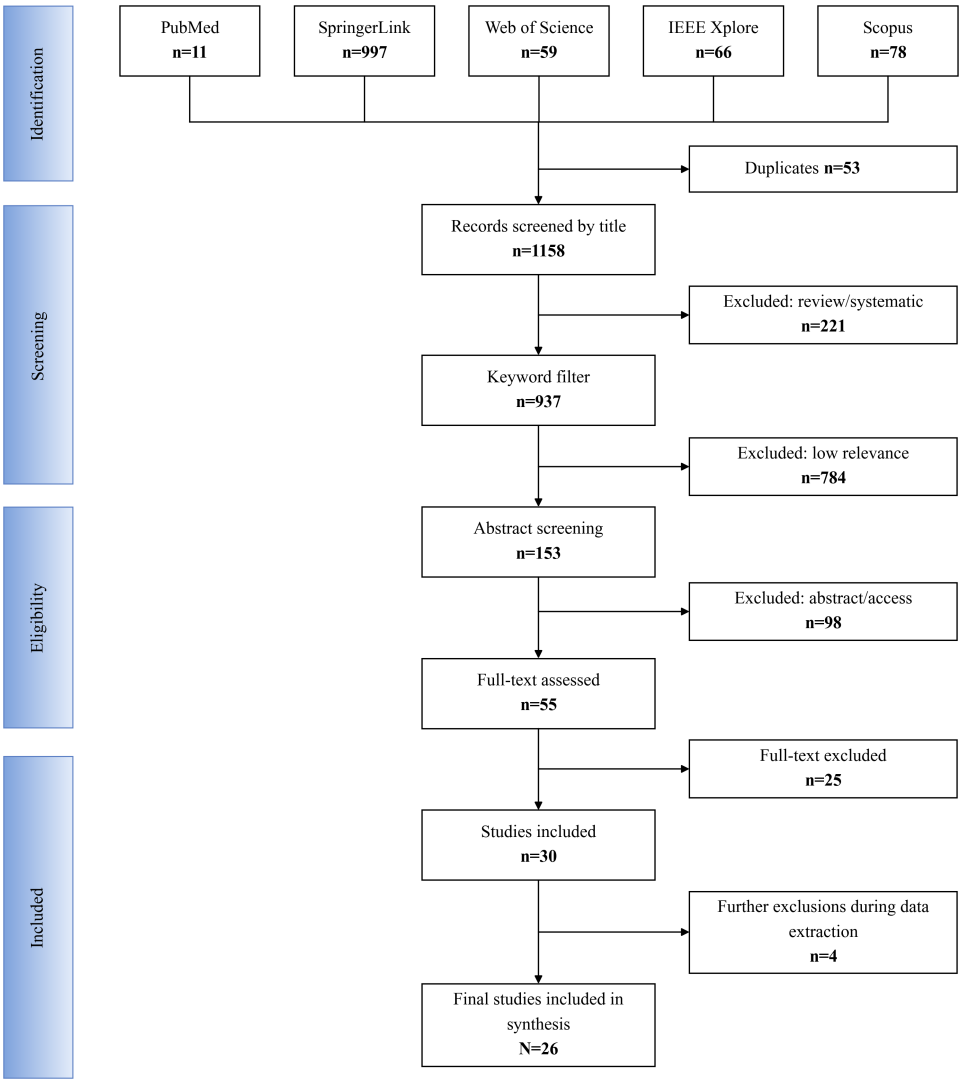
Data Charting and Synthesis

Data were charted using a standardized form covering author, year, country, sample size, clinical task, imaging modality, nonimaging modality, model architecture, fusion strategy, validation design, primary performance metric, public dataset use, and code availability. Findings were synthesized narratively by clinical task, modality, fusion strategy, validation level, reproducibility, and translational readiness. Formal risk-of-bias ratings were not used to exclude or rank studies because this scoping review aimed to map the breadth of evidence across diagnostic, prognostic, and treatment-selection designs, and a duplicate tool-based appraisal had not been prespecified. Prediction Model Risk Of Bias Assessment Tool+AI (PROBAST+AI) is designed for prediction model development and evaluation, whereas the Quality Assessment of Diagnostic Accuracy Studies–AI (QUADAS-AI) source identified for this review was a development protocol [24,25]. Neither provided a single, fully applicable framework for every study type in the corpus. We therefore did not retrofit a summary quality grade. To improve standardization without implying a formal risk-of-bias assessment, the charting framework grouped prespecified indicators across participants and data sources, predictors and modalities, outcomes and reference standards, and analysis and validation, aligned where applicable with PROBAST+AI domains, while retaining design-specific fields for sample size, external testing, calibration, uncertainty quantification, public code, and public data.

Results

The search identified 1211 records: 997 from SpringerLink, 78 from Scopus, 66 from IEEE Xplore, 59 from Web of Science, and 11 from PubMed. Deduplication removed 53 records, leaving 1158 records; 221 review or systematic-review records were then removed. Of the 937 records entering the automated title filter, 153 were retained, and 784 were excluded. The 153 retained records underwent manual title and abstract screening. Of these, 98 were excluded on the basis of abstract content or because the full-text was unavailable, and 55 proceeded to full-text assessment. After 25 full-text exclusions, with reasons documented, 30 studies were considered eligible, and 4 were excluded during final eligibility verification because they did not address csPCa as a primary clinical end point. The final corpus comprised 26 studies for evidence synthesis (Figure 1). Retrospective human assessment of a deterministic random sample of 200 excluded records identified no eligible studies (0/200; 0%; exact 95% CI, 0%-1.8%). The known-positive check showed 100% recall among the included studies that could be traced to the historical input (24/24); 2 studies could not be linked reliably and were excluded from that denominator. Full-text exclusions with reasons, the complete search logic, the title-filtering criteria, and the validation checks are provided in Multimedia Appendix 1.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram of study identification, screening, eligibility assessment, and inclusion.



The included studies were published from 2021 to 2025, with most appearing in 2024 or 2025, reflecting rapid acceleration in the field. The temporal, geographic, and validation distribution of the evidence base is important for interpretation. Twenty of the 26 included studies (76.9%) were published in 2024 or 2025 [26-45]. Studies originated predominantly from China (12/26, 46.2%) [30-32,34-37,39-41,46,47], followed by the United States (4/26, 15.4%) [33,42,48,49], Canada and France (2 each) [28,43], and single studies from Japan, Germany, Iran, India, and multicountry settings. This geographic skew has direct implications for model transportability because prostate cancer prevalence, screening intensity, MRI acquisition, scanner vendors, biopsy practice, genetic risk, and treatment thresholds differ substantially across health systems. Analytic cohorts ranged from 77 participants to 10,207 MRI examinations (9129 patients), with a median of 342 participants or cases. The clinical tasks were diverse. Five studies addressed prostate cancer (PCa)/csPCa detection or biopsy triage [26-29,48]; 4 [26-28,48] specifically targeted csPCa-related detection or triage, whereas Sherafatmandjoo et al [29] evaluated general PCa diagnosis and PI-RADS classification. Eight focused on Gleason score or International Society of Urological Pathology (ISUP) grade group prediction [30-34,46,50,51]. Three addressed adverse pathology, extraprostatic extension, seminal vesicle invasion, or positive surgical margins [35,36,47]. Two evaluated biochemical recurrence [37,38], 3 examined progression to castration-resistant or hormone-sensitive states [39-41], 2 used randomized clinical trial data for treatment personalization [42,49], one performed multitask prognosis [43], and 2 evaluated lesion detection or segmentation [44,45].

Clinical variables were incorporated in 24 of 26 studies [26-31,33-49,51], making them the most common nonimaging input. Only 2 of the 26 included studies (7.7%) relied exclusively on multisequence MRI fusion without incorporating non-MRI data sources [32,50]. Both addressed prostate cancer grading. Accordingly, 24 studies (92.3%) met a stricter cross-domain definition of multimodality, indicating that the broader inclusion of multisequence MRI-only approaches had a limited effect on the overall composition of the evidence base. mpMRI was used in 17 studies [29,31-33,35-37,39-41,44-48,50,51], usually combining T2-weighted imaging, diffusion-weighted imaging, apparent diffusion coefficient maps, and sometimes dynamic contrast

enhancement. PSMA PET/CT was included in 4 studies [30,31,34,35], fluorodeoxyglucose-PET/CT in one study, biparametric MRI in 3 studies [26-28], ultrasound in one study [45], and whole-slide histopathology in 6 studies [38,39,41,42,44,49]. Molecular data were uncommon: proteomics [41] and genomic expression data [38] each appeared in one study. Less frequent examples encountered during full-text mapping included *FOXA1* mutation prediction, tumor-abnormal-protein-enhanced mpMRI radiomics, B-mode/contrast-enhanced ultrasound radiomics, and PSMA PET/CT uptake classification [52-55]. This distribution shows that most current multimodal prostate AI is not yet broadly multiomic; rather, it typically combines MRI-derived information with routine clinical variables.

Fusion strategies varied. Late or decision-level fusion was the most common approach, followed by feature-level fusion, hybrid fusion, and early input-level fusion. More novel designs included conditional convolution, cross-attention transformer models, contrastive cross-modal learning, and Bayesian sequential modeling [33,35,37,38,43]. Only one study systematically compared early, intermediate, and late fusion on the same dataset, finding late fusion most effective for lesion detection while also offering practical robustness when some modalities are absent [44]. This lack of head-to-head fusion comparisons limits inference about which architecture is best for a given clinical task.

The distributional pattern in Table 1 has practical implications. First, the field is not evenly distributed across the prostate cancer pathway. Detection and grading dominate, whereas longitudinal prognosis, treatment response, economic evaluation, and implementation studies remain sparse. Second, clinical variables are not auxiliary; they are central model inputs in nearly all studies and sometimes account for large incremental gains. Third, the strongest evidence comes from a small number of large, well-curated cohorts, including prostate imaging: cancer AI (PI-CAI) [26] and NRG Oncology trials, whereas many architecturally innovative studies remain small, single-center, and retrospective. This split between scale and novelty should shape how results are interpreted.

Table 1. Key distributional findings across included studies.

Domain	Frequency	Interpretation
Clinical variables	24/26	Routine clinical data were almost universal and often drove performance gains without requiring additional imaging.
mpMRI[a]	17/26	MRI[b] remains the dominant imaging backbone for detection, grading, staging, and recurrence prediction.
Whole-slide histopathology	6/26	Pathology enabled prognostic and treatment-personalization tasks, especially in trial-derived cohorts.
PSMA[c] PET[d]/CT[e]	4/26	PET/CT was concentrated in grading and adverse pathology prediction, where high discrimination was reported, but costs and availability limit generalizability.
External validation	7/26	External testing was uncommon and frequently small, limiting confidence in transportability.
Public code/components	7/26	Reproducibility remains weak despite rapid technical development.
Uncertainty quantification	2/26	Few studies offered mechanisms for flagging unreliable predictions before clinical use.

[a]mpMRI: multiparametric magnetic resonance imaging.
[b]MRI: magnetic resonance imaging.
[c]PSMA: prostate-specific membrane antigen.
[d]PET: positron emission tomography.
[e]CT: computed tomography.

The detailed evidence map is presented in Tables 2 and 3. Table 2 summarizes the clinical task, modalities, cohort size, key performance metric, and validation design for each included study, while Table 3 reports the corresponding architecture, fusion strategy, and public code availability. An expanded narrative synthesis, extended descriptive statistics, and a thematic analysis organized by the 5 review questions are provided in Multimedia Appendix 1.

Table 2. Clinical characteristics of the 26 included studies, grouped by task category.

Study	Clinical task	Modalities	N	Key metric	Validation
PCa[a]/csPCa[b] detection and biopsy triage (n=5)					
Saha et al [26]	csPCa detection; AI vs radiologists	bpMRI[c] (T2WI[d], DWI[e]/ADC[f]); PSA[g], age, and scanner info	9129 patients (10,207 MRI[h] examinations)	AUROC[i] 0.91 vs 0.86 (62 radiologists)	Multireader; 11 centers (Netherlands/Norway)
Grabke et al [27]	Biopsy triage; workload reduction	bpMRI (T2WI, DWI/ADC); PSA density, age, prostate volume	2895 patients	NPV[j] 0.862; 12.5% workload reduction; Sensitivity 0.964	Simulated workflow; 2 Canadian centers
Taguelmimt et al [28]	csPCa detection+uncertainty quantification	bpMRI (T2WI, DWI/ADC); PSA, prostate volume, age	1295 patients	AUC[k] 0.82; uncertainty accuracy 0.85	PI-CAI[l] subset; 5-fold CV[m]+held out test
Sherafatmandjoo et al [29]	PCa diagnosis; PI-RADS[n] classification	mpMRI[o] (T2WI, T2-sag[p], DWI, ADC); PSA, age, and prostate volume	343 patients	Accuracy 0.96	Single-center (Iran); no external validation

Study	Clinical task	Modalities	N	Key metric	Validation
Zheng et al [48]	Biopsy prediction in MRI-negative cases	mpMRI (T2WI, DWI, ADC); PSA density, age, prostate volume	330 patients	NPV 0.99; Sensitivity 0.88	Single-center (UCLA)[q]; retrospective
Gleason score/ISUP[r] grade group prediction (n=8)					
Miao et al [31]	ISUP grading; few-shot learning	mpMRI (T2WI, DWI, ADC, DCE[s])+PSMA PET[t]/CT[u]; PSA, SUVmax[v]	341 patients	AUC 0.865 (vs 0.808 mpMRI; 0.818 PET/CT)	Single-center+small external (n=34)
Jiang et al [30]	psISUP[w] ≥4	PSMA[x] PET/CT; PSA density, clinical stage	222 patients	AUC 0.850 (test); 0.833 (external n=21)	10-fold CV+external; 2 Chinese hospitals
Liu et al [32]	18 Gleason grading (≤3+4 vs ≥4+3)	mpMRI (T2WI+ADC); peritumoral radiomics	650 patients	Accuracy: 0.907; F_1-score: 0.918	2-center retrospective; China
Khanfari et al [50]	PCa grading (Gleason) on mpMRI	mpMRI (T2WI, DWI, ADC, DCE); multisequence imaging fusion	111 patients	AUC approximately 0.95 (0.020); balanced accuracy 0.78	10-fold CV; public PROSTATEx-2
Zhong et al [33]	Gleason score/csPCa prediction	mpMRI (multisequence); PSA, PI-RADS, age	921 patients	AUROC 0.90 vs 0.69 imaging only	Internal test; public MRI dataset
Yang et al [34]	ISUP grade group prediction (PET radiomics)	^{18}F-PSMA[y] PET/CT; tPSA[z], SUVmax, clinical stage	356 patients	AUC 0.879 (internal); 0.861 (external)	External test; 2 Chinese centers
Bao et al [46]	Noninvasive ISUP grade assessment	mpMRI (T2WI, DWI, ADC, DCE); PSA, age, volume, PI-RADS	1442 patients	Macro-AUC 0.798 (internal); 0.762 (external)	Internal+external; 2 Chinese hospitals
Basso Dias et al [51]	Pathological grade group prediction	^{18}F-DCFPyL[aa] PET/MRI (PET +ADC+T2WI); PSA, age	105 patients	AUC 0.85 (combined)	5-fold CV; single center (Canada)
Adverse pathology, EPE[ab], ECE[ac], and staging prediction (n=3)					
Lin H et al [35]	Adverse pathology: EPE, SVI[ad], PSM[ae]	mpMRI+^{18}F-PSMA PET/CT; BMI, tPSA, volume, T stage[af]	341 patients	AUC 0.955 (95% CI: 0.932-0.975)	10-fold CV + internal; single center (China)
Yao et al [36]	EPE prediction	mpMRI+^{18}F-PSMA PET/CT; clinical data	233 patients	AUC 0.82 (vs 0.72 mpMRI alone)	External test (n=36); 2 Chinese hospitals
Hou et al [47]	ECE detection	mpMRI (T2WI, high-b DWI, ADC); PSA, GS[ag], PI-RADS, T stage	849 patients	AUC 0.857 (train); 0.807 (internal); 0.728 (external)	2-center+ external; China
Biochemical recurrence prediction (n=2)					
Lian et al [37]	Preoperative BCR[ah] prediction	Multisequence MRI (T2WI-FS, DWI); PSA, age	232 patients	AUC 0.835 (external); vs 0.612 clinical alone	External test; 3 Chinese hospitals
Dudhabhate et al [38]	BCR prediction; contrastive learning	WSI[ai] (H&E[aj]); genomic expression; clinical text	370 patients	AUC 0.821 (vs 0.748 WSI alone)	Hold-out split; TCGA-PRAD[ak] (public)
CRPC[al] and HSPC[am] progression prediction (n=3)					
Zhou et al [39]	CRPC progression risk prediction	mpMRI+H&E histopathology; PSA, GS, T stage, age	399 patients	AUC 0.86 (trimodal vs 0.72-0.78 unimodal)	Multicenter retrospective; 3 Chinese hospitals
He et al [40]	CRPC progression after 12 mo ADT[an]	mpMRI (T2WI, ADC, DCE); PSA, GS, TNM[ao] stage	96 patients	Accuracy: 94.2%; AUC: 0.945[ap]	5-fold CV; single center (China)
Lu et al [41]	18 HSPC → CRPC progression	MRI+H&E WSI+serum proteomics	77 patients	C-index:[aq] 0.906; AUC: 0.97[ar]	LOO-CV[as]; single center (China)
Therapy personalization and risk stratification (n=2)					
Esteva et al [49]	Therapy personalization; distant metastasis	H&E WSI; GS, T stage, PSA, age, treatment arm	5654 patients	5-yr DM[at] AUC 0.84 vs 0.74 NCCN[au]; 9.2%-14.6% relative AUC gain	Held-out validation; 5 NRG Oncology RCTs[av]
Tward et al [42]	Risk stratification; 10-year distant metastasis	H&E WSI; GS, T stage, PSA, age, NCCN risk	9787 patients	HR[aw] 3.2 (high vs low AI-risk); 42% reclassified	Cross-trial validation; 8 NRG Oncology RCTs
Multitask prognosis (n=1)					

Study	Clinical task	Modalities	N	Key metric	Validation
Larose et al [43]	Multitask prognosis (LNI[ax], BCR-FS[ay], MFS[az], dADT-FS[ba], CRPC-FS[bb], PCSS)[bc]	FDG-PET/CT[bd] (SUVmax, MTV[be], TLG)[bf]; PSA, GS, T/N-stage, age	295 patients	LNI AUC 0.71; outperformed clinical baselines in 5/6 tasks	Internal validation; single center (France)
csPCa lesion detection and segmentation (n=2)					
Gunashekar et al [44]	Lesion detection; fusion strategy comparison	mpMRI (T2WI, ADC, high-b DWI); PSA, GTV[bg], GPV[bh]	118 patients	DSC[bi] 0.36 (decision-level best); F_1-score 0.76	Internal test; single center (Germany)
Takeda et al [45]	csPCa prediction in low–intermediate PSA	US[bj] (B-mode)+ bpMRI (T2WI, DWI, ADC); PSA, age, DRE[bk]	178 patients	≤AUC 0.878; 0.862 (PSA≤20 ng/mL)	Train/test (80/20); single center (Japan)

[a]PCa: prostate cancer.
[b]csPCa: clinically significant prostate cancer.
[c]bpMRI: biparametric magnetic resonance imaging.
[d]T2WI: T2-weighted imaging.
[e]DWI: diffusion-weighted imaging.
[f]ADC: apparent diffusion coefficient.
[g]PSA: prostate-specific antigen.
[h]MRI: magnetic resonance imaging.
[i]AUROC: area under the receiver operating characteristic curve.
[j]NPV: negative predictive value.
[k]AUC: area under the curve.
[l]PI-CAI: prostate imaging: cancer AI.
[m]CV: cross validation.
[n]PI-RADS: Prostate Imaging Reporting and Data System.
[o]mpMRI: multiparametric magnetic resonance imaging.
[p]T2-sag: sagittal T2-weighted imaging.
[q]UCLA: University of California, Los Angeles.
[r]ISUP: International Society of Urological Pathology.
[s]DCE: dynamic contrast-enhanced.
[t]PET: positron emission tomography.
[u]CT: computed tomography.
[v]SUVmax: maximum standardized uptake value.
[w]psISUP: postsurgical International Society of Urological Pathology grade.
[x]PSMA: prostate-specific membrane antigen.
[y]^{18}F-PSMA: fluorine-18–labeled prostate-specific membrane antigen.
[z]tPSA: total prostate-specific antigen.
[aa]^{18}F-DCFPyL: Fluorine-18–labeled 2-(3-.
[ab]EPE: extraprostatic extension.
[ac]ECE: extracapsular extension.
[ad]SVI: seminal vesicle invasion.
[ae]PSM: positive surgical margin.
[af]T stage: tumor stage.
[ag]GS: Gleason score.
[ah]BCR: biochemical recurrence.
[ai]WSI: whole-slide image.
[aj]H&E: hematoxylin and eosin.
[ak]TCGA-PRAD: The Cancer Genome Atlas Prostate Adenocarcinoma dataset.
[al]CRPC: castration-resistant prostate cancer.
[am]HSPC: hormone-sensitive prostate cancer.
[an]ADT: androgen-deprivation therapy.
[ao]TNM: tumor-node-metastasis stage.
[ap]Small sample (n=96); high overfitting risk; no external validation.
[aq]C-index: concordance index.
[ar]AUC 0.97 from LOO-CV with n=77; C-index 0.906 is the more conservative metric
[as]LOO-CV: leave-one-out cross-validation.
[at]DM: distant metastasis.
[au]NCCN: National Comprehensive Cancer Network.
[av]RCT: randomized clinical trial.
[aw]HR: hazard ratio.
[ax]LNI: lymph-node involvement.
[ay]BCR-FS: biochemical recurrence-free survival.
[az]MFS: metastasis-free survival.
[ba]dADT-FS: definitive androgen-deprivation therapy-free survival.
[bb]CRPC-FS: castration-resistant prostate cancer-free survival.

[bc]PCSS: prostate cancer-specific survival.
[bd]FDG: fluorodeoxyglucose.
[be]MTV: metabolic tumor volume.
[bf]TLG: total lesion glycolysis.
[bg]GTV: gross tumor volume.
[bh]GPV: gross prostate volume.
[bi]DSC: Dice similarity coefficient.
[bj]US: ultrasound.
[bk]DRE: digital rectal examination.

Table 3. Architectural and methodological details of the 26 included studies. Architecture names are retained as reported in the source articles.

Study	Architecture	Fusion strategy	Code
Saha et al [26]	nnU-Net+nnDetection (12-model ensemble)	Early/intermediate; ensemble	Part[a]
Grabke et al [27]	nnU-Net (3D)+LR[b]	Late (AI score+clinical LR)	Yes[c]
Taguelmimt et al [28]	U-Net+TabNet; Bayesian MC[d] dropout	Late (bimodal prediction combination)	No[e]
Sherafatmandjoo et al [29]	†4× ResNet-50; FC[f] late fusion	Late (parallel multibranch)	No
Zheng et al [48]	Random Forest/SVM[g]/LR ensemble	Feature-level (radiomics+clinical)	No
Miao et al [31]	Siamese CL-MSNet; ResNet3D-34; few-shot KNN[h]	Intermediate (contrastive 5-modality)	No
Jiang et al [30]	ResNet-50; LASSO[i]; SHAP[j] nomogram	& lt; Late (DTL[k]+radiomics➞LR)	No
Liu et al [32]	CPCB+SE[l] Net CNN[m]+radiomics	Intermediate (radiomics+DL[n])	No
Khanfari et al [50]	ResNet/VGG/custom CNN; voting ensemble	Feature-level (8 fusion techniques)	Yes
Zhong et al [33]	ResNet-18+EmbedCondConv	Conditional (clinical modulates CNN)	Yes
Yang F et al [34]	Radiomics pipeline; LR/SVM	Feature-level (DL segmentation+radiomics+clinical)	No
Bao et al [46]	DenseNet-121+ResNet-50+VGG-16 (stacked)	Intermediate (feature concatenation+stacking)	No
Basso Dias et al [51]	Random Forest/SVM/LR	Feature-level (PET[o]+MRI[p] radiomic concat.)	No
Lin et al [35]	ResNet-50+Vision Transformer; LR nomogram	Hybrid (per-modality CNN+ViT[q]; decision LR)	No
Yao et al [36]	ResNet-50 (input-channel concatenation)	Early (multichannel input fusion)	No
Hou et al [47]	PAGNet:[r] ResNeXt+CBAM[s]+prior-attention	Intermediate (attention; expert prior)	No
Lian et al [37]	CNN+Transformer (MRMS-CNNFormer)	Intermediate (CNN+Transformer)	No
Dudhabhate et al [38]	Vision Transformer+text encoder (PCa-CoViT)[t]	Cross-modal contrastive (vision-text)	Part
Zhou et al [39]	ResNet-50+XGBoost[u] nomogram	Hybrid (intermediate+late; trimodal)	No
He et al [40]	DenseNet+SE SPP[v]+XGBoost	Intermediate (DL+clinical ➞ XGBoost)	Yes
Lu et al [41]	LASSO+Cox regression; ResNet-50	Feature-level (multiomics; nomogram)	No
Esteva et al [49]	Self-supervised CNN+CatBoost[w]	Late (DL features+clinical)	No
Tward et al [42]	Locked MMAI[x] DL model (pretrained)	Late (DL features+clinical risk scores)	No
Larose et al [43]	Bayesian Sequential Network; multitask	Intermediate (joint PET/CT[y]+clinical)	Yes
Gunashekar et al [44]	U-Net variants (2D and 3D); all fusions compared	Early/intermediate/ late (head-to-head)	No
Takeda et al [45]	Custom CNN+SVM	Feature-level (per-modality ➞ SVM)	No

[a]Part: only partial code or model components were identified.
[b]LR: logistic regression.
[c]Yes: a public study-specific implementation was identified.
[d]MC: Monte Carlo.
[e]No: no public study-specific implementation was identified.
[f]FC: fully connected.
[g]SVM: support vector machine.
[h]KNN: k-nearest neighbors.
[i]LASSO: least absolute shrinkage and selection operator.
[j]SHAP: Shapley Additive Explanations.
[k]DTL: deep transfer learning.
[l]SE: squeeze-and-excitation.
[m]CNN: convolutional neural network.
[n]DL: deep learning.
[o]PET: positron emission tomography.
[p]MRI: magnetic resonance imaging.
[q]ViT: vision transformer.
[r]PAGNet: prior-attention guided network.
[s]CBAM: convolutional block attention module.
[t]PCa: prostate cancer.

[u]XGBoost: extreme gradient boosting.
[v]SPP: spatial pyramid pooling.
[w]CatBoost: categorical boosting.
[x]MMAI: multimodal AI.
[y]CT: computed tomography.

The evidence map also shows that multimodal AI in prostate cancer is better understood as a pathway of decision-specific tools than as a single diagnostic technology. In suspected localized disease, the relevant clinical decision is whether a biopsy can be avoided, deferred, or prioritized without missing csPCa. In that setting, large-scale biparametric magnetic resonance imaging (bpMRI) or mpMRI systems and MRI-negative triage models are most directly aligned with the clinical problem [26-29,48]. Saha et al [26] provided the strongest benchmark-scale evidence by comparing AI with radiologists in the PI-CAI setting, whereas Grabke et al [27] and Taguelmimt et al [28] provided evidence more representative of deployable bpMRI triage systems. Zheng et al [48] addressed a clinically important gap by focusing on patients with negative MRI, where biopsy decisions remain difficult despite reassuring imaging.

For grading and ISUP prediction, the evidence base is more technically heterogeneous. Several studies integrated PSMA PET/CT, mpMRI sequences, radiomics, or clinical variables to predict ISUP grade group or Gleason stratification [30-34,46,50,51]. These models are attractive because grade group influences active surveillance, surgery, radiotherapy, and systemic therapy decisions, but most studies remain retrospective and task-specific. PET/MRI or PET/CT plus mpMRI approaches may improve discrimination in selected high-risk patients; however, their incremental value must be weighed against cost, availability, and the fact that grade confirmation still depends on pathology. By contrast, MRI plus clinical data may be more scalable for prebiopsy or preoperative risk refinement, especially when the model outputs can be mapped to explicit downstream decisions.

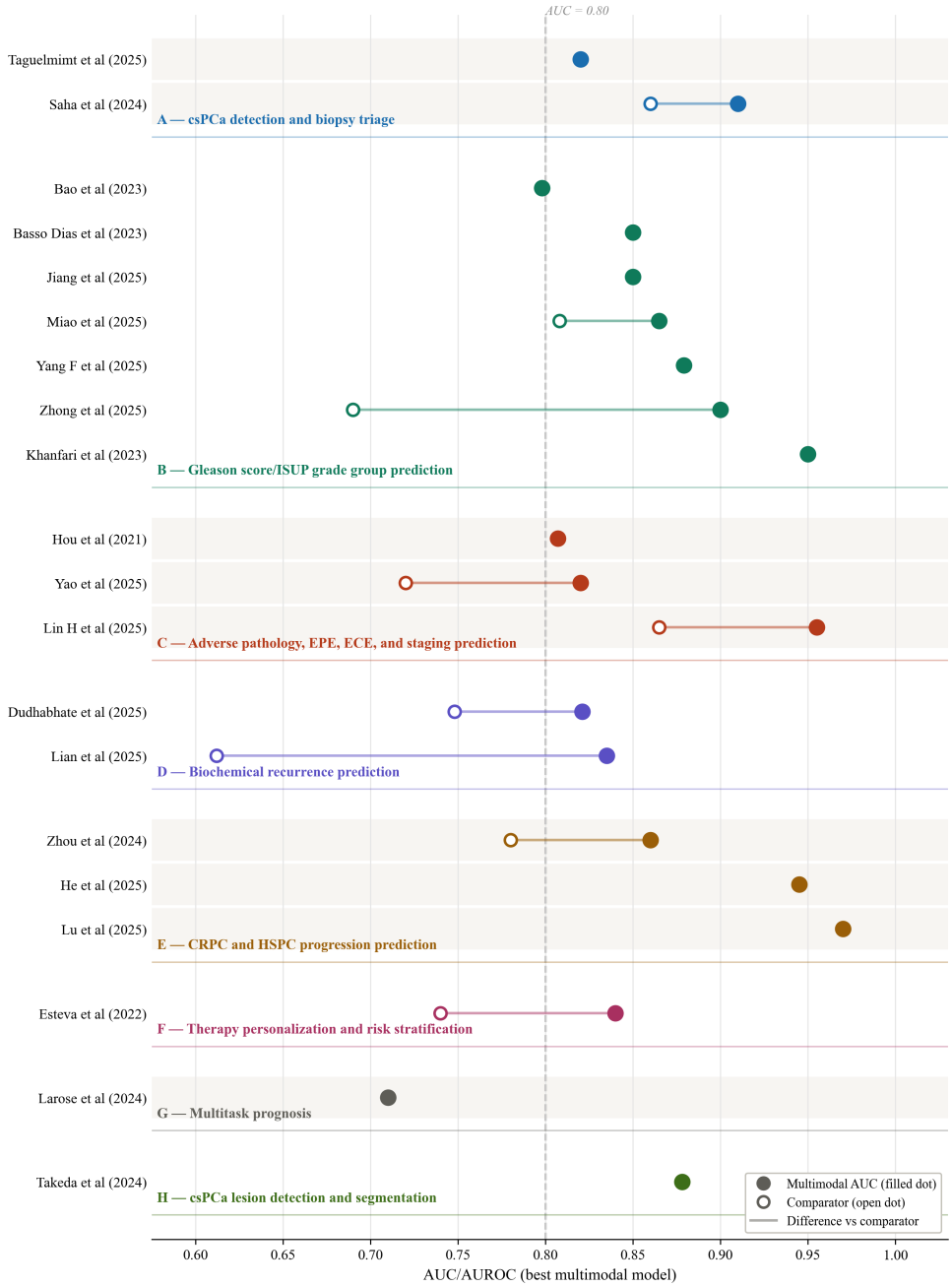
Adverse pathology and staging models formed a smaller but clinically coherent group. Lin et al [35], Yao et al [36], and Hou et al [47] focused on extraprostatic extension, seminal vesicle invasion, positive surgical margins, or broader adverse pathology outcomes. These end points are particularly relevant to surgical planning, nerve-sparing decisions, lymph-node dissection, and counseling. Biochemical recurrence and progression models targeted a later segment of the pathway, combining imaging, pathology, genomic, or clinical features to estimate recurrence or transition to resistant disease [37-41]. The clearest translational message from these studies is that multimodal AI may be most valuable when conventional clinical risk groups compress biologically diverse patients into broad categories.

The treatment-personalization studies were methodologically distinctive because they used randomized clinical trial resources and long follow-up, rather than small single-center imaging cohorts [42,49]. These studies suggest that digital histopathology combined with clinical trial data may support risk stratification across clinically meaningful end points. Larose et al [43] extended this idea to multitask prognosis across several survival and progression outcomes. Lesion detection and segmentation studies addressed a different problem: how modalities should be fused spatially when the model must localize disease rather than only assign a patient-level risk score [44,45]. Together, these categories reinforce that performance should not be compared across tasks without considering the decision, end point, comparator, and acceptable error profile.

The full-text audit identified additional adjacent studies that informed eligibility boundaries but were not part of the 26-study evidence synthesis. These included reports on multimodal clinical diagnostic systems, anatomically guided bpMRI detection, PI-RADS estimation, benign or non-csPCa end points, molecular mutation prediction, tumor-abnormal-protein-enhanced MRI radiomics, ultrasound radiomics, PSMA uptake classification, and bone metastasis prediction [18-22,52-57]. Their exclusion does not imply low technical quality. Rather, it reflects the narrower review question, which required extractable multimodal AI evidence directly aligned with csPCa detection, grading, risk stratification, recurrence, progression, or treatment personalization. Retaining these references in the manuscript is important because they document how the boundary between eligible multimodal prostate AI and adjacent AI literature was applied.

Performance was generally favorable to multimodal integration in studies that reported paired comparisons, but the strength of evidence varied substantially (Figure 2). The PI-CAI study provided the strongest detection evidence, reporting an area under the receiver operating characteristic curve (AUROC) of 0.91 for AI versus 0.86 for the median of 62 radiologists in a 400-case reader study nested within 10,207 MRI examinations from 9129 patients [26]. Trimodal PSMA PET/CT, mpMRI, and clinical modeling reached AUC 0.955 for adverse pathology prediction in one internally validated study [35]. Conditional convolution integrating clinical information improved AUROC from 0.69 to 0.90 without adding a new imaging modality [33]. For prognostic and treatment-selection tasks, large NRG Oncology analyses suggested that multimodal AI could reclassify risk beyond conventional clinical groupings [42,49].

Figure 2. Table 2Area under the curve (AUC) or area under the receiver operating characteristic curve (AUROC) of the best multimodal model across the 20 included studies that reported AUC as their primary performance metric, organized by the clinical task categories of Table 2 (A-H) and ordered within each category by ascending multimodal AUC (top to bottom) [26,28,30,31,33-41,43,45-47,49-51]. AUC: area under the curve; AUROC: area under the receiver operating characteristic curve; csPCa: clinically significant prostate cancer; CRPC: castration-resistant prostate cancer; ECE: extracapsular extension; EPE: extraprostatic extension; HSPC: hormone-sensitive prostate cancer; ISUP: International Society of Urological Pathology.



In Figure 2, filled dots denote the multimodal model AUC, and open dots denote the comparator reported by the same study (the best unimodal or clinical-only model; median radiologist performance for Saha et al [26]; and National Comprehensive Cancer Network risk groups for Esteva et al [49]); connecting lines indicate the difference between them, and the dashed vertical line marks AUC=0.80 as a reference threshold. Multimodal models exceeded the comparator in every paired comparison, with the largest gains in biochemical recurrence (Lian et al [37]) and Gleason or International Society of Urological Pathology grading (Zhong et al [33]). Five studies enrolled fewer than 200 patients (Basso Dias et al [51], Khanfari et al [50], Takeda et al [45], He et al [40], and Lu et al [41]); these carry an elevated risk of optimistic bias, and their results should be interpreted with particular caution. Six studies reporting primary outcomes in metrics other than AUC (Dice similarity coefficient, negative predictive value, accuracy, or hazard ratio) are not shown in Figure 2 and are detailed in Table 2. Reported values should be interpreted in relation to task, end point, cohort size, and validation design, because high retrospective discrimination does not necessarily indicate clinical readiness; Figure 2 is therefore descriptive and is not a cross-study ranking of model performance.

The magnitude and meaning of performance gains differed by task. In biopsy triage, the clinically relevant target is not maximum discrimination alone but safe reduction of unnecessary biopsy while preserving sensitivity for csPCa. In grade-group and adverse pathology prediction, multimodal AI may support preoperative counseling and surgical planning, but many studies used internal validation or small external sets. In biochemical recurrence and castration-resistant prostate cancer (CRPC) progression, reported gains are clinically attractive because they could influence surveillance intensity or systemic treatment planning, yet these end points require longer follow-up and stronger temporal validation. The trial-derived MMAI studies are distinctive because they link AI risk groups to clinically meaningful treatment and metastasis outcomes rather than only diagnostic classification.

Several patterns were consistent across the detailed evidence map. First, the strongest detection evidence came from large-scale, multireader evaluation rather than from the highest numerical AUC. PI-CAI is therefore important not only because of its AUROC, but also because it compared AI with 62 radiologists in a 400-case reader study and was developed and tested across a large multicenter bpMRI dataset [26]. Second, some of the highest AUCs were reported in smaller cohorts, including leave-one-out or internal validation designs, which should be interpreted as hypothesis-generating rather than clinically definitive [40,41]. Third, the studies with randomized controlled trial (RCT)–derived pathology inputs provided a different type of evidence by connecting AI-derived strata with distant metastasis and treatment-effect heterogeneity [42,49]. This distinction matters: diagnostic AI and prognostic AI require different validation end points, and they should not be judged by a single performance metric.

These findings should be interpreted cautiously. Six studies enrolled fewer than 200 patients, including some reporting AUCs above 0.94 [40,41,50]. External validation was present in only a minority of studies and often used small test sets. No included study prospectively evaluated AI in a clinical workflow, measured downstream biopsy reduction with patient outcomes, or quantified cost-effectiveness. Some form of public code or model component was identified for 7 studies [26,27,33,38,40,43,50], including 5 [27,33,40,43,50] with a public study-specific implementation; public datasets were used in only 4 studies [26,33,38,50], and uncertainty quantification was reported in only 2 [28,43], limiting reproducibility and clinical safety assessment. Calibration, decision curve analysis, and failure analysis were inconsistently reported, despite their importance for clinical decision support.

Validation strategies were heterogeneous and generally not strong enough for clinical deployment claims. Twelve studies [29,33,35,38,40,41,43-45,48,50,51] used internal validation only, 7 [26-28,32,39,42,49] reported multicenter retrospective validation, and 7 [30,31,34,36,37,46,47] included an external test cohort. External test sets were often small, and no included study conducted prospective validation in a live clinical workflow. Reproducibility was also limited: 5 [27,33,40,43,50] studies had a public study-specific implementation, 2 [26,38] had only partial public code or model components, and 4 [26,33,38,50] used public datasets. These findings indicate that the literature is technically active but remains methodologically immature for routine clinical use.

Model architectures reflected the same mixture of technical momentum and translational immaturity. Convolutional neural networks (CNNs) were the dominant backbone, with ResNet variants used frequently for classification and U-Net variants used for segmentation-based detection. Transformer-based or hybrid CNN-transformer architectures appeared in newer studies, as did conditional, contrastive, and Bayesian approaches. Classical machine learning classifiers, including extreme gradient boosting (XGBoost), categorical boosting (CatBoost), support vector machines, random forests, logistic regression, and least absolute shrinkage and selection operator (LASSO)–based models, often served as final-stage classifiers after imaging or pathology feature extraction. These choices show that multimodal prostate AI is not a single technical paradigm; rather, it spans deep feature learning, radiomics, clinical nomograms, ensemble learning, uncertainty-aware modeling, and trial-derived risk stratification.

The size and provenance of datasets strongly influenced interpretation of performance claims. Large cohorts such as PI-CAI and the NRG Oncology trial-derived studies provide stronger evidence because they combine scale, clinically meaningful end points, and broader validation. In contrast, several smaller studies reported very high AUC values despite enrolling fewer than 200 patients, sometimes with leave-one-out or internal cross-validation only. Such results are useful for hypothesis generation, but they should not be interpreted as evidence of clinical readiness. Across the corpus, few studies reported calibration, decision curve analysis, subgroup performance, or failure analysis, even though these measures are central for clinical decision support.

Discussion

Principal Findings

This scoping review identifies a field with strong technical momentum but incomplete translational evidence. Across diverse prostate cancer tasks, multimodal AI frequently outperformed unimodal baselines when paired comparisons were available. The most clinically important insight is that useful gains were not restricted to expensive or specialized multimodal acquisitions. Several studies showed that structured clinical variables, already available in routine care, can materially improve imaging-based AI. For urologists, this suggests that the most deployable near-term systems may be those combining biparametric or mpMRI with PSA, age, prostate volume, PI-RADS, and prior biopsy information, rather than systems requiring PET/CT or molecular assays for every patient.

The most mature application appears to be biopsy triage and csPCa detection. PI-CAI demonstrated large-scale AI performance comparable with experienced radiologists, while

bpMRI-based triage studies suggest potential reductions in workload and unnecessary biopsy [26-28]. However, even these promising systems require prospective testing before changing practice. A clinically useful triage model must preserve sensitivity for csPCa, reduce unnecessary biopsy, maintain equity across patient subgroups, and function reliably across scanners, centers, and acquisition protocols. Retrospective AUROC alone cannot establish these outcomes.

Grading, staging, recurrence prediction, and treatment personalization are less mature but clinically compelling. Multimodal models integrating MRI, PET/CT, histopathology, and clinical variables reported promising discrimination for ISUP grade group, adverse pathology, biochemical recurrence, and trial-derived risk stratification [31,35,37,42,49]. Yet these tasks often used smaller datasets, single-center designs, and heterogeneous end points. For clinical adoption, future studies need external validation in diverse populations, calibration analyses, decision curve analysis, and direct comparison against current clinical nomograms or guideline-based pathways.

A clinically important finding is that multimodal does not always mean technologically complex. The largest practical opportunity may be integration of existing clinical variables with routinely acquired MRI. Such models could be implemented earlier than PET/MRI or multiomic systems because they depend on data already present in urology, radiology, and pathology workflows. Conversely, PET/CT and molecular modalities may be most valuable for selected high-risk contexts, such as adverse pathology prediction, metastatic-risk stratification, or treatment personalization, where the clinical consequence justifies higher acquisition cost. Future studies should therefore report not only model accuracy but also the marginal value of each modality relative to its availability, cost, and clinical burden.

The 2 multisequence MRI-only studies reported strong grading performance, but both remained retrospective and lacked independent external validation, limiting direct conclusions about clinical translation [32,50]. By contrast, cross-domain multisource studies showed performance gains from integrating clinical, pathological, PET, or molecular information across a broader range of clinical tasks. Direct quantitative comparison between these groups was not appropriate because of the small MRI-only subgroup and substantial heterogeneity in tasks, metrics, and validation designs. Their translational barriers also differ: MRI-only models are primarily affected by sequence availability, scanner and protocol variation, whereas cross-domain models additionally face missing data, data integration, infrastructure, cost, and interoperability challenges.

Barriers to Clinical Translation

Several recurring barriers explain the gap between benchmark performance and clinical readiness. First, all included studies were retrospective. Prospective trials are needed to test whether multimodal AI changes decisions, reduces unnecessary biopsy, improves csPCa detection, affects treatment selection, or improves patient-reported outcomes. Second, geographic concentration was substantial, with nearly half of studies originating from China and limited representation of high-burden groups, including African American men. Models trained in narrow populations may underperform in settings with different disease prevalence, genetics, screening practices, scanners, or treatment pathways. Third, missing-modality robustness remains underdeveloped. In real practice, PET/CT may be absent, pathology slides may not be digitized, clinical variables may be incomplete, and image quality may vary. Models requiring every input at inference will be difficult to deploy safely.

Generalizability is especially important in prostate cancer because screening intensity, MRI acquisition, PI-RADS expertise, biopsy practice, genetic risk, and treatment thresholds differ substantially across health systems. A model trained in a single center may learn local scanner characteristics, reporting habits, referral patterns, or case mix rather than generalizable tumor biology. The external performance decline reported in some studies illustrates this risk, and the concentration of evidence in a limited number of countries makes equity a central translational issue rather than a secondary concern. Future validation should therefore include subgroup performance, calibration, and failure analyses across ancestry, age, prostate volume, PSA range, scanner vendor, field strength, and institution type.

Fourth, reproducibility is weak. Public code and public data were uncommon, and reporting of preprocessing, hyperparameters, calibration, uncertainty, and failure modes was inconsistent. Adoption of Transparent Reporting of Multivariable Prediction Models with AI (TRIPOD+AI Checklist for AI in Medical Imaging (CLAIM), Standards for Reporting Diagnostic Accuracy-AI (STARD-AI), and PROBAST+AI would improve interpretability, replication, and comparison across studies. Fifth, the economic and regulatory dimensions of multimodal AI remain underexplored. A model that improves AUC but requires costly imaging or complex data integration may not be practical in lower-resource settings. Conversely, a model using routine clinical variables and bpMRI could be highly valuable if it prospectively reduces unnecessary biopsy without increasing missed csPCa.

Methodological Gaps

Missing-modality robustness remains one of the most important unresolved barriers. Real clinical workflows rarely contain a complete set of ideal inputs. PET/CT may not be available, pathology slides may not be digitized, genomic or proteomic assays may be absent, and structured clinical variables may be incomplete or inconsistently recorded. Models that require all modalities at inference time are therefore fragile. Late-fusion and selective-prediction approaches may be more deployable because they can support partial-input operation or route uncertain cases to expert review, but this capability was rarely tested systematically.

Scanner and domain shift also require more explicit treatment. MRI acquisition protocols, field strength, vendor, reconstruction pipeline, radiologist reporting habits, biopsy

technique, and disease prevalence differ across institutions. Radiomics features are particularly vulnerable to acquisition and preprocessing variation. External performance declines reported in several studies suggest that models may learn local technical or workflow signatures rather than generalizable tumor biology. Future studies should report scanner/vendor distributions, harmonization procedures, subgroup calibration, and performance under realistic protocol variation.

Unstructured clinical reports create a related data-engineering barrier for models that require staging variables. In a feasibility study of 80 Turkish-language PSMA PET/CT reports, Gemini 2.5 Pro (Google) and ChatGPT-4o (OpenAI) achieved 93.8% and 91.3% accuracy, respectively, for the composite clinical-stage and ChemoHormonal Therapy Versus Androgen Ablation Randomized Trial for Extensive Disease in Prostate Cancer (CHAARTED) disease-volume task when prompted with an embedded knowledge base and few-shot examples [58]. These results show that large language models may support conversion of narrative reports into structured inputs, but the single-language, single-center design, prompt dependence, and absence of prospective workflow testing preclude assumptions about transportability or autonomous staging.

Interpretability was commonly presented but rarely validated as a clinical tool. Heat maps, attention visualizations, Shapley Additive Explanations (SHAP) values, and feature-importance plots may help developers understand a model, but they do not automatically establish clinician trust or patient safety. For deployment, interpretability should be tied to explicit decisions: when to accept a prediction, when to request expert review, which modality drove the output, and whether the explanation identifies plausible anatomy or pathology. Uncertainty quantification is especially important because it can convert model limitations into workflow rules rather than leaving clinicians with a single deterministic score.

Economic and regulatory evidence was almost absent. A multimodal model can improve AUC yet remain impractical if it requires expensive imaging, specialized molecular assays, complex data integration, or infrastructure unavailable outside academic centers. Conversely, a model that combines routine clinical data with bpMRI may be valuable even with more modest discrimination if it safely reduces unnecessary biopsies. Future studies should include health economic analysis, implementation burden, data-governance requirements, and a clear regulatory pathway for models that combine radiology, pathology, molecular, and clinical data.

Patient-centered outcomes and equity were also underdeveloped. No included study prospectively measured patient anxiety, quality of life, shared decision-making, biopsy complications avoided, treatment decisional regret, or downstream oncological outcomes after AI-guided care. The limited representation of high-burden populations, including African American men, raises safety and fairness concerns. Prospective evaluation should therefore include subgroup performance, calibration across demographic and clinical strata, and patient-relevant outcomes rather than relying exclusively on benchmark discrimination.

Clinical Translation and Implementation Priorities

The most plausible near-term application is AI-assisted biopsy triage using biparametric or mpMRI combined with routinely available clinical variables. This use case depends on data already present in most diagnostic pathways and could reduce unnecessary biopsies if sensitivity for csPCa is preserved. PET/CT, digital pathology, and molecular data may be most valuable in higher-risk contexts, including adverse pathology prediction, metastatic-risk stratification, recurrence forecasting, or treatment personalization, where the clinical consequence may justify additional acquisition cost and infrastructure. Any implementation should include uncertainty thresholds, escalation rules for unreliable predictions, local calibration monitoring, and subgroup performance assessment.

Implementation should also be matched to the clinical decision being supported. A model intended for biopsy triage must prioritize sensitivity, negative predictive value, and safe escalation of uncertain cases. A grading or staging model must be compared with current radiological assessment, biopsy pathology, and clinical nomograms rather than only with a unimodal AI baseline. A prognostic or treatment-selection model must demonstrate value against guideline risk groups and should ideally show that it changes management in a way that improves patient-centered outcomes. This decision-centered framing is especially important for multimodal AI because each additional modality adds cost, operational complexity, missing-data risk, and potential inequity in access.

The agenda in Table 4 is intended to shift the field from benchmark optimization toward clinical evaluation. For JMIR readers, the key question is not whether multimodal AI can produce high retrospective AUCs; several studies show that it can. The unresolved question is where these tools should sit in the patient pathway. In suspected localized disease, the most plausible role is selective triage and second-reader support. In intermediate- and high-risk disease, AI may refine staging or recurrence risk. In trial-derived pathology models, AI may identify biologically meaningful risk strata for treatment intensification. Each use case needs its own validation standard because the acceptable balance between sensitivity, specificity, uncertainty, and cost differs across the pathway.

Table 4. Translational research agenda for multimodal AI in prostate cancer.

Priority	Why it matters	Minimum evidence needed
Prospective clinical testing	Retrospective discrimination does not show whether AI improves biopsy, staging, treatment, or follow-up decisions.	Workflow-embedded studies measuring csPCa[a] detection, biopsy avoidance, decision change, safety, and patient outcomes.
Diverse external validation	Models trained in narrow populations may fail under different prevalence, biology, scanners, and practice patterns.	Multicenter cohorts with prespecified subgroup analyses by ancestry, age, scanner vendor, and clinical setting.
Missing-modality robustness	Real clinics often lack PET[b]/CT[c], digitized pathology, complete variables, or high-quality images.	Modality-dropout experiments, partial-input testing, and calibrated uncertainty reporting.
Transparent reporting	Poor reporting prevents replication and weakens trust.	Reporting aligned with TRIPOD+AI[d], CLAIM[e], or STARD-AI[f], alongside design-appropriate risk-of-bias assessment (eg, PROBAST+AI[g]), with public code and preprocessing details.
Implementation value	Expensive multimodal systems may not be scalable even when accurate.	Decision curve analysis, calibration, health economic evaluation, and comparison against existing nomograms.

[a]csPCa: clinically significant prostate cancer.
[b]PET: magnetic resonance imaging.
[c]CT: computed tomography.
[d]TRIPOD+AI: Transparent Reporting of Multivariable Prediction Models with AI.
[e]CLAIM: Checklist for AI in Medical Imaging.
[f]STARD-AI: Standards for Reporting Diagnostic Accuracy-AI.
[g]PROBAST+AI: Prediction Model Risk Of Bias Assessment Tool+AI.

Regulatory clearance should be treated as the beginning rather than the end of evaluation. Postdeployment governance should prospectively define monitored clinical outcomes, subgroup and site-specific calibration, input-data quality and drift, changes in disease prevalence and workflow, alert thresholds, investigation and rollback procedures, and adverse-event reporting. Updating or retraining can alter the benefit-risk profile; adaptive systems therefore require version control, revalidation on representative data, transparent communication to users, and a predetermined change control plan where applicable [59]. For multimodal systems, surveillance should also track modality availability and missingness because deterioration in one input stream can degrade the joint model without an obvious software failure.

Strength of Evidence and Generalizability

The overall strength of evidence is uneven. The review includes a small number of high-value evidence anchors, including large benchmark or trial-derived resources, but most studies remain early-stage model-development reports. PI-CAI provides scale, multireader comparison, and clinically relevant csPCa detection end points [26]. The NRG Oncology trial-derived analyses provide long follow-up, trial-quality outcome ascertainment, and comparison against established risk-stratification systems [42,49]. These studies are therefore more informative for translation than small retrospective cohorts, even when smaller studies report higher AUC values.

Generalizability is limited by population, site, and technology concentration. Nearly half of the included studies originated from China, whereas several populations with high prostate cancer burdens were poorly represented or not explicitly analyzed. Few studies reported ancestry, socioeconomic variables, or subgroup calibration, even though prostate cancer incidence, mortality, screening exposure, and access to MRI or molecular tests differ across demographic groups and health systems. A model trained in a tertiary referral center with high MRI quality, specialized radiology expertise, and enriched disease prevalence may not behave similarly in community practice, opportunistic screening settings, or lower-resource systems.

End point heterogeneity also restricts synthesis. The included models predicted csPCa, ISUP grade, adverse pathology, extraprostatic extension, recurrence, progression, multitask survival outcomes, and treatment benefit. Because populations, outcome definitions, reference standards, validation designs, sample sizes, thresholds, and reported metrics differed substantially, AUCs and other performance measures cannot be compared directly across studies or used to rank clinical tasks, modalities, or fusion architectures. Figure 2 is descriptive; only within-study paired multimodal-unimodal comparisons provide an interpretable estimate of incremental performance, and even these do not establish transportability or clinical utility. The end points also occur at different points in the care pathway and imply different clinical actions. A false negative in biopsy triage may delay diagnosis, whereas a false positive in recurrence prediction may intensify surveillance or treatment. For this reason, the review emphasizes structured evidence mapping rather than meta-analysis. Future studies should align model outputs with prespecified decisions, such as avoiding biopsy, selecting nerve-sparing surgery, escalating staging imaging, intensifying therapy, or changing follow-up intervals.

The available evidence also shows a tension between technical sophistication and clinical usability. Architectures based on cross-attention, contrastive learning, conditional convolution, and Bayesian modeling may capture richer cross-modal relationships, but they are not necessarily easier to implement or validate. Conversely, models that combine routinely collected clinical variables with bpMRI or mpMRI may offer a more realistic path to adoption even when their performance gains are more modest. For health systems, the relevant question is not whether every possible modality

improves discrimination, but whether the marginal gain changes management enough to justify acquisition burden, integration cost, and inequity risk.

Comparison With Prior Work

Relative to the broad narrative overview of prostate cancer AI by Baydoun et al [15] and the multimodality-imaging survey by Ayyad et al [16], this review uses a scoping framework to focus on explicit fusion across imaging, histopathology, molecular data, and structured clinical variables and to map validation, reproducibility, and translational readiness across the clinical pathway. The evidence map includes PET/CT-based frameworks, digital pathology, RCT-derived risk-stratification models, biochemical recurrence, castration-resistant progression, and treatment-personalization studies. The emergence of conditional convolution, cross-attention Transformer architectures, contrastive cross-modal learning, and Bayesian sequential modeling shows that the field is moving beyond simple feature concatenation. Nevertheless, the same structural weaknesses remain: small retrospective cohorts, geographic concentration, incomplete external validation, limited calibration reporting, and scarce reproducibility.

This review also highlights a design principle for future multimodal systems: the added modality should be justified by an added clinical decision. Adding PSMA PET/CT or molecular data may be reasonable when the task is adverse pathology prediction, metastatic-risk stratification, or treatment personalization, but it may be excessive for first-line biopsy triage if similar performance can be achieved with bpMRI and routine clinical variables. Similarly, pathology-based AI may be more appropriate for postbiopsy risk refinement than for initial diagnostic triage. Framing multimodality around the clinical decision would make studies easier to interpret and would help identify which systems are scalable across community, academic, and resource-constrained settings.

Limitations

This review has limitations. The search used 5 databases and English-language restrictions, so relevant studies using nonstandard terminology or other languages may have been missed. The historical automated filter was fully reproducible and retained all final included studies that could be traced reliably to its input. Two-reviewer assessment of 200 randomly sampled exclusions found no eligible studies. However, the filter evaluated titles only, was not prospectively validated, and the remaining 584 excluded records were not manually reassessed. Eligible records using terminology not captured by the literal title substrings may therefore have been excluded, introducing potential selection bias. Conference proceedings were included when peer-reviewed, but their reporting depth may be lower than full journal articles. Consistent with scoping review methodology, no meta-analysis or formal risk-of-bias exclusion was performed. The maturity indicators were narratively mapped rather than independently rated with a validated tool and should not be interpreted as formal risk-of-bias judgments. Finally, because multimodal AI definitions vary, the decision to include studies fusing multiple MRI sequences may be debated; this choice was made transparently and should be considered when interpreting the evidence map.

Publication bias and selective outcome reporting may also make the evidence appear more favorable. Studies with positive performance are more likely to be published, while investigators may preferentially report the best architecture, split, end point, threshold, or metric. Meta-research in oncology machine learning prediction studies found method-results inconsistencies related to additional or selectively reported analyses and frequent overinterpretation [60]. We did not formally assess reporting bias; the evidence map should therefore be interpreted as a map of published claims rather than an unbiased estimate of underlying model performance.

Future Directions

Future research should prioritize prospective clinical trials, diverse external validation, missing-modality robustness, scanner harmonization, open benchmarking, uncertainty quantification, and patient-centered outcomes. The highest-value studies will not simply report better AUROC. They will show that multimodal AI works across centers, improves clinically meaningful decisions, remains safe when uncertain, and delivers value to patients and health systems. In practical terms, the next generation of studies should predefine the clinical decision to be changed, the comparator pathway, the acceptable false-negative rate, the action taken when predictions are uncertain, and the patient-centered outcomes that justify deployment.

Near-term research should focus on prospective workflow studies, large-scale external validation, and missing-modality experiments. Prospective studies should predefine the clinical decision to be changed, the comparator pathway, the acceptable false-negative rate, the action triggered by an uncertain prediction, and the patient-centered outcomes that justify deployment. External validation should be deliberately diverse rather than opportunistic, including multiple vendors, field strengths, referral patterns, ancestry groups, age ranges, prostate volumes, PSA ranges, and clinical settings. Missing-modality testing should become a routine part of reporting because partial data are the rule rather than the exception in real practice.

Medium-term research should develop harmonized, open, and federated evaluation infrastructure. Shared benchmarks such as PI-CAI and PROSTATEx have already shaped detection and grading research, but comparable public resources are limited for recurrence, treatment response, pathology, molecular, and multiomic end points. Federated learning may allow institutions to train and validate models across diverse cohorts without transferring sensitive patient data, which is particularly relevant for rare molecular inputs and high-risk subgroups. Foundation models pretrained on large radiology, pathology, text, and structured clinical datasets may also improve generalization, but they will require careful evaluation for site-specific shortcuts and inequitable performance.

Reporting standards should evolve from optional aspiration to publication norm. TRIPOD+AI, CLAIM, and STARD-AI provide relevant anchors for model-development and diagnostic-accuracy reporting, while PROBAST+AI supports structured assessment of quality, risk of bias, and applicability for prediction models [24]. For multimodal AI, reports should also specify modality availability, missing-input handling, fusion strategy, calibration, uncertainty, preprocessing, scanner distribution, code availability, and the marginal contribution of each modality. This would make future systematic reviews and meta-analyses more informative and would allow clinicians to judge whether additional modalities improve decisions enough to justify their burden.

Uncertainty handling should also become routine rather than optional. A deterministic score may be adequate for retrospective ranking, but deployment requires knowing when the model is operating outside its reliable domain. Selective prediction, uncertainty thresholds, and escalation rules could allow AI systems to support rather than replace expert review. This is particularly relevant when inputs are incomplete, images are degraded, or patients differ from the development cohort. Without such safeguards, multimodal AI risks appearing accurate in aggregate while failing silently in precisely the cases where clinical support is most needed.

Conclusions

Multimodal AI for prostate cancer has advanced rapidly and often improves performance over unimodal models in retrospective paired comparisons. The most practical near-term opportunity is likely integration of MRI with routinely available clinical variables for biopsy triage and risk stratification. However, no included study prospectively tested multimodal AI in routine clinical care, and evidence remains limited by small cohorts, geographic concentration, incomplete reproducibility, and inconsistent reporting of calibration and uncertainty. The field should now move from retrospective benchmarking toward transparent, prospective, diverse, and patient-centered evaluation before multimodal AI is used to guide routine prostate cancer decisions.

Generative AI Use

ChatGPT (OpenAI, GPT-5.6 Sol) was used to assist with journal formatting, language editing, and preparation of submission materials. The tool was not used to generate scientific content, interpret data, or draft conclusions. All text, data, references, interpretations, and final wording were reviewed, verified, and approved by the authors, who take full responsibility for the integrity of the work.

Funding

This work was partially supported by the Asociación Española Contra el Cáncer (AECC) through the Ayuda Predoctoral AECC Valencia 2025, grant ID PRDVA258654ALZA. The funder had no role in study design, data collection, data analysis, data interpretation, manuscript preparation, or the decision to submit the article for publication.

Data Availability

No new patient-level data were generated for this scoping review. The full-text assessment audit table, reproducibility notes, screening scripts, and figure-generation notebooks are available in an open repository [61].

Authors' Contributions

JAA-G, MI-V, FG-G, RT-S, and JMG-G conceived and designed the review and contributed to funding acquisition. JAA-G, FOO, JDRO, and JDRT acquired the data; JAA-G and FOO analyzed and interpreted the data and drafted the manuscript. JDRO and JDRT contributed to statistical analysis and technical support. MP-T, ANB, and CRT provided administrative, technical, or material support. FOO, JDRO, JDRT, MP-T, ANB, CRT, FG-G, RT-S, JMG-G, and MI-V critically revised the manuscript for important intellectual content. FG-G, RT-S, JMG-G, and MI-V supervised the work. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Extended search and audit.
[DOCX File (Microsoft Word File), 377 KB-Multimedia Appendix 1]

Checklist 1

PRISMA-ScR checklist.
[PDF File (Adobe File), 5 KB-Checklist 1]

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263. [doi: 10.3322/caac.21834] [Medline: 38572751]

2. Global cancer observatory: prostate. International Agency for Research on Cancer; World Health Organization; 2024. URL: https://gco.iarc.who.int/media/globocan/factsheets/cancers/27-prostate-fact-sheet.pdf [Accessed 2026-10-28]
3. Global cancer observatory: cancer tomorrow homepage. International Agency for Research on Cancer. 2024. URL: https://gco.iarc.who.int/tomorrow/ [Accessed 2026-10-28]
4. Cornford P, van den Bergh RCN, Briers E, et al. EAU-EANM-ESTRO-ESUR-ISUP-SIOG guidelines on prostate cancer-2024 update. Part I: screening, diagnosis, and local treatment with curative intent. Eur Urol. Aug 2024;86(2):148-163. [doi: 10.1016/j.eururo.2024.03.027] [Medline: 38614820]
5. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med. May 10, 2018;378(19):1767-1777. [doi: 10.1056/NEJMoa1801993] [Medline: 29552975]
6. Ahdoot M, Wilbur AR, Reese SE, et al. MRI-targeted, systematic, and combined biopsy for prostate cancer diagnosis. N Engl J Med. Mar 5, 2020;382(10):917-928. [doi: 10.1056/NEJMoa1910038] [Medline: 32130814]
7. Stabile A, Giganti F, Rosenkrantz AB, et al. Multiparametric MRI for prostate cancer diagnosis: current status and future directions. Nat Rev Urol. Jan 2020;17(1):41-61. [doi: 10.1038/s41585-019-0212-4] [Medline: 31316185]
8. Merriel SWD, Pocock L, Gilbert E, et al. Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (PSA) for the detection of prostate cancer in symptomatic patients. BMC Med. Feb 7, 2022;20(1):54. [doi: 10.1186/s12916-021-02230-y] [Medline: 35125113]
9. Yang C, Li B, Luan Y, et al. Deep learning model for the detection of prostate cancer and classification of clinically significant disease using multiparametric MRI in comparison to PI-RADs score. Urol Oncol. May 2024;42(5):158. [doi: 10.1016/j.urolonc.2024.01.021] [Medline: 38388243]
10. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med. Jan 2019;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
11. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nat Rev Clin Oncol. Feb 2022;19(2):132-146. [doi: 10.1038/s41571-021-00560-7] [Medline: 34663898]
12. Hofman MS, Lawrentschuk N, Francis RJ, et al. Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. The Lancet. Apr 2020;395(10231):1208-1216. [doi: 10.1016/S0140-6736(20)30314-7] [Medline: 32649919]
13. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. Sep 2022;28(9):1773-1784. [doi: 10.1038/s41591-022-01981-2] [Medline: 36109635]
14. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med. 2020;3(1):136. [doi: 10.1038/s41746-020-00341-z] [Medline: 33083571]
15. Baydoun A, Jia AY, Zaorsky NG, et al. Artificial intelligence applications in prostate cancer. Prostate Cancer Prostatic Dis. Mar 2024;27(1):37-45. [doi: 10.1038/s41391-023-00684-0] [Medline: 37296271]
16. Ayyad SM, Abdel-Hamid NB, Ali HA, Labib LM. Multimodality imaging in prostate cancer diagnosis using artificial intelligence: basic concepts and current state-of-the-art. Multimed Tools Appl. 2025;84(34):42649-42678. [doi: 10.1007/s11042-025-20786-2]
17. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. Oct 2, 2018;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Chen T, Hu W, Zhang Y, et al. A multimodal deep learning nomogram for the identification of clinically significant prostate cancer in patients with gray-zone PSA levels: comparison with clinical and radiomics models. Acad Radiol. Feb 2025;32(2):864-876. [doi: 10.1016/j.acra.2024.10.009] [Medline: 39496535]
19. Esteban LM, Borque-Fernando Á, Escorihuela ME, et al. Integrating radiological and clinical data for clinically significant prostate cancer detection with machine learning techniques. Sci Rep. Feb 4, 2025;15(1):4261. [doi: 10.1038/s41598-025-88297-6] [Medline: 39905119]
20. Yildirim K, Yildirim M, Eryesil H, et al. Deep learning-based PI-RADS score estimation to detect prostate cancer using multiparametric magnetic resonance imaging. Comput Electr Eng. Sep 2022;102:108275. [doi: 10.1016/j.compeleceng.2022.108275]
21. Ma Q, Chen B, Seifert R, et al. PSMA PET/CT based multimodal deep learning model for accurate prediction of pelvic lymph-node metastases in prostate cancer patients identified as candidates for extended pelvic lymph node dissection by preoperative nomograms. Eur J Nucl Med Mol Imaging. May 2025;52(6):2063-2075. [doi: 10.1007/s00259-024-07065-2] [Medline: 39865180]
22. Zhang YF, Zhou C, Guo S, et al. Deep learning algorithm-based multimodal MRI radiomics and pathomics data improve prediction of bone metastases in primary prostate cancer. J Cancer Res Clin Oncol. Feb 5, 2024;150(2):78. [doi: 10.1007/s00432-023-05574-5] [Medline: 38316655]

23. Engelage L, Behnel N, Bashkanov O, et al. Artificial intelligence in urology: combining imaging and clinical data for improved prostate diagnostics. Presented at: Advanced Photonics in Urology 2025; Jan 25-31, 2025; San Francisco, United States. 2025.[doi: 10.1117/12.3042008]
24. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. Mar 24, 2025;388:e082505. [doi: 10.1136/bmj-2024-082505] [Medline: 40127903]
25. Guni A, Sounderajah V, Whiting P, Bossuyt P, Darzi A, Ashrafian H. Revised tool for the quality assessment of diagnostic accuracy studies using AI (QUADAS-AI): protocol for a qualitative study. JMIR Res Protoc. Sep 18, 2024;13:e58202. [doi: 10.2196/58202] [Medline: 39293047]
26. 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. Jul 2024;25(7):879-887. [doi: 10.1016/S1470-2045(24)00220-1] [Medline: 38876123]
27. Grabke EP, Heming CAM, Hadari A, et al. Using AI to triage patients without clinically significant prostate cancer using biparametric MRI and PSA. Abdom Radiol (NY). Dec 2025;50(12):5924-5933. [doi: 10.1007/s00261-025-05019-2] [Medline: 40445307]
28. Taguelmimt K, Andrade-Miranda G, Harb H, et al. Towards more reliable prostate cancer detection: incorporating clinical data and uncertainty in MRI deep learning. Comput Biol Med. Aug 2025;194:110440. [doi: 10.1016/j.compbiomed.2025.110440] [Medline: 40505283]
29. Sherafatmandjoo H, Safaei AA, Ghaderi F, Allameh F. Prostate cancer diagnosis based on multi-parametric MRI, clinical and pathological factors using deep learning. Sci Rep. Jun 28, 2024;14(1):14951. [doi: 10.1038/s41598-024-65354-0] [Medline: 38942817]
30. Jiang J, Fan Z, Jiang S, et al. Interpretable multimodal deep learning model for predicting post-surgical international society of urological pathology grade in primary prostate cancer. Eur J Nucl Med Mol Imaging. Aug 2025;52(10):3730-3743. [doi: 10.1007/s00259-025-07248-5] [Medline: 40183953]
31. Miao C, Yao F, Fang J, et al. Exploring the role of multimodal [^{18}F]F-PSMA-1007 PET/CT and multiparametric MRI data in predicting ISUP grading of primary prostate cancer. Eur J Nucl Med Mol Imaging. May 2025;52(6):2087-2095. [doi: 10.1007/s00259-025-07099-0] [Medline: 39871017]
32. Liu F, Zhao Y, Song J, et al. A hybrid classification model with radiomics and CNN for high and low grading of prostate cancer Gleason score on mp-MRI. Displays. Jul 2024;83:102703. [doi: 10.1016/j.displa.2024.102703]
33. Zhong J, Chen F, Chen L, Shung D, Onofrey JA. Conditional convolution of clinical data embeddings for multimodal prostate cancer classification. Presented at: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI); Apr 14-17, 2025:1-5; Houston, TX, USA. [doi: 10.1109/ISBI60581.2025.10981307]
34. Yang F, Wang C, Shen J, et al. End-to-end [^{18}F]PSMA-1007 PET/CT radiomics-based pipeline for predicting ISUP grade group in prostate cancer. Abdom Radiol (NY). Apr 2025;50(4):1641-1652. [doi: 10.1007/s00261-024-04601-4] [Medline: 39349643]
35. Lin H, Yao F, Yi X, et al. Prediction of adverse pathology in prostate cancer using a multimodal deep learning approach based on [^{18}F]PSMA-1007 PET/CT and multiparametric MRI. Eur J Nucl Med Mol Imaging. Jul 2025;52(8):2814-2825. [doi: 10.1007/s00259-025-07134-0] [Medline: 39969539]
36. Yao F, Lin H, Xue YN, et al. Multimodal imaging deep learning model for predicting extraprostatic extension in prostate cancer using MpMRI and 18 F-PSMA-PET/CT. Cancer Imaging. Aug 19, 2025;25(1):103. [doi: 10.1186/s40644-025-00927-4] [Medline: 40830810]
37. Lian T, Zhou M, Shao Y, Chen X, Zhao Y, Feng Q. MRMS-CNNFormer: a novel framework for predicting the biochemical recurrence of prostate cancer on multi-sequence MRI. Bioengineering (Basel). May 16, 2025;12(5):538. [doi: 10.3390/bioengineering12050538] [Medline: 40428157]
38. Dudhabhate S, Chattopadhyay N, Parekh A, Singhal N. Multi-modal integration for predicting biochemical recurrence in prostate cancer. Presented at: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI); Apr 14-17, 2025:1-4; Houston, TX, USA. [doi: 10.1109/ISBI60581.2025.10980823]
39. Zhou C, Zhang YF, Guo S, et al. Multimodal data integration for predicting progression risk in castration-resistant prostate cancer using deep learning: a multicenter retrospective study. Front Oncol. 2024;14:1287995. [doi: 10.3389/fonc.2024.1287995] [Medline: 38549937]
40. He D, Zhuang H, Ma Y, et al. A multiparametric MRI and baseline-clinical-feature-based dense multimodal fusion artificial intelligence (MFAI) model to predict castration-resistant prostate cancer progression. Cancers (Basel). May 3, 2025;17(9):1556. [doi: 10.3390/cancers17091556] [Medline: 40361482]
41. Lu X, Pan C, Yao L, et al. Integrating multimodal data to predict the progression of hormone-sensitive prostate cancer. Clin Proteomics. May 29, 2025;22(1):21. [doi: 10.1186/s12014-025-09543-7] [Medline: 40442579]

42. Tward JD, Huang HC, Esteva A, et al. Prostate cancer risk stratification in NRG oncology phase III randomized trials using multimodal deep learning with digital histopathology. JCO Precis Oncol. Oct 2024;8(8):e2400145. [doi: 10.1200/PO.24.00145] [Medline: 39447096]
43. Larose M, Archambault L, Touma N, et al. Multi-task Bayesian model combining FDG-PET/CT imaging and clinical data for interpretable high-grade prostate cancer prognosis. Sci Rep. Nov 6, 2024;14(1):26928. [doi: 10.1038/s41598-024-77498-0] [Medline: 39505979]
44. Gunashekar DD, Bielak L, Oerther B, et al. Comparison of data fusion strategies for automated prostate lesion detection using mpMRI correlated with whole mount histology. Radiat Oncol. Jul 29, 2024;19(1):96. [doi: 10.1186/s13014-024-02471-0] [Medline: 39080735]
45. Takeda H, Akatsuka J, Kiriyama T, et al. Clinically significant prostate cancer prediction using multimodal deep learning with prostate-specific antigen restriction. Curr Oncol. Nov 15, 2024;31(11):7180-7189. [doi: 10.3390/curroncol31110530] [Medline: 39590160]
46. Bao J, Hou Y, Qin L, et al. High-throughput precision MRI assessment with integrated stack-ensemble deep learning can enhance the preoperative prediction of prostate cancer Gleason grade. Br J Cancer. Mar 2023;128(7):1267-1277. [doi: 10.1038/s41416-022-02134-5] [Medline: 36646808]
47. Hou Y, Zhang YH, Bao J, et al. Artificial intelligence is a promising prospect for the detection of prostate cancer extracapsular extension with mpMRI: a two-center comparative study. Eur J Nucl Med Mol Imaging. Nov 2021;48(12):3805-3816. [doi: 10.1007/s00259-021-05381-5] [Medline: 34018011]
48. Zheng H, Miao Q, Raman SS, Scalzo F, Sung K. Integrative radiomics models to predict biopsy results for negative prostate MRI. Presented at: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI); Apr 13-16, 2021:877-881; Nice, France. [doi: 10.1109/ISBI48211.2021.9433879]
49. Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. NPJ Digit Med. Jun 8, 2022;5(1):71. [doi: 10.1038/s41746-022-00613-w] [Medline: 35676445]
50. Khanfari H, Mehranfar S, Cheki M, et al. Exploring the efficacy of multi-flavored feature extraction with radiomics and deep features for prostate cancer grading on mpMRI. BMC Med Imaging. Nov 22, 2023;23(1):195. [doi: 10.1186/s12880-023-01140-0] [Medline: 37993801]
51. Basso Dias A, Mirshahvalad SA, Ortega C, et al. The role of [^{18}F]-DCFPyL PET/MRI radiomics for pathological grade group prediction in prostate cancer. Eur J Nucl Med Mol Imaging. Jun 2023;50(7):2167-2176. [doi: 10.1007/s00259-023-06136-0] [Medline: 36809425]
52. Lin S, Deng L, Hu Z, et al. Predicting FOXA1 gene mutation status in prostate cancer through multi-modal deep learning. Biomed Signal Process Control. Aug 2025;106:107739. [doi: 10.1016/j.bspc.2025.107739]
53. Zhang C, Wang Z, Shang P, et al. Combining multi-parametric MRI radiomics features with tumor abnormal protein to construct a machine learning-based predictive model for prostate cancer. Sci Rep. 2025;15(1):22816. [doi: 10.1038/s41598-025-05718-2]
54. Sun Y, Fang J, Shi Y, et al. Machine learning based on radiomics features combing B-mode transrectal ultrasound and contrast-enhanced ultrasound to improve peripheral zone prostate cancer detection. Abdom Radiol. ;49(1):141-150. [doi: 10.1007/s00261-023-04050-5]
55. Li Y, Imami MR, Zhao L, et al. An automated deep learning-based framework for uptake segmentation and classification on PSMA PET/CT imaging of patients with prostate cancer. J Digit Imaging Inform med. 2024;37(5):2206-2215. [doi: 10.1007/s10278-024-01104-y]
56. Zhang H, Ji J, Liu Z, et al. Artificial intelligence for the diagnosis of clinically significant prostate cancer based on multimodal data: a multicenter study. BMC Med. 2023;21(1):270. [doi: 10.1186/s12916-023-02964-x]
57. Karagoz A, Alis D, Seker ME, et al. Anatomically guided self-adapting deep neural network for clinically significant prostate cancer detection on bi-parametric MRI: a multi-center study. Insights Imaging. Jun 19, 2023;14(1):110. [doi: 10.1186/s13244-023-01439-0] [Medline: 37337101]
58. Ismayilov R, Aktas A, Gencoglu EA, Oguz A, Altundag O, Akcali Z. Staging prostate cancer with AI: a comparative study of large language models and expert interpretation on PSMA PET-CT reports. Mol Imaging Biol. Feb 2026;28(1):93-105. [doi: 10.1007/s11307-025-02072-7] [Medline: 41350971]
59. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions: guidance for industry and food and drug administration staff. U.S. Food and Drug Administration; 2025. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence [Accessed 2026-10-28]
60. Dhiman P, Ma J, Andaur Navarro CL, et al. Overinterpretation of findings in machine learning prediction model studies in oncology: a systematic review. J Clin Epidemiol. May 2023;157:120-133. [doi: 10.1016/j.jclinepi.2023.03.012] [Medline: 36935090]

61. Alzate-Grisales JA. Multimodal-ai-prostate-cancer-scoping-review. Github. 2026. URL: https://github.com/jesusalzate/multimodal-ai-prostate-cancer-scoping-review [Accessed 2026-10-28]

Abbreviations

AUC: area under the curve
AUROC: area under the receiver operating characteristic curve
bpMRI: biparametric magnetic resonance imaging
CatBoost: categorical boosting
CHAARTED: ChemoHormonal Therapy Versus Androgen Ablation Randomized Trial for Extensive Disease in Prostate Cancer
CLAIM: Checklist for AI in Medical Imaging
CNN: convolutional neural network
CRPC: castration-resistant prostate cancer
csPCa: clinically significant prostate cancer
CT: computed tomography
ISUP: International Society of Urological Pathology
LASSO: least absolute shrinkage and selection operator
mpMRI: multiparametric magnetic resonance imaging
MRI: magnetic resonance imaging
PCa: prostate cancer
PET: positron emission tomography
PI-CAI: prostate imaging: cancer AI
PI-RADS: Prostate Imaging Reporting and Data System
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PROBAST+AI: Prediction Model Risk of Bias Assessment Tool+AI
PSA: prostate-specific antigen
PSMA: prostate-specific membrane antigen
QUADAS-AI: Quality Assessment of Diagnostic Accuracy Studies–AI
RCT: randomized clinical trial
SHAP: Shapley Additive Explanations
STARD-AI: Standards for Reporting Diagnostic Accuracy-AI
TRIPOD+AI: Transparent Reporting of Multivariable Prediction Models with AI
XGBoost: extreme gradient boosting

Edited by Matthew Balcarras; peer-reviewed by Morteza Alipour, Rashad Ismayilov; submitted 01.Jul.2026; initial comments to author 26.Aug.2026; final revised version received 07.Sep.2026; accepted 10.Sep.2026; published 27.Nov.2026

Please cite as:
Alzate-Grisales JA, Ocampo Osorio F, Restrepo Orozco JD, Rendon Tabares JD, Perán-Teruel M, Navarro Beltrán A, Ruiz Torres C, García-García F, Tabares-Soto R, García-Gómez JM, de la Iglesia-Vayá M
Multimodal AI for Prostate Cancer Diagnosis and Risk Stratification: Scoping Review
J Med Internet Res 2026;28:e105912
URL: https://www.jmir.org/2026/1/e105912
doi: 10.2196/105912

© Jesus Alejandro Alzate-Grisales, Felipe Ocampo Osorio, Jose Daniel Restrepo Orozco, Juan Diego Rendon Tabares, Miguel Perán-Teruel, Azahar Navarro Beltrán, Clara Ruiz Torres, Francisco García-García, Reinel Tabares-Soto, Juan M García-Gómez, Maria de la Iglesia-Vayá. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.