

Zehn Jahre nach der Prognose vom Ende der Radiologie fehlen Radiologen

Zehn Jahre nach der Prognose vom Ende der Radiologie ist das Fach gewachsen. Die stärkste Evidenz stützt den assistierenden Einsatz von KI im Screening. Eine Einordnung der Evidenzlage.

ZUSAMMENFASSUNG

Hintergrund. 2016 wurde prognostiziert, künstliche Intelligenz werde den Radiologen überflüssig machen. Zehn Jahre später ist das Fach gewachsen, und in mehreren Ländern fehlen Radiologen. Zu klären ist, was die künstliche Intelligenz im radiologischen Alltag belegbar leistet.

Evidenz. Die stärkste Evidenz stammt aus dem Mammografie-Screening. In der randomisierten MASAI-Studie mit 105.934 Frauen senkte die KI-Assistenz die Leselast um 44 Prozent und entdeckte 29 Prozent mehr Karzinome ohne zusätzliche Fehlalarme. Die Intervallkarzinome sanken um 12 Prozent. An der Mayo Clinic stieg die Zahl der Radiologen seit 2016 um 55 Prozent, begleitet von mehr als 250 eingesetzten Algorithmen.

Offen. Die belegte Wirkung betrifft die assistierte Befundung, nicht den Ersatz des Arztes. Außerhalb des Screenings hoben KI-Vorschläge die diagnostische Leistung im Mittel nicht. Prognosen sehen die Nachfrage nach Bildgebung rascher steigen als das Angebot an Radiologen.

Schlussfolgerung. Belegt ist ein assistierender Einsatz: Die künstliche Intelligenz übernimmt definierte Aufgaben, der Arzt verantwortet Befund und Konsequenz. Ein Ersatz der ärztlichen Rolle ist weder eingetreten noch durch die Studienlage gedeckt.

Dr. med. Christian Mauch · Facharzt für Orthopädie und Unfallchirurgie · SummitGO
· Lesezeit etwa 8 Minuten

+55 %

Radiologen der Mayo Clinic seit
2016

-44 %

Leselast der Ärzte (MASAI)

+29 %

entdeckte Karzinome (MASAI)

Der Testerfolg und seine Grenzen

Die Annahme stützte sich auf Ergebnisse, die für sich genommen stark waren. 2020 veröffentlichte ein Team von Google Health in Nature eine Auswertung,

in der ein KI-System Mammografien las.³ In einer Vergleichsstudie mit sechs Radiologen übertraf das System jeden einzelnen. Gemessen an der Fläche unter der ROC-Kurve, dem Standardmaß für die Trennschärfe eines Tests zwischen Kranken und Gesunden, lag es um 11,5 Prozentpunkte über dem Durchschnitt der Ärzte. Der Vorsprung war groß. Nachprüfen ließ er sich nicht. Programmcode und vollständige Trainingsdaten waren nicht öffentlich zugänglich, eine unabhängige Reproduktion der Arbeit blieb damit unmöglich. Eine Gruppe um Benjamin Haibe-Kains kritisierte in Nature vor allem diese fehlende Transparenz und Reproduzierbarkeit.⁴ Die Autoren ergänzten später in einem Addendum genauere Angaben zum Training, ohne Code und Daten offenzulegen.³ Hinzu kommt die Anlage der Vergleichsstudie. Die sechs Radiologen lasen einen für die Auswertung zusammengestellten Testsatz mit erhöhtem Karzinomanteil. Ein solcher Aufbau prüft die Trennschärfe unter günstigen Bedingungen und sagt wenig über die Leistung im Alltagsstrom eines Screening-Programms, in dem die allermeisten Aufnahmen unauffällig sind.

Ein zweiter Einwand betrifft den Vergleich selbst. In der US-Teilstichprobe der Google-Studie lag die Rate übersehener Karzinome bei den Radiologen deutlich höher als in den anerkannten Vergleichswerten der digitalen Screening-Mammografie. Ein Kommentar in European Radiology zog daraus den Schluss, der Vorsprung des Systems könne teils auf eine ungewöhnlich schwache Vergleichsleistung der Ärzte zurückgehen, weshalb die Studie kein gültiger Kopf-an-Kopf-Vergleich sei.⁵ Die Zahl von 11,5 Prozentpunkten darf deshalb nicht als allgemeiner Leistungsabstand zwischen KI und Radiologen gelesen werden. Sie beschreibt eine Auswertung unter kuratierten Bedingungen, die sich nicht unabhängig überprüfen ließ.

Wo die KI im Alltag geprüft wurde

Belastbarer ist die Evidenz dort, wo KI im laufenden Screening geprüft wurde. Die schwedische MASAI-Studie randomisierte über 100.000 Frauen auf zwei Wege: die eine Hälfte der Aufnahmen wurde mit KI-Unterstützung gelesen, die andere nach dem üblichen Vier-Augen-Prinzip zweier Radiologen. Die 2023 im Lancet Oncology veröffentlichte Sicherheitsanalyse zeigte eine um rund 44 Prozent geringere Leselast der Ärzte.⁶ Eine weitere Auswertung in Lancet Digital Health 2025 fand 29 Prozent mehr entdeckte Karzinome, ohne dass die Zahl der Fehllarme stieg.⁷ Die 2026 veröffentlichten Endergebnisse ergänzten einen weiteren Prüfstein: Die Rate der Intervallkarzinome, also der zwischen zwei Screening-Runden diagnostizierten Tumoren, betrug in der KI-Gruppe 1,55 gegenüber 1,76 je 1.000 Frauen und lag damit numerisch um 12 Prozent niedriger.⁸ Die Studie war auf den Nachweis der Nichtunterlegenheit ausgelegt, nicht auf den einer statistisch gesicherten Überlegenheit. Sie zeigt belastbar, dass die höhere Entdeckungsrate nicht mit einer höheren Intervallkarzinomrate erkaufte wurde. Ob KI diese Tumoren tatsächlich verringert, muss weiter untersucht werden.

Als randomisierte, prospektive Studie ist MASAI die aussagekräftigste Arbeit zum Thema. Ihr wichtigstes Merkmal steht quer zur Ersetzungsthese: Das System las nicht allein. Es bewertete das Risiko, markierte Auffälligkeiten und steuerte, ob eine oder zwei ärztliche Lesungen erfolgten. Die Befundentscheidung blieb in den ärztlich verantworteten Screening-Ablauf eingebettet.⁸ Der Gewinn an Treffsicherheit und Effizienz entstand aus einer Anordnung, die den Arzt im Ablauf hielt. Die stärkste klinische Evidenz betrifft damit einen ärztlich eingebetteten Einsatz. Ein autonomer war es nicht.

Das Paradox des Arbeitsmarkts

Der Arbeitsmarkt bestätigt dieses Bild. Ein Fach, das automatisiert, stellte mehr Ärzte ein. Die Mayo Clinic hat ihren Bestand aufgestockt, während sie ihre KI-Modelle ausbaute.² Die Ursache liegt in einer Schere zwischen Angebot und Nachfrage. Eine Projektion des Neiman Health Policy Institute im Journal of the American College of Radiology erwartet, dass die Zahl der US-Radiologen bis 2055 je nach Zahl der Weiterbildungsstellen um 25,7 bis 40,3 Prozent gegenüber 2023 steigt.¹ Eine begleitende Projektion derselben Forschungsgruppe erwartet je nach Modalität einen Anstieg der Bildgebungsleistungen um 16,9 bis 26,9 Prozent bis 2055.⁹ Angebot und Nachfrage wachsen in vergleichbarer Größenordnung, weshalb der Mangel fortbesteht. In dieser Lage nimmt KI dem Fach Arbeit ab, ohne es überflüssig zu machen.

Ein Arbeitspapier des National Bureau of Economic Research verschiebt den Blick auf eine andere Frage. Die Ökonomen um Nikhil Agarwal untersuchten, wie Radiologen ihre eigene Einschätzung mit der Empfehlung eines KI-Systems verbinden.¹⁰ Die Assistenz führte nicht bei allen Ärzten zu derselben Verbesserung. Ein Teil des möglichen Gewinns ging verloren, weil die Radiologen die Zuverlässigkeit der KI und den Zusammenhang zwischen beiden Informationsquellen nicht angemessen gewichteten. Das Problem ist nicht das blinde Vertrauen, und es ist auch nicht die Ablehnung des Systems. Es kommt darauf an, ob der Arzt weiß, wann die Empfehlung zusätzliche Information liefert und wann sie nur die eigene Einschätzung wiederholt. Die übertragbare Lehre lautet deshalb, dass eine hohe Modellgüte allein nicht genügt. Auch die Schnittstelle, die Schulung und die Regeln für den Umgang mit abweichenden Ergebnissen bestimmen die klinische Qualität.

DIE EVIDENZ IM ÜBERBLICK

Arbeit	Design	Kernbefund	Belastbarkeit
McKinney, Nature 2020	Reader-Studie, kuratierter Testsatz mit erhöhtem Karzinomanteil,	KI 11,5 Prozentpunkte über dem Durchschnitt der Ärzte (ROC-Fläche)	Nicht unabhängig reproduzierbar, Code und Daten unter Verschluss, Vergleichsarm

Arbeit	Design	Kernbefund	Belastbarkeit
	sechs Radiologen		umstritten
MASAI, Lancet 2023-2026	Randomisiert, prospektiv, über 100.000 Frauen im laufenden Screening	44 Prozent weniger Leselast, 29 Prozent mehr entdeckte Karzinome ohne mehr Fehlalarme, Intervallkarzinomrate nicht unterlegen	Stärkste Evidenz, gilt für den ärztlich eingebetteten Einsatz
NBER 31422, 2023	Experiment zur Verbindung von ärztlicher Einschätzung und KI-Empfehlung	Assistenz verbessert nicht bei allen Ärzten, entscheidend ist die Gewichtung der Empfehlung	Zeigt Schnittstelle und Schulung als eigene Qualitätsfaktoren

Die rechtliche Schwelle

Von der Leistungsfrage ist die regulatorische und haftungsrechtliche Einordnung zu trennen. Software fällt unter die Verordnung (EU) 2017/745, wenn sie nach ihrer vom Hersteller festgelegten Zweckbestimmung für einen medizinischen Zweck eingesetzt wird, etwa zur Diagnose oder zur Unterstützung diagnostischer Entscheidungen.¹¹ Bevor ein solches System auf dem europäischen Markt bereitgestellt wird, muss es das für seine Risikoklasse vorgesehene Konformitätsbewertungsverfahren durchlaufen und eine CE-Kennzeichnung tragen. Dazu gehört eine klinische Bewertung seiner Sicherheit und Leistung innerhalb der vorgesehenen Anwendung.

Die CE-Kennzeichnung beantwortet jedoch nicht abschließend, wer für einen konkreten Diagnosefehler haftet. Maßgeblich sind unter anderem die Ursache des Fehlers, die Zweckbestimmung und die Gebrauchsanweisung, die technische und organisatorische Einbindung sowie das Verhalten von Hersteller, Betreiber und behandelndem Arzt. Eine ärztliche Prüfung und ihre nachvollziehbare Dokumentation bleiben deshalb praktisch bedeutsam. Sie dürfen nicht zu einer formalen Bestätigung eines maschinellen Vorschlags verkürzt werden. Wer eine ärztliche Freigabe vorsieht, muss organisatorisch sicherstellen, dass für eine eigenständige Prüfung ausreichend Information, Zeit und fachliche Kompetenz vorhanden sind.

Ärztliche Einordnung

Der realistische Entwicklungspfad zeichnet sich in der Evidenz bereits ab. Die überzeugendsten Ergebnisse finden sich bei klar umrissenen, standardisierten Aufgaben mit großen Datenmengen. Das Screening der Mammografie ist der am besten belegte Fall. In solchen Aufgaben hebt ein KI-Vorleser die Entdeckungsrate und senkt die Leselast, während der Befund eine ärztliche Aussage bleibt. Automatisiert werden einzelne Arbeitsschritte, nicht das gesamte Fach. Offen bleibt, welche Aufgaben sich sicher übertragen lassen und wie der Alltag verhindert, dass aus der Entlastung eine unkritische Übernahme wird. An dieser Stelle endet die belastbare Evidenz. Sie benennt das Problem genauer, als sie es löst.

Die SummitGO-Perspektive

Die Radiologie gehört zu den frühesten und am gründlichsten untersuchten Anwendungsfeldern klinischer KI, und darin liegt ihr Wert über das Fach hinaus. Das Muster wiederholt sich in anderen Gebieten: Die Leistung unter Testbedingungen überzeichnet die klinische Reife, und der dauerhafte Nutzen entsteht aus Anordnungen, die den Arzt in der Verantwortung halten. SummitGO ordnet diese Entwicklung als Herausgeber ein und teilt die Grundhaltung, die sich aus der Radiologie ablesen lässt: Die Maschine unterstützt, der Arzt urteilt. Dieselbe Haltung, das System auf die Zuarbeit zu begrenzen und die Verantwortung beim Arzt zu belassen, prägt die Entwicklung ärztlich freigegebener KI-Anwendungen über die Bildgebung hinaus. Der klinische Gewinn entsteht aus der Verbindung technischer Leistung mit ärztlicher Beurteilung.

Fazit

Belegt ist, dass die Radiologie zehn Jahre nach ihrer angekündigten Abschaffung gewachsen ist und dass die stärkste Evidenz, eine randomisierte Studie, den assistierenden Einsatz von KI im Screening stützt. Die Behauptung einer eigenständigen Überlegenheit der Maschine ruht dagegen auf schwächeren, teils nicht reproduzierbaren Arbeiten. Offen bleiben die schwierigen Fragen: welche Aufgaben sich sicher delegieren lassen, wie die Verantwortung im Schadensfall zugeordnet wird und wie Ärzte die Empfehlung eines Systems richtig gewichten, damit der mögliche Gewinn nicht verloren geht. Die Prognose vom Ende des Fachs hat sich nicht bestätigt. An ihre Stelle tritt eine genauere und schwierigere Frage nach der richtigen Arbeitsteilung zwischen Arzt und Maschine.

QUELLEN

1. Christensen EW, Drake AR, Parikh JR, et al. Projected US radiologist supply, 2025 to 2055. *J Am Coll Radiol.* 2025;22. doi.org/10.1016/j.jacr.2024.10.019
2. Lohr S. Your A.I. radiologist will not be with you soon. *The New York Times*, 14. Mai 2025. nytimes.com
3. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature.* 2020;577:89-94 (mit Addendum, *Nature* 2020;586:E19). doi.org/10.1038/s41586-019-1799-6
4. Haibe-Kains B, Adam GA, Hosny A, et al. Transparency and reproducibility in artificial intelligence. *Nature.* 2020;586:E14-E16. doi.org/10.1038/s41586-020-2766-y
5. Autier P, Burrion JB, Grivegnée AR. AI for reading screening mammograms: the need for circumspection. *Eur Radiol.* 2020;30:4783-4784. doi.org/10.1007/s00330-020-06833-6
6. Lång K, Josefsson V, Larsson A-M, et al. Artificial intelligence-supported screen reading versus standard double reading in the MASAI trial: a clinical safety analysis. *Lancet Oncol.* 2023;24:936-944. doi.org/10.1016/S1470-2045(23)00298-X
7. Hernström V, Josefsson V, Sartor H, et al. Screening performance and characteristics of breast cancer detected in the MASAI trial. *Lancet Digit Health.* 2025;7:e175-e183. doi.org/10.1016/S2589-7500(24)00267-X
8. Gommers J, Hernström V, Josefsson V, et al. Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study. *Lancet.* 2026;407:505-514. doi.org/10.1016/S0140-6736(25)02464-X
9. Christensen EW, Drake AR, Parikh JR, et al. Projected US imaging utilization, 2025 to 2055. *J Am Coll Radiol.* 2025;22:151-158. doi.org/10.1016/j.jacr.2024.10.017
10. Agarwal N, Moehring A, Rajpurkar P, Salz T. Combining human expertise with artificial intelligence: experimental evidence from radiology. NBER Working Paper 31422. 2023 (rev. 2025). nber.org/papers/w31422
11. Verordnung (EU) 2017/745 (Medizinprodukteverordnung). eur-lex.europa.eu/eli/reg/2017/745/oj

Eine erste, kürzere Fassung dieses Themas erschien im Newsletter „KI und Medizin“, Ausgabe 3.

IMPRESSUM

Summitgo GmbH & Co. KG, Relenbergstraße 65, 70174 Stuttgart. Komplementärin Dr. Mauch Beteiligung GmbH (AG Stuttgart HRB 805274), Geschäftsführer Dr. Christian Mauch. Registergericht AG Stuttgart, HRA 743417. USt-IdNr. nach § 27a UStG: [folgt]. Verantwortlich nach § 18 Abs. 2 MStV: Dr. Christian Mauch. Interessenkonflikt: SummitGO entwickelt Software zur ärztlich verantworteten KI-gestützten Dokumentation. Dieser Beitrag stellt kein Produkt dar. © 2026 SummitGO.