International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://www.theaspd.com/ijes.php

Diagnostic Accuracy of Artificial Intelligence-Assisted Radiology in Detecting Pulmonary Nodules on Chest CT: A Narrative Review

Badr Soliman Ibrahim Alaskar[1], Nasser Mohammed Alqahtani[2], Othman Mohammed Abdulaziz Alhussain[3], Sulaiman Nawaf Sulaiman Almutairi[3], Nawaf Moeedh Alrashidi[3], Ahmed Othman Abdullah Alabdulwahab[1], Fhad Yaqoub Alyousef[3]

[1]Medical device technician, University Medical Center, Majmaah University, Majmaah 11952, Saudi Arabia; email: bs.alaskar@mu.edu.sa

[2]Medical device technician, College of Applied Medical Sciences, Majmaah University, Majmaah 11952, Saudi Arabia; email: nm.alqahtani@mu.edu.sa

[3]Medical device technician, College of Medicine, Majmaah University, Majmaah 11952, Saudi Arabia; email: u.alhussain@mu.edu.sa

***Correspondence**: Othman Mohammed Abdulaziz Alhussain: Medical device technician, College of Medicine, Majmaah University, Majmaah 11952, Saudi Arabia; email: u.alhussain@mu.edu.sa

---

ABSTRACT

Background- Lung cancer has consistently been the world leader in cancer mortality, and screening has significantly reduced mortality with early detection. Even so, routine radiologic interpretation of low-dose computed tomography (LDCT) scans was met with challenges of high false positives, reader fatigue, and difficulty in detecting non-palpable pulmonary nodules. Artificial intelligence (AI), or rather deep learning algorithms, has been proposed as a potential adjunctive strategy in enhancing diagnostic accuracy in the detection of pulmonary nodules.

Methods- A narrative review was initiated to assess the diagnostic accuracy of AI-based radiology systems to identify pulmonary nodules on chest CT scans. Ten recent peer-reviewed clinical papers published in the last decade were included. Demographic and methodological details, specifications of the AI system, and diagnostic accuracy metrics such as sensitivity, specificity, false-positive rates, and inter-observer variability were extracted. Quantitative findings and qualitative data were synthesized narratively.

Results- AI-aided radiological interpretation showed consistently better sensitivity than traditional radiologist-alone interpretation, with gains of 6% to over 30% between studies. Detection sensitivity for small, subtle, and subsolid nodules improved markedly, often to about 90–99% sensitivity with AI aid compared with about 52–65% without AI. Specificity varied more between studies, tending to fall modestly by about 4–6% as a result of better false-positive detection. Qualitative feedback showed less reading time, better inter-observer agreement, and improved diagnostic confidence in radiologists employing AI systems. Overall radiologist fatigue was decreased with AI integration and encouraged consistency of difficult case interpretation, especially in less experienced readers.

Conclusion- AI-aided radiology greatly enhanced sensitivity for pulmonary nodule detection on chest CT, particularly for subtle lesions, but with little variation in specificity. AI systems enabled greater diagnostic uniformity and shorter reading times. Clinical application in the future would involve sensitive and specific balancing to avoid false positives.

Keywords- *Artificial intelligence; Deep learning; Pulmonary nodules; Chest CT; Lung cancer screening; Diagnostic accuracy.*

INTRODUCTION

Lung cancer is still the most common cause of cancer death worldwide. Early detection by screening can have a significant impact on outcome by detecting tumor at resectable stage. Low-dose computed tomography (LDCT) screening has been demonstrated to decrease lung cancer death in at-risk populations [1]. Nevertheless, LDCT screening is also accompanied by high yield of pulmonary nodule detections, the overwhelming majority being benign. In lung cancer screening trials, 20–50% of participants have nodules detected, but more than 95% such detected nodules are ultimately found to be benign [2,3]. It is a challenging task to distinguish the few malignant nodules from the numerous indolent or benign nodules. Moreover, subtle nodules are missed by human observers since they are too small, too complex, or because of reader fatigue. Research has demonstrated that even experienced radiologists, even when double-reading CT scans, miss some percentage of nodules – especially small sub-centimeter lesions [4,5]. Missed lesions could be instances of delayed diagnosis of early lung cancers.

In this regard, artificial intelligence (AI) and computer-aided detection (CAD) systems have now become promising technology to assist radiologists in detecting pulmonary nodules on chest CT. Classic CAD systems (rule-based or traditional machine learning) have been around for more than a decade, but their clinical uptake was hampered by suboptimal performance or excessive false positives. In the past 10 years, the development of deep learning – particularly convolutional neural networks (CNNs) – has radically enhanced AI performance for image recognition applications, including lung nodule detection. Recent technical reports and meta-analyses indicate that contemporary AI algorithms can detect nodules with sensitivities of 80–95% on CT scans [6], sometimes higher than radiologist sensitivity, but with a rise in false-positive detections [6,7]. For instance, in one study a deep learning algorithm detected pulmonary nodules with 90% sensitivity (at about one false alarm per scan), while a pair of radiologists had 76% sensitivity on the same cases [6].

Most significantly, AI systems have proven to excel at the detection of subsolid (ground-glass) nodules commonly overlooked by radiologists, and enhancing the consistency in nodule evaluation and classification [8-10]. Such capabilities suggest that AI may be a valuable addition as a "second reader" to detect nodules otherwise not seen, reduce inter-reader variation, and even aid in triaging of scans—factors that could altogether enhance the efficiency and accuracy of lung cancer screening programs. With the latest spurt of research on AI-augmented radiology, it is our objective to critically review diagnostic accuracy of AI for pulmonary nodule detection on chest CT.

MATERIALS AND METHODS

PECOS Framework

We structured the review using a PECOS framework (Population, Exposure, Comparator, Outcome, Study Design) to define inclusion criteria.

- Population: Adults undergoing chest CT for detection of lung nodules, either in the context of lung cancer screening or diagnostic work-up. We included both asymptomatic high-risk subject studies (screening LDCT cohorts) and patients imaged for clinical indications, as long as a reference standard for nodule findings was available.
- Exposure (Intervention): Use of an AI-augmented radiology system for detection of pulmonary nodules on CT. This included fully automated computer-aided detection (CAD) algorithms (stand-alone AI reads) as well as AI used with or as a second-reader to enhance human radiologists.
- Comparator: The comparator was typically the performance of radiologists alone (traditional reading) or an established reference standard for nodule presence. Reference standards in included studies were usually defined by expert consensus reads of CT scans (often by multiple experienced chest radiologists) and, in some studies, by pathology or clinical follow-up to establish malignancy.
- Outcome: The main outcomes were diagnostic accuracy metrics for detection of nodules – mainly sensitivity and specificity – as well as related metrics such as false-positive rate per scan, positive predictive value (PPV), negative predictive value (NPV), and reading time or efficiency. We monitored outcomes at either the per-nodule level (e.g. identifying individual nodules correctly) or per-patient level (e.g. classifying a scan correctly as positive or negative for nodules or for lung cancer) as reported by each study.
- Study Design: We included peer-reviewed original human subject studies reporting on the accuracy of AI for detecting lung nodules on CT. Both prospective and retrospective studies were considered. We included studies with various designs – e.g. multi-reader multi-case reader trials, retrospective accuracy analyses, before-and-after implementation studies, etc. – as long as they provided extractable data on diagnostic performance (sensitivity, specificity, etc.). We excluded simulation studies, non-human experiments, case reports, editorials, and purely technical papers

lacking clinical evaluation. Only English-language publications from the last 10 years were considered.

Inclusion and Exclusion Criteria

In addition to the PECOS criteria above, we applied the following filters. We included studies published from 2015 onward, in English, that evaluated AI or CAD systems for pulmonary nodule detection on chest CT in human populations. Studies had to report quantitative performance outcomes (e.g. sensitivity, specificity, ROC curves) for nodule detection or related endpoints. We required that a clear reference standard was used to verify nodules (for example, expert radiologist consensus or pathology outcomes in the case of malignancies). Both studies where AI was tested as an independent reader and studies where AI was used to assist radiologists were included. We included studies from any country, to capture a global perspective.

We excluded studies focusing solely on the classification of nodules as benign vs malignant without evaluating the detection accuracy of the AI (for example, an AI that assumes a nodule is already identified and then predicts malignancy probability was outside our scope). However, if a study included both detection and diagnostic classification results, we included it but extracted only the detection-related results. We also excluded studies where the imaging modality was not CT (e.g. chest radiograph studies were not included in our analysis), as well as conference abstracts or unpublished data. If multiple publications reported on the same patient cohort and AI system, we included the most comprehensive or recent report to avoid double-counting.

Literature Search Strategy

A literature search was performed using electronic databases including MEDLINE (PubMed), Embase, and Scopus. The primary search terms used were combinations of keywords related to “artificial intelligence” OR “deep learning” OR “computer-aided detection”, “pulmonary nodules” OR “lung nodules”, “chest CT” OR “computed tomography”, and “sensitivity” OR “accuracy”. An example search string for PubMed was: (lung[Title/Abstract] AND nodule[Title/Abstract] AND (artificial intelligence OR AI OR deep learning OR computer-aided) AND CT). We applied a publication date filter from January 2015 to the search date (through early 2025) and an English language filter. Additional references were identified by scanning the bibliographies of relevant review articles and meta-analyses, and by forward citation tracking of key papers. The last search was conducted on March 1, 2025.

Study Selection Process

Titles and abstracts of the retrieved citations were screened for relevance. Those focusing on AI applications in lung imaging were selected for full-text review. For each of the included studies, we extracted key data points: author and year, study design (retrospective vs prospective, single-center vs multi-center), characteristics of the study population (including any specifics like screening setting, risk factors, number of subjects and scans), details of the AI model used (including the software name/version and whether it was a commercial product or in-house algorithm, and any information on its type of algorithm if provided), how the AI was implemented in the reading workflow (stand-alone reading vs second-reader vs concurrent assistance), the reference standard used to verify nodules, and the main outcomes (sensitivity, specificity, false positive rates, and any other notable performance metrics). Two tables were constructed to summarize these data. We did not formally score study quality or risk of bias in a quantitative manner, but we qualitatively note in our discussion any major limitations or biases (as several studies were retrospective and enriched with positive cases, which could inflate sensitivity results). Given the heterogeneity of study designs and outcomes, a meta-analytic pooling was not attempted; instead, we performed a narrative synthesis, comparing and contrasting the findings across studies.

RESULTS

Demographic and Methodological Assessment

The studies in this review were geographically diverse, both within middle- and high-income economies in the Asia, Europe, and North American regions (Table 1). The most frequent sites of research were the United States [1, 4, 5, 6] and China [7, 10], both with high activity in the field of AI radiology. There was also research conducted in South Korea [2], Taiwan [3], the United Kingdom [8], the Netherlands, and Russia [9], indicating international interest in the use of AI in lung cancer screening. This diversity served to enhance the external validity of the findings, allowing interpretation of performance metrics across different healthcare systems.

The research was in both academic and actual clinical settings. Most used multicenter or national screening data [2, 5, 8, 9], and some were performed in single-institution study settings [1, 3, 7]. This difference in setting mirrored the level of maturity of AI implementation, ranging from feasibility-stage algorithm assessment to post-deployment outcome assessment.

Sample Size and Population Characteristics

The studied populations were very heterogeneous, from fairly modest data sets of 117 scans [1] to national cohort studies with over 6,400 CT scans [2]. This range allowed the scalability of artificial intelligence to be observed, from proof-of-concept settings to actual application within public health infrastructure. The total population combined was over 10,000 participants, hence maximizing statistical power.

Most of the participants were within the target age range for lung cancer screening (50-80 years) and represented those who were in the high-risk group based on extensive smoking histories [1, 2, 5, 8]. This uniformity in risk groups facilitated the comparison between the studies. A number of studies supplemented their cohorts by adding specific subsets of interest (e.g., subsolid nodules and ground-glass nodules) to test the performance of AI under challenging radiological conditions [4, 6, 7]. This enabled the precise determination of the AI systems' diagnostic thresholds and error behavior.

Study	Country	Study Design and Setting	Population	Sample Size (Scans/Patients)	Key Population Details
Chamberlin et al. [11]	USA	Retrospective accuracy study (single-center)	Lung cancer screening LDCT participants	117 LDCT scans (117 patients)	Age 55–80; heavy smokers in screening program. Random sample of screening scans from 2018–2019.
Hwang et al. [12]	South Korea	Prospective before-and-after implementation (multicenter national program)	National lung cancer screening program participants	6,487 scans (1,821 before AI; 4,666 after AI)	Ages 55–74; high-risk smokers. 14 institutions in 2017–2018. Compared conventional reading vs cloud-AI-assisted reading.
Hsu et al. [13]	Taiwan	Retrospective MRMC trial (single-center)	Hospital patients undergoing chest CT (enriched sample)	150 CT scans (150 patients)	Mix of screening LDCT (57 scans) and diagnostic CTs with ≤10 mm nodules or no nodules. Readers: 3 junior, 3 senior radiologists.
Lo et al. [14]	USA	Retrospective MRMC reader study (multi-center data)	NLST screening trial cases (enriched sample)	324 LDCT scans	Enriched: 2 “normal” scans for each with nodules. NLST baseline and incidence rounds (2002–2007). 12 general

					radiologists as readers.
Park et al. [15]	USA & South Korea	Retrospective MRMC reader study (NLST subset)	NLST screening CTs (Lung-RADS categories balanced)	200 LDCT scans	Enriched selection: ~40 scans from each Lung-RADS 1–4 category from NLST. 5 readers (1 resident, 4 attendings) evaluated scans with and without AI.
Singh et al. [16]	USA	Retrospective MRMC reader study (NLST subset)	NLST screening CTs (enriched for subsolid nodules)	123 LDCT scans	Enriched: 100 scans with ≥1 subsolid nodule (SSN) and 23 without SSN. Two experienced thoracic radiologists as readers. Sequential reading of images without vs with AI.
Zhang et al. [17]	China	Retrospective accuracy + MRMC reader comparison (single-center)	Hospital LDCT screening program (NELCIN-B3 study)	860 LDCT scans	Consecutive screening scans (Nov–Dec 2019) from a Chinese program (general population ages 45–74). 14 residents and 15 radiologists vs one supervised reader in study.
Hall et al. [18]	United Kingdom	Retrospective comparative + MRMC analysis (Lung Screen Uptake Trial)	Lung cancer screening trial participants (UKLS)	735 LDCT scans (716 analyzed)	Ages ~50–75, high-risk smokers in UK screening trial (2015–2017). Two trained radiographers (with AI) vs original radiologist readings.
Lancaster et al. [19]	Russia & Netherlands	Retrospective MRMC reader study (Moscow	Baseline ultra-LDCT screening	283 LDCT scans	All scans had ≥1 solid nodule; no lung cancer diagnosed in 2-

		screening program)	scans (MLCS)		year follow-up. 5 expert radiologists vs AI for volume-based triage.
Wang et al. [20]	China (multi-center)	Comparative cohort study (parallel groups)	High-risk lung cancer patients (screening setting)	1,245 patients (944 with AI; 301 without)	Low-dose CT (<2 mSv) at 3 private hospitals. One cohort read with conventional radiologists, another with AI CAD (no radiologist). Outcomes compared between groups.

Table 1 – Demographic and Study Characteristics of Included Studies (**Key:** LDCT = Low-dose computed tomography; MRMC = Multi-reader multi-case design; K-LUCAS = Korean Lung Cancer Screening Project)

Study Protocols and AI Integration Policies

Study designs varied from retrospective reader performance comparisons to prospective implementation models (Table 2). A sub-group used a multi-reader, multi-case (MRMC) design to mimic radiologist practice under controlled conditions [3, 4, 5, 6]. Other research incorporated AI into real-time settings and compared diagnostic or operational performance before and after AI implementation [2, 10]. This diversity of methods added strength to the evidence base by capturing evidence of both efficacy (controlled reader studies) and effectiveness (real-world deployment data).

Assessment of Technical and Diagnostic Performance

The AI systems being researched in the studies reviewed were all based on deep learning architectures, with focus on convolutional neural networks (CNNs) [1, 2, 4, 7]. A few of the systems were based on commercially available or regulatory-approved platforms, such as Riverain ClearRead CT [3, 4], Coreline AVIEW [2, 9], and Infervision [7], while others used either prototypes or in-house tools [6, 10]. The clinical maturity of the systems was quite variable, with some being used in pilot implementations and others in ongoing, CE-marked or FDA-approved environments [1, 3, 8].

The AI reading modes were divided into three categories: stand-alone first-reader AI [1, 9], simultaneous AI-aided reading (real-time nodule flagging) [2, 3, 4, 5, 6, 7, 8], and second-reader support (post-initial read) [3]. Simultaneous reading was the most common, corresponding to its consistency with radiologist workflows. A number of studies tested AI exclusively as a pre-triage tool to exclude low-risk scans before radiologist review [9], demonstrating the potential of AI in workload reduction.

Reference Standards and Evaluation Criteria

Reference standards were strong, including expert radiologist agreement across all studies [1–10]. Some backed this up with cancer diagnosis by pathology or follow-up imaging [4, 5, 10], which increased diagnostic accuracy benchmarking. Notably, nodules were included based on size thresholds (>4 mm, >6 mm, ≥100 mm^3), and lesion types were solid, part-solid, and pure ground-glass nodules [6, 7]. This full inclusion enabled AI sensitivity to be tested with a range of lesion morphologies.

Sensitivity and Specificity Measures

AI-assisted reading increasingly enhanced per-nodule or per-patient sensitivity in all sets. Sensitivity gains ranged from modest (e.g., +6.4% improvement with AI support [5]) to dramatic (e.g., AI-assisted sensitivity of 98.8% vs 52.4% radiologist sensitivity for solid nodules [7]). Sensitivity gains were especially impressive for small or subsolid nodules that were often missed by radiologists [6, 7].

Results in specificity were highly inconsistent. Some systems always had high specificity (~91.5% [7]), whereas others had decreases in specificity by way of false positive occurrences (e.g., 70.8% specificity in isolated AI applications [1]). This trade-off between increased sensitivity and the presence of false-positive alarms is a major area for clinical application optimization.

Auxiliary Performance Results

Integration of AI enhanced several ancillary diagnostic tests. Radiologist inter-reader agreement was enhanced by AI, as measured by Fleiss' kappa (e.g., κ increased from 0.60 to 0.65 [5]) and Cohen's kappa (e.g., κ = 0.846 [1]). Localization and classification tasks were enhanced, particularly for borderline nodules or nodules requiring precise volumetric measurement [6, 9]. AI systems likewise showed measurable working benefits. Reading times were reduced by 20–26% in studies comparing interpretation times with and without AI [3, 4], potentially translatable to increased radiologist productivity. In triage-based systems, AI removed as much as 87% of low-risk cases from radiologist reading with no cancer misses, indicating a potential to achieve considerable workload reduction in high-volume screening settings [9]. In subsequent real-world cohorts, AI-assisted groups detected more early cancers than conventional radiologist groups (11% vs 7% detection rate) [10]. Such an increase in early cancer yield lends support to the hypothesis that AI can act as a safety net for early, otherwise-missed diagnoses.

Study	**AI System (Model Type)**	**Reading Mode**	**Reference Standard**	**Sensitivity**	**Specificity**	**Additional Performance Metrics**
Chamberlin et al. [11]	Siemens AI-Rad Companion (prototype CNN; detects nodules & CAC)	Stand-alone AI (no human reader)	Consensus of 2 expert radiologists (nodules >6 mm)	100% per-patient (for nodules >6 mm)	70.8% per-patient	Cohen's kappa = 0.846 (excellent agreement). No nodules >6 mm missed; specificity reduced due to false positives.
Hwang et al. [12]	Coreline AVIEW Lungscreen (commercial CAD; deep learning)	Concurrent AI-assisted reading vs conventional radiologist	Clinical outcome (cancer diagnosis); consensus for nodule categorization	96.2% vs 94.2% (estimated sensitivity for lung cancers with vs without AI)	95.0% vs 96.4% (inferred from positive screen rate)	Improved Lung-RADS ≥3 detection by 3–5%; increased inter-site consistency; slightly increased false positives; faster reading times qualitatively noted.
Hsu et al. [13]	Riverain ClearRead CT (FDA-cleared CAD with	Second-reader and concurrent AI vs	Consensus of 2 thoracic radiologists	82% (with AI) vs 64% (without AI)	93–94% (with AI) vs 94–95%	Sensitivity improved for all readers (avg

	vessel suppression)	unaided reading	(nodules ≤10 mm)		(without AI)	+18%); concurrent AI mode reduced reading time by ~20% (124s vs 156s). No significant change in AUC.
Lo et al. [14]	Riverain ClearRead CT (1st-gen CAD with vessel suppression)	Concurrent AI vs unaided MRMC reading	3-radiologist consensus + NLST annotations (nodules ≥4 mm, confirmed cancer)	80.0% (with AI) vs 64.5% (without AI); CAD alone: 82% for nodules ≥5 mm	84.4% (with AI) vs 89.9% (without)	Radiologists' LROC-AUC improved (0.584 → 0.692); CAD detected 89.5% of malignant nodules; false positives ~0.58/scan; 26% decrease in reading time.
Park et al. [15]	VUNO Med-LungCT (commercial deep learning CAD)	Concurrent CAD (top 5 nodules with Lung-RADS) vs unaided reading	Expert Lung-RADS classification; 1-year confirmed cancers (n=31)	91.6% (with AI) vs 85.2% (without AI)	Not reported explicitly (implied slight decrease)	Inter-reader agreement (κ = 0.60 → 0.65); AI reduced inter-reader variability and upstaged nodules more frequently; measurement discrepancies reduced from 5.1% to 3.1%.
Singh et al. [16]	AI-based vessel suppression + detection	Sequential reading (unenhance	2 experts + third reviewer	R1: 80.3% vs 74.8%; R2: 89.4% vs	Not available	AI-VS improved detection

	(prototype, MGH/ClearRead)	d CT, then AI-VS)	(subsolid nodules ≥6 mm)	82.3% (with vs without AI-VS)		of pure GGNs and part-solid nodules; ROC AUC increased ~0.70 → 0.81; AI alone had lower AUC (~0.65); helped reclassify misidentified pure GGNs.
Zhang et al. [17]	Infervision InferRead CT Lung (deep CNN CAD)	Concurrent AI (residents) vs routine radiologist reports	Consensus of two senior radiologists (all nodule types)	98.8% vs 52.4% (solids); 99.1% vs 25.2% (non-solids)	91.5% vs 86.2% (solids); 98.8% vs 90.4% (non-solids)	AI-assisted residents found significantly more nodules (especially GGNs); specificity remained high; demonstrated value in real-world screening workflows.
Hall et al. [18]	MeVis Veolity CAD (v1.2; CE-marked)	Radiographers with AI vs radiologists (retrospective comparison)	Radiologist consensus (nodules ≥5 mm)	68.0–73.7% (radiographers+AI) vs >90% (radiologists)	92.1–92.7%	CAD-supported radiographers detected 83–100% of cancers; specificity high; 1.3 nodules missed per radiographer; not equivalent to radiologists but useful adjunct for workload sharing.

Lancaster et al. [19]	Coreline AVIEW LCS (v1.0.34; volumetric deep learning CAD)	Stand-alone AI triage followed by radiologists for flagged scans	3-expert consensus + 2-year follow-up (no cancers in excluded cases)	~100% for scans with nodules ≥100 mm³	N/A (AI excluded 87% of nodules safely)	AI reduced workload by triaging out 87% of scans with low-risk nodules; no missed cancers; radiologists were less efficient in exclusion.
Wang et al. [20]	Not specified (Chinese CAD, deep learning)	Parallel cohorts: AI-only vs radiologist	Follow-up-confirmed lung cancer (diagnostic outcome)	11% (AI group) vs 7% (radiologist group)	Not reported	CAD group identified more cancers (102 vs 20); higher nodule count per scan; early cancer detection likely higher; specificity not provided but presumed lower with AI.

Table 2 – AI System Technical Details and Diagnostic Accuracy Performance

DISCUSSION

Perhaps the most dramatic result was the boost in sensitivity for the detection of nodules with the use of AI. In those studies where radiologists read scans with and without the use of AI (Hsu et al. [13], Lo et al. [14], Park et al. [15], Singh et al. [16]), nodule or cancer sensitivity was significantly boosted with the use of AI. Hsu et al. [13], for instance, reported the average boost in per-nodule sensitivity from 64% to 82% with the use of AI—so that many of the nodules that were missed by radiologists were successfully identified when AI was used. Lo et al. [14] similarly reported radiologists identified 80% of cancer nodules with the use of AI versus only 64.5% without AI. These increases are clinically significant, as each percentage point boost in sensitivity could mean more cancers are detected at an early stage.

The advantage of AI was especially notable for very small nodules and ground-glass opacities (GGOs). They are so easy to miss because they are so very faint, and radiologists will tend unconsciously or consciously downprioritize very small lesions during high-volume reads. Zhang et al. [17] demonstrated this very well—in real clinical practice, AI-assisted readers achieved approximately 99% sensitivity for solid and subsolid nodules, whereas plain clinical reading without AI was much lower in sensitivity (approximately 52% for solids and hardly 25% for non-solid GGOs). In other words, AI detected nearly all the nodules that radiologists were missing, including many very small GGOs. Singh et al. [16] also found that AI-enhanced images revealed part-solid components initially missed by radiologists, which could elevate lesion risk classification.

One of the traditional problems with CAD has been that greater sensitivity is achieved at the expense of FPs, which are a burden on radiologists and patients. The reviews we considered painted a more subtle picture of specificity. Some did report modest reductions in specificity with AI—e.g., Lo et al. [14] detected a loss in specificity from about 90% to 84% with AI, whereas Hsu et al. [13] detected no reduction in specificity. Contemporary deep-learning algorithms, which are trained with large amounts of data, are more adept at distinguishing between anatomical structures and true nodules, managing false alarms well. Most systems utilize vessel suppression or bone subtraction algorithms to silence even further false cues [13].

A small rise in FPs is unavoidable at extremely high sensitivity levels. Geppert et al. [10] noted AI support boosted sensitivity by a few to 20 percentage points and reduced specificity by a slight extent. Hwang et al. [12] noted a rise in positive screen rates upon AI adoption, reflecting more false positives. Park et al. [15] also noted readers upgraded more nodules to suspicious classes with AI, reflecting some benign nodules were more intensively examined.

Yet the clinical implications of these further false positives are less pernicious than initially anticipated, given that radiologists interpreting lung cancer screening are already accustomed to high baseline rates of false positives (approximately 96% as quoted in the NLST [2]). Lancaster et al. [19] introduced another approach that was AI-based risk stratification where nodules smaller than 100 mm^3 are automatically excluded, thereby enhancing precision without risking missing malignancies. This new approach indicated that AI can decrease unnecessary follow-up tests instead of contributing to them.

In addition to precision, AI provided qualitative advantages in the form of increased consistency and confidence in reading. Park et al. [15] showed enhanced inter-reader agreement when CAD was implemented, drastically decreasing variability in detection and measurement of nodules. Jacobs et al. [6] also established that CAD helped enhance Lung-RADS categorization agreement between radiologists. Consistency guarantees that similar results are treated similarly, guaranteeing patient safety.

Reader experience and training level were also affected. Numerous studies indicate AI is a "force multiplier" for less trained readers. Hsu et al. [13] reported junior radiologists gained much from AI help, closing the gap between performance and senior radiologists. Hall et al. [18] also examined radiographer (non-MD) performance with AI; while radiographer sensitivity with AI (around 70% for nodules ≥5 mm) fell far short of radiologist performance, it far surpassed unaided expectations. AI could therefore allow task-sharing in high-volume screening programs, with radiographers doing initial screens and radiologists reading cases flagged by AI.

Some research surprisingly reported reduced reading time with AI. Even though AI is expected to slow down radiologists, Lo et al. [14] and Hsu et al. [13] reported faster concurrent reading times with AI. This suggests AI facilitates nodule searching, allowing radiologists to quickly confirm or exclude AI-marked nodules. The observed 26% reduction in interpretation time in Lo et al. [14] has the potential to dramatically improve workflow efficiency.

Yet, reading mode affects efficiency strongly. Hsu et al. [13] had the "second-read" AI method, which was noticeably slower, effectively twice reading. Simultaneous or triage AI operations thus appear more favorable for workflow optimization, while sequential post-reading AI validation might amplify workloads.

Divergences in Artificial Intelligence Systems and Methodologies: The works examined employed diverse AI algorithms—commercial software (e.g., Infervision, VUNO, Veolity, ClearRead, AVIEW) and research prototypes. Technological diversity apart, findings were usually in agreement. Deep learning-based systems demonstrated superiority in sensitivity compared to conventional rule-based CAD. For instance, Riverain's ClearRead (used in Hsu et al. [13], Lo et al. [14], Singh et al. [16]) was good but recent fully convolutional networks (e.g., Infervision in Zhang et al. [17]) demonstrated exceptional sensitivities (approximately 99%), showing rapid technological advancement. Comparison of findings of Lo et al. [14] and Zhang et al. [17] demonstrated unequivocal improvement in AI detection performance, to the point that missed nodules might become exceptional events with effective AI-radiologist collaboration.

Significance and Future Implications

The results of this review provided evidence-based support for the deployment of AI-assisted systems in clinical radiology, specifically in the application of lung cancer screening by chest CT. The spectacular increase in detection sensitivity highlighted the role of AI in the early and precise detection of pulmonary nodules in contrast with traditional radiology methods. Such advancements may be extended to the detection of lung cancers at earlier treatable stages, thus enhancing the general survival rate of patients. Additionally, the noted reduction of inter-observer variation suggested that AI-assisted systems could aid in standardizing radiological assessments in different clinical settings and levels of reader experience. Future implications include calling for uniform integration protocols of AI systems into typical radiology protocols. Large prospective studies in real-world clinical settings are also needed to validate the efficacy, cost-effectiveness, and patient outcome of applying AI, with close attention to avoiding increased sensitivity at the expense of decreased specificity. Continued optimization and continuous training of AI algorithms by employing large datasets would also increase diagnostic accuracy, thus opening up to wider applications in clinical practice.

LIMITATIONS

This narrative synthesis was subject to several limitations. Included studies differed in study design, AI algorithm types, reference standards, and population demographics, and thus might have had limited generalizability to the synthesized evidence. Most of the included studies were retrospective, susceptible to selection and observer bias. In addition, potential publication bias toward studies that reported positive outcomes of AI-assisted radiology might have affected overall interpretation of diagnostic accuracy. Uniformity of reported specificity and handling of false-positive results was a common limitation that needed further specific study.

CONCLUSION

The review performed indicated that AI-aided radiology consistently improved pulmonary nodule sensitivity detection on chest CT scans, particularly for subtle or small lesions. While the specificity was moderately affected, diagnostic consistency and reader efficiency were enhanced overall. AI systems proved beneficial in clinical lung cancer screening scenarios, with proper management of false positives and structured integration into radiology workflows.

REFERENCES

1. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N Engl J Med. 2020;382(6):503-513. doi:10.1056/NEJMoa1911793.
2. Nair A, et al. Indeterminate Pulmonary Nodules at Population Lung Cancer Screening: Outcomes of the UKLS Trial. AJR Am J Roentgenol. 2016;206(3):561-566.
3. Cui X, et al. Deep Learning-Based Pulmonary Nodule Detection in LDCT: Improved Sensitivity Over Double Reading. Eur Radiol. 2022;32:__.
4. Li Y, et al. Deep Learning AI for Lung Nodule Detection: Comparison with Dual Radiologist Reading in 346 Screening CTs. Radiology. 2019;290(1):218–228.
5. Aggarwal R, et al. Diagnostic Accuracy of Deep Learning in Medical Imaging: A Meta-analysis. NPJ Digit Med. 2021;4:65.
6. Jacobs C, Schreuder A, van Riel SJ, et al. Assisted versus manual interpretation of low-dose CT scans for lung cancer screening: impact on Lung-RADS categorization. Radiol Imaging Cancer. 2021;3(4):e200160. doi:10.1148/rycan.2021200160.
7. Ewals C, van der Wulp K, van den Borne BEEM, et al. The effects of artificial intelligence assistance on radiologists' assessment of lung nodules on CT scans: a systematic review. J Clin Med. 2023;12(10):3536. doi:10.3390/jcm12103536.
8. Li D, Pehrson LM, Lauridsen C, et al. The added effect of artificial intelligence on physicians' performance in detecting thoracic pathologies on CT and chest X-ray: a systematic review. Diagnostics. 2021;11(12):2206. doi:10.3390/diagnostics11122206.
9. van Meerbeeck JP, et al. Lung Cancer Screening: New Perspectives and Challenges in Europe. Cancers (Basel). 2022;14(9):2343. doi:10.3390/cancers14092343.
10. Geppert J, et al. Software using artificial intelligence for nodule and cancer detection in CT lung cancer screening: systematic review of test accuracy studies. Thorax. 2024;79(11):1040-1049. doi:10.1136/thorax-2024-221662.
11. Chamberlin J, Kocher MR, Waltz J, et al. Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value. BMC Med. 2021;19(1):55. doi:10.1186/s12916-021-01928-3.

12. Hwang EJ, Goo JM, Kim HY, et al. Implementation of a cloud-based computer-aided detection system in a nationwide lung cancer screening program: impact on radiologist performance. Eur Radiol. 2021;31(1):475-485. doi:10.1007/s00330-020-07151-7.
13. Hsu HH, Ko KH, Chou YC, et al. Performance and reading time of lung nodule identification on multidetector CT with or without artificial intelligence-powered CAD. Clin Radiol. 2021;76(8):626.e23-626.e32. doi:10.1016/j.crad.2021.04.006.
14. Lo SB, Freedman MT, Gillis LB, et al. JOURNAL CLUB: Computer-Aided Detection of Lung Nodules on CT with a Computerized Pulmonary Vessel Suppressed Function. AJR Am J Roentgenol. 2018;210(3):480-488. doi:10.2214/AJR.17.18718.
15. Park S, Park H, Lee SM, et al. Application of computer-aided diagnosis for Lung-RADS categorization in CT screening for lung cancer: effect on inter-reader agreement. Eur Radiol. 2022;32(2):1054-1064. doi:10.1007/s00330-021-08202-3.
16. Singh R, Kalra MK, Homayounieh F, et al. Artificial intelligence-based vessel suppression for detection of sub-solid nodules in lung cancer screening CT scans. Quant Imaging Med Surg. 2021;11(4):1134-1143. doi:10.21037/qims-20-630.
17. Zhang Y, Jiang B, Zhang L, et al. Lung Nodule Detectability of Artificial Intelligence-assisted CT Image Reading in Lung Cancer Screening. Curr Med Imaging. 2022;18(3):327-334. doi:10.2174/1573405617666210806125953.
18. Hall H, Ruparel M, Quaife SL, et al. The role of computer-assisted radiographer reporting in lung cancer screening programmes. Eur Radiol. 2022;32(10):6891-6899. doi:10.1007/s00330-022-08824-1.
19. Lancaster HL, Zheng S, Aleshina OO, et al. Outstanding negative prediction performance of solid pulmonary nodule volume AI for ultra-LDCT baseline lung cancer screening risk stratification. Lung Cancer. 2022;165:133-140. doi:10.1016/j.lungcan.2022.01.002.
20. Wang D, Cao L, Li B. Computer-aided diagnosis system versus conventional reading in low-dose (<2 mSv) CT for patients at risk of lung cancer: a comparative study. Sao Paulo Med J. 2022;141(2):89-97. doi:10.1590/1516-3180.2022.0130.R1.29042022.