INPLASY

INPLASY2025110042

doi: 10.37766/inplasy2025.11.0042

Received: 15 November 2025

Published: 16 November 2025

Comparison of Artificial Intelligence and Radiologists in MRI-Based Prostate Cancer Diagnosis: A Meta-Analysis of Accuracy and Effectiveness

Chen, HQ; Li, EW; Christos, PJ; Zhu, YS.

Corresponding author:
Huiqi Chen

huc4007@alumni.weill.cornell.edu

Author Affiliation:
Weill Cornell Medicine, Cornell University.

ADMINISTRATIVE INFORMATION

Support - UL1TR002384；KL2TR002385.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY2025110042

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 16 November 2025 and was last updated on 16 November 2025.

INTRODUCTION

Review question / Objective Population: Adult men undergoing MRI for prostate cancer evaluation. Intervention: Artificial intelligence models (AI/DL-based MRI interpretation). Comparator: Radiologists using PI-RADS v2.1. Outcomes: Diagnostic accuracy, sensitivity, specificity, AUROC. Study design: Clinical diagnostic studies directly comparing AI vs radiologists.

Review Question:
Does artificial intelligence (AI) perform as well as or better than radiologists in MRI-based detection of clinically significant prostate cancer?

Rationale Prostate cancer remains a major global health burden, and accurate MRI interpretation is essential for detecting clinically significant prostate cancer (csPCa). However, diagnostic performance among radiologists varies widely due to differences in experience and inter-reader inconsistency. Artificial intelligence (AI), particularly deep learning methods, has emerged as a potential tool for improving diagnostic accuracy and reducing variability.

Despite the rapid development of AI algorithms for prostate MRI, existing studies differ considerably in design, patient populations, MRI acquisition parameters, and reported diagnostic metrics. These inconsistencies make it difficult for clinicians and researchers to understand whether AI can reliably match or surpass radiologists in real-world diagnostic performance.

While narrative reviews exist, few quantitative meta-analyses directly compare AI with radiologists across multiple studies. The lack of consolidated evidence leaves uncertainty about AI's true diagnostic utility. Therefore, a structured and transparent systematic review and meta-analysis is needed to synthesize the current evidence, evaluate comparative diagnostic performance, and guide future clinical adoption of AI-assisted MRI reading.

Condition being studied Prostate cancer is the most commonly diagnosed cancer among men and a leading cause of cancer-related death. Multiparametric MRI (mpMRI) is a key diagnostic tool used before biopsy to detect suspicious lesions and guide clinical decision-making. Clinically significant prostate cancer (csPCa) typically refers to Gleason Grade Group ≥2. Accurate MRI interpretation is essential for avoiding unnecessary biopsies and improving early detection.

METHODS

Search strategy This review followed PRISMA 2020 guidelines. A systematic search was conducted in PubMed on March 14, 2025. The search strategy included structured MeSH terms and keyword combinations such as "Artificial Intelligence," "Machine Learning," "Prostate Cancer," "Prostatic Neoplasms," "Magnetic Resonance Imaging," and "Radiologists." Four structured queries were used, including MeSH-guided expressions. PubMed filters for Clinical Study, Randomized Controlled Trial, and Comparative Study were applied. Reference lists of included studies were manually screened to identify additional eligible publications.

Participant or population Adult male patients (≥18 years) undergoing MRI evaluation for suspected or known prostate cancer.

Intervention Artificial intelligence or deep learning–based models used to interpret prostate MRI images and detect clinically significant prostate cancer.

Comparator Radiologists interpreting prostate MRI using PI-RADS (v2.1) or other standard clinical reading approaches.

Study designs to be included Retrospective, prospective, comparative, or randomized clinical diagnostic studies directly comparing AI with radiologists.

Eligibility criteria --Inclusion: Adult patients; MRI-based diagnostic studies; direct comparison of AI vs radiologists; reported sensitivity, specificity, or AUROC; sample size reported; peer-reviewed full text; English language.
--Exclusion: AI development studies without clinical comparison; non-MRI modalities; no radiologist comparison; missing diagnostic metrics; case reports, reviews, abstracts; preclinical or animal studies; procedural or AI-guided biopsy studies unless diagnostic accuracy was assessed.

Information sources PubMed was the primary database. Reference lists from included studies were manually reviewed to ensure comprehensive coverage.

Main outcome(s) Primary diagnostic outcomes included sensitivity, specificity, and AUROC for AI and radiologists. Lesion-level outcomes were extracted when available; otherwise, patient-level metrics were used.

Additional outcome(s) Workflow efficiency, reading time, and radiologist variability improvements when reported.

Data management Records were screened using EndNote. Data were extracted manually into structured tables capturing study characteristics, performance metrics, reference standards, sample sizes, and diagnostic thresholds.

Quality assessment / Risk of bias analysis Risk of bias was assessed using QUADAS-2 across four domains: patient selection, index test, reference standard, and flow and timing. ROBVIS was used for visualization.

Strategy of data synthesis Pooled sensitivity and specificity were calculated using a random-effects model. Heterogeneity was assessed using Cochran's Q and I^2. AUROC differences were evaluated using a weighted paired approach, with standard errors extracted or back-calculated. Forest plots, funnel plots, and paired difference plots were generated using R (v4.5.1).

Subgroup analysis Subgroup analysis was performed when studies included multiple internal or external test sets, or when both lesion-level and patient-level data were reported.

Sensitivity analysis Sensitivity analyses excluded studies with zero standard error in sensitivity calculations and evaluated robustness of results by weighting based on precision.

Language restriction English only.

Country(ies) involved United States.

Other relevant information This review uses publicly available published data only.

Keywords Prostate cancer, MRI, Machine learning, Artificial Intelligence, Deep Learning, Radiologists, Meta-Analysis.

Contributions of each author

Author 1 - Huiqi Chen.
Email: huc4007@alumni.weill.cornell.edu
Author 2 - Erwang Li.
Email: erl4012@alumni.weill.cornell.edu
Author 3 - Paul J Christos.
Email: pac2001@med.cornell.edu
Author 4 - Yuan-Shan Zhu.
Email: yuz2002@med.cornell.edu