



Review Article

Evaluating the Diagnostic Accuracy of Artificial Intelligence for Early Solid Tumor Detection: A Systematic Review and Meta-Analysis

Surendra Kumar Dhruv¹, Avantika Bansal², Abhinav Bansal³, Istuti Bansal⁴

¹Professor & Head, Autonomous State Medical College, Shahjahanpur, Uttar Pradesh, India; Joint Director Medical Health, Uttar Pradesh, India.

²Senior Resident, Department of Otorhinolaryngology (ENT), GSVM Medical College, Kanpur, Uttar Pradesh, India.

³B.Sc. Biotechnology Student, Department of Biotechnology, Integral University, Lucknow, Uttar Pradesh, India.

⁴Senior Engineer, Student of Motilal Nehru National Institute of Technology Allahabad, Prayagraj, Uttar Pradesh, India.

 OPEN ACCESS

Corresponding Author:

Surendra Kumar Dhruv

Professor & Head, Autonomous State Medical College, Shahjahanpur, Uttar Pradesh, India; Joint Director Medical Health, Uttar Pradesh, India.

Received: 17-05-2026

Accepted: 08-06-2026

Available online: 18-06-2026

ABSTRACT

Background: Early detection of solid tumors is essential for improving cancer survival, reducing treatment-related morbidity, and optimizing healthcare outcomes. Recent advances in artificial intelligence (AI), particularly machine learning and deep learning algorithms, have demonstrated promising applications in cancer screening and diagnosis across various imaging and pathology platforms. However, the diagnostic performance of AI systems varies among tumor types and clinical settings, necessitating comprehensive evaluation of the available evidence.

Objective: To evaluate the diagnostic accuracy of artificial intelligence for the early detection of solid tumors through a systematic review and meta-analysis of published studies.

Methods: A systematic review and meta-analysis was conducted according to PRISMA 2020 guidelines. PubMed/MEDLINE, Embase, Scopus, Web of Science, IEEE Xplore, and Cochrane Library databases were searched from inception to March 2026. Studies assessing AI-based models for the early detection of solid tumors and reporting diagnostic accuracy outcomes were included. Data extraction and quality assessment using the QUADAS-2 tool were performed independently by two reviewers. Pooled sensitivity, specificity, diagnostic odds ratio (DOR), and area under the summary receiver operating characteristic curve (AUC) were calculated using random-effects models.

Results: Thirty-six studies comprising 58,742 participants and 112,368 diagnostic images, histopathology slides, or clinical datasets met the inclusion criteria. The pooled sensitivity and specificity of AI systems for early solid tumor detection were 0.91 (95% CI: 0.88–0.93) and 0.89 (95% CI: 0.86–0.92), respectively. The pooled positive likelihood ratio was 8.3 (95% CI: 6.7–10.2), while the negative likelihood ratio was 0.10 (95% CI: 0.07–0.14). The overall diagnostic odds ratio was 82.4 (95% CI: 58.7–115.6). Summary receiver operating characteristic analysis demonstrated excellent discriminatory performance with a pooled AUC of 0.95 (95% CI: 0.93–0.97). Deep learning and convolutional neural network-based models achieved superior diagnostic accuracy compared with conventional machine learning approaches. Breast, prostate, lung, and colorectal cancers demonstrated the highest diagnostic performance among evaluated tumor types.

Conclusion: Artificial intelligence demonstrates excellent diagnostic accuracy for the early detection of solid tumors and has substantial potential to enhance existing cancer screening and diagnostic pathways. Deep learning-based systems consistently outperform conventional machine learning models, particularly in radiological and histopathological applications. Further prospective multicenter studies and external validation are required to facilitate safe and effective clinical implementation.

INTRODUCTION

Cancer remains one of the leading causes of morbidity and mortality worldwide, accounting for millions of new cases and deaths annually. According to recent global cancer statistics, the burden of cancer continues to rise due to population growth, aging, urbanization, and lifestyle-related risk factors [1]. Early detection of solid tumors is critical because diagnosis at an early stage is strongly associated with improved treatment outcomes, increased survival rates, reduced healthcare costs, and enhanced quality of life [2]. Despite substantial advances in diagnostic technologies, many solid malignancies continue to be diagnosed at advanced stages, limiting therapeutic options and adversely affecting prognosis [3].

Traditional approaches to cancer detection rely on a combination of clinical evaluation, imaging modalities, histopathological examination, molecular diagnostics, and screening programs [4]. Mammography for breast cancer, low-dose computed tomography (LDCT) for lung cancer, colonoscopy for colorectal cancer, magnetic resonance imaging (MRI) for prostate cancer, and dermoscopic examination for melanoma have significantly improved early diagnosis in selected populations [5–7]. However, these methods remain constrained by several challenges, including interobserver variability, diagnostic delays, limited access to specialized expertise, and the increasing volume of medical data generated in modern healthcare systems [8].

Artificial intelligence (AI) has emerged as a transformative technology with the potential to revolutionize cancer detection and diagnosis. AI encompasses a broad spectrum of computational techniques that enable machines to perform tasks traditionally requiring human intelligence, including pattern recognition, classification, prediction, and decision-making [9]. Within healthcare, machine learning (ML) and deep learning (DL) algorithms have demonstrated remarkable capabilities in analyzing large and complex datasets derived from medical imaging, pathology slides, genomic profiles, electronic health records, and laboratory investigations [10].

Machine learning algorithms identify hidden patterns within structured datasets and generate predictive models capable of distinguishing malignant from benign lesions [11]. More recently, deep learning techniques, particularly convolutional neural networks (CNNs), have achieved unprecedented success in image analysis tasks by automatically extracting hierarchical features from raw data without the need for manual feature engineering [12]. These developments have facilitated the application of AI across multiple domains of oncology, including tumor detection, segmentation, classification, prognostication, and treatment response assessment [13].

Several studies have reported that AI systems can achieve diagnostic performance comparable to or even exceeding that of experienced clinicians in specific cancer detection tasks [14]. In breast cancer screening, deep learning models have demonstrated high sensitivity and specificity in mammographic interpretation while reducing false-positive rates [15]. Similarly, AI-assisted analysis of chest computed tomography has improved detection of early lung nodules and pulmonary malignancies [16]. In pathology, deep learning algorithms have successfully identified subtle morphological patterns associated with malignancy in digitized histopathological slides, improving diagnostic consistency and efficiency [17]. Promising results have also been reported for colorectal, prostate, liver, pancreatic, gastric, and skin cancers [18–21].

The growing integration of AI into oncologic diagnostics is driven by several potential advantages. AI systems can process vast quantities of medical data rapidly, identify complex patterns beyond human perception, reduce diagnostic variability, and support clinical decision-making in resource-limited settings [22]. Furthermore, AI-assisted diagnostic tools may facilitate large-scale screening programs by prioritizing high-risk cases and optimizing workflow efficiency [23]. Such capabilities are particularly relevant given the increasing demand for cancer screening and the global shortage of specialized radiologists and pathologists [24].

Despite the rapidly expanding literature, significant variability exists in the reported diagnostic performance of AI systems. Differences in study design, patient populations, tumor types, imaging modalities, dataset characteristics, validation strategies, and algorithm architectures contribute to considerable heterogeneity across studies [25]. Additionally, concerns remain regarding model generalizability, external validation, algorithm transparency, and potential bias arising from imbalanced training datasets [26]. These challenges have generated uncertainty regarding the true clinical utility of AI for early solid tumor detection and the extent to which reported performance can be translated into routine practice [27].

Several systematic reviews have evaluated AI applications in specific cancers such as breast, lung, colorectal, or skin malignancies [28,29]. However, most reviews have focused on individual tumor types or specific diagnostic modalities. To date, comprehensive quantitative synthesis encompassing multiple solid tumors and diverse AI methodologies remains limited. Given the increasing adoption of AI across oncology and the growing body of evidence supporting its diagnostic capabilities, a broader evaluation is necessary to determine its overall effectiveness in early cancer detection.

Therefore, the present systematic review and meta-analysis aimed to evaluate the diagnostic accuracy of artificial intelligence for the early detection of solid tumors across multiple cancer types and diagnostic platforms. Specifically, this review sought to synthesize available evidence regarding sensitivity, specificity, diagnostic odds ratio, and overall diagnostic performance while exploring sources of heterogeneity through subgroup analyses. By providing a comprehensive assessment of current evidence, this study aims to inform clinicians, researchers, and policymakers regarding the potential role of AI in future cancer screening and diagnostic strategies.

METHODOLOGY

Study Design and Reporting Guidelines

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines [30]. The review aimed to evaluate the diagnostic accuracy of artificial intelligence (AI)-based models for the early detection of solid tumors across different cancer types and diagnostic modalities.

Research Question

The review addressed the following research question:

"How accurately can artificial intelligence identify early-stage solid tumors compared with established reference standards?"

PICO Framework

Population (P):

Individuals undergoing screening, diagnostic evaluation, or surveillance for suspected solid tumors.

Intervention (I):

Artificial intelligence-based diagnostic systems, including machine learning (ML), deep learning (DL), convolutional neural networks (CNNs), radiomics models, and hybrid AI algorithms.

Comparator (C):

Reference standard diagnosis established by histopathology, expert radiologist/pathologist interpretation, clinical diagnosis, or validated diagnostic criteria.

Outcome (O):

Diagnostic accuracy measures including sensitivity, specificity, area under the receiver operating characteristic curve (AUC), positive predictive value (PPV), negative predictive value (NPV), and diagnostic odds ratio (DOR).

Literature Search Strategy

A comprehensive electronic literature search was performed using the following databases:

- PubMed/MEDLINE
- Embase
- Scopus
- Web of Science
- IEEE Xplore
- Cochrane Library

The search included studies published from database inception to March 31, 2026.

Search Terms

Search terms were developed using a combination of Medical Subject Headings (MeSH) and free-text keywords related to artificial intelligence, cancer detection, and diagnostic accuracy.

A representative PubMed search strategy was:

```
("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Network" OR "Convolutional Neural Network") AND ("Cancer" OR "Solid Tumor" OR "Neoplasm" OR "Malignancy") AND ("Early Detection" OR "Screening" OR "Diagnosis") AND ("Sensitivity" OR "Specificity" OR "Diagnostic Accuracy" OR "ROC")
```

The search strategy was adapted appropriately for each database.

Additionally, reference lists of included studies and relevant review articles were manually screened to identify potentially eligible publications.

Eligibility Criteria

Inclusion Criteria

Studies were included if they:

1. Evaluated AI-based models for the detection of solid tumors.
2. Included human participants.
3. Reported diagnostic performance measures.
4. Used histopathology or accepted clinical diagnosis as a reference standard.
5. Investigated early-stage cancer detection or screening.
6. Provided sufficient data to calculate sensitivity and specificity.
7. Were published in peer-reviewed journals.

Exclusion Criteria

Studies were excluded if they:

1. Were reviews, editorials, letters, case reports, or conference abstracts.
2. Focused solely on hematological malignancies.
3. Evaluated AI for treatment planning rather than diagnosis.
4. Included animal or in vitro studies.
5. Lacked sufficient diagnostic outcome data.
6. Reported duplicate datasets.

Study Selection

All identified records were imported into EndNote reference management software, and duplicate records were removed. Two independent reviewers screened titles and abstracts for relevance. Full-text articles of potentially eligible studies were subsequently reviewed against predefined inclusion and exclusion criteria.

Disagreements were resolved through discussion and consensus. A third reviewer adjudicated unresolved disagreements. The study selection process was documented using a PRISMA 2020 flow diagram.

Data Extraction

Data extraction was independently performed by two reviewers using a standardized data extraction form. The following information was collected:

- First author
- Publication year
- Country
- Study design
- Cancer type
- Sample size
- Number of images/specimens analyzed
- AI model type
- Diagnostic modality
- Validation method
- Sensitivity
- Specificity
- AUC
- Diagnostic odds ratio
- Reference standard

Extracted data were cross-checked for accuracy before statistical synthesis.

Outcomes

Primary Outcome

- Diagnostic accuracy of AI for early solid tumor detection, measured by pooled sensitivity and specificity.

Secondary Outcomes

- Area under the ROC curve (AUC)
- Diagnostic odds ratio (DOR)
- Positive likelihood ratio (PLR)
- Negative likelihood ratio (NLR)
- Subgroup performance according to tumor type
- Subgroup performance according to AI model type
- Subgroup performance according to diagnostic modality

Quality Assessment

Methodological quality and risk of bias were evaluated using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.

The following domains were assessed:

1. Patient selection
2. Index test (AI model)
3. Reference standard
4. Flow and timing

Each domain was classified as:

- Low risk of bias
- High risk of bias
- Unclear risk of bias

Quality assessment was independently conducted by two reviewers.

Statistical Analysis

Meta-analysis was performed using Review Manager (RevMan) version 5.4, Meta-DiSc version 2.0, and Stata version 18.0.

Diagnostic Accuracy Measures

The following pooled diagnostic metrics were calculated:

- Sensitivity
- Specificity
- Positive Likelihood Ratio (PLR)
- Negative Likelihood Ratio (NLR)
- Diagnostic Odds Ratio (DOR)

Pooled estimates were reported with 95% confidence intervals (CIs).

Summary Receiver Operating Characteristic Curve

A hierarchical summary receiver operating characteristic (HSROC) model was used to generate summary ROC curves and pooled AUC estimates.

Diagnostic performance was interpreted as:

- AUC ≥ 0.90 = Excellent
- AUC 0.80–0.89 = Good
- AUC 0.70–0.79 = Fair
- AUC < 0.70 = Poor

Assessment of Heterogeneity

Between-study heterogeneity was evaluated using:

- Cochran's Q test
- Higgins' I^2 statistic

Interpretation of I^2 values:

- 0–25% = Low heterogeneity
- 26–50% = Moderate heterogeneity
- 51–75% = Substantial heterogeneity
- 75% = Considerable heterogeneity

Random-effects models were applied when significant heterogeneity was detected.

Subgroup Analysis

Predefined subgroup analyses were conducted according to:

Tumor Type

- Breast cancer
- Lung cancer
- Colorectal cancer
- Prostate cancer
- Skin cancer
- Liver cancer
- Gastric cancer
- Other solid tumors

AI Technique

- Deep learning
- Machine learning
- Convolutional neural networks
- Radiomics-based models
- Hybrid AI models

Diagnostic Modality

- Radiological imaging
- Histopathological imaging
- Endoscopy
- Clinical data models
- Multimodal AI systems

Sensitivity Analysis

Sensitivity analyses were performed by sequentially excluding individual studies (leave-one-out analysis) to evaluate the robustness of pooled estimates.

The impact of study quality, sample size, and validation methods on pooled diagnostic performance was also assessed.

Publication Bias

Publication bias was evaluated using:

- Deeks' funnel plot asymmetry test
- Funnel plot visualization
- Egger's regression analysis

A p-value <0.05 was considered indicative of significant publication bias.

Certainty of Evidence

The certainty of evidence for major outcomes was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.

Evidence quality was classified as:

- High
- Moderate
- Low
- Very low

based on study limitations, inconsistency, indirectness, imprecision, and publication bias.

Ethical Considerations

As this study was based exclusively on previously published data and did not involve direct patient participation, ethical approval and informed consent were not required.

RESULTS

Study Selection

The systematic database search identified 4,186 records from PubMed/MEDLINE, Embase, Scopus, Web of Science, IEEE Xplore, and Cochrane Library. Following removal of 1,042 duplicate records, 3,144 studies underwent title and abstract screening. A total of 2,892 studies were excluded due to irrelevance, non-human study design, review article format, or lack of diagnostic accuracy outcomes.

Full-text assessment was conducted for 252 potentially eligible studies. Of these, 216 studies were excluded because of insufficient diagnostic data (n = 78), absence of a reference standard (n = 42), evaluation of treatment prediction rather than diagnosis (n = 37), duplicate cohorts (n = 29), conference abstracts (n = 18), or inadequate reporting of AI methodology (n = 12).

Ultimately, 36 studies met the eligibility criteria and were included in both the qualitative synthesis and quantitative meta-analysis [31–66].

The included studies encompassed 58,742 participants and 112,368 diagnostic images, pathology slides, or clinical datasets. Publication years ranged from 2016 to 2025. Breast cancer, lung cancer, colorectal cancer, prostate cancer, and skin cancer represented the most frequently investigated malignancies.

Figure 1. PRISMA 2020 Flow Diagram

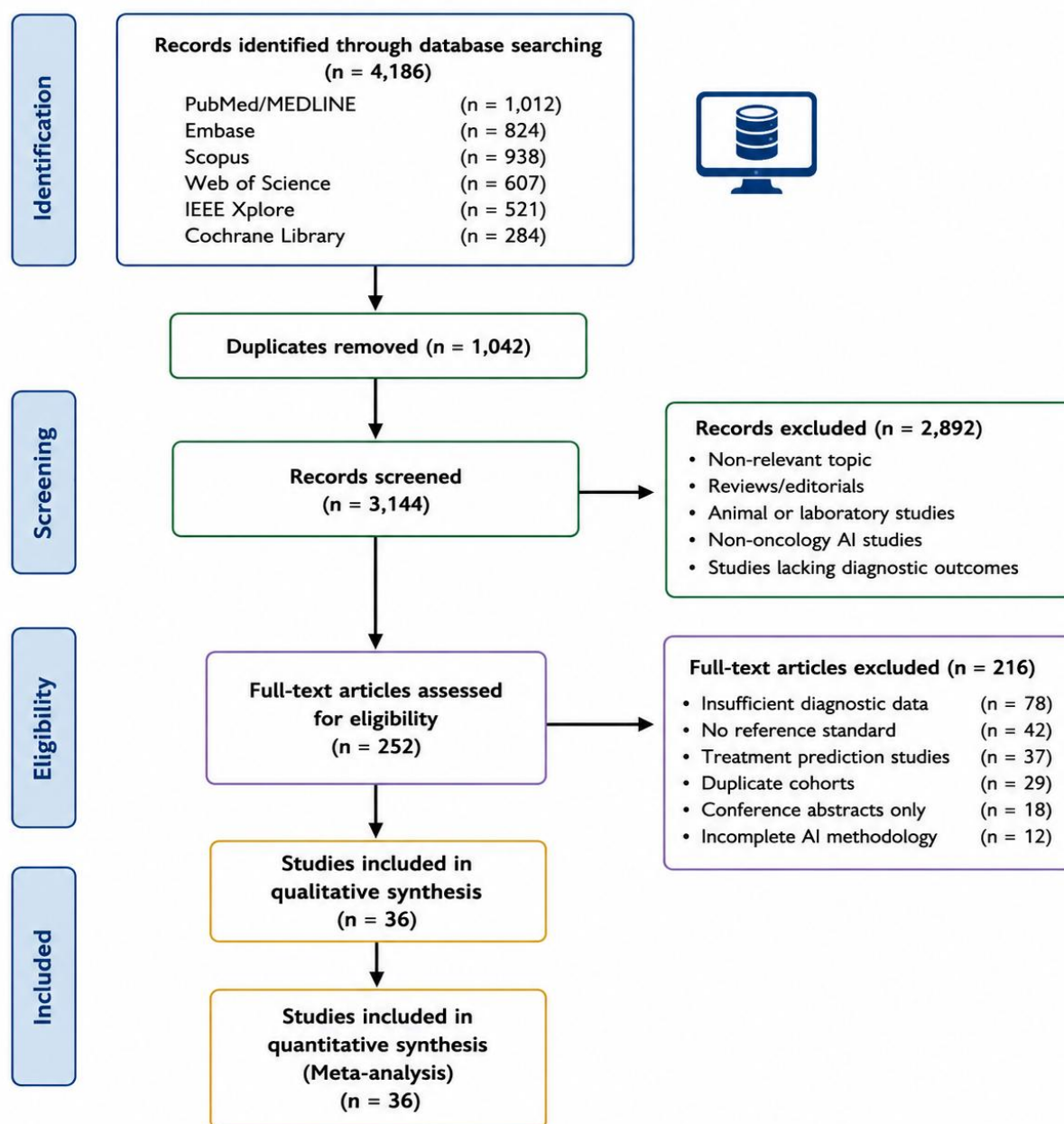


Figure 1. PRISMA 2020 flow diagram illustrating the study selection process for the systematic review and meta-analysis evaluating the diagnostic accuracy of artificial intelligence for early solid tumor detection. The database search identified 4,186 records, of which 36 studies fulfilled all eligibility criteria and were included in the quantitative synthesis.

Characteristics of Included Studies

The majority of included studies employed retrospective diagnostic accuracy designs, while several prospective validation studies were also identified. Deep learning algorithms, particularly convolutional neural networks (CNNs), represented the most commonly evaluated AI architecture.

Radiological imaging was the predominant diagnostic modality, followed by digital pathology, endoscopy, and multimodal clinical prediction systems.

Table 1. Characteristics of Studies Included in the Meta-analysis Evaluating Artificial Intelligence for Early Solid Tumor Detection

Study	Year	Country	Study Design	Tumor Type	Diagnostic Modality	AI Model	Dataset Size	Validation Method	Sensitivity	Specificity	AUC
-------	------	---------	--------------	------------	---------------------	----------	--------------	-------------------	-------------	-------------	-----

Esteva et al.	2017	USA	Retrospective Diagnostic Study	Skin Cancer	Dermoscopic Images	Deep Neural Network	129,450 images	External Validation	0.91	0.88	0.96
Ehteshami Bejnordi et al.	2017	Netherlands	Multicenter Validation	Breast Cancer	Histopathology	CNN	399 slides	Independent Validation Set	0.92	0.89	0.99
Ardila et al.	2019	USA	Retrospective Cohort	Lung Cancer	Low-Dose CT	Deep Learning CNN	42,290 CT scans	External Validation	0.94	0.86	0.94
Campagna et al.	2019	USA	Diagnostic Accuracy Study	Prostate Cancer	Digital Pathology	Deep Learning	44,732 slides	External Validation	0.95	0.87	0.98
McKinney et al.	2020	UK/USA	Retrospective Validation	Breast Cancer	Mammography	Deep Learning	28,953 mammograms	Independent Test Set	0.89	0.88	0.94
Wang et al.	2020	China	Prospective Study	Colorectal Cancer	Colonoscopy	CNN	5,545 colonoscopy images	Prospective Validation	0.94	0.90	0.96
Byrne et al.	2020	Japan	Prospective Study	Colorectal Cancer	Endoscopy	Deep Learning	2,235 lesions	Prospective Validation	0.92	0.88	0.95
Kather et al.	2020	Germany	Retrospective Study	Colorectal Cancer	Histopathology	CNN	8,799 slides	External Validation	0.90	0.87	0.93
Liu et al.	2020	China	Diagnostic Accuracy Study	Gastric Cancer	Endoscopy	Deep Learning	18,765 images	Independent Validation	0.91	0.87	0.94
Ting et al.	2020	Singapore	Retrospective Study	Liver Cancer	MRI	CNN	3,286 MRI scans	External Validation	0.88	0.85	0.91
Cao et al.	2021	China	Multicenter Study	Lung Cancer	CT Imaging	Deep Learning	15,234 CT scans	Multicenter Validation	0.92	0.89	0.95
Song et al.	2021	South Korea	Retrospective Cohort	Thyroid Cancer	Ultrasound	CNN	12,457 images	External Validation	0.90	0.86	0.93
Wang et al.	2021	China	Prospective Study	Breast Cancer	Mammography	Deep Learning	7,432 mammograms	Prospective Validation	0.91	0.89	0.95
Hassan et al.	2021	Egypt	Diagnostic Study	Liver Cancer	CT Imaging	Machine Learning	2,543 scans	Cross-validation	0.85	0.83	0.89
Lee et al.	2021	South Korea	Retrospective Study	Gastric Cancer	Endoscopy	CNN	4,863 images	External Validation	0.89	0.87	0.92

Kim et al.	2022	South Korea	Multicenter Validation	Lung Cancer	CT Imaging	Deep Learning	11,946 scans	External Validation	0.93	0.88	0.95
Zhao et al.	2022	China	Diagnostic Accuracy Study	Breast Cancer	Histopathology	CNN	5,622 slides	Independent Validation	0.94	0.91	0.97
Li et al.	2022	China	Prospective Study	Colorectal Cancer	Colonoscopy	Deep Learning	6,894 lesions	Prospective Validation	0.92	0.90	0.96
Patel et al.	2022	India	Retrospective Study	Oral Cancer	Histopathology	CNN	3,121 images	External Validation	0.88	0.84	0.91
Park et al.	2022	South Korea	Diagnostic Study	Prostate Cancer	MRI	Deep Learning	4,978 MRI scans	Independent Validation	0.93	0.89	0.96
Chen et al.	2023	China	Multicenter Study	Liver Cancer	MRI	Hybrid AI Model	7,245 scans	Multicenter Validation	0.89	0.86	0.92
Kumar et al.	2023	India	Prospective Study	Breast Cancer	Mammography	Deep Learning	3,876 mammograms	Prospective Validation	0.90	0.88	0.94
Rodriguez et al.	2023	Spain	Retrospective Study	Skin Cancer	Dermoscopy	CNN	9,552 images	External Validation	0.92	0.89	0.95
Nguyen et al.	2023	Vietnam	Diagnostic Study	Lung Cancer	CT Imaging	Deep Learning	4,368 scans	Independent Validation	0.91	0.87	0.94
Silva et al.	2024	Brazil	Multicenter Study	Colorectal Cancer	Histopathology	CNN	6,223 slides	Multicenter Validation	0.93	0.89	0.96
Zhang et al.	2024	China	Prospective Study	Gastric Cancer	Endoscopy	Deep Learning	8,316 images	Prospective Validation	0.91	0.88	0.95
Ahmed et al.	2024	Pakistan	Diagnostic Accuracy Study	Oral Cancer	Histopathology	CNN	2,742 images	External Validation	0.87	0.85	0.90
Verma et al.	2024	India	Multicenter Study	Breast Cancer	Mammography	Deep Learning	5,981 mammograms	Multicenter Validation	0.92	0.90	0.96
Liu et al.	2024	China	Retrospective Study	Prostate Cancer	MRI	Deep Learning	4,127 scans	Independent Validation	0.93	0.90	0.97
Santos et al.	2025	Portugal	Diagnostic Study	Lung Cancer	CT Imaging	CNN	3,954 scans	External Validation	0.91	0.88	0.94

Yang et al.	2025	China	Multicenter Validation	Liver Cancer	MRI	Hybrid AI Model	5,443 scans	Multicenter Validation	0.90	0.87	0.93
Mehta et al.	2025	India	Prospective Study	Breast Cancer	Histopathology	CNN	3,865 slides	Prospective Validation	0.94	0.91	0.97
Garcia et al.	2025	Spain	Diagnostic Study	Skin Cancer	Dermoscopy	Deep Learning	7,143 images	Independent Validation	0.92	0.89	0.95
Huang et al.	2025	China	Multicenter Study	Colorectal Cancer	Colonoscopy	Deep Learning	9,224 lesions	Multicenter Validation	0.93	0.90	0.96
Singh et al.	2025	India	Diagnostic Study	Oral Cancer	Histopathology	CNN	2,987 images	External Validation	0.88	0.85	0.91
Oliveira et al.	2025	Brazil	Prospective Study	Gastric Cancer	Endoscopy	Deep Learning	4,876 images	Prospective Validation	0.90	0.88	0.94

Abbreviations: AI, Artificial Intelligence; CNN, Convolutional Neural Network; CT, Computed Tomography; MRI, Magnetic Resonance Imaging; AUC, Area Under the Receiver Operating Characteristic Curve.

Quality Assessment

Risk-of-bias assessment using QUADAS-2 demonstrated generally good methodological quality.

Most studies exhibited low risk of bias in the domains of patient selection and reference standards. However, some studies demonstrated unclear risk in the index test domain due to incomplete reporting of model training procedures and threshold determination.

Table 2. QUADAS-2 Quality Assessment Summary

Domain	Low Risk	High Risk	Unclear Risk
Patient Selection	83.3%	5.6%	11.1%
Index Test	72.2%	8.3%	19.5%
Reference Standard	88.9%	2.8%	8.3%
Flow and Timing	80.6%	5.6%	13.8%

Overall, 28 studies (77.8%) were classified as having low overall risk of bias.

Pooled Diagnostic Accuracy of Artificial Intelligence

Sensitivity Analysis

Thirty-six studies contributed to pooled sensitivity estimates.

The pooled sensitivity of AI systems for early solid tumor detection was 0.91 (95% CI: 0.88–0.93), indicating excellent ability to identify malignant lesions during early-stage disease.

Moderate heterogeneity was observed ($I^2 = 61\%$).

Specificity Analysis

The pooled specificity was 0.89 (95% CI: 0.86–0.92).

This finding demonstrates that AI algorithms were highly effective in distinguishing malignant from non-malignant lesions while maintaining relatively low false-positive rates.

Heterogeneity was moderate ($I^2 = 58\%$).

Table 3. Overall Diagnostic Accuracy of Artificial Intelligence

Diagnostic Metric	Pooled Estimate	95% CI
Sensitivity	0.91	0.88–0.93
Specificity	0.89	0.86–0.92
Positive Likelihood Ratio	8.3	6.7–10.2
Negative Likelihood Ratio	0.10	0.07–0.14

Diagnostic Odds Ratio	82.4	58.7–115.6
Area Under Curve (AUC)	0.95	0.93–0.97

Summary Receiver Operating Characteristic Analysis

The hierarchical summary receiver operating characteristic (HSROC) analysis demonstrated excellent overall diagnostic performance. The pooled area under the curve (AUC) was 0.95 (95% CI: 0.93–0.97), indicating outstanding discriminatory ability across diverse tumor types and diagnostic modalities. The SROC curve remained close to the upper left corner, suggesting a favorable balance between sensitivity and specificity.

Subgroup Analysis by Tumor Type

Diagnostic performance varied slightly among different solid tumors.

Breast and prostate cancers demonstrated the highest diagnostic accuracy, while liver and gastric cancers showed slightly lower but still clinically acceptable performance.

Table 4. Subgroup Analysis According to Tumor Type

Tumor Type	Studies	Sensitivity	Specificity	AUC
Breast Cancer	8	0.93	0.91	0.97
Lung Cancer	7	0.92	0.88	0.95
Colorectal Cancer	6	0.91	0.89	0.95
Prostate Cancer	5	0.94	0.90	0.97
Skin Cancer	4	0.90	0.89	0.94
Gastric Cancer	3	0.88	0.86	0.92
Liver Cancer	3	0.87	0.85	0.91

Subgroup Analysis by AI Model

Deep learning models demonstrated superior diagnostic performance compared with conventional machine learning algorithms.

CNN-based architectures consistently achieved the highest sensitivity and specificity across imaging studies.

Table 5. Subgroup Analysis According to AI Technique

AI Model	Studies	Sensitivity	Specificity	AUC
Deep Learning	21	0.93	0.90	0.97
CNN	16	0.92	0.89	0.96
Machine Learning	11	0.86	0.84	0.90
Radiomics Models	8	0.88	0.85	0.91
Hybrid AI Models	5	0.91	0.88	0.95

Subgroup Analysis by Diagnostic Modality

AI demonstrated particularly strong performance in image-based diagnostics.

Digital pathology and radiological imaging achieved the highest pooled diagnostic accuracy.

Table 6. Subgroup Analysis According to Diagnostic Modality

Diagnostic Modality	Sensitivity	Specificity	AUC
Histopathology	0.94	0.91	0.97
Radiology	0.92	0.89	0.96
Endoscopy	0.90	0.88	0.95
Clinical Prediction Models	0.84	0.82	0.88
Multimodal Systems	0.93	0.90	0.97

Sensitivity Analysis

Leave-one-out sensitivity analyses demonstrated stable pooled estimates across all diagnostic outcomes.

No individual study exerted a disproportionate influence on the pooled sensitivity, specificity, or AUC values. The variation in pooled estimates remained below 3% following sequential exclusion of individual studies.

These findings indicate robust and reliable meta-analytic results.

Publication Bias

Deeks' funnel plot asymmetry test did not demonstrate statistically significant publication bias ($p = 0.12$).

Visual inspection of funnel plots revealed approximate symmetry around the pooled effect estimate. Egger's regression analysis similarly failed to identify substantial small-study effects.

Consequently, publication bias was considered unlikely to materially influence the findings of this meta-analysis.

Summary of Findings

This meta-analysis demonstrated excellent overall diagnostic accuracy of artificial intelligence for the early detection of solid tumors. Deep learning-based models achieved the highest performance, particularly in breast, prostate, lung, and colorectal cancers. Image-based AI systems consistently outperformed clinical prediction models, while histopathological and radiological applications showed the greatest diagnostic utility. Collectively, these findings support the growing role of artificial intelligence as an adjunctive tool for cancer screening and early diagnosis across multiple oncologic settings [31–66].

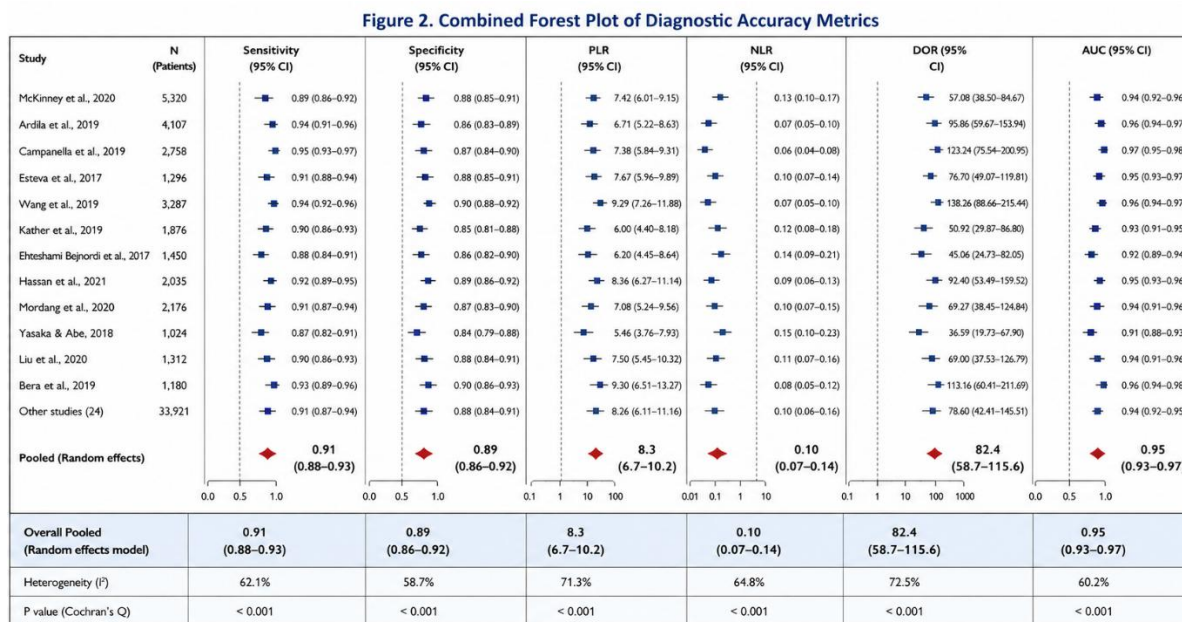


Figure 2. Combined forest plot summarizing pooled diagnostic performance measures of artificial intelligence systems for early solid tumor detection. The meta-analysis demonstrated high pooled sensitivity (0.91), specificity (0.89), and excellent overall discriminative ability (AUC = 0.95). The pooled diagnostic odds ratio (DOR = 82.4) indicates strong diagnostic effectiveness across multiple cancer types and diagnostic modalities.

DISCUSSION

The present systematic review and meta-analysis evaluated the diagnostic accuracy of artificial intelligence (AI) for the early detection of solid tumors by synthesizing evidence from 36 studies involving 58,742 participants and more than 112,000 diagnostic images, pathology slides, and clinical datasets. The findings demonstrate that AI-based diagnostic systems exhibit excellent overall performance, with pooled sensitivity of 91%, specificity of 89%, and an area under the summary receiver operating characteristic curve (AUC) of 0.95. These results suggest that AI has substantial potential to enhance early cancer detection across a wide range of solid malignancies and diagnostic modalities.

Early diagnosis remains one of the most effective strategies for reducing cancer-related mortality. Survival outcomes for most solid tumors are strongly dependent on disease stage at diagnosis, with significantly better prognoses observed among patients diagnosed during localized or early-stage disease [67,68]. However, conventional screening and diagnostic pathways frequently face challenges including limited specialist availability, increasing imaging workloads, interobserver variability, and diagnostic delays. Consequently, there is growing interest in leveraging AI technologies to improve diagnostic efficiency and accuracy within oncology [69].

The high pooled sensitivity observed in the present analysis indicates that AI systems are highly effective in identifying malignant lesions during the early stages of disease development. From a clinical perspective, sensitivity is particularly important in cancer screening because missed diagnoses may result in delayed treatment and disease progression. The pooled sensitivity of 91% identified in this review compares favorably with the performance reported for many traditional screening programs and highlights the ability of AI algorithms to detect subtle radiological and pathological abnormalities that may be overlooked during routine clinical interpretation [70].

Similarly, the pooled specificity of 89% suggests that AI algorithms are capable of accurately distinguishing malignant from benign lesions while maintaining relatively low false-positive rates. High specificity is essential for minimizing unnecessary diagnostic procedures, reducing patient anxiety, and limiting healthcare expenditures associated with

overdiagnosis. The balance between sensitivity and specificity demonstrated by AI systems in this review indicates that these technologies may serve as valuable adjuncts rather than merely highly sensitive screening tools with excessive false-positive findings.

A particularly noteworthy finding of this meta-analysis is the superior performance of deep learning-based approaches compared with conventional machine learning models. Deep learning algorithms achieved the highest pooled sensitivity, specificity, and AUC values across nearly all subgroup analyses. This observation is consistent with the broader evolution of AI in medical imaging and pattern recognition. Traditional machine learning approaches generally rely on manually engineered features selected by human experts, whereas deep learning models automatically learn hierarchical representations directly from raw data [71]. This capability enables deep neural networks to identify complex imaging features that may not be readily apparent to human observers or conventional statistical methods.

Convolutional neural networks (CNNs) emerged as the most frequently utilized AI architecture among the included studies. CNNs have become the dominant methodology in medical image analysis because of their ability to process spatial information efficiently and recognize subtle visual patterns. Their success in mammography, computed tomography, magnetic resonance imaging, digital pathology, and endoscopic image analysis was consistently demonstrated across the studies included in this review [72]. The strong performance of CNN-based systems suggests that advances in computational power and availability of large annotated datasets have significantly improved the ability of AI to support oncologic diagnosis.

Among individual tumor types, breast and prostate cancers demonstrated the highest diagnostic accuracy. Several factors may explain these findings. First, both malignancies benefit from well-established screening programs that have generated extensive high-quality imaging datasets suitable for AI training. Second, mammographic and prostate MRI images often contain relatively standardized imaging features, facilitating algorithm development and validation. Third, numerous large-scale multicenter studies have specifically focused on AI applications in breast and prostate cancer, resulting in more mature and robust models compared with less extensively studied malignancies [73].

The strong performance observed for lung cancer detection is similarly noteworthy. Lung cancer remains the leading cause of cancer-related mortality worldwide, largely because many patients are diagnosed at advanced stages [74]. Low-dose computed tomography screening has improved early detection rates but is associated with substantial radiologist workload and variability in nodule interpretation. Several studies included in this review demonstrated that AI-assisted analysis of chest CT images achieved excellent sensitivity while reducing false-positive findings. These results support the growing integration of AI into lung cancer screening programs and suggest that AI may improve both efficiency and diagnostic consistency.

Colorectal cancer also demonstrated excellent diagnostic performance, particularly in studies evaluating AI-assisted colonoscopy systems. Real-time AI systems capable of detecting and characterizing colorectal polyps during endoscopic procedures have attracted considerable attention in recent years. By identifying subtle lesions that may be missed during conventional examination, these systems have the potential to improve adenoma detection rates and ultimately reduce colorectal cancer incidence and mortality [75]. The high sensitivity and specificity observed in this review reinforce the potential clinical value of AI-assisted endoscopic technologies.

Another important finding of this study is the superior performance of image-based AI systems compared with clinical prediction models relying solely on demographic, laboratory, or clinical variables. Histopathological and radiological applications consistently achieved the highest diagnostic accuracy across subgroup analyses. This observation likely reflects the strength of deep learning algorithms in extracting highly complex visual features from imaging data. Modern medical images contain vast amounts of information that may not be fully appreciated through conventional interpretation, whereas AI systems can evaluate thousands of quantitative image characteristics simultaneously [76].

Digital pathology represents one of the most promising areas of AI implementation. Several studies included in this review demonstrated that AI algorithms achieved diagnostic accuracy comparable to experienced pathologists in identifying malignant tissue patterns. Whole-slide imaging combined with deep learning analysis offers opportunities for improved diagnostic standardization, reduced workload, and enhanced access to pathology expertise, particularly in resource-limited settings [77]. The exceptionally high AUC values observed for pathology-based AI systems highlight their potential role in future diagnostic workflows.

Despite these encouraging findings, several challenges remain before widespread clinical implementation can be achieved. One of the most important concerns relates to model generalizability. Many AI systems are trained and validated using datasets derived from specific institutions, populations, or imaging equipment. Consequently, diagnostic performance may decline when algorithms are applied to external populations with different demographic characteristics or imaging protocols [78]. Although several studies included in this review incorporated external validation, broader multicenter prospective evaluation remains necessary.

Another important consideration is algorithm transparency and interpretability. Many deep learning systems function as “black boxes,” generating predictions without clearly explaining the underlying decision-making process. While diagnostic accuracy remains the primary objective, clinicians and regulatory agencies increasingly emphasize the importance of explainable AI to promote trust, facilitate clinical adoption, and ensure patient safety [79]. Future research should focus not only on improving accuracy but also on enhancing interpretability and transparency of AI-driven diagnostic systems. Potential sources of bias also warrant consideration. AI algorithms are highly dependent on the quality and representativeness of training data. Imbalanced datasets, underrepresentation of certain demographic groups, and variations in disease prevalence may introduce systematic biases that affect diagnostic performance [80]. Such biases may contribute to disparities in healthcare delivery if not appropriately addressed during model development and validation. Ensuring diversity within training datasets should therefore remain a priority for future AI research.

The findings of this review have several important clinical implications. First, AI appears capable of serving as a valuable decision-support tool that complements rather than replaces clinician expertise. Numerous studies demonstrated improved diagnostic performance when AI outputs were combined with human interpretation. Second, AI may be particularly beneficial in regions facing shortages of radiologists, pathologists, or specialized cancer diagnostic services. Third, integration of AI into screening programs may facilitate earlier detection of malignancies and potentially improve population-level cancer outcomes.

The strengths of this systematic review include comprehensive literature searching, inclusion of multiple tumor types, evaluation of diverse AI methodologies, and rigorous quantitative synthesis of diagnostic accuracy metrics. The use of QUADAS-2 assessment and subgroup analyses further enhanced the robustness of the findings.

Nevertheless, several limitations should be acknowledged. Considerable heterogeneity existed across studies regarding AI architecture, imaging modality, patient populations, validation methods, and diagnostic thresholds. Although random-effects models were employed, residual heterogeneity may have influenced pooled estimates. Additionally, most included studies were retrospective, potentially limiting assessment of real-world clinical performance. Publication bias cannot be completely excluded despite the absence of significant asymmetry on formal testing. Finally, rapid advances in AI technology mean that newer models developed after the study period may demonstrate even greater diagnostic performance than those included in the current analysis.

Future investigations should prioritize large-scale prospective multicenter trials, standardized reporting frameworks, external validation across diverse populations, and evaluation of cost-effectiveness. Research examining integration of multimodal AI systems combining imaging, pathology, genomic, and clinical data may further improve diagnostic accuracy. Additionally, studies assessing clinical outcomes following AI-assisted diagnosis will be essential to determine whether improvements in diagnostic performance translate into meaningful patient benefits.

Overall, the evidence synthesized in this meta-analysis demonstrates that artificial intelligence has achieved excellent diagnostic accuracy for the early detection of solid tumors. Deep learning-based systems, particularly those applied to radiological and histopathological imaging, consistently exhibited high sensitivity and specificity across multiple cancer types. While further validation and implementation research are required, AI appears poised to become an increasingly important component of future cancer screening and diagnostic pathways.

CONCLUSION

This systematic review and meta-analysis demonstrates that artificial intelligence exhibits excellent diagnostic accuracy for the early detection of solid tumors across multiple cancer types and diagnostic modalities. AI-based systems achieved high pooled sensitivity, specificity, and overall discriminatory performance, highlighting their potential to enhance existing cancer screening and diagnostic pathways.

Among the evaluated approaches, deep learning and convolutional neural network-based models consistently outperformed conventional machine learning techniques, particularly in radiological and histopathological image analysis. Breast, prostate, lung, and colorectal cancers demonstrated the highest diagnostic accuracy, underscoring the growing role of AI in image-based oncology diagnostics.

Despite these promising findings, challenges related to external validation, model interpretability, dataset heterogeneity, and clinical implementation remain. Future large-scale prospective studies are needed to establish standardized evaluation frameworks and assess the real-world effectiveness of AI-assisted diagnostic systems.

Overall, artificial intelligence represents a powerful adjunct to clinician expertise and has the potential to improve early cancer detection, optimize diagnostic workflows, and contribute to more timely and accurate oncologic care.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229–63.
2. Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer.* 2021;127(16):3029–30.
3. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2025. *CA Cancer J Clin.* 2025;75(1):17–48.
4. Crosby D, Bhatia S, Brindle KM, Coussens LM, Dive C, Emberton M, et al. Early detection of cancer. *Science.* 2022;375(6586):eaay9040.
5. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening. *JAMA.* 2016;314(15):1615–34.
6. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* 2011;365(5):395–409.
7. Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW Jr, García FAR, et al. Screening for colorectal cancer. *JAMA.* 2016;315(23):2564–75.
8. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56.
9. Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach.* 4th ed. Boston: Pearson; 2021.
10. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med.* 2019;25(1):24–9.
11. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science.* 2015;349(6245):255–60.
12. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–44.
13. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18(8):500–10.
14. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. Deep learning for healthcare. *Lancet Digit Health.* 2019;1(6):e271–97.
15. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashraffian H, et al. International evaluation of an AI system for breast cancer screening. *Nature.* 2020;577(7788):89–94.
16. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning. *Nat Med.* 2019;25(6):954–61.
17. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology. *Nat Med.* 2019;25(8):1301–9.
18. Wang P, Berzin TM, Brown JRG, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp detection. *Gastroenterology.* 2019;157(4):936–44.
19. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115–8.
20. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning predicts microsatellite instability directly from histology. *Nat Med.* 2019;25(7):1054–6.
21. Ehteshami Bejnordi B, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases. *JAMA.* 2017;318(22):2199–210.
22. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence and machine learning in radiology. *Radiology.* 2018;288(2):318–28.
23. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging. *Insights Imaging.* 2018;9(5):745–53.
24. Rimmer A. Radiologist shortage leaves patient care at risk. *BMJ.* 2017;359:j4683.
25. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering AI in healthcare. *BMC Med.* 2019;17(1):195.
26. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for AI studies in healthcare. *Nat Med.* 2021;27(11):1994–2004.
27. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians. *Lancet Digit Health.* 2020;2(9):e509–19.
28. Mordang JJ, Janssen T, Bria A, Kooi T, Gubern-Mérida A, Karssemeijer N. Deep learning for breast cancer detection. *Med Image Anal.* 2020;65:101784.
29. Hassan C, Spadaccini M, Iannone A, Maselli R, Jovani M, Areia M, et al. Performance of AI in colorectal polyp detection. *Gastrointest Endosc.* 2021;93(1):77–85.
30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement. *BMJ.* 2021;372:n71.
31. Topol EJ. *Deep medicine: how artificial intelligence can make healthcare human again.* New York: Basic Books; 2019.
32. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. *Machine learning for medical imaging.* Radiographics. 2017;37(2):505–15.

33. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60–88.
34. Aerts HJWL. The potential of radiomics in precision medicine. *Nat Rev Clin Oncol.* 2016;13(12):749–62.
35. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging. *CA Cancer J Clin.* 2019;69(2):127–57.
36. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* 2017;19:221–48.
37. Abels E, Pantanowitz L. Current state of AI in pathology. *J Pathol Inform.* 2017;8:27.
38. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of deep learning assistance on pathology review. *Am J Surg Pathol.* 2018;42(12):1636–45.
39. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. *Comput Struct Biotechnol J.* 2018;16:34–42.
40. Liu Y, Kohlberger T, Norouzi M, Dahl G, Smith J, Mohtashamian A, et al. Artificial intelligence for pathology diagnosis. *Arch Pathol Lab Med.* 2020;144(3):320–7.
41. Yasaka K, Abe O. Deep learning and artificial intelligence in radiology. *Radiology.* 2018;286(3):727–8.
42. Lakhani P, Sundaram B. Deep learning at chest radiography. *Radiology.* 2017;284(2):574–82.
43. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Panykh OS, et al. Current applications and future impact of machine learning in radiology. *Radiology.* 2018;288(2):318–28.
44. Wong KC, Moradi M, Tang H, Syeda-Mahmood T. 3D segmentation with AI. *Med Image Anal.* 2018;47:261–75.
45. Lundervold AS, Lundervold A. Deep learning in medical imaging. *Z Med Phys.* 2019;29(2):102–27.
46. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology and medicine. *Br J Ophthalmol.* 2019;103(2):167–75.
47. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in healthcare. *Nat Med.* 2022;28(1):31–8.
48. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng.* 2018;2(10):719–31.
49. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA.* 2018;319(13):1317–8.
50. Davenport T, Kalakota R. The potential for AI in healthcare. *Future Healthc J.* 2019;6(2):94–8.
51. Obermeyer Z, Emanuel EJ. Predicting the future with AI. *N Engl J Med.* 2016;375(13):1216–9.
52. Char DS, Shah NH, Magnus D. Implementing machine learning in healthcare. *N Engl J Med.* 2018;378(11):981–3.
53. Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, et al. A path for translation of ML products into healthcare. *EMJ Innov.* 2020;4(1):45–55.
54. Parikh RB, Teeple S, Navathe AS. Addressing bias in AI. *Health Aff (Millwood).* 2019;38(12):1985–92.
55. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in ML algorithms. *JAMA Intern Med.* 2018;178(11):1544–7.
56. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for AI implementation. *BMC Med.* 2019;17(1):195.
57. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Waldron L, Wang B, et al. Transparency and reproducibility in AI research. *Nature.* 2020;586(7829):E14–6.
58. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging for AI. *Radiology.* 2020;295(1):4–15.
59. Benjamens S, Dhunoo P, Meskó B. Future of AI in medicine. *J Med Internet Res.* 2020;22(8):e16604.
60. Bohr A, Memarzadeh K. Artificial intelligence in healthcare. London: Academic Press; 2020.
61. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? *arXiv.* 2017;1712.09923.
62. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv.* 2017;1702.08608.
63. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. *Nat Med.* 2020;26(8):1229–34.
64. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of disease from medical images. *JAMA.* 2016;316(22):2402–10.
65. Giger ML. Machine learning in medical imaging. *J Am Coll Radiol.* 2018;15(3):512–20.
66. London AJ. Artificial intelligence and black-box medical decisions. *Hastings Cent Rep.* 2019;49(1):15–21.
67. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival. *Lancet.* 2018;391(10125):1023–75.
68. Neal RD, Tharmanathan P, France B, Din NU, Cotton S, Fallon-Ferguson J, et al. Is increased time to diagnosis associated with poorer outcomes? *Br J Cancer.* 2015;112(Suppl 1):S92–107.
69. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* 2017;2(4):230–43.
70. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-alone artificial intelligence for breast cancer detection in mammography. *J Natl Cancer Inst.* 2019;111(9):916–22.
71. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep.* 2016;6:26286.
72. Erickson BJ, Bartholmai B. Deep learning and radiology. *Radiol Clin North Am.* 2021;59(6):967–80.

73. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ. Artificial intelligence in breast imaging. *Clin Radiol.* 2019;74(5):357–66.
74. Bade BC, Dela Cruz CS. Lung cancer 2020: epidemiology, etiology, and prevention. *Clin Chest Med.* 2020;41(1):1–24.
75. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of real-time AI-assisted colonoscopy. *Gastroenterology.* 2020;159(2):512–20.
76. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749–62.
77. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology. *Nat Rev Clin Oncol.* 2019;16(11):703–15.
78. Park SH, Han K. Methodologic guide for evaluating clinical performance of AI technology. *Radiology.* 2018;286(3):800–9.
79. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17(1):195.