

Supplementary material 1: Detailed table of characteristics of summarised (but ineligible) studies

Study details	Population	Interventions	Notes
<p>Dissez 2022²²</p> <p>UK</p> <p>Retrospective cohort study, single centre</p> <p>Database dates: 2020</p> <p>Aim: To evaluate the impact of an AI algorithm in augmenting the ability of clinicians to identify lung cancer on CXR</p> <p>No funding. Several authors are employed by and/or have stock/stock options in Behold.ai</p>	<p>400 CXRs taken from 400 adults with either:</p> <ul style="list-style-type: none"> - clinical text report indicating potentially malignant CXR and follow-up CT (random sample of n=200) - clinical text report of no urgent findings (random sample of n=200) <p>132/400 CXR ground-truthed as suspicious for lung cancer</p> <p>72/400 CXR clinically confirmed lung cancers</p> <p>CXRs taken from hospital databases</p>	<p>Red Dot (Behold.ai) + radiologists</p> <p>Comparator: 11 clinicians (3 FRCR consultant radiologists, 2 board-certified radiologists, 2 radiology trainees, 4 reporting radiographers) not involved in ground-truthing</p> <p>Two sessions with 4-week washout: CXRs reviewed without AI assistance in first session and with AI assistance in second image. Clinicians provided with basic clinical information, including age and sex</p> <p>Reference standard: lung cancer diagnosis (clinically confirmed outcomes collated by radiologist including repeat CXR and CT outcomes, lung cancer diagnosis, TNM staging and biopsy outcomes)</p> <p>Comparison of relevance extracted here is average accuracy of radiologists in identifying lung cancer (versus reference standard) and average performance of radiologists + AI software in identifying lung cancer (versus reference standard)</p>	<p>Referral route unclear, not known if symptomatic or incidental.</p> <p>Population from retrospective CXR collected in one UK NHS hospital during 2020.</p> <p>Participating clinicians had a range of 1 to 18 years' experience. Each of 11 clinicians reviewed each x-ray with and without AI, unclear if 4-week washout is sufficient.</p> <p>CT referrals were hypothetical rather than actual, as CXRs were retrospectively selected from databases.</p> <p>Full details and responses of the clinician survey not reported.</p>
<p>Nam 2020²³</p> <p>Korea</p>	<p>218 CXRs from 218 people with pathologically confirmed lung cancers at percutaneous</p>	<p>Lunit INSIGHT version 1.0.1.1 + radiologists</p> <p>Comparator: four experienced thoracic</p>	<p>Referral route unclear, not known if symptomatic or incidental.</p>

<p>Retrospective cohort study, single centre</p> <p>Database dates: 2017 to 2018</p> <p>Aim: To evaluate a deep learning–based algorithm for detecting lung cancers not reported on CXR</p> <p>Non-commercial funding</p>	<p>lung biopsy from single hospital AND false-negative posteroanterior CXR prior to biopsy (n=168)</p> <p>Nodules <5mm excluded</p> <p>Normal true-negative CXR confirmed on same day CT (n=50)</p> <p>CXRs taken from hospital database</p>	<p>radiologists not involved in ground truth</p> <p>Each reader reviewed CXR and made judgement (test 1), then reviewed results of algorithm and initial decision, and modified decision (test 2)</p> <p>Ground truth: CT (3-38 days from CXR) and/or re-evaluation of CXR reviewed by two experienced thoracic radiologists</p> <p>Comparison of relevance extracted here are average of radiologists (versus ground truth) and average of radiologist with AI (versus ground truth)</p>	<p>Korean population likely low generalisability to UK population.</p> <p>Population mainly confirmed lung cancer and false negative CXR prior to biopsy, some with true negative CXR.</p> <p>Nodules smaller than 5 mm were excluded</p> <p>Reader aware of initial decision on second read with algorithm</p> <p>Experience of radiologists ranged between 5 and 9 years</p> <p>Readers were aware of the characteristics of the CXRs but not the proportion of positive to negative cases</p>
<p>Jang 2020²⁴</p> <p>Korea</p> <p>Retrospective cohort study, single centre</p> <p>Database dates: 2010 to 2014</p> <p>Aim: To evaluate the efficacy of a deep learning–based automatic detection</p>	<p>351 CXRs taken from 351 people diagnosed with lung cancer at a single tertiary hospital AND visible cancer on prior CXR at least 3 months before diagnosis when reviewed retrospectively by radiologist (n=117):</p> <ul style="list-style-type: none"> - Detected without misinterpretation n=12 - Overlooked cancers n=105 (detected with misinterpretation n=23, undetected n=82) 	<p>Lunit INSIGHT version 1.2.0.0 + radiologists</p> <p>Comparator: six experienced thoracic radiologists and three radiology residents not involved in reference standard</p> <p>Each reader reviewed each CXR twice, once with and once without algorithm, with ≥4 week interval between sessions</p> <p>Reference standard: lung cancer lesion areas identified on CXRs and CT scans at time of diagnosis marked in consensus by two authors</p>	<p>Referral route unclear, not known if symptomatic or incidental, population is people with lung cancer and cancer visible on CXR prior to diagnosis, control group is those with normal CXR</p> <p>Korean population likely low generalisability to UK population.</p> <p>Experience of radiologists ranged between 1 and 12 years, radiology residents were either 2nd or 3rd years</p> <p>Observers were blind to clinical</p>

<p>algorithm in observer performance for detection of lung cancers on CXR</p> <p>Non-commercial funding</p>	<p>Healthy control with normal CXR confirmed at CT (n=234)</p> <p>CXRs taken from hospital database</p>	<p>Comparison of relevance extracted here are average accuracy of radiologists (versus reference standard) and average of radiologists with AI (versus reference standard)</p>	<p>information</p> <p>A web-based tool was used to document the readers results and calculate agreement between the AI and non-AI reading of the CXRs</p> <p>CT referrals were hypothetical rather than actual, as CXRs were retrospectively selected from databases</p>
<p>Koo et al 2021²⁵</p> <p>Korea</p> <p>Retrospective cohort study, single centre</p> <p>Database dates: 2016 to 2018</p> <p>Aim: To assess a deep convolutional neural network algorithm for pulmonary nodules on CXR</p> <p>Non-commercial funding</p>	<p>434 CXRs from 378 adults from a tertiary hospital with ≤ 3 nodules on both CXR and CT with ≥ 1 nodule pathologically confirmed on biopsy as either benign (n=246) or malignant (n=132) and nodules evident on chest CT visible on CXR</p> <p>CXRs taken from hospital database</p>	<p>Lunit INSIGHT CXR version 1.00 + radiologist</p> <p>Comparator: Two radiology residents and two thoracic radiologists</p> <p>Each reader reviewed CXR without AI and then ≥ 3 weeks later re-evaluated with the AI data</p> <p>Reference standard: Consensus from two thoracic radiologists with 10 and 7 years of experience using CR or CT</p> <p>Comparison of relevance extracted here are average accuracy of radiologists (versus reference standard) and average of radiologists with AI (versus reference standard)</p>	<p>Referral route unclear, participants from hospital setting in Korea.</p> <p>Likely low generalisability to UK population.</p> <p>Radiologist expertise differed (between 7 and 10 years of thoracic imaging experience for 2 radiologists, and 2 and 4 years of experience for 2 radiological residents), results for overall group extracted only</p> <p>Readers were blind to clinical information but were aware that CXRs would exhibit more nodules than CXRs from a normal clinical setting but not how many CXRs featured nodules</p> <p>Unclear if radiologists had their original decisions at the second reading</p>
<p>Homayounieh et al 2021²⁶</p>	<p>100 CXRs taken from 100 adults with posterior-anterior</p>	<p>AI-Rad Companion Chest X-ray (Siemens Healthineers) + Radiologist</p>	<p>Referral route unclear and generalisability to a UK primary care referred population</p>

<p>USA and Germany</p> <p>Retrospective cohort study, Two centres</p> <p>Database dates: 2000 to 2010</p> <p>Aim: To assess the ability of an AI algorithm to detect pulmonary nodules from CXR</p> <p>Commercial funding: (Lunit Inc; Riverain Technologies Inc; Siemens Healthineers AG)</p>	<p>CXRs taken between 2000 and 2010 (n=25 with absence of any abnormality, n=50 presence of pulmonary nodules of varying detection difficulties (20 challenging, 7 moderate, 23 easy), n=25 non-nodular abnormalities)</p> <p>CXRs taken from two databases, an ambulatory health care centre and the Lung Image Database Consortium</p>	<p>Comparator: Radiologist alone (7 staff radiologists and 3 radiology residents)</p> <p>One month period between readings without AI first, and then original decisions available at the second reading</p> <p>Ground truth: consensus from 2 thoracic radiologists (with 14 and 16 years of experience respectively)</p> <p>Comparison of relevance extracted here are average of radiologists (versus ground truth) and average of radiologist with AI (versus ground truth)</p>	<p>unclear</p> <p>Radiologist expertise differed widely (between 2.5 years and 35 years for staff radiologists; radiology residents were in first year)</p> <p>Radiologists had their original decisions at the second reading (although described as a washout period)</p> <p>CXRs were selected to ensure negative and positive cases and the detection level varied (nodule sizes between 4 to 28 mm). Readers were aware that there were positive and negative CXRs but not the ratio of positive or negative CXRs</p> <p>The order CXRs were read was randomised across readers and reading sessions and findings recorded on an electronic case record form. One test reader was excluded from the analysis as did not follow the exact instructions</p>
<p>[REDACTED]</p>	<p>[REDACTED]</p>	<p>Prototype AI Rad Companion Chest X-ray algorithm (Siemens Healthineers) + Radiologist</p> <p>Comparator: Seven radiologists</p> <p>Each radiologist assessed CXR unaided and with the AI after a four-week washout</p>	<p>Ongoing study with limited detail of early results provided</p> <p>Referral route unclear</p> <p>Radiologist expertise differed (four with > 4 years' experience and 3 with <4 years' experience)</p>

<p>[Redacted]</p> <p>[Redacted]</p> <p>[Redacted]</p>	<p>[Redacted]</p>	<p>Ground truth: Two thoracic radiologists using CXR and CT</p> <p>Comparison of relevance extracted here are average accuracy of radiologists (versus reference standard) and average of radiologists with AI (versus reference standard)</p>	<p>Unclear if prototype is commercially available AI</p> <p>Generalisability to a UK primary care referred population unclear</p>
---	-------------------	--	---