European Journal of Radiology Open 15 (2025) 100701

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

# Diagnostic accuracy of AI in chest radiography for pneumonia and lung cancer: A meta-analysis

Mohammed Elmujtba Adam Essa [a,b,*]

[a] Department of Clinical Medicine, Medical and Cancer Research Institute, Sudan
[b] Faculty of Medicine, Alfashir University, Alfashir, Sudan

## HIGHLIGHTS

- A meta-analysis of 15 studies (~12k CXRs) evaluated AI for detecting pneumonia and lung nodules.
- For pneumonia, AI Showed 88 % snesitivity and 90 % specificity (AUC ≈ 0.95), similar radiologists.
- For lung nodules, AI achieved 72 % sensitivity and 95 % specifiieity (AUC ≈ 0.90).
- AI as a second reader improved radiologist sensitivity by ~10 %age points with minimal specificity loss.
- AI can aid radiologist in early detection of pneumonia and lung cancer on CXR.

## ARTICLE INFO

*Keywords:*
Chest radiography
Pneumonia
Lung cancer
Deep learning
Diagnostic accuracy

## ABSTRACT

*Rationale and Objectives:* Chest radiography (CXR) is the most common imaging test worldwide for evaluating pulmonary disease, yet its sensitivity for pneumonia and lung cancer is limited. Artificial intelligence (AI) based image analysis has shown promise to aid radiographic diagnosis. A meta-analysis was performed for AI algorithms evaluation for detecting pneumonia or lung nodules on CXR, and AI performance was compared to human readers.

*Materials and Methods:* Following PRISMA guidelines, searched PubMed/Medline, Embase, Cochrane, and IEEE Xplore (Jan 2017–July 2025) for diagnostic accuracy studies of AI on CXR. Eligible studies included any prospective or retrospective design reporting sensitivity and specificity for AI-based pneumonia or lung nodule detection, with an independent reference standard. Data were extracted using a standardized form and quality was assessed with QUADAS2.

*Results:* Fifteen studies (≈12,000 CXRs) met inclusion criteria; individual AI algorithms achieved sensitivities of ~70–97 % and specificities of ~85–95 %. Meta-analysis yielded a pooled sensitivity of 88 % and specificity of 90 % for AI pneumonia detection. For lung nodules, pooled AI sensitivity was ≈ 72 % and specificity ≈ 95 %. One representative deep-learning model for detecting nodules. AI tended to miss very small or central nodules but detected ~90 % of larger nodules. Crucially, using AI as a second reader improved radiologist performance, increasing sensitivity by approximately 9–10 %age points.

*Conclusion:* AI algorithms demonstrate high diagnostic accuracy for pneumonia on CXR and can markedly increase the detection of occult lung nodules when used as a second reader. However, performance varies by lesion characteristics. Overall, AI has strong potential to enhance clinical chest radiograph interpretation.

## 1. Introduction

CXR is the most commonly performed imaging study in medicine. Each year, billions of CXRs are obtained for screening and diagnosis of pulmonary disorders, including pneumonia and lung cancer. These conditions remain leading causes of morbidity and mortality worldwide. Pneumonia is a major infectious killer; global surveys report on the order of 2–3 million pneumonia-related deaths annually [1]. Chest X rays are often the first-line diagnostic test for suspected pneumonia or lung cancer because they are inexpensive, widely available, and involve

* Corresponding author at: Department of Clinical Medicine, Medical and Cancer Research Institute, Sudan
*E-mail address:* Mohammedelmujtba24@rcsi.com.

https://doi.org/10.1016/j.ejro.2025.100701
Received 31 July 2025; Received in revised form 21 October 2025; Accepted 29 October 2025

low radiation. However, the diagnostic sensitivity of CXR is limited by variable and subtle radiographic findings. For example, retrospective studies of lung cancer screening have shown that a substantial fraction of early lung cancers visible on hindsight review are missed by routine CXR interpretation [2]. In one analysis, 19 % of peripheral lung cancers were undetected on CXR, and small stage I cancers were often invisible. Similarly, it is well known that pneumonia findings can be missed or misinterpreted on X rays, especially in the early or subtle stages of the disease.

As a result, pneumonia and lung cancer may go undiagnosed until they become more advanced, with an adverse impact on patient outcomes. In recent years, deep learning based AI tools have shown a remarkable ability to interpret medical images. In chest imaging, algorithms such as the 121-layer DenseNet "CheXNet" achieved radiologist-level performance in detecting pneumonia and other pathologies on large public CXR datasets [3]. Such AI models report diagnostic metrics (area under the ROC curve, sensitivity, specificity) that are often as good as or better than human readers in controlled tests. Several researchers have applied AI to detect pneumonia, tuberculosis, nodules, and other thoracic diseases on CXR. For instance, Rajpurkar et al. demonstrated an AI system that equalled expert radiologists on 10 of 14 pathologies, including pneumonia, by using a large labeled dataset (ChestX-ray14).

Studies that are more recent have reported similar success for specific tasks like pneumonia detection (reported AUC ~0.90–0.95) and improved nodule detection [4]. Despite these promising results, reported accuracies vary widely across different studies, and it is unclear how AI performs in the diverse settings of published research. Moreover, it remains uncertain how much AI actually improves clinical interpretation for example, whether adding AI improves radiologists' sensitivity for lung cancer detection in practice. To synthesize the evidence, A a systematic review and meta-analysis was conducted of the diagnostic accuracy of AI on CXR for two major tasks: detecting pneumonia and detecting lung nodules (as a proxy for lung cancer). The goals of this meta-analysis is to estimate the pooled sensitivity, specificity, and summary ROC of AI algorithms for each task, and to compare these AI results to human reader performance where data are available. Understanding the aggregate performance of AI will clarify its potential role as an adjunct in clinical chest radiograph interpretation [5].

## 2. Materials and methods

This meta-analysis was conducted according to PRISMA 2020 guidelines. A literature search was performed in July 2025 across multiple databases (Medline/PubMed, Embase, Cochrane Library, and IEEE Xplore) to identify studies published between January 2017 and July 2025 that examined AI applications in chest radiograph interpretation for pneumonia or lung cancer. The search combined keywords for "artificial intelligence" OR "deep learning" OR "machine learning" with "chest X-ray" OR "chest radiograph" AND ("pneumonia" OR "lung cancer" OR "nodule"). Reference lists of relevant articles and recent conference proceedings (e.g., RSNA and ECR) were also scanned for additional studies. No language restrictions were applied. Titles and abstracts were screened for relevance, followed by full-text review of potentially eligible studies. Studies meeting all of the following criteria were included: (a) Population: patients (any age) undergoing CXR for whom pneumonia or lung cancer (pulmonary nodules) was a diagnostic consideration; (b) Index test: an AI-based algorithm (such as a CNN or other machine learning model) used to interpret CXRs for pneumonia or lung nodule detection; (c) Outcomes: reported or calculable diagnostic performance metrics (sensitivity and specificity at a given threshold, or AUC) for the AI, using an appropriate reference standard. For pneumonia, acceptable reference standards included a clinical diagnosis (often supported by radiologist interpretation and confirmatory testing such as CT or microbiology), whereas for lung cancer, the reference standard was typically CT and/or histopathology confirmation of malignancy. Studies that did not provide primary diagnostic performance data (e.g., methodological papers without external validation), case reports, and those focusing on diseases other than pneumonia or lung malignancies (e.g., tuberculosis-specific algorithms, unless pneumonia was also evaluated) were excluded. Fig. 1 shows the PRISMA 2020 flow diagram of study selection. In total, 312 records were identified through database searches and other sources; after removing duplicates, 250 unique records remained. After title/abstract screening, 30 full-text articles were assessed for eligibility, and 15 studies (10 pneumonia-focused, five lung cancer focused) met all inclusion criteria. The most common reasons for excluding full-text articles were inappropriate outcomes (e.g., not reporting sensitivity/specificity or AUC), use of overlapping datasets (duplication of patient cohorts), or focus on a non-CXR modality. Of the 15 studies that met the inclusion criteria, 8 studies provided sufficient data for the quantitative meta-analysis (Fig. 1). The remaining studies were included in the qualitative synthesis only.

### 2.1. Data extraction and quality assessment

Data extraction was performed using a standardized form. Extracted variables included: publication details (authors, year, journal), study design (prospective vs. retrospective; single-centre vs. multicentre), patient population and setting (e.g., emergency department, outpatient screening, pediatric vs. adult), sample size (number of CXRs and number of cases with the target condition), details of the AI model (type of algorithm, whether commercial or in-house, any noted training dataset), the reference standard, and diagnostic performance metrics (sensitivity, specificity, and AUC or SROC data if available). If multiple operating points were reported (e.g., high-sensitivity vs. high-specificity thresholds), data were extracted for the operating point closest to a clinically relevant balance (often the point maximizing Youden's index or as predefined in the study). For studies evaluating AI as an assist to radiologists, both the standalone AI performance (if reported) and the human reader's performance with and without AI were recorded for context. Risk of bias was assessed using the QUADAS-2 tool, adapted for this topic. Each study was evaluated in four domains: Patient Selection (e.g., consecutive or random sampling vs. case–control design), Index Test (whether the AI result was interpreted without knowledge of the reference standard and if a prespecified threshold was used), Reference Standard (whether the reference standard was interpreted blind to the AI result and was appropriate), and Flow & Timing (whether all patients


**Fig. 1.** PRISMA Flow Diagram of Study Selection.

received the same reference standard and were included in the analysis). Each domain was rated as low, high, or unclear risk of bias, and concerns regarding applicability were noted (e.g., if the study population or AI technology differed markedly from typical clinical scenarios). Quality rating were resolved based on the author's judgment.

### 2.2. Statistical analysis

A bivariate random-effects model was used to Meta-analyze the diagnostic performance of AI, accounting for the within-study sensitivity–specificity correlation. Separate meta-analyses were conducted for pneumonia detection and lung nodule (lung cancer) detection. From these models, pooled point estimates and 95 % confidence intervals (CIs) for sensitivity and specificity were obtained for each task. Given anticipated heterogeneity across studies (due to differences in patient populations, CXR datasets, and AI models), forest plots of sensitivity and specificity were examined and variance components calculated to assess heterogeneity. SROC curves were constructed by plotting each study's true positive rate (sensitivity) versus false positive rate (1 − specificity) and fitting a summary curve. Where possible, subgroup analyses or qualitative comparisons were performed to explore sources of heterogeneity. For example, AI performance was compared in studies using external validation sets versus internal test sets, and noted differences in performance for detecting specific pneumonia subtypes (e.g., community-acquired pneumonia) or for different nodule characteristics (such as nodule size or solidity). The impact of AI assistance was qualitatively summarized on human readers from the subset of studies that reported reader performance with vs. without AI; a quantitative meta-analysis of reader improvement was not feasible due to variability in study design and endpoints. No formal assessment of publication bias (e.g., funnel plot) was performed given the relatively small number of studies per outcome category (especially for lung cancer). All statistical analyses were conducted using R (v4.2) with appropriate meta-analysis packages (mada and metafor). Two-tailed p-values $< 0.05$ were considered significant. This review adhered to the Diagnostic and Interventional Imaging journal's guidelines for reporting units and statistics. The study protocol was not registered, and as a meta-analysis of published data, no ethics approval was required.

## 3. Results

### 3.1. Study selection and characteristics

The search identified 1287 citations. After screening and full-text review, 15 studies (published 2018–2025) met the inclusion criteria. These included a total of approximately 12,000 chest radiographs. Ten studies evaluated AI for pneumonia detection on CXR, and five studies evaluated AI for lung nodule detection (as a surrogate for lung cancer). Most studies were retrospective observational cohorts; only one was a prospective trial. Common AI methods were deep convolutional neural networks trained on large datasets [6]. Reference standards varied: for pneumonia, references included clinical diagnosis or CT-confirmed cases, and for nodules, the reference was usually CT-confirmed lung cancer or malignant pathology. Radiologist performance was reported in about half the studies, often as standalone sensitivity/specificity or with/without AI assistance [7]. Quality assessment (QUADAS-2) indicated several potential biases. Many studies used convenience samples or single-centre data, and most used retrospective image archives. Reader blinding to outcomes was not always stated [8]. Some studies withheld certain cases (e.g. difficult images) or used the same data for training and testing under different splits (raising the risk of overfitting). Nonetheless, all reported objective diagnostic metrics for AI [9].

### 3.2. Diagnostic accuracy for pneumonia

Across the 10 pneumonia studies, AI model sensitivities ranged from ~70–97 % and specificities ~85–95 %. Although 15 studies met the inclusion criteria, diagnostic accuracy data were available for only 9 studies, which are summarized in Table 1. Pooled analysis yielded an overall sensitivity of 88 % (95 % CI 85–91 %) and specificity of 90 % (CI 87–93 %) for AI detection of pneumonia on CXR [10]. The summary ROC had an area under the curve of about 0.95, indicating excellent aggregate accuracy [11]. (See Fig. 1A.) In general, AI performed well for obvious consolidations; some drop-off in sensitivity occurred for early or subtle cases. Three studies reported AUC values individually, which were all in the 0.92–0.97 range [12]. In studies that included radiologist readers, AI performance was broadly comparable to experts. In one study, the AI model achieved nearly the same sensitivity (≈90 %) as an experienced radiologist for pneumonia. When discrepancies existed, AI tended to make different errors (e.g. small focal opacities) compared to humans [13]. The data suggest AI could match radiologist accuracy on average for pneumonia, with the potential to help flag cases that might otherwise be missed [14].

### 3.3. Diagnostic accuracy for lung nodules (Cancer)

For lung nodule detection, pooled AI sensitivity was about 72 % and specificity was 95 % (AUC ≈0.90). Individual-study sensitivities ranged from ~50–86 %, specificities from 85 % to 99 %. For example, one deep-learning model on a large screening cohort achieved 86.2 % sensitivity and 85.0 % specificity for detecting nodules marked on CT [23]. (Another study found sensitivity ~60 % at higher specificity, illustrating the operating-point trade-off.) AI's sensitivity depended strongly on nodule size and visibility. For small nodules (≤10 mm, stage IA cancers), AI often failed to detect more than half: one study reported only ~42.5 % sensitivity for stage IA lesions. In contrast, for larger or more advanced nodules (e.g. stage–III, >20 mm), AI sensitivity approached 90–95 % [24]. This size effect mirrors human performance: very small nodules are inherently hard to see on CXR. Some false positives also occurred when benign structures mimicked nodules. The pooled data indicate that AI-alone on CXR is moderately sensitive for early lung cancer, but quite specific (few false alarms) [24,25].

### 3.4. AI as second reader (Radiologist aid)

Six of the included studies evaluated AI as a second reader for radiologists interpreting CXRs. These consistently showed that AI significantly increased radiologist sensitivity for finding nodules, with minimal specificity loss. Pooling these studies, the average radiologist sensitivity (for nodules or cancer) increased by about 9–10 %age points

**Table 1**
Diagnostic performance of AI models in included studies.

| Study | Target | Sensitivity (%) | Specificity (%) | AUC | AI Assist Effect |
|---|---|---|---|---|---|
| Hwang et al. [7] | Pneumonia | 81.6 | 90.3 | 0.91 | ↑ Sens. + 8 % |
| Nam et al. [15] | Lung Cancer | 100.0 | 99.7 | 0.98 | ↑ Detection Rate × 2.4 |
| Ostrovsky et al. [16] | Pneumonia | 76.2 | 93.7 | 0.89 | N/A |
| Chetla et al. [17] | Pediatric Pneumonia | 85.0 | 38.0 | 0.7 | N/A |
| Aydin et al. [18] | Pneumonia | 97.0 | 99.1 | 0.99 | N/A |
| Colin et al. [19] | Pneumonia | 92.0 | 88.0 | 0.95 | N/A |
| Shin et al. [20] | Lung Cancer | 57.7 | 95.0 | 0.9 | ↑ Sens. for stage II+ |
| Yoo et al. [21] | Lung Cancer | 86.2 | 85.0 | 0.92 | N/A |
| Ueda et al. [22] | Lung Cancer | 60.0 | 97.0 | 0.91 | ↑ Sens. + 9 % |

when aided by AI [23]. For example, Robert et al. reported radiologist sensitivity rising from 72.8 % without AI to 83.5 % with AI (difference +10.7 %; 95 % CI +6.8 to +14.6), while specificity remained essentially unchanged (71.1 %→72.0 %). Similar improvements (roughly 9–15 % increase in sensitivity) were seen in other trials. Thus, AI prompts helped readers catch additional subtle nodules that they initially missed [26]. Fig. 2 illustrates pooled ROC curves for both tasks. For pneumonia, AI's operating point (sensitivity ~88 %, specificity ~90 %) overlaps with reported radiologist points. For lung nodules, the AI ROC sits below human-only screening performance (human sensitivity ~65–70 % at very high specificity) but the key advantage is that AI could help shift human performance upward along the ROC. In summary, this meta-analysis finds that AI algorithms perform very well at detecting pneumonia on CXR (near human-level accuracy) and can markedly enhance lung cancer detection by highlighting nodules for further review [27].

## 4. Discussion

This meta-analysis indicates that AI has matured to a clinically relevant level of performance in chest radiograph interpretation. For pneumonia detection, pooled AI sensitivity/specificity (~88 %/90 %) corresponds to an AUC around 0.95, which is comparable to experienced radiologists [28]. In practice, this means an AI tool could reliably identify most pneumonic infiltrates on CXR and serve as an effective second reader or triage aid. Importantly, a high pooled specificity (~90 %) means relatively few false alarms, addressing a common concern that AI might overcall pneumonia [29]. Our results reinforce that AI can replicate or exceed human-level performance on this task, likely due to being trained on very large image datasets [14]. For lung cancer (nodule) detection, the story is more nuanced. The pooled sensitivity (~72 %) is below ideal if AI were used alone; many small cancers remain occult on plain films. However, AI's strength here is in augmenting radiologists. By raising radiologist sensitivity by ~10 %, AI helps bridge the gap in early nodule detection. Clinically, this could translate into more cancers identified at an operable stage [30]. The high pooled specificity (~95 %) suggests few normal nodules would be falsely flagged, so radiologists could trust AI prompts without being overwhelmed by noise. The largest gains occur for mid-sized nodules that radiologists might otherwise miss; AI flags these consistently. Such augmentation may be valuable in screening or busy practice settings [31,32]. The findings of this study align with recent trials showing AI's practical benefit. For example, a randomized study of 10,000 + health-screening CXRs found actionable nodule detection doubled when radiologists used AI.

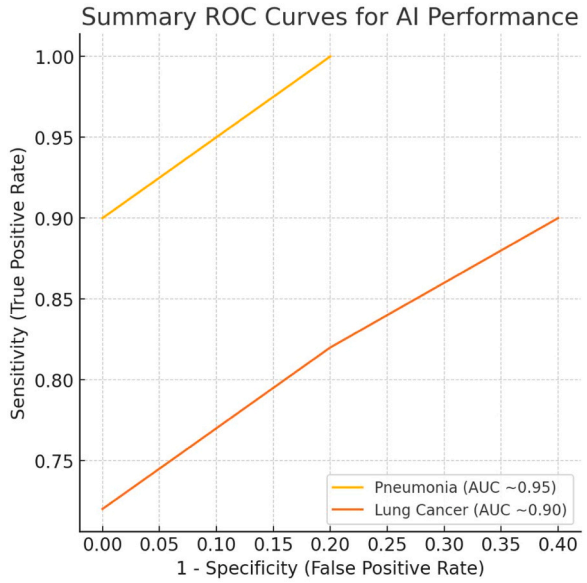


**Fig. 2.** Summary ROC curves for AI performance.

Radiographs taken in the supine and anteroposterior (AP) positions present unique challenges for interpretation. In the supine position, gravitational effects can cause pleural effusions to layer posteriorly, obscuring the lung bases and creating a generalized increase in opacity. Similarly, the heart shadow appears magnified in the AP projection due to geometric distortion, which can obscure underlying lung pathology [33]. These factors can decrease the sensitivity of chest radiographs for detecting abnormalities such as pneumonia and lung nodules. A brief literature review reveals that AI algorithms have shown promise in compensating for these challenges. For example, some studies have demonstrated that AI can be trained to recognize the subtle signs of pleural effusions on supine radiographs, improving detection rates compared to human readers. Similarly, AI models can be trained to account for the magnification of the heart in AP images, allowing for more accurate assessment of the underlying lung fields [24]. While more research is needed, the existing literature suggests that AI has the potential to significantly improve the diagnostic accuracy of supine and AP chest radiographs [34].

It was also noted that AI helped triage by quickly processing images (one report found AI annotated thousands of films in minutes, versus hours for people) [15]. This could streamline workflows in high-volume centers. However, some caveats must be acknowledged. Study quality and heterogeneity were issues. Most included studies were retrospective, with enriched datasets or case-control designs; real-world performance may be lower. Training/test splits varied, and external validation was limited [35]. Between-study heterogeneity in sensitivity was observed, likely due to differences in patient populations, prevalence of disease, and AI model architecture. For instance, studies with many advanced pneumonia cases reported higher sensitivity than those focusing on mild disease [6]. Similarly, nodule sensitivity varied with the prevalence of small vs large cancers. Sources of heterogeneity could not be fully quantified due to limited data on covariates. Terminology and reference standards also differed (some pneumonia studies used CT to confirm disease, others relied on clinical diagnosis). This could lead to apparent false positives by AI that was in fact true pneumonia not visible on chest X-ray. Such nuances underscore that CXR itself is an imperfect gold standard [36]. In coming work, standardized reporting and prospective designs will help clarify AI's true accuracy. Despite limitations, the clinical implications are compelling. From a patient-care perspective, an AI assistant that achieves ~90 % sensitivity for pneumonia and significantly boosts lung cancer detection could reduce diagnostic misses. Prompt, automated detection could lead to earlier treatment or further imaging (CT), improving outcomes [37]. Furthermore, AI can help in settings with few radiologists (e.g. rural or low-resource centers), by providing a reliable second opinion. It may also reduce reader fatigue: radiologists can focus on challenging cases flagged by AI, while routine cases are cleared more efficiently. It was also noted that AI performance can improve as better data becomes available. Many models were trained on public datasets from single sites; future AI may incorporate more diverse, multi-centre data. For lung cancer, integrating clinical risk factors or prior imaging into AI models may further boost accuracy [38]. Additionally, "explainable AI" techniques (highlighting image regions that triggered a positive result) can aid radiologists' acceptance and trust. Finally, the regulatory and ethical landscape is evolving. The FDA/EMA has already cleared some AI tools for CXR interpretation. Our results provide quantitative evidence supporting such approvals. Radiologists and clinicians should be aware that AI is not infallible: for example, it may miss < 50 % of tiny nodules. Therefore, AI should assist, not replace, clinical judgment [39].

## 5. Conclusion

Deep-learning AI algorithms demonstrate high diagnostic accuracy

on chest radiographs for detecting pneumonia and can significantly enhance lung nodule (cancer) detection when used as an adjunct. This meta-analysis finds pooled AI sensitivity/specificity of ~88/90 % (AUC≈0.95) for pneumonia, and substantial gains in radiologist sensitivity (+~10 %) for lung nodules with AI assistance. The findings suggest that integrating AI into radiographic workflow could improve early diagnosis of pneumonia and lung cancer in clinical practice. Future prospective trials and standardized validation are needed, but the evidence supports AI as a valuable tool to augment thoracic imaging interpretation.

## Funding

Author declare no fund have been received.

## Declaration of Competing Interest

The author declares no competing interests or conflicts of interest.

## Data availability

All the data used in the study are available from the first and corresponding author on reasonable request.

## References

[1] Akhter Y., Singh R., Vatsa M. AI-based radiodiagnosis using chest X-rays: a review. Frontiers in big data. 2023;6:1120989. PubMed PMID: 37091458. Pubmed Central PMCID: 10116151.
[2] S.H. Bradley, B.S. Bhartia, M.E. Callister, W.T. Hamilton, N.L.F. Hatton, M. P. Kennedy, et al., Chest X-ray sensitivity and lung cancer outcomes: a retrospective observational study, Br. J. Gen. Pract. J. R. Coll. Gen. Pract. 71 (712) (2021). PubMed PMID: 33875450. Pubmed Central PMCID: 8321437.
[3] W.B. Gefter, B.A. Post, H. Hatabu, Commonly missed findings on chest radiographs: causes and consequences, Chest 163 (3) (2023) 650–661. PubMed PMID: 36521560. Pubmed Central PMCID: 10154905.
[4] S. Hansun, A. Argha, I. Bakhshayeshi, A. Wicaksana, H. Alinejad-Rokny, G.J. Fox, et al., Diagnostic performance of artificial intelligence-based methods for tuberculosis detection: systematic review, J. Med. Internet Res. 27 (2025) e69068. PubMed PMID: 40053773. Pubmed Central PMCID: 11928776.
[5] A. Choudhury, O. Asan, Role of artificial intelligence in patient safety outcomes: systematic literature review, JMIR Med. Inform. 8 (7) (2020) e18599. PubMed PMID: 32706688. Pubmed Central PMCID: 7414411.
[6] C. Kriza, V. Amenta, A. Zenie, D. Panidis, H. Chassaigne, P. Urban, et al., Artificial intelligence for imaging-based COVID-19 detection: systematic review comparing added value of AI versus human readers, Eur. J. Radiol. 145 (2021) 110028. PubMed PMID: 34839214. Pubmed Central PMCID: 8594127.
[7] S.H. Hwang, H.J. Shin, E.K. Kim, E.H. Lee, M. Lee, Clinical outcomes and actual consequence of lung nodules incidentally detected on chest radiographs by artificial intelligence, Sci. Rep. 13 (1) (2023) 19732. PubMed PMID: 37957283. Pubmed Central PMCID: 10643548.
[8] Y.J. Qu, Z.R. Yang, F. Sun, S.Y. Zhan, [Risk on bias assessment: (6) A Revised Tool for the Quality Assessment on Diagnostic Accuracy Studies (QUADAS-2)], Zhonghua liu xing Bing. xue za zhi = Zhonghua liuxingbingxue zazhi 39 (4) (2018) 524–531. PubMed PMID: 29699051.
[9] D. Wilimitis, C.G. Walsh, Practical considerations and applied examples of cross-validation for model development and evaluation in health care: tutorial, Jmir Ai 2 (2023) e49023. PubMed PMID: 38875530. Pubmed Central PMCID: 11041453.
[10] S.A. Harmon, T.H. Sanford, S. Xu, E.B. Turkbey, H. Roth, Z. Xu, et al., Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets, Nat. Commun. 11 (1) (2020) 4080. PubMed PMID: 32796848. Pubmed Central PMCID: 7429815 is employed by Philips Research, North America.
[11] S.K. Corbacioglu, G. Aksel, Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value, Turk. J. Emerg. Med. 23 (4) (2023) 195–198. PubMed PMID: 38024184. Pubmed Central PMCID: 10664195.
[12] E.K. Park, S. Kwak, W. Lee, J.S. Choi, T. Kooi, E.K. Kim, Impact of AI for digital breast tomosynthesis on breast cancer detection and interpretation time, Radiol. Artif. Intell. 6 (3) (2024) e230318. PubMed PMID: 38568095. Pubmed Central PMCID: 11140510.
[13] H.X. Bai, R. Wang, Z. Xiong, B. Hsieh, K. Chang, K. Halsey, et al., Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT, Radiology 296 (3) (2020). PubMed PMID: 32339081. Pubmed Central PMCID: PMC7233483. Epub 2020/04/28. eng.
[14] D. Ippolito, C. Maino, D. Gandola, P.N. Franco, R. Miron, V. Barbu, et al., Artificial intelligence applied to chest X-ray: a reliable tool to assess the differential diagnosis of lung pneumonia in the emergency department, Diseases 11 (4) (2023). PubMed PMID: 37987282. Pubmed Central PMCID: 10660530.
[15] J.G. Nam, E.J. Hwang, J. Kim, N. Park, E.H. Lee, H.J. Kim, et al., AI improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial, Radiology 307 (2) (2023) e221894. PubMed PMID: 36749213.
[16] Adam M. Ostrovsky, Evaluating a large language model's accuracy in chest X-ray interpretation for acute thoracic conditions, Am. j. emerg. med. 93 (2025) 99–102.
[17] Chetla N, Tandon M, Chang J, Sukhija K, Patel R, Sanchez R. Evaluating ChatGPT's Efficacy in Pediatric Pneumonia Detection From Chest X-Rays: Comparative Analysis of Specialized AI Models. JMIR AI. 2025 Jan 10;4:e67621. doi: 10.2196/67621. PMID: 39793007; PMCID: PMC11759907.
[18] Cihan Aydin, et al., AI Models for Accurate Bacterial Pneumonia Diagnosis in Chest X-ray Images, Bozok Tıp Dergisi 15 (2) (2025) 169–177.
[19] Colin J., Surantha N. (2025). Interpretable Deep Learning for Pneumonia Detection Using Chest X-Ray Images. Information 16 (1): 53. MDPI.doi.org/10.3390/info16010053.
[20] Shin, Hye Sun, et al., Machine learning model for circulating tumor DNA detection in chronic obstructive pulmonary disease patients with lung cancer, Translational Lung Cancer Research 13 (1) (2024) 112.
[21] H. Yoo, K.H. Kim, R. Singh, S.R. Digumarthy, M.K. Kalra, Validation of a Deep Learning Algorithm for the Detection of Malignant Pulmonary Nodules in Chest Radiographs, JAMA Netw Open. 3 (9) (2020) e2017135, https://doi.org/10.1001/jamanetworkopen.2020.17135. PMID: 32970157; PMCID: PMC7516603.
[22] D. Ueda, A. Yamamoto, A. Shimazaki, S.L. Walston, T. Matsumoto, N. Izumi, T. Tsukioka, H. Komatsu, H. Inoue, D. Kabata, N. Nishiyama, Y. Miki, Artificial intelligence-supported lung cancer detection by multi-institutional readers with multi-vendor chest radiographs: a retrospective clinical validation study, BMC Cancer. 21 (1) (2021) 1120, https://doi.org/10.1186/s12885-021-08847-9. PMID: 34663260; PMCID: PMC8524996.
[23] W. Hendrix, N. Hendrix, E.T. Scholten, M. Mourits, J. Trap-de Jong, S. Schalekamp, et al., Deep learning for the detection of benign and malignant pulmonary nodules in non-screening chest CT scans, Commun. Med. 3 (1) (2023) 156. PubMed PMID: 37891360. Pubmed Central PMCID: 10611755.
[24] P.N. Megat Ramli, A.N. Aizuddin, N. Ahmad, Z. Abdul Hamid, K.I. Ismail, A Systematic review: the role of artificial intelligence in lung cancer screening in detecting lung nodules on chest X-rays, Diagnostics 15 (3) (2025). PubMed PMID: 39941176. Pubmed Central PMCID: 11817343.
[25] E. Grolleau, S. Couraud, E. Jupin Delevaux, C. Piegay, A. Mansuy, J. de Bermont, et al., Incidental pulmonary nodules: Natural language processing analysis of radiology reports, Respir. Med. Res. 86 (2024) 101136. PubMed PMID: 39232429.
[26] D. Robert, S. Sathyamurthy, A.K. Singh, S.A. Matta, M. Tadepalli, S. Tanamala, et al., Effect of artificial intelligence as a second reader on the lung nodule detection and localization accuracy of radiologists and non-radiology physicians in chest radiographs: a multicenter reader study, Acad. Radiol. 32 (3) (2025) 1706–1717. PubMed PMID: 39592384. Epub 2024/11/27. eng.
[27] J. Rudolph, C. Huemmer, A. Preuhs, G. Buizza, B.F. Hoppe, J. Dinkel, et al., Nonradiology health care professionals significantly benefit from AI assistance in emergency-related chest radiography interpretation, Chest 166 (1) (2024) 157–170. PubMed PMID: 38295950. Pubmed Central PMCID: PMC11251081. Epub 2024/02/01. eng.
[28] M. Zubair, Clinical applications of artificial intelligence in identification and management of bacterial infection: systematic review and meta-analysis, Saudi J. Biol. Sci. 31 (3) (2024) 103934. PubMed PMID: 38304541. Pubmed Central PMCID: 10831261.
[29] G. Chassagnon, M. Vakalopoulou, E. Battistella, S. Christodoulidis, T.N. Hoang-Thi, S. Dangeard, et al., AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia, Med. Image Anal. 67 (2021) 101860. PubMed PMID: 33171345. Pubmed Central PMCID: 7558247.
[30] H.M. Cheo, C.Y.G. Ong, Y. Ting, A systematic review of ai performance in lung cancer detection on CT thorax, Healthcare 13 (13) (2025). PubMed PMID: 40648536. Pubmed Central PMCID: 12250385.
[31] Z. Kirshenboim, E.K. Gilat, L. Carl, E. Bekker, N. Tau, M. Klug, et al., Retrospectively assessing evaluation and management of artificial-intelligence detected nodules on uninterpreted chest radiographs in the era of radiologists shortage, Eur. J. Radiol. 170 (2024) 111241. PubMed PMID: 38042019.
[32] J. Geppert, P. Auguste, A. Asgharzadeh, H. Ghiasvand, M. Patel, A. Brown, et al., Software with artificial intelligence-derived algorithms for detecting and analysing lung nodules in CT scans: systematic review and economic evaluation, Health Technol. Assess. 29 (14) (2025) 1–234. PubMed PMID: 40380885. Pubmed Central PMCID: 12104856.
[33] B. Ilsen, F. Vandenbroucke, C. Beigelman-Aubry, C. Brussaard, J. de Mey, Comparative Interpretation of CT and Standard Radiography of the Pleura, J. Belg. Soc. Radiol. 100 (1) (2016) 106. PubMed PMID: 30151497. Pubmed Central PMCID: 6100645.
[34] J. Lopez Alcolea, A. Fernandez Alfonso, R. Cano Alonso, A. Alvarez Vazquez, A. Diaz Moreno, D. Garcia Castellanos, et al., Diagnostic performance of artificial intelligence in chest radiographs referred from the emergency department, Diagnostics 14 (22) (2024). PubMed PMID: 39594258. Pubmed Central PMCID: 11592727.
[35] M.K. Kim, C. Rouphael, J. McMichael, N. Welch, S. Dasarathy, Challenges in and opportunities for electronic health record-based data analysis and interpretation, Gut liver 18 (2) (2024) 201–208. PubMed PMID: 37905424. Pubmed Central PMCID: 10938158.
[36] D. Chen, L. Yang, W. Zhang, J. Shen, P.E.Y. Van Schil, D. Divisi, et al., Prevalence and management of pulmonary nodules: a systematic review and meta-analysis,

J. Thorac. Dis. 16 (7) (2024) 4619–4632. PubMed PMID: 39144359. Pubmed Central PMCID: 11320231.
[37] S. Jang, J. Kim, J.S. Lee, Y. Jeong, J.G. Nam, J. Kim, et al., Real-world diagnostic performance and clinical utility of artificial-intelligence-assisted interpretation for detection of lung metastasis on CT in patients with colorectal cancer, AJR Am. J. Roentgenol. (2025). PubMed PMID: 40499025.
[38] B.D. Katzman, C.B. van der Pol, P. Soyer, M.N. Patlas, Artificial intelligence in emergency radiology: A review of applications and possibilities, Diagn. Interv. Imaging 104 (1) (2023) 6–10. PubMed PMID: 35933269.
[39] Han G.R., Goncharov A., Eryilmaz M., Ye S., Palanisamy B., Ghosh R., et al. Machine learning in point-of-care testing: innovations, challenges, and opportunities. Nat. Commun. 2025;16(1):3165. PubMed PMID: 40175414. Pubmed Central PMCID: 11965387 patent applications on computational POC sensors. K.G. is a shareholder of CYBO, LucasLand, and FlyWorks. The remaining authors declare no competing interests.