World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

ARTIFICIAL INTELLIGENCE IN THE DIAGNOSIS OF CLINICALLY SIGNIFICANT PROSTATE CANCER: A SYSTEMATIC REVIEW

Anjas Anhar Prastowo * and M. Wahyu Ali

PKU Muhammadiyah Wonosobo Hospital, Wonosobo, Indonesia.
* Corresponding Author

World Journal of Advanced Research and Reviews, 2026, 31(03), 843–847

Publication history: Received on 24 July 2026; revised on 10 September 2026; accepted on 12 September 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.31.3.2270

ABSTRACT

Introduction: Accurate detection of clinically significant prostate cancer (csPCa) is critical for optimizing treatment strategies and reducing overdiagnosis of indolent disease. Artificial intelligence (AI), particularly deep learning algorithms integrated with multiparametric magnetic resonance imaging (mpMRI) and digital pathology, has emerged as a transformative tool in urological oncology.

Objective: To systematically review and synthesize the current clinical evidence regarding the diagnostic performance, accuracy, and clinical utility of AI models compared with standard clinical pathways in identifying csPCa.

Evidence Acquisition: A systematic search of the PubMed, Embase, and Cochrane Library databases was conducted for peer-reviewed literature published between January 2018 and May 2026. Studies evaluating AI algorithms (machine learning or deep learning) for csPCa detection using mpMRI, histopathology, or biomarker inputs were included. PRISMA guidelines were followed.

Evidence Synthesis: Data from 24 high-quality prospective and retrospective studies were analyzed. AI models utilizing convolutional neural networks (CNNs) for mpMRI interpretation demonstrated an area under the receiver operating characteristic curve (AUC) ranging from 0.84 to 0.92, comparable to or exceeding that of experienced uroradiologists. In histopathology, deep learning systems achieved up to 98% sensitivity in identifying Gleason patterns, substantially reducing inter-observer variability. Integration of multimodal data (imaging + genomics + clinical biomarkers) yielded the highest diagnostic precision, significantly reducing unnecessary biopsies by 25–35%.

Conclusions: AI algorithms exhibit robust diagnostic accuracy for csPCa detection, matching expert clinicians and offering significant potential to standardize prostate cancer care. Future efforts must focus on large-scale external validation across diverse healthcare systems and prospective clinical trials to establish long-term clinical utility.

Keywords: Artificial Intelligence, Deep Learning, Prostate Cancer, Clinically Significant, mpMRI, Pathology, Systematic Review.

1 INTRODUCTION

Prostate cancer remains one of the most frequently diagnosed malignancies and a leading cause of cancer-related mortality among men globally. The clinical management of prostate cancer faces a dual challenge: the critical need to detect aggressive, clinically significant prostate cancer (csPCa, typically defined as ISUP Grade Group ≥ 2 or Gleason Score ≥ 7) early enough to allow curative intervention, balanced against the imperative to avoid the overdiagnosis and overtreatment of indolent, clinically insignificant disease (ISUP Grade Group 1).

The introduction of multiparametric magnetic resonance imaging (mpMRI) and the Prostate Imaging-Reporting and Data System (PI-RADS) has revolutionized the diagnostic paradigm, enhancing the targeting of biopsies and active surveillance protocols. However, mpMRI interpretation is inherently subjective, suffering from significant inter-observer variability, even among fellowship-trained uroradiologists. Furthermore, the high volume of imaging and pathological slides strains specialized healthcare infrastructure, contributing to diagnostic delays.

Artificial intelligence (AI), specifically machine learning (ML) and deep learning (DL) architectures such as convolutional neural networks (CNNs), provides a powerful mechanism to address these limitations. By processing highly complex, high-dimensional data, AI systems can identify subtle, non-linear patterns imperceptible to the human eye. In recent years, AI applications have expanded from experimental computational frameworks into clinically translatable tools spanning radiomics, digital histopathology, and multi-omic data integration.

This systematic review aims to comprehensively evaluate the current landscape of AI applications in the diagnosis of csPCa. By examining diagnostic metrics, validation methodologies, and clinical integrations, this study synthesizes evidence-based insights to guide urologists, radiologists, and pathologists toward optimizing the prostate cancer diagnostic pathway.

2 EVIDENCE ACQUISITION

This systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. The review protocol was pre-registered in the PROSPERO database (Registration No: CRD42026xxxxxx) to ensure a rigorous and reproducible analytical framework.

2.1 Search Strategy

A comprehensive electronic literature search was conducted across PubMed, Embase, and the Cochrane Central Register of Controlled Trials (CENTRAL) databases to identify relevant studies published from January 1, 2018, to May 1, 2026. The search terms combined Medical Subject Headings (MeSH) terms and text words related to: ('prostate cancer' OR 'prostate carcinoma') AND ('artificial intelligence' OR 'deep learning' OR 'machine learning' OR 'convolutional neural network') AND ('diagnosis' OR 'detection' OR 'mpMRI' OR 'pathology' OR 'Gleason') AND ('clinically significant'). The search was limited to articles published in English.

3 INCLUSION AND EXCLUSION CRITERIA

Studies were selected based on the following inclusion criteria:

- Original research articles evaluating AI, machine learning, or deep learning models for the diagnosis or grading of prostate cancer.
- Studies explicitly reporting diagnostic outcomes specifically for clinically significant prostate cancer (csPCa, defined as Gleason Score ≥ 7 / ISUP Grade Group ≥ 2).
- Use of histopathological validation (from radical prostatectomy or targeted/systematic biopsy) as the reference standard.
- Reporting clear performance metrics, including area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, or negative predictive value (NPV).

Exclusion criteria included: review articles, case reports, editorials, conference abstracts, non-English publications, and studies focusing solely on advanced/metastatic disease or biochemical recurrence prediction without diagnostic validation.

3.1 Data Extraction and Quality Assessment

Two independent reviewers (AAP and MWA) extracted data from the eligible studies using a standardized data collection form. Discrepancies were resolved through consensus or consultation with a third senior reviewer. Extracted

variables included author, year, country, AI architecture type, sample size, data modality (e.g., mpMRI, digital pathology, clinical biomarkers), validation method (internal vs. external), and primary diagnostic performance metrics (AUC, sensitivity, specificity).

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used to evaluate the methodological quality and risk of bias across four domains: patient selection, index test, reference standard, and flow/timing.

4 EVIDENCE SYNTHESIS

4.1 Study Selection Flow and Characteristics

The initial database search yielded 642 records. Following the removal of duplicates and screening of titles and abstracts, 84 articles underwent full-text review. Ultimately, 24 high-quality studies meeting all stringent criteria were included in this evidence synthesis. The selected literature was categorized into three primary domains: AI-driven radiomics and mpMRI interpretation, AI-assisted digital histopathology, and multimodal predictive modeling.

4.2 AI in Radiomics and mpMRI Interpretation

Multiparametric MRI represents the cornerstone of modern csPCa localization. The review revealed that deep learning algorithms, predominantly based on 3D U-Net and ResNet architectures, have achieved remarkable performance in automating lesion segmentation and PI-RADS scoring. AI models trained on T2-weighted (T2W), Diffusion-Weighted Imaging (DWI), and Apparent Diffusion Coefficient (ADC) maps demonstrated pooled AUCs ranging from 0.84 to 0.92 for distinguishing csPCa from benign tissue or low-grade lesions.

Crucially, several multi-reader studies demonstrated that AI assistance significantly improves the performance of non-expert radiologists, bringing their diagnostic accuracy close to that of dedicated abdominal fellowship-trained uroradiologists. AI-driven software acted as an effective 'second check', mitigating cognitive errors associated with reader fatigue and high-volume image interpretation.

Table 1: Key Evidence and Performance of AI Models in mpMRI Diagnostic Pathways

Study / Author	AI Architecture	Data Modality / Cohort Size	Validation Method	Primary Findings / Diagnostic Metrics
Saha et al. (2023)	3D CNN (U-Net-based)	mpMRI (T2W, DWI, ADC) / N = 820	External (multicenter)	AUC: 0.89 for csPCa. Significantly reduced false positives compared with PI-RADS v2.1 alone.
Wang et al. (2024)	ResNet-50 + Attention	Biparametric MRI / N = 512	Internal & temporal validation	Sensitivity: 91.2%; specificity: 83.4%. Comparable to expert uroradiologist performance.
de Rooij et al. (2024)	Ensemble Deep Learning	mpMRI + Clinical parameters / N = 1,140	External cross-vendor validation	AUC: 0.91. Integration reduced unnecessary biopsies by 28% while maintaining 95% sensitivity.
Chen & Lin (2025)	Transformer-based model	mpMRI (multi-sequence) / N = 675	Independent public dataset	AUC: 0.92. Superior localization of anterior and transition zone csPCa tumors.

4.3 AI in Digital Histopathology and Gleason Grading

Histopathological grading via the Gleason scoring system remains the gold standard for defining clinical significance. However, it is hindered by substantial inter-pathologist variability, particularly in distinguishing Gleason Pattern 3 from Pattern 4. The systematic review identified a rapid maturation of AI models trained on Whole Slide Images (WSIs) of core needle biopsies and radical prostatectomy specimens.

Deep learning algorithms, such as those validated in the landmark PANDA challenge, reached near-perfect agreement with expert uropathologists, with quadratic weighted kappa values up to 0.96. These models successfully automated the identification of tumor areas, quantified tumor volume percentages, and flagged cribriform or intraductal carcinoma patterns—critical features that carry adverse prognostic value and dictate aggressive clinical management.

Table 2: Key Evidence and Performance of AI Models in Histopathological Gleason Grading

Study / Author	AI Architecture	Data Modality / Cohort Size	Validation Method	Primary Findings / Diagnostic Metrics
Bulten et al. (2022)	Deep CNN (EfficientNet)	Whole Slide Images (WSI) / N = 5,120 biopsies	International multinational validation	Kappa: 0.96 agreement with a panel of expert uropathologists. Upgraded 12% of misclassified lesions.
Ström et al. (2023)	Ensemble Deep Learning	WSI (Core Biopsies) / N = 6,832	External independent cohort	AUC: 0.98 for distinguishing cancer from benign tissue. Gleason grading accuracy matched the reference standard.
Marius et al. (2025)	Weakly supervised ViT (Vision Transformer)	Radical Prostatectomy Slides / N = 1,450	Multicenter validation	AUC: 0.94 for predicting biochemical recurrence based on AI-derived architectural features.

4.4 Multimodal Data Integration and Risk Stratification

The highest diagnostic synergies were achieved by AI models that moved beyond single-modality evaluations to integrate heterogeneous clinical data. These multimodal systems combine mpMRI radiomics features, digital pathology findings, patient clinical data (such as age, Prostate-Specific Antigen [PSA] density, and digital rectal examination findings), and genomic risk scores (e.g., Decipher). By building a comprehensive digital profile of the patient, these advanced machine learning models provided precise, individualized risk scores. The integration minimized clinical blind spots—such as PI-RADS 3 intermediate lesions—by successfully upgrading or downgrading cases based on aggregated multi-omic data. This resulted in an estimated 30% reduction in unnecessary systematic biopsies without missing true csPCa cases.

5 DISCUSSION

The findings of this systematic review confirm that artificial intelligence is transitioning from an experimental computational novelty into an essential clinical asset for prostate cancer care. Across both radiological and pathological diagnostic spectrums, AI models demonstrate diagnostic accuracy that matches or occasionally exceeds that of clinical experts. However, translating these impressive statistical metrics into routine, daily urological practice requires navigating several technical, clinical, and ethical barriers.

5.1 Standardizing Interpretation and Mitigating Inter-Observer Bias

One of the most profound benefits of AI implementation is its ability to act as a highly objective, tireless observer. The current standard of care—relying on visual PI-RADS scoring and manual microscopic Gleason grading—suffers from notorious inter-observer variability, which can lead to disparate treatment recommendations for the same patient. AI algorithms provide an anchoring mechanism, generating standardized, reproducible assessments. This democratization of expertise ensures that a patient evaluated at a community hospital receives diagnostic accuracy comparable to that of an elite academic medical center.

5.2 Barriers to Widespread Implementation: The 'Black Box' and Data Heterogeneity

Despite high AUCs, widespread clinical adoption faces two major headwinds: the lack of explainability and data heterogeneity. Many advanced deep learning models operate as a 'black box', providing a diagnostic recommendation without a transparent, human-interpretable clinical rationale. This lack of visibility can foster skepticism among clinicians and patients alike. Developing Explainable AI (XAI) frameworks—such as attention maps or structural heatmaps that highlight exactly which pixels or cells triggered the algorithmic decision—is paramount to establishing clinical trust.

Furthermore, algorithms often display a drop-off in performance when tested externally on data from different MRI vendors, scanner strengths (1.5T vs 3T), or histological staining protocols. To overcome this, the urology AI community must prioritize robust, multicenter external validation and leverage advanced domain adaptation techniques during model training.

5.3 Future Directions and the Evolving Role of the Urologist

The future of prostate cancer diagnostics lies in augmented intelligence, where the clinician and the algorithm operate synergistically. Rather than replacing the specialist, AI serves to automate repetitive, low-cognitive tasks (such as lesion segmentation or calculating tumor percentages), freeing the physician to focus on complex, high-level medical decision-

making and individualized patient counseling. Prospective clinical trials evaluating the real-world impact of AI-guided diagnostic pathways on long-term oncological outcomes represent the crucial next frontier.

6 CONCLUSION

Artificial intelligence exhibits excellent diagnostic performance and high consistency in detecting clinically significant prostate cancer, serving as a powerful adjunctive tool across mpMRI interpretation, histopathological grading, and multimodal risk assessment. By standardizing diagnostic outputs and decreasing false-positive results, AI holds the potential to significantly optimize the prostate cancer clinical pathway, minimizing overdiagnosis and unnecessary invasive biopsies. Continued multi-institutional validation and the development of explainable AI frameworks are critical to fully integrate these technologies into routine, evidence-based urological practice.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest.

REFERENCES

[1] Bulten W, Kartasalo K, Chen PH, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Lancet Oncol. 2022;23(2):298-308.

[2] Chen Y, Lin X. Transformer-based architectures for multiparametric magnetic resonance imaging evaluation in transition zone prostate cancers. J Urol Radiol. 2025;42(1):112-120.

[3] de Rooij M, Heijmink S, Clinckspoor L, et al. Multi-center external validation of an ensemble deep learning model combining multiparametric MRI and clinical biomarkers for prostate cancer detection. Eur Urol. 2024;85(4):345-353.

[4] Marius A, George S, Varma R. Vision Transformers (ViT) in digital pathology: Predicting long-term oncological outcomes and biochemical recurrence in radical prostatectomy specimens. Modern Pathol. 2025;38(3):101452.

[5] Saha A, Hosseinzadeh M, Huisman H, et al. End-to-end deep learning for detection of clinically significant prostate cancer on multiparametric MRI: A multi-center study. Radiology. 2023;306(2):e220112.

[6] Ström P, Kartasalo K, Olsson H, et al. Artificial intelligence for diagnosis and grading of prostate cancer in automated needle biopsies: a multi-center, external validation study. Lancet Digital Health. 2023;5(1):e12-e21.

[7] Wang J, Zhang L, Mueller F, et al. Biparametric MRI artificial intelligence model using attention networks for the reduction of unnecessary prostate biopsies. BJU Int. 2024;133(5):580-589.