HIGHLY CONFIDENTIAL

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Diagnostics Advisory Committee – Thursday 22 June 2023

Artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer in primary care referrals: early value assessment – 2nd discussion

An early value assessment report has been produced by Warwick Evidence and following the committee meeting in March 2023, draft recommendations have been made.

Concerns have been raised that the literature review inclusion criteria for the assessment were too strict and that the potential benefits of AI-derived software have not been fully captured. Additional analysis has therefore been undertaken to review and summarise studies on AI-derived software for analysing chest X-rays that report results for clinician review alone compared with AI-derived software alone.

The following additional document was made available to the Committee:

1. **Additional analysis** prepared by Cedar Health Technology Research Centre

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Early Value Assessment Addendum

Artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer in primary care referrals: early value assessment addendum

External Assessment Group report

Produced by: Cedar Health Technology Research Centre

Authors:

Michael Beddard, Senior Researcher, Cedar Health Technology Research Centre

Dr Meg Kiseleva, Systematic Reviewer, Specialist Unit for Review Evidence

Dr Susan O'Connell – Principal Researcher, Cedar Health Technology Research Centre

Ayesha Rahim, Researcher, Cedar Health Technology Research Centre

Correspondence to: Cedar Health Technology Research Centre

Cardiff Medicentre
University Hospital of Wales
Heath Park Campus
Cardiff
CF14 4UJ

Date completed: 16/06/2023

Contains confidential information: Yes

Number of attached appendices: 5

Purpose of the assessment report

The purpose of this External assessment group (EAG) report is to review the evidence currently available for included technologies and advise what further evidence should be collected to help inform decisions on whether the technologies should be widely adopted in the NHS. The report may also include additional analysis of the submitted evidence or new clinical evidence. NICE has commissioned this work and provided the template for the report.

Declared interests of the authors

None

Acknowledgements

Robin Alcock, Consultant Radiologist, Salisbury District Hospital

Alberto Alonso, Consultant Thoracic Radiologist, Manchester University NHS Foundation Trust

Fergus Gleeson, Consultant Radiologist. Oxford University Hospital

Christopher Johns, Consultant Thoracic Radiologist. Northern General Hospital, Sheffield

Emma O'Dowd, Consultant Respiratory Physician, Nottingham City Hospital

Ian Woolhouse, Consultant in Respiratory and Internal Medicine, Queen Elizabeth Hospital Birmingham

Alan Thomson, Lay specialist committee member

Responsibility for report

The views expressed in this report are those of the authors and not those of NICE. Any errors are the responsibility of the authors.

Contents

Abbreviations list

Abbreviation	Definition
ACM	Association for Computing Machinery
AI	Artificial intelligence
AP	Anterior posterior
AUC	Area under the curve
AUROC	Area under the receiver operating characteristic curve
CENTRAL	Cochrane Central Register of Controlled Trials
CDSR	Cochrane Database of Systematic Reviews
CI	Confidence interval
CS	Confidence score
CT	Computerised tomography
CXR	Chest X-ray
EAG	External Assessment Group
EVA	Early value assessment
FDR	False Discovery Rate
FOR	False Omission Rate
GP	General Practitioner
HCN	High confidence normal
HRQoL	Health-related quality of life
ICD-10	International Classification of Diseases 10th revision
ICTRP	International Clinical Trials Registry Platform
MAUDE	Manufacturer and User Facility Device Experience
MDT	Multidisciplinary team
MHRA	Medicines and Healthcare products Regulatory Agency
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NPV	Negative predictive value
NR	Not reported
PA	Posterior anterior
PACS	Picture archiving and communication system
PARD	Public Access Registration Database
PPV	Positive predictive value
PROSPERO	International prospective register of systematic reviews
SCM	Specialist committee member
WHO	World Health Organisation
WR	Written report

Executive summary

This report is an addendum to the NICE early value assessment (EVA) report by Warwick Evidence, focusing on AI-based software for the analysis of chest X-rays to detect lung cancer in primary care referrals. The addendum aimed to review the comparative evidence on 14 AI technologies (AI alone) versus clinician review (Clinician alone). The primary population of interest was individuals referred from primary care, with some studies including a mixed population or unclear referral route.

Literature searches were conducted in May 2023, along with discussions with clinical experts and a lay Specialist Committee Member. The review followed the same criteria as the original EVA report. Out of the sifted studies, nine were relevant to the addendum scope, including one systematic review, three ongoing studies, and five published studies. Three technologies were covered in the published evidence: AI-Rad Companion Chest X-ray (Siemens Healthineers), Red Dot (Behold.ai), and Lunit INSIGHT CXR (Lunit). Of the three ongoing studies, two were identified for qXR (Qure.ai) and one for Lunit INSIGHT CXR (Lunit).

All five published studies were retrospective cohort studies, except for one with a prospective evaluation. Two studies reported GP referral, while the referral route was unknown for the others. The reported outcome measures focused on diagnostic accuracy, concordance with clinician review, and one on time taken to report CXR results. However, no outcomes specifically related to lung cancer were reported. The evidence demonstrated high sensitivity and specificity for the technologies, but with a notable rate of false positives. The technologies were also limited in their use with poor-quality radiographs and lateral view images, raising concerns among clinical experts.

Clinical experts highlighted the need for training on these technologies if they are to be successfully adopted within the NHS. They expressed optimism about the future potential of AI technologies in clinical practice, particularly in triaging urgent cases to alleviate workload. However, they cautioned against relying solely on AI and recommended research on their implementation in a radiographer setting as an initial step in the pathway.

Overall, significant evidence gaps remain regarding the use of these AI technologies for chest X-ray analysis. Further research, training, and a user-centered rollout are needed to support broader adoption of AI technologies in the future.

1 Scope of the addendum

This report serves as an addendum to the NICE early value assessment (EVA) produced by Warwick Evidence that examined the application of artificial intelligence (AI)-derived software in the analysis of chest X-rays for suspected lung cancer in primary care referrals (Stinton et al., 2023). The EVA primarily focused on studies that evaluated AI software as an adjunct to clinician review, comparing it with clinician review alone. However, the availability of evidence was limited, and feedback from stakeholders highlighted concerns about the overly restrictive inclusion criteria, which may have overlooked the potential of AI-derived software. The diagnostic committee also concluded that further research was needed on how using AI-derived software alongside clinician review of chest x-rays affects the accuracy of detecting lung cancer.

In this addendum, we aim to reassess the evidence by examining comparative evidence on the use of Chest X-ray (CXR) interpreted by any of the 14 AI technologies in the scope of the review or ‘AI alone’ versus the comparator - CXR interpreted by a radiology specialist (radiologist, reporting radiographer) or 'Clinician alone'. We will also make note of comparative evidence on the intervention versus intervention in conjunction with Clinician review, or ‘Clinician + AI’ versus ‘AI alone’. The primary population of interest is people referred from primary care, but studies with a mixed population or unclear referral route have also been included.

It is important to note that AI software is not intended for autonomous use without the review and approval of clinicians and is solely employed for research purposes. To address stakeholders' concerns and uncover the potential benefits of AI-derived software, we employed existing search strategies used in the EVA, but in a less restrictive approach to identify relevant studies that explore the use of AI-derived software alone versus the comparators.

This addendum supports committee decision making regarding AI-based software used to analyse CXRs for suspected lung cancer in primary care referrals. The EAG contacted each specialist committee member (SCM) and sent them a questionnaire to complete. Five discussions were held with SCMs, including one with a lay SCM. Based on discussions with SCMs and advice from NICE, the focus of the review was

narrowed to Clinician vs AI alone. Modified questions were emailed to SCMs who had not been interviewed or responded. Three additional SCM responses were received. Both versions of the questionnaire are provided in appendix C.

2 Clinical evidence selection

2.1 *Evidence search strategy and study selection*

Searches were originally run between the 12th and 16th May 2023 on Ovid Medline, Ovid Embase, Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), Epistemonikos, Association for Computing Machinery (ACM) Digital Library, PROSPERO, and the WHO International Clinical Trials Registry Platform (ICTRP). The original search strategy developed for the EVA was used and an AI search filter (Ayiku & Finnegan, 2023) was applied to the Medline and Embase Searches – see Appendix D for background information on the development of the NICE AI search filter.

This search produced 2,491 records, which were imported into EndNote and deduplicated, resulting in 1,908 references. However, as the AI filter is not yet validated, both the EAG and NICE decided that the evidence should be searched for without the AI filter to avoid missing any relevant records. A supplementary search was run on the 25th May 2023 to retrieve the records omitted when the AI search filter was applied. The original EVA search strategy was used to retrieve records from Medline and Embase, resulting in 1,667 additional references after deduplication. The total number of screened records from both searches combined was 3,582.

As this work is part of an early value assessment and the short timescale for the review in time for the committee discussion, pragmatic rapid review methods were used rather than full systematic review methods. One reviewer screened by all records by title and abstract using the inclusion, exclusion and post-hoc inclusion criteria used in the EAG EVA report for comparability (Appendix B). No date limit was applied to the searches, but only records published in or after 2012 were screened.

Eight sources of additional evidence were provided by companies in the consultation document to Warwick Evidence during the original EVA. This evidence was all reviewed by Cedar during the addendum and details of each study and reasons for inclusion/exclusion are summarised in table 1.

An additional 5 records not submitted to Warwick evidence during the original EVA were sent to Cedar on 02/06/2023 from four of the companies in the scope of the

review. Only 1 study (Niehoff et al., 2023) had already been identified by the searches conducted by Cedar for the EVA addendum. Two were conference abstracts/preliminary results, 1 pre-print and 2 publications and are summarised in table 2 with reasons for inclusion/exclusion in the addendum. This totaled 76 records to be screened at full-text. During full-text screening, 67 records were excluded. Reasons for study exclusion can be found in Appendix E.

After full-text screening, 9 were remaining; 1 systematic review, 3 ongoing studies, and 5 published studies. A study flowchart of the screening and sifting process is available in figure 1 of Appendix A. The supplementary search without the use of the NICE AI filter identified no additional relevant evidence to the search with the AI filter.

Table 1: Additional company evidence provided in original EVA consultation document

Device & manufacturer	Study author and year	Study title	Type of publication	Identified in Cedar addendum searches	Included in addendum	Comment
Lunit INSIGHT CXR (Lunit)	Kwak et al., 2023	Incidentally found resectable lung cancer with the usage of artificial intelligence on chest radiographs	Published study	Yes	X	Excluded at full-text by Cedar as study is AI alone with no eligible comparator
Lunit INSIGHT CXR (Lunit)	Ahn et al., 2022	Association of Artificial Intelligence–Aided Chest Radiograph Interpretation with Reader Performance and Efficiency	Published study	Yes	X	Excluded at title and abstract by Cedar as comparator is clinician alone and out of scope for addendum
Lunit INSIGHT CXR (Lunit)	Nam et al,. 2023	AI Improves Nodule Detection on Chest Radiographs in a Health Screening Population: A Randomised Controlled Trial	Published study	Yes	X	Excluded at title and abstract by Cedar as study population is screening population and out of scope for addendum
Red dot (Behold.ai)	Behold.ai	Prospective Validation of a diagnostic AI in accelerating Lung Cancer Referral Pathways	Abstract	No	X	Exclude at title and abstract by Cedar as it is a poster/abstract and not peer reviewed and is therefore out of scope for the addendum
Red dot (Behold.ai)	Dyer et al., 2022	Robustness of an Artificial Intelligence Solution for Diagnosis of Normal Chest X-Rays	Pre-print	No	X	Excluded at full text by Cedar as although software and comparators (Clinician vs AI) are relevant, the study does not have extractable data on outcomes relevant to addendum scope. No data on lung cancer or lung nodules.
Red dot (Behold.ai)	Dyer at al., 2021	Diagnosis of normal chest radiographs using an	Published study	Yes	X	Excluded at full text by Cedar as study does not name intervention and

		autonomous deep-learning algorithm				population is unclear and taken from A&E, GP & outpatients
Red dot (Behold.ai)	Dissez et al., 2022	Enhancing Early Lung Cancer Detection on Chest Radiographs with AI-assistance: A Multi-Reader Study	Pre-print	No	X	Excluded at full text by Cedar as study assesses the use of clinician review with and without AI (Clinician + AI vs Clinician) and is therefore out of scope for the addendum
Red dot (Behold.ai)	Tam et al., 2021	Augmenting lung cancer diagnosis on chest radiographs: positioning artificial intelligence to improve radiologist performance	Published study	Yes	✓	Including in addendum as comparator (Clinician + AI vs AI alone) is relevant. However, referral route is not reported.

Table 2: Additional company evidence submissions provided to Cedar for EVA addendum

Device & manufacturer	Study author & year	Study title	Type of publication	Identified in Cedar addendum searches	Included in addendum	Comment
[redacted]	[redacted]	[redacted]	[redacted]	No	X	Excluded as it is an abstract
[redacted]	[redacted]	[redacted]	[redacted]	No	X	Excluded as it is an abstract
Red dot (Behold.ai)	Smith et al. 2023	Real-World Performance of Autonomously Reporting Normal Chest Radiographs in NHS Trusts Using a Deep-Learning Algorithm on the GP Pathway	Publication pre-print	No	✓	Include as study is a retrospective study analysing the performance of the Red Dot (Behold.ai) software as a diagnostic decision support software in two NHS trusts in active clinical pathways. Study assesses 'AI alone' with a subset reviewed by a clinician.
Lunit INSIGHT CXR (Lunit)	Shin et al., 2023	The impact of artificial intelligence on the reading times of radiologists for chest radiographs. *NPJ Digital Medicine*, *6*(1), 82. https://doi.org/10.1038/s41746-023-00829-4	Prospective study	No	X	Excluded at full text by Cedar as study compares AI-unaided vs. AI-aided or 'Clinician + AI vs. Clinician', which is out of scope for this addendum. The population is also in- and outpatients and would likely have been excluded in the EVA report conducted by Warwick Evidence.
AI-Rad Companion Chest X-ray (Siemens Healthineers)	Niehoff et al., 2023	Evaluation of the clinical performance of an AI-based application for the automated analysis of chest X-rays. *Scientific Reports*, *13*(1), 3680.	Scientific report	Yes	✓	Study had already been identified in the literature search and included at full text.

2.2 *Included and excluded studies*

Aggarwal 2021 is a systematic review and was used to check for any missing studies on AI technologies within the scope of the review. No additional studies were found to those found by the literature searches.

Table 3 summarises the included studies by type and AI software technology, along with indications of their presence in and relevance to the original EAG EVA report conducted by Warwick evidence.

Table 3: Included studies in review

Author and date	Study type	AI technology	Identified in original EAG EVA report	Comment
Author: Aggarwal, 2021 Title: Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis	Systematic review	N/A	Yes	A review used for reference checking in the original EAG EVA report by Warwick Evidence
Jagirdar et al., 2023 CTRI/2020/08/027488 (2020) Title: Use of artificial intelligence to interpret chest X-rays	Ongoing study	qXR (Qure.ai)	Yes	Mentioned as an ongoing study in the original EAG EVA report by Warwick Evidence
Avery et al., 2022 NCT05489471 Title: A Study to Assess the Impact of an Artificial Intelligence (AI) System on Chest X-ray Reporting	Ongoing Study	Lunit INSIGHT CXR (Lunit)	Yes	Mentioned as an ongoing study in the original EAG EVA report by Warwick Evidence

Baldwin et al., 2023 – (LungIMPACT study) ISRCTN78987039 Title: Impact of immediate AI enabled patient triage to chest CT on the lung cancer pathway	Ongoing study	qXR (Qure.ai)	No	Not registered at the time of the original EAG EVA search
Niehoff et al., 2023 Title: Evaluation of the clinical performance of an AI-based application for the automated analysis of chest X-rays	Published study	AI-Rad Companion Chest X-ray (Siemens Healthineers)	No	Not published at the time of the original EAG EVA search
Smith et al., 2023 Title: Real-World Performance of Autonomously Reporting Normal Chest Radiographs in NHS Trusts Using a Deep-Learning Algorithm on the GP Pathway	Pre-print study publication	Red Dot (Behold.ai)	No	Study submitted by the company for the addendum and was not was available at the time of the original EAG EVA.
Tam et al., 2021 Augmenting lung cancer diagnosis on chest radiographs: positioning artificial intelligence to improve radiologist performance	Published study	Red Dot (Behold.ai)	Yes	Study excluded by Warwick Evidence in original EVA report with the following reason: *"Software eligible. Population referral route not reported. Includes CXRs with difficult to locate nodules and CXRs with no nodules. includes AI+clinician vs AI alone but is simulating what might happen if the AI alone was used as triage"* Included by Cedar for addendum as comparator and software are eligible with extractable data related to lung nodules.

Van Beek et al., 2023 Title: Validation study of machine-learning chest radiograph software in primary and emergency medicine	Published study	Lunit INSIGHT CXR (Lunit)	No	Not published at the time of the original EAG EVA search
Vasilev et al., 2023 Title: AI-Based CXR First Reading: Current Limitations to Ensure Practical Value	Published study	Lunit INSIGHT CXR (Lunit)	No	Not published at the time of the original EAG EVA search

3 Clinical evidence overview

Table 4 below summarises each of the five published studies relevant to the review.

Table 4: Studies selected by the EAG as the evidence base

Study name and location	Study design	Participants and referral route	Reference standard / Ground truth	Outcomes
Author: Niehoff et al., 2023 **Location:** Germany	**Study design:** Retrospective observational cohort study **Intervention:** AI-Rad Companion Chest X-ray (Siemens Healthineers) Version: VA23A **Comparator:** WR by two radiologists unaware of study	**No. patients:** 499 consecutive patients examined between August and September 2021 **Demographics** Age: 65.4 ± 17.0 (median: 67.6, range 22-97) **No. CXRs:** 499 **Referral route:** Not specified/unclear	Defined in a consensus by two radiologists using further images (additional radiographs, previous and/or follow-up CXR or CT scans if available). In 375/499 cases additional examinations and/or CT scans were available.	• Sensitivity and specificity of AI-Rad vs radiologist for the detection of lung lesions, consolidation, atelectasis, pneumothorax and pleural effusion.
Author: Smith et al., 2023 **Location:** UK	**Study design:** Retrospective observational cohort study **Intervention:** Red Dot (Behold.ai) V2.2 **Comparator:** Audit by independent radiologists	**No. patients:** 4,076 **Demographics:** Female: 2,205 Male: 1,870 Mean age: 62.1 **No.CXRs:** 4,654 radiographs collected between April and May 2023 **Referral route:** Referred by GP	NR	• NPV of the AI algorithm. • Discrepancy rate between the algorithm and auditing radiologists. • Time taken to report result

Study name and location	Study design	Participants and referral route	Reference standard / Ground truth	Outcomes
Author: Tam et al., 2021 **Location:** UK	**Study design:** Retrospective observational study **Intervention:** Red dot (Behold.ai). Version NR **Comparator:** Clinician review (consultant radiologists)	**No. patients:** NR **Demographics:** Tumour set: Female: 108 Male: 92 Age: Mean 72.6±10.4 (range 32-92) Control set: Female: 113 Male: 87 Age: Mean 61.8±15.6 (range 34-98) **No.CXRs:** 400 **Referral route:** Unclear	Established by a combination of the cancer registry database records, the electronic clinical record, and review of both subsequent and prior imaging.	• Standalone tumour classification performance for radiologists and AI algorithm (Accuracy, Sensitivity, Specificity, Precision, True Positives, False Positives, False negatives)

Study name and location	Study design	Participants and referral route	Reference standard / Ground truth	Outcomes
Author: Van Beek et al., 2023 **Location:** UK	**Study design:** Retrospective validation study **Intervention:** Lunit INSIGHT CXR (Lunit). Version 3.1.2.0 **Comparator:** Two chest radiologists	**No. patients:** NR **Demographics:** Primary care: 554 female, 438 male, mean age 60 years, range 13-96 years. ED: 474 female, 494 male, mean age 64 years, range 13-102 years. **No.CXRs:** Primary care: 1,046,ED: 1,072 initial, Primary care: 992, ED: 968 after poor quality CXR removed. Total: 1,960. **Referral route:** Primary care and emergency department (ED)	All radiographs were reviewed and annotated by two independent expert chest radiologists both with >20 years' experience, blinded to the original report and the annotations, to reach a consensus.	• Sensitivity and specificity of AI versus clinician in detecting 10 pathological findings • Accuracy defined as correctly identified cases/total number of cases

Study name and location	Study design	Participants and referral route	Reference standard / Ground truth	Outcomes
Author: Vasilev et al., 2023 **Location:** Russia	**Study design:** Combined multicentre retrospective case-control study and prospective validation study **Intervention:** Lunit INSIGHT CXR (Lunit) Version 3.110 **Comparator:** Clinician (Radiologist)	**No. patients:** Retrospective: 73 Prospective: 4,752 **Demographics:** **Retrospective:** Male: 30 Female: 42 Unknown: 1 **Prospective:** Male: 1,746 Female: 3,005 Unknown: 1 **No.CXRs:** Retrospective: 73 Prospective: 4,752 **Referral route:** Unclear	A subset of radiographs (378/4,752) were interpreted by three experts.	• Sensitivity, specificity and AUROC for AI and clinician. • Concordance rate for radiologists and AI.

Abbreviations – AUROC: Area under the receiver operating characteristic curve; CT: Computed tomography; CXR: Chest X-ray; NPV: Negative Predictive Value; NR: Not Reported; WR: Written Report

3.1 *Results from the evidence base*

Of the relevant published evidence found by the EAG in this addendum, there is comparative evidence available on the use of the following AI technologies: AI-Rad Companion Chest X-ray (Siemens Healthineers), Lunit INSIGHT CXR (Lunit), and Red Dot (behold.ai). There are three ongoing studies using Lunit INSIGHT CXR (Lunit) and two using qXR (Qure.ai) technology.

Outcomes of interest

The EAG engaged with the clinicians to discuss the suggested outcomes outlined in the final NICE scope and sought their input on key focus areas for the review. Based on their expertise, clinicians emphasised the importance of concordance between the intervention and comparator, as well as the diagnostic accuracy for both lung cancer and nodules. While turnaround time could be supported by published data, secondary care clinicians expressed that time to X-ray report, time to CT scan, and time to diagnosis were of lesser significance to them compared to policymakers. Another crucial outcome identified by clinicians was the ease of use and acceptability of the technology. They emphasised that for successful adoption in practice, the AI software must be user-friendly and easily integrated into existing workflows. The impact of the software's output on clinical decision-making and the number of false positives were highlighted as paramount considerations.

Clinicians felt that the interpretation of technical failure rate in the context of AI software was unclear but could represent any potential instances where the software might be unable to analyse an image. Regarding additional outcomes, clinicians were doubtful that there would be much published data on morbidity/mortality and health-related quality of life (HRQoL). Time to surgery, while not within the scope of the review, was deemed an interesting outcome, although its availability in the existing data was considered unlikely. Furthermore, one clinician stressed the need to consider ethical components within the outcomes. Adding AI into practice was seen as an additional intervention to standard protocols, potentially leading to unnecessary CT scans and exposing patients to radiation that could be avoided.

During discussions with the lay SCM who is a person with lung nodules, the EAG inquired about the outcomes within the scope of the review that would be of most importance to patients in both existing and future studies focusing on AI for CXRs in lung cancer diagnosis. Concordance between the intervention and the clinician was identified as a crucial factor in fostering trust in the AI software. Additionally, the timeframe for CXR, CT scans, or receiving a diagnosis took precedence, along with the number of undetected cancer cases and the patient's health-related quality of life (HRQoL). Patients expressed that HRQoL is frequently disregarded in trials, while improvements in turnaround time are likely to have a substantial positive impact on a patient's mental well-being and overall quality of life. Consequently, addressing these aspects was considered a priority for future research in this domain.

On the other hand, patients assigned less importance to the impact of software outputs on clinical decision-making, ease of use/acceptability among clinicians, and the number of individuals referred for CXR or CT scans or discharged from further testing. Similarly, false positives were viewed as less concerning to patients due to their perception of it serving as an additional precautionary measure that provides added certainty and acts as a safety net, even if it means exposing the patient to additional radiation.

Table 5 reports the outcomes relevant to the NICE addendum scope for each of the four published studies.

Table 5: Outcomes relevant to scope for included studies

Study	Accuracy	Sensitivity	Specificity	PPV	NPV	FDR	FOR	AUROC	Concordance between intervention and comparator	Time taken to report result
Niehoff et al., 2023	NR	For detection of lung lesions: AI: 0.83$^{\Delta}$ (0.28 at CS = 10) WR: 0.52	For detection of lung lesions: AI: 0.83$^{\Delta}$ (0.99 at CS = 10) WR: 0.98	For detection of lung lesions: AI: 0.38$^{\Delta}$ (0.80 CS = 10) WR:0.79	For detection of lung lesions: AI: 0.97$^{\Delta}$ (0.91 at CS = 10) WR: 0.94	For detection of lung lesions: AI: 0.62$^{\Delta}$ (0.20 at CS = 10) WR: 0.21	For detection of lung lesions: AI: 0.03$^{\Delta}$ (0.09 at CS = 10) WR: 0.06	AI: 0.867 WR: 0.750	50.3%*	NR
Smith et al., 2023	NR	NR	NR	NR	AI: 0.96 Clinician: NR	NR	NR	NR	0.77%•	AI: Mean 7.1 seconds (range 5.0-17.0 seconds) Clinician: Mean 3 hours 50 minutes

Study	Accuracy	Sensitivity	Specificity	PPV	NPV	FDR	FOR	AUROC	Concordance between intervention and comparator	Time taken to report result
Tam et al., 2021	Rad 1: 0.90 Rad 2: 0.87 Rad 3: 0.84 AI alone: 0.87 Rad 1 + AI: 0.91 Rad 2 + AI: 0.90 Rad 3 + AI: 0.91	Rad 1: 0.86 Rad 2: 0.79 Rad 3: 0.69 AI alone: 0.8 Rad 1 + AI: 0.94 Rad 2 + AI: 0.91 Rad 3 + AI: 0.89	Rad 1:0.94 Rad 2: 0.95 Rad 3: 0.99 AI alone: 0.93 Rad 1 + AI: 0.88 Rad 2 + AI: 0.90 Rad 3 + AI: 0.92	Reported as 'True positives': Rad 1: 171 Rad 2: 157 Rad 3: 136 AI alone: 159 Rad 1 + AI: 186 Rad 2 + AI: 180 Rad 3 + AI: 176	Reported as 'False negatives': Rad 1: 27 Rad 2: 41 Rad 3: 62 AI alone: 39 Rad 1 + AI: 12 Rad 2 + AI: 18 Rad 3 + AI: 22	Reported as 'False positives': Rad 1:12 Rad 2: 9 Rad 3: 1 AI alone: 14 Rad 1 + AI: 23 Rad 2 + AI: 20 Rad 3 + AI: 15	Reported as 'Precision': Rad 1: 0.93 Rad 2: 0.95 Rad 3: 0.99 AI alone: 0.92 Rad 1 + AI: 0.89 Rad 2 + AI: 0.90 Rad 3 + AI: 0.92	NR	Overall combined radiologist + AI concordance: 92%	NR
Van Beek et al., 2023	For detection of urgent lung nodules: AI: 0.8653 (95% CI: (0.8494-0.8801)	For detection of lung nodules: AI: **ED** CXR: 0.794 (95% CI: 0.621-0.913), **GP** CXR: 0.833 (95% CI: 0.653-0.944)	For detection of lung nodules: AI: **ED** CXR: 0.848 (95% CI: 0.823-0.87) **GP** CXR: 0.886 (95% CI: 0.864-0.905)	NR	NR	NR	NR	For detection of lung nodules: AI: **ED**: 0.881 (95% CI: 0.814-0.949), **GP**: 0.905 (95% CI:0.84-0.97)	NR	NR

Study	Accuracy	Sensitivity	Specificity	PPV	NPV	FDR	FOR	AUROC	Concordance between intervention and comparator	Time taken to report result
Vasilev et al., 2023		**Retrospective:** AI: 0.9 (95% CI: 0.79-1.0) Clinicians: 0.9 (95% CI: 0.79-1.0) **Prospective:** AI: 0.77 (CI: 0.73-0.80) Clinicians: 0.86 (95% CI: 0.82-0.91)	**Retrospective:** AI: 0.89 (95% CI: 0.79-0.98) Clinicians: 0.95 (95% CI: 0.89-1.0) **Prospective:** AI: 0.81 (CI: 0.80-0.82) Clinicians: 0.92 (95% CI: 0.88-0.96)	NR	NR	NR	NR	**Retrospective:** AI: 0.94 (95% CI: 0.87-1.0) Clinicians: 0.97 (95% CI: 0.94-1.0) **Prospective:** AI: 0.84 (95% CI: 0.82-0.86) Clinicians: 0.89 (95% CI: 0.86-0.92)	**Retrospective:** 86% **Prospective:** 81%	NR

Abbreviations: AUROC: Area under the receiver operating characteristic curve; CI: Confidence interval; CS: Confidence score; ED: Emergency Department ; FDR: False Discovery Rate; FOR:False Omission Rate; NPV: Negative Predictive Value; NR: Not reported; PPV: Positive predictive value; Rad: Radiologist; WR: Written report

*Concordance for all pathologies, not just lung lesions

$^{\Delta}$CS ≥6

•Discrepancy rate defined as the proportion of all processed exams that were incorrectly classified as HCN according to auditing radiologists

4 Clinical evidence review and critical appraisal

Niehoff et al., 2023

The study aimed to evaluate the performance of the AI-Rad system in detecting lung lesions on chest X-rays compared to clinician evaluations. The AI-Rad system demonstrated promising results with high sensitivity and specificity, along with an excellent negative predictive value. However, it had a relatively high false discovery rate, indicating false-positive results. Increasing the confidence threshold improved the false discovery rate but significantly reduced sensitivity.

In comparison, the written reports by clinicians showed lower sensitivity but excellent specificity and negative predictive value. Overall, the AI-Rad system showed potential in assisting radiologists by accurately detecting lung lesions. However, there was a trade-off between reducing false positives and maintaining sensitivity.

It is worth noting that the study used a consensus agreement of two radiologists to create the written reports, which may not align directly with clinical practice. The inclusion of additional radiographs and CT examinations in the written reports may have also influenced the results.

The patient referral process was not clearly explained, and the retrospective enrolment of patients within a short timeframe raises some uncertainties. Additionally, the study focused solely on the detection of lung lesions and did not evaluate the AI-Rad system's performance with chest radiographs of poor image quality.

While the AI-Rad system has the potential to assist radiologists and increase their confidence, careful consideration is needed when setting confidence thresholds to balance false positives and sensitivity.

Smith et al., 2023

This study is a pre-print submitted to the EAG by the company. It reports on a retrospective study conducted in the UK aimed at evaluating the performance of the Red dot (Behold.ai) software, deployed as diagnostic decision support software, in

two NHS Trusts. The study sample consisted of all processed CXR data collected over a continuous six-week period from the two sites. The algorithm employed by the software assigned abnormality scores to classify the CXR images as high confidence normal (HCN) or abnormal. Examinations classified as HCN underwent review and assessment by independent radiologists. The software served as a rule-out test by providing clinical diagnosis reports for examinations classified as normal with high confidence, thus reducing the need for further human interpretation or intervention. The primary focus of the study was to evaluate the algorithm's performance in classifying HCN examinations within an active clinical pathway, aiming to assess its effectiveness as an automated diagnostic tool.

The results of the study demonstrated that the algorithm autonomously diagnosed 20% of the examined CXRs as normal, effectively decreasing the workload and alleviating pressure on radiology departments. However, during the audit conducted by independent radiologists, it was discovered that 36 abnormal examinations were incorrectly classified as normal. Although none of these misclassifications were deemed potentially serious, their presence emphasises the importance of cautious interpretation when relying on algorithmic results. However the audit results were promptly communicated to the trusts at a delivery rate of 99.3% within 24 hours, contributing to the overall confidence and trust among clinicians in the algorithm's performance.

While the study successfully reduced the turnaround time, its design had some limitations. The retrospective nature of the study and the utilisation of anonymised data limits the generalisability of the findings to real-time clinical practice. Moreover, the exclusion of lateral chest X-rays in the study raises concerns about the external validity of the findings, as these are commonly performed in clinical practice. Additionally, the study solely focused on HCN diagnoses and did not report on the abnormal findings, which constituted 80% of the analysed dataset. This lack of information introduces uncertainties regarding the algorithm's capability to detect abnormalities beyond normal findings.

In summary, this study demonstrated the Red dot (Behold.ai) software’s ability to autonomously and confidently diagnose a specific subset of CXRs as normal within real-world clinical pathways in the NHS. By effectively reducing the workload for

radiologists and expediting result delivery, the software demonstrated its potential to streamline the diagnostic process which aligned with the perspectives of the clinical experts consulted by the EAG.

Tam et al., 2021

Tam and colleagues conducted a study to evaluate the impact of the Red Dot algorithm (Behold.ai) in an augmented cancer-triaging pathway using a curated dataset of 400 CXRs, with 50% of cases positive for lung cancer. The AI algorithm served as the 'first reader' of the CXRs, identifying high confidence tumour (HCT) cases for immediate triage, while radiologists reviewed and reported the remaining CXRs as normal. The study retrospectively simulated the influence of AI on the subsequent pathway.

The findings revealed that AI-based triage has the potential to improve the accuracy and efficiency of tumour detection on CXRs. The standalone AI algorithm demonstrated performance equivalent to that of consultant radiologists, achieving an overall accuracy of 87%. However, the most substantial improvements were observed when the AI algorithm was integrated into the proposed triage pathway. This combined approach enhanced the performance of every radiologist, resulting in an average accuracy, sensitivity, and specificity of 90.67%, 91.33%, and 90%, respectively. It is important to note that the combined approach also resulted in an increase in false-positive results, with an overall precision of 92%.

The study has several limitations. The study was a retrospective design simulating the use of AI in a clinical workflow. The referral route was unclear, as radiographs were selected from a seven-year period of lung cancers. Such selection of cases potentially introduces selection bias as the characteristics of those referred through primary care may differ to those referred via other routes. The dataset's composition of 50% tumours was known by radiologists which may have biased the interpretation and impacts the study's external validity. Additionally, upon radiologist review, the absence of patient history and clinical presentation information does not reflect real-world practice, but is arguably more of a direct comparison of the AI technology and radiologist review as AI software does not incorporate either. The study's findings on increased false positives with the AI software also raise concerns regarding the

impact on subsequent services such as CT, primary care and follow-up. Furthermore, it is important to note the potential conflict of interest among the study authors, with stock ownership in Behold.ai.

Van Beek et al., 2023

This retrospective validation study aimed to assess the performance of the Lunit INSIGHT CXR (Lunit) machine learning algorithm in analysing chest radiographs (CXRs) compared to expert chest radiologists. A total of 1,960 consecutive CXRs from primary care referrals and the emergency department were collected from a UK hospital in 2015. The reference standard was established by two independent chest radiologists who reviewed all the images. The algorithm's performance was evaluated in detecting various pathologies including atelectasis, fibrosis, calcification, consolidation, lung nodules, cardiomegaly, mediastinal widening, pleural effusion, pneumothorax, and pneumoperitoneum.

However, the study utilised a retrospective design and relied on historical clinical cases, which may not fully reflect real-world clinical practice and the population is likely to differ to those referred from primary care. Furthermore, the exclusion of poor quality CXRs may introduce a bias in the results, as poor quality images are commonly encountered in routine clinical scenarios. Additionally, while the study employed independent expert radiologists to establish a consensus reference standard, there is a lack of direct comparison between the AI algorithm and the clinicians. Moreover, although the study reports results on lung nodules, this made up only 7.5% of the findings within the GP cohort and the omission of cancer as an outcome limits the applicability of the findings to this review.

Vasilev et al., 2023

This study aimed to assess the effectiveness of Lunit INSIGHT CXR AI software in analysing CXRs through a combined retrospective and prospective evaluation. The retrospective evaluation involved 160 radiologists who interpreted 73 cases from a locally collected test set obtained from outpatient hospitals in Moscow. The AI software independently analysed the CXR images and its diagnostic accuracy was compared to that of the radiologists. The AI demonstrated comparable performance

to the radiologists in most aspects of the retrospective evaluation, with no statistically significant differences observed.

In the prospective evaluation, CXRs from 107 inpatient and outpatient departments in Moscow were processed by an on-stream AI system. This evaluation included 4752 cases, which were then compared to the reports provided by 226 radiologists. To establish consensus, a subset of prospective cases (378/4752) was interpreted by three expert radiologists. The AI's performance in the prospective evaluation was slightly lower than that in the retrospective evaluation. The authors claim that the decrease in accuracy was primarily due to clinically insignificant false-positive findings and the AI's inability to detect specific abnormalities (such as ‘opacity’, ‘nodule’, and calcification) that were reported by human radiologists.

Although this study was the only prospective study, there are several limitations. Firstly, the retrospective evaluation used a small sample size of only 73 cases from a Russian population, potentially impacting the generalisability of the findings. Additionally, the referral route for the prospective cases was unclear, and the study included a mixed population of inpatient and outpatient cases. Notably, 621 cases with lateral viewpoint or patient rotation were excluded, limiting the analysis to unambiguous cases which are routinely encountered in clinical practice. Also the software was capable of detecting ten pathological conditions, not including lung cancer/lesions, although it could identify lung nodules, however no subgroup analysis of the software's diagnostic accuracy specifically for interpreting lung nodule data was reported.

5 Technical failures

In the original EAG EVA (Stinton et al., 2023), none of the reviewed studies provided information regarding technical failures of AI-derived software. Similarly, in this addendum, no studies included in the analysis reported on technical failures, however several studies excluded CXRs that were not taken at an anteroposterior angle due to the inability of the AI software to interpret the CXR.

The EAG for the addendum also conducted a thorough search of the Public Access Registration Database (PARD), Medicines and Healthcare products Regulatory Agency (MHRA), and Manufacturer and User Facility Device Experience (MAUDE)

databases, specifically focusing on the use of the 14 AI technologies. The search was performed on 25/05/2023, and no reports of technical failures or adverse events related to the use of these technologies were found.

6 Evidence synthesis

Five studies were included in this addendum: three from the UK, one from Germany, and one from Russia. All five studies were retrospective cohort studies, although one of them also included a prospective diagnostic accuracy study. The AI software used in the studies included two studies using Lunit INSIGHT CXR (Lunit), two using Red Dot (Behold.ai), and another study using AI-Rad Companion (Siemens Healthineers). The combined population across all studies consisted of approximately 10,792 patients. Regarding the referral routes, three were unclear but one study involved referrals from GP and aligned with the scope of this review. Another study had a mixed population comprising both primary care and emergency department referrals.

In three of the studies, the ground truth was established through independent review by radiologists and another used a combination of registry data, clinical records and additional imaging. The primary outcomes assessed in all five studies focused on the diagnostic accuracy of the AI software, including accuracy, sensitivity and specificity. One study, conducted by Niehoff et al. in 2023, additionally reported values for positive predictive value (PPV), negative predictive value (NPV), false discovery rate (FDR), and false omission rate (FOR). Tam et al., 2021 reported data on both Clinician and AI software precision. Four of the studies reported the concordance rate between the AI software and clinicians, while one study specifically examined the time taken to report results. It is worth noting that all included studies were either evaluation studies or diagnostic accuracy studies, providing limited or no data on technical failure rates, ease of use/acceptability, the impact of false positives on workflow, health-related quality of life, and morbidity and mortality rates.

7 Interpretation of the clinical evidence

The interpretation of the evidence from the included studies provides valuable insights into the performance and diagnostic accuracy of three AI software systems for chest x-ray analysis; AI-Rad Companion Chest X-ray (Siemens Healthineers),

Lunit INSIGHT CXR (Lunit), Red dot (Behold.ai). Niehoff et al. focused on multiple pathologies but reported separate diagnostic accuracy for lung lesions, showing that AI-Rad had high sensitivity and specificity, albeit with a high false discovery rate for false positive results. In comparison, clinicians' written reports demonstrated lower sensitivity but excellent specificity and negative predictive value.

Smith et al. used the Red dot (Behold.ai) technology as a rule-out test within an active clinical pathway, effectively reducing clinicians' workload and enabling better prioritisation of abnormal cases. However, an audit of the high confidence normal (HCN) group revealed some abnormal examinations that were incorrectly reported as normal by the AI, mostly due to bone abnormalities or sub-optimal images.

Similarly, Tam et al., 2021 used Red Dot (Behold.ai) technology as the ‘first reader’ of the CXR in two retrospective datasets of confirmed lung tumour and normal radiographs in a simulated referral pathway. The study reported outcomes related to diagnostic accuracy for Clinician alone, AI alone, and combined Clinician + AI alone. Both Clinician alone and AI alone had similar accuracy in detection of tumours, but accuracy improved when combined.

Van Beek's validation study of Lunit INSIGHT CXR only had a small proportion of results relevant to the scope of this review, as it focused on multiple pathologies. The study did not compare the accuracy of clinicians versus AI alone, instead analysing two datasets separately: GP and ED referrals. Similarly, the Vasilev study using Lunit INSIGHT CXR found a higher rate of false positives with the AI software, although this was on a prospective dataset.

Several clinical experts raised concerns about the potential for false positive results, which could impact workflow and increase the need for CT scans. However, from the published studies, the high false discovery rate often resulted from the identification of benign pre-existing conditions that were of minimal concern to radiologists. This limitation highlights the current inability of AI software to consider the full medical history and previous scans, presenting a trade-off between reducing false positives and maintaining sensitivity.

The studies also had limitations, such as excluding lateral view CXRs and mostly analysing unambiguous radiographs, which raises questions about the external

validity and utility of these AI algorithms in routine clinical practice. Additionally, none of the available evidence allowed for subgroup analysis based on ethnicity, age, or socio-economic status, highlighting the need for future research to address these factors.

Overall, while the studies provide valuable insights, there is still a lack of evidence specifically focused on lung cancer/lesions and lung nodules. Future research should aim for comparative studies that assess technical failure rates, workflow impact, patient-related outcomes, and long-term outcomes such as mortality and morbidity to comprehensively evaluate the potential of AI software in clinical practice.

8 Integration into the NHS

In current practice, chest X-rays are typically performed and then reviewed by radiologists after a period of several hours or even days. They are subsequently reported either to a lung multidisciplinary team (MDT) or back to the GP. Some centers have implemented a protocol where individuals promptly flag a CXR for radiologists to examine. However, these individuals are often inexperienced and not reporting radiographers. Consequently, they can detect large, obvious cancers on scans, but struggle with identifying subtle, smaller cancers that require more time and pose greater difficulty. These more difficult cancers are the ones that radiologists feel could be the ones where AI would be most helpful in identifying.

One clinical expert, who has previously used Auto Lung Nodule Detection (Samsung) software in a research capacity involving approximately 6,000 patients, has submitted a paper that is currently undergoing peer review. During this research, the software yielded a significant number of false positives, raising concerns about the increased incidence of unnecessary and costly CT scans, which could strain limited resources. The expert plans to evaluate the same dataset with ten other AI software, most of which fall within the NICE scope for this review. However, the clinician expressed skepticism about finding an AI software that would offer substantial clinical value.

One clinician who has experience with AI software for CT scans in clinical practice but not for CXRs, spoke specifically about how best to ensure successful adoption of such software in the NHS. They attributed this success to effective implementation

strategies, which included the presence of local champions and proactive engagement with stakeholders beforehand. The clinician emphasised the significance of on-the-ground support throughout the rollout process. They highlighted key elements of a successful rollout, such as involving all stakeholders, appointing a local champion to advocate for and support the software's adoption, and establishing a regular feedback loop between the AI software company and the clinical team. This feedback mechanism allows clinicians to share their perspectives on the software's strengths and weaknesses, while also encouraging the company to incorporate the feedback and provide statistical summaries of its usage within the department or trust.

However, all clinicians acknowledged that training is a challenge with these technologies, as many of them are still works in progress. Additionally, the versions used for training may differ from the ones in practice, and upgrades often lack accompanying training, leading to inadequate understanding of the technology and difficulty in interpreting results.

Discussions centered around the fact that when using AI software trained on datasets derived from clinical practice, it does not provide a direct comparison. One clinician highlighted the existence of a ‘red spot’ test during a radiologist's training, consisting of 30-50 plain radiographs with a mixture of abnormal and normal cases, including both easy and challenging scans for interpretation. Trainees analyse these images to achieve maximum accuracy. The clinician proposed a similar approach for CXR interpretation with AI software, suggesting that the integration of AI technology into practice could formalise this ‘red spot’ training, ensuring urgent findings are promptly flagged for radiologists.

Clinicians expressed concerns regarding the issue of accountability if something goes wrong when relying solely on AI. They also discussed data-related challenges, as many companies require data to be uploaded to an external, cloud-based server for interpretation before being sent back to the hospital. While individual patient permission is sought within a research context, it is not a common practice in routine healthcare settings.

A clinician who had responded to the questionnaire commented that they could see AI alone as a ‘first read’ step in the pathway to try and improve workflow/prioritisation. Another clinician responded that when integrating an AI software into the NHS, it is important that it integrates into current radiology systems, which all differ depending on the hospital or trust.

Definition of ‘AI alone’ and ‘Clinician + AI’

Clinical experts highlighted the significant workload and pressures faced by their departments and the national workflow shortage of experienced radiologists. This they felt explains the necessity for AI technology in this context. When asked about the practicality of ‘Clinician + AI’ collaboration and whether clinicians could retrospectively review a subset of radiographs, none of the clinicians considered it realistic. Regarding the integration of AI into clinicians' workflow, one clinician noted that most technologies involve clinicians examining the CXR first, forming an opinion, and then consulting the algorithm. However, in practice, clinicians tend to use the algorithm first and then assess its agreement with their own evaluation.

Clinicians emphasised that AI software that prioritises images as likely normal or abnormal could be perceived as AI acting autonomously, but in reality, a clinician still reviews the results. For instance, one model mentioned by a clinician identifies a high likelihood of normality and generates a notification within the hospital's picture archiving and communication system (PACS). This can be seen as independent practice. Abnormal cases are placed in the standard reporting pile, creating mostly with the urgency of those cases unknown until they are reviewed. Clinicians suggested that AI alone could potentially triage critical results, aiding workload prioritisation.

Clinicians discussed the benefits of using AI in their practice, particularly in efficiently handling normal radiographs, reducing fatigue, and allowing them to address the backlog of cases more quickly. The ability of AI to highlight findings for review was seen as advantageous. One clinician who had previous experience with AI technologies for CT scans mentioned that it eliminated the need for manual measurements on radiographs, as the technology performed these tasks automatically. However, they cautioned that clinicians can become biased by the algorithms, potentially focusing on less important findings and overemphasising

irrelevant results. While AI as an adjunct to clinician review was considered a safety net, relying on AI alone was deemed a high-risk "pipedream" because CXRs still require clinical assessment. One clinician pointed out that AI alone without a clinician reviewing the CXR is impossible, as radiographers are responsible for examining the scans as part of the scanning process. Clinicians envisioned AI working alongside clinicians during the early stage of patient scans by integrating AI software into radiographers' practice. This integration could aid prioritisation and expedite the identification of urgent cases for radiologists.

When discussing the possibility of AI integration into the NHS involving retrospective review, all clinicians expressed consensus that it is highly unlikely and unrealistic. They cited their full capacity, time constraints, and lack of resources for quality assurance/audit processes, except within the context of funded research studies specifically dedicated to auditing.

One clinician emphasised the importance of intensive discussions between AI technology providers and organisations like NHS Digital when considering software integration into the NHS. They highlighted the challenges they face in implementing new software in their trust due to data protection risks. Engaging with NHS Digital and obtaining agreement for seamless integration with necessary support would facilitate the process. The clinician stressed the significance of involving not only clinical teams but also IT departments responsible for installation and in-house maintenance. NHS Digital was regarded as a means to gain evidence in many UK trusts but often posed obstacles that hindered progress and dampened interest. This slowed down the adoption of novel technologies such as AI and contributed to the lack of real-world evidence.

In terms of funding these technologies, clinicians suggested the need for a regional or nationally agreed funding structure for effective rollout. Often, companies offer free trials of software, but once it becomes embedded in clinical practice, trusts are faced with substantial bills for continued usage, making payment approval challenging. Consequently, software may be removed, creating difficulties for clinicians who have grown accustomed to its usefulness and must revert to their previous software or methods.

One clinician proposed that the NHS should adopt the perspective that, instead of waiting for a finished AI product, there should be a framework to support research that allows clinical teams to contribute to shaping and developing the software. They suggested that NICE could provide a research framework for AI development, incorporating benchmarking and quality assurance.

Perception of AI software

One clinician raised concerns regarding the design of trials conducted with AI software, stating that they are generally poorly designed and involve mostly start-ups or larger companies with subgroups. Clinicians further noted that while many algorithms have been tested on large retrospective datasets, these datasets often exclude CXRs with factors such as patient rotation, poor image quality, obesity, or complex comorbidities—factors that are commonly encountered in clinical practice. As clinicians are required to analyse every CXR, the usefulness of AI software diminishes when it struggles or fails to assist in interpretation of more challenging cases. Additionally, even CXRs labelled as 'normal' in general practice may reveal underlying issues.

Another clinician remarked that current AI technologies focus solely on standalone CXRs at a single timepoint, which does not replicate the comprehensive clinical practice. Incorporating a patient's previous imaging and medical history, a standard practice during CXR interpretation, is crucial for understanding and contextualising findings. Software often detects pre-existing, benign abnormalities that the clinician may already be aware of and not be of concern.

Clinicians expressed concerns about the potential for AI technologies to flag numerous insignificant findings, shifting the problem from a pile of CXRs to a pile of more costly and time-consuming CT scans. They emphasised the need for research to investigate the impact of AI technologies on clinical workflow.

Regarding the development stage of AI technologies, one clinician likened them to a trainee doctor in a constant learning phase: *"Current technologies resemble early-stage 'trainee radiologist registrars' that require supervision and tend to be overly cautious in challenging cases. However, a well-established AI software would*

incorporate user feedback, provide accurate interpretations, and instil trust without extensive supervision—similar to a senior radiology registrar."

During discussions with the lay SCM, when asked how they felt about retrospective study designs not seeking informed consent from the patients, they commented: *"people are not overly concerned by this. There is an element of trust that the data is not being abused and an understanding that to do research, you have to have data and so there is support for it, providing personal details are not used".* There were also discussions around the use of data such as CXRs being sent to third party organisations external to the NHS for analysis during research. They commented that this is more of an issue, especially if it goes outside of the UK, and if this is the case, the patient must be made aware.

From the patient perspective, the lay SCM commented that people are likely to endorse the use of AI software in their care if there is a benefit to them such as a quicker turnaround time as there is a lot of anxiety felt by patients while waiting for scans. If there were to be an AI software that prioritises scans based on potential urgency, it is very important that this is discussed with the patient so that they can manage expectations. If it means that a longer wait is likely to mean it is less urgent, this might help alleviate some of the anxiety around waiting for the result. The lay member raised concerns that without effective communication between the system and the patient, prioritisation of scans may increase anxiety for patients. The biggest concern for the lay SCM was around trusting AI technologies, but once it can be trusted, people are likely to prefer it.

Special considerations, including issues related to equality

The following are potential considerations associated with the autonomous use of AI-derived software (AI Alone) to analyse chest X-rays for suspected lung cancer:

Bias in training data: AI algorithms rely on large datasets for training, and if the data used to develop these algorithms is biased, it can lead to disparities in accuracy and outcomes across different populations. It is important to ensure that the training data used for AI models is diverse and representative of various demographic groups to mitigate bias.

Disparities in access to AI technology: There may be inequalities in access to AI technology, such as limited availability in certain healthcare settings or regions/UK nations. It is crucial to address these disparities to ensure that all patients, regardless of their socioeconomic status or geographic location, have equal access to accurate and timely lung cancer diagnoses.

Differential performance across population subgroups: AI algorithms may perform differently across various demographic groups due to differences in the data used for training or variations in disease characteristics. It is essential to evaluate the algorithm's performance across diverse populations to avoid potential disparities in diagnostic accuracy.

7.2 Ongoing Studies

Three ongoing studies have been identified; one in India (Jagirdar et al., 2023) and two in the United Kingdom (Avery et al., 2022; Baldwin et al., 2023). The Indian study is a retrospective cross-sectional observational study that aims to assess the accuracy of an AI algorithm in interpreting chest X-rays. It uses the qXR (Qure.ai) deep learning system to identify various abnormalities and compares the AI-generated reports with those provided by radiologists. The study focuses on validating algorithms for detecting abnormal CXRs, particularly targeting the detection of pulmonary nodules.

The study by Avery and colleagues is a UK-based study evaluating the impact of the Lunit INSIGHT CXR (Lunit) AI system on CXR reporting through a predominantly prospective design with an initial retrospective component. It compares the performance of the AI system with clinician review in detecting abnormalities and explores its influence on radiological reports, recommended imaging, and patient management. The study population consists of patients aged 16 or older who underwent chest X-rays either requested by GPs or performed in the Emergency Department radiology unit, with potential subgroup analysis for primary care referrals. The primary outcomes of both studies aim to evaluate the accuracy of AI detection compared to clinicians in detecting abnormalities including lung nodules, but do not include lung cancer. The completion date for the Indian study is not specified, while the UK study is expected to be completed in 2023.

The second UK study, named the LungIMPACT study, led by Baldwin and colleagues is a prospective diagnostic multicentre randomised controlled trial. The study will use qXR technology to analyse CXRs from patients referred from primary care. The planned sample size is 150,000 patients to be recruited over a period of 12 months from eight NHS trusts in England. The primary objective of the study is to assess the effectiveness of AI ‘immediate read’ and worklist prioritisation for immediate review on the time to diagnosis of lung cancer the time to CT chest following abnormal CXR. All patients will have their CXR read by qXR software, the only difference between the intervention and non-intervention arms is the timing of the information from the AI. The main outcomes measure how AI assistance at the point of CXR acquisition and prioritisation for immediate review of CXRs referred

from primary care on the time to CT chest and the time to diagnosis of lung cancer. It will also assess the agreement between the AI algorithm and CXR reports for normal/abnormal decisions. The study will also provide data on the impact of software output on clinical decision making, turnaround time, and time to treatment. It is also the only ongoing study to assess lung cancer as an outcome. Recruitment for the study started on 01/05/2023 and is due to finish on 01/08/2023, with an intention to publish results on 31/05/2025. A protocol for the study is available via this link: https://www.isrctn.com/editorial/retrieveFile/f44a654e-9ad4-4a10-8784-707abbbf8c78/43324. All three ongoing studies are summarised in table 6.

Table 6: Summary of relevant ongoing studies

Study	AI software (manufacturer)	Study design	Status	Outcomes	EAG Comments	How the study could address the research need
CTRI/2020/08/027488 **Public title:** Use of artificial intelligence to interpret chest X-rays. **Scientific title:** Can Artificial Intelligence reliably report chest X-Rays? Radiologist validation of an algorithm trained on 2.3 million chest X-rays. **Country:** India.	qXR (Qure.ai)	**Design:** Retrospective cross-sectional observational study. **Intervention:** A deep learning system trained on 2.3 million chest X-rays and their corresponding radiology reports to identify abnormal X-rays and the following specific abnormalities: blunted costophrenic angle, cardiomegaly, cavity, consolidation, fibrosis, hilar enlargement, nodule, opacity and pleural effusion. **Comparator:** Radiologist reports. **Population:** ICD-10 Condition: J989 Respiratory disorder, unspecified; aged 16–99. Anonymised Chest X-rays with View: PA/AP, Patient Position: Erect, File format: Valid DICOM. **Setting:** Virtual study done using online servers.	Reported as *Not yet recruiting as of 08/06/2023*, but there is a pre-print from 2018 – see EAG Comments section. **Estimated completion date:** Not stated	**Primary outcome(s):** Retrospective assessment of 1. A three- radiologist majority on an independent, retrospectively collected set of 2000 anonymised chest X-rays (CQ2000); 2. The radiologist report on a separate validation set of anonymised 100,000 scans (CQ100k). The primary accuracy measure was area under the ROC curve (AUC), estimated separately for each abnormality as well as for normal versus abnormal scans. **Secondary outcome(s):** Validation of algorithms for detecting abnormal chest X-rays from normal ones.	Named commercial software (Qure.ai), unclear population and is a validation study. Qure.ai is the primary sponsor. Nature of support: Setting up of servers for study. There is a pre-print from 2018 (Putha et al., 2018). The trial is registered in 2020, marked as prospective.	Compares AI alone vs Clinician alone. AI trained to detect pulmonary nodules.

NCT05489471 **Brief Title:** A Study to Assess the Impact of an Artificial Intelligence (AI) System on Chest X-ray Reporting. **Official Title:** A Prospective Study to Assess the Impact of an Artificial Intelligence System on Reporting of Chest X-rays, Evaluate the Ability of AI Driven Worklists to Improve Reporting Times and Improve Same Day CT Pathway for Suspected Lung Cancer. **Country:** UK.	Lunit INSIGHT CXR (Lunit)	**Design:** Observational predominately prospective study with initial short retrospective component. **Intervention:** Artificial intelligence review. The AI looks for ten different abnormalities on each chest X-ray and produces a heat map and percentage confidence score if it detects an abnormality. **Comparator:** Clinician review. **Population:** Patients 16 years or older; posterior-anterior and Anterior-posterior chest radiographs requested by GP or performed in the Emergency Department radiology unit. **Setting:** Hospital (Hull University Teaching Hospitals NHS Trust).	Not yet recruiting as of August 2022 **Estimated completion date:** July 2023	**Primary outcome(s):** 1. For each finding present in the chest radiograph and/or the AI output, readers will record: Missed finding by AI, but detected by reporter; Correctly detected finding by AI; Missed finding by the reporter but detected by AI; Finding detected by AI but disputed by the reporter. 2. AI's impact on: Radiological report, Further recommended imaging; Altering patient management. **Secondary outcome(s):** Whether AI has increased confidence in reporting an abnormality or reporting a study as normal.	Primary care referrals or taken in the emergency department or inpatient.	Clinician alone vs AI alone (phase 1), Clinician alone vs Clinician + AI (phase 2).
ISRCTN78987039 **Trial acronym: LungIMPACT** **Brief title:** Impact of immediate AI enabled patient triage to chest	qXR (Qure.ai)	**Design:** Interventional randomised controlled trial (diagnostic) **Intervention:** Immediate reporting of AI CXR report alert	Reported as *Not yet recruiting* as of 17/05/2023. **Estimated completion date:** October 2024	**Primary outcomes:** 1. Time from chest X-ray to lung cancer diagnosis in days from the cancer waiting time database 2. Time from chest X-ray to CT (when performed) in days from the	Referral route is in scope and those not referred from primary care are excluded.	Study looks at Clinician + AI, but implementing the AI at different times during the scan. CXRs are randomised to be either checked by the radiographer

CT on the lung cancer pathway **Official Title:** Does triage of chest X-rays with artificial intelligence shorten the time to lung cancer diagnosis: a randomised controlled trial **Country:** UK		**Comparator:** Usual reporting of CXR report + AI **Population:** Patients 18 years or older who have been referred by their GP for an anteroposterior or posteroanterior CXR at the trial participating centres **Setting:** Hospital (Nottingham University Hospitals NHS Trust)		radiology information system. **Secondary outcomes:** 1. Time to first respiratory cancer outpatient appointment in days from the cancer waiting time database 2. Time to treatment start for lung cancer patients in days from the cancer waiting time database 3. Agreement between AI (qXR) and human readers for normal/abnormal interpretation of chest X-ray as an agree/disagree decision with discordance review by a thoracic radiologist where required 4. Number of urgent lung cancer referrals from the cancer waiting time database 5. The incidence of lung cancer from the cancer waiting time database 6. The stage of lung cancer diagnosis from the cancer waiting time database 7. Cost-effectiveness of AI support at the time of CXR acquisition and prioritisation for immediate review of		with qXR AI read, or without.

				CXRs; to be measured by difference in costs per patient diagnosed, per percentage increase in early-stage diagnosis and potentially per QALY subject to the availability of health utilities in the published studies.		

8.1 *Evidence gap analysis*

The EAG found no evidence that met the addendum scope and inclusion criteria for the following AI technologies: Annalise CXR (annalise.ai), Auto Lung Nodule Detection (Samsung), Chestlink Radiology Automation (Oxipit), Chestview (GLEAMER), Chest X-ray (Rayscape), InferRead DR Chest (Infervision), Milvue Suite (Milvue), SenseCare-Chest DR Pro (Sensetime), and VUNO Med-Chest X-ray (VUNO).

From the published evidence included in this addendum, there are no retrospective or prospective real-world studies looking specifically at any of the below outcomes with respect to lung cancer or lung nodules:

- Technical failure rate
- Impact of software output on clinical decision-making
- Number of people referred for a CT scan
- Number of people referred for follow-up X-ray
- Number of people identified as 'normal'/discharged
- Stage of cancer at detection
- Time to CT scan
- Time to diagnosis
- Ease of use/acceptability of the software by clinicians
- Morbidity
- Mortality
- Health-related quality of life

The ongoing study in India using qXR technology is likely to address the gap in diagnostic accuracy of the technology for analysing abnormalities, including lung nodules. However, it is important to note that as the study is conducted in India, it may not be directly applicable to the UK healthcare setting. The UK study using Lunit INSIGHT CXR (Lunit) will provide additional diagnostic accuracy evidence for the technology to detect lung nodules, but does not look to address any more of the outcomes within the addendum scope that have not already been addressed by published research on the technology. Furthermore, both studies primarily rely on

retrospective data, which may not accurately represent the patient population referred from primary care. There is also an absence of information in the study protocols regarding subgroup analysis based on factors such as ethnicity, age, sex, or socio-economic status.

The second UK study using qXR (Qure.ai) by Baldwin and colleagues is likely to address many of the outcomes not yet addressed by the published evidence or ongoing studies. This includes the following:

- Time to diagnosis
- Time to X-ray report
- Time to CT scan
- Turnaround time (time from start of image review to radiology report)
- Impact of software output on clinical decision-making
- Impact of false positives on the workflow
- Number of people referred for a CT scan
- Number of people identified as normal/discharged
- Stage of cancer at detection

After all ongoing studies have completed and published, outcomes in the addendum scope that will remain unaddressed will include:

- Technical failure rate
- Number of people referred for follow-up X-ray
- Ease of use/acceptability of the software by clinicians
- Morbidity
- Mortality
- Health-related quality of life.

9 Conclusions

In conclusion, there are significant gaps evidence in comparing AI technologies and clinician review for diagnosing lung cancer using chest X-rays. While retrospective studies have shown some promise in terms of AI software's sensitivity and specificity in detecting lung lesions and nodules, the lack of real-world evidence and limited

focus on lung cancer pose challenges to its practical implementation. The presence of false positives and negatives further hampers the suitability of autonomous AI analysis, which could disrupt clinical workflows and lead to adverse patient outcomes, including unnecessary CT scans and missed cancer diagnoses. Therefore, integrating autonomous AI analysis into routine NHS clinical practice is currently unfeasible outside the context of research. Ongoing studies will begin to address recommendations made by clinical experts that these technologies are likely to be best placed as adjuncts to clinician review, but during the early radiographer stage for triaging urgent cases. If successful, such integration into the NHS has the potential to improve efficiency, reduce turnaround time, detect cancers quicker and at an earlier stage, and alleviate clinician workload, paving the way for broader adoption of AI technologies in the future.

10 References

Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S. W., Karthikesalingam, A., King, D., Ashrafian, H. & Darzi, A. Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *NPJ Digital Medicine*, 2021;4(65).

Avery, G. A Study to Assess the Impact of an Artificial Intelligence (AI) System on Chest X-ray Reporting. *ClinicalTrials.gov*, NCT05489471. 2022. Available at: https://clinicaltrials.gov/show/NCT05489471 (Accessed 15 May 2023).

Baldwin, D. Does triage of chest X-rays with artificial intelligence shorten the time to lung cancer diagnosis: a randomised trial. ISRCTN, ISRCTN78987039. 2023. Available at: https://trialsearch.who.int/Trial2.aspx?TrialID=ISRCTN78987039 (Accessed 14/06/2023).

Colquitt, J., Jordan, M., Court, R., Loveman, E., Parr, J., Ghosh, I., Auguste, P., Patel, M. & Stinton, C. Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer. Warwick Evidence. 2023.

Jagirdar, A. Use of artificial intelligence to interpret chest X-rays. *The Clinical Trials Registry - India (CTRI)*, CTRI/2020/08/027488. 2020. Available at: http://www.ctri.nic.in/Clinicaltrials/pmaindet2.php?trialid=44155 (Accessed 15 May 2023).

Niehoff, J. H., Kalaitzidis, J., Kroeger, J. R., Schoenbeck, D., Borggrefe, J. & Michael, A. E. Evaluation of the clinical performance of an AI-based application for the automated analysis of chest X-rays. *Scientific Reports*, 2023:13(3680).

Smith, J., Naunton Morgan, T., Williams, P., Malik, Q. & Rasalingham, S. Real World Performance of Autonomously Reporting Normal Chest Radiographs in NHS Trusts. 2023. Pre-print.

Tam, M. D. B. S., Dyer, T., Dissez, G., Morgan, T. N., Hughes, M., Illes, J., ... & Rasalingham, S. (2021). Augmenting lung cancer diagnosis on chest radiographs: Positioning artificial intelligence to improve radiologist performance. *Clinical Radiology*, *76*(8), 607-614.

Van Beek, E. J. R., Ahn, J. S., Kim, M. J. & Murchison, J. T. Validation study of machine-learning chest radiograph software in primary and emergency medicine. *Clinical Radiology*, 2023;**78**, pp.1-7.

Vasilev, Y., Vladzymyrskyy, A., Omelyanskaya, O., Blokhin, I., Kirpichev, Y. & Arzamasov, K. AI-Based CXR First Reading: Current Limitations to Ensure Practical Value. *Diagnostics*, 2023;13:1430.

11 Appendices

Appendix A – EAG Search strategy audit

Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer

Date	Database Name	Searcher	Total Number of records retrieved	Total number of records loaded into library (Duplicates not imported)
16/05/2023	Medline ALL + Filter	MK	534	529
16/05/2023	Embase Classic+Embase + Filter	MK	1,086	751
12/05/2023	Cochrane Library CDSR	MK	0	0
12/05/2023	Cochrane Library CENTRAL	MK	67	52
12/05/2023	Epistemonikos Systematic Reviews Broad Synthesis	MK	 133 6	 83 3
15/05/2023	ACM Digital Library Reviews Primary Studies #1 Primary Studies #2	MK	 20 509 1	 16 488 1
12/05/2023	PROSPERO	MK	46	46
15/05/2023	ICTRP #1 #2 #3	MK	 39 16 34	 39 14 24
			2,491	2,046
Total	**Manual deduplication: 1,908**			

Supplementary search (filter removed)

Date	Database Name	Searcher	Total Number of records retrieved	Total number of records loaded into library

				(Duplicates not imported)
25/05/2023	Medline ALL	MK	1,217	1,213
25/05/2023	Embase Classic+Embase	MK	2,527	1,842
			3,744	3,055
Total	**Manual deduplication: 2,873** **Total imported into the main library after automatic EndNote deduplication: 1,667**			

With the Medline and Embase filter: 1,908
Additional without the filter: 1,667
Total: 3,575

EAG Search Strategy

DATABASES

Medline with the filter
Ovid MEDLINE(R) ALL <1946 to May 15, 2023>

1 exp artificial intelligence/ or exp machine learning/ or exp deep learning/ or exp supervised machine learning/ or exp support vector machine/ or exp unsupervised machine learning/ 171989
2 ai.kf,tw. 44022
3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 145821
4 exp Neural Networks, Computer/ 58285
5 (neural network* or convolutional or CNN or CNNs).kf,tw. 99612
6 exp Diagnosis, Computer-Assisted/ 86689
7 Pattern Recognition, Automated/ 26471
8 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 35551
9 (support vector machine* or random forest* or black box learning).kf,tw. 41534
10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 [AI] 426876
11 exp Radiography, Thoracic/ 40604
12 X-Rays/ 31577
13 (((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).kf,tw. 67662
14 11 or 12 or 13 [CXR] 123339
15 10 and 14 [AI and CXR] 4251
16 limit 15 to "reviews (best balance of sensitivity and specificity)" 389
17 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 304963
18 (systematic* adj3 (review* or overview* or search or literature)).mp. 352947
19 17 or 18 487967
20 15 and 19 [AI and CXR and SRs] 47
21 16 or 20 [AI and CXR and Reviews / SRs] 402
22 exp Lung Neoplasms/ or Solitary Pulmonary Nodule/ 274186
23 ((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)).kf,tw. 336580
24 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 960
25 (sclc or nsclc).kf,tw. 67383
26 22 or 23 or 24 or 25 [Lung Cancer / Nodule] 409190
27 10 and 26 [AI and Lung Cancer / Nodule] 7345
28 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 304963
29 (systematic* adj3 (review* or overview* or search or literature)).mp. 352947
30 28 or 29 [SRs] 487967
31 27 and 30 [AI and Lung Cancer / Nodule and SRs] 109
32 10 and 14 and 26 [AI and CXR and Lung Cancer / Nodule] 753
33 AI-Rad Companion Chest X-ray*.kf,tw,in. 2
34 Annalise CXR*.kf,tw,in. 1
35 Auto Lung Nodule Detection*.kf,tw,in. 0
36 ChestView*.kf,tw,in. 0

37 (Chest X-Ray Classifier* or Quibim*).kf,tw,in. 54
38 CheXVision*.kf,tw,in. 0
39 (ClearRead Xray* adj2 Detect).kf,tw,in. 0
40 InferRead DR Chest*.kf,tw,in. 0
41 JLD-02K*.kf,tw,in. 0
42 Lunit INSIGHT CXR*.kf,tw,in. 8
43 Milvue Suite*.kf,tw,in. 0
44 ChestEye Quality*.kf,tw,in. 0
45 (qXR* or Qure*).kf,tw,in. 7219
46 (red dot* or behold*).kf,tw,in. 1123
47 SenseCare-Chest DR Pro*.kf,tw,in. 0
48 VUNO Med-Chest X-Ray*.kf,tw,in. 0
49 (X1* and Visionairy Health).kf,tw,in. 0
50 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 [Technology Names / Companies] 8405
51 50 and 14 [Technology Names / Companies and CXR] 71
52 50 and 26 [Technology Names / Companies and Lung Cancer / Nodules] 98
53 51 or 52 [Technology Names / Companies and CXR / Lung Cancer / Nodules] 152
54 21 or 31 or 32 or 53 1291
55 limit 54 to english language 1233
56 limit 54 to no language specified 0
57 55 or 56 1233
58 exp animals/ not humans.sh. 5121777
59 57 not 58 1227
60 limit 59 to (comment or editorial or letter) 10
61 59 not 60 1217
62 algorithm*.ti,kf. 64191
63 (algorithm* adj1 (learn* or automate* or detect* or treatment* or therap* or radiolog* or ai or dl or data or dataset* or base*)).ab. 58001
64 artificial intelligen*.ti,ab,kf. 34593
65 ai.ti,kf. 7111
66 machine learning*.ti,ab,kf. 86619
67 deep learn*.ti,ab,kf. 46817
68 convolutional neural network*.ti,ab,kf. 23183
69 automate*.ti. 45501
70 (automate* adj3 (system* or score* or software* or analysis* or analyse* or risk* or evaluat* or tool* or detect* or process*)).ab,kf. 39980
71 or/62-70 303809
72 61 and 71 534

Medline without the filter

Ovid MEDLINE(R) ALL <1946 to May 24, 2023>

1 exp artificial intelligence/ or exp machine learning/ or exp deep learning/ or exp supervised machine learning/ or exp support vector machine/ or exp unsupervised machine learning/ 172258
2 ai.kf,tw. 44209
3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 146698
4 exp Neural Networks, Computer/ 58378
5 (neural network* or convolutional or CNN or CNNs).kf,tw. 99955
6 exp Diagnosis, Computer-Assisted/ 86704

7 Pattern Recognition, Automated/ 26475
8 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 35625
9 (support vector machine* or random forest* or black box learning).kf,tw. 41680
10 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 [AI] 428040
11 exp Radiography, Thoracic/ 40602
12 X-Rays/ 31583
13 (((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).kf,tw. 67693
14 11 or 12 or 13 [CXR] 123376
15 10 and 14 [AI and CXR] 4255
16 limit 15 to "reviews (best balance of sensitivity and specificity)" 387
17 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 305475
18 (systematic* adj3 (review* or overview* or search or literature)).mp. 353697
19 17 or 18 488839
20 15 and 19 [AI and CXR and SRs] 47
21 16 or 20 [AI and CXR and Reviews / SRs] 400
22 exp Lung Neoplasms/ or Solitary Pulmonary Nodule/ 274312
23 ((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)).kf,tw. 336978
24 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 960
25 (sclc or nsclc).kf,tw. 67524
26 22 or 23 or 24 or 25 [Lung Cancer / Nodule] 409605
27 10 and 26 [AI and Lung Cancer / Nodule] 7368
28 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 305475
29 (systematic* adj3 (review* or overview* or search or literature)).mp. 353697
30 28 or 29 [SRs] 488839
31 27 and 30 [AI and Lung Cancer / Nodule and SRs] 109
32 10 and 14 and 26 [AI and CXR and Lung Cancer / Nodule] 753
33 AI-Rad Companion Chest X-ray*.kf,tw,in. 2
34 Annalise CXR*.kf,tw,in. 1
35 Auto Lung Nodule Detection*.kf,tw,in. 0
36 ChestView*.kf,tw,in. 0
37 (Chest X-Ray Classifier* or Quibim*).kf,tw,in. 54
38 CheXVision*.kf,tw,in. 0
39 (ClearRead Xray* adj2 Detect).kf,tw,in. 0
40 InferRead DR Chest*.kf,tw,in. 0
41 JLD-02K*.kf,tw,in. 0
42 Lunit INSIGHT CXR*.kf,tw,in. 8
43 Milvue Suite*.kf,tw,in. 0
44 ChestEye Quality*.kf,tw,in. 0
45 (qXR* or Qure*).kf,tw,in. 7227
46 (red dot* or behold*).kf,tw,in. 1127
47 SenseCare-Chest DR Pro*.kf,tw,in. 0
48 VUNO Med-Chest X-Ray*.kf,tw,in. 0
49 (X1* and Visionairy Health).kf,tw,in. 0
50 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 [Technology Names / Companies] 8417
51 50 and 14 [Technology Names / Companies and CXR] 71
52 50 and 26 [Technology Names / Companies and Lung Cancer / Nodules] 97

53 51 or 52 [Technology Names / Companies and CXR / Lung Cancer / Nodules] 152
54 21 or 31 or 32 or 53 1291
55 limit 54 to english language 1233
56 limit 54 to no language specified 0
57 55 or 56 1233
58 exp animals/ not humans.sh. 5123334
59 57 not 58 1227
60 limit 59 to (comment or editorial or letter) 10
61 59 not 60 1217

Embase with the filter
Embase Classic+Embase <1947 to 2023 Week 19>
1 exp artificial intelligence/ or exp machine learning/ 427451
2 ai.kf,tw. 61597
3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 177976
4 (neural network* or convolutional or CNN or CNNs).kf,tw. 121703
5 computer assisted diagnosis/ or computer assisted radiography/ 45381
6 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 48268
7 (support vector machine* or random forest* or black box learning).kf,tw. 52417
8 1 or 2 or 3 or 4 or 5 or 6 or 7 [AI] 593239
9 exp thorax radiography/ 238701
10 X ray/ 124436
11 (((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).kf,tw. 111054
12 9 or 10 or 11 [CXR] 393905
13 8 and 12 [AI and CXR] 6555
14 limit 13 to "reviews (best balance of sensitivity and specificity)" [AI and CXR and Reviews] 768
15 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 452584
16 (systematic* adj3 (review* or overview* or search or literature)).mp. 580491
17 15 or 16 763900
18 13 and 17 [AI and CXR and SRs] 141
19 14 or 18 [AI and CXR and Reviews / SRs] 798
20 exp lung tumor/ or lung nodule/ 520743
21 ((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)).kf,tw. 516454
22 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 1364
23 (sclc or nsclc).kf,tw. 123593
24 20 or 21 or 22 or 23 [Lung Cancer / Nodule] 684899
25 8 and 24 [AI and Lung Cancer / Nodule] 14804
26 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 452584
27 (systematic* adj3 (review* or overview* or search or literature)).mp. 580491
28 26 or 27 [SRs] 763900

29 25 and 28 [AI and Lung Cancer / Nodule and SRs] 384
30 8 and 12 and 24 [AI and CXR and Lung Cancer / Nodule] 1246
31 AI-Rad Companion Chest X-ray*.kf,tw,in. 2
32 Annalise CXR*.kf,tw,in. 1
33 Auto Lung Nodule Detection*.kf,tw,in. 0
34 ChestView*.kf,tw,in. 0
35 (Chest X-Ray Classifier* or Quibim*).kf,tw,in. 67
36 CheXVision*.kf,tw,in. 0
37 (ClearRead Xray* adj2 Detect).kf,tw,in. 0
38 InferRead DR Chest*.kf,tw,in.0
39 JLD-02K*.kf,tw,in. 0
40 Lunit INSIGHT CXR*.kf,tw,in. 9
41 Milvue Suite*.kf,tw,in. 0
42 ChestEye Quality*.kf,tw,in. 2
43 (qXR* or Qure*).kf,tw,in. 15149
44 (red dot* or behold*).kf,tw,in. 1587
45 SenseCare-Chest DR Pro*.kf,tw,in. 0
46 VUNO Med-Chest X-Ray*.kf,tw,in. 0
47 (X1* and Visionairy Health).kf,tw,in. 0
48 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 [Technology Names / Companies] 16814
49 48 and 12 [Technology Names / Companies and CXR] 286
50 48 and 24 [Technology Names / Companies and Lung Cancer / Nodules] 253
51 49 or 50 [Technology Names / Companies and CXR / Lung Cancer / Nodules] 501
52 19 or 29 or 30 or 51 2697
53 limit 52 to english language 2602
54 limit 52 to no language specified 1
55 53 or 54 2603
56 animal experiment/ not (human experiment/ or human/) 2577488
57 55 not 56 2592
58 limit 57 to (editorial or letter) 74
59 57 not 58 2518
60 algorithm*.ti,kf. 83566
61 (algorithm* adj1 (learn* or automate* or detect* or treatment* or therap* or radiolog* or ai or dl or data or dataset* or base*)).ab. 78005
62 artificial intelligen*.ti,ab,kf. 42697
63 ai.ti,kf. 8946
64 machine learning*.ti,ab,kf. 105471
65 deep learn*.ti,ab,kf. 56062
66 convolutional neural network*.ti,ab,kf. 28282
67 automate*.ti. 61974
68 (automate* adj3 (system* or score* or software* or analysis* or analyse* or risk* or evaluat* or tool* or detect* or process*)).ab,kf. 60829
69 or/60-68 394309
70 59 and 69 1086

Embase without the filter

Embase Classic+Embase <1947 to 2023 Week 20>
1 exp artificial intelligence/ or exp machine learning/ 429519
2 ai.kf,tw. 61945

3 ((artificial or machine or deep) adj5 (intelligence or learning or reasoning)).kf,tw. 179222
4 (neural network* or convolutional or CNN or CNNs).kf,tw. 122210
5 computer assisted diagnosis/ or computer assisted radiography/ 45399
6 ((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).kf,tw. 48420
7 (support vector machine* or random forest* or black box learning).kf,tw. 52626
8 1 or 2 or 3 or 4 or 5 or 6 or 7 [AI] 595641
9 exp thorax radiography/ 239084
10 X ray/ 124597
11 (((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).kf,tw. 111225
12 9 or 10 or 11 [CXR] 394483
13 8 and 12 [AI and CXR] 6578
14 limit 13 to "reviews (best balance of sensitivity and specificity)" [AI and CXR and Reviews] 768
15 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 453907
16 (systematic* adj3 (review* or overview* or search or literature)).mp. 582547
17 15 or 16 766250
18 13 and 17 [AI and CXR and SRs] 140
19 14 or 18 [AI and CXR and Reviews / SRs] 798
20 exp lung tumor/ or lung nodule/ 521731
21 ((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)).kf,tw. 517349
22 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumor* or tumour* or syndrome*)).kf,tw. 1365
23 (sclc or nsclc).kf,tw. 123835
24 20 or 21 or 22 or 23 [Lung Cancer / Nodule] 686057
25 8 and 24 [AI and Lung Cancer / Nodule] 14873
26 (metaanalys* or meta analys* or NMA* or MAIC* or indirect comparison* or mixed treatment comparison*).mp. 453907
27 (systematic* adj3 (review* or overview* or search or literature)).mp. 582547
28 26 or 27 [SRs] 766250
29 25 and 28 [AI and Lung Cancer / Nodule and SRs] 385
30 8 and 12 and 24 [AI and CXR and Lung Cancer / Nodule] 1255
31 AI-Rad Companion Chest X-ray*.kf,tw,in. 2
32 Annalise CXR*.kf,tw,in. 1
33 Auto Lung Nodule Detection*.kf,tw,in. 0
34 ChestView*.kf,tw,in. 0
35 (Chest X-Ray Classifier* or Quibim*).kf,tw,in. 68
36 CheXVision*.kf,tw,in. 0
37 (ClearRead Xray* adj2 Detect).kf,tw,in. 0
38 InferRead DR Chest*.kf,tw,in. 0
39 JLD-02K*.kf,tw,in. 0
40 Lunit INSIGHT CXR*.kf,tw,in. 9
41 Milvue Suite*.kf,tw,in. 0
42 ChestEye Quality*.kf,tw,in. 2
43 (qXR* or Qure*).kf,tw,in. 15193
44 (red dot* or behold*).kf,tw,in. 1589
45 SenseCare-Chest DR Pro*.kf,tw,in. 0
46 VUNO Med-Chest X-Ray*.kf,tw,in. 0
47 (X1* and Visionairy Health).kf,tw,in. 0

48 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 [Technology Names / Companies] 16861
49 48 and 12 [Technology Names / Companies and CXR] 286
50 48 and 24 [Technology Names / Companies and Lung Cancer / Nodules] 254
51 49 or 50 [Technology Names / Companies and CXR / Lung Cancer / Nodules] 503
52 19 or 29 or 30 or 51 2708
53 limit 52 to english language 2612
54 limit 52 to no language specified 1
55 53 or 54 2613
56 animal experiment/ not (human experiment/ or human/) 2579406
57 55 not 56 2602
58 limit 57 to (editorial or letter) 75
59 57 not 58 2527

Cochrane Library CDSR

Date Run: 12/05/2023 10:04:35

ID Search Hits

#1 [mh "artificial intelligence"] OR [mh "machine learning"] OR [mh "deep learning"] OR [mh "supervised machine learning"] OR [mh "support vector machine"] OR [mh "unsupervised machine learning"] 2832

#2 ai:ti,ab,kw 5198

#3 ((artificial OR machine OR deep) NEAR/5 (intelligence OR learning OR reasoning)):ti,ab,kw 4296

#4 [mh "Neural Networks, Computer"] 518

#5 (("neural" NEXT network*) OR convolutional OR CNN OR CNNs):ti,ab,kw 1838

#6 [mh "Diagnosis, Computer-Assisted"] 2195

#7 [mh ^"Pattern Recognition, Automated"] 218

#8 ((automat* OR autonomous OR "computer aided" OR "computer assisted") NEAR/3 (detect* OR identif* OR diagnos*)):ti,ab,kw 2174

#9 (("support vector" NEXT machine*) OR ("random" NEXT forest*) OR "black box learning"):ti,ab,kw1008

#10 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 14531

#11 [mh "Radiography, Thoracic"]438

#12 [mh ^X-Rays] 76

#13 (((chest OR lung* OR thora*) NEAR/3 (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*):ti,ab,kw6134

#14 #11 OR #12 OR #13 6219

#15 #10 AND #14 142

#16 [mh "Lung Neoplasms"] OR [mh ^"Solitary Pulmonary Nodule"] 10362

#17 ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) NEAR/3 (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)):ti,ab,kw 29446

#18 ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") NEAR/4 (tumor* OR tumour* OR syndrome*)):ti,ab,kw 17

#19 (sclc OR nsclc):ti,ab,kw 12548

#20 #16 OR #17 OR #18 OR #19 30061

#21 #10 AND #20 390

#22 #15 OR #21 473

Cochrane Reviews 0

Cochrane Library CENTRAL

Date Run: 12/05/2023 10:23:29

ID	Search	Hits
#1	[mh "artificial intelligence"] OR [mh "machine learning"] OR [mh "deep learning"] OR [mh "supervised machine learning"] OR [mh "support vector machine"] OR [mh "unsupervised machine learning"]	2832
#2	ai:ti,ab,kw	5198
#3	((artificial OR machine OR deep) NEAR/5 (intelligence OR learning OR reasoning)):ti,ab,kw	4296
#4	[mh "Neural Networks, Computer"]	518
#5	(("neural" NEXT network*) OR convolutional OR CNN OR CNNs):ti,ab,kw	1838
#6	[mh "Diagnosis, Computer-Assisted"]	2195
#7	[mh ^"Pattern Recognition, Automated"]	218
#8	((automat* OR autonomous OR "computer aided" OR "computer assisted") NEAR/3 (detect* OR identif*OR diagnos*)):ti,ab,kw	1946
#9	(("support vector" NEXT machine*) OR ("random" NEXT forest*) OR "black box learning"):ti,ab,kw	1008
#10	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9	14355
#11	[mh "Radiography, Thoracic"]	438
#12	[mh ^X-Rays]	76
#13	((chest OR lung* OR thora*) NEAR/3 (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film* OR CXR*)):ti,ab,kw	5972
#14	#11 OR #12 OR #13	6057
#15	[mh "Lung Neoplasms"] OR [mh ^"Solitary Pulmonary Nodule"]	10362
#16	((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) NEAR/3 (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)):ti,ab,kw	29446
#17	((pancoast* OR "superior sulcus" OR "pulmonary sulcus") NEAR/4 (tumor* OR tumour* OR syndrome*)):ti,ab,kw	17
#18	(sclc OR nsclc):ti,ab,kw	12548
#19	#15 OR #16 OR #17 OR #18	30061
#20	#10 and #14 and #19	59
#21	("AI-Rad Companion Chest" NEXT X-ray*)	0
#22	("Annalise" NEXT CXR*)	0
#23	("Auto Lung Nodule" NEXT Detection*)	0
#24	ChestView*	0
#25	(("Chest X-Ray" NEXT Classifier*) OR Quibim*)	0
#26	CheXVision*	0
#27	(("ClearRead" NEXT Xray*) NEAR/2 Detect)	0
#28	("InferRead DR" NEXT Chest*)	0
#29	JLD-02K*	0
#30	("Lunit INSIGHT" NEXT CXR*)	3
#31	("Milvue" NEXT Suite*)	0
#32	("ChestEye" NEXT Quality*)	0
#33	(qXR* OR Qure*)	947
#34	(("red" NEXT dot*) OR behold*)	76
#35	("SenseCare-Chest DR" NEXT Pro*)	0
#36	("VUNO Med-Chest" NEXT X-Ray*)	1
#37	(X1* AND "Visionairy Health")	0
#38	#21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37	1027
#39	#14 and #38	8
#40	#19 and #38	8

#41 #39 or #40 12
#42 #20 or #41 68
Trials 67

Epistemonikos

(title:(("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND ((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)) OR ((pancoast* OR superior sulcus OR pulmonary sulcus) AND (tumor* OR tumour* OR syndrome*)))) OR abstract:(("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND ((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)) OR ((pancoast* OR superior sulcus OR pulmonary sulcus) AND (tumor* OR tumour* OR syndrome*)))))

Systematic Review: 133

Broad Synthesis: 6

ACM Digital Library

Searched ACM Guide to Computing Literature

<u>Reviews</u>

Title:((("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR (automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*) OR "support vector machine" OR

"support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND ((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*))))) OR Abstract:((("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR (automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND ((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*)))))

Limited publication type to: Review Article

20 Results

Primary Studies #1

Title:((("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR (automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) AND ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*)))) OR Abstract:((("AI" OR "artificial intelligence" OR "artificial learning" OR "artificial reasoning" OR "machine intelligence" OR "machine learning" OR "machine reasoning" OR "deep intelligence" OR "deep learning" OR "deep reasoning" OR "neural network" OR "neural networks" OR "neural networking" OR convolutional OR "CNN" OR "CNNs" OR (automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*) OR "support vector machine" OR "support vector machines" OR "support vector network" OR "support vector networks" OR "random forest" OR "random forests" OR "black box learning") AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) AND ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR

malignan* OR adenocarcinom* OR blastoma*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*))))

509 Results

Primary Studies #2

Title:(((ChestView* OR “Chest X-Ray Classifier” OR Quibim* OR CheXVision* OR (“ClearRead Xray” AND Detect) OR “InferRead DR Chest” OR JLD-02K* OR “Lunit INSIGHT CXR” OR “Milvue Suite” OR “ChestEye Quality” OR qXR* OR Qure* OR “red dot” or behold* OR “SenseCare-Chest DR Pro” OR “VUNO Med-Chest X-Ray” OR (X1* AND “Visionairy Health”)) AND ((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*))))) OR Abstract:(((ChestView* OR “Chest X-Ray Classifier” OR Quibim* OR CheXVision* OR (“ClearRead Xray” AND Detect) OR “InferRead DR Chest” OR JLD-02K* OR “Lunit INSIGHT CXR” OR “Milvue Suite” OR “ChestEye Quality” OR qXR* OR Qure* OR “red dot” or behold* OR “SenseCare-Chest DR Pro” OR “VUNO Med-Chest X-Ray” OR (X1* AND “Visionairy Health”)) AND ((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*) OR ((lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) AND (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumor* OR tumour* OR carcino* OR malignan* OR adenocarcinom* OR blastoma*)) OR ((pancoast* OR "superior sulcus" OR "pulmonary sulcus") AND (tumor* OR tumour* OR syndrome*)))))

1 Result

PROSPERO

Line	Search for	Hits
#1	MeSH DESCRIPTOR Artificial Intelligence EXPLODE ALL TREES	477
#2	MeSH DESCRIPTOR machine learning EXPLODE ALL TREES	154
#3	MeSH DESCRIPTOR deep learning EXPLODE ALL TREES	23
#4	MeSH DESCRIPTOR supervised machine learning EXPLODE ALL TREES	1
#5	MeSH DESCRIPTOR support vector machine EXPLODE ALL TREES	0
#6	MeSH DESCRIPTOR unsupervised machine learning EXPLODE ALL TREES	0
#7	ai	2253
#8	(artificial or machine or deep) AND (intelligence or learning or reasoning)	2211
#9	MeSH DESCRIPTOR Neural Networks, Computer EXPLODE ALL TREES	28
#10	"neural network" or "neural networks" or convolutional or CNN or CNNs	566
#11	MeSH DESCRIPTOR Diagnosis, Computer-Assisted EXPLODE ALL TREES	15

#12 MeSH DESCRIPTOR Pattern Recognition, Automated EXPLODE ALL TREES 1
#13 ((automat* or autonomous or "computer aided" or "computer assisted") AND (detect* or identif* or diagnos*)) 4314
#14 "support vector machine" or "support vector machines" or "random forest" or "black box learning" 182
#15 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 7901
#16 MeSH DESCRIPTOR Radiography, Thoracic EXPLODE ALL TREES 10

#17 MeSH DESCRIPTOR X-Rays 29
#18 ((chest or lung* or thora*) and (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR* 1194
#19 #18 OR #17 OR #16 1210
#20 #15 AND #19 111
#21 MeSH DESCRIPTOR Lung Neoplasms EXPLODE ALL TREES 574
#22 MeSH DESCRIPTOR Solitary Pulmonary Nodule 6
#23 (lung or lungs or pulmon* or intrapulmon* or bronch*) AND (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*) 6688
#24 (pancoast* or "superior sulcus" or "pulmonary sulcus") and (tumor* or tumour* or syndrome*) 6
#25 sclc or nsclc 1004
#26 #21 OR #22 OR #23 OR #24 OR #25 6738
#27 #26 AND #15 301
#28 #27 OR #20 366
#29 #15 AND #19 AND #26 46

WHO ICTRP

Targeted search #1

((lung* OR pulmonary OR intrapulmon* or bronch*) AND (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumor* or tumour* or carcino* or malignan* or adenocarcinom* or blastoma*)) in the Condition

AND

(((artificial or machine or deep) AND (intelligence or learning or reasoning)) OR (AI OR "neural network*" or convolutional or CNN or CNNs OR "support vector machine*" or "random forest*" or "black box learning") OR ((automat* or autonomous or "computer aided" or "computer assisted") AND (detect* or identif* or diagnos*))) in the Intervention

AND

Recruitment status is All

40 records for 39 trials

Targeted search #2

((((artificial or machine or deep) AND (intelligence or learning or reasoning)) OR (AI OR "neural network*" or convolutional or CNN or CNNs OR "support vector machine*" or "random forest*" or "black box learning") OR ((automat* or autonomous or "computer aided" or "computer assisted") AND (detect* or identif* or diagnos*))) AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*)) in the Intervention

AND

Recruitment status is All

16 records for 16 trials

Targeted search #3

((((artificial or machine or deep) AND (intelligence or learning or reasoning)) OR (AI OR "neural network*" or convolutional or CNN or CNNs OR "support vector machine*" or "random forest*" or "black box learning") OR ((automat* or autonomous or "computer aided" or "computer assisted") AND (detect* or identif* or diagnos*))) AND (((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR x-ray* OR xray* OR film*)) OR CXR*))

AND

Recruitment status is All

34 records for 34 trials

Study selection flowchart

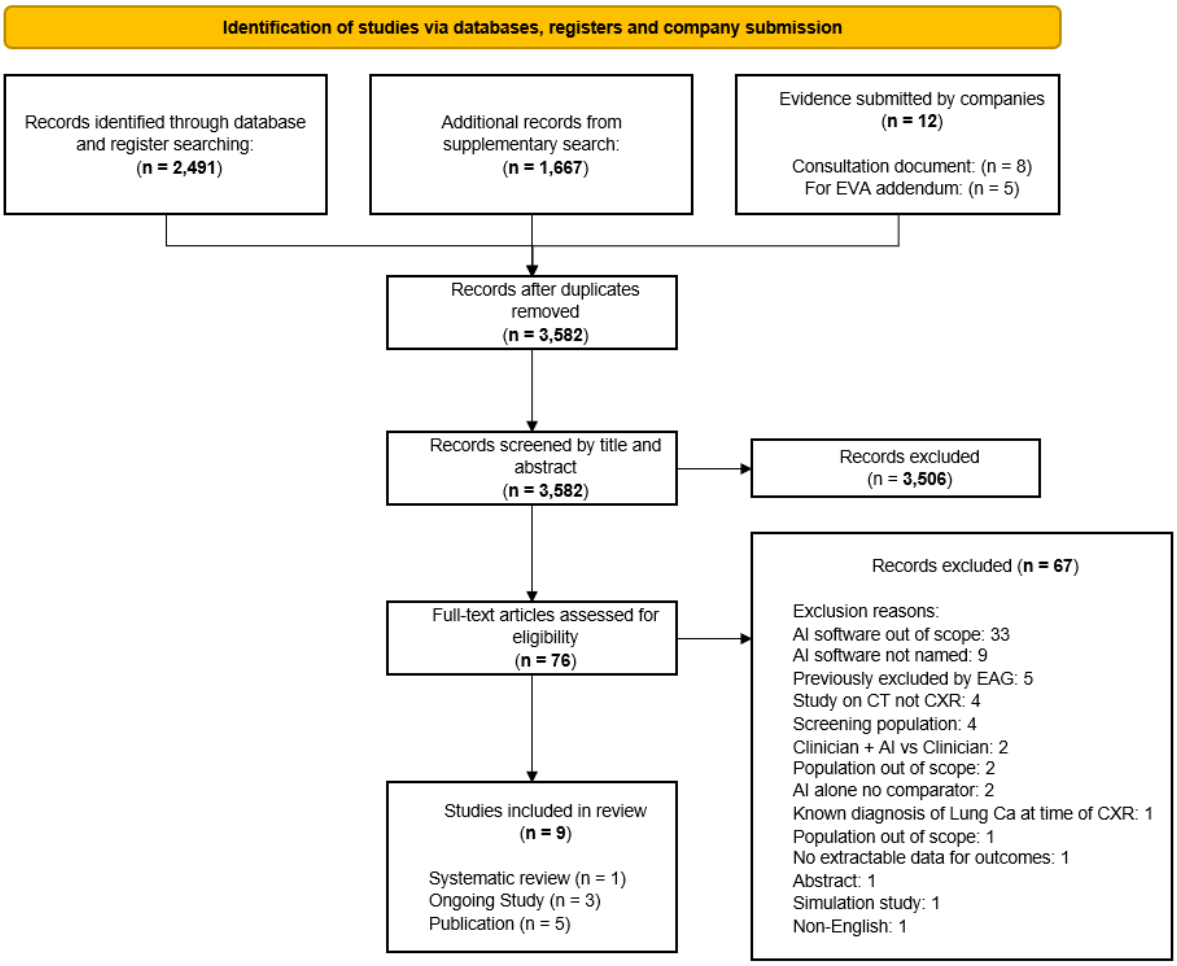

Figure 1: Study selection flowchart

Appendix B – Inclusion/Exclusion criteria

Population	Adults referred from primary care who are: 1. undergoing CXR due to symptoms suggestive of lung cancer, e.g., cough, fatigue, shortness of breath, chest pain, weight loss, appetite loss, persistent or recurrent chest infection, finger clubbing, supraclavicular lymphadenopathy or persistent cervical lymphadenopathy, chest signs consistent with lung cancer and/or thrombocytosis (symptomatic population) 2. undergoing CXR for reasons unrelated to lung cancer (incidental population). Where data permits, subgroups will be considered based on: • Ethnicity • Age • Sex • Socio-economic status
Target condition	Lung Cancer
Intervention	CXR interpreted by one of the following AI software: AI-Rad Companion Chest X-ray (Siemens Healthineers), Annalise CXR (annalise.ai), Auto Lung Nodule Detection (Samsung), ChestLink Radiology Automation (Oxipit), ChestView (GLEAMER), Chest X-ray (Rayscape), ClearRead Xray – Detect (Riverain Technologies), InferRead DR Chest (Infervision), Lunit INSIGHT CXR (Lunit), Milvue Suite (Milvue), qXR (Qure.ai), red dot (behold
Comparator	CXR interpreted by radiology specialist (radiologist, reporting radiographer) 'Clinician Alone' OR in conjunction with one of the above AI softwares 'Clinician + AI'
Reference standard	For accuracy of lung cancer detection: Lung cancer confirmed by histological analysis of lung biopsy, or diagnostic methods specified in NICE guideline 122, 8 where biopsy is not applicable. For accuracy of nodule detection: Radiology specialist (single reader or consensus of more than one reader)

Outcome	• Test accuracy for the detection of lung cancer (sensitivity, specificity, positive predictive value, numbers of true positive, false positive, true negative, false negative results, number of lung cancers diagnosed)
	• Test failures (rates, and data on inconclusive, indeterminate, and excluded samples, failure due to any other reason)
	• Characteristics of discordant cancers cases
	• Test accuracy for the detection of lung nodules
	• Concordance in lung nodule detection between radiology specialist with and without adjunct AI software
	• Practical implications:
	  ○ Time to x-ray report, CT scan, diagnosis, turnaround time (image review to radiology report),
	  ○ Acceptability of software to clinicians

	○ Impact on clinical decision-making ○ Impact of false positives on workflow • Mortality • Morbidity • Health-related quality of life
Study design	Comparative study designs
Publication type	Peer reviewed papers
Language	English
Exclusion	Versions of AI software that are not commercially available, are not named in the protocol, or are not specified in the study publication. Computer aided detection that does not include AI software. Non-human studies. Letters, editorials, communications, conference abstracts, qualitative studies. People with a known diagnosis of lung cancer at the time of CXR. Studies of children. Study designs that do not include a control/comparator arm. Simulation studies or studies using synthetic images. Studies not applicable to primary care patients, e.g., neurosurgery, transplant, or plastic surgery patients, people in secure forensic mental health services. Studies where more than 10% of the sample do not meet our inclusion criteria. Studies without extractable numerical data. Studies that provided insufficient information for assessment of methodological quality/risk of bias. Articles not available in the English language. Studies using index tests or reference standards other than those specified in the inclusion criteria. Studies of people who do not have signs and symptoms of cancer or a suspected condition or trauma (i.e., people undergoing health screening). Studies where it cannot be determined if the inclusion criteria are met.

Appendix C – EAG questionnaire

NICE Early Value Assessment - EAG literature review Topic: Artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer in primary care referrals (Early Value Assessment)
https://www.nice.org.uk/guidance/indevelopment/gid-hte10018 Cedar are an external assessment group (EAG) for NICE and have been commissioned to conduct a literature review as part of the above early value assessment (EVA) to support committee decision making on artificial intelligence (AI)-derived software to analyse chest x-rays for suspected lung cancer in primary care referrals. The objective is to review studies on AI-derived software for analysing chest X-rays that report results for clinician review (Clinician + AI) compared with AI-derived software alone (AI alone). The EAGs findings will be summarised in a short report. Any responses you provide may be used in the report which will be shared with stakeholders, the diagnostics advisory committee and will be published on the NICE website as part of the committee papers supporting guidance development.

From previous committee discussions during the EVA process, what was the rationale for suggesting ‘AI alone’ as a comparator, considering it is not current practice?

1. How would you define adjunct AI software or 'Clinician + AI' in practice?

2. How would you define 'AI alone' in practice, and at what point, in your opinion, does the transition from ‘AI alone’ to ‘Clinician + AI’ occur?

3. Where in the primary care referral pathway for suspected lung cancer could you envision ‘AI alone’ software being used, if at all, and what benefits might it bring?

4. NICE have proposed some potential outcomes to look at in the evidence seen in the image. Which of the outcomes would be a priority for us to focus on when collating evidence?

5. Are you aware of any ongoing/upcoming studies or published comparative evidence on 'clinician + AI' vs 'AI alone' that we should be aware of?

6. Please use this text box to provide us with any other comments

EAG questionnaire – Updated version

Topic: Artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer in primary care referrals (Early Value Assessment)

https://www.nice.org.uk/guidance/indevelopment/gid-hte10018

Cedar are an external assessment group (EAG) for NICE and have been commissioned to conduct a rapid literature review as part of the above early value assessment (EVA) to support committee decision making on artificial intelligence (AI)-derived software to analyse chest x-rays for suspected lung cancer in primary care referrals.

The objective is to review comparative studies on Clinician review of CXR **(Clinician alone)** versus AI-derived software for analysing CXR **(AI alone)**.

We are also interested in adjunctive AI-derived software for analysing chest X-rays that report results for clinician review **(Clinician + AI)** compared with AI-derived software alone **(AI alone)**. The EAGs findings will be summarised in a short addendum report.

Any responses you provide may be used in the report which will be shared with stakeholders, the diagnostics advisory committee and will be published on the NICE website as part of the committee papers supporting guidance development.

1. What experience, if any, do you have in using AI software for analysing CXRs for suspected lung cancer?

2. From previous committee discussions during the EVA process, what was the rationale for suggesting ‘AI alone’ as a comparator, considering it is not current practice?

3. How would you define adjunct AI software or 'Clinician + AI' in practice?

4. How would you define 'AI alone' in practice, and at what point, in your opinion, does the transition from ‘AI alone’ to ‘Clinician + AI’ occur?

5. Where in the primary care referral pathway for suspected lung cancer could you envision ‘AI alone’ software being used, if at all, and what benefits might it bring?

6. When considering how an AI software would be integrated into the NHS, what concerns or challenges do you foresee in clinical practice?

7. NICE have proposed some potential outcomes to look at in the evidence seen in the image.

8. Which of the outcomes would be a priority for us to focus on when collating evidence?

9. Are you aware of any ongoing/upcoming studies or published comparative evidence on either 'Clinician vs AI alone' or 'Clinician + AI' vs 'AI alone' that we should be aware of?

10. Please use this text box to provide us with any other comments.

Appendix D – Development of the draft NICE AI medical device intervention search filters

The information in the below section was provided to the EAG by the NICE information services team. The draft NICE AI medical device intervention search filters for MEDLINE and Embase (Ovid) were designed to retrieve evidence about the interventions effectively to inform NICE medical technology and diagnostics topics.

The draft MEDLINE filter was developed first using 'second-generation' filter methodology (1). This method involves using a set of known relevant references on the filter topic to identify the search terms for the filters and then calculating the recall (also known as sensitivity) of the filters against the references (1). Recall is the proportion of relevant references retrieved by a search filter (1, 2). Included references from published NICE medical technology and diagnostics products were used to inform the search terms for the draft MEDLINE filter and the draft filter's recall against the included references was calculated. The draft MEDLINE filter achieved 100% recall of the 53 included references with AI identifiers in their database records. The draft MEDLINE filter was then translated to create a draft Embase filter for use in the Embase bibliographic database.

Following the development of the draft MEDLINE and Embase filters, a 'watchful waiting' period of at least 6 months was implemented. The purpose of this was to further assess the retrieval performance of the filters by testing them on active NICE AI topics and to allow for the potential identification of new AI-related terminology. Following the 'watchful waiting' period, the draft filters will be revised (if required) and validated using 'third generation' filter development methodology (1). This method will involve assessing the external validity of filters by calculating their recall against an independent set of relevant references that have not been used for their development (1, 2).

Appendix E – Studies excluded at full-text

Study author and year	Study title	Reason for exclusion at full-text
Aggarwal et al., 2014	Content based image retrieval approach in creating an effective feature index for lung nodule detection with the inclusion of expert knowledge and proven pathology.	AI software out of scope
Ahmad et al., 2023	Machine Learning Augmented Interpretation of Chest X-rays: A Systematic Review	Systematic review - AI technologies largely not relevant and of those that are, they have already been screened and excluded
Amir et al., 2016	After Detection: The Improved Accuracy of Lung Cancer Assessment Using Radiologic Computer-aided Diagnosis	Systematic review but no mention of individual AI technologies
Aoki et al., 2012	Usefulness of computerized method for lung nodule detection on digital chest radiographs using similar subtraction images from different patients	Technology not stated and doesn't look to be AI. Also excluded by Warwick Evidence during original EVA
Astley et al., 2022	Deep learning in structural and functional lung image analysis	Only 12.2% of included studies are on CXRs - none of which include AI technologies relevant to the scope of the review
Bhattarai et al.,	Diagnostic accuracy of Artificial Intelligence assisted chest x-ray compared to chest x-ray alone in detection and classification of lung cancer: A systematic review	Prospero registration of systematic review but AI software not named
Cha et al., 2019	Performance of Deep Learning Model in Detecting Operable Lung Cancer With Chest Radiographs	AI software out of scope
Chen et al., 2013	Computerized detection of lung nodules by means of "virtual dual-energy" radiography	AI software out of scope
Chen et al., 2016	A parameterized logarithmic image processing method with Laplacian of Gaussian filtering for lung nodule enhancement in chest radiographs	AI software out of scope. No radiologist input and so not relevant to comparator and intervention. Validation study for the algorithm. Also excluded by Warwick Evidence during original EVA
Choi et al., 2012	The cut-off values for auto-detection of lung cancer in chest radiography: An example using an unsupervised method	AI software out of scope. Also exclude by Warwick Evidence during original EVA
CTRI/2022/09/046002	An open label, retrospective, observational, non-inferior study to Investigate the Potential of Deep Learning in Assessing the chest diseases Depicted on Digital Chest Radiographs and to Compare Its Performance with Certified Radiologists	Clinician vs AI but AI software not named. Referral route unclear as data from retrospective hospital dataset. Inclusion criteria is patients with any of 18 conditions
Dasegowda et al., 2023	Radiologist-Trained AI Model for Identifying Suboptimal Chest-Radiographs	AI software out of scope
Dissez et al., 2022	Enhancing Early Lung Cancer Detection on Chest Radiographs with AI-assistance: A Multi-Reader Study	Study assesses the use of clinician review with and without AI (Clinician + AI vs Clinician) and is therefore out of scope for the addendum

Dyer et al., 2021	Diagnosis of normal chest radiographs using an autonomous deep-learning algorithm	AI software not named although authors employed by Behold.ai company. Also excluded by Warwick Evidence during original EVA due to mixed population of A&E, GP and outpatients. Also no subgroup analysis of sample recruited from primary care
Dyer et al., 2022	Robustness of an Artificial Intelligence Solution for Diagnosis of Normal Chest X-Rays	Software and comparators (Clinician vs AI) are relevant but the study does not have extractable data on outcomes relevant to addendum scope. No data on lung cancer or lung nodules.
Farhan et al., 2023	MCLSG:Multi-modal classification of lung disease and severity grading framework using consolidated feature engineering mechanisms	AI software out of scope
Fischer et al., 2022	Computer-Aided Detection of Seven Chest Pathologies on Standard Posteroanterior Chest X-Rays Compared to Radiologists Reading Dual-Energy Subtracted Radiographs	Clinician vs AI but excluded as population is inpatients and outpatients with most having CXR pre-surgery so unlikely 90% were referred from primary care. Also no subgroup analysis of outpatient group. Also excluded by Warwick Evidence during original EVA
Forte et al., 2022	Deep Learning Algorithms for Diagnosis of Lung Cancer: A Systematic Review and Meta-Analysis	Systematic review evaluating the diagnostic performance of AI networks for lung cancer on CT and not CXR
France et al., 2020	Classification and retrieval of thoracic diseases using patch-based visual words: A study on chest X-rays	Non comparative and no mention of any AI softwares in scope
Govindarajan et al., 2022	Role of an Automated Deep Learning Algorithm for Reliable Screening of Abnormality in Chest Radiographs: A Prospective Multicenter Quality Improvement Study	AI alone with no comparator. Population also includes children (age >6 years), subgroup results reported but only for normal/abnormal, and states setting is routine screening. Also excluded by Warwick Evidence during original EVA
Guo et al., 2012	A computerized scheme for lung nodule detection in multiprojection chest radiography	Software not named. Also excluded by Warwick Evidence during original EVA
Hassen et al., 2013	Automatic detection of lesions in lung regions that are segmented using spatial relations	AI software out of scope
Juan et al, 2023	Computer-assisted diagnosis for an early identification of lung cancer in chest X rays	AI software out of scope
Kang et al., 2022	Development of a multipotent diagnostic tool for chest X-rays by multi-object detection method	AI software out of scope
Kaviani et al., 2022a	Performance of a Chest Radiography AI Algorithm for Detection of Missed or Mislabeled Findings: A Multicentre Study	Population not described, CXRs taken from a database and no information that these would be primary care referrals. Also excluded by Warwick Evidence during original EVA
Kaviani et al., 2022b	Frequency of Missed Findings on Chest Radiographs (CXRs) in an International, Multicentre Study: Application of AI to Reduce Missed Findings	Excluded as population is not described and it's unclear if referred with symptoms but only those with 'normal' CXR were used. Also excluded by Warwick Evidence during original EVA

KCT0004147	A single-centre, randomized, crossover and retrospective pivotal trial to evaluate the efficacy of VUNO Med – Chest X-ray in screening of abnormalities on chest radiograph	Screening population and comparator is without AI assistance (Clinician alone) and out of scope
KCT0005051	Diagnosis of lung nodule and lung cancer on screening chest radiographs: comparative clinical trial for evaluation of artificial intelligence-integrated PACS versus conventional PACS	Screening population
KCT0008153	Artificial intelligence based prioritization of interpretation for chest radiographs with suspected pneumothorax and pneumoperitoneum: A prospective controlled before after study	Study is AI based prioritisation in radiographs with suspected pneumothorax and pneumoperitoneum and not lung cancer or nodules. Also Clinician alone vs Clinician
Kim et al., 2020	Test-retest reproducibility of a deep learning-based automatic detection algorithm for the chest radiograph.	Population undergoing pre-operative CXR and so not primary care referrals. Also excluded by Warwick Evidence during original EVA
Klarenbeek et al., 2020	The Effect of Higher Level Computerized Clinical Decision Support Systems on Oncology Care: A Systematic Review	Systematic review of cancers and not specific to Lung Ca, Does not include any studies within the scope or any with technologies in scope
Kvak et al., 2023	Leveraging Deep Learning Decision-Support System in Specialized Oncology Centre: A Multi-Reader Retrospective Study on Detection of Pulmonary Lesions in Chest X-ray Images	AI software out of scope
Kwak et al., 2023	Incidentally found resectable lung cancer with the usage of artificial intelligence on chest radiographs	Study is AI alone with no comparator
Li et al., 2015	Computer-aided nodule detection system: results in an unselected series of consecutive chest radiographs	Non comparative and 'AI alone'. Population is unclear
Li et al, 2021	Assessing the predictive accuracy of lung cancer, metastases, and benign lesions using an artificial intelligence-driven computer aided diagnosis system	Study on CT scans and AI software trained on CT scans
Li et al., 2018	A Solitary Feature-Based Lung Nodule Detection Approach for Chest X-Ray Radiographs	AI software not named. CXRs are also from two databases and study is validation study. Also excluded by Warwick Evidence during original EVA
Li et al., 2020	Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection	AI software out of scope
Liang et al., 2019	Dense networks with relative location awareness for thorax disease identification	AI software out of scope

Liang et al., 2020	Identifying pulmonary nodules or masses on chest radiography using deep learning: external validation and strategies to improve clinical practice	AI software out of scope
Liu et al., 2023	The value of artificial intelligence in the diagnosis of lung cancer: A systematic review and meta-analysis	Systematic review of studies on CT
Majkowska et al., 2020	Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation	Software not named. Also excluded by Warwick Evidence in the original EVA due to referral route unclear but one database is consecutive inpatient and outpatient images and the others are all CXRs from multiple different hospitals
Mercy Theresa et al., 2016	A Survey on CAD technique for various abnormality classification in chest radiography	AI software out of scope & survey
Nagendran et al., 2020	Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies	Systematic review but no AI softwares in scope included
Nam et al., 2019	Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs	Population unclear, also a simulation validation from CXRs from emergency department CXRs. Also excluded by Warwick Evidence during original EVA
Nam et al., 2020	Undetected Lung Cancer at Posteroanterior Chest Radiography: Potential Role of a Deep Learning-based Detection Algorithm	Interventions are Clinician + AI vs Clinician and therefore out of scope. CXRs from people with confirmed lung CA initially undetected on CXR - unclear referral route or if CXR for symptoms or no symptoms, also unclear where the 'normal' x-rays are from. Also excluded by Warwick Evidence during original EVA
Nicholson et al., 2022	Could simplified stimuli change how the brain performs visual search tasks? A deep neural network study	AI software out of scope
Park et al., 2020	Deep learning-based detection system for multiclass lesions on chest radiographs: comparison with observer readings	Software not named. Also excluded by Warwick Evidence during original EVA
Pesce et al., 2019	Learning to detect chest radiographs containing pulmonary lesions using visual attention networks	AI software out of scope
Pham et al., 2021	Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels	AI software out of scope

Rajogopalan et al., 2020	The detection of lung cancer using massive artificial neural network based on soft tissue technique	AI software out of scope
Schalekamp et al., 2013	Computer aided detection shows added value to bone suppression imaging for the detection of lung nodules in chest radiographs	Abstract
Schalekamp et al., 2014a	New methods for using computer-aided detection information for the detection of lung nodules on chest radiographs	Looks to be Clinician + AI vs Clinician (ClearRead detect with ClearRead bone suppression + radiologist). Also excluded by Warwick Evidence in original EVA as CXRs are retrospectively selected, derived from clinically indicated examinations
Schalekamp et al., 2014b	Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the support by bone-suppressed images	Exclude as Clinician + AI vs Clinician. Also excluded by Warwick evidence in original EVA as *'ClearRead Detect with ClearRead Bone suppression + radiologist. CXR retrospectively selected, derived from clinically indicated examinations, and referral route unclear'*.
Sheng et al., 2014	Separation of bones from chest radiographs by means of anatomically specific multiple massive-training ANNs combined with total variation minimization smoothing	AI software out of scope
Shi et al., 2015	Evaluation of MTANNs for eliminating false-positive with different computer aided pulmonary nodules detection software	Validation of an additional algorithm applied to named interventions but on simulations and not patients. Also excluded by Warwick Evidence in original EVA
Shi et al., 2014	A new method based on MTANNs for cutting down false-positives: an evaluation on different versions of commercial pulmonary nodule detection CAD software	AI software out of scope and non comparative
Shin et al., 2023	The impact of artificial intelligence on the reading times of radiologists for chest radiographs	Study compares AI-unaided vs. AI-aided or 'Clinician + AI vs. Clinician'.
Sim et al., 2020	Deep Convolutional Neural Network-based Software Improves Radiologist Detection of Malignant Lung Nodules on Chest Radiographs	Population includes normal CXR from health screening populations and CXR with lung cancer at tertiary hospital. Also excluded by Warwick Evidence in original EVA
Singh et al., 2018	Deep learning in chest radiography: Detection of findings and presence of change	No extractable outcome data on lung nodules or cancer. Also excluded by Warwick Evidence in original EVA
Suarez-Cuenca et al., 2017	A cad scheme for early lung cancer detection in chest radiography	Non comparative and CAD scheme not named

Subapriya et al., 2020	A computer aided diagnosis of lung disease using machine learning approach	AI software out of scope
Vidal-Mondejar et al., 2023	Methodological evaluation of systematic reviews based on the use of artificial intelligence systems in chest radiography	Study in Spanish and is a methodological critique of systematic reviews on the use of AI in chest radiography
Wang et al., 2018	Automated chest screening based on a hybrid model of transfer learning and convolutional sparse denoising autoencoder	AI software out of scope
Weiss et al., 2023	Deep learning to estimate lung disease mortality from chest radiographs	AI software out of scope
Yuan et al., 2022	Application of logistic regression and convolutional neural network in prediction and diagnosis of high-risk populations of lung cancer	AI software out of scope
Zheng et al., 2021	Natural Language Processing to Identify Pulmonary Nodules and Extract Nodule Characteristics From Radiology Reports	AI software out of scope
Zheng et al., 2022	Diagnostic Accuracy of Deep Learning and Radiomics in Lung Cancer Staging: A Systematic Review and Meta-Analysis	Systematic review on studies using CT and none on CXRs