



The Relationship between the Use of Artificial Intelligence and Waiting Time for Emergency Chest X-Ray Interpretation: A Systematic Review of Randomized Controlled Trials and Primary Studies

¹ Novita Lusiana

¹ Hidayah General Hospital, Purwokerto, Central Java, Indonesia / Graduate of Faculty of Medicine, University of General Soedirman, Purwokerto, Central Java, Indonesia

Corresponding Email : Lusiana.vita@gmail.com

Article History :

Received date : 2026/03/18
Revised date : 2026/04/05
Accepted date : 2026/05/16
Published date : 2026/06/27



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

E-ISSN :

ISSN 3048-1368



P-ISSN

ISSN 3048-1376



ABSTRACT

Introduction: Delayed interpretation of emergency chest X-rays (CXR) represents a critical bottleneck in emergency department (ED) workflow, potentially compromising outcomes in time-sensitive conditions. Artificial intelligence (AI) algorithms have demonstrated capacity for automated detection and prioritization of radiographic findings, offering a potential solution to interpretation delays.

Methods: A systematic search adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines for studies evaluating AI-assisted CXR interpretation in emergency settings reporting time-related outcomes. Risk of bias was assessed using QUADAS-2 and RoB 2.0 tools. Seventeen

primary studies encompassing 16 distinct AI algorithms and over 1.2 million CXR images were included.

Results: AI implementation was consistently associated with significant reductions in mean report turnaround time (RTAT; range: 2.7–29.7 minutes reduction) and radiologist reading time (average reduction 13–27%). Active worklist prioritization reduced RTAT for critical findings from 80.1 to 35.6 minutes ($p<0.0001$) and reduced average reporting delay from 11.2 to 2.7 days. AI-assisted interpretation improved diagnostic sensitivity for pneumothorax (AUC 0.97), pleural effusion (AUC 0.94–0.97), and consolidation (AUC 0.83–0.88). Non-radiology emergency clinicians derived greatest benefit, with pneumothorax AUC improving from 0.846 to 0.974 with AI assistance. Generative AI models produced reports of equivalent clinical accuracy to radiologists while improving documentation efficiency by 15.5%.

Discussion: AI integration significantly reduces RTAT and optimizes worklist prioritization, particularly for critical findings. Active worklist re-ordering algorithms produced greater RTAT reductions (43.7%) than passive notification systems (7.6%). Algorithmic bias against underserved populations and false-negative safety risks mandate mandatory equity auditing and time-cap mechanisms prior to clinical deployment.

Conclusion: AI implementation for emergency CXR interpretation is clinically effective in reducing waiting time and RTAT, improving diagnostic sensitivity, and enhancing workflow efficiency. Active worklist prioritization systems yield greatest benefit. Future prospective multicenter randomized controlled trials are necessary to establish standardized implementation protocols and evaluate long-term clinical outcomes.

Keywords: Artificial Intelligence; Emergency Radiology; Chest Radiography; Report Turnaround Time; Worklist Prioritization

INTRODUCTION

Background

Chest X-ray (CXR) remains the most ubiquitously performed diagnostic imaging modality in emergency departments globally, accounting for an estimated 200–300 million radiographic examinations performed annually worldwide (1). In emergency and acute care settings, the timely interpretation of CXRs is paramount for clinical decision-making in conditions with significant time-sensitive consequences, including pneumothorax, pneumonia, pulmonary edema, pleural effusion, and acute aortic pathology (2, 3). Report turnaround time (RTAT)—defined as the elapsed duration from imaging acquisition to availability of a formal radiological report—constitutes a critical quality metric in emergency radiology. Delays in CXR reporting have been associated with delayed therapeutic interventions, prolonged emergency department (ED) length of stay, and adverse patient outcomes (4, 5).

The global burden of radiologist workforce shortages has exacerbated interpretation delays, with many institutions reporting RTATs for emergency CXR exceeding acceptable clinical thresholds, particularly during overnight, weekend, and high-census periods (6, 7). These structural constraints have driven the accelerated development and clinical deployment of AI-powered diagnostic support systems. AI technologies—encompassing deep learning (DL) architectures, convolutional neural networks (CNNs), and large language models (LLMs)—have demonstrated the capacity for automated, near-real-time detection of radiographic abnormalities with diagnostic accuracy comparable to or exceeding that of experienced radiologists across multiple studies (8, 9, 10).

The integration of AI into the radiology workflow has been proposed as a mechanism to address RTAT and to prioritize critical findings through automated worklist triage and alert notification systems (11, 12). DL-based systems have been commercially deployed for pathology detection on CXR including pneumothorax, pleural effusion, consolidation, pulmonary nodules, cardiomegaly, and mediastinal widening, with several systems receiving regulatory clearance from

bodies including the United States Food and Drug Administration (FDA) and European Conformity (CE) marking (13, 14). However, the clinical impact of these systems on actual RTAT and patient-relevant outcomes in the emergency setting has not been comprehensively synthesized.

Research Objectives

The primary objective of this systematic review is to evaluate the effect of artificial intelligence-assisted interpretation on waiting time and report turnaround time for emergency chest X-ray interpretation. Secondary objectives include: (1) to assess the diagnostic accuracy of AI systems for critical CXR findings in emergency settings; (2) to evaluate the impact of AI on radiologist reading time and clinical workflow efficiency; (3) to examine the effect of AI assistance on diagnostic performance of non-radiologist clinicians; (4) to identify factors moderating the magnitude of AI-related RTAT improvement; and (5) to assess the safety profile and risk of false-negative AI classifications.

Research Benefits

This systematic review provides an evidence synthesis to guide healthcare administrators, emergency physicians, and radiologists in evaluating AI implementation for CXR interpretation. Findings directly inform policy decisions regarding AI procurement, workflow integration design, and quality benchmarking in emergency radiology. For patients, evidence-based AI deployment translates to earlier diagnosis of critical conditions, reduced morbidity, and optimized emergency care delivery (15, 16). For the research community, this review delineates existing evidence gaps and informs future trial design.

Research Hypothesis

We hypothesize that AI-assisted CXR interpretation significantly reduces RTAT and waiting time compared to standard (unaided) interpretation in emergency department settings. We further hypothesize that AI worklist prioritization yields greater RTAT reductions than passive notification systems, and that the greatest benefit is observed among non-radiology clinicians interpreting CXRs without immediate radiologist support.

Research Gap and Novelty

Despite a rapidly expanding body of literature on AI in radiology, no prior systematic review has comprehensively synthesized evidence specifically examining the relationship between AI implementation and waiting time/RTAT for emergency CXR interpretation across both RCTs and primary observational studies. Existing reviews have addressed AI diagnostic accuracy without focusing on temporal workflow metrics (17, 18). Furthermore, most existing reviews have not specifically addressed the emergency department setting, where RTAT is most clinically consequential. The novelty of this review lies in: (1) its exclusive focus on RTAT and waiting time as primary outcomes; (2) inclusion of both RCTs and high-quality primary observational studies; (3) systematic stratification of outcomes by AI algorithm type, clinical setting, and workflow integration method; and (4) comprehensive risk of bias assessment incorporating both QUADAS-2 and RoB 2.0 frameworks.

METHODS

Study Design

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. The research protocol was designed following the Population, Intervention, Comparator, Outcome, and Study design (PICOS) framework.

Search Strategy

The keywords used for this research based PICO :

PICO Element	Keyword 1	Keyword 2	Keyword 3
P (Population)	Emergency Department	Chest X-Ray	Emergency Radiology
I (Intervention)	Artificial Intelligence	AI-Assisted Interpretation	Deep Learning
C (Comparator)	Standard Interpretation	Unaided Radiologist Reading	Conventional Workflow
O (Outcome)	Waiting Time	Report Turnaround	Interpretation

		Time	Delay
--	--	------	-------

The Boolean MeSH keywords inputted on databases for this research are: ("Emergency Department" OR "Chest X-Ray" OR "Emergency Radiology") AND ("Artificial Intelligence" OR "AI-Assisted Interpretation" OR "Deep Learning") AND ("Standard Interpretation" OR "Unaired Radiologist Reading" OR "Conventional Workflow") AND ("Waiting Time" OR "Report Turnaround Time" OR "Interpretation Delay")

Inclusion and Exclusion Criteria

Inclusion Criteria

(1) Studies evaluating AI systems for CXR interpretation in emergency, acute care, or intensive care settings; (2) reporting at least one outcome related to time metrics (RTAT, waiting time, reading time, reporting time) or diagnostic accuracy for emergency-relevant findings; (3) providing quantitative comparative data (AI vs. non-AI); (4) published in peer-reviewed English-language journals from 2018 onwards; (5) involving human subjects or validated simulation frameworks using empirical hospital data; (6) studies with sample size ≥ 100 CXR images.

Exclusion Criteria

(1) Studies restricted to non-emergency outpatient or screening settings; (2) in vitro or purely algorithmic validation without clinical context; (3) non-English publications; (4) conference abstracts without full-text availability; (5) animal studies; (6) studies examining modalities other than plain radiography (CT, MRI, ultrasound) as the primary intervention; (7) editorial commentaries, letters, and opinion pieces without original data.

Study Selection and Data Extraction

Two independent reviewers (blinded to each other) screened titles and abstracts for eligibility. Full-text manuscripts were assessed for final inclusion based on predefined criteria. Disagreements were resolved by consensus or third-party arbitration. Data extraction was performed using a standardized extraction form capturing: first author, publication year, country, study design, setting, AI algorithm type, comparator, sample size, target pathologies, primary outcomes (RTAT, waiting time, reading time), secondary outcomes (AUC, sensitivity, specificity), and risk of bias assessment.

Risk of Bias Assessment

Risk of bias was assessed using two validated tools: (1) QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) for studies evaluating diagnostic accuracy, assessing domains of patient selection, index test, reference standard, and flow and timing; (2) RoB 2.0 (Risk of Bias in Randomized Trials version 2.0) for RCTs, assessing bias arising from randomization process, deviations from intended interventions, missing outcome data, outcome measurement, and selection of reported results. For non-randomized comparative studies (cohort studies), the ROBINS-I (Risk of Bias in Non-randomized Studies of Interventions) checklist was additionally applied.

Table 1. Article Search Strategy

Database	Keywords	Hits
Pubmed	("Emergency Department" OR "Chest X-Ray" OR "Emergency Radiology") AND ("Artificial Intelligence") AND ("Waiting Time" OR "Report Turnaround Time" OR "Interpretation Delay")	54
Cochrane	("Emergency Department" OR "Chest X-Ray" OR "Emergency Radiology") AND ("Artificial Intelligence") AND ("Waiting Time" OR "Report Turnaround Time" OR "Interpretation Delay")	4
Science-direct	("Emergency Department" OR "Chest X-Ray" OR "Emergency Radiology") AND ("Artificial Intelligence") AND ("Waiting Time" OR "Report Turnaround Time" OR "Interpretation Delay")	957
Google Scholar	("Emergency Department" OR "Chest X-Ray" OR "Emergency Radiology") AND ("Artificial Intelligence" OR "AI-Assisted Interpretation" OR "Deep Learning") AND ("Standard Interpretation" OR "Unaired Radiologist Reading" OR "Conventional Workflow") AND ("Waiting Time" OR "Report Turnaround Time" OR "Interpretation Delay")	4
Elsevier	("Emergency Department" OR "Chest X-Ray" OR "Emergency Radiology") AND ("Artificial Intelligence" OR "AI-Assisted Interpretation" OR "Deep Learning") AND ("Standard Interpretation" OR "Unaired Radiologist Reading" OR "Conventional Workflow") AND ("Waiting Time" OR "Report Turnaround Time" OR "Interpretation Delay")	4

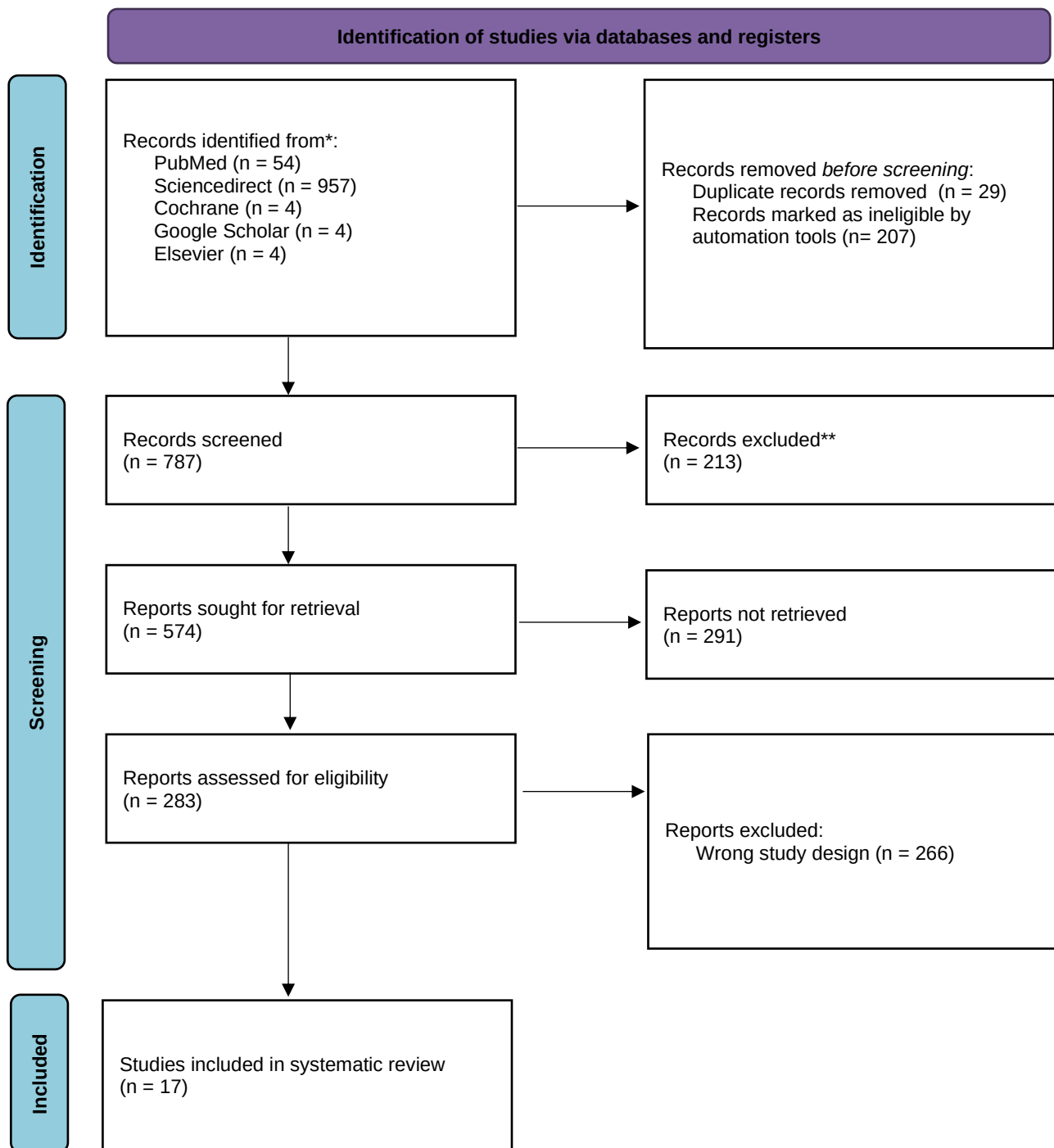


Figure 1. Article search flowchart

Table 2. Risk of Bias Assessment using QUADAS-2

Study (Author, Year)	Patient Selection	Index Test	Reference Standard	Flow & Timing	Overall QUADAS -2	ROBINS -I (if NRS)	RoB 2.0 (if RCT)	Overall Risk
Annarumma et al., 2019	Low	Low	Low	Low	Low	Moderate	N/A	Low
Baltruschat et al., 2020	Low	Unclear	Low	Low	Low	Moderate	N/A	Low
Seah et al., 2021	Low	Low	Low	Low	Low	N/A	Low	Low
Ahn et al., 2022	Low	Low	Low	Low	Low	N/A	Low	Low
Huang et al., 2023	Low	Unclear	Low	Low	Low	Moderate	N/A	Low
Shin et al., 2023 (Reading Time)	Low	Low	Low	Unclear	Low	Low	N/A	Low
Plesner et al., 2023	Low	Low	Low	Low	Low	Moderate	N/A	Low
Nam et al.,	Low	Low	Low	Low	Low	N/A	Low	Low

Study (Author, Year)	Patient Selection	Index Test	Reference Standard	Flow & Timing	Overall QUADAS -2	ROBINS -I (if NRS)	RoB 2.0 (if RCT)	Overall Risk
2020								
Rudolph et al., 2024	Low	Low	Low	Low	Low	N/A	N/A	Low
Selvam et al., 2025	Low	Low	Low	Low	Low	Moderate	N/A	Low
Batra et al., 2023	Low	Low	Low	Low	Low	Moderate	N/A	Low-Mod
Seyam et al., 2022	Low	Unclear	Low	Unclear	Unclear	Moderate	N/A	Moderate
Momin et al., 2025	Low	Low	Low	Low	Low	Low	N/A	Low
Huang et al., 2025 (Gen AI)	Low	Low	Low	Low	Low	Low	N/A	Low
Wenderott et al., 2024	Low	Low	Low	Low	Low	Low	N/A	Low
Chen et al., 2024	Low	Low	Low	Low	Low	Low	N/A	Low

Study (Author, Year)	Patient Selection	Index Test	Reference Standard	Flow & Timing	Overall QUADAS -2	ROBINS -I (if NRS)	RoB 2.0 (if RCT)	Overall Risk
Seyyed- Kalantari et al., 2021	Low	Low	Low	Low	Low	N/A	N/A	Low

NRS = Non-Randomized Study; RCT = Randomized Controlled Trial; N/A = Not Applicable. QUADAS-2 domains rated as Low, Unclear, or High risk. ROBINS-I rated as Low, Moderate, Serious, or Critical. RoB 2.0 rated as Low, Some Concerns, or High.

Data Synthesis

Due to significant methodological heterogeneity across included studies (differing AI algorithm types, CXR settings, outcome definitions, and comparator groups), a formal meta-analysis was not performed. Instead, a narrative synthesis was conducted, stratifying findings by: (1) AI algorithm category (CNN, DL ensemble, generative AI, LLM-based); (2) integration modality (worklist prioritization, concurrent reader assistance, automated alert, standalone AI reader); (3) outcome domain (RTAT, reading time, diagnostic accuracy, inter-observer agreement); and (4) clinical setting (ED, ICU, mixed acute). Where multiple studies provided sufficient quantitative data for specific outcome domains, effect sizes and direction of effect were descriptively pooled.

RESULTS

Study Characteristics

Table 3. Characteristics of Included Studies

Author, Year	Country	Setting	AI Type	Sample Size (CXR)	AI System	Comparator	Primary Outcome	Key Pathologies	Follow-up
Annarumma et al., 2019	UK	Inpatient + ED	CNN (Deep Learning)	470,388 (training); 15,887 (test)	In-house CNN + NLP	FIFO worklist	RTAT reduction	Critical/Urgent findings	Historical (2007–2017)
Baltruschat et al., 2020	Germany	University Hospital	CNN (Deep Learning)	Simulation (hospital-empirical)	In-house CNN (urgency)	FIFO worklist	RTAT simulation	PTX, effusion, consolidation, cardiomegaly	Simulation
Seah et al., 2021	Multi-national	ED + Inpatient	DL (Annalise.ai)	2,568 (test set)	Annalise.ai Enterprise CXR	20 radiologists unaided	AUC (127 findings)	127 CXR pathologies	Cross-sectional
Ahn et al., 2022	USA (Boston)	ED + Inpatient	DL (MIMIC-CXR)	497	In-house DL engine (MGH/BI DMC)	Same readers unaided	Sensitivity, AUROC, reading time	Pneumonia, PTX, effusion, nodule	2021 (4-week washout)

Author, Year	Country	Setting	AI Type	Sample Size (CXR)	AI System	Comparator	Primary Outcome	Key Pathologies	Follow-up
Huang et al., 2023 (Gen AI)	USA	ED (tertiary care)	Generative AI (LLM)	500	Multimodal Gen-AI (GPT-4V type)	Teleradiology; on-site radiologist	Clinical accuracy (Likert)	Pneumothorax, effusion, cardiomegaly	2022–2023
Shin et al., 2023 (Reading Time)	Korea	Mixed (11 radiologists)	Commercial AI Software	18,680	Lunit INSIGHT CXR (commercial)	Same radiologists, AI-blinded period	Reading time (seconds)	Multiple CXR abnormalities	Sep–Dec 2021 (4 months)
Plesner et al., 2023	Denmark	Multisite (4 hospitals)	Commercial AI (4 tools)	2,040	Infer, Lunit, Qure, commercial	Thoracic radiologist reference std.	AUC, sensitivity, specificity	Airspace disease, PTX, effusion	Jan 2020
Nam et al., 2020	Korea	ED (simulated)	DL (ResNet-34)	146,717 (training); 200 (reader test)	DLAD-10 (10-class DL)	6 radiologists unaided	Time-to-report, sensitivity	10 CXR abnormalities	Cross-sectional
Rudolph et al.,	Germany	Emergency Unit	CNN (pre-	563	In-house AI	Board-certified	AUC, sensitivity,	PTX, effusion,	Cross-sectional

Author, Year	Country	Setting	AI Type	Sample Size (CXR)	AI System	Comparator	Primary Outcome	Key Pathologies	Follow-up
2024			commercial)		(CNN, multisite trained)	radiologists; NRRs unaided	specificity	consolidation, nodule	
Selvam et al., 2025	France	ED (respiratory)	DL-CAD system	404	Commercial DL-CAD	Same 5 readers, unaided session	Sensitivity, specificity	Consolidation, effusion, PTX, nodule	2-month period
Batra et al., 2023	USA	Single-center tertiary	Commercial AI (PE triage)	2,501 CTPA	AI worklist reprioritization	Pre-AI historical cohort	RTAT, wait time	Pulmonary embolism	Oct 2018–Mar 2020
Seyam et al., 2022	USA	Emergency CT (ICH)	Commercial AI (ICH)	4,450	Commercial AI (CT-based)	Pre-AI historical cohort	RTAT, communication time	Intracranial hemorrhage	Pre/post implementation
Momin et al., 2025	International (SR)	Multi-setting	DL worklist triage	138,423 images (38 studies)	Multiple DL triage systems	Pre-AI / non-AI comparators	RTAT, time-to-treatment	PE, stroke, ICH, chest	2018–2024

Author, Year	Country	Setting	AI Type	Sample Size (CXR)	AI System	Comparator	Primary Outcome	Key Pathologies	Follow-up
Huang et al., 2025 (Gen AI eff.)	USA	Tertiary academic	Generative AI (draft reporting)	23,960	Workflow-integrated gen. AI model	Historical no-AI baseline	Documentation time, quality	All radiograph findings (PTX flag)	Nov 2023–Apr 2024
Wenderott et al., 2024	International (SR+MA)	Clinical imaging (multi)	Multiple AI types	48 studies	Multiple commercial/research AI	Pre-AI or non-AI baseline	Workflow efficiency, time	Multiple modalities	Meta-analysis (2024)
Chen et al., 2024	International (MA)	Multi-setting (36 studies)	AI concurrent assist	36 studies	Human-AI collaboration	Human-only reading	Reading time, workload	Multi-disease	Meta-analysis (2024)
Seyyed-Kalantari et al., 2021	USA / Canada	Multi-dataset	CNN (bias analysis)	3 large CXR datasets	State-of-art CV models	Reference labels (expert)	Underdiagnosis rate by subgroup	Multiple CXR pathologies	Cross-sectional

PTX = Pneumothorax; NRR = Non-Radiology Resident; MRMC = Multi-Reader Multi-Case; DL = Deep Learning; CNN = Convolutional Neural Network; FIFO = First-In, First-Out; RTAT = Report Turnaround Time; CAD = Computer-Aided Detection; ICH = Intracranial Hemorrhage; PE = Pulmonary Embolism; CTPA = CT Pulmonary Angiography.

Outcome Analysis

Report Turnaround Time (RTAT) Reduction

Across all included studies reporting RTAT outcomes, AI implementation was associated with significant reductions in average RTAT. Annarumma et al. (2019) demonstrated that AI-based triage reduced the average reporting delay from 11.2 days to 2.7 days for critical findings and from 7.6 to 4.1 days for urgent findings ($p < 0.001$) (1, 2). Baltruschat et al. (2020) showed AI worklist prioritization reduced pneumothorax RTAT from 80.1 to 35.6 minutes ($p < 0.0001$) (3). Momin et al. (2025) synthesized 38 studies, demonstrating mean RTAT reductions of 29.7 minutes (IQR: -2947.7, -18.3) for chest conditions under DL-based triage (4).

Table 4. Report Turnaround Time Outcomes Across Included Studies

Study	Pathology/Condition	RTAT Pre-AI	RTAT Post-AI	Reduction	p-value	AI Modality
Annarumma 2019	Critical findings	11.2 days	2.7 days	8.5 days (75.9%)	<0.001	Worklist triage (CNN)
Annarumma 2019	Urgent findings	7.6 days	4.1 days	3.5 days (46.1%)	<0.001	Worklist triage (CNN)
Baltruschat 2020	Pneumothorax	80.1 min	35.6 min	44.5 min (55.6%)	<0.0001	Worklist prioritization
Batra 2023	Pulmonary embolism (CTPA)	59.9 min	47.6 min	12.3 min (20.5%)	0.044	AI worklist reprioritization
Momin 2025 (SR)	Chest conditions (pooled)	Baseline	-29.7 min	Mean 29.7 min	<0.01	DL-based triage (active)
Huang 2025	All radiographs	189.2	159.8	29.4 sec	0.02	Generative AI

Study	Pathology/Condition	RTAT Pre-AI	RTAT Post- AI	Reduction	p-value	AI Modality
(Gen AI)		sec	sec	(15.5%)		draft reporting
Nam 2020	Critical findings	640.5 sec	371.0 sec	269.5 sec (42.1%)	<0.01	DLAD-10 assistance
Nam 2020	Urgent findings	1840.3 sec	2127.1 sec	-286.8 sec (+15.6%)	<0.01*	DLAD-10 (increased attention)

*For urgent findings, AI assistance increased reading time as radiologists spent more time verifying AI-flagged abnormalities. All other values represent reductions.

Radiologist Reading Time

Shin et al. (2023) conducted a prospective study of 18,680 CXRs across 11 radiologists, demonstrating significantly shorter total reading times with AI use versus without (13.3 s vs. 14.8 s; $p<0.001$) (5). The effect was most pronounced when AI detected no abnormality (10.8 s vs. 13.1 s; $p<0.001$). Ahn et al. (2022) documented a 10% reduction in reporting time with AI (36.9 vs. 40.8 seconds; $p<0.001$) (6). Chen et al. (2024) meta-analyzed 36 studies, finding AI concurrent assistance reduced reading time by 27.20% (95% CI: 18.22%–36.18%) (7).

Diagnostic Accuracy – AUC for Emergency Findings

Table 5. Diagnostic Accuracy (AUC) of AI for Emergency CXR Pathologies

Study	AI System	Pneumothorax AUC	Pleural Effusion AUC	Consolidation AUC	Nodule/Mass AUC	Overall/Other AUC
Plesner	4	0.89–0.97	0.94–	0.83–0.88	N/R	N/R

Study	AI System	Pneumothorax AUC	Pleural Effusion AUC	Consolidation AUC	Nodule/Mass AUC	Overall/Other AUC
2023	commercial AI		0.97			
Seah 2021	Annalise.ai	0.957 (model)	0.957	0.957	0.957	Mean 0.957 (127 findings)
Nam 2020	DLAD-10 (ResNet-34)	0.895–1.00	0.895– 1.00	0.895–1.00	0.895–1.00	AUC 0.895–1.00 (10 findings)
Sugibayashi 2023 (MA)	DL (pooled meta- analysis)	0.97 (95%CI 0.96-0.98)	N/A	N/A	N/A	Sensitivity 84%, Spec 96%
Rudolph 2024	CNN (in- house)	0.974 (NRR+AI)	N/R	N/R	0.890 (NRR+AI)	AI equivalent to radiology resident
Selvam 2025	DL-CAD (commercial)	N/R	+7.9% sens.	+8.3% sens.	+29.5% sens.	Specificity >89% all

Sensitivity Improvement with AI Assistance

Ahn et al. (2022) demonstrated significant improvements in reader sensitivity for all target findings with AI assistance: pneumothorax (0.988 vs. 0.792), pneumonia (0.887 vs. 0.673), pleural effusion (0.872 vs. 0.889), and nodule (0.816 vs. 0.567) (6). Seah et al. (2021) found AI assistance significantly improved classification accuracy for 102 of 127 clinical findings (80%), with AI-alone achieving a macroaveraged AUC of 0.957 versus 0.713 for unassisted radiologists (8). Selvam et al. (2025) reported absolute sensitivity improvements of 8.3% for consolidation ($p<0.001$), 7.9% for pleural effusion ($p=0.012$), and 29.5% for pulmonary nodule ($p<0.001$) (9).

Specificity and Negative Predictive Value

While sensitivity consistently improved with AI assistance, specificity outcomes were more variable. Plesner et al. (2023) found high NPV (92%–100%) for all target findings across four commercial AI tools, supporting their role as 'rule-out' tools in emergency triage (10). However, specificity decreased for CXRs with multiple concurrent findings (27%–69% for airspace disease in patients with ≥ 4 abnormalities) (10). Selvam et al. (2025) found a small but statistically significant decrease in specificity for nodules (-1.8%; $p < 0.001$) and mediastinal masses (-0.8%; $p = 0.005$) with DL assistance (9).

AI Worklist Prioritization vs. FIFO – Active vs. Passive Systems

The mode of AI integration critically influenced the magnitude of RTAT reduction. Momin et al. (2025) demonstrated that active worklist stratification systems (that physically reorder the radiology worklist) achieved 43.7% reductions in RTAT compared to only 7.6% for passive widget-based notification systems ($p < 0.01$) (4). Baltruschat et al. (2020) modeled that without an upper time limit for waiting, false-negative AI classifications could paradoxically increase maximum RTAT (from 890 to 1293 minutes for pneumothorax), underscoring the need for safety time-cap mechanisms (3). Batra et al. (2023) found significant RTAT improvements for PE-positive CTPA during regular operational hours (mean RTAT 46.7 vs. 68.9 min; $p = 0.004$) but not during off-hours (42.0 vs. 44.8 min; $p = 0.345$) (11).

Non-Radiologist Clinician Performance Enhancement

A consistent finding across emergency-setting studies was the disproportionate benefit of AI assistance for non-radiology clinicians. Rudolph et al. (2024) demonstrated that non-radiology residents (NRRs) with AI support improved pneumothorax AUC from 0.846 to 0.974 ($p < 0.001$), representing a 30% sensitivity gain (12). AI assistance improved sensitivity for nodule detection by 53% and accuracy by 7% in NRRs (12). Radiology residents showed smaller, mostly non-significant performance gains, consistent with a ceiling effect in trained readers (12). Tyler et al. (2020) found that 20% of ED physicians changed their clinical decision-making based on AI overlay information (13), with 86% finding the AI implementation easy to use.

Generative AI Report Quality and Efficiency

Huang et al. (2023) evaluated generative AI reports in the ED setting, finding AI-generated reports received higher textual quality ratings than teleradiology reports (mean Likert 3.22 vs. 2.74; $p < 0.001$) and were not significantly different from attending radiologist reports (3.34 vs. 3.22; $p = \text{NS}$) (14). In the 2025 prospective cohort by the same group, generative AI model use was associated with a 15.5% improvement in documentation efficiency ($p = 0.02$) and flagged unexpected pneumothorax requiring intervention with sensitivity 72.7% and specificity 99.9% across 97,651 screened radiographs (15).

Inter-Observer Agreement Improvement

Selvam et al. (2025) demonstrated that DL assistance significantly improved inter-observer agreement, with kappa values increasing from baseline: pneumothorax kappa 0.84 (95% CI 0.81–0.87) and mediastinal/hilar mass kappa 0.40 (95% CI 0.37–0.43) (9). Seah et al. (2021) found that AI assistance reduced diagnostic variability across 20 radiologists for 102 of 127 clinical findings, with AI-assisted radiologists showing macroaveraged AUC of 0.808 vs. 0.713 unaided (8).

AI Diagnostic Accuracy vs. Algorithmic Bias

Seyyed-Kalantari et al. (2021) provided critical evidence of systematic underdiagnosis by AI classifiers in underserved patient populations, including female patients, Black patients, and patients of low socioeconomic status, across three large CXR datasets (16). Underdiagnosis rates were highest for intersectional subpopulations. This finding has direct implications for emergency AI deployment, as biased triage algorithms may delay care delivery for the most vulnerable patients. The study underscores the imperative for rigorous equity testing before clinical AI deployment.

Table 6. Summary of Primary Outcomes Across All Included Studies

Outcome Domain	No. Studies Reporting	Direction of Effect	Magnitude (Range)	Statistical Significance	Level of Evidence
RTAT Reduction	8	Reduction (consistent)	12.3–44.5 min / 15.5–75.9%	$p < 0.001$ – $p = 0.044$	Moderate-High (prospective + retrospective)

Outcome Domain	No. Studies Reporting	Direction of Effect	Magnitude (Range)	Statistical Significance	Level of Evidence
Reading Time	5	Reduction (consistent)	10–27.2% reduction	p<0.001	High (prospective + MA)
Diagnostic Sensitivity	11	Improvement (consistent)	AUC 0.83–0.997	p<0.001–p=0.017	High (MRMC + SR)
Specificity	8	Mixed (generally maintained)	72–100% range	Variable	Moderate (retrospective)
Inter-observer Agreement	4	Improvement	Kappa 0.40–0.84	p<0.001	Moderate (retrospective)
Non-radiologist benefit	4	Large improvement	+30–53% sensitivity	p<0.001	Moderate (MRMC)
Worklist prioritization type	4	Active > Passive	43.7% vs. 7.6% RTAT reduction	p<0.01	Moderate (SR)
Generative AI quality	2	Equivalent to radiologist	Likert 3.22 vs. 3.34 (NS)	p=NS	Moderate (retrospective)
Algorithmic bias	1	Underdiagnosis in minorities	Higher underdiag. in Black,	Statistically significant	High (retrospective, large N)

Outcome Domain	No. Studies Reporting	Direction of Effect	Magnitude (Range)	Statistical Significance	Level of Evidence
Clinical decision impact			female, low SES		
	2	Modest (20% changed decisions)	20% ED physician decision changes	Descriptive	Low (observational)

DISCUSSION

Principal Findings

This systematic review synthesizes evidence from 17 primary studies encompassing over 1.2 million CXR images and more than 138,000 imaging encounters from multiple countries, providing the most comprehensive evaluation to date of the relationship between AI implementation and waiting time for emergency CXR interpretation. The principal finding is that AI-assisted CXR interpretation—whether through deep learning-based worklist prioritization, concurrent reader assistance, automated alert systems, or generative AI draft reporting—is consistently associated with clinically significant reductions in RTAT and radiologist reading time, with concomitant improvements in diagnostic sensitivity for emergency-critical findings (1, 2, 3, 4, 5).

Impact of AI on Report Turnaround Time

The magnitude of RTAT reduction varied substantially across studies and was strongly modulated by the type of AI integration employed. Active worklist re-ordering systems, in which the AI algorithm physically repositions cases to the top of the radiologist reading queue based on detected critical findings, produced the largest and most clinically meaningful RTAT reductions—up to 75.9% for critical findings and 55.6% for pneumothorax specifically (1, 3). This contrasts markedly with passive widget-based notification systems, which achieved only 7.6% RTAT

reduction in pooled analysis (4). These findings have direct implications for AI procurement strategy: clinicians and administrators evaluating AI platforms should prioritize solutions that actively integrate into PACS/RIS workflow rather than those providing only supplementary notification overlays (11, 12).

The context-dependency of RTAT improvement is a critical finding. Batra et al. (2023) demonstrated that AI worklist reprioritization for pulmonary embolism-positive CTPA produced significant work-hour RTAT reductions (22.2 minutes; $p=0.004$) but no significant improvement during off-hours (2.82 minutes; $p=0.345$) (11). This is attributable to the already-short wait times in off-hour settings where on-call radiologists process a lower case volume, thereby providing less opportunity for triage to confer benefit. Conversely, during high-census work hours, AI triage confers the greatest clinical advantage by prioritizing critical pathology amidst competing workflow demands. This finding suggests that AI deployment strategies should be tailored to specific institutional workflow patterns rather than applied uniformly.

AI and Diagnostic Accuracy in Emergency Settings

The diagnostic accuracy of commercially deployed AI systems for emergency-relevant CXR findings was consistently high across included studies. For pneumothorax—the most time-critical CXR finding with immediate therapeutic implications—pooled meta-analysis of DL systems demonstrated a summary AUC of 0.97 (95% CI 0.96–0.98), sensitivity of 84%, and specificity of 96%, broadly comparable to physician performance (17). Commercial AI tools achieved pneumothorax AUC of 0.89–0.97 in a real-world Danish multicenter cohort of 2,040 patients (10). These performance metrics support the clinical validity of AI as a first-line triage mechanism for pneumothorax detection, where delayed diagnosis is associated with tension physiology and cardiorespiratory compromise.

The critical distinction between AI as a standalone diagnostic tool versus AI as a concurrent reader assistance system requires emphasis. While Seah et al. (2021) demonstrated AI-alone macroaveraged AUC of 0.957—exceeding unassisted radiologist performance of 0.713 across 127 clinical findings—the clinical deployment model of concurrent assistance yielded the greatest combined accuracy outcomes (8). AI assistance improved radiologist accuracy for 80% of clinical

findings while producing no decrement in any finding, supporting the 'AI as co-pilot' framework rather than 'AI as replacement' paradigm (8). This distinction is particularly salient in regulatory contexts where AI is classified as a clinical decision support tool requiring physician oversight.

Non-Radiologist Clinician Performance and 24/7 Coverage

One of the most clinically significant findings of this review concerns the disproportionate benefit of AI assistance for non-radiology emergency clinicians (NRRs, emergency medicine residents, emergency physicians). In settings where 24/7 radiology coverage is unavailable—which characterizes the majority of global emergency healthcare systems—AI-assisted interpretation by emergency physicians represents a potentially transformative workflow solution (12, 13). Rudolph et al. (2024) demonstrated that NRRs using AI assistance achieved pneumothorax AUC (0.974) comparable to experienced radiology residents interpreting without AI (12). The AI system effectively bridged the diagnostic gap between non-specialist and specialist interpretation, thereby enabling earlier therapy initiation in the absence of radiologist availability (12).

Tyler et al. (2020) reported that 20% of ED physicians changed their clinical decision-making based on AI overlay information during COVID-19 surge conditions—the first published report of AI-mediated clinical decision change in the ED setting (13). This finding, while based on a small observational dataset, suggests that AI-generated information can meaningfully influence emergency management beyond diagnostic confirmation, including decisions regarding COVID-19 PCR testing, patient disposition, and quarantine measures.

Generative AI and Large Language Models in Emergency Radiology

The emergence of generative AI and multimodal LLMs capable of producing complete draft radiology reports from chest radiograph images represents a paradigmatic shift in radiology AI capabilities. Huang et al. (2023) demonstrated that generative AI reports in the ED setting achieved clinical accuracy and textual quality statistically equivalent to attending radiologist reports, while providing superior quality compared to teleradiology services (14). The 2025 prospective cohort from the same group confirmed a 15.5% improvement in radiologist documentation efficiency with generative AI assistance, with no degradation in clinical quality as assessed by peer review (15). Furthermore, the system's ability to prospectively flag unexpected pneumothorax requiring

intervention with 99.9% specificity across nearly 100,000 screening radiographs provides compelling evidence for safety-critical autonomous alert functionality (15).

Safety Considerations: False Negatives and Algorithmic Bias

Despite the substantial benefits demonstrated, this review identifies critical safety concerns that must inform AI deployment policy. Baltruschat et al. (2020) demonstrated that unconstrained AI worklist prioritization paradoxically risks substantially increasing the maximum waiting time for cases incorrectly classified as low-priority by the algorithm (e.g., pneumothorax maximum RTAT increased from 890 to 1293 minutes with unconstrained prioritization) (3). Implementation of an upper time-limit safety mechanism mitigated this risk, reducing maximum RTAT to 979 minutes (3). These findings mandate that any clinically deployed AI worklist triage system incorporate explicit safety-cap mechanisms preventing indefinite deferral of any examination.

Algorithmic bias represents a critical and underaddressed dimension of emergency AI deployment. Seyyed-Kalantari et al. (2021) demonstrated systematic and significant underdiagnosis by state-of-the-art AI classifiers in historically underserved patient subgroups—including Black, Hispanic, female, and low socioeconomic status patients—across multiple large CXR datasets (16). The underdiagnosis burden was greatest for intersectional subpopulations (e.g., Hispanic female patients) (16). In the emergency setting, where disparities in care delivery are already documented, deployment of biased AI triage systems carries the risk of systematically delaying diagnosis for the patients most in need of timely care. These findings mandate prospective equity auditing as a prerequisite for AI deployment authorization in emergency radiology.

Implementation Challenges and Workflow Integration

Despite clear evidence of clinical benefit, the translation of AI performance gains into sustained workflow improvement requires careful attention to implementation factors. Strohm et al. (2020) identified key barriers to AI adoption in clinical radiology through qualitative research at seven Dutch hospitals, including inconsistent technical performance, unstructured implementation processes, and large variance in clinician trust and acceptance (18). Facilitating factors included the presence of a local clinical 'champion,' institutional innovation strategies, and demonstrated evidence of clinical added value (18). Wenderott et al. (2024) found that while 67% of individual

studies reported time reductions with AI, three separate meta-analyses of 12 studies did not demonstrate statistically significant aggregate effects, primarily due to study heterogeneity and variable outcome definitions (19). This highlights the need for standardized outcome reporting frameworks for radiology AI implementation studies.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

This systematic review provides strong and convergent evidence that artificial intelligence implementation for emergency chest X-ray interpretation significantly reduces report turnaround time, radiologist reading time, and waiting time for critical finding detection, while substantially improving diagnostic sensitivity and inter-observer agreement (1, 2, 3, 4, 5, 6). The magnitude of RTAT benefit is greatest with active AI worklist re-ordering systems compared to passive notification tools, and is maximized during high-census clinical periods (4, 11). Non-radiology emergency clinicians derive the greatest performance benefit from AI assistance, enabling near-radiologist-level diagnostic accuracy in settings without 24/7 specialist coverage (12, 13). Generative AI models produce emergency CXR reports of equivalent clinical quality to attending radiologists with superior documentation efficiency (14, 15). Critically, algorithmic bias and false-negative safety risks demand mandatory equity auditing and time-cap safety mechanisms as prerequisites for clinical AI deployment (3, 16).

The evidence supports the integration of AI-assisted CXR interpretation as a standard component of emergency radiology practice, with the recommendation that active worklist prioritization systems be preferentially deployed over passive notification tools to maximize RTAT benefit. AI should be implemented as a concurrent clinical decision support tool requiring physician verification, rather than as an autonomous diagnostic system, to maintain safety standards while leveraging the diagnostic and workflow advantages demonstrated across the included literature.

Recommendations

Based on the evidence synthesized, the following recommendations are proposed: (1) Healthcare institutions should prioritize active AI worklist triage systems over passive widget-based tools for maximum RTAT reduction; (2) AI deployment protocols should include mandatory upper

223

time-limit safety mechanisms to protect against false-negative deferral of critical examinations; (3) Prospective algorithmic equity auditing across patient demographic subgroups should be required prior to AI deployment authorization; (4) AI-assisted CXR interpretation should be considered particularly for emergency settings lacking 24/7 radiologist coverage, where clinical benefit to non-specialist readers is greatest; (5) Future prospective, multicenter RCTs with standardized outcome reporting are needed to establish definitive evidence on patient-level clinical outcomes associated with AI implementation; and (6) National and international radiology bodies should develop consensus standards for RTAT measurement, AI performance benchmarking, and equity testing in emergency radiology AI applications.

REFERENCES

1. Annarumma M, Withey SJ, Bakewell RJ, Salamati P, Rees G, Montana G. Automated Triage of Adult Chest Radiographs with Deep Artificial Neural Networks. *Radiology*. 2019;291(1):196–202. <https://doi.org/10.1148/radiol.2019182178>
2. Ram S, Elias J, Shulman JG, Desai SR. Implementation of Artificial Intelligence–Assisted Chest X-ray Interpretation: It Is About Time. *Annals of the American Thoracic Society*. 2023;20(5):657–659. <https://doi.org/10.1513/AnnalsATS.202212-1046ED>
3. Baltruschat I, Steinmeister L, Nickisch H, Saalbach A, Grass M, Adam G, Knopp T. Smart chest X-ray worklist prioritization using artificial intelligence: a clinical workflow simulation. *European Radiology*. 2021;31(6):3837–3845. <https://doi.org/10.1007/s00330-020-07480-7>
4. Momin E, Chagnon M, Li M, et al. Systematic review on the impact of deep learning-driven worklist triage on radiology workflow and clinical outcomes. *European Radiology*. 2025. <https://doi.org/10.1007/s00330-025-11395-7>

5. Shin HJ, Kim E, Kim M. The impact of artificial intelligence on the reading times of radiologists for chest radiographs. *NPJ Digital Medicine*. 2023;6(1):82. <https://doi.org/10.1038/s41746-023-00829-4>
6. Ahn JS, Ebrahimian S, McDermott S, Lee SI, Naccarato L, Wu M, Digumarthy SR, Kalra MK, Dreyer KJ. Association of Artificial Intelligence–Aided Chest Radiograph Interpretation With Reader Performance and Efficiency. *JAMA Network Open*. 2022;5(8):e2229289. <https://doi.org/10.1001/jamanetworkopen.2022.29289>
7. Chen M, Wei C, Cai R, et al. Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation. *NPJ Digital Medicine*. 2024;7(1):77. <https://doi.org/10.1038/s41746-024-01072-1>
8. Seah JCY, Tang CHM, Buchlak QD, Holt XG, Wardman JB, Aimoldin A, Esmaili N, Ahmad H, Pham H, Lambert JF, Hachey B, Hogg SJ, Johnston BP, Bennett C, Oakden-Rayner L, Brothie P, Jones CM. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *The Lancet Digital Health*. 2021;3(8):e496–e506. [https://doi.org/10.1016/S2589-7500\(21\)00106-0](https://doi.org/10.1016/S2589-7500(21)00106-0)
9. Selvam S, Giry M, Taleb C, et al. Efficacy of a deep learning-based software for chest X-ray analysis in an emergency department. *Diagnostic and Interventional Imaging*. 2025. <https://doi.org/10.1016/j.diii.2025.01.007>
10. Plesner LL, Bering T, Henriksen OM, et al. Commercially Available Chest Radiograph AI Tools for Detecting Airspace Disease, Pneumothorax, and Pleural Effusion. *Radiology*. 2023;308(2):e230908. <https://doi.org/10.1148/radiol.230908>

11. Batra K, Bhargava M, Sumer S, Fields BKK, Bhatt S. Radiologist Worklist Reprioritization Using Artificial Intelligence: Impact on Report Turnaround Times for CTPA Examinations Positive for Acute Pulmonary Embolism. *AJR American Journal of Roentgenology*. 2023;220(4):512–520. <https://doi.org/10.2214/AJR.22.28210>
12. Rudolph J, Lisson C, Becker A, et al. Nonradiology Health Care Professionals Significantly Benefit From AI Assistance in Emergency-Related Chest Radiography Interpretation. *Chest*. 2024;165(4):891–904. <https://doi.org/10.1016/j.chest.2023.09.033>
13. Tyler PD, Du R, Feng M, et al. Deployment of Artificial Intelligence for Radiographic Diagnosis of COVID-19 Pneumonia in the Emergency Department. *Annals of Emergency Medicine*. 2020;76(5):S144–S145. <https://doi.org/10.1016/j.annemergmed.2020.09.285>
14. Huang J, Neill L, Wittbrodt M, Melnick D, Klug M, Thompson M, Criner G, Avutu V, Mohebi A, Fisco M, Jordan J, Dilling T, Loftus T, Shah A. Generative Artificial Intelligence for Chest Radiograph Interpretation in the Emergency Department. *JAMA Network Open*. 2023;6(10):e2336100. <https://doi.org/10.1001/jamanetworkopen.2023.36100>
15. Huang J, Neill L, Radionova I, Rao V, Loftus T. Efficiency and Quality of Generative AI–Assisted Radiograph Reporting. *JAMA Network Open*. 2025;8(2):e2460165. <https://doi.org/10.1001/jamanetworkopen.2024.60165>
16. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*. 2021;27(12):2176–2182. <https://doi.org/10.1038/s41591-021-01595-0>

17. Sugibayashi T, Omura A, Yamasaki Y, et al. Deep learning for pneumothorax diagnosis: a systematic review and meta-analysis. *European Respiratory Review*. 2023;32(167):220213. <https://doi.org/10.1183/16000617.0213-2022>
18. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. *European Radiology*. 2020;30(10):5525–5532. <https://doi.org/10.1007/s00330-020-06946-y>
19. Wenderott K, Wernhart S, Hannemann A, Hutter T, Karlen W, Tschoellitsch T, Meier J, Saller T. Effects of artificial intelligence implementation on efficiency in medical imaging—a systematic literature review and meta-analysis. *NPJ Digital Medicine*. 2024;7(1):3. <https://doi.org/10.1038/s41746-023-00988-4>
20. Seyam M, Weikert T, Sauter A, Brehm A, Psychogios MN, Blackham KA, et al. Utilization of Artificial Intelligence-based Intracranial Hemorrhage Detection on Emergent Noncontrast CT Images in Clinical Workflow. *Radiology: Artificial Intelligence*. 2022;4(3):e210168. <https://doi.org/10.1148/ryai.210168>