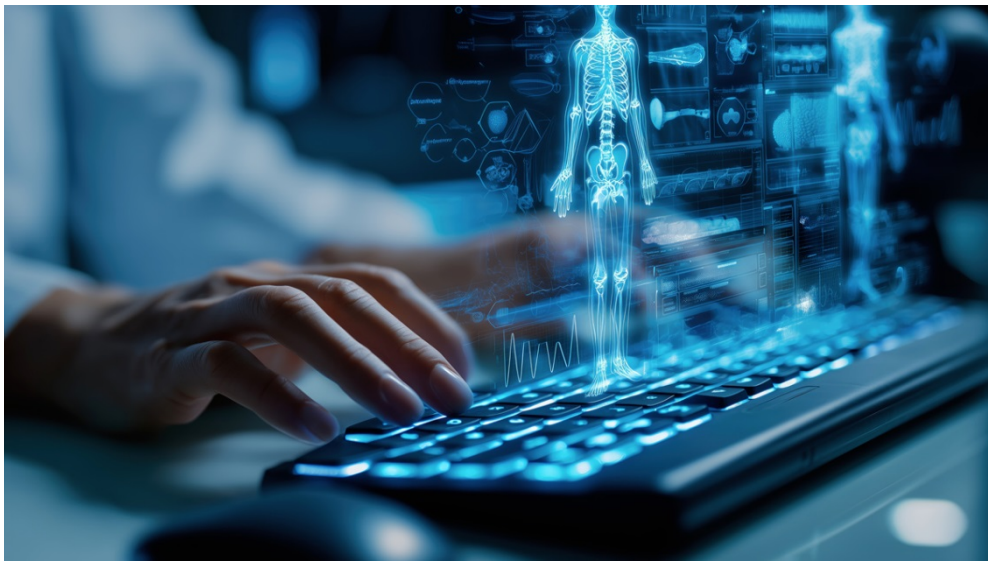
HTA Austria
Austrian Institute for
Health Technology Assessment
GmbH

AI-supported Chest X-Ray Analysis for Lung Cancer Detection



Systematic Review of clinical outcomes and organisational implications

Final report

AIHTA Project Report No.: 171b | ISSN: 1993-0488 | ISSN-online: 1993-0496

HTA Austria
Austrian Institute for
Health Technology Assessment
GmbH

AI-supported Chest X-Ray Analysis for Lung Cancer Detection

Systematic Review of clinical outcomes and organisational implications

Vienna, February 2026

Project Team

Project leader: Judit Erdos, MA

Authors: Judit Erdös, MA; Lena Grabenhofer, MSc

Project Support

Systematic literature search: Tarquin Mittermayr, MA

Internal review: Dr.PH, MSSc, MPH Gregor Götz

External Review: Assoc. Prof. Priv.-Doz. Dr. Helmut Prosch, Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna
Miguel Ángel Armengol de la Hoz, PhD; Head of the Data Science Lab, Fundación Progreso y Salud, Ministry of Health and Consumer Affairs, Regional Government of Andalusia

Correspondence: Judit Erdos, Judit.erdos@aihta.at

Cover photo: @ ERiK – stock.adobe.com

This report should be referenced as follows:

Erdos J, Grabenhofer L. AI-supported Chest X-Ray Analysis for Lung Cancer Detection. Systematic Review of clinical outcomes and organisational implications. AIHTA Project Report No.: 171b; 2026. Vienna: HTA Austria – Austrian Institute for Health Technology Assessment GmbH.

Conflict of interest

All authors and the reviewers involved in the production of this report have declared they have no conflicts of interest in relation to the technology assessed according to the Uniform Requirements of Manuscripts Statement of Medical Journal Editors (www.icmje.org).

Disclaimer

The external reviewers did not co-author the scientific report and do not necessarily all agree with its content. Only the AIHTA is responsible for errors or omissions that could persist. The final version and the policy recommendations are under the full responsibility of the AIHTA.

During the preparation of this work, the authors used Claude.ai and ChatGPT.com to enhance the writing process. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

IMPRINT

Publisher:
HTA Austria – Austrian Institute for Health Technology Assessment GmbH
Josefstädter Straße 39 | 1080 Vienna – Austria
https://www.aihta.at/

Responsible for content:
Dr. rer. soc. oec. Ingrid Zechmeister-Koss, MA, managing director

AIHTA Project Reports do not appear on a regular basis and serve to publicize the research results of the Austrian Institute for Health Technology Assessment.

AIHTA Project Reports are only available to the public via the Internet at http://eprints.aihta.at/view/types/hta_report.html.

AIHTA Project Report No.: 171b

ISSN 1993-0488

ISSN online 1993-0496

Content

List of figures

List of tables

List of abbreviations

AI ..... Artificial Intelligence
AIHTA ..... Austrian Institue for Health Technology Assessment
AP ..... Anteroposterior
ASPECTS ..... Alberta Stroke Program Early CT Score
AUROC ..... Area Under the Receiver Operating Characteristic Curve
AVIEW LCS+ .... Advanced Visualisation for Lung Cancer Screening Plus
CADe ..... Computer-Aided Detection
CADx ..... Computer-Aided Diagnosis
CAST ..... Computer-Assisted Triage
CINA ..... Computed Imaging Neuro Analysis
CI ..... Confidence Interval
CE ..... Conformité Européenne (European Conformity)
CT ..... Computed Tomography
CTCA ..... Coronary Computed Tomography Angiography
CXR ..... Chest X-Ray
DB-MIPS ..... Digital Dermoscopy Melanoma Imaging Processing System
DBT ..... Digital Breast Tomosynthesis
DHT ..... Digital Health Technology
DL ..... Deep Learning
DR ..... Digital Radiography
EVA ..... Early Value Assessment
EBUS-TBNA ..... Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration
ED ..... Emergency Department
EUS-FNA ..... Endoscopic Ultrasound-Guided Fine Needle Aspiration
EU ..... European Union
EUnetHTA ..... European Network for Health Technology Assessment
FDR ..... False Discovery Rate
FF ..... Forschungsfrage
FN ..... False Negative
FOR ..... False Omission Rate

FP .....False Positive
GRADE.....Grading of Recommendations Assessment, Development and Evaluation
HCN .....High-Confidence Nodules
HCT .....High-Contrast Thorax
HrQoL.....Health-Related Quality of Life
HTA .....Health Technology Assessment
HTD .....Health Technology Developer
INAHTA .....International Network of Agencies for Health Technology Assessment
IT .....Information Technology
KOALA .....Knee OsteoArthritis Labeling Assistant
LUAD .....Lung Adenocarcinoma
MDR .....Medical Device Regulation
MAIDA .....Medical Artificial Intelligence for Detection in ICU Chest X-rays
MDT.....Multidisciplinary Team
MRI .....Magnetic Resonance Imaging
NA/N.A. .....Not Applicable/Not Available
NICE .....National Institute for Health and Care Excellence
NIHR .....National Institute for Health and Care Research
NHS .....National Health Service
NPV.....Negative Predictive Value
NR .....Not Reported
NSCLC .....Non-Small Cell Lung Cancer
PA.....Posteroanterior
PET .....Positron Emission Tomography
PET-CT .....Positron Emission Tomography – Computed Tomography
PPV .....Positive Predictive Value
PRISMA.....Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Rad .....Radiologist
ROBIS.....Risk of Bias Assessment Tool for Systematic Reviews
RQ .....Research Question
SCLC .....Small Cell Lung Cancer
SD.....Standard Deviation
SHTG .....Scottish Health Technology Group
SR.....Systematic Review
TAT.....Turnaround Time
TP.....True Positive
TN .....True Negative
UK.....United Kingdom
US.....United States

Executive Summary

Background

Artificial intelligence (AI) applications in diagnostic imaging have expanded rapidly in recent years, with a growing number of systems developed to support image interpretation, workflow efficiency, and clinical decision-making in hospital settings. Radiology has been identified as one of the clinical domains most affected by increasing workloads, staff shortages, and rising diagnostic complexity, making it a priority area for AI-supported digital health technologies (DHTs).

Across imaging modalities, AI systems are most frequently designed to assist with the detection, classification, or prioritisation of suspected pathologies. Many of these tools are marketed as decision-support or triage systems and are regulated as CE-marked medical devices. Their potential benefits include improved diagnostic accuracy, reduced turnaround times, and more efficient use of radiology resources. Given the increasing availability and intended clinical use of such systems, structured assessment can support informed decisions on whether, where, and under which conditions they should be implemented. This includes consideration of test performance (diagnostic and technical), patient-relevant clinical outcomes, organisational implications, and resource requirements.

The aim of this assessment was therefore twofold: first, to identify and prioritise relevant AI-supported diagnostic imaging applications for hospital use in Austria; and second, to evaluate the clinical effectiveness, organisational implications, and resource considerations of the selected application. In addition, the assessment aimed to explore the applicability of the procurement checklist for AI-enabled DHTs developed by the Austrian Institute for Health Technology Assessment (AIHTA) and the ASSESS-DHT taxonomy and guidance to this topic.

Methods

First, relevant AI applications in diagnostic imaging were identified and prioritised using expert consultation and a structured shortlist drawn from prior reports by AIHTA and Gesundheit Österreich GmbH (GÖG).

Second, a systematic review was conducted following a predefined and publicly available protocol. The evidence identification followed a staged approach: we first searched for high-quality HTAs and systematic reviews and then updated the evidence through a supplementary search for primary studies (search date: August 2025). Study screening, selection, extraction, and risk-of-bias assessment were performed in duplicate. Eligible studies included those assessing AI alone or AI-assisted radiologist reading versus radiologist-only interpretation in adults undergoing CXR for suspected lung cancer.

Outcomes included diagnostic accuracy for lung cancer and nodule detection, technical performance (failure rates and concordance), patient-relevant clinical outcomes, organisational implications, and cost/resource considerations. Evidence was summarised narratively.

Results

Following the priorisation process, AI-supported chest X-ray (CXR) analysis for suspected lung cancer was chosen for detailed assessment based on clinical relevance, routine hospital use, and expected applicability. Furthermore, lung cancer remains among the leading causes of cancer-related mortality, and CXR is widely used as a first-line imaging modality in symptomatic patients. In addition, in Austria AI-supported CXR was identified as either already implemented in clinical settings or currently in the testing phase.

Three HTAs from UK organisations (Scottish Health Technology Group, National Institute for Health and Care Research, and Cedar) were included, comprising 13 unique primary studies. Two additional primary studies were identified through the supplementary update search, resulting in 15 unique primary studies overall. Thirteen of the 15 studies were retrospective cohort analyses, while two studies combined a retrospective and a prospective phase; most studies were conducted in single-centre settings. Across the evidence base, 14 software products were evaluated. No study was conducted in Austria.

Clinical and technical performance

No study reported patient-relevant clinical outcomes (mortality, morbidity, health-related quality of life, or safety) or demonstrated that AI use leads to earlier lung cancer diagnosis, stage shift or faster treatment initiation.

Lung cancer detection

AI-assistance increased sensitivity in some studies without affecting specificity. Stand-alone AI showed variable performance across settings, with consistently high negative predictive values but often low positive predictive values, particularly in low prevalence populations. In practical terms, when the AI classified a CXR as not suspicious, this was generally associated with a low likelihood of lung cancer being present according to the study reference standard (histological confirmation or radiologist assessment), whereas a CXR classified as suspicious frequently represented false alarms.

Nodule detection

Stand-alone AI systems performed similarly or slightly better than radiologists in sensitivity, with widely varying specificity. AI-assisted radiologist reading consistently improved sensitivity and led to a small increase in the proportion of correctly classified cases. None of the studies evaluated performance in patient subgroups (age, sex, ethnicity) or across multiple clinical settings.

Organisational outcomes

Evidence on organisational implications was sparse. AI integration had a mixed impact on workflow: some studies suggested improved efficiency in reporting or triage, while others highlighted increased CT referrals due to false positives. Clinician surveys indicated cautious support for AI use but concerns regarding accuracy, transparency, and potential distancing effects on patient care. No studies reported staff training needs, human–AI interaction outcomes, or long-term workflow changes.

Cost and resource implications

Based on the available studies the following cost categories could be outlined: software licensing, implementation and integration costs, training costs, staff time costs, potential increases in diagnostic and downstream healthcare costs, and cancer treatment costs.

Discussion

Across the current evidence base, AI-supported CXR interpretation has not yet demonstrated meaningful clinical benefit in lung cancer pathways. Evidence is limited to surrogate outcomes (abnormality or nodule detection) rather than patient-relevant endpoints. Improvements in sensitivity are inconsistently reported and often offset by lower specificity, especially for stand-alone AI, which may lead to increased CT workload and patient anxiety.

Major limitations are predominantly retrospective study designs, study populations with higher disease prevalence than routine care, inconsistent reference standards (histological confirmation versus radiologist assessment), limited transparency on training data, and a lack of real-world implementation studies.

The role of AI within the reading workflow (first-reader vs second-reader vs triage) remains unclear and likely influences both benefits and risks. Additionally, many commercially available AI tools have no published evidence at all. Substantial evidence gaps remain regarding clinical effectiveness, safety, workflow integration, resource use, and long-term economic impact. Transparency about model development and dataset composition is insufficient across most tools.

Conclusion

Current evidence is insufficient to demonstrate that AI-assisted CXR interpretation provides added value in lung cancer pathways in Austrian hospitals. Available studies focus on technical outcomes, with no demonstrated improvements in patient-relevant outcomes or healthcare efficiency. While AI may offer potential benefits in assisting radiologists or improving sensitivity for nodule detection, uncertainties around accuracy, workflow integration, transparency, equity, and costs remain substantial.

Future research should include prospective, real-world evaluations that follow patients throughout the diagnostic pathway. These studies should evaluate clinical and organisational outcomes, ensure transparency of training datasets, and assess local calibration and applicability. Without such evidence, the overall added value of AI-supported CXR for lung cancer detection in Austria remains undetermined.

Zusammenfassung

Hintergrund

Künstliche Intelligenz (KI) wird in der medizinischen Bildgebung immer häufiger eingesetzt. In den letzten Jahren wurden zahlreiche Systeme entwickelt, die bei der Interpretation von Bildbefunden helfen, Arbeitsabläufe effizienter zu gestalten und klinische Entscheidungen in Krankenhäusern zu unterstützen. Die Radiologie gehört zu jenen Fachbereichen, die besonders stark von steigenden Arbeitsbelastungen, Personalmangel und zunehmend komplexen Diagnoseaufgaben betroffen sind. Daher ist sie ein vorrangiger Bereich für KI-gestützte digitale Gesundheitstechnologien.

KI-Systeme für verschiedene bildgebende Verfahren werden vor allem dafür entwickelt, verdächtige Befunde zu erkennen, einzuordnen oder nach Dringlichkeit zu reihen. Viele dieser Anwendungen werden als Entscheidungshilfen oder Triage-Systeme vermarktet und sind als CE-zertifizierte Medizinprodukte zugelassen. Zu ihren möglichen Vorteilen zählen eine verbesserte Diagnosegenauigkeit, kürzere Bearbeitungszeiten und ein effizienterer Einsatz radiologischer Ressourcen. Angesichts der wachsenden Verfügbarkeit und des geplanten klinischen Einsatzes solcher Systeme kann eine strukturierte Bewertung dabei helfen, fundierte Entscheidungen darüber zu treffen, ob, wo und unter welchen Bedingungen sie eingeführt werden sollten. Berücksichtigt werden dabei die Testleistung (diagnostisch und technisch), patient:innenrelevante klinische Ergebnisse, organisatorische Auswirkungen und Ressourcenbedarf. Das vorliegende Review verfolgte daher zwei Ziele: Erstens sollte eine relevante KI-gestützte Anwendung für die diagnostische Bildgebung im österreichischen Krankenhausbereich identifiziert und priorisiert werden. Zweitens sollte die ausgewählte Anwendung hinsichtlich ihrer klinischen Wirksamkeit, organisatorischen Auswirkungen und des Ressourcenbedarfs detailliert untersucht werden. Ein zusätzliches Ziel war es, die Anwendbarkeit der vom Austrian Institute for Health Technology Assessment (AIHTA) entwickelten Checkliste für die Beschaffung von KI-Anwendungen sowie der ASSESS-DHT-Taxonomie und des dazugehörigen Handbuchs zu prüfen. Unter Testleistung wird dabei die Erkennungsgenauigkeit der KI verstanden, gemessen wird sie anhand von Sensitivität, Spezifität und weiteren diagnostischen Kennzahlen.

Methode

Es wurde ein zweistufiger Ansatz angewandt. Zunächst wurden relevante KI-Anwendungen in der diagnostischen Bildgebung identifiziert und mittels Experten:innenkonsultation sowie einer strukturierten Auswahlliste aus früheren Berichten von AIHTA und der Gesundheit Österreich GmbH (GÖG) priorisiert.

Im zweiten Schritt wurde ein systematisches Review nach einem vordefinierten und öffentlich zugänglichen Protokoll durchgeführt. Zuerst wurden hochwertige Health Technology Assessments (HTAs) identifiziert und wenn vorhanden durch eine ergänzende Suche nach Primärstudien aktualisiert (Suchzeitpunkt: August 2025). Das Screening der Studien, die Auswahl, Datenextraktion und Bewertung des Verzerrungsrisikos mittels ROBIS erfolgten im Zweipersonenprinzip. Eingeschlossen wurden Studien, die entweder eigenständige KI-Systeme oder KI-unterstützte radiologische Befundung mit der alleinigen Befundung durch Radiolog:innen bei erwachsenen Patient:innen mit Thorax-Röntgenaufnahmen bei Verdacht auf Lungenkrebs verglichen.

Die untersuchten Endpunkte umfassten die diagnostische Genauigkeit bei der Erkennung von Lungenkrebs und Lungenknoten, die technische Testleistung, klinische Ergebnisse, organisatorische Auswirkungen sowie Kosten- und Ressourcenaspekte. Die diagnostische Genauigkeit wurde anhand folgender Kriterien bewertet: Sensitivität (Erkennung tatsächlich Erkrankter), Spezifität (Erkennung tatsächlich Gesunder), positiven und negativen prädiktiven Werten (Wahrscheinlichkeit, dass ein KI-Befund korrekt ist), Genauigkeit (Anteil korrekt klassifizierter Fälle), AUROC (Diskriminierungsfähigkeit) und Konkordanz (Übereinstimmung mit Referenzstandard). Zur technischen Testleistung gehörten die technische Ausfallrate (Fälle, in denen die Software ein Bild nicht analysieren kann) sowie die Konkordanz – das Ausmaß, in dem KI- und Nicht-KI-Technologien vergleichbare Ergebnisse liefern. Letztere gilt als wichtiger Vertrauensindikator für die Leistungsfähigkeit von KI-Software. Die verfügbare Evidenz wurde in narrativer Form zusammengefasst.

Ergebnisse

Nach der Identifikation und Priorisierung von KI-Instrumenten in der diagnostischen Bildgebung wurde die KI-gestützte Analyse von Thorax-Röntgenaufnahmen bei Verdacht auf Lungenkrebs für die detaillierte Bewertung ausgewählt, da sie klinisch relevant ist, routinemäßig in Krankenhäusern eingesetzt wird und voraussichtlich gut anwendbar ist. Lungenkrebs zählt nach wie vor zu den häufigsten krebsbedingten Todesursachen, und die Thorax-Röntgenaufnahme wird üblicherweise als bildgebendes Verfahren der ersten Wahl bei Patient:innen mit entsprechenden Symptomen verwendet. Darüber hinaus wurde festgestellt, dass KI-gestützte Thorax-Röntgenaufnahmen in Österreich bereits in einer Pilotphase erprobt werden.

Drei Health Technology Assessments (HTAs) von britischen Organisationen (Scottish Health Technology Group, National Institute for Health and Care Research und Cedar) wurden eingeschlossen und umfassten insgesamt 13 eigenständige Primärstudien. Durch eine ergänzende systematische Literatursuche konnten zwei weitere Primärstudien identifiziert werden, sodass insgesamt 15 Primärstudien vorlagen. Die meisten Studien (13 von 15) waren retrospektive Kohortenanalysen, die überwiegend als Single-Center-Studien durchgeführt wurden, wobei die zwei anderen sowohl retrospektive als auch prospektive Phasen enthielten. Die Studien umfassten insgesamt 14 verschiedene Softwareprodukte. Keine der Studien wurde in Österreich durchgeführt.

Klinische und technische Leistungsfähigkeit

Keine der eingeschlossenen Studien berichtete über patienten:innenrelevante klinische Endpunkte (Mortalität, Morbidität, gesundheitsbezogene Lebensqualität oder Sicherheit) oder wies nach, dass der Einsatz von KI zu einer früheren Diagnose von Lungenkrebs, einer Verschiebung des Stadiums oder einem schnelleren Behandlungsbeginn führt.

Lungenkrebs-Detektion

Die KI-Unterstützung bei der Interpretation von Befunden erhöhte in sieben Studien die Sensitivität, ohne die Spezifität (zu beeinträchtigen. Die Konkordanz – die Übereinstimmung zwischen Befundenden und Referenzstandard – war mit KI-Assistenz höher (z. B. 57 % vs. 42 %), was bedeutet, dass Radiolog:innen mit KI-Unterstützung häufiger mit dem Referenzstandard übereinstimmten als ohne Unterstützung. Eigenständig arbeitende KI-Systeme zeigten je nach Studiensetting unterschiedliche Testleistungen. Dabei wurden durchgehend hohe negative prädiktive Werte erreicht, während die positiven prädiktiven Werte häufig niedrig ausfielen, insbesondere in Populationen mit geringer Prävalenz. In der Praxis bedeutet dies: Wenn die KI eine Thorax-Röntgenaufnahme als unauffällig einstufte, war die Wahrscheinlichkeit für das Vorliegen von Lungenkrebs gemäß dem jeweiligen Referenzstandard der Studie in der Regel gering. Wurde eine Aufnahme hingegen als auffällig klassifiziert, handelte es sich häufig um Fehlalarme. In Studien aus der Routineversorgung zeigten sich sehr niedrige positive prädiktive Werte (PPV) (1-6 %), was bedeutet, dass die meisten positiven KI-Befunde falsch-positiv waren, während negativen prädiktiven Werte (NPV) sehr hoch blieben (~99 %), was auf eine starke Ausschlussleistung hinweist. Eine Studie zeigte jedoch in einer Population mit höherer Krankheitsprävalenz einen hohen PPV (97 %), aber einen niedrigen NPV (62 %), was einen wichtigen Trade-off verdeutlicht: Entweder produziert das System viele Falschpositive (niedrige Spezifität/PPV) oder es zeigt eine schwache Ausschlussleistung (niedriger NPV). Als alleiniges Triage-Instrument ist die eigenständige KI daher nur eingeschränkt zuverlässig; KI-Assistenz kann hingegen die Übereinstimmung mit dem Referenzstandard verbessern.

Knotendetektion

Alleinstehende KI-Systeme zeigten ähnliche oder leicht bessere Sensitivität als Radiolog:innen, mit stark variierender Spezifität. Eine Studie verglich sieben kommerzielle KI-Systeme mit Radiolog:innen und fand KI-Sensitivitäten von 64-93 % gegenüber 81 % bei menschlichen Befundenden. Die leistungsstärksten Systeme (Lunit INSIGHT und Annalise.ai) erreichten Sensitivitäten über 90 %, während die Spezifität breiter streute (50-89 % für KI vs. 71 % für Radiolog:innen). Die KI-assistierte radiologische Befundinterpretation verbesserte die Sensitivität konsistent und führte zu einem leicht höheren Anteil kor-

rekt klassifizierter Fälle. Sensitivitätsverbesserungen lagen typischerweise bei 5-13 %, das bedeutet, die KI übersah seltener tatsächliche Fälle. Die Spezifität stieg moderat an (von etwa 78-93 % auf 82-97 %). Der Anteil korrekt klassifizierter Fälle erhöhte sich geringfügig (z. B. von 70 % auf 75 % oder von 84-90 % auf 90-91 %). Allerdings war die Treffsicherheit bei unauffälligen Aufnahmen sehr unterschiedlich – je nach System gab es mehr oder weniger Fehlalarme.

Keine der Studien evaluierte die Leistungsfähigkeit in Patient:innensubgruppen (Alter, Geschlecht, Ethnizität) oder über mehrere klinischen Settings hinweg.

Organisatorische Ergebnisse

Die Evidenz zu organisatorischen Auswirkungen war begrenzt. Die Integration von KI-Systemen zeigte unterschiedliche organisatorische Versorgungseffekte auf die Arbeitsabläufe: Einige Studien deuteten auf eine verbesserte Effizienz bei der Befundung oder Triage hin, während andere auf eine Zunahme von CT-Überweisungen aufgrund falsch-positiver Befunde hinwiesen. Befragungen von Kliniker:innen zeigten eine vorsichtig positive Haltung gegenüber dem KI-Einsatz, gleichzeitig wurden jedoch Bedenken hinsichtlich der Genauigkeit, Transparenz und möglicher negativer Auswirkungen auf die Arzt-Patient:innen-Beziehung geäußert. Keine der Studien berichtete über den Schulungsbedarf des Personals, Ergebnisse der Mensch-KI-Interaktion oder langfristige Veränderungen der Arbeitsabläufe.

Kosten- und Ressourcenaspekte

Auf Basis der verfügbaren Studien konnten folgende Kostenkategorien identifiziert werden: Softwarelizenzen, Implementierungs- und Integrationskosten, Schulungskosten, Personalzeitkosten, potenzielle Kostensteigerungen in der Diagnostik und nachgelagerten Gesundheitsversorgung sowie Kosten für die Krebsbehandlung.

Eine HTA-Bewertung (NHS Scotland) verglich die Kosten eines KI-gestützten Versorgungspfads mit der Standardversorgung. Es zeigten sich geringe Mehrkosten, die hauptsächlich auf KI-Software und zusätzlichen Personalaufwand zurückzuführen waren. Weitere Kosteneffekte durch nachgelagerte Versorgungsleistungen konnten aufgrund fehlender Evidenz nicht bestimmt werden. Österreichische Daten lagen nicht vor.

Diskussion

Auf Grundlage der derzeit verfügbaren Evidenz konnte für die KI-gestützte Befundung von Thorax-Röntgenaufnahmen noch kein Zusatznutzen in der Versorgung von Lungenkrebspatienten nachgewiesen werden. Die Evidenz beschränkte sich auf Surrogatendpunkte (indirekte Endpunkte) wie die Erkennung von Auffälligkeiten oder Lungenknoten, während patient:innenrelevante Ergebnisse nicht untersucht wurden. Verbesserungen der Sensitivität wurden uneinheitlich berichtet und häufig durch eine geringere Spezifität ausgeglichen, insbesondere bei eigenständig arbeitenden KI-Systemen. Diese Entwicklung kann zu einer erhöhten CT-Untersuchungslast und Patient:innenängsten führen.

Wesentliche Einschränkungen umfassen überwiegend retrospektive Studiendesigns, Studienpopulationen mit höherer Krankheitsprävalenz als in der Routineversorgung, inkonsistente Referenzstandards (histologische Bestätigung versus radiologische Beurteilung), eingeschränkte Transparenz bezüglich der Trainingsdaten sowie das Fehlen von Implementierungsstudien.

Die Rolle der KI innerhalb des Befundungsprozesses (Erstbefunder, Zweitbefunder oder Triage) wurde nicht systematisch untersucht, kann jedoch sowohl Nutzen als auch Risiken beeinflussen. Für zahlreiche kommerzielle KI-Anwendungen liegen keine publizierten Studien vor.

Erhebliche Evidenzlücken bestehen hinsichtlich der klinischen Wirksamkeit, der Sicherheit, der Workflow-Integration, der Ressourcennutzung und der langfristigen ökonomischen Auswirkungen. Die Transparenz bezüglich der Modellentwicklung und Zusammensetzung der Datensätze ist bei den meisten Anwendungen unzureichend. Für die meisten kommerziellen Systeme existieren kaum öffentlich zugängliche Informationen zu Trainingsdatensätzen, zu den verwendeten Algorithmen oder zu Verfahren für Modell-Updates und Post-Market-Performance-Monitoring.

Im Rahmen dieses Berichts wurde die vom AIHTA entwickelte Checkliste für die Beschaffung von KI-Anwendungen erstmals auf ihre Anwendbarkeit bei KI-gestützten Thorax-Röntgen-Instrumenten getestet. Die Analyse zeigte, dass die Checkliste grundsätzlich geeignet ist, jedoch für medizinische KI-Systeme zwei zusätzliche Aspekte berücksichtigen sollte: Eine detaillierte Validierung der Datensatzrepräsentativität und transparente Dokumentation des Modell-Lebenszyklus und Post-Market-Performance-Monitoring.

Die fehlende Transparenz zu Trainingsdaten erschwert die Bewertung, ob die Systeme für alle Patient:innengruppen gleichermaßen zuverlässig funktionieren. Equity-Bedenken entstehen, wenn Trainingsdatensätze nicht ethnisch, demographisch oder klinisch repräsentativ für die Zielpopulation sind. Keine der identifizierten Studien führte Subgruppenanalysen durch (z. B. nach Alter, Geschlecht, Ethnizität, Komorbiditäten oder sozioökonomischem Status), und es liegen keine Kalibrierungsstudien für die österreichische Population vor. Dies unterstreicht die Notwendigkeit lokaler Validierung und Fairness-Assessments vor großflächigem Einsatz.

Schlussfolgerung

Die verfügbare Evidenz ist unzureichend, um einen Zusatznutzen KI-gestützter Thorax-Röntgenbefundung bei Verdacht auf Lungenkrebs in österreichischen Krankenhäusern zu belegen. Die eingeschlossenen Studien konzentrieren sich auf technische Endpunkte; Verbesserungen bei patient:innenrelevanten Ergebnissen oder der Effizienz der Gesundheitsversorgung konnten nicht nachgewiesen werden. Obwohl KI potenziell Radiolog:innen unterstützen oder die Sensitivität bei der Erkennung von Lungenknoten verbessern könnte, bestehen erhebliche Evidenzlücken hinsichtlich Genauigkeit, Workflow-Integration, Transparenz, Chancengleichheit und Kosten.

Um fundierte Entscheidungen über die Einführung solcher Systeme treffen zu können, sollte zukünftige Forschung prospektive Evaluierungen unter realen Versorgungsbedingungen umfassen, die Patient:innen entlang des gesamten Diagnosepfads begleiten, klinische und organisatorische Ergebnisse bewerten, Transparenz der Trainingsdatensätze gewährleisten sowie die lokale Kalibrierung und Anwendbarkeit prüfen. Insbesondere fehlen Studien, die Evidenz zu technischen Ausfallraten, Veränderungen in der klinischen Entscheidungsfindung, Kosten und Ressourcennutzung berichten. Auch algorithmische Verzerrungen, Fairness und langfristige Outcomes sollten bewertet werden. Ohne eine solche Evidenzbasis bleibt der tatsächliche Mehrwert KI-gestützter Thorax-Röntgenaufnahmen zur Lungenkrebserkennung in Österreich ungeklärt.

1 Background

KI in diagnostischen Bildgebung

Artificial intelligence (AI) applications in diagnostic imaging have expanded rapidly in recent years, with a growing number of systems developed for use in hospital settings [1-3]. Radiology has been identified as one of the clinical domains most affected by increasing workloads, staff shortages, and rising diagnostic complexity, making it a priority area for AI-supported digital health technologies (DHTs) [4].

klinische Rolle und erwartete Effekte

Across imaging modalities, AI systems are most frequently designed to assist with the detection, classification, or prioritisation of suspected pathologies [5, 6]. Many of these tools are marketed as decision-support or triage systems intended to assist clinicians in image interpretation and workflow prioritisation [6, 7]. Their potential benefits include improved diagnostic accuracy, shorter reporting or turnaround times, and more efficient use of radiology resources [7, 8]. Given the increasing availability and intended clinical use of such systems, structured assessment can support informed decisions on whether, where, and under which conditions they should be implemented. This includes consideration of diagnostic and technical test performance, patient-relevant clinical outcomes, organisational implications, and resource requirements.

2 Scope

The aim of this report is to provide an overview of AI-enabled DHTs in the field of diagnostic imaging and prioritise those currently in use or considered most relevant in Austrian hospitals, and to evaluate the clinical and organisational impacts, as well as the types of resources to be considered of the prioritised AI-enabled DHT. This systematic review addresses these aims through two research questions (RQ).

Ziel: Überblick & Priorisierung KI-DHTs in Bildgebung; Bewertung der priorisierten KI-DHT

2.1 Research questions

Forschungsfragen

RQ1: Which AI-enabled DHTs in the fields of diagnostic imaging are considered most relevant in Austrian hospitals by Austrian healthcare experts?

RQ2: What is the clinical effectiveness, what are the organisational implications, and what types of resources are needed for implementing the selected AI-enabled DHT in diagnostic imaging? Specifically, sub-questions for the selected applications.

Fokus auf klinische und organisatorische Auswirkungen

In particular, the review aimed to examine how selected AI applications in diagnostic imaging influence diagnostic accuracy and efficiency in hospital settings, including potential risks, limitations, and unintended consequences. In addition, it sought to explore the organisational implications of integrating AI into imaging workflows, such as its impact on staff training, resource allocation, and workflow structure. Finally, the review assessed resource-related impacts, including those associated with the acquisition, implementation, and interaction of AI technologies with existing diagnostic resources.

präregistriertes Protokoll ohne Abweichungen

The objectives, inclusion criteria and methods for this review were specified in advance and documented in a protocol on the Austrian Institute for Health Technology assessment (AIHTA) website as well as on the Open Science Framework platform. There were no protocol deviations.

ASSESS-DHT-Anwendbarkeit für KI-Diagnostik geprüft

Additionally, the ASSESS DHT guidance documents are piloted to examine their applicability for the assessment of AI-supported diagnostic imaging tools and to identify potential adaptations to improve their usability and relevance.

3 Methods

3.1 Selection of the AI-enabled DHT for assessment (RQ1)

FF1: GÖG-Bericht und HTA-Reviews als Evidenzquellen

Stakeholder basierte Priorisierung mittels Online-Survey

Fokus auf diagnostische Bildgebung

Erstellung strukturierter „Long List“

FF2: EUnetHTA Core Model® und ASSESS-DHT als Framework

To support prioritisation, we identified candidate topics from two complementary sources. First, we screened the Gesundheit Österreich GmbH (GÖG) report to identify Austrian use cases or pilot projects using AI and retained only those in diagnostic imaging. Second, we used the AIHTA scoping report's evidence base to compile HTA reports and systematic reviews assessing AI technologies and again retained only those in diagnostic imaging. These two subsets were merged into a single, structured list ("long list") that served as the basis for subsequent prioritisation. Selected Austrian stakeholders (providers, healthcare professionals, chief IT officers in selected Austrian hospitals) were consulted via a brief online survey with targeted e-mail follow-up. Experts were asked to prioritise items on the long list based on criteria such as clinical relevance (novelty of the technology, addressing clinical need, potential to improve patient outcomes or clinical workflows, and availability of evidence), resource implications (frequency of use, costs, expected impact on healthcare resource use), and feasibility of implementation (including potential barriers such as organisational resistance, infrastructure limitations, data security concerns) and potential risks or unintended consequences (e.g. diagnostic errors, increased workload, ethical concerns). The "long list" (Table A-1) and structured survey ("Survey: KI-Anwendungsbereiche") can be found in the Appendix.

3.2 Assessment of the selected AI-enabled DHT (RQ2)

The review is registered on OSF (registration number D458b).

FF2: EUnetHTA Core Model® und ASSESS-DHT als Framework

The European Network for Health Technology Assessment (EUnetHTA) Core Model® was used as reporting framework. Also, interim version of methods and taxonomy documents of the ongoing European project for the assessment of DHTs, called ASSESS-DHT, was piloted in the present SR.

Multi-Domain-Bewertungsansatz

This assessment employed a multi-domain assessment approach, following the EUnetHTA methodology [9] (see guiding question in the Appendix Research questions).

3.2.1 Inclusion criteria

Einschlusskriterien

To answer RQ2, the inclusion criteria for relevant studies are summarised in Table 3-1:

Table 3-1: Inclusion criteria

Population	Patients with *suspected* lung cancer referred to chest X-ray from primary care
Intervention	AI software used in the interpretation of chest X-ray images (possible software, including but not restricted to: Annalise Enterprise CXR, qXR, AI-Rad Companion Chest X-ray, Auto Lung Nodule Detection, ChestLink, ChestView, Chest X-ray, ClearRead Xray, InferRead DR Chest, Lunit INSIGHT CXR, Milvue Suite, Red dot, SenseCare-Chest DR PRO, VUNO Med-Chest X-Ray, Gleamer, Veolity) (AI alone or AI in conjunction with a radiologist)[1]
Control	**Interpretation** of chest X-ray images by radiologists only
Reference standard	For accuracy of *lung cancer* detection: lung cancer confirmed by histological analysis of lung biopsy. For accuracy of *suspicious nodule* detection: radiology specialist (single reader or consensus of more than one reader).
Outcomes	
Efficacy and Safety	*Clinical and test performance:* ■ 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 accuracy for the detection of lung nodules, ■ Concordance in lung nodule detection between radiology specialist with and without adjunct AI, ■ Technical failure, ■ Mortality, ■ Morbidity, ■ Health-related quality of life (HrQoL) ■ Safety *Organisational:* ■ Turnaround time (image review to radiology report), ■ Timeframe for follow-up CT scans, or receiving a diagnosis, ■ Acceptability of AI software to clinicians (e.g., user-friendliness) ■ Impact on clinical decision-making, ■ Impact on use of resources (e.g., staff training, integration into existing systems) ■ Impact of false positives on the workflow *Costs (types of resources)*
Study design	
Efficacy and Safety	Two-phase-approach: ■ HTA reports and systematic reviews ■ Randomized and non-controlled trials, prospective observational studies, retrospective cohort studies

Abbreviations: AI … Artificial Intelligence, CT … Computed Tomography, HrQoL … Health-relatedQuality of Life, HTA … Health Technology Assessment

Ausschluss: nicht KI-CAD und asymptomatisches Screening

Studies were excluded if they named computer-aided detection that does not include AI software. Also, studies of people who do not have signs and symptoms of cancer or a suspected condition or trauma (i.e. people undergoing health screening) were out of scope.

[1] 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.

3.2.2 Literature search

zweistufige systematische Literatursuche in 4 Datenbanken

The systematic literature search was undertaken in two steps. First, health technology assessments (HTAs) and systematic reviews (SRs) were searched on 1st August, 2025 in four bibliographic databases:

HTAs & SRs

- Medline via Ovid
- Embase
- The Cochrane Library
- HTA (INAHTA)

Sprache: Englisch & Deutsch

The systematic search was limited to English or German language publications. The specific search strategy employed can be found in the Appendix.

3 HTAs/SRs identifiziert, SHTG-Review aktualisiert;

Qualitätsbewertung: ROBIS

The objective of this initial search was to identify existing high-quality evidence syntheses to avoid duplication of work and to determine whether updates to existing reviews were warranted. Identified reviews and HTAs were critically appraised using ROBIS (Risk of Bias Assessment Tool for Systematic Reviews), and the most comprehensive and methodologically robust review was selected for updating. In this step we identified three HTAs/SRs fitting our scope. We concluded that the review by SHTG was appropriate for inclusion and update (risk of bias assessment can be found in the Appendix, Table A-4).

zweiter Schritt: Suche nach Primärstudien; Zeitraum: nach Juli 2024

In the second step, primary studies published after the latest search date of the selected review (July 2024) were searched on 8th August, likewise in the four bibliographic databases. Search terms and search strategies were taken from the review we chose to update and are available in the Appendix.

ergänzende Handsuche

A targeted hand search complemented the systematic search. The search strategy combined controlled vocabulary (MeSH) and free-text terms for *artificial intelligence*, *machine learning*, *deep learning*, *radiology*, *chest X-ray*, and *lung cancer*. Reference lists of relevant publications and websites of HTA bodies (e.g., NICE, CADTH, and IQWiG) were also screened to identify additional publications reporting on organisational and cost outcomes. The search and selection processes were documented according to PRISMA standards.

Dokumentation nach PRISMA-Standards

Literatur zu Technologie & Gesundheitsproblematik

Literature for the description of the technology and literature concerning health problems and current use of the technology were identified through the systematic search complemented by hand search.

ergänzende Suche nach laufenden Studien

For ongoing studies, a search was conducted on ClinicalTrials.gov on 5th November, 2025, using ‘Lung Cancer’ as the main term, combined with relevant AI and imaging terms (e.g., ‘AI Chest X-ray’, ‘AI Software’, ‘AI imaging’, ‘Chest X-ray interpretation’).

3.2.3 Literature selection

Literaturauswahl

In the first step, 815 hits were found through database search five through hand search. In the second step, 537 hits were found through database search and one through hand search. The abstracts and titles, as well as the full text articles were screened by two independent assessors (JE, LG). Differences were discussed and solved with the involvement of a third assessor. The selection process of systematic reviews and HTAs is displayed in Figure 3-1 and that of primary studies is displayed in Figure 3-2.

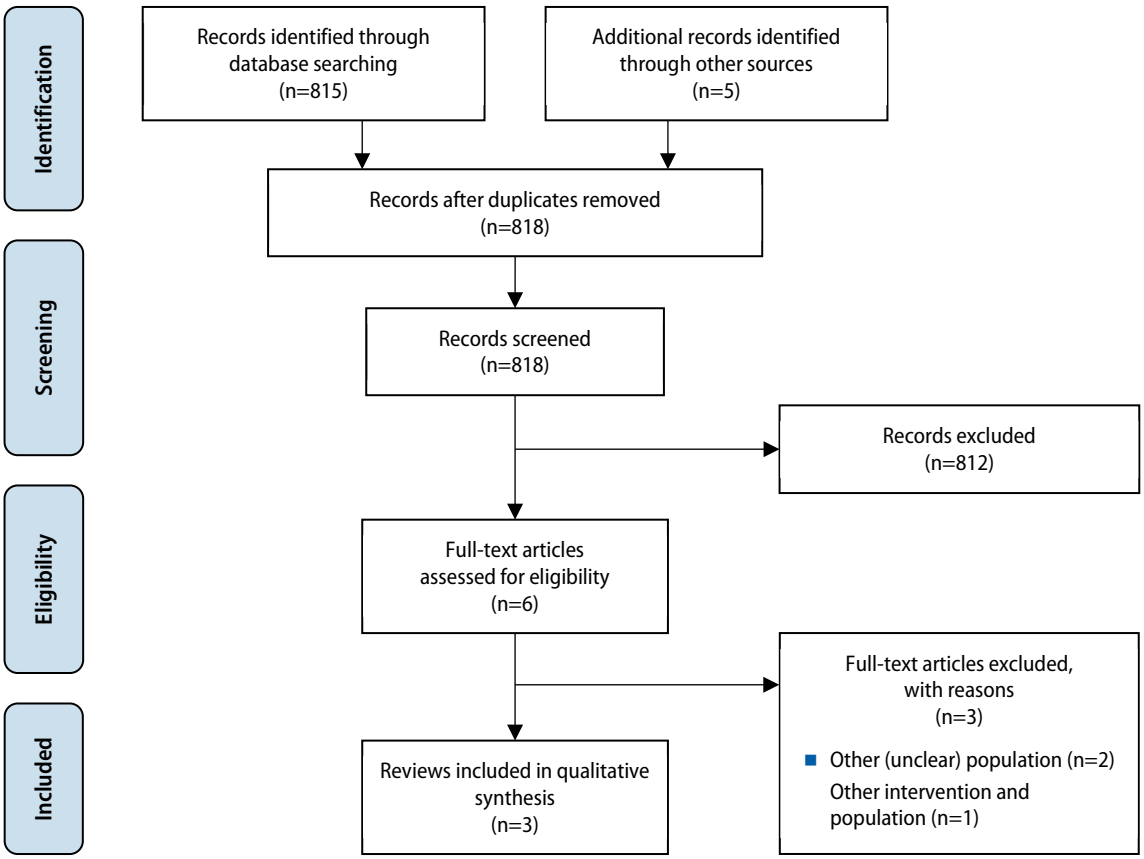


Figure 3-1: Selection process (PRISMA Flow Diagram) of systematic reviews and HTAs

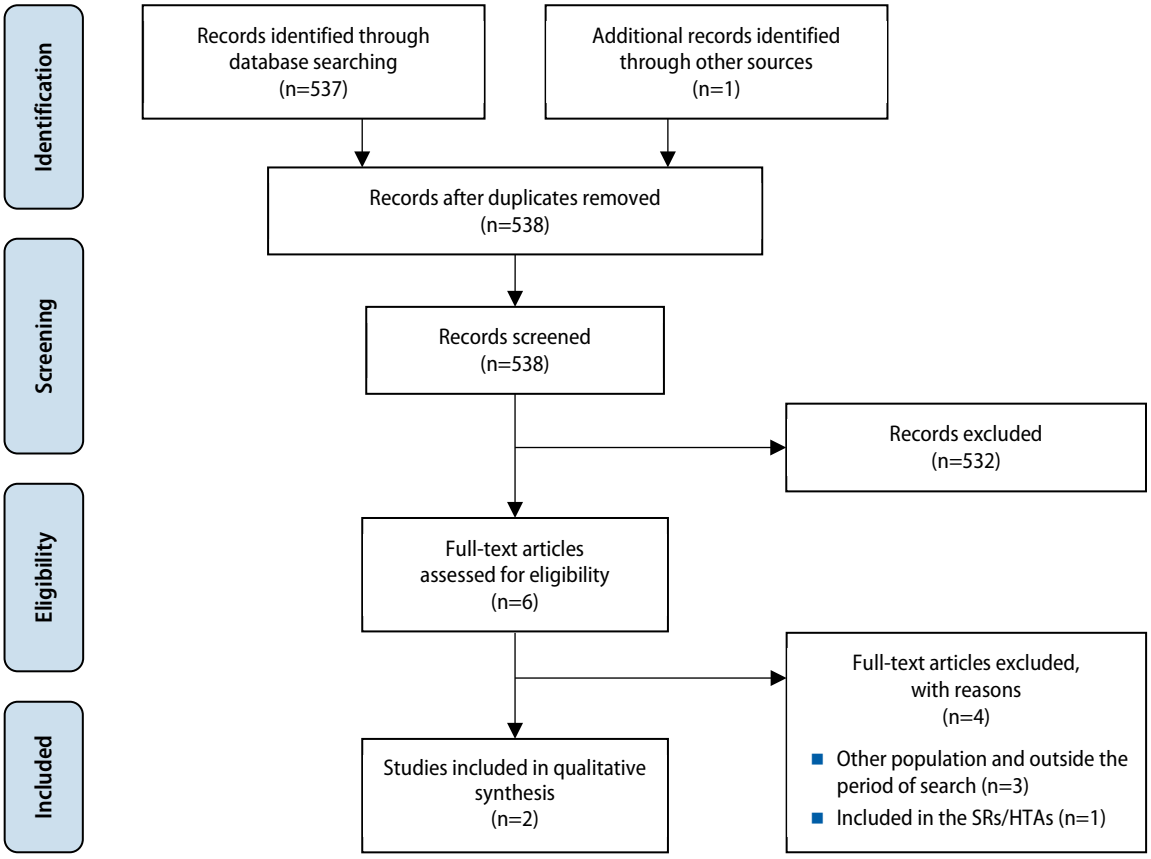


Figure 3-2: Selection process (PRISMA Flow Diagram) of primary studies

3.2.4 Analysis and synthesis of the evidence

Datenextraktion;

Bias-Bewertung

One reviewer (JE) systematically extracted relevant data from the included studies into standardised extraction tables, and a second reviewer (LG) cross-checked all entries against the original sources. Data were extracted from the primary studies included in the SRs. Risk of bias was assessed independently by two researchers (JE, LG) and differences were settled via consensus.

strukturierte Synthese und narrative Zusammenfassung

For *clinical outcomes*, the use of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to rate certainty of evidence was planned; for *non-clinical outcomes*, evidence limitations were summarised narratively. Based on the extraction tables (see Appendix, Table A-2, Table A-3), data for each outcome category were, where applicable, synthesised across studies and summarised narratively. The resource implications RQ was addressed by narratively summarising cost components and resource use reported in existing HTA reports and service evaluations.

4 Results

4.1 Selection of the AI-enabled DHT for assessment (RQ1)

„Long List" mit 16 KI-Anwendungen in Bildgebung

Radiologie dominiert (10/16)

The "long list" of AI-enabled DHTs identified from the AIHTA and GÖG reviews comprised 16 AI-enabled DHTs in the field of diagnostic imaging: 10 in radiology, one in internal medicine, one in pathology, one in dermatology, two in ophthalmology and one in general medicine. Six of them were identified from the GÖG report and eight from the AIHTA report, and two were mentioned in both reports (the full list can be found in the Appendix, Table A-1).

„Short List": 4 in Österreich pilotierte KI-Anwendungen

Despite repeated follow-up efforts, full completion of the structured expert survey could not be achieved. To ensure that expert input was nonetheless incorporated, the research team refined the list of candidate applications by examining which AI tools had been explicitly mentioned multiple times in the GÖG report as being piloted in Austria, and by verifying this information through targeted web searches. This process resulted in a short list of four applications with confirmed or likely ongoing local activities:

- AI-assisted chest X-ray for lung cancer detection,
- AI-supported brain CT analysis for stroke detection,
- AI-assisted X-ray interpretation for bone fracture detection, and
- AI-aided colonoscopy image analysis.

klinische Expertise priorisiert Thorax-Röntgen-KI

This shortlist was then discussed with a clinical expert not included in the initial consultation pool, who provided qualitative feedback and identified *AI-assisted chest X-ray analysis for lung cancer detection* as the most relevant topic for further assessment.

4.2 Assessment of the selected AI-enabled DHT (RQ2)

4.2.1 Overview of the health problem

Overview of the health problem and target population[2]

Zielgruppe: Erwachsene aus der Primärversorgung; CXR bei Krebsverdacht oder aus anderen Gründen

The target population in the scope of this assessment is adults referred from primary care who are: either undergoing chest X-ray (CXR) due to symptoms suggestive of lung cancer, for example cough, fatigue, shortness (symptomatic population) or undergoing CXR for reasons unrelated to lung cancer (incidental population).

Lungenkrebs: häufigste krebsbedingte Todesursache

hohe Krankheitslast in Österreich

Lung cancer is the most common cause of cancer-related death and the second most frequently diagnosed cancer worldwide [10, 11]. It is also a major health burden in Austria, with approximately 5,000 new cases annually (e.g., 5,232 in 2023), accounting for 11-12% of all cancer diagnoses [12-14]. While incidence in men has been declining for years, rates in women have increased and recently stabilised, narrowing the gender gap [14]. Lung cancer remains

[2] A0007, A0023, A0002, A0003, A0005, A0018.

the leading cause of cancer-related death, responsible for around 21% of male and 18% of female cancer deaths, and has one of the lowest survival rates among major cancers [15]. Prevalence in Austria is projected to rise substantially by 2030, with an estimated 23,700 people living with lung cancer, with the largest relative increases expected in the age groups 45-59 and over 75 years [16].

Histological subtypes

3 histologische Typen; 2 klinische Gruppen (NSCLC vs. SCLC)

Morphologie & Genetik variieren

Lung cancer is broadly classified into three histological subtypes: adenocarcinoma, squamous cell lung carcinoma – both grouped as non-small cell lung cancer (NSCLC) – and small cell lung cancer (SCLC). These subtypes differ in morphology and underlying genetic alterations. All are highly lethal, although notable progress has been made through targeted therapies, particularly for adenocarcinomas, and immunotherapy [10]. Lung adenocarcinoma (LUAD) is the most common subtype and the most frequently diagnosed in never-smokers [10].

Risk factors

Hauptrisiken: Rauchen, Alter, Umwelt- und Berufsexposition

Key risk factors include age, tobacco use (including second-hand smoke), radiation, air pollution, and occupational exposure to substances such as asbestos, arsenic, chromium, beryllium, and nickel. Smokers have about tenfold higher risk of developing lung cancer compared to never-smokers. Quitting smoking reduces precancerous changes and the overall risk of developing lung cancer [17]. Those at highest risk include smokers and individuals with occupational exposures, as well as people with a family history of the disease; less commonly, those with previous lung conditions [18-21].

Symptoms, natural course and burden of disease

langjähriger, stiller Krankheitsverlauf; lokale Tumoren oft zufällig entdeckt

Lung cancer typically develops over many years and is characterised by a long asymptomatic phase, during which small, or even moderate-sized tumours, especially localised ones, often remain clinically silent and are frequently detected only incidentally on imaging performed for unrelated reasons [20, 22]. Variable, non-specific symptoms might be present, such as [18, 20, 23]:

unspezifische Symptomvielfalt

- Cough: new or changing chronic cough.
- Bloody sputum: blood in the mucus when coughing.
- Shortness of breath on exertion or at rest.
- Pain in the chest, shoulders, or arms.
- Bone pain: might indicate metastases.
- Swelling in the face or neck.
- Weight loss: unintentional loss of over 5 kg.
- General weakness: fatigue, loss of appetite.
- Fever.

NSCLC: variable Wachstumsraten, heterogener Verlauf

SCLC: extrem aggressiv, schnelle Dissemination …

Progression commonly involves early spread to hilar and mediastinal lymph nodes, which substantially worsens prognosis. With further progression, distant metastases frequently occur, particularly to the brain, bones, liver, adrenal glands and the contralateral lung [24]. NSCLC typically shows variable growth rates, with tumour doubling times ranging from several weeks to many months, leading to considerable heterogeneity in its natural course [25]. In contrast, SCLC follows an extremely aggressive trajectory, characterised by rapid tumour growth, early dissemination and short survival without treatment – measured in weeks to a few months [26]. Overall, untreated advanced

NSCLC has a median survival of approximately 4-12 months depending on stage and performance status [25]. Early-stage NSCLC (stage I-II), although often asymptomatic, is potentially curable, with markedly better outcomes if detected before metastatic spread [27]. Early diagnosis is therefore critical. In Austria, survival rates have improved in recent years: one-year survival increased from 45% (2004-2008) to 60% (2022), and five-year survival from 17% to 24%, with women showing higher five-year survival than men (28% vs. 21%) [12].

… Früherkennung entscheidend;

Überlebensraten Österreich:
1-Jahr ↑ 45 → 60 %,
5-Jahr ↑ 17 → 24 %

The economic burden of lung cancer in Austria is reflected in several areas: per capita healthcare spending on cancer care is roughly at the EU average, while productivity losses due to cancer exceed the European mean. Treatment costs for NSCLC recently amounted to approximately €471 million [28]. For comparison, in Germany, lung cancer causes the highest disease burden among all cancers, with average treatment costs ranging from €7,600 to €20,200 per life-year gained [29].

ökonomische Belastung hoch;

EU:
NSCLC-Behandlungskosten:
~471 Mio. €

Current clinical diagnosis[3]

Lung cancer pathways are complex and contain many routes to diagnosis. Despite the existence of national guidelines and timelines for diagnosis in some countries, clinical practice can still vary widely across radiology departments.

vielfältige Zugänge zur Diagnosestellung

CXR remains the standard first-line diagnostic tool for symptomatic patients with suspected lung cancer in guidelines from Germany (S3 Leitlinie) [30] and the UK (NICE) [31]. However, a normal X-ray does not exclude cancer: if symptoms persist or risk factors are present, guidelines recommend further investigation, usually with computer tomography (CT) imaging or specialist referral. The German S3 guideline emphasises that CXR should be followed by contrast-enhanced chest CT in appropriate cases, especially before invasive diagnostics or treatment decisions. Diagnosis also involves a thorough medical history and clinical examination, assessment of risk factors (e.g., smoking, family history), tissue sampling through biopsy (typically via bronchoscopy or CT-guided puncture), and molecular analysis in cases of NSCLC to guide targeted therapies. The guideline stresses the importance of an interdisciplinary approach for optimal diagnostic accuracy and treatment planning [30].

CXR: Standard-Diagnostik bei Verdacht

persistierende Symptome → CT/Fachüberweisung

Anamnese, Risikofaktoren, klinische Untersuchung

interdisziplinärer Ansatz empfohlen

Since early-stage lung cancer rarely causes symptoms, abnormalities, particularly suspicious nodules, are often found incidentally. Pulmonary nodules typically defined as rounded opacities in the lung parenchyma measuring <10 mm in diameter. They are usually benign and asymptomatic, but a proportion – especially larger or morphologically suspicious nodules – represent early-stage malignancy. Nodules exceeding eight millimetres in diameter warrant closer monitoring, as they are more likely to be cancerous [32-34]. For this reason, their detection on CXR is a key trigger for further diagnostic workup, most commonly chest CT, positron emission tomography – computed tomography (PET-CT), or tissue sampling. A CXR may be flagged as suspicious for lung cancer when a nodule, lung mass, hilar enlargement, or a combination of these findings is present [35].

Frühstadium oft asymptomatisch

Zufallsbefunde: Lungenrundherde

CT, PET-CT, Biopsie zur Abklärung

The National Institute of Health and Care Excellence (NICE) recommends that all patients potentially eligible for curative treatment should undergo PET-CT before starting therapy. Contrast-enhanced CT of the chest, liver,

NICE-Leitlinie: PET-CT und multimodale Staging-Verfahren

[3] A0024.

adrenals, and lower neck should precede biopsy procedures. Endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) is preferred for mediastinal or peri-bronchial lesions, while endoscopic ultrasound-guided FNA (EUS-FNA) or surgical mediastinal staging is considered if nodal involvement remains uncertain. The guidelines emphasise an interdisciplinary approach to ensure accurate staging, guide treatment decisions, and optimise patient outcomes [31].

diagnostischer Pfad nach deutscher S3-Leitlinie

The diagnostic pathway illustrated in this report is based on the German S3 Guideline (Version 4.0, 2025) [30] and reflects the clinical practice context in Germany and Austria. A schematic presentation, created by the review authors (LG, JE) is depicted in Figure 4-1.

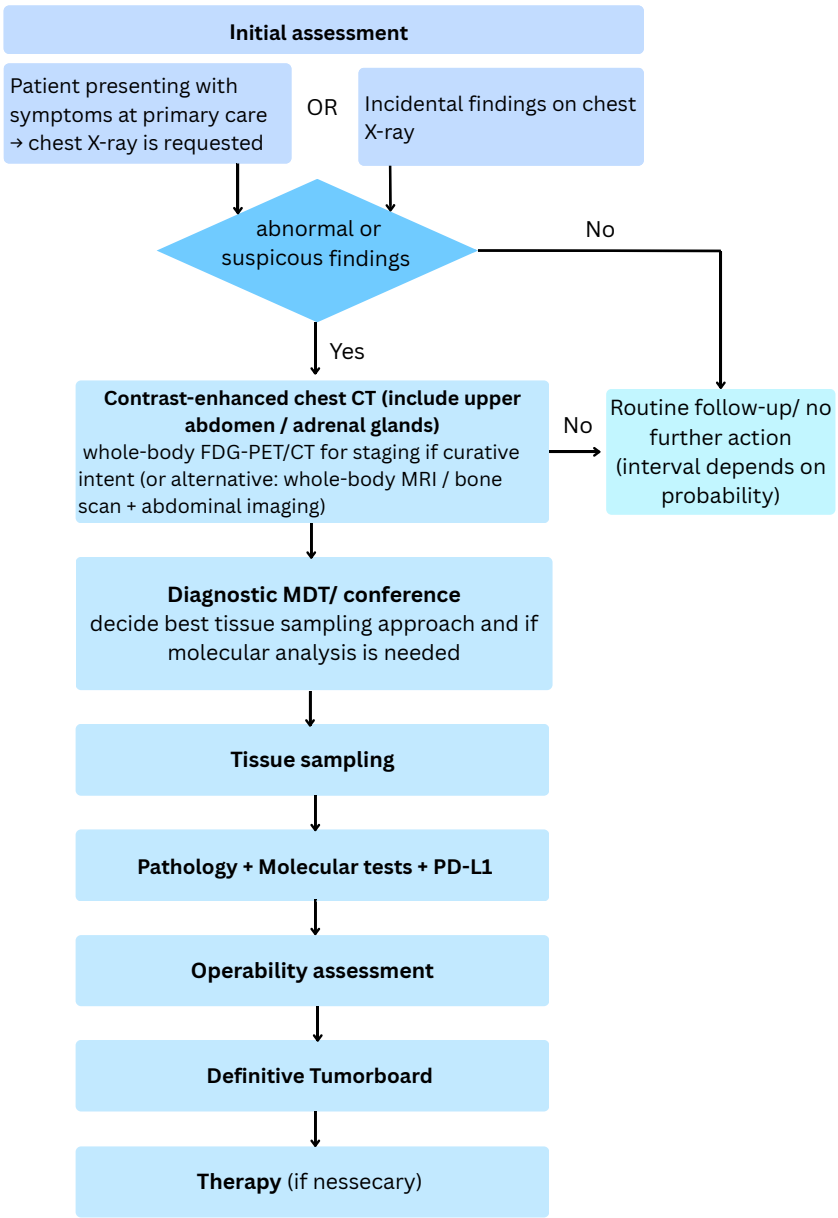


Figure 4-1: Diagnostic pathway (Source: adapted and created by AIHTA, based on S3 Guideline (Version 4.0, 2025) [30])

Österreich: Erstkontakt Hausarzt, Überweisung bei Bedarf (z. B. CXR)

In Austria, the family doctor is usually the first point of contact. They assess symptoms, carry out basic diagnostics and only issue a referral if a special examination or specialist treatment is necessary (e.g. for CXR). Patients with respiratory symptoms raising concern for lung cancer are typically referred to outpatient radiology providers, which may be hospital-based radiology departments or private radiology institutes, while hospitals are primarily reserved for acute emergencies and more complex cases. For the clinical question addressed in this assessment, the referral pathway and resulting case-mix are therefore more relevant than the physical location where the chest radiograph is performed.

4.2.2 Description and technical characteristics of the DHT[4]

Features of the intervention (DHT)[5]

Description and technical characteristics

KI-gestützte DHTs als radiologisches Assistenzsystem für CXR

This assessment covers the use of AI-enabled digital health technology (DHT) as an adjunct to a radiology specialist to help identify suspected lung cancer. Artificial intelligence (AI) technologies subject to this assessment are standalone software platforms developed with deep-learning algorithms to interpret CXRs. The algorithms are fixed but regularly updated. The AI software automatically interprets radiology images from the CXR to identify abnormalities or suspected abnormalities. The abnormalities detected and the methods of flagging the location and type of abnormalities differ between different AI technologies.

unterschiedliche Detektions- und Klassifikationsmethoden

For example, a CXR may be flagged as suspected lung cancer when a lung nodule, lung mass or hilar enlargement, or a combination of these, is identified. Depending on the technology employed, CXRs may be classified dichotomously into nodule-positive and nodule-negative cases, or alternatively, the system may detect and differentiate multiple abnormalities or pulmonary pathologies.

Table 4-1: Features of the intervention (DHT)

Feature	General characteristics of the DHTs with some examples
Input data	Chest radiographs (minimum one frontal AP/PA view) in Digital Imaging and Communications in Medicine (DICOM) format. Images are typically transmitted from hospital Picture Archive and Communications System (PACS) and associated metadata from the Radiology Information System (RIS) using the DICOM protocol.
Output data	Results are usually provided as secondary capture DICOM objects and/or DICOM structured reports, accompanied by a graphical user interface displaying the image and detected findings (e.g., AI-Rad Companion [36]).

[4] B0001, B0002, B0003, B0004, B0007, B0010, B0012, B0013, A0002.

[5] Considerations for DHTs from the ASSESS DHT manual: key components and how they interact (input and output data, use of algorithms and their type, function the DHT performs and its features, who interprets the health content), required connectivity (internet, mobile data), hardware requirements (operating system and platform, compatible devices) and user experience with the DHT (language, alert options etc.). Additional considerations for DHTs with AI component: static or adaptive, role of the AI within the DHT, role of the human, tasks automated by AI, type of model and learning used to develop the DHT, ability for retraining, planned updates and retraining, on-market retraining (continuous or periodic).

Feature	General characteristics of the DHTs with some examples
Type of AI and training dataset	Based on deep convolutional neural networks (CNNs) trained on large datasets from various digital radiology systems from different geographical regions.
Function/intended purpose	■ *Decision-support* (adjunctive use): AI does not make diagnosis or rank cases but purely assists image analysis and interpretation, for example by highlighting abnormalities or acting as a "second viewer" after the radiologist's initial review or ■ *Triage/prioritisation:* AI automatically analyses and triages images to support case prioritisation and workflow, acting as a "first viewer" before the radiologist, who then validates the AI output. In both functions, the software is not intended to provide a standalone diagnosis.
Deployment/technical components	Deployment models: ■ *On-premise* solution (all data processed locally) or ■ *Cloud-based/hybrid systems* (e.g., hosted on Microsoft Azure). Technical components: ■ Connection via hospital PACS/RIS or ■ Integration into digital health platforms (e.g., Siemens teamplay, Lunit AI Engine).
Verification and validation procedures	Validation procedures vary: most report *non-clinical testing* (unit, integration, system-level validation) (e.g., AI-Rad Companion [36] and *clinical performance testing* (e.g., ROC AUC, sensitivity, specificity) (e.g., Lunit Insight CXR [37], Gleamer ChestView [38]. Some conduct *reader studies* comparing AI-assisted and unaided radiologists (e.g., Gleamer ChestView [38]).
Tasks automated by AI	■ If any findings are suspected (abnormalities), the image is *flagged*, and a *passive notification* is provided to the user (e.g., ClearRead Xray Detect [39], Auto Lung Nodule Detection [40]). ■ Automated *image analysis and triage* (not diagnosis); automated *case prioritisation* for workflow (e.g., Lunit Insight CXR [37]).
Who interprets output	A human radiologist interprets and validates AI output.
Type of model	Static ("locked") models (= same input gives the same results every time) assumed[6]. The images are typically sourced from data providers based on collaboration agreements (e.g., Lunit Insight CXR [37]).
Ability for retraining	Retraining requires regulatory re-submission (new CE marking). No public reporting on re-training for any of the tools.
Planned updates/ retraining frequency	No public schedules disclosed.
On-market retraining	No information available; none of the identified products are known to perform autonomous continuous learning.
Post-market performance monitoring	Subject to general post-market surveillance requirements under EU MDR, though details of implementation are not reported.
Alert options	Generally *passive notifications* are provided (e.g., visual flags directly on the image or prioritisation cues) when abnormalities are detected (e.g., VUNO [43], ClearRead Xray [39], Lunit Insight CXR [37]).
Language support	Multiple language interfaces are typically available (e.g., Annalise.ai).

Abbreviations: AI … Artificial Intelligence, AP … Anteposteroanterior, CNN … convolutional neural networks, CE … Conformité Européenne, CXR … chest X-ray, DHT … Digital Health Technology, DICOM … Digital Imaging and Communications in Medicine, MDR … Medical Device Regulation, PA … Posteroanterior, PACS … Picture Archiving and Communication System, RIS … Radiology Information System, ROC AUC … Receiver Operating Characteristic – Area Under the Curve

[6] At present, we found no definitive source clearly stating whether commercially available CXR-AI systems are implemented as static or adaptive models; available literature mainly indicates that they are trained on fixed datasets, with little publicly reported information on post-deployment re-training, calibration or continuous learning, and several authors highlight this lack of transparency around model updating and lifecycle management as an important gap in the current evidence base [41, 42].

Current use of the DHT[7]

KI-Systeme: Detektion, Diagnose und Triage

In the CXR evaluation, the assessed DHT is used to support the clinician in the review of the image and help inform the need for further examinations such as a CT scan. The DHT means an AI software, which includes computer-aided detection (CADe), computer-aided diagnosis (CADx) and computer-assisted triage (CAST). CADe and CADx are used to diagnose cancer or to detect abnormalities on a CXR. CAST is used to prioritise and triage CXRs for review by a healthcare professional.

DHT umfasst CADe, CADx, CAST

Regulatory aspects

14 KI-Unternehmen für CXR identifiziert; 3× Risikoklasse 2b, 11× Risikoklasse 2a

Fourteen companies producing AI software for analysing CXR images were identified. Most products were CADe with some CADx and CAST products as well[8]. All listed products are CE-marked: three had class 2b category, and 11 had 2a risk class[9].

Table 4-2: AI software for analysing CXRs

Name and type of the software	HTD	Regulatory status
AI-Rad Companion CXR, CADx	Siemens Healthineers	Class 2a
Annalise CXR, CADe/CAST	annalise.ai	Class 2b
Auto Lung Nodule Detection, CADe	Samsung	Class 2a
ChestLink Radiology Automation, CADe/CAST	Oxipit	Class 2b
ChestView, CADe	GLEAMER	Class 2a
Chest X-ray, CADe	Rayscape	Class 2a
ClearRead Xray, CADe	Riverain Technologies	Class 2a
InferRead DR Chest, CADe	Infervision	Class 2a
Lunit INSIGHT CXR, CADe	Lunit	Class 2a
Milvue Suite, CADe/CAST	Milvue	Class 2a
qXR, CADe	Qure.ai	Class 2a
red dot, CADe/CADx	behold.ai	Class 2a
SenseCare-Chest DR Pro, CADe	SenseTime	Class 2b
VUNO Med-CXR, CADe	VUNO	Class 2a

Abbreviations: CADe … computer-aided detection, CADx … computer-aided diagnosis, CAST … computer-aided traigeCXR … chest X-ray, HTD … health technology developer

[7] A0001

[8] Definitions to be found in the Appendix Glossary.

[9] Under Regulation (EU) 2017/745 (MDR), **Class 2a** devices are generally considered **low to medium risk**. Requirements are the EU declaration of conformity, the technical documentation and a conformity assessment procedure carried out by a EU notified body. **Class 2b** devices are generally **medium to high risk**. Similar compliance route to class 2a devices, with an additional requirement on the assessment of the technical documentation. According to Rule 11, referring to "software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes", it is classified as **Class 2a**, unless such decisions have an impact that may cause serious deterioration of a person's health or require surgical intervention (in this case it is **Class 2b**) or cause death or irreversible deterioration (in this case it is Class 3).

ASSESS-DHT Taxonomy

Applying the ASSESS-DHT combination of criteria (purpose, risk and AI component), the DHT is positioned as having the medical purpose *diagnosis*. Respecting directness and patient vulnerability, applying the taxonomy risk matrix, the risk category is *serious* for that AI software, which have medical device regulation (MDR) class 2a category and *critical* for those having MDR class 2b (see Figure 4-2). Class 2a are decision-support or triage tools (adjunctive use), while Class 2b are AI systems influencing diagnostic decisions via partially automated and autonomous processes, e.g. autonomous preliminary reporting. The following ASSESS-DHT taxonomy applies (Figure 4-3).

DHT medizinischer Zweck: Diagnose

Risikokategorie: 2a: ernst, 2b: kritisch

Bewertung via ASSESS-DHT Kriterien

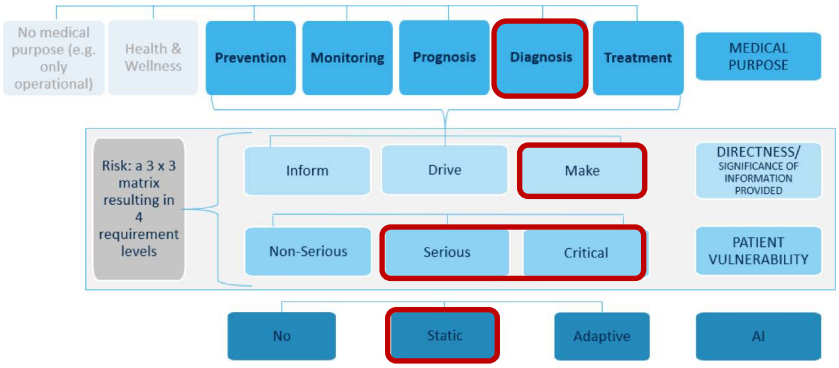


Figure 4-2: Categorisation of the AI software by the ASSESS-DHT taxonomy

Vulnerability of patients (state of healthcare situation/ condition)	'Directness' Intention of using DHT-collected or -produced information for health care decisions (by health care provider or patient)		
	Make (treat or diagnose)	Drive (clinical) management	Inform (clinical) management
Critical	IV Provides information to treat or diagnose a disease or conditions in a critical situation or condition → very high impact on patient or public health	III Provides information to drive clinical management of a disease or conditions in a critical situation or condition → high impact on patient or public health	II Provides information to inform clinical management of a disease or conditions in a critical situation or condition → medium impact on patient or public health
Serious	III Provides information to treat or diagnose a disease or conditions in a serious situation or condition → high impact on patient or public health	II Provides information to drive clinical management of a disease or conditions in a serious situation or condition → medium impact on patient or public health	I Provides information to inform clinical management for a disease or conditions in a serious situation or condition → low impact on patient or public health
Non-serious	II Provides information to treat or diagnose a disease or conditions in a non-serious situation or condition → medium impact on patient or public health t	I Provides information to drive clinical management of a disease or conditions in a non-serious situation or condition → low impact on patient or public health	I Provides information to inform clinical management for a disease or condition in a non-serious situation or condition → low impact on patient or public health

Figure 4-3: ASSESS-DHT taxonomy risk matrix

With respect to lifecycle of the technology and the ASSESS-DHT flowchart classification algorithm, the technology under assessment is in their *early clinical* study stage.

Features of the comparator

Description and technical characteristics

The comparator for this assessment is CXR images reviewed by a radiology specialist (e.g. radiologist or radiographer) without AI assistance. These are considered the reference standard.

4.2.3 Outcomes

The following outcomes were considered in this report:

diagnostische, klinische, technische und organisatorische Endpunkte definiert

Diagnostic accuracy:

- Both in nodule and in cancer detection, measured by sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratios, area under the ROC curve (AUROC), and the number of lung cancers diagnosed.[10]

Technical performance:

- Technical failure measured by failure rate due to inconclusive, indeterminate and excluded samples, and failures for any other reason.
- Concordance in lung nodule detection between radiology specialist with and without adjunct AI software.

Clinical:

- Mortality,
- Morbidity,
- Health-related quality of life (HrQoL),
- Safety or harm outcomes
  (e.g., consequences from false-positives/negatives).

Organisational implications:

- Turnaround time (image review to radiology report),
- Timeframe for follow-up CT scans, or receiving a diagnosis,
- Impact on clinical decision-making,
- Impact on use of resources
  (e.g., staff training, integration into existing systems),
- Impact of false positives on the workflow,
- Acceptability of AI software to clinicians (e.g., user-friendliness),

Costs (types of resources).

diagnostische Genauigkeit: Sensitivität/Spezifität dominieren, NPV/PPV seltener

Diagnostic accuracy parameters are those commonly used in diagnostic accuracy studies [44]. The included studies mostly reported sensitivity and specificity, and to a lesser extent NPV, PPV and AUROC. Frequently, instead of NPV and PPV, the true positive (TP), false positive (FP), true negative (TN),

[10] Definitions of terms related to diagnostic test accuracy are provided in the Appendix.

false negative (FN) values were reported, which allow calculation of these test accuracy metrics. False discovery rate (FDR) and false omission rate (FOR) were also reported in one study.

technische Performance: unklare Definition, keine Studiendaten vorhanden

Regarding technical performance, the interpretation of technical failure rate in the context of AI software remains unclear. In this review it is defined as instances where the software is unable to analyse an image [45]. None of the included studies reported this outcome.

Konkordanz als Vertrauensindikator für KI-Performance

Concordance is defined as the extent to which AI and non-AI technologies produce identical or comparable results. This parameter is considered particularly relevant for establishing confidence and trust in the performance of AI software.

keine klinischen oder Safety-Outcomes verfügbar

Regarding clinical outcomes, (all-cause) mortality, morbidity and HRQoL were not reported in the literature. Neither were any safety-related outcomes identified. One of the included reviews [45] consulted experts about outcomes of interest; although they found these outcomes relevant, they noted that it would be highly unlikely to find them reported in the current literature.

TAT inkonsistent definiert; Review folgt Definition der Originalstudien

In terms of organisational outcomes, report turnaround time (TAT) [46] is inconsistently defined in the literature and may refer to different intervals in the diagnostic workflow (e.g., from image acquisition, image review, or study availability to report completion or validation). For this review, we will use the definitions applied by the respective study authors.

KI in CXR kann klinische Entscheidungen beeinflussen; Potenziale: Früherkennung vs. Risiken (Überdiagnostik); Patientenangst, höhere Kosten & Ressourcennutzung

The integration of AI into CXR interpretation may influence clinical decision-making and organisational workflows in several ways. Potential benefits include earlier identification of suspicious findings, which may prompt timelier confirmatory testing and facilitate prioritisation of cases requiring urgent follow-up. This, in turn, could enable earlier discharge of patients without significant findings and help free staff time and radiology reporting capacity. At the same time, potential harms and unintended consequences must be considered. Increased AI-supported detection of benign pulmonary nodules may lead to a higher number of follow-up CT scans, some of which may not be clinically necessary. This may result in additional radiation exposure, greater patient anxiety following positive or indeterminate CXR results, and increased healthcare costs and resource utilisation [45].

Kliniker:innen-Akzeptanz und technische Integration zentral

Ease of use and acceptability by clinicians and healthcare personnel [45], as well as seamless integration into existing systems of the technology are key considerations for successful adoption in practice.

Einschluss aktueller Daten zu Preisen & Kostenwirksamkeit

For our review, we considered the types of cost categories associated with the implementation and use of AI-supported chest X-ray analysis, including recently published data related to prices, cost-effectiveness and budget impact analysis. Cost-effectiveness and budget impact analyses from one country are not transferable to others; however, certain elements – such as the main cost drivers – are relevant across healthcare settings.

4.2.4 Included studies

Basis: 3 HTA-Berichte mit variierenden Komparatoren

Three HTAs were included: (1) the Cedar Health Technology Research Centre review (2023) [45], evaluating AI alone versus clinician alone or clinician + AI; (2) the National Institute for Health and Care Research (NIHR) review (2024) [47], focusing specifically on adjunct use (clinician + AI versus clinician alone), and (3) the Scottish Health Technology Group (SHTG) assessment (2025) [48], evaluating both AI alone and clinician + AI versus clinician alone.

NICE EVA als konzeptuelle Basis, unterschiedliche Scope-Anpassungen

Cedar erweitert, NIHR behält ursprünglichen NICE-Scope bei

SHTG: Update von NIHR & Cedar, inkl. lokale NHS-Daten

All three HTAs build on the NICE Early Value Assessment (EVA) [49] on AI-derived CXR software, which is cited for context but excluded as it is not a full HTA. It addressed adjunct AI software for analysing CXR for suspected lung cancer and developed a conceptual cost-effectiveness model to inform discussion of what would be required to develop a fully executable cost-effectiveness model for future economic evaluation. The population comprised primary care populations referred for CXR due to symptoms suggestive of lung cancer or reasons unrelated to lung cancer. Comparative studies were eligible (radiology specialists assessing CXR with adjunct AI software versus radiology specialists alone), assessing outcomes related to test accuracy, practical implications of using AI software and patient-related outcomes. 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 were not fully captured. This resulted in an adjusted scope for the Cedar assessment [45], while NIHR [47] continued with the original scope as defined in the NICE EVA [49]. SHTG [48] updated both the NIHR and Cedar reviews and additionally incorporated evidence from a local service evaluation in the National Health Service (NHS) Grampian region, which had not been published in peer-reviewed form.

ROBIS: niedriges Bias-Risiko, Einschränkungen bei Eligibility/Extraction

The three HTAs [45, 47, 48] were judged to be at low risk of bias across most ROBIS domains, with two exceptions: (1) “Study eligibility criteria” in the SHTG assessment, where risk was unclear due to no public protocol/registration and no information on protocol deviations; and (2) “Data collection & study appraisal”, where risk was unclear for all there HTAs because procedures for duplicate data extraction, and independent risk-of-bias assessment were not described, and in some instances a single-reviewer process was reported. Detailed ROBIS assessments for each HTA are provided in the Appendix Table A-4.

Überblick über die eingeschlossenen HTAs

An overview of the characteristics of the included HTAs is provided in Table 4-3.

Table 4-3: Overview of included health technology assessments

Author (Institution), year	**Colquitt et al. (NIHR), 2024 [47]**	**Beddard et al. (Cedar), 2023 [45]**	**Moss et al. (SHTG), 2025 [48]**
Title	Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer: an evidence synthesis early value assessment	Artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer in primary care referrals: early value assessment addendum	Artificial intelligence supported clinician review of chest x-rays from patients with suspected lung cancer
Country	UK	UK	Scotland (UK)
Population	**Target condition:** lung cancer. 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 and socio-economic status.		
Intervention	**Clinician + AI:** CXR interpreted by radiology specialist (e.g. radiologist or radiographer) in conjunction with AI software.	**AI alone:** CXR interpreted by AI software.	**Clinician + AI** OR **AI alone**
Comparator	**Clinician alone**	**Clinician alone** OR **Clinician + AI**	**Clinician alone**
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, where biopsy is not applicable. For accuracy of *nodule* detection: Radiology specialist (single reader or consensus of more than one reader).		
Outcomes	■ Test accuracy for the detection of lung cancer (sensitivity, specificity, PPV, NPV, true positives, false positives, true negatives, false negatives, number diagnosed lung cancers), ■ Test accuracy for the detection of suspicious lung nodules, ■ Test failures (rates, and data on inconclusive, indeterminate and excluded samples, failure for any other reason), ■ Characteristics of discordant cancer cases, ■ Concordance in lung nodule detection between radiology specialist with and without adjunct AI software.		
	NA	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.	
Eligible study designs	Comparative studies	Comparative studies	NR
Number and type of included studies	No eligible studies, but 6 ineligible retrospective diagnostic accuracy studies were analysed	5 retrospective and prospective cohort/validation studies	2 diagnostic accuracy and feasibility studies (retrospective and prospective)
Economic analysis	Conceptual cost-effectiveness model developed – no empirical data used; budget impact analysis framework outlined, but no full economic evaluation conducted.	Not included.	Resource impact analysis comparing the *traditional diagnostic pathway* with the *AI-enabled pathway*

Abbreviations: AI … artificial intelligence, CT … Computed Tomography, CXR … chest X-ray, HTA … Health Technology Assessment, NA … not applicable, NIHR … National Institute for Health and Care Research, NICE … National Institute for Health and Care Excellence, NPV … negative predictive value, NR … not reported, PPV … positive predictive value, SHTG … Scottish Health Technologies Group, UK … United Kingdom

In total, the three HTAs included and analysed 13 unique primary studies [48, 50-61]. The update-search for primary studies identified two additional primary studies [62, 63], resulting in 15 primary studies overall. All studies were retrospective; in two of them [48, 61], a prospective component was also conducted. Most studies were single-centre (11/15) [50, 52-56, 58-62], two were conducted in two or more centres [51, 63], and for two studies [48, 57] the number and location of centres were not reported. Six studies were conducted in the UK [48, 50, 54, 58-60], three in the Republic of Korea [52, 53, 55], one in Germany [56], one in Germany and the US [51], one in Russia [61], one in the Netherlands [63], one in South Africa [62], and for one study the country was not reported [57]. One study was a company submission by Siemens [57], in which key details (country, centres, study design and number of patients/CXR) were removed as confidential.

15 Primärstudien: überwiegend retrospektiv, single-center, aus 7 Ländern – UK-dominant

Three studies focused solely on lung cancer detection [48, 50, 62], 11 studies analysed suspicious lung nodules [51-53, 55-61, 63], and one study addressed both [54]. The number of CXRs analysed ranged from 100 to over 5,700, and the number of patients from 100 to almost 5,600. Three studies compared standalone AI with radiologist alone [48, 54, 63], 11 studies compared radiologist + AI with radiologist alone [50-53, 55, 56, 58-62], and the Siemens company submission study compared a prototype AI + radiologist to a non-prototype AI + radiologist [57]. Five studies assessed Lunit [52, 53, 55, 60, 61], three Red Dot [50, 58, 59], three AI-Rad Companion [51, 56, 57], one the Auto Lung Nodule Detection software from Samsung [54], one Annalise [48], one qXR [62] and one study assessed multiple software [63].

3 Studien: reine Lungenkrebsdetektion

11 Studien: verdächtige Lungenherde

1 Studie: beide Fragestellungen

CXR-Anzahl: 100-5.700, Patient:innen: 100-5.600

An overview of all primary studies included in the HTAs and the update search is presented in Table 4-4.

Table 4-4: Overview of included primary studies from reviews and the update search

First author and year	Country	Study design	Population	Index test	Comparator	Reference standard
Lung cancer						
Dissez 2022 [50]	UK	Retrospective cohort study, one centre	400 CXRs from 400 adults	Red Dot (Behold.ai) + radiologists	Radiologists, radiographers	Blind reads of CXRs by two consultant radiologists.
Nxumalo 2024 [62]	South Africa	Retrospective cohort study, one centre	127 CXRs Number of patients: NR	qXR (Qure.ai) + radiologist	Radiologists	Histologically confirmed lung cancer diagnosis.
NHS Grampian service evaluation 2025 [48]	UK	Cohort study (retrospective + prospective phases), no information on the centres	Total CXR number: NR Prospective cohort: 68 lung cancer patients who reached treatment stage Comparator retrospective cohort: 113 patients.	Annalise Enterprise CXR AI module	Clinical review alone (retrospective 2019 baseline). For technical metrics, also compared against clinician-confirmed diagnosis (reference standard).	Clinician-confirmed lung cancer diagnosis.
Lung cancer + lung nodule						
Maiter 2023 [54]	UK	Retrospective cohort study, one centre	5722 CXRs from 5592 adults	Auto Lung Nodule Detection software (Samsung Electronics Version V.1.0)	Radiologists	Radiologists for nodule detection, multidisciplinary team consensus for lung cancer diagnosis
Lung nodule						
Nam 2020 [55]	Republic of Korea	Retrospective cohort study, one centre	218 CXRs from 218 adults	Lunit INSIGHT version 1.0.1.1 + radiologists	Radiologists	CT scan.
Jang 2020 [52]	Republic of Korea	Retrospective cohort study, one centre	351 CXRs from 351 adults	Lunit INSIGHT version 1.2.0.0 + radiologists	Radiologists	CXR and CT images.
Koo 2021 [53]	Republic of Korea	Retrospective cohort study, one centre	434 CXRs from 378 adults	Lunit INSIGHT version 1.00 + radiologist	Radiologists	Consensus from two thoracic radiologists using CXR or CT.
Homayounieh 2021 [51]	Germany; USA	Retrospective cohort study, two centres	100 CXRs from 100 adults	AI-Rad Companion (Siemens Healthineers) + radiologist	Radiologists	Consensus from two thoracic radiologists using all available clinical data.
Siemens 2022 [57]	Confidential information has been removed.			Prototype AI-Rad Companion + radiologist	CXR algorithm (Siemens Healthineers) + radiologist	Consensus from two thoracic radiologists using CXR or CT.
Niehoff 2023 [56]	Germany	Retrospective cohort study, one centre	499 CXRs from 499 adults	AI-Rad Companion (Siemens Healthineers Version VA23A) + radiologist	Radiologists	Consensus by two radiologists using additional radiographs, previous and/or follow-up CXR or CT scans.
Smith 2023 [58]	UK	Retrospective cohort study, one centre	4654 CXRs from 4076 adults	Red Dot (Behold.ai, V2.2) + radiologist	Radiologists	NR

First author and year	Country	Study design	Population	Index test	Comparator	Reference standard
Tam 2021 [59]	UK	Retrospective cohort study, one centre	400 CXRs from NR patients	Red dot (Behold.ai, Version NR) + radiologist	Clinician review (consultant radiologists)	Combination of the cancer registry database records, the electronic clinical record, and review of both subsequent and prior imaging.
Van Beek 2023 [60]	UK	Retrospective cohort study, one centre	1960 CXRs from NR patients	Lunit INSIGHT (Lunit Version 3.1.2.0) + radiologist	Radiologists	Consensus by two radiologists.
Vasilev 2023 [61]	Russia	Combined multicentre retrospective case-control study and prospective validation study, one centre	4825 CXRs from 4825 patients	Lunit INSIGHT (Lunit Version 3.110) + radiologist	Radiologists	A subset of radiographs (378/4,752) were interpreted by three experts.
Van Leeuwen 2024 [63]	The Netherlands	Retrospective cohort study, seven centres	561 CXRs from 386 patients	Annalise Enterprise CXR (annalise.ai), InferRead DR Chest (Infervision), INSIGHT CXR (Lunit), Milvue Suite–SmartUrgencesqXR (Milvue), ChestEye (Oxipit), AI-Rad Companion Chest X-ray (Siemens Healthineers), Med-Chest X-ray (VUNO)	Radiologist	Expert read of CT scan

Abbreviations: AI … Artificial Intelligence, CT … computer tomography, CXR … chest X-ray, NR … Not Reported, UK … United Kingdom, USA … United States of America.

4.2.5 Clinical effectiveness and safety[11]

keine klinischen Effektivitäts- oder Safety-Outcomes berichtet

None of the included studies assessed clinical effectiveness outcomes, such as health-related quality of life, mortality or morbidity.

Neither the included reviews, nor the update search identified any studies which reported safety outcomes.

4.2.6 Diagnostic accuracy[12] and technical performance

Sensitivität/Spezifität am häufigsten; Unterschiedliche Metriken erschweren den direkten Vergleich

The included studies analysed diagnostic accuracy of the AI software using diverse metrics. Sensitivity and specificity were reported in 13 studies (counting Maiter et al. once) [48, 50-56, 59-63]; predictive values (PPV, NPV, or both) in six [48, 50, 54, 56, 58, 62], TN/TP/FN/FP in seven [50-55, 59], accuracy in five [50, 51, 54, 59, 60], AUROC in seven [52, 53, 55, 56, 60, 61, 63], FDR explicitly in one [56] (FP available in seven, allowing derivation), FOR explicitly in one [56] (TN available in seven, allowing derivation), and concordance in six [50, 56, 58, 59, 61, 63].

Lung cancer

2 Vergleichstypen: KI-assistiert vs. Radiolog:in allein, Stand-alone KI vs. Radiolog:in

Across the four studies evaluating explicitly the performance of AI for *lung cancer* detection, two comparison types emerged: ***AI-assisted radiologist reading versus radiologist alone*** [48, 50] and ***stand-alone AI versus radiologist or confirmed diagnosis*** [54, 62].

KI-Assistenz erhöht Sensitivität, Spezifität unverändert, viele Falschpositive, gutes Rule-out

fehlender Radiolog:innen-allein-Vergleich limitiert Interpretation

When ***AI assisted the radiologist***, one study showed that sensitivity increased (77% vs 66%); while no difference was observed for specificity. Overall accuracy was similar (~75% in both arms), and there was no difference in PPV [50]. Another study from the real-world practice (NHS Grampian evaluation) showed that sensitivity of AI-assisted radiologist reading was 78% when compared with clinical review (urgent-CT decision) and 82% when compared with clinician-confirmed diagnosis; specificity was 91% in both comparisons. The AI system showed very low positive predictive values (PPV 1-3%), indicating that most positive flags were false positives, while the negative predictive value was consistently high (NPV ~100%), suggesting strong rule-out performance. However, because the study did not report radiologist-alone accuracy metrics, it is not possible to determine whether AI improved or diminished diagnostic performance [48].

stand-alone KI: variable Leistung, kontextabhängig

Trade-off: entweder viele Falschpositive (↓ Spezifität/PPV) oder schwaches Rule-out (↓ NPV) …

For ***stand-alone AI***, performance was more variable and context dependent. In a real-world cohort [54], AI showed no significant differences in sensitivity compared to radiologists (61% vs 66%) but substantially lower specificity (83% vs 98%). The AI produced substantially more false positives, resulting in a much lower PPV (6% vs 36%). In practical terms, when the AI flagged a CXR as "positive", it was correct only in a minority of cases, whereas radiologists' positive findings were correct more than one-third of the time. By contrast, NPV was very high and similar for both (≈99%), meaning that a negative result from either AI or radiologists almost always corresponded to the

[11] C0006, C0008

[12] D0001, D0005, D0032, D0011, D0012, D0013, D1001, D1004, D1005, D1006, D1008, D0020, D0021, D0022. Additionally, from the ASSESS-DHT manual: Is the system's performance *consistent across subgroups* (age, sex, ethnicity)? Was the *external validity* assessed in multiple clinical settings or populations?

absence of disease. Overall accuracy was also clearly lower for AI than for radiologists (83% vs 98%), indicating that, in this setting, the stand-alone AI system generated more incorrect classifications and would likely increase unnecessary follow-up investigations without improving cancer detection. In contrast, another real-world cohort [62] reported high PPV (97%), good sensitivity (84%) and specificity (91%), but a modest NPV (62%). In practice, this means that when the AI flagged a CXR as "positive", it was correct in almost all cases (very few false positives), but when it classified a CXR as "negative", it was wrong in nearly four out of ten cases, which is likely explained by the underlying study population (higher prevalence of disease in a selectively enriched study population of this specific study). Thus, while the system was highly reliable in confirming disease when it raised an alert, it was much less reliable in ruling out disease, limiting its usefulness as a stand-alone triage tool in that context. AUROC was not reported in any of these studies. Concordance – the degree to which two readers agree on whether a finding is present – was reported in one study, favouring AI assistance (57% vs 42%) [50]. This means that radiologists supported by AI agreed more often with the reference standard than when reading alone.

… als alleinige Triage nur eingeschränkt belastbar;

Assistenz kann Übereinstimmung verbessern

Lung nodules

Eleven unique studies evaluating the detection of suspicious *lung nodule* detection covered three comparison types: six studies evaluated the performance of ***stand-alone AI systems*** [54, 56, 58, 60, 61, 63], three studies compared ***stand-alone AI versus radiologist alone*** [59, 61, 63], and five studies examined ***AI-assisted radiologist reading versus radiologist alone*** [51-53, 55, 59]. Counts by comparison are not mutually exclusive because some studies reported more than one comparison.

stand-alone-KI (6), KI vs. Radiolog:in (3), KI-assistiert (5)

Across the six studies [54, 56, 58, 60, 61, 63] evaluating AI systems without a comparator, diagnostic performance was broadly similar. Reported accuracy ranged from 80% to 86%, with sensitivity between 55% and 94% and specificity between 79% and 98%. The study with the lower sensitivity data is a real-world study, which found a low PPV [54] of 6% and a high NPV of 99%. In practical terms, this means that when the AI flagged a CXR as "positive", it was correct only in a small minority of cases (most alerts were false positives), whereas a "negative" AI result was highly reliable in ruling out disease. Another study [61] analysing both a retrospective and a prospective study arm found similar results in the prospective cohort (sensitivity 84% and specificity 81%), whereas the retrospective cohort showed higher sensitivities and specificities (94% and 89%). Only one study [58] reported an AUROC, which was 0.77. This indicates a moderate ability of the AI system to discriminate between cases with and without the target abnormality across all possible decision thresholds (≥0.9 would be typically considered strong for diagnostic tools [64]).

stand-alone-KI: Accuracy 80-86 %, Sensitivität 55-94 %

moderate Diskriminierungsfähigkeit

Real-World: gutes Rule-out, viele Falschpositive

retrospektive Kohorten: höhere Performance; prospektive Daten: geringere Accuracy als retrospektive

In the three studies comparing ***stand-alone AI versus human readers*** [59, 61, 63], overall the AI systems performed similarly or slightly better than radiologists in terms of sensitivity, while specificity varied across algorithms. One study reported accuracy results [59], showing that the AI system achieved accuracy 87%, comparable to radiologists (84-90%). Another study [63] benchmarked seven commercial AI systems against radiologists, showing AI sensitivities ranging from 64% to 93%, compared to 81% (77-85%) for human readers. The best-performing systems (Lunit INSIGHT and Annalise.ai) achieved sensitivities above 90%, while specificity ranged more widely (50-89% for AI vs 71% for radiologists). The third study [61] also found no statistically sig-

Standalone-KI vs. Radiolog:in: vergleichbar oder besser

Sensitivität ähnlich, Spezifität variable

Leistungsstarke KI-Systeme teils ≥ Radiolog:innen, aber heterogen

nificant differences in sensitivity or specificity between AI and radiologists, both in a retrospective and in a prospective cohort.

All five studies [51-53, 55, 59] examining ***AI-assisted radiologist reading versus radiologist alone*** reported increased sensitivity and accuracy when AI assistance was used. In each study, sensitivity increased with AI, typically by about 5-13% (e.g. from the mid-40% to mid-50% range [51, 52, 55], or from the low-90% to mid-90% [53, 59]), while specificity showed modest gains (generally from around 78-93% without AI to 82-97% with AI) [51-53, 55, 59]. Two studies reporting accuracy found a minor increase with AI assistance (from roughly 70% to 75% with AI assistance [51] and from 84-90% to 90-91% [59].

KI-Assistenz steigert Sensitivität, geringe Effekte auf Spezifität

konsistente Genauigkeitsverbesserung

None of the studies – whether focused on cancer or nodule detection – evaluated whether system performance was consistent across patient subgroups (e.g., age, sex, ethnicity), nor did they assess external validity across different clinical settings or populations. In addition, none of the included studies reported technical failure rates.

keine Subgruppenanalysen oder externe Validierung; technische Ausfallraten nicht berichtet

Detailed performance results per study can be found in the Appendix, Table A-2.

4.2.7 Organisational outcomes[13]

None of the included studies directly assessed all aspects related to resource use, staff training, quality assurance, or management processes of AI-supported clinician review of CXRs. However, the available evidence provides insights into several organisational aspects relevant to current work processes, acceptance among clinical users, and potential implications for resource use and workflow efficiency

indirekte Evidenz zu Workflow, Akzeptanz und Ressourcennutzung

In one study [50], implementation of AI tool was associated with a simulated increase in the number of patients referred for CT (from 29% to 36%), though this did not result in a statistically significant change in the proportion of diagnostic CTs for lung cancer (from 39% to 38%, p=0.22). Participant feedback indicated overall positive user experience: 8 out of 10 clinicians indicated that reporting speed was not negatively affected by AI use, and 9 out of 10 found the AI-generated heatmaps useful for understanding the model's focus areas.

1 Studie: Anstieg bei CT-Überweisungen, diagnostische Ausbeute stabil, positive User Experience

Another study [58] reported stable service levels following AI implementation. The AI algorithm provided results within a mean of 7.1 seconds (range 5-17 seconds), and 99.3% of radiographs flagged as high-confidence-negative – meaning the algorithm assigned a high probability to the absence of abnormal findings- were audited by radiologists within 24 hours, with an average TAT of 3 hours and 50 minutes. Based on these findings, the authors indicated that AI integration did not disrupt, and may have enhanced, workflow efficiency and timeliness of reporting.

1 Studie: stabile Servicequalität, potenzielle Effizienzsteigerung

[13] E0001, D0023, G0001, G0003, G0012, G0006, G0008, G0010. Additionally, from the ASSESS-DHT manual: How does *integration into existing IT systems* affect clinical operations? How do *users* (clinicians, patients) interact with the technology? Are there *human factors or cognitive load* implications?

A third study [59] discussed that positioning AI as the first reader of CXR could improve overall diagnostic sensitivity and reduce radiologist workload through triaging of positive cases. However, this approach may also lead to an increased number of false positives requiring further review. The authors emphasised that continued clinician oversight and interaction with AI systems could mitigate such effects and foster knowledge exchange between clinicians and algorithms.

1 Studie: KI als "First Reader": Triage-Potenzial vs. falsch positiv

In addition, the SHTG assessment [48] reported findings from a large Health Foundation survey that explored attitudes towards AI use in healthcare among NHS staff (n=1,292) and the general public (n=7,201). The survey found broad support for AI use, particularly for administrative tasks (NHS staff 81%, public 61%) and to a lesser extent for clinical applications (staff 76%, public 54%). More than half of NHS staff (57%) expressed that they were looking forward to using AI in their role. However, both groups also expressed notable concerns: 17% of the public and 10% of NHS staff believed that AI could worsen the quality of care, while 53% of the public and 65% of NHS staff were concerned that AI might make staff feel more distant from patients or colleagues. Furthermore, 26-28% expressed worries about inaccuracy, and both groups emphasised a strong desire for transparency, stating that people should be informed when AI is being used.

SHTG: Umfrage zu KI-Nutzung im NHS & Öffentlichkeit

breite Akzeptanz, aber Bedenken zu Distanz und Genauigkeit

Public vs. Staff: unterschiedliche Akzeptanzniveaus

Detailed extraction can be found in the Appendix, Table A-3.

4.2.8 Cost implications

This section aims to outline the cost and resource implications of introducing adjunct AI for detecting lung cancer on CXR and to reflect on what would be required to estimate budget impact in Austria.

Kostenimplikationen und Budget-Impact-Überlegungen für Österreich

We identified two reviews with economic content: the NIHR [47] and the SHTG [48] assessments, both in NHS settings. The SHTG review presents quantitative resource-impact data from an NHS Grampian service evaluation using the *Annalise Enterprise CXR*. The NIHR report present a conceptual framework for cost-effectiveness but reports no completed economic evaluations directly comparing AI-supported versus standard CXR pathways. No Austrian primary cost-effectiveness or budget-impact analyses were identified, and we found no newer economic studies beyond those covered by the reviews.

2 UK-basierte Reviews, keine österreichischen Kostendaten

Although cost estimates were presented for the UK context, these values are not directly transferable to Austria due to differences in healthcare structures, salary levels, and procurement processes. Nonetheless, these reviews provide useful guidance in defining relevant cost categories and data needs.

UK-Kosten nicht übertragbar, aber Kostenkategorien relevant

Accordingly, the following cost categories and associated resource-use considerations are proposed for the Austrian context (drawing on NIHR/SHTG for structure, without UK unit prices):

5 Kostenkategorien: Software, Training, Personal, Diagnostik, Therapie

1. **AI software costs**: AI vendors apply either fixed annual subscription fees or volume-based pricing models. There is a one-time implementation fee covering software installation, integration with radiology IT systems, and staff training. Ongoing subscription costs typically cover licensing, maintenance, technical support, and updates. The NIHR report noted that some detailed pricing information was confidential.

2. **Training costs**: training is typically brief and is incorporated into the implementation fee. Some vendors delivered training via a customised "train-the-trainer" approach. Training durations for radiologists reported by NIHR ranged from approximately 30 minutes to one hour.
3. **Staff time costs:** particularly radiologist and reporting radiographer time, and potential pathway-related resource needs such as a dedicated lung pathway coordinator or extended/out-of-hours CXR reporting. In the UK these costs can be derived from published sources and reference rates.
4. **Additional diagnostic and downstream healthcare costs**: following CXR, further tests may be needed, including repeat CXR, CT, PET, bronchoscopy, or biopsy, with associated clinical input from GPs, radiologists, respiratory physicians, and multidisciplinary team (MDT) meetings. The UK analysis identified potential cost sources for these tests but found no evidence that AI use alters referral rates, follow-up imaging, or diagnostic pathways. Thus, any downstream resource impact remains uncertain.
5. **Cancer treatment costs**, assigned by cancer stage. These costs are estimates and in the NIHR report they were used conceptually, as there is no direct evidence that AI changes stage at diagnosis or treatment intensity.

keine empirischen Daten zu Ressourcennutzung durch KI

Using the above cost categories, the NIHR review found that unit cost sources are generally identifiable but that no empirical evidence exists on how AI-assisted CXR interpretation changes resource use; therefore, the *overall cost impact is indeterminate*, with the only certain addition being the cost of purchasing and implementing AI software and uncertain downstream effects.

initiale Mehrkosten dokumentiert, Gesamtbilanz unklar

The SHTG review, applying the same cost categories in the NHS Scotland setting, estimated a *small net short-term cost increase* for an AI-enabled CXR pathway versus usual care. The incremental costs were mainly driven by the AI software and operational add-ons (e.g., pathway coordination, extended/out-of-hours reporting, and additional CT list capacity). Downstream effects were likewise concluded to be uncertain due to limited evidence. Some pricing details were commercially confidential.

österreichische Tarife und Gehaltsdaten erforderlich

For an Austrian analyses, equivalent information would need to be derived from national hospital tariffs, reimbursement catalogues, and local wage data.

5 Discussion

Summary and interpretation of the evidence

CXR: erste Wahl bei thorakalen Symptomen

Arbeitsbelastung, Personalknappheit, Übersehensrisiko

KI-Software: Unterstützung der Radiolog:innen

Workflow-Frage: KI als Erst- vs. Zweitleser

Wahl der Workflow-Konfiguration beeinflusst Effizienz & Sicherheit

Lung cancer is one of the most common cancers in Austria. Chest X-ray (CXR) is frequently the modality of choice for primary chest evaluation, as it allows rapid clarification of thoracic symptoms with relatively low radiation exposure and lower costs compared to computer tomography (CT) [65]. However, CXR interpretation can be challenging, especially in the context of increasing radiology workload, radiological staff shortages, and the associated risk of delayed or missed diagnoses [66, 67]. A key goal is therefore to improve the detectability (or confident exclusion) of lung nodule on CXR, especially when nodules are subtle or hidden by over- or underlying anatomical structures. More accurate nodule assessment could help to better target indications for lung CT and avoid unnecessary radiation exposure. Commercial artificial intelligence (AI) software solutions aim to increase the diagnostic accuracy of CXR and supporting radiologists by improving workflow efficiency. A further question is how AI is positioned within the reading workflow. Most available studies evaluated AI as an adjunct or triage tool but did not clearly distinguish between AI as a first reader (pre-screening and flagging examinations) and AI as a second reader (used after an initial human read). The choice of workflow configuration is likely to influence both efficiency and safety, yet no robust comparative data are available to guide whether, and how, AI should be integrated into existing reporting pathways.

keine Studien zur klinischen Effektivität

HTAs: unzureichende Evidenz für Kosten-Effektivität

Real-World-Daten deskriptiv, kausale Effekte ungeklärt

This review synthesised evidence on the clinical effectiveness, organisational and cost and resource implications of AI-assisted radiologist interpretation of CXRs in patients with suspected lung cancer, summarising and updating reviews conducted by health technology assessment (HTA) organisations in the UK (NIHR, Cedar, and SHTG) [45, 47, 48]. Across all sources, no published studies were identified to demonstrate clinical effectiveness (e.g., improving health related quality of life (HrQoL) or survival outcomes), and the included HTAs likewise did not find sufficient evidence to support any conclusions on the cost-effectiveness of AI-assisted radiologist review of CXRs in lung cancer. Real-world service evaluations, such as the National Health Service (NHS) Grampian analysis [48], enrich the evidence by providing descriptive data on the time until diagnosis and the time to treatment initiation, but do not establish casual impact and hence, clinical benefit cannot be inferred. Cedar review authors also highlighted concerns from clinical experts about the risk of false positives, potentially increasing CT demand and workflow burden. However, the high false discovery rates observed in some studies mainly reflected benign pre-existing findings that radiologists generally recognised as non-problematic. This highlights a current limitation of AI systems: they cannot incorporate patients' clinical history or prior imaging, creating a trade-off between reducing false positives and maintaining sensitivity.

fehlender klinischer Kontext limitiert KI-Spezifität

KI-Systeme primär auf Auffälligkeiten/ Lungenherde trainiert

KI primär zur Kennzeichnung auffälliger Befunde …

A consistent pattern across the available evidence is that most AI systems were evaluated for their ability to detect abnormalities or pulmonary nodules rather than for their contribution to the diagnosis of lung cancer at the patient level. This reflects the practice that CXR-AI models are typically trained on datasets with abundant image-level labels for abnormalities (e.g., nodules, opacities or masses), whereas confirmed cancer diagnoses are less frequent and more complex to annotate [68]. As a result, current systems primarily function as *flagging instruments* – whether used as a first reader (triage/prioritisation)

… indirekter Beitrag zur Früherkennung,

Downstream-Effekte und Falsch-Positiv-Schäden nicht untersucht

or a second reader (decision-support) – highlighting suspicious findings for radiologists rather than providing a definitive cancer diagnosis [65]. This distinction has important clinical implications: although detecting a pulmonary nodule is not equivalent to identifying lung cancer, such findings are key triggers for further diagnostic work-up. AI may therefore contribute indirectly to earlier detection of malignancies if nodule identification improves, but the downstream impact on patient-important outcomes and the potential harms of increased false-positive findings (e.g. unnecessary imaging, invasive procedures, anxiety, and costs) remain untested.

breitere Bildgebungs-Evidenz bestätigt Ergebnisse

Evidence from broader imaging research supports these observations. A recent systematic review and meta-analysis across multiple imaging modalities [5] showed that while many studies reported improvements in abnormality detection or workflow efficiency, pooled estimates did not demonstrate consistent reductions in reporting time, and most studies suffered from substantial heterogeneity and methodological weaknesses. Importantly, the review highlighted that improvements in detection or workflow have rarely been linked to downstream diagnostic accuracy, timeliness of diagnosis, or clinical outcomes – an observation consistent with the gaps identified in this assessment.

Onkologie-KI: vielversprechend in Studien, Real-World-Effektivität unklar

Additional contextual evidence from the wider AI-in-oncology literature aligns with the findings of this assessment. A recent systematic review and meta-analysis [69] showed that many AI models for lung cancer imaging – primarily CT-based – demonstrate promising diagnostic accuracy under experimental conditions but are rarely validated prospectively or across diverse clinical settings, limiting conclusions on real-world effectiveness. Earlier publications [70, 71] similarly underline that despite rapid advances in AI for lung cancer detection, most systems remain insufficiently evaluated in routine care, with challenges related to dataset representativeness, generalisability and integration into clinical workflows.

Limitations of the evidence

retrospektiv, Studien mit höherer Krankheitsrate als im Versorgungsalltag

The internal validity of the available studies is limited by several methodological shortcomings. Most studies relied on retrospective designs and often used enriched datasets with a higher prevalence of cancer or nodules than typically seen in primary care, which limits the applicability of test accuracy estimates. Reference standards varied considerably, ranging from expert radiologist consensus to clinical diagnosis or mixed imaging modalities, making direct comparisons difficult. Some studies excluded poor-quality images or small nodules, introducing selection and spectrum bias, and several allowed radiologists to view their initial unaided reading while conducting the AI-assisted reading, thereby introducing carryover and recall bias. Reporting was often incomplete, with insufficient information about patient selection, referral pathways or blinding procedures [45, 47, 48].

heterogene Referenzstandards

unklare Überweisungswege

Rundherddetektion ≠ Krebsdiagnose …

Generalisability of findings is also limited. Many of the studies summarised in the National Institute for Health and Care Research (NIHR) review [47] did not report referral routes or symptom status, making it unclear whether the populations resemble those typically seen in primary care – an important consideration for early lung cancer detection. Several studies originated from non-European settings with different diagnostic pathways, imaging protocols and healthcare structures, limiting transferability to Austrian practice. Most importantly, many studies evaluated nodule detection rather than confirmed cancer diagnosis, creating a mismatch between reported outcomes and the

clinical decision problem. Exclusion of clinically challenging images (such as poor-quality or lateral radiographs) further limits generalisability, as does the lack of true real-world implementation studies. Additional factors may further limit applicability to Austrian practice. First, test performance metrics such as PPV and NPV depend strongly on the pre-test probability of lung cancer, which varies across settings. Most included studies did not report prevalence or case-mix, making it unclear whether study samples were enriched or reflective of Austrian practice. Second, the prevalence of tuberculosis in training and evaluation datasets is a relevant factor which was not reported by the included studies but in some countries (e.g., South Africa, Russia) might be considerably different than in Austria, potentially influencing false-positive detections of benign nodules or scarring. Third, most studies did not report whether models were evaluated on bedside (portable) radiographs, which constitute a substantial share of CXR imaging in hospitals and differ from standard PA/AP acquisitions in quality and projection. The absence of subgroup analyses for these factors further restricts the transferability of results to Austrian clinical pathways.

… PPV/NPV prävalenzabhängig

fehlende Subgruppen

strukturelle und epidemiologische Unterschiede in Österreich

Economic extrapolation is also constrained because resource use, workflow patterns and cost structures differ substantially across healthcare systems. Collectively, these factors reduce the applicability of study findings to Austrian clinical pathways.

Ressourcennutzung und Kostenstrukturen nicht übertragbar

Equity concerns also arise in the development, validation and deployment of AI tools. Bias may be introduced if training datasets are not ethnically, demographically or clinically representative of the populations in which the tools will be used [2, 72]. This is particularly relevant for CXR imaging, where anatomical characteristics, comorbidities and disease prevalence may vary across populations. For most commercial AI systems, however, only limited information was available on the training datasets, and the fact that a study was conducted in a given country does not necessarily mean that the underlying model was trained on data from that setting [45, 48]. If Austrian patient populations differ from those used for model development, performance may vary. Yet none of the identified studies assessed subgroup performance (e.g. by age, sex, ethnicity, comorbidity or socioeconomic status), and no calibration studies for the Austrian population were identified. It underlines the need for local validation and fairness assessments to ensure that deployed algorithms perform reliably and equitably in the Austrian context. These equity concerns require careful consideration prior to large-scale deployment.

Equity-Bedenken

keine Subgruppenanalysen

keine Kalibrierung für Österreich

lokale Validierung & Fairness-Assessment nötig vor Einsatz

Evidence gaps

No studies assessed whether AI-assisted CXR interpretation leads to earlier diagnosis of lung cancer, stage shift, faster treatment initiation or improved patient outcomes. Likewise, there is no evidence on the long-term clinical or cost and resource implications of AI-assisted CXR interpretation. Studies focused on intermediate (e.g. nodule detection) and test performance outcomes, and none linked AI-flagged abnormalities to full diagnostic pathways or patient follow-up. Organisational implications remain uncertain; existing studies were too small or methodologically limited to infer workflow or system-level impact. Economic evidence is sparse, and there are insufficient data on clinical effectiveness, resource use and Austrian care pathways to populate a robust cost-effectiveness model. Furthermore, many commercially available AI products (e.g., ChestLink, Rayscape's Chest X-ray, ClearRead, InferRead, SenseCare-Chest) have no published evidence at all.

Evidenzlücken:

technische Daten vorhanden, patient:innenrelevante Endpunkte und Implementierung ungeklärt

Transparenzdefizit: Trainingsdaten, Architektur & Updates nicht dokumentiert

In addition, transparency regarding the underlying AI models remains limited. For most commercial systems, little publicly available information exists on the composition of training datasets, algorithmic architecture, model type (static or adaptive), or procedures for model updates, re-training and post-market performance monitoring. These details were generally absent both from the published studies included in this assessment and from publicly available user manuals or technical documentation.

4 laufende Studien zu KI-CXR bei Lungenkrebs/ Lungenherden

In a search for ongoing studies, we identified four trials investigating the real-world performance of AI tools for CXR interpretation in lung cancer and pulmonary nodule detection. In the UK, two major studies are underway. The AID-CXR study (NCT06075836) evaluates *Lunit INSIGHT CXR* using approximately 500 retrospectively collected CXR from emergency departments and inpatient settings across two hospital trusts. Healthcare professionals from various specialties interpret images with and without AI assistance, with primary outcomes measuring diagnostic accuracy (sensitivity, specificity, PPV and NPV) and secondary outcomes assessing reader speed, efficiency, and confidence. The estimated end date of the study was 06/2025, but the study is still marked as active. The RADICAL trial (NCT06044454) employs a stepped-wedge cluster-randomised design to assess *qXR* (Qure.ai) in Scotland, measuring reduction in reporting time for suspicious CXR alongside test performance, safety, and cost-effectiveness. The estimated end date of the study is 11/2025. A Czech study (NCT05594485), completed in 2022 but not yet reported results and not specific to lung cancer, retrospectively examined *Carebot AI CXR* performance on 127 CXR, comparing AI diagnostic accuracy against five radiologists across twelve predefined abnormalities. The *ACER trial* (NCT06456203), not yet recruiting, will compare AI-assisted versus standard CXR interpretation for lung cancer and pneumonia detection in a randomised clinical trial. The expected completion date of the study is 12/2025. Detailed information on these studies is provided in the Appendix, Table A-5.

Limitations of our review

HTAs mit teilweise unklarem Bias eingeschlossen

One limitation is that we included HTAs even when the risk-of-bias assessment was unclear in some domains. To mitigate potential bias in those domains, we ran an update search from January 2024 and applied independent duplicate screening, study selection and data extraction.

organisatorische Outcomes: nur durch Handsuche

Another limitation is the use of a "clinical trials" filter in the systematic search for primary studies, which reduced the large number of hits but may have led to the omission of relevant non-trial designs. This risk was partly mitigated by systematically checking ongoing studies listed in the included HTAs and following up these records. For organisational outcomes, we did not perform a separate systematic search, instead, we relied on the primary studies already identified and supplemented these with targeted hand searches. The same approach was used for the cost and resource needs analysis.

Excurse: checklist for hospital procurement decisions

AIHTA-Checklist anwendbar, 2 CXR-spezifische Ergänzungen nötig

The checklist developed by the Austrian Institute of Health Technology Assessment (AIHTA) and published in 2024 [2] provides a structured and comprehensive overview of key considerations relevant for hospital procurement of AI-enabled DHTs. We piloted the applicability of this checklist for AI-supported CXR tools. All major domains covered in the original table – regulatory requirements, privacy, technical characteristics, safety, ethical considerations, organisational implications and economic aspects – are relevant. However, for AI-supported CXR tools specifically, two additional considerations are particularly important: (1) dataset representativeness and local validation, and (2) model lifecycle transparency.

lokale Validierung nötig

Transparenz zu Updates, Re-Training & Post-Market-Surveillance nötig

Because many commercial AI systems are trained on datasets from other countries and populations, procurement decisions should explicitly require evidence of local validation or calibration to the Austrian population. In addition, hospitals should request documentation on model updates, re-training policies and post-market surveillance procedures, as these elements are rarely disclosed in published studies but are essential for safe long-term deployment. Overall, the checklist is suitable for supporting procurement decisions in Austrian hospitals, but it should be applied alongside local clinical expertise, IT governance processes and regulatory compliance checks. An extended checklist with the additional considerations is presented in the Appendix, Table A-5.

Checkliste ergänzend zu lokaler Expertise & Governance

6 Conclusion

unzureichende Evidenzbasis

Overall, the current evidence base is insufficient to demonstrate that AI-assisted chest x-ray (CXR) interpretation provides added value in lung cancer pathways. Existing studies largely focus on intermediate and test performance outcomes, such as nodule or abnormality detection, and do not show whether AI support leads to earlier diagnosis, stage shift, faster treatment initiation, improved survival, or better quality of life. While AI has potential to support radiologists and may improve reporting efficiency in some settings, concerns remain about accuracy in real-world settings, transparency, workflow impact and the inability of current systems to incorporate clinical history and prior imaging.

keine belegten Patient:innenvorteile

Potenzial, aber Unsicherheiten

Forschungsbedarf in Realwelt

Future research should therefore prioritise prospective, real-world comparative studies that follow patients across the full diagnostic pathway and report clinical, organisational and economic outcomes. This includes analysing and reporting on training datasets, ensuring that models are trained and calibrated on populations representative of local clinical practice, and assessing diagnostic accuracy in relevant subgroups. Further progress in assessing algorithmic bias, fairness and long-term outcomes depends on the availability of high-quality comparative studies, including randomised trials where feasible. Evidence is also needed on technical failure rates, equity, changes in clinical decision-making, and costs and resource use. Until such data become available, the role of AI-assisted CXR interpretation in lung cancer care in Austria remains uncertain.

lokale Validierung erforderlich

unklare Rolle in Österreich

7 References

[1] Topol E. J. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019;25(1):44-56. DOI: 10.1038/s41591-018-0300-7.

[2] Riegelnegg M G. D., Goetz G. Artificial Intelligence in Health Care with a Focus on Hospitals: Methodological Considerations for Health Technology Assessment. A scoping review. AIHTA Project Report No.: 142. 2024 [cited 15.07.2025]. Available from: https://eprints.aihta.at/1546/1/HTA-Projektbericht_Nr.164.pdf.

[3] Pesapane F., Codari M. and Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp. 2018;2(1):35. Epub 20181024. DOI: 10.1186/s41747-018-0061-6.

[4] Reiner B., Siegel E. and Carrino J. A. Workflow optimization: current trends and future directions. J Digit Imaging. 2002;15(3):141-152. Epub 20021217. DOI: 10.1007/s10278-002-0022-7.

[5] Wenderott K., Krups J., Zaruchas F. and Weigl M. Effects of artificial intelligence implementation on efficiency in medical imaging—a systematic literature review and meta-analysis. npj Digital Medicine. 2024;7(1):265. DOI: 10.1038/s41746-024-01248-9.

[6] van Leeuwen K. G., Hedderich D. M., Harvey H. and Schalekamp S. How AI should be used in radiology: assessing ambiguity and completeness of intended use statements of commercial AI products. Insights Imaging. 2024;15(1):51. Epub 20240216. DOI: 10.1186/s13244-024-01616-9.

[7] Hosny A., Parmar C., Quackenbush J., Schwartz L. H. and Aerts H. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500-510. DOI: 10.1038/s41568-018-0016-5.

[8] van Leeuwen K. G., de Rooij M., Schalekamp S., van Ginneken B. and Rutten M. How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr Radiol. 2022;52(11):2087-2093. Epub 20210612. DOI: 10.1007/s00247-021-05114-8.

[9] European Network For Health Technology Assessment (EUnetHTA). EUnetHTA Core Model® for the full assessment of Diagnostic Technologies, Medical and Surgical Interventions, Pharmaceuticals and Screening Technologies (Version 3.0) 2015.

[10] Schulz W. A. Molekularbiologie menschlicher Krebserkrankungen. In: Schulz W. A., editor. Lungenkrebs: Springer Nature Switzerland; 2024. p. 375-385.

[11] Lang-Stöberl A., Fabikan H., Hochmair M., Kirchbacher K., Rodriguez V. M., Ay L., et al. The Landsteiner lung cancer research platform (LALUCA). Wiener klinische Wochenschrift. 2025;137(7):197-204. DOI: 10.1007/s00508-024-02351-3.

[12] Statistik Austria. Krebserkrankungen in Österreich 2024. 2024 [cited 06.10.2025]. Available from: https://www.statistik.at/fileadmin/publications/Krebs-2024_Webversion-barrierefrei_Rev1.pdf.

[13] Statistik Austria. Krebserkrankungen. 2025 [cited 06.10.2025]. Available from: https://www.statistik.at/statistiken/bevoelkerung-und-soziales/gesundheit/krebserkrankungen.

[14] Statistik Austria. Krebserkrankungen in Österreich 2025. 2025 [cited 06.10.2025]. Available from: https://www.statistik.at/fileadmin/publications/Krebserkrankungen-in-Oesterreich-2025_barrierefrei.pdf.

[15] Statistik Austria. Überlebenschancen nach Krebsdiagnose steigen weiter. Pressemitteilung: 13 519-013/25. 2025 [cited 06.10.2025]. Available from: https://www.statistik.at/fileadmin/announcement/2025/01/20250123Krebsstatistik.pdf.

[16] Hackl M., Hanika A. and Klotz J. Prognose der Krebsprävalenz bis 2030. Wien: Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz,, 2016.

[17] Geiger-Gritsch S., Absenger G., Endel F., Flicker M., Hermann A., Kocher F., et al. Health Services Research in Oncology: Patients with Non-Small Cell Lung Cancer Treated with Anti-PD-1/PD-L1 Therapy in Real-World Practice. Wien: 2021 [cited 10.09.2025]. Available from: https://eprints.aihta.at/1227/.

[18] American Cancer Society. Lung Cancer Risk Factors. 2024 [cited 06.10.2025]. Available from: https://www.cancer.org/cancer/types/lung-cancer/causes-risks-prevention/risk-factors.

[19] Coté M. L., Liu M., Bonassi S., Neri M., Schwartz A. G., Christiani D. C., et al. Increased risk of lung cancer in individuals with a family history of the disease: a pooled analysis from the International Lung Cancer Consortium. Eur J Cancer. 2012;48(13):1957-1968. Epub 20120319. DOI: 10.1016/j.ejca.2012.01.038.

[20] Ruano-Raviña A., Provencio M., Calvo de Juan V., Carcereny E., Moran T., Rodriguez-Abreu D., et al. Lung cancer symptoms at diagnosis: results of a nationwide registry study. ESMO Open. 2020;5(6):e001021. DOI: 10.1136/esmoopen-2020-001021.

[21] Spyratos D., Zarogoulidis P., Porpodis K., Tsakiridis K., Machairiotis N., Katsikogiannis N., et al. Occupational exposure and lung cancer. J Thorac Dis. 2013;5 Suppl 4(Suppl 4):S440-445. DOI: 10.3978/j.issn.2072-1439.2013.07.09.

[22] Bredtoft E. N., Madsen H. H. and Rasmussen T. R. Stage I lung cancer patients with or without symptoms – are the patients different and should we treat them differently? Acta Oncologica. 2021;60(9):1169-1174. DOI: 10.1080/0284186X.2021.1931959.

[23] Krebsinformationsdienst des Deutsches Krebsforschungszentrums. Lungenkrebs: Symptome und Früherkennung. 2025 [cited 06.10.2025]. Available from: https://www.krebsinformationsdienst.de/lungenkrebs/symptome-und-frueherkennung.

[24] Perisano C., Spinelli M. S., Graci C., Scaramuzzo L., Marzetti E., Barone C., et al. Soft tissue metastases in lung cancer: a review of the literature. Eur Rev Med Pharmacol Sci. 2012;16(14):1908-1914.

[25] Detterbeck F. C. and Gibson C. J. Turning Gray: The Natural History of Lung Cancer Over Time. Journal of Thoracic Oncology. 2008;3(7):781-792. DOI: 10.1097/JTO.0b013e31817c9230.

[26] Basumallik N. and Agarwal M. Small Cell Lung Cancer. Treasure Island (FL): 2023 [cited 23.10.2025]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK482458.

[27] Postmus P. E., Kerr K. M., Oudkerk M., Senan S., Waller D. A., Vansteenkiste J., et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2017. DOI: 10.1093/annonc/mdx222.

[28] Malmberg C. and Hofmarcher T. Krebs-Dashboard für Österreich. 2024 [cited 06.10.2025]. Available from: https://ihe.se/app/uploads/2024/10/IHE-Cancer-Dashboard-Austria_.pdf.

[29] Hernandez D., Giri P., von Both A. and Schlander M. Krankheitslast von Lungenkrebs in Deutschland: Epidemiologie und Kosten. Lungenkarzinome. 2022;5.

[30] Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft D. K., AWMF), . Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms, Langversion 4.0. AWMF-Registernummer: 020-007OL 2025 [cited 10.10.2025]. Available from: https://register.awmf.org/assets/guidelines/020-007OLl_S3_Praevention-Diagnostik-Therapie-Nachsorge-Lungenkarzinom_2025-04.pdf.

[31] National Institute for Health and Care Excellence (NICE). Lung cancer: diagnosis and management. 2024 [cited 10.10.2025]. Available from: https://www.nice.org.uk/guidance/ng122/chapter/diagnosis-and-staging.

[32] Aberdeen Royal Infirmary. Nodules. 2024 [cited 23.10.2025]. Available from: https://www.aberdeenlungs.com/for-patients/nodules.

[33] City of Hope. Lung nodules. 2022 [cited 23.10.2025]. Available from: https://www.cancercenter.com/cancer-types/lung-cancer/risk-factors/lung-nodules.

[34] Mosleh B., Sarova P., Prosch H., Widder J., Aigner C., Idzko M., et al. Management and Outcomes of Pulmonary Nodules in a Real-World Setting. Diagnostics (Basel). 2025;15(13). Epub 20250701. DOI: 10.3390/diagnostics15131677.

[35] MacMahon H., Naidich D. P., Goo J. M., Lee K. S., Leung A. N. C., Mayo J. R., et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. Radiology. 2017;284(1):228-243. DOI: 10.1148/radiol.2017161659.

[36] Food and Drug Administration (FDA). 510(k) Summary for AI-Rad Companion. 2024 [cited 30.10.2025]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf23/K233753.pdf.

[37] Food and Drug Administration (FDA). 510(k) Summary Lunit Insight CXR Triage. 2021 [cited 02.11.2025]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf21/K211733.pdf.

[38] Food and Drug Administration (FDA). 510(k) Summary for Gleamer - ChestView 2025 [cited 02.11.2025]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf24/K241620.pdf.

[39] Food and Drug Administration (FDA). 510(k) Summary ClearRead Xray Pneumothorax. 2022 [cited 03.11.2025]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf21/K213566.pdf.

[40] Food and Drug Administration (FDA). 510(k) Summary Auto Lung Nodule Detection. 2021 [cited 03.11.2025]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf20/K201560.pdf.

[41] Houssein E. H., Gamal A. M., Younis E. M. G. and Mohamed E. Explainable artificial intelligence for medical imaging systems using deep learning: a comprehensive review. Cluster Computing. 2025;28(7):469. DOI: 10.1007/s10586-025-05281-5.

[42] Guan H., Bates D. and Zhou L. Keeping Medical AI Healthy: A Review of Detection and Correction Methods for System Degradation. 2025 [cited 22.11.2025]. Available from: https://arxiv.org/html/2506.17442v1.

[43] Food and Drug Administration (FDA). 510(k) Summary VUNO Med-Chest X-ray Triage/VUNO Med-CXR Link Triage. 2024 [cited 03.11.2025]. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf24/K241439.pdf.

[44] Deeks J., Bossuyt P., Leeflang M. and Takwoingi Y. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). 2023 [cited 15.10.2025]. Available from: https://www.cochrane.org/authors/handbooks-and-manuals/handbook-systematic-reviews-diagnostic-test-accuracy/chapter-pdfs-cochrane-handbook-systematic-reviews-diagnostic-test-accuracy-v20/4-understanding-test-accuracy.

[45] Beddard M., Kiseleva M., O'Connell S. and Rahim A. Artificial intelligence-derived software to analyse chest X-rays for suspected lung cancer in primary care referrals: early value assessment addendum. National Institute for Health and Care Excellence (NICE). 2023.

[46] Breil B., Fritz F., Thiemann V. and Dugas M. Mapping Turnaround Times (TAT) to a Generic Timeline: A Systematic Review of TAT Definitions in Clinical Domains. BMC Medical Informatics and Decision Making. 2011;11(1):34. DOI: 10.1186/1472-6947-11-34.

[47] Colquitt J., Jordan M., Court R., Loveman E., Parr J., Ghosh I., et al. Artificial intelligence software for analysing chest X-ray images to identify suspected lung cancer: an evidence synthesis early value assessment. Health Technology Assessment (Winchester, England). 2024;28(50):1-75. DOI: https://dx.doi.org/10.3310/LKRT4721.

[48] Moss R., Chappell J., Herbert P., Frank L., Stewart J., Emengo H., et al. Artificial intelligence supported clinician review of chest X-rays for suspected lung cancer. Glasgow/Edinburgh: 2025 [cited 09.01.2025]. Available from: https://shtg.scot/our-advice/artificial-intelligence-supported-clinician-review-of-chest-x-rays-from-patients-with-suspected-lung-cancer/.

[49] National Institute for Health and Care Excellence (NICE). Artificial intelligence (AI)-derived software to analyse chest X-rays for suspected lung cancer in primary care referrals: early value assessment (HTE12). 2023 [cited 12.09.2025]. Available from: https://www.nice.org.uk/guidance/hte12/resources/artificial-intelligencederived-software-to-analyse-chest-xrays-for-suspected-lung-cancer-in-primary-care-referrals-early-value-assessment-pdf-50261967918277.

[50] Dissez G., Tay N., Dyer T., Tam M., Dittrich R. and Doyne D. Enhancing Early Lung Cancer Detection on Chest Radiographs with AI-Assistance: A Multi-Reader Study [Preprint]. 2022 [cited 07.09.2025]. Available from: https://arxiv.org/ftp/arxiv/papers/2208/2208.14742.pdf.

[51] Homayounieh F., Digumarthy S., Ebrahimian S., Rueckel J., Hoppe B. F., Sabel B. O., et al. An Artificial Intelligence–Based Chest X-ray Model on Human Nodule Detection Accuracy From a Multicenter Study. JAMA Network Open. 2021;4(12):e2141096-e2141096. DOI: 10.1001/jamanetworkopen.2021.41096.

[52] Jang S., Song H., Shin Y. J., Kim J., Kim J., Lee K. W., et al. Deep Learning–based Automatic Detection Algorithm for Reducing Overlooked Lung Cancers on Chest Radiographs. Radiology. 2020;296(3):652-661. DOI: 10.1148/radiol.2020200165.

[53] Koo Y. H., Shin K. E., Park J. S., Lee J. W., Byun S. and Lee H. Extravalidation and reproducibility results of a commercial deep learning-based automatic detection algorithm for pulmonary nodules on chest radiographs at tertiary hospital. Journal of Medical Imaging and Radiation Oncology. 2021;65(1):15-22. DOI: https://doi.org/10.1111/1754-9485.13105.

[54] Maiter A., Hocking K., Matthews S., Taylor J., Sharkey M., Metherall P., et al. Evaluating the performance of artificial intelligence software for lung nodule detection on chest radiographs in a retrospective real-world UK population. BMJ Open. 2023;13(11):e077348. Epub 20231108. DOI: 10.1136/bmjopen-2023-077348.

[55] Nam J. G., Hwang E. J., Kim D. S., Yoo S.-J., Choi H., Goo J. M., et al. Undetected Lung Cancer at Posteroanterior Chest Radiography: Potential Role of a Deep Learning–based Detection Algorithm. Radiology: Cardiothoracic Imaging. 2020;2(6):e190222. DOI: 10.1148/ryct.2020190222.

[56] Niehoff J. H., Kalaitzidis J., Kroeger J. R., Schoenbeck D., Borggrefe J. and Michael A. E. Evaluation of the clinical performance of an AI-based application for the automated analysis of chest X-rays. Sci Rep. 2023;13(1):3680. Epub 20230305. DOI: 10.1038/s41598-023-30521-2.

[57] Siemens. Company submission on Prototype AI-Rad Companion (redacted confidential information). 2022.

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

[59] Tam M., Dyer T., Dissez G., Morgan T. N., Hughes M., Illes J., et al. Augmenting lung cancer diagnosis on chest radiographs: positioning artificial intelligence to improve radiologist performance. Clin Radiol. 2021;76(8):607-614. Epub 20210511. DOI: 10.1016/j.crad.2021.03.021.

[60] van Beek E. J. R., Ahn J. S., Kim M. J. and Murchison J. T. Validation study of machine-learning chest radiograph software in primary and emergency medicine. Clin Radiol. 2023;78(1):1-7. Epub 20220925. DOI: 10.1016/j.crad.2022.08.129.

[61] Vasilev Y., Vladzymyrskyy A., Omelyanskaya O., Blokhin I., Kirpichev Y. and Arzamasov K. AI-Based CXR First Reading: Current Limitations to Ensure Practical Value. Diagnostics (Basel). 2023;13(8). Epub 20230416. DOI: 10.3390/diagnostics13081430.

[62] Nxumalo Z. Z., Irusen E. M., Allwood B. W., Tadepalli M., Bassi J. and Koegelenberg C. F. N. The utility of artificial intelligence in identifying radiological evidence of lung cancer and pulmonary tuberculosis in a high-burden tuberculosis setting. S Afr Med J. 2024;114(6):e1846. Epub 20240531. DOI: 10.7196/SAMJ.2024.v114i6.1846.

[63] van Leeuwen K. G., Schalekamp S., Rutten M. J. C. M., Huisman M., Schaefer-Prokop C. M., de Rooij M., et al. Comparison of Commercial AI Software Performance for Radiograph Lung Nodule Detection and Bone Age Prediction. Radiology. 2024;310(1):e230981. DOI: 10.1148/radiol.230981.

[64] Nahm F. S. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol. 2022;75(1):25-36. DOI: 10.4097/kja.21209.

[65] Ohlmann-Knafo S., Ramanauskas N., Huettinger S., Jeyakumar J. E., Barušauskas D., Bielskienė N., et al. AI-based software for lung nodule detection in chest X-rays - Time for a second reader approach? 2022 [cited 04.11.2025]. Available from: https://arxiv.org/pdf/2206.10912.

[66] European Cancer Organisation. Radiation Therapist, Austria. [cited 15.11.2025]. Available from: https://www.europeancancer.org/timetoact/impact/resource/radiation-therapist-austria.html.

[67] Österreichisches Onkologie Forum (ÖOF). Versorgungsreport Bronchialkarzinom 2025 [cited 08.11.2025]. Available from: https://www.oesterreichisches-onkologie-forum.at/wp-content/uploads/2025/05/OeOF_Versorgungsreport_Bronchialkarzinom_ohne-Anzeigen_Web.pdf.

[68] Shin H. J., Han K., Ryu L. and Kim E.-K. The impact of artificial intelligence on the reading times of radiologists for chest radiographs. npj Digital Medicine. 2023;6(1):82. DOI: 10.1038/s41746-023-00829-4.

[69] Yuan X., Xu H., Zhu J., Yang Z., Pan B., Wu L., et al. Systematic review and meta-analysis of artificial intelligence for image-based lung cancer classification and prognostic evaluation. npj Precision Oncology. 2025;9(1):300. DOI: 10.1038/s41698-025-01095-1.

[70] Cellina M., Cacioppa L. M., Cè M., Chiarpenello V., Costa M., Vincenzo Z., et al. Artificial Intelligence in Lung Cancer Screening: The Future Is Now. Cancers (Basel). 2023;15(17). Epub 20230830. DOI: 10.3390/cancers15174344.

[71] Svoboda E. Deep learning delivers early detection. 2020 [cited 20.11.2025]. Available from: https://www.nature.com/articles/d41586-020-03157-9.

[72] ASSESS-DHT. Real-World Data Validation methods for AI-based decision support systems. 2025.

Appendix

Glossary[14]

Artificial intelligence (AI): The ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.

CXR: an X-ray image of the chest area, including the lungs, airways, heart and ribs.

Computer-aided detection (CADe): software that can detect abnormalities on a CXR.

Computer-aided diagnosis (CADx): software that can diagnose abnormalities on CXR.

Computer-assisted triage (CAST): supports the prioritisation of medical images that require urgent review.

Negative Predictive Value (NPV): The likelihood that a person who has a negative test result indeed does not have the disease, condition, biomarker, or mutation (change) in the gene being tested. The negative predictive value is a way of measuring how accurate a specific test is.

Positive predictive value (PPV) The likelihood that a person who has a positive test result does have the disease, condition, biomarker, or mutation (change) in the gene being tested. The positive predictive value is a way of measuring how accurate a specific test is.

Sensitivity: The proportion of people who test positive for a disease among people who have the disease of interest. The ratio between the true-positive value and (true-positive value + false-negative value).

$$\text{Sensitivity} = \frac{a}{a+c} = \frac{TP}{TP+FN}$$

Specificity: The proportion of people who test negative for a disease among people who do not have the disease of interest. The ratio between the true-negative value and (true-negative value + false-positive value)

$$\text{Specificity} = \frac{d}{b+d} = \frac{TN}{TN+FP}$$

Reference standard: The test, combination of tests or procedure that is considered the best available method of categorising participants in a study of diagnostic test accuracy as having or not having a target condition

True-negative value: The number of cases in which the index test has correctly indicated the patient as being disease-free. TN = d.

True-positive value: The number of cases in which the index test has correctly indicated the patient as having the disease. TP = a.

False-negative value: The number of cases in which the index test has wrongly suggested the patient as being disease-free when they do have the disease. FN = c.

False-positive value: The number of cases in which the index test has wrongly indicated the patient as having the disease when they do not have the disease. FP = b.

[14] Definitions adopted from NIHR and SHTG reviews [47, 48].

Table A-1: Topics on the prioritisation list ("long list")

Medical field	Specific application area (disease, if known)	AI-tool name (if known)
Radiology	Mammography (breast cancer detection)	Transpara Mammography, Transpara DBT, HealthMammo, ProFound AI for 2D mammography, ProFound AI for DBT
Radiology	Brain CT (stroke detection e.g., vessel occlusion, ischemia)	ELVO, StroCare Suite, Accipio, Aidoc, BioMind, BrainScan CT, Cercare Perfusion, CINA Head, CT Perfusion 4D, e-Stroke, icobrain ct, Neuro Solutaion, qER, Viz LVO, Viz ICH, head, ASPECTS, Zebra triage, DLCExpert, e-CTA/e-ASPECTS, Avicenna CINA LVO, RapidAI®/CTA/LVO/CTP, ischemaView
Radiology	Chest CT (CTCA scan for coronary artery disease early detection and risk stratification)	AI-Rad Companion, CaRi-Heart
Radiology	Chest CT (lung disease/nodules detection and characterization)	Contectflow Search Lung CT, AI-Rad Companion Chest CT, AVIEW LCS+, ClearRead CT, InferRead CT lung, LD-01K, Lung AI, Lung Nodule AI, qCT, SenseCare-Lung Pro, Veolity, Veye Lung Nodules, VUNO Med-Lung CT AI, Veye Chest, Icolung
Radiology	Chest X-ray (lung cancer detection)	Annalise Enterprise CXR, qXR, AI-Rad Companion Chest X-ray, Auto Lung Nodule Detection, ChestLink, ChestView, Chest X-ray, ClearRead Xray, InferRead DR Chest, Lunit INSIGHT CXR, Milvue Suite, Red dot, SenseCare-Chest DR PRO, VUNO Med-Chest X-Ray, Gleamer, Veolity
Radiology	Chest X-ray (infection detection in intensive care units)	MAIDA
Radiology	Hip and knee X-rays (arthritis detection)	KOALA (Knee OsteoArthritis Labeling Assistant), Imaging Biopsy Lab (knee and hip)
Radiology	X-rays (fracture detection)	Gleamer, Radiobotics Fracture
Radiology	X-rays (spinal cord injury detection)	KiaMed
Radiology	MRI (atypical Parkinson syndrome diagnosis)	NA
Internal medicine	Colonoscopy for detecting and diagnosis of polyps (CADe=computer-aided detection, CADx=computer-aided diagnosis)	GI Genius™ , CAD EYE
Dermatology	Digital dermoscopy (malignant melanoma detection)	DB-MIPS, DermoGenious, DermoGenius Basic II, FotoFinder bodyscan ATBM, MicroDERM, Mole Max II, Mole Expert, SolarScan, nomela, DERM, Molenanalyzer pro, Skin Vision
Ophthalmology	Retina scan (diabetic retinopathy detection)	CARA/Neoretina, EyeArt, IDx-DR V2.0, Retmarker, OpthtAI, SELANA+, RetinaLyze
Ophthalmology	Retinal imaging (rare hereditary retinal disease detection)	NA
Pathology	Prostate biopsy (prostate cancer detection)	Galen Prostate Solution, Paige Prostate, DeepDx
General medicine	Wound measurements (3D imaging of wounds, automatic assessment, centralised digital dashboard)	Minuteful for Wounds, insight, Cares4Wounds, Tissue Analytics, Swift Wound, Wound Viewer, ImageJ software

Abbrevations: AI … Artificial Intelligence, ASPECTS … Alberta Stroke Program Early CT Score, AVIEW LCS+ … Advanced Visualization for Lung Cancer Screening Plus, CADe … Computer-Aided Detection, CADx … Computer-Aided Diagnosis, CINA … Computed Imaging Neuro Analysis, CT … Computed Tomography, CTCA … Coronary Computed Tomography Angiography, CXR … Chest X-Ray, DB-MIPS … Digital Dermoscopy Melanoma Imaging Processing System, DBT … Digital Breast Tomosynthesis, DR … Digital Radiography, ELVO … Emergent Large Vessel Occlusion, HTA … Health Technology Assessment, MAIDA … Medical Artificial Intelligence for Detection in ICU Chest X-rays, MRI … Magnetic Resonance Imaging, NA … Not Available, KOALA … Knee OsteoArthritis Labeling Assistant.

Survey: KI-Anwendungsbereiche

Bitte stufen Sie die Relevanz von 16 KI-Anwendungen (für Krankenhäusern in Österreich) für die Unterstützung der diagnostischen Bildgebung ein und wählen Sie davon drei aus, die Sie anhand der folgenden Kriterien als am wichtigsten oder relevantesten einstufen. Stufen Sie bitte anschließend 2 Anwendungsbereiche der Dokumentations-/Verwaltungsunterstützung (für Krankenhäusern in Österreich) anhand derselben Kriterien ein. Ihre Auswahl sollte Ihr fachliches Urteil widerspiegeln, das durch die nachstehenden Kriterien gestützt wird:

- **Klinische Relevanz:** Dazu gehören die Neuartigkeit der Technologie, die Frage, ob sie einem klinischen Bedarf entspricht, ihr Potenzial zur Verbesserung von Patient:innenergebnissen oder klinischen Arbeitsabläufen sowie die Verfügbarkeit von Evidenz.
- **Auswirkungen auf die Ressourcen:** Berücksichtigt die Häufigkeit der Anwendung, die damit verbundenen Kosten und die erwarteten Auswirkungen auf die Nutzung der Gesundheitsressourcen.
- **Durchführbarkeit der Implementierung:** Betrachtet potenzielle Hindernisse wie organisatorische Widerstände, Einschränkungen der Infrastruktur oder Bedenken hinsichtlich der Datensicherheit.
- **Potenzielle Risiken oder unbeabsichtigte Folgen:** Wie z. B. Diagnosefehler, erhöhte Arbeitsbelastung oder ethische Fragen.

Bitte bewerten Sie jedes Element auf einer Skala von 1 (geringste Relevanz) bis 10 (höchste Relevanz).

Wenn Sie ein Tool nicht kennen, um es zu beurteilen, wählen Sie bitte „keine Antwort“.

Welche der folgenden Rollen beschreibt Ihre derzeitige berufliche Tätigkeit am besten?

- Klinische Anwender:in
- IT-Verantwortliche:r Im Krankenhaus
- Andere klinische Frachkraft
- Andere (bitte Angeben)

In Österreich tätig?

- Ja
- Nein
- Keine Antwort

Unterstützung der diagnostischen Bildgebung

Bitte bewerten Sie jedes Element auf einer Skala von 1 (geringste Relevanz) bis 10 (höchste Relevanz).

1. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Mammographie (Brustkrebserkennung)

   KI-Anwendungen (falls bekannt): Transpara Mammography, Transpara DBT, HealthMammo, ProFound AI for 2D mammography, ProFound AI for DBT .

2. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Gehirn-CT (Schlaganfallerkennung, z. B. Gefäßverschluss, Ischämie).

   KI-Anwendungen (falls bekannt): ELVO, StroCare Suite, Accipio, Aidoc, BioMind, BrainScan CT, Cercare Perfusion, CINA Head, CT Perfusion 4D, e-Stroke, icobrain ct, Neuro Solutaion, qER, Viz LVO, Viz ICH, head, ASPECTS, Zebra triage, DLCExpert, e-CTA/e-ASPECTS, Avicenna CINA LVO, RapidAI®/CTA/LVO/CTP, ischemaView

3. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Thorax-CT (CTCA-Scan zur Früherkennung und Risikostratifizierung koronarer Herzkrankheiten).

   KI-Anwendungen (falls bekannt): AI-Rad Companion, CaRi-Heart

4. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Thorax-CT (Erkennung und Charakterisierung von Lungenkrankheiten/Knötchen).

   KI-Anwendungen (falls bekannt): Contectflow Search Lung CT, AI-Rad Companion Chest CT, AVIEW LCS+, ClearRead CT, InferRead CT lung, LD-01K, Lung AI, Lung Nodule AI, qCT, SenseCare-Lung Pro, Veolity, Veye Lung Nodules, VUNO Med-Lung CT AI, Veye Chest, Icolung.

5. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Röntgenaufnahme der Brust (Erkennung von Lungenkrebs).

   KI-Anwendungen (falls bekannt): Annalise Enterprise CXR, qXR, AI-Rad Companion Chest X-ray, Auto Lung Nodule Detection, ChestLink, ChestView, Chest X-ray, ClearRead Xray, InferRead DR Chest, Lunit INSIGHT CXR, Milvue Suite, Red dot, SenseCare-Chest DR PRO, VUNO Med-Chest X-Ray, Gleamer, Veolity

6. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Thorax-Röntgenbilder (Erkennung von Infektionen auf der Intensivstation).

   KI-Anwendung (falls bekannt): MAIDA

7. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Röntgenaufnahmen von Hüfte und Knie (Erkennung von Arthrose).

   KI-Anwendungen (falls bekannt): KOALA (Knee OsteoArthritis Labeling Assistant), Imaging Biopsy Lab (Knie und Hüfte)

8. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Röntgenaufnahmen (Erkennung von Frakturen).

   KI-Anwendungen (falls bekannt): Gleamer, Radiobotics Fracture

9. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): Röntgenaufnahmen (Erkennung von Rückenmarksverletzungen).

   KI-Anwendung (falls bekannt): KiaMed

10. Medizinisches Fachgebiet: Radiologie; Anwendungsbereich (Krankheit, falls bekannt): MRT (Diagnose atypisches Parkinson-Syndrom).

    KI-Anwendungen (falls bekannt): -

11. Medizinisches Fachgebiet: Innere Medizin; Anwendungsbereich (Krankheit, falls bekannt): Koloskopie (Erkennung von Polypen, Diagnose von Darmkrebs).

    KI-Anwendungen: GI Genius™ , CAD EYE

12. Medizinisches Fachgebiet: Dermatologie; Anwendungsbereich (Krankheit, falls bekannt): Digitale Dermatoskopie (Erkennung von malignen Melanomen).

    KI-Anwendungen (falls bekannt): DB-MIPS, DermoGenious, DermoGenius Basic II, FotoFinder bodyscan ATBM, MicroDERM, Mole Max II, Mole Expert, SolarScan, nomela, DERM, Molenanalyzer pro, Skin Vision

13. Medizinisches Fachgebiet: Ophthalmologie; Anwendungsbereich (Krankheit, falls bekannt): Netzhautscan (Erkennung von diabetischer Retinopathie).

    KI-Anwendungen (falls bekannt): CARA/Neoretina, EyeArt, IDx-DR V2.0, Retmarker, OpthtAI, SELANA+, RetinaLyze

14. Medizinisches Fachgebiet: Ophthalmologie; Anwendungsbereich (Krankheit, falls bekannt): Netzhautbildgebung (frühzeitige Diagnostik von Netzhauterkrankungen).

    KI-Anwendungen (falls bekannt): -

15. Medizinisches Fachgebiet: Pathologie; Anwendungsbereich (Krankheit, falls bekannt): Prostata-Biopsie (Erkennung von Prostatakrebs).

    KI-Anwendungen (falls bekannt): Galen Prostate Solution, Paige Prostate, DeepDx

16. Medizinisches Fachgebiet: Allgemeine Medizin; Anwendungsbereich (Krankheit, falls bekannt): Wundmessungen (3D-Darstellung von Wunden, automatische Bewertung, zentralisiertes digitales Dashboard).

    KI-Anwendungen (falls bekannt): Minuteful for Wounds, insight, Cares4Wounds, Tissue Analytics, Swift Wound, Wound Viewer, ImageJ software

Bitte wählen Sie nun **die 3 KI-gestützten Technologien** zur Unterstützung der diagnostischen Bildgebung aus, **die Sie am höchsten bewertet haben**. Geben Sie im Kommentarfeld an, **welche davon am wichtigsten (1), die zweitwichtigste (2) und die drittwichtigste (3) ist**. Falls Sie eine bestimmte KI-Anwendungen hervorheben können, ergänzen Sie diese bitte ebenfalls im Kommentarfeld. Bitte lassen Sie die anderen Kommentarfelder leer.

Dokumentations-/Verwaltungsunterstützung

Bitte bewerten Sie jedes Element auf einer Skala von 1 (geringste Relevanz) bis 10 (höchste Relevanz).

1. Speech-to-Text-Tools (Spracherkennung)

   Anwendungbereich: Verschriftlichung von gesprochener Sprache in Text, (Konvertieren Gesprochenes von Klinikern in Berichte/Entlassungsschreiben)

   KI-Anwendung (falls bekannt): Dragon Medical One

2. KI-generierter Text/Berichterstattung

   Anwendungsbereich: Generieren oder vervollständigen von Texten mit Hilfe Large Language Models (LLMs - z. B. Zusammenfassungen, Berichte, Briefe)

   KI-Anwendung (falls bekannt): VNAI-Broker, ChatGPT oder andere Generative KI-Modelle

Welche KI-Anwendungen, die nicht erwähnt wurden, werden in ihrer Organisation genutzt/getestet?

Gibt es KI-Anwendungsfelder, die Ihrer Einschätzung nach in Ihrer Organisation künftig an Bedeutung gewinnen werden/können?

*Table A-2: Test performance results from primary studies (Legend: **red** from NIHR; **green** from SHTG; **blue** from CEDAR; **not coloured** from update search.)*

Study	AI name	N patients	N CXRs	Accuracy (95% CI)	PPV (95% CI)	NPV (95% CI)	FDR (95% CI)	FOR (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Concordance
Lung cancer detection												
NHS Grampian service evaluation (SHTG 2025)	Annalise Enterprise CXR (Annalise ai)	Retrospective cohort: 113 Prospecive cohort: 68 *cancer* patients	NR	NR	AI + Rad vs Rad review: 3% AI + Rad vs confirmed diagnosis: 1%	AI + Rad vs Rad review: 99.99% AI + Rad vs confirmed diagnosis: 100%	NR	NR	NR	AI + Rad vs Rad review: 78% AI + Rad vs confirmed diagnosis: 82%	91% (both comparisons)	NR
Dissez 2022	Red Dot (Behold.ai)	400	400	AI + Rad: 75% (71% to 79%) Rad: 75% (69% to 79%)	AI + Rad: 41% (38% to 43%) Rad: 44% (40% to 48%)[15]	Reported as *FN*[16] Rad: 24 AI + Rad: 17	Reported as *FP*[17] Rad: 266 AI + Rad: 82	Reported as *TN*[18] Rad: 62 AI + Rad: 246	NR	AI + Rad: 77% (75% to 80%) Rad: 66% (59% to 71%)	AI + Rad: 75% (71% to 77%) Rad: 81% (77% to 85%)	AI + Rad: 57% Rad: 42%
Maiter 2023	ALND AI software (Samsung Electronics Version V.1.0)	5,592	5,722	AI: 83% (82% to 84%) Rad: 98% (97% to 98%)	AI: 6% (5% to 7%) Rad: 36% (31% to 41%)[19]	AI: 99% (99% to 99%) Rad: 99% (99% to 99%)[20]	Reported as *FP:* AI: 943 Rad: 110	Reported as *TN:* AI: 4687 Rad: 5520	NR	AI: 61% (50% to 70.9%) Rad: 66% (56% to 76%)	AI: 83% (82% to 84%) Rad: 98% (98% to 98%)	NR
Nxumalo 2024	qXR (Qure.ai)	127	NR	NR	AI: 97% (95% to 99%)	AI: 62% (54% to 71%)	NR	NR	NR	AI: 84% (80% to 87%)	AI: 91% (85% to 97%)	NR
Suspicious lung nodule detection												
Maiter 2023	ALND AI software (Samsung Electronics Version V.1.0)	5,592	5,722	AI: 83% (82% to 84%)	AI: 6% (5% to 7%)	AI: 99% (99% to 99%)	NR	NR	NR	AI: 55% (44% to 64%)	AI: 83% (82% to 84%)	NR
Homayounieh 2021	AI-Rad Companion CXR (Siemens Healthineers)	100	100	Rad: 69% (62% to 77%) AI + Rad: 75% (70% to 81%)	Reported as *TP:* Rad: 23.6 AI + Rad: 26.4	Reported as *FN:* Rad: 26.4 AI + Rad: 23.6	Reported as *FP:* Rad: 4.1 AI + Rad: 2.5	Reported as *TN:* Rad: 45.5 AI + Rad: 47.5	NR	Rad: 45% (38% to 53%) AI + Rad: 55% (48% to 63%)	Rad: 93% (89% to 96%) AI + Rad: 95% (91% to 9%)	NR

[15] TP was not reported by Dissez et al. but calculated by the NIHR review authors. AI + Rad: 55, Rad: 48.

[16] FN was not reported by Dissez et al. but calculated by the NIHR review authors.

[17] FP was not reported by Dissez et al. but calculated by the NIHR review authors.

[18] TN was not reported by Dissez et al. but calculated by the NIHR review authors.

[19] TP also reported: AI: 56, radiologists: 61.

[20] FN also reported: AI: 36, radiologists: 31.

Study	AI name	N patients	N CXRs	Accuracy (95% CI)	PPV (95% CI)	NPV (95% CI)	FDR (95% CI)	FOR (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Concordance
Koo 2021	Lunit INSIGHT CXR version 1.00	378	434	NR	Reported as *TP*: Rad: 152 AI + Rad: 157	Reported as *FN*: Rad: 13 AI + Rad: 8	Reported as *FP*: Rad: 15 AI + Rad: 6	Reported as *TN*: Rad: 198 AI + Rad: 207	Rad: 93% AI + Rad: 96%	Rad: 92% (87% to 96%) AI + Rad: 95% (91% to 98%)	Rad: 93% (89% to 96%) AI + Rad: 97% (94% to 99%)	NR
Nam 2020	Lunit INSIGHT version 1.0.1.1	218	218	NR	Reported as *TP*: AI: 117 Rad: 316 AI + Rad: 357	NR	Reported as *FP*: AI vs Rad: 21% vs 19% Rad vs AI + Rad: 19% vs 19%	Reported as *TN*: AI: 47 Rad: 156 AI + Rad: 164	AI: 90% Rad: 63% to 66% AI + Rad: 69% to 72%	AI: 70% Rad: 47% (43% to 51%) AI + Rad: 53% (49% to 57%)	AI: 94% Rad: 78% (72% to 84%) AI + Rad: 82% (77% to 87%)	NR
Jang 2020	Lunit INSIGHT version 1.2.0.0	351	351	NR	Reported as *TP*[21]: Rad: 50 AI + Rad: 66	Reported as *FN*: Rad: 67 AI + Rad: 51	Reported as *FP*: Rad: 24 AI + Rad: 19	Reported as *TN*: Rad: 210 AI + Rad: 215	Rad: 67% (62% to 72%) AI + Rad: 76% (71% to 81%)	Rad: 43% (34% to 52%) AI + Rad: 56% (47% to 65%)	Rad: 90% (86% to 94%) AI + Rad: 92% (88% to 95%)	NR
Niehoff 2023[22]	AI-Rad Companion CXR (Siemens Healthineers Version VA23A)	499	499	NR	AI: At CS[23] ≥6: 38% (80 at CS=10) Rad: 79%	AI: At CS ≥6: 97% (91 at CS=10) Rad: 0.94	AI: At CS ≥6: 62% (20 at CS=10) Rad: 21%	AI: At CS ≥6: 3% (9 at CS=10) Rad: 6%	AI: 87% Rad: 75%	AI: At CS ≥6: 83% (28 at CS=10) Rad: 52%	AI: At CS ≥6: 83% (99 at CS=10) Rad: 98%	50.3%[24]
Smith 2023	Red Dot (Behold.ai, V2.2)	4,654	4,076	NR	NR	AI: 96% Rad: NR	NR	NR	NR	NR	NR	0.77%[25]
Tam 2021	Red dot (Behold.ai, Version NR)	400	NR	Rad: 84% to 90% AI: 87% Rad + AI: 90% to 91%	Reported as *TP*: Rad: 136 to 171 AI: 159 Rad + AI: 176 to 186	Reported as *FN*: Rad: 27 to 62 AI: 39 Rad + AI: 12 to 22	Reported as *FP*: Rad: 1 to 12 AI: 14 Rad + AI: 15 to 23	Reported as *Precision*: Rad: 93% to 99% AI: 92% Rad + AI: 89% to 92%	NR	Rad: 69% to 86% AI: 80% Rad + AI: 89% to 94%	Rad: 94% to 99% AI: 93% Rad + AI: 88% to 92%	Overall combined Rad + AI: 92%

[21] TP, FN, FP and TN were not reported by Jang et al. but calculated by the NIHR review authors.

[22] This study investigated the detection of lung lesions, not lung nodules. A nodule is a small, round, well-defined type of lung lesion (≤3 cm), whereas lesion is a broader term for any abnormal area in lung tissue.

[23] AI-Rad provides a "confidence score" (CS) on a scale of 1 (low) to 10 (high) for each finding, which expresses the algorithm´s certainty for the presence of that particular finding. The manufacturer has preset the AI-Rad only to report findings with a CS ≥ 6, whilst findings with a CS ≤ 5 are not displayed.

[24] Concordance for all pathologies, not just lung lesions.

[25] Discrepancy rate defined as the proportion of all processed exams that were incorrectly classified as HCN according to auditing radiologists.

Study	AI name	N patients	N CXRs	Accuracy (95% CI)	PPV (95% CI)	NPV (95% CI)	FDR (95% CI)	FOR (95% CI)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Concordance
Van Beek 2023	Lunit INSIGHT CXR (Lunit Version 3.1.2.0)	1,960	NR	AI: 86% (85% to 88%)	NR	NR	NR	NR	AI: ED: 88% (81% to 95%) GP: 90% (84% to 97%)	AI: ED: 79% (62% to 91%) GP: 83% (65% to 94%)	AI: ED: 85% (82% to 87%) GP: 89% (86% to 90%)	NR
Vasilev 2023	Lunit INSIGHT CXR (Lunit Version 3.110)	4,825	4,825	NR	NR	NR	NR	NR	*Retrospective*: AI: 94% (87% to 100%) Rad: 97% (94% to 100%) *Prospecti ve*: AI: 84% (82% to 86%) Rad: 89% (86% to 92%)	*Retrospective*: AI: 90% (79% to 100%) Rad: 90% (79% to 100%) *Prospective*: AI: 77% (73% to 80%) Rad: 86% (82% to 91%)	*Retrospective*: AI: 89% (79% to 98%) Rad: 95% (89% to 100%) *Prospective*: AI: 81% (80% to 82%) Rad: 92% (88% to 96%)	*Retrospective*: 86% *Prospective*: 81%
Van Leeuwen 2024	Infervision, Milvue, VUNO, Lunit, Siemens Healthineers, Annalise.ai, Oxipit	386		NR	NR	NR	NR	NR	Rad: 81% (77% to 85%) AI: Annalise.ai: 90% (87% to 94%); Lunit: 93% (91% to 96%); Milvue: 86% (82% to 90%); Oxipit: 88% (85% to 92%); Infervision: 79% (74% to 84%); Siemens Healthineers: 80% (75% to 85%); VUNO: 84% (80% to 88%).	Rad: 71% (66% to 75%) AI: Infervision: 64% (56% to 72%), Milvue: 50% (42% to 58%), Siemens Healthineers: 66% (58% to 74%), VUNO: 75% (68% to 82%), Lunit: 89% (84% to 94%)	Clinicians: 80% (73% to 85%) AI: Infervision: 83% (79% to 88%), Milvue: 99% (97% to 100%), Siemens Healthineers: 87% (83% to 91%), VUNO: 88% (83% to 92%), Lunit: 80% (75% to 85%)	All nodule detection algorithms and the reader mean too showed performance decline with decreasing nodule conspicuity class. Nodule size showed limited correlation with AUC for most algorithms and the reader mean.

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.

Table A-3: Organisational implications from primary studies (Legend: **red** *from NIHR;* **blue** *from CEDAR;* **not coloured** *from update search.)*

First author and year	Workflow	Participant feedback	Time-related outcomes
Dissez 2022	With red dot® the number of patients potentially referred to CT would increase from 117 (29%) (95% CI 93 to 147) to 144 (36%) (95% CI 119 to 172) on average. However, the simulated increase in CT referral would have resulted in an increase of 17.4% lung cancers detected, resulting in the proportional change in lung cancer diagnostic CTs being negligible (from 39% to 38%) and statistically non-significant (p= 0.22). [26]	Upon completion of the study, participants were asked to take part in a survey to collect their feedback on the AI model. 8 of the 10 participants declared that reporting was not slower when using the algorithm, and 9 out of 10 reported that the heatmaps produced by the AI model were helpful to understand the algorithm's attention points.	NR
Jang 2020	With Lunit INSIGHT the number of patients potentially referred to CT would increase from 80 (23%, 95% CI 18.5% to 27.5%) to 96 (27%, 95% CI 22.8% to 32.3%)[27] on average.	NR	Average reading time: per image with AI 22.5 [standard deviation (SD) 40.3] seconds versus 24.3 (SD 27.4) seconds without AI.
Koo 2020		NR	Average reading time: to read 434 CXRs with AI 171 (SD 33.8) minutes versus 211.25 (SD 38.4) minutes without AI.
Smith	Service levels throughout the study period remained high. Upon either site submitting a radiograph, results of the DL algorithm were returned in a mean time of 7.1 seconds (range 5-17 seconds) Radiographs classed as HCN by the DL algorithm were audited by independent radiologists and results were returned to the hospital. Of these exams, 99.3% were audited within 24 hours of the radiograph's submission. The average time taken from exam submission to audit was 3 hours and 50 minutes.	NR	Time taken to report the result: with AI mean 7.1 sec (range 5 to 17 sec) versus mean 3 hours 50 min without AI.
Tam	Positioning AI as the first reader of examinations stands to improve the overall accuracy and sensitivity to potential cancer cases. In the workflow, direct triaging of positive cases will also reduce the CXR reporting burden; however, the increase in false positivesmay be passed onto CT and other follow-up services. Full clinical implementation of this algorithm may still involve clinical review of HCT-positive examinations, meaning reader disagreement may decrease the falsepositive rate. This clinicianeAI interaction may give the additional benefit of knowledge transfer from the algorithm to readers, an area of huge potential for clinical algorithms and worthy of further study (this is in the discussion only).	NR	NR

[26] It is important to note that these are hypothetical referrals. We found no evidence on the impact of AI on the readers' behaviour in real-world clinical practice.

[27] Percentages and Cis calculated by the NIHR assessment team. It is important to note that these are hypothetical referrals. We found no evidence on the impact of AI on the readers' behaviour in real-world clinical practice.

First author and year	Workflow	Participant feedback	Time-related outcomes
SHTG 2025	Based on a Health Foundation Survey: High acceptance for **admin use** of AI (NHS staff 81%, general public 61%) but operational risk that AI could reduce patient–clinician contact (staff 65%, public 53%) Broad support for **clinical use** (staff 76%, public 54%).	Based on a Health Foundation Survey: Persistent concerns about inaccuracy (26-28%) and strong desire for transparency (people would like to be told when AI is used). 57% of NHS staff was looking forward to use AI in their job. 17% public and 10% of NHS staff thought that use of AI would make quality of care worse. 53% public and 65% of NHS staff were concerned that use of AI technologies makes staff feel more distant from patients or clinicians.	NHS Grampian evaluation: patients in NHS Grampian received a CT scan 6 days more quickly following a CXR report, which was statistically significant (95% CI [3.647,7.369], p<0.001) there was a 7-day reduction in average time to treatment from the pre-pandemic baseline (mean=58 days, SD=35) to post-implementation (mean=51 days, SD=20), but this was not statistically significant (95% CI [-1.62,14.418], p=0.117)

Abbreviations: AI … Artificial Intelligence, CI … Confidence Interval, CXR … Chest X-Ray, DL … Deep Learning, HCT … High-Contrast Thorax, HCN … High-Confidence Nodules, NHS … National Health Service, NR … Not Reported, SD … Standard Deviation, SHTG … Scottish Health Technologies Group, UK … United Kingdom.

Table A-4: Risk of bias assessment of the included HTAs

Reference	Concerns with the review process
NIHR	1. Study eligibility criteria: low. 2. Identification & selection of studies: low. 3. Data collection & study appraisal: unclear, data extraction and appraisal methods were described only briefly; with single-reviewer processes and missing information, bias cannot be ruled out but is not demonstrated. 4. Synthesis & findings: low.
Cedar	1. Study eligibility criteria: low. 2. Identification & selection of studies: low. 3. Data collection & study appraisal: unclear, study details extracted sufficiently, but no procedural information on dual extraction or independent risk-of-bias appraisal; no formal RoB tool. 4. Synthesis & findings: low.
SHTG	1. Study eligibility criteria: unclear, no public protocol or registration, no information about post hoc changes or deviations; restrictions (UK context, English language). 2. Identification & selection of studies: low. 3. Data collection & study appraisal: unclear, procedural detail on extraction and quality appraisal is limited. Study details extracted sufficiently, but no procedural information on dual extraction or independent risk-of-bias appraisal; some qualitative bias comments only. 4. Synthesis & findings: low.

Table A-5: Ongoing studies

Study	AI software (manufacturer)	Study Design	Status	Outcomes	Comments	How the study could address the research need
ID: NCT06075836 *Public Title:* AI Assisted Detection of Chest X-Rays (AID-CXR) *Official Title:* Utility of an AI-based CXR Interpretation Tool in Assisting Diagnostic Accuracy, Speed, and Confidence of Healthcare Professionals: a Study Using 500 Retrospectively Collected Inpatient and Emergency Department CXRs From Two UK Hospital Trusts *County:* UK	Lunit INSIGHT CXR (Lunit, Inc.)	Retrospective observational reader-study: ~500 retrospectively collected inpatient & emergency department chest X-rays (CXRs) from two UK hospital trusts; readers from various clinical groups (emergency medicine, ICU, general medicine, radiographers, general radiologists) will interpret CXRs without and with AI assistance; ground truth by two thoracic radiologists (third senior thoracic arbitration)	Active, not recruiting	*Primary outcomes:* diagnostic accuracy metrics (sensitivity, specificity, PPV, NPV) of AI alone and reader + AI vs reader alone *secondary outcomes:* reader speed/efficiency, reader confidence/trust, performance across reader types/specialties; the abnormal findings include: pulmonary nodules/mass, consolidation, pneumothorax, atelectasis, calcification, cardiomegaly, fibrosis, mediastinal widening, pleural effusion, pneumoperitoneum.	The retrospective design limits assessment of AI in real-time workflow; reader study may not reflect full clinical complexity of inpatient/ emergency settings; ground truth from thoracic radiologists may differ from generalist clinical interpretation; sample size (~500 images) may restrict subgroup analyses (reader types, pathology types) and might be under-powered for rarer abnormalities.	How the study could address the research need: This study directly evaluates an AI tool (Lunit INSIGHT CXR) in a real hospital dataset of inpatients/emergency CXRs, with a heterogeneous reader cohort (not only expert radiologists). By comparing reader performance with and without AI support, it provides evidence on how AI might improve diagnostic accuracy, speed and confidence in non-ideal conditions (emergency/inpatient). Thus it addresses the gap between algorithm performance on curated datasets and actual clinical interpretative workflow, especially for common chest-XR abnormalities including lung nodules/masses relevant for lung cancer detection.
ID: NCT05594485 *Public Title:* Retrospective Study of Carebot AI CXR Performance in Preclinical Practice *Official Title:* Chest X-Ray Abnormality Detection Using Artificial Intelligence: Retrospective Study of Carebot AI CXR Performance in Preclinical Practice *Country:* Czech Republic	Carebot AI CXR (Carebot s.r.o., Czechia)	Retrospective observational study: 127 anonymised chest X-rays collected between 15-17 Aug 2022 from one municipal hospital; five independent radiologists of varying experience annotated presence of 12 predefined abnormalities; excluded paediatric (<18 yrs), lateral projections, technically poor images.	Completed (data collection 15-17 Aug 2022, presumably completed by 20 Oct 2022)	Diagnostic performance of AI vs radiologists on the set of 12 abnormalities; evaluation of "clinical impact" of false negatives by AI.	The small sample size (127 images) severely limits general-isability; single hospital/single region dataset may have limited diversity; retrospective design means no assessment of workflow integration or real-time usage; focusing on 12 selected abnormalities may not reflect full spectrum of chest X-ray findings in lung cancer screening/diagnosis.	Although limited, this study offers early real-world data (though retrospective) on how a commercial AI (Carebot AI CXR) performs on actual hospital CXR images (including lung disease findings) compared with human readers. It begins to fill the gap between AI algorithm development and clinical practice by providing performance metrics in a non-ideal setting.
ID: NCT06044454 *Public Title:* Radiograph Accelerated Detection and Identification of Cancer in the Lung (RADICAL) *Official Title:* RADICAL: A Mixed Methods Study to Assess the Clinical Effectiveness and Acceptability of an Artificial Intelligence Software to Prioritise Chest X-ray (CXR) Interpretation *Country:* UK: Scotland,	qXR (Qure.ai, India)	Mixed-methods study: stepped-wedge cluster-randomised design with retrospective technical evaluation then prospective clinical effectiveness, qualitative acceptability work and cost-utility analysis	Active, not recruiting (per registry)	*Primary outcomes:* reduction in reporting time for CXRs flagged for suspicion of lung cancer. *Secondary outcomes:* technical performance of AI, safety, health economics, acceptability.	Strong design with prospective clinical component; however, real-world deployment risks (workflow change, user acceptance) may affect outcomes; cluster design may have contamination; focusing on prioritisation rather than full diagnosis may limit generalisability	By embedding AI into actual hospital CXR workflow and measuring real-world effectiveness (reporting time, prioritisation of suspicious cases) and economic/ acceptability outcomes, this study helps bridge the gap between AI algorithm evaluation and clinical impact in lung cancer (or suspected lung cancer) detection on chest X-rays.

Study	AI software (manufacturer)	Study Design	Status	Outcomes	Comments	How the study could address the research need
ID: NCT06456203 *Public Title:* Trial of Artificial Intelligence for Chest Radiography (ACER) *Official Title:* Artificial Intelligence for Chest Radiography: Impact on Economics, Patient Outcomes and Radiology Service Delivery *Country:* USA	N.A	Randomised clinical trial for chest X-ray interpretation using AI vs standard reading.	Not yet recruiting	Diagnostic performance of AI vs standard reading in chest X-rays (lung cancer/ pneumonia)	lack of manufacturer specification reduces transparency; early phase before recruitment limits insight into real-world workflow; combining pneumonia & lung cancer may complicate specificity of findings.	This study could help fill the gap by assessing AI tool performance in chest X-ray interpretation for lung-cancer relevant cases (and pneumonia), in a randomised setting – thus moving closer to real-world evidence of AI assistance in thoracic imaging.

Implementation checklist

Table A-6: Checklist for decision-makers

Checklist	
Purpose	
	What is the main purpose of the AI and what is the main utility?
	Which specific healthcare processes will be affected?
	Who are the intended users (healthcare professionals, patients, administrators)?
Regulatory Requirements	
Medical Device Classification	
	Is it considered a medical device under MDR?
	What is its risk classification under MDR (Class I, IIa, IIb, or III)
	What is its risk classification under EU AI Act (high-risk, low-risk)?
	Does the AI-system adhere to high-risk AI systems transparency and safety requirements? (see MDR, EU AI Act)
	Is a valid CE marking present?
Data Protection and Privacy	
	Does the AI-enabled DHT comply with GDPR requirements?
	Are there procedures for patient consent and data rights?
	Consider the EHDS once fully implemented.
HTA Evaluation	
	Reflect on who will conduct the assessment, if HTA-reports are not yet available
AI relevant considerations (covered in standard methodology[28])	
CUR	What are the main characteristics of the health problem, including the proposed AI solution, and the specific patient populations and clinical settings where it can be implemented?
TEC	What are the main characteristics of the AI-enabled DHT?
EFF	What are the clinical benefits and quality of life impact of the AI-enabled DHT, and are the benefits superior to those of existing alternatives?
SAF	Are there risks or possible undesirable effects caused by the AI-enabled DHT that could lead to physical or psychological harm to patients or professionals?
ETH	Does the AI-enabled DHT have an impact on inequalities?
SOC	What is the user experience of the AI-enabled DHT?
ORG	Does the implementation of the AI-enabled DHT involve the training of the professional team?
ECO	What are the costs of acquiring, maintaining and using the AI-enabled technology at the patient and health system level?
AI-specific considerations (not covered in standard methodology)	
TEC	Which data sets were used for training and validating the DHT? Is there a strategy how to handle incomplete data? What is the type of machine learning? How will the performance be measured? *Is there evidence that training and validation datasets are representative of the target clinical population (e.g. age, sex, disease spectrum)? Has the AI been externally validated in a setting comparable to the Austrian hospital (or locally)?*
EFF	*Are performance results reported for relevant subgroups (e.g. sex, age, comorbidity)?*
SAF	Are there strategies on data risk management foreseen? How can anomalies of the AI-enabled DHT in operational use be detected?
ETH	Are there strategies to mitigate algorithmic bias in the AI-enabled DHT?
ORG	What is the level of professional oversight? Is staff's approval needed for action, proposed by the AI-enabled DHT? Has the output been cross-checked by a qualified human?
ECO	Is it clear what ongoing support is available for adopters and what it would cost?

[28] E.g. the EUnetHTA Core Model

Monitoring of performance	
	Define strategies on post-deployment for the AI-enabled DHT.
	How often will the AI-enabled DHT be monitored and by whom? *Is there a documented process for software updates and model re-training, including version control and change logs?* *Is there a plan for post-market performance monitoring (e.g. periodic audits, drift detection, incident reporting)?*
	How will changes in performance be detected and measured? *Is it clear whether the model is static or adaptive, and how changes will be communicated to the hospital?*
	When should a re-assessment of the AI-enabled DHT be conducted?
Check again in case of changes in performance and purpose	

Legend: Questions in italics represent additions made in this pilot, reflecting issues identified as particularly important for AI-supported CXR tools in this assessment.

Abbreviations: AI … Artificial Intelligence, CUR … Current Use, DHT … Digital Health Technology, ECO … Economic, EFF … Effectiveness, EHDS … Electronic Health Data Space, ETH … Ethical, EU … European Union, GDPR … General Data Protection Regulation, HTA … Health Technology Assessment, MDR … Medical Device Regulation, ORG … Organisational, SAF … Safety, SOC … Social; TEC … Technical.

Research questions

Description of the technology	
Element ID	**Research question**
A0002	Who manufactures the technology?
F0001	Is the technology a new, innovative mode of care, an add-on to, or modification of a standard mode of care, or a replacement of a standard mode of care?
B0001	What is the technology and the comparator(s)?
B0002	What is the claimed benefit of the technology in relation to the comparator(s)?
B0003	What is the phase of development and implementation of the technology and the comparator(s)?
B0004	Who will be using the technology and the comparator(s)?
B0007	What kind of special premises are needed to use the technology and the comparator(s)?
B0010	What kind of data and records are needed to monitor the use of the technology and the comparator(s)?
B0012	What kind of qualification, training and quality assurance processes are needed for the use or maintenance of the technology and the comparator(s)?
B0013	What is the regulatory status of the technology?

Health problem and Current Use	
Element ID	**Research question**
A0007	What is the target population in this assessment?
A0023	How many people belong to the target population?
A0002	What is the disease or health condition in the scope of this assessment?
A0003	What are the known risk factors for lung cancer?
A0005	What are the symptoms and the burden of lung cancerfor the patient?
A0018	What are the other typical or common alternatives to the current technology
A0024	How is lung cancer currently diagnosed according to published guidelines and in practice?

Clinical and test performance	
Element ID	**Research question**
D0001	What is the expected beneficial effect of the technology on mortality?
D0005	How does telemonitoring affect symptoms and findings (severity, frequency) of lung cancer?
D0032	How does the technology modify the magnitude and frequency of Morbidity?
D0011	What is the effect of the technology on patients' body functions?
D0012	What is the effect of the technology on generic health-related quality of life?
D0013	What is the effect of the technology on disease-specific quality of life?
D1001	What is the accuracy of the test against reference standard?
D1004	What are the requirements for accuracy in the context where the technology will be used?
D1005	What is the optimal threshold value in this context?
D1006	Does the test reliably rule in or rule out the target condition?
D1008	What is known about the intra- and inter-observer variation in test interpretation?
D0020	Does use of the test lead to improved detection of the condition?
D0021	How does use of the test change physicians' management decisions?
D0022	Does the test detect other potential health conditions that can impact the subsequent management decisions?

Safety	
Element ID	**Research question**
C0006	What are the consequences of false positive, false negative and incidental findings generated by using the technology from the viewpoint of patient safety?
C0008	How safe is the technology in relation to the comparator(s)?

Organisational outcomes	
Element ID	**Research question**
E0001	What types of resources are used when delivering the assessed technology and its comparators (resource-use identification)?
D0023	How does the technology modify the need for other technologies and use of resources?
G0001	How does the technology affect the current work processes?
G0003	What kind of process ensures proper education and training of staff?
G0012	In What way is the quality assurance and monitoring system of the new technology organised?
G0006	What are the costs of processes related to acquisition and setting up the new technology?
G0008	What management problems and opportunities are attached to the technology?
G0010	How is the technology accepted?

Search strategy

1. Systematic reviews

Database: Cochrane

Search Name: AI in X-ray analysis	
Search date: 01.08.2025	
ID	Search
#1	MeSH descriptor: [Artificial Intelligence] this term only
#2	("artificial intelligence")
#3	(AI):ti,ab,kw
#4	((artificial OR machine OR deep) NEAR/5 (intelligence OR learning OR reasoning)):ti,ab,kw
#5	[mh "Neural Networks, Computer"]
#6	(("neural" NEXT network*) OR convolutional OR CNN OR CNNs):ti,ab,kw
#7	[mh "Diagnosis, Computer-Assisted"]
#8	[mh "Pattern Recognition, Automated"]
#9	((automat* OR autonomous OR "computer aided" OR "computer assisted") NEAR/3 (detect* OR identif* OR diagnos*)):ti,ab,kw
#10	(("support vector" NEXT machine*) OR ("random" NEXT forest*) OR "black box learning"):ti,ab,kw
#11	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10
#12	[mh "Radiography, Thoracic"]
#13	[mh X-Rays]
#14	(((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
#15	#12 OR #13 OR #14
#16	#11 AND #15
#17	[mh "Lung Neoplasms"]
#18	[mh "Solitary Pulmonary Nodule"]
#19	((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
#20	((pancoast* OR "superior sulcus" OR "pulmonary sulcus") NEAR/4 (tumor* OR tumour* OR syndrome*)):ti,ab,kw
#21	sclc:ti,ab,kw
#22	nsclc:ti,ab,kw
#23	#17 OR #18 OR #19 OR #20 OR #21 OR #22
#24	#11 AND #23
#25	#16 OR #24
#26	#16 OR #24 in Cochrane Reviews, Cochrane Protocols
#27	#26 with Cochrane Library publication date Between Jan 2020 and Aug 2025
Total hits: 4	

Database: Embase

Search Name: AI in X-ray analysis		
Search date: 01.08.2025		
No.	Query Results	Results
#61.	#57 NOT #60	596
#60.	#58 OR #59	533,431
#59.	'clinical trial':dtype	533,362
#58.	#57 AND 'Conference Abstract'/it	69

#57.	#55 NOT #56	676
#56.	'animal experiment'/de NOT ('human experiment'/de OR 'human'/de)	2,767,976
#55.	#53 AND [2020-2025]/py AND ([english]/lim OR [german]/lim)	677
#54.	#53 AND [2020-2025]/py	678
#53.	#50 OR #52	761
#52.	#49 AND #51	716
#51.	'systematic review'/de OR 'systematic review (topic)'/de OR (('comprehensive':ti,ab,kw OR 'mapping':ti,ab,kw OR 'methodology':ti,ab,kw OR 'scoping':ti,ab,kw OR 'systematic':ti,ab,kw) AND ('search':ti,ab,kw OR 'searched':ti,ab,kw OR 'searches':ti,ab,kw OR 'studies':ti,ab,kw) AND ('cinahl':ti,ab,kw OR 'cochrane':ti,ab,kw OR 'embase':ti,ab,kw OR 'psycinfo':ti,ab,kw OR 'pubmed':ti,ab,kw OR 'medline':ti,ab,kw OR 'scopus':ti,ab,kw OR 'web of science':ti,ab,kw OR 'bibliographic review':ti,ab,kw OR 'bibliographic reviews':ti,ab,kw OR 'literature review':ti,ab,kw OR 'literature reviews':ti,ab,kw OR 'literature search':ti,ab,kw OR 'literature searches':ti,ab,kw OR 'qualitative review':ti,ab,kw OR 'qualitative reviews':ti,ab,kw OR 'quantitative review':ti,ab,kw OR 'quantitative reviews':ti,ab,kw)) OR 'comprehensive review':ti,ab,kw OR 'comprehensive reviews':ti,ab,kw OR 'comprehensive search':ti,ab,kw OR 'comprehensive searches':ti,ab,kw OR 'critical review':ti,ab,kw OR 'critical reviews':ti,ab,kw OR (('electronic database':ti,ab,kw OR 'electronic databases':ti,ab,kw OR (databases NEAR/3 searched)) AND (eligibility:ti,ab,kw OR excluded:ti,ab,kw OR exclusion:ti,ab,kw OR included:ti,ab,kw OR inclusion:ti,ab,kw)) OR 'evidence assessment':ti,ab,kw OR 'evidence review':ti,ab,kw OR 'exploratory review':ti,ab,kw OR 'framework synthesis':ti,ab,kw OR 'mapping review':ti,ab,kw OR 'meta-review':ti,ab,kw OR 'meta-synthesis':ti,ab,kw OR 'methodology review':ti,ab,kw OR 'mixed methods review':ti,ab,kw OR 'mixed methods synthesis':ti,ab,kw OR (overview NEAR/4 reviews) OR 'prisma':ab OR ('preferred':ti,ab,kw AND reporting:ti,ab,kw) OR 'prognostic review':ti,ab,kw OR 'psychometric review':ti,ab,kw OR 'rapid evidence assessment':ti,ab,kw OR 'rapid literature review':ti,ab,kw OR 'rapid literature search':ti,ab,kw OR 'rapid realist':ti,ab,kw OR 'rapid review':ti,ab,kw OR 'rapid reviews':ti,ab,kw OR 'realist review':ti,ab,kw OR 'review of reviews':ti,ab,kw OR 'scoping review':ti,ab,kw OR 'scoping reviews':ti,ab,kw OR 'scoping study':ti,ab,kw OR 'systematic evidence map':ti,ab,kw OR 'systematic evidence mapping':ti,ab,kw OR 'systematic literature':ti,ab,kw OR 'systematic medline':ti,ab,kw OR 'systematic pubmed':ti,ab,kw OR 'systematic review':ti,ab,kw OR 'systematic reviews':ti,ab,kw OR 'systematic search':ti,ab,kw OR 'systematic searches':ti,ab,kw OR 'systematical literature review':ti,ab,kw OR 'systematical review':ti,ab,kw OR 'systematical reviews':ti,ab,kw OR 'systematically identified':ti,ab,kw OR 'systematically review':ti,ab,kw OR 'systematically reviewed':ti,ab,kw OR 'umbrella review':ti,ab,kw OR 'umbrella reviews':ti,ab,kw OR '13616137':is OR 'cochrane database of systematic reviews'/jt	854,022
#50.	#49 AND ([cochrane review]/lim OR [systematic review]/lim OR [meta analysis]/lim)	535
#49.	#15 OR #23 OR #24 OR #48	22,384
#48.	#46 OR #47	89
#47.	#22 AND #45	41
#46.	#14 AND #45	73
#45.	#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 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44	249
#44.	veolity*	12
#43.	gleamer*	33
#42.	'visionary health'	5
#41.	'vuno med-chest x-ray*'	1
#40.	sensecare*:dn,tn	2
#39.	'red dot':dn,tn	59
#38.	'red dot'/exp	19
#37.	qxr*:dn,tn	11
#36.	'chesteye quality*'	
#35.	'milvue suite*'	1
#34.	'lunit insight*'	72
#33.	'jld-02k*'	1
#32.	'inferread dr*'	7
#31.	'clearread xray*'	2
#30.	chexvision*	-
#29.	'chest x-ray classifier*'	6
#28.	chestview*	1
#27.	'auto lung nodule detection*'	2
#26.	'annalise cxr*'	2

#25.	'ai-rad companion*'	46
#24.	#10 AND #14 AND #23	1,540
#23.	#10 AND #22	17,056
#22.	#16 OR #17 OR #18 OR #19 OR #20 OR #21	734,235
#21.	nsclc:ta,ab,kw	132,732
#20.	sclc:ta,ab,kw	20,447
#19.	(pancoast* OR 'superior sulcus' OR 'pulmonary sulcus') NEAR/3 (tumo$r* OR syndrome*)	1,701
#18.	(lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) NEAR/2 (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumo$r* OR carcino* OR malignan* OR adeno$carcinom* OR blastoma*)	717,314
#17.	'lung nodule'/de	34,323
#16.	'lung tumor'/mj/exp	380,009
#15.	#10 AND #14	6,863
#14.	#11 OR #12 OR #13	319,213
#13.	((chest OR lung* OR thora*) NEAR/2 (radiograph* OR radiogram* OR radiology OR ro$ntgen* OR 'x ray*' OR xray* OR film*)) OR cxr*	292,400
#12.	'x ray'/mj/exp	22,754
#11.	'thorax radiography'/mj/exp	34,230
#10.	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9	688,631
#9.	'support vector machine*' OR 'random forest*' OR 'black box learning'	101,877
#8.	(automat* OR autonomous OR 'computer aided' OR 'computer assisted') NEAR/2 (detect* OR identif* OR diagnos*)	87,363
#7.	'computer assisted diagnosis'/mj	24,667
#6.	'neural network*' OR convolutional OR cnn OR cnns	208,062
#5.	(artificial OR machine OR deep) NEAR/2 (intelligence OR learning OR reasoning)	420,657
#4.	ai:ti,ab	114,829
#3.	'machine learning'/mj	75,358
#2.	'artificial intelligence'	167,322
#1.	'artificial intelligence'/mj/exp	76,137
Total hits: 596		

Database: Ovid MEDLINE(R)

Search Name: AI in X-ray analysis	
Search date: 31.07.2025	
ID	Search
1	exp Artificial Intelligence/ (245026)
2	artificial intelligence.mp. (105528)
3	AI.mp. (82927)
4	((artificial or machine or deep) adj3 (intelligence or learning or reasoning)).mp. (308299)
5	exp Neural Networks, Computer/ (86820)
6	(neural network* or convolutional or CNN or CNNs).mp. (158236)
7	exp Diagnosis, Computer-Assisted/ (91608)
8	exp Pattern Recognition, Automated/ (26977)
9	((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).mp. (65277)
10	(support vector machine* or random forest* or black box learning).mp. (65247)
11	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 (612102)
12	exp Radiography, Thoracic/ (41445)
13	X-Rays/ (32784)
14	(((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).mp. (93307)
15	12 or 13 or 14 (129734)

16	11 and 15 (5712)
17	exp Lung Neoplasms/ (298724)
18	exp Solitary Pulmonary Nodule/ (5151)
19	((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumo?r* or carcino* or malignan* or adenocarcinom* or blastoma*)).mp. (458525)
20	((pancoast* or superior sulcus or pulmonary sulcus) adj3 (tumo?r* or syndrome*)).mp. (1195)
21	sclc.mp. (11783)
22	nsclc.mp. (72456)
23	17 or 18 or 19 or 20 or 21 or 22 (461758)
24	11 and 23 (10941)
25	11 and 15 and 24 (908)
26	AI-Rad Companion*.mp. (23)
27	Annalise CXR*.mp. (1)
28	Auto Lung Nodule Detection*.mp. (0)
29	ChestView*.mp. (1)
30	Chest X-Ray Classifier*.mp. (5)
31	CheXVision*.mp. (0)
32	ClearRead Xray*.mp. (0)
33	InferRead DR*.mp. (4)
34	JLD-02K*.mp. (0)
35	Lunit INSIGHT*.mp. (31)
36	Milvue Suite*.mp. (1)
37	ChestEye Quality*.mp. (0)
38	qXR*.ti,ab. (98)
39	Qure*.mp. (146)
40	red dot*.mp. (304)
41	SenseCare-Chest DR*.mp. (0)
42	VUNO Med-Chest X-Ray*.mp. (0)
43	Visionary Health.mp. (3)
44	Gleamer*.mp. (12)
45	Veolity*.mp. (2)
46	26 or 27 or 28 or 29 or 30 or 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 (608)
47	15 and 46 (82)
48	23 and 46 (28)
49	47 or 48 (91)
50	16 or 24 or 25 or 49 (15752)
51	limit 50 to (meta analysis or "systematic review") (161)
52	(((comprehensive* or integrative or systematic*) adj3 (bibliographic* or review* or literature)) or (meta-analy* or metaanaly* or "research synthesis" or ((information or data) adj3 synthesis) or (data adj2 extract*))).ti,ab. or (cinahl or (cochrane adj3 trial*) or embase or medline or psyclit or (psycinfo not "psycinfo database") or pubmed or scopus or "sociological abstracts" or "web of science").ab. or ("cochrane database of systematic reviews" or evidence report technology assessment or evidence report technology assessment summary).jn. or Evidence Report: Technology Assessment*.jn. or ((review adj5 (rationale or evidence)).ti,ab. and review.pt.) or meta-analysis as topic/ or Meta-Analysis.pt. (897539)
53	50 and 52 (539)
54	51 or 53 (546)
55	limit 54 to yr="2020 - 2025" (447)
56	limit 55 to (english or german) (445)
57	exp animals/ not humans.sh. (5361400)
58	56 not 57 (443)
59	remove duplicates from 58 (439)
Total hits: 439	

Database: INAHTA

Search Name: AI in X-ray analysis	
Search date: 01.08.2025	
ID #	Search query,"Hits","Searched At"
18	((((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR rontgen* OR x-ray* OR xray* OR film* OR CXR*)) OR ("X-Rays"[mhe]) OR ("Radiography Thoracic"[mhe])) AND (("support vector machine" OR "support vector machines" OR "random forest" OR "random forests" OR "black box learning") OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR ("Pattern Recognition Automated"[mhe]) OR ("Diagnosis Computer-Assisted"[mhe]) OR ("neural network" OR "neural networks" OR convolutional OR CNN OR CNNs) OR ("Neural Networks Computer"[mhe]) OR ((artificial OR machine OR deep) AND (intelligence OR learning OR reasoning)) OR ("artificial intelligence") OR ("Artificial Intelligence"[mhe]))) FROM 2020 TO 2040) AND (English OR German)[Language],"4","2025-08-01T10:42:20.000000Z"
17	(((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR rontgen* OR x-ray* OR xray* OR film* OR CXR*)) OR ("X-Rays"[mhe]) OR ("Radiography Thoracic"[mhe])) AND (("support vector machine" OR "support vector machines" OR "random forest" OR "random forests" OR "black box learning") OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR ("Pattern Recognition Automated"[mhe]) OR ("Diagnosis Computer-Assisted"[mhe]) OR ("neural network" OR "neural networks" OR convolutional OR CNN OR CNNs) OR ("Neural Networks Computer"[mhe]) OR ((artificial OR machine OR deep) AND (intelligence OR learning OR reasoning)) OR ("artificial intelligence") OR ("Artificial Intelligence"[mhe]))) FROM 2020 TO 2040,"5","2025-08-01T10:41:50.000000Z"
16	((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR rontgen* OR x-ray* OR xray* OR film* OR CXR*)) OR ("X-Rays"[mhe]) OR ("Radiography Thoracic"[mhe])) AND (("support vector machine" OR "support vector machines" OR "random forest" OR "random forests" OR "black box learning") OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR ("Pattern Recognition Automated"[mhe]) OR ("Diagnosis Computer-Assisted"[mhe]) OR ("neural network" OR "neural networks" OR convolutional OR CNN OR CNNs) OR ("Neural Networks Computer"[mhe]) OR ((artificial OR machine OR deep) AND (intelligence OR learning OR reasoning)) OR ("artificial intelligence") OR ("Artificial Intelligence"[mhe])),"10","2025-08-01T10:41:23.000000Z"
15	((((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR rontgen* OR x-ray* OR xray* OR film* OR CXR*)) OR ("X-Rays"[mhe]) OR ("Radiography Thoracic"[mhe])) AND (("support vector machine" OR "support vector machines" OR "random forest" OR "random forests" OR "black box learning") OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR ("Pattern Recognition Automated"[mhe]) OR ("Diagnosis Computer-Assisted"[mhe]) OR ("neural network" OR "neural networks" OR convolutional OR CNN OR CNNs) OR ("Neural Networks Computer"[mhe]) OR ((artificial OR machine OR deep) AND (intelligence OR learning OR reasoning)) OR ("artificial intelligence") OR ("Artificial Intelligence"[mhe])),"10","2025-08-01T10:40:12.000000Z"
14	(((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR rontgen* OR x-ray* OR xray* OR film* OR CXR*)) OR ("X-Rays"[mhe]) OR ("Radiography Thoracic"[mhe]),"101","2025-08-01T10:40:02.000000Z"
13	((chest OR lung* OR thora*) AND (radiograph* OR radiogram* OR radiology OR roentgen* OR rontgen* OR x-ray* OR xray* OR film* OR CXR*),"52","2025-08-01T10:39:31.000000Z"
12	"X-Rays"[mhe],"46","2025-08-01T10:37:45.000000Z"
11	"Radiography Thoracic"[mhe],"12","2025-08-01T10:37:23.000000Z"
10	("support vector machine" OR "support vector machines" OR "random forest" OR "random forests" OR "black box learning") OR ((automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*)) OR ("Pattern Recognition Automated"[mhe]) OR ("Diagnosis Computer-Assisted"[mhe]) OR ("neural network" OR "neural networks" OR convolutional OR CNN OR CNNs) OR ("Neural Networks Computer"[mhe]) OR ((artificial OR machine OR deep) AND (intelligence OR learning OR reasoning)) OR ("artificial intelligence") OR ("Artificial Intelligence"[mhe]),"382","2025-08-01T10:36:50.000000Z"
9	"support vector machine" OR "support vector machines" OR "random forest" OR "random forests" OR "black box learning","0","2025-08-01T10:36:35.000000Z"
8	(automat* OR autonomous OR "computer aided" OR "computer assisted") AND (detect* OR identif* OR diagnos*),"202","2025-08-01T10:34:34.000000Z"
7	"Pattern Recognition Automated"[mhe],"1","2025-08-01T10:33:17.000000Z"
6	"Diagnosis Computer-Assisted"[mhe],"80","2025-08-01T10:32:50.000000Z"
5	"neural network" OR "neural networks" OR convolutional OR CNN OR CNNs,"5","2025-08-01T10:32:16.000000Z"
4	"Neural Networks Computer"[mhe],"0","2025-08-01T10:31:21.000000Z"
3	(artificial OR machine OR deep) AND (intelligence OR learning OR reasoning),"44","2025-08-01T10:29:58.000000Z"
2	"artificial intelligence","33","2025-08-01T10:26:32.000000Z"
1	"Artificial Intelligence"[mhe],"135","2025-08-01T10:26:03.000000Z"
Total hits: 4	

2. Trials

Database: Cochrane

Search Name: AI in X-ray analysis	
Search date: 07.08.2025	
ID	Search
#1	MeSH descriptor: [Artificial Intelligence] this term only
#2	("artificial intelligence")
#3	(AI):ti,ab,kw
#4	((artificial OR machine OR deep) NEAR/5 (intelligence OR learning OR reasoning)):ti,ab,kw
#5	[mh "Neural Networks, Computer"]
#6	(("neural" NEXT network*) OR convolutional OR CNN OR CNNs):ti,ab,kw
#7	[mh "Diagnosis, Computer-Assisted"]
#8	[mh "Pattern Recognition, Automated"]
#9	((automat* OR autonomous OR "computer aided" OR "computer assisted") NEAR/3 (detect* OR identif* OR diagnos*)):ti,ab,kw
#10	(("support vector" NEXT machine*) OR ("random" NEXT forest*) OR "black box learning"):ti,ab,kw
#11	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10
#12	[mh "Radiography, Thoracic"]
#13	[mh X-Rays]
#14	(((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
#15	#12 OR #13 OR #14
#16	#11 AND #15
#17	[mh "Lung Neoplasms"]
#18	[mh "Solitary Pulmonary Nodule"]
#19	((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
#20	((pancoast* OR "superior sulcus" OR "pulmonary sulcus") NEAR/4 (tumor* OR tumour* OR syndrome*)):ti,ab,kw
#21	sclc:ti,ab,kw
#22	nsclc:ti,ab,kw
#23	#17 OR #18 OR #19 OR #20 OR #21 OR #22
#24	#11 AND #23
#25	#16 OR #24
#26	#16 OR #24 in Trials
#27	#16 OR #24 with Publication Year from 2024 to 2025, in Trials
#28	(conference proceeding):pt
#29	(abstract):so
#30	(((clinicaltrials OR trialsearch OR ANZCTR OR ensaiosclinicos OR Actrn OR chictr OR cris OR ctri OR registroclinico OR clinicaltrialsregister OR DRKS OR IRCT OR Isrctn OR rctportal OR JapicCTI OR JMACCT OR jRCT OR JPRN OR Nct OR UMIN OR trialregister OR PACTR OR R.B.R.OR REPEC OR SLCTR OR Tcr))):so
#31	#28 OR #29 OR #30
#32	#27 NOT #31
Total hits: 46	

Database: Embase

Search Name: AI in X-ray analysis		
Search date: 07.08.2025		
No.	Query Results	Results
#59.	#55 NOT #58	331
#58.	#56 OR #57	533,481
#57.	'clinical trial':dtype	533,362
#56.	#55 AND 'conference abstract'/it	119
#55.	#54 AND ([english]/lim OR [german]/lim)	550
#54.	#53 AND [2024-2025]/py	559
#53.	#50 OR #52	1,943
#52.	#49 AND #51	1,921
#51.	'clinical trial'/de OR 'randomized controlled trial'/de OR 'randomization'/de OR 'single blind procedure'/de OR 'double blind procedure'/de OR 'crossover procedure'/de OR 'placebo'/de OR 'prospective study'/de OR ('randomi?ed controlled' NEXT/1 trial*) OR rct OR 'randomly allocated' OR 'allocated randomly' OR 'random allocation' OR (allocated NEAR/2 random) OR (single NEXT/1 blind*) OR (double NEXT/1 blind*) OR ((treble OR triple) NEAR/1 blind*) OR placebo*	3,761,627
#50.	#49 AND ([controlled clinical trial]/lim OR [randomized controlled trial]/lim)	465
#49.	#15 OR #23 OR #24 OR #48	22,557
#48.	#46 OR #47	89
#47.	#22 AND #45	41
#46.	#14 AND #45	73
#45.	#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 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44	249
#44.	veolity*	12
#43.	gleamer*	33
#42.	'visionary health'	5
#41.	'vuno med-chest x-ray*'	1
#40.	sensecare*:dn,tn	2
#39.	'red dot':dn,tn	59
#38.	'red dot'/exp	19
#37.	qxr*:dn,tn	11
#36.	'chesteye quality*'	3
#35.	'milvue suite*'	1
#34.	'lunit insight*'	72
#33.	'jld-02k*'	1
#32.	'inferread dr*'	7
#31.	'clearread xray*'	2
#30.	chexvision*	-
#29.	'chest x-ray classifier*'	6
#28.	chestview*	1
#27.	'auto lung nodule detection*'	2
#26.	'annalise cxr*'	2
#25.	'ai-rad companion*'	46
#24.	#10 AND #14 AND #23	1,547
#23.	#10 AND #22	17,207
#22.	#16 OR #17 OR #18 OR #19 OR #20 OR #21	736,036
#21.	nsclc:ta,ab,kw	133,468
#20.	sclc:ta,ab,kw	20,568

#19.	(pancoast* OR 'superior sulcus' OR 'pulmonary sulcus') NEAR/3 (tumo$r* OR syndrome*)	1,703
#18.	(lung OR lungs OR pulmon* OR intrapulmon* OR bronch*) NEAR/2 (abnormal* OR nodul* OR lesion* OR mass OR masses OR cancer* OR neoplas* OR tumo$r* OR carcino* OR malignan* OR adeno$carcinom* OR blastoma*)	719,098
#17.	'lung nodule'/de	34,392
#16.	'lung tumor'/mj/exp	381,118
#15.	#10 AND #14	6,892
#14.	#11 OR #12 OR #13	319,586
#13.	((chest OR lung* OR thora*) NEAR/2 (radiograph* OR radiogram* OR radiology OR ro$ntgen* OR 'x ray*' OR xray* OR film*)) OR cxr*	292,759
#12.	'x ray'/mj/exp	22,771
#11.	'thorax radiography'/mj/exp	34,252
#10.	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9	691,831
#9.	'support vector machine*' OR 'random forest*' OR 'black box learning'	102,355
#8.	(automat* OR autonomous OR 'computer aided' OR 'computer assisted') NEAR/2 (detect* OR identif* OR diagnos*)	87,468
#7.	'computer assisted diagnosis'/mj	24,677
#6.	'neural network*' OR convolutional OR cnn OR cnns	208,782
#5.	(artificial OR machine OR deep) NEAR/2 (intelligence OR learning OR reasoning)	423,234
#4.	ai:ti,ab	115,788
#3.	'machine learning'/mj	75,840
#2.	'artificial intelligence'	168,578
#1.	'artificial intelligence'/mj/exp	76,711
Total hits: 331		

Database: Ovid MEDLINE(R)

Search Name: AI in X-ray analysis	
Search date: 07.08.2025	
ID	Search
1	exp Artificial Intelligence/ (245756)
2	artificial intelligence.mp. (106028)
3	AI.mp. (83333)
4	((artificial or machine or deep) adj3 (intelligence or learning or reasoning)).mp. (309698)
5	exp Neural Networks, Computer/ (87173)
6	(neural network* or convolutional or CNN or CNNs).mp. (158653)
7	exp Diagnosis, Computer-Assisted/ (91620)
8	exp Pattern Recognition, Automated/ (26978)
9	((automat* or autonomous or computer aided or computer assisted) adj3 (detect* or identif* or diagnos*)).mp. (65344)
10	(support vector machine* or random forest* or black box learning).mp. (65451)
11	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 (613876)
12	exp Radiography, Thoracic/ (41443)
13	X-Rays/ (32786)
14	(((chest or lung* or thora*) adj3 (radiograph* or radiogram* or radiology or roentgen* or x-ray* or xray* or film*)) or CXR*).mp. (93333)
15	12 or 13 or 14 (129762)
16	11 and 15 (5709)
17	exp Lung Neoplasms/ (298881)
18	exp Solitary Pulmonary Nodule/ (5153)
19	((lung or lungs or pulmon* or intrapulmon* or bronch*) adj3 (abnormal* or nodul* or lesion* or mass or masses or cancer* or neoplas* or tumo?r* or carcino* or malignan* or adenocarcinom* or blastoma*)).mp. (458872)

20	((pancoast* or superior sulcus or pulmonary sulcus) adj3 (tumo?r* or syndrome*)).mp. (1195)
21	sclc.mp. (11793)
22	nsclc.mp. (72534)
23	17 or 18 or 19 or 20 or 21 or 22 (462114)
24	11 and 23 (10986)
25	11 and 15 and 24 (908)
26	AI-Rad Companion*.mp. (23)
27	Annalise CXR*.mp. (1)
28	Auto Lung Nodule Detection*.mp. (0)
29	ChestView*.mp. (1)
30	Chest X-Ray Classifier*.mp. (5)
31	CheXVision*.mp. (0)
32	ClearRead Xray*.mp. (0)
33	InferRead DR*.mp. (4)
34	JLD-02K*.mp. (0)
35	Lunit INSIGHT*.mp. (30)
36	Milvue Suite*.mp. (1)
37	ChestEye Quality*.mp. (0)
38	qXR*.ti,ab. (98)
39	Qure*.mp. (146)
40	red dot*.mp. (306)
41	SenseCare-Chest DR*.mp. (0)
42	VUNO Med-Chest X-Ray*.mp. (0)
43	Visionary Health.mp. (3)
44	Gleamer*.mp. (12)
45	Veolity*.mp. (2)
46	26 or 27 or 28 or 29 or 30 or 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 (609)
47	15 and 46 (81)
48	23 and 46 (28)
49	47 or 48 (90)
50	16 or 24 or 25 or 49 (15794)
51	limit 50 to clinical trial, all (297)
52	((randomized controlled trial or controlled clinical trial).pt. or randomized.ab. or placebo.ab. or clinical trials as topic.sh. or randomly.ab. or trial.ti.) not (exp animals/ not humans.sh.) (1571076)
53	50 and 52 (830)
54	51 or 53 (951)
55	limit 54 to yr="2024 - 2025" (248)
56	limit 55 to (english or german) (245)
57	remove duplicates from 56 (244)
Total hits: 244	

HTA Austria
Austrian Institute for
Health Technology Assessment
GmbH