

Gefördert durch

STAATSMINISTERIUM FÜR
SOZIALES, GESUNDHEIT UND
GESELLSCHAFTLICHEN ZUSAMMENHALT

KI als Entscheidungsunterstützung im Brustkrebs-Früherkennungs-Programm

Sprechernotizen – Folie 01: Titelfolie

Sehr geehrte Damen und Herren, liebe Kolleginnen und Kollegen,

bevor mein Kollege Dr. Hamm Ihnen gleich die konkreten Auswirkungen von KI auf die Qualitätsindikatoren im Chemnitzer Screening zeigt, möchte ich die Grundlagen legen: Was steckt hinter den Begriffen Künstliche Intelligenz und maschinelles Lernen? Warum ist gerade das Mammographie-Screening das stärkste Evidenzfeld? Und wo liegen die Grenzen?

Warum KI im Mammascreeing?

> 10 Mio.

Einzelbilder jährlich im
deutschen MSP

3,1 Mio.

untersuchte Frauen
im Jahr 2023

19.439

entdeckte Karzinome
im MSP 2023

- Doppelbefundung als Qualitätsstandard bindet erhebliche radiologische Kapazität
- Steigende Untersuchungszahlen bei begrenztem Fachpersonal
- Befundungsvariabilität: Inter- und intraindividuelle Unterschiede
- Ziel: Konsistente Befundung unabhängig von Tageszeit und Arbeitsbelastung

Sprechernotizen – Folie 02: Warum KI im Mammascreeing?

Warum ist das Thema jetzt relevant? Die Zahlen des deutschen Mammographie-Screening-Programms verdeutlichen die Dimension: Im Jahr 2023 wurden über 3,1 Millionen Frauen untersucht, 24 Millionen Einzelbilder befundet, knapp 19.500 Karzinome entdeckt. Diese Zahlen stammen aus dem aktuellen Jahresbericht Evaluation 2023.

Die Doppelbefundung, unser europäischer Qualitätsstandard, bindet erhebliche radiologische Kapazität. Gleichzeitig steigen die Untersuchungszahlen bei begrenztem Fachpersonal. Und wir wissen, dass die Befundungsvariabilität ein reales Problem ist.

KI wird dort relevant, wo Volumen, Standardisierung und Qualitätsanspruch gleichzeitig hoch sind.

KI, maschinelles Lernen, Deep Learning

Künstliche Intelligenz

Oberbegriff: Systeme mit kognitiven Fähigkeiten

Maschinelles Lernen

Lernt Muster aus Daten statt expliziter Programmierung

Deep Learning / CNN

Neuronale Netze mit vielen Schichten;
CNNs als Kernarchitektur für Bilddaten

Wichtige Klarstellung

Die KI im Mammascreening ist **bildanalytische Mustererkennung mit Deep Learning**.

Das hat nichts mit ChatGPT oder Large Language Models zu tun.

- CNNs (Convolutional neural networks) analysieren Bilddaten (Mammographien).
- LLMs verarbeiten Textdaten (Sprache).
- Unterschiedliche Architekturen, Daten, Regulierung.

Sprechernotizen – Folie 03: KI, maschinelles Lernen, Deep Learning

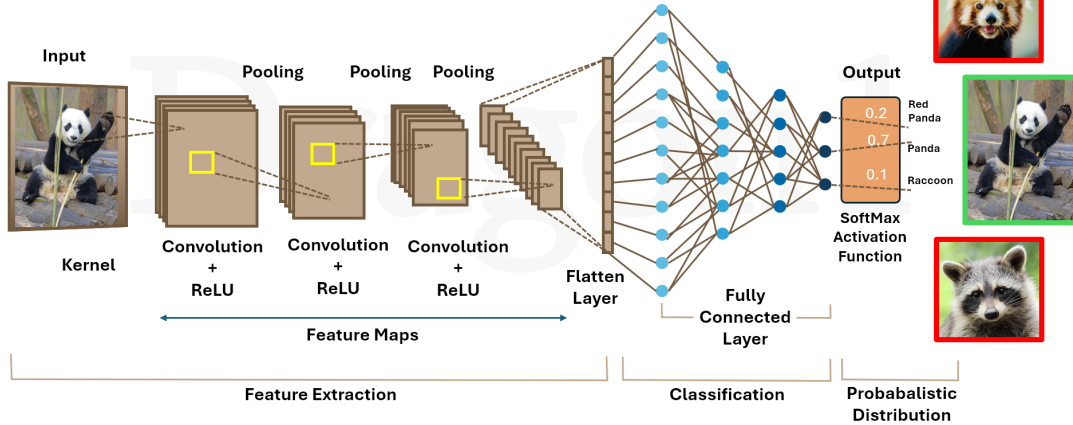
Beginnen wir mit den Begriffen – und einer wichtigen Klarstellung.

Künstliche Intelligenz ist der Oberbegriff für Systeme mit kognitiven Fähigkeiten. Maschinelles Lernen ist ein Teilgebiet – hier lernt das System Muster aus Daten, statt dass jede Regel explizit programmiert wird. Deep Learning nutzt neuronale Netze mit vielen Schichten. Für die Bildanalyse sind CNNs die entscheidende Architektur – sie extrahieren relevante Bildmerkmale automatisch.

Die Klarstellung: Die KI, über die wir heute sprechen, ist bildanalytische Mustererkennung. Das hat nichts mit ChatGPT zu tun. CNNs analysieren Bilddaten, LLMs verarbeiten Textdaten. Unterschiedliche Architekturen, unterschiedliche Regulierung. Wenn Patientinnen fragen, ob jetzt die KI die Mammographie übernimmt, sollten wir klar sagen können, wovon wir sprechen.

Convolutional Neural Networks (AI Deep Learning)

D

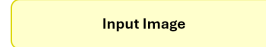


© Copyright 2025, Dragon1, www.dragon1.com

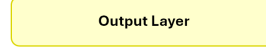
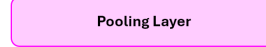
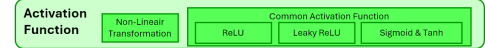
Dragon1 – Conceptual Architecture CNN (Convolutional Neural Network)

D

Core Building Blocks



Core Sub-Blocks

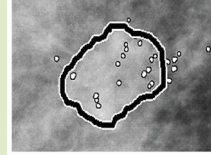
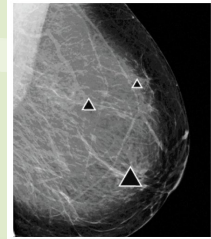
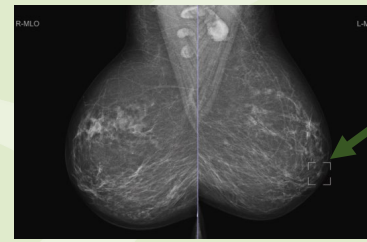


Sprechernotizen – Folie 04: KI, maschinelles Lernen, Deep Learning

CNNs sind eine spezielle Form neuronaler Netze, die vor allem für Bilddaten eingesetzt werden. Sie können visuelle Merkmale wie Kanten, Muster, Texturen und Formen selbstständig erkennen und schrittweise zu komplexeren Bildinformationen zusammensetzen. Dadurch eignen sie sich besonders gut für Anwendungen wie Klassifikation, Objekterkennung und Segmentierung in der medizinischen Bildanalyse.

Waschbär = Racoon

Kleiner Panda: Red Panda (*Ailurus fulgens*) ist ein etwa katzen großes, rotbraunes Säugetier aus den Bergwäldern des Himalaya und Südwest-Chinas. Er lebt vor allem in Bäumen, ernährt sich überwiegend von Bambus und gilt laut IUCN als gefährdet mit weniger als 10 000 erwachsenen Tieren. Daher kann es sein, dass er bald nicht mehr als zu diskriminierendes Merkmal vorkommen wird.



<https://www.hugoboss.com/en-us/products/mimgsticker-2d-cad-technology>

Von CAD zu Deep Learning – Mammographie als Pionierfeld

1998

Erstes CAD-System
(R2 ImageChecker);
FDA-Zulassung

~2003

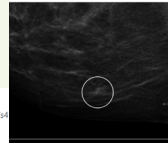
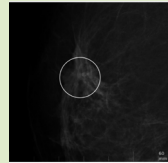
Medicare-Vergütung;
90 % der US-Einrich-
tungen nutzen CAD

~2016

Deep-Learning-basierte
Systeme lösen
klassisches CAD ab

2021–25

Prospektive Studien:
MASAI, PRAIM,
ScreenTrustCAD



Paradigmenwechsel: Klassisches CAD hatte hohe Falsch-positiv-Raten und begrenzten Nettonutzen. Deep-Learning-Systeme zeigen signifikant höhere Sensitivität (ca. 1 zusätzlicher Brustkrebs pro 1.000 Frauen) bei stabiler Spezifität – und liefern erstmals prospektive Evidenz auf Bevölkerungsniveau.

Observational Study | Nat Med. 2025 Mar;31(3):917-924. doi: 10.1038/s41591-025-01000-1. Epub 2025 Jan 7.

Nationwide real-world implementation of AI for cancer detection in population-based mammography screening

Nora Esemann ¹, Stefan Bunk ², Tamas Mukama ³, Hannah Baltus ¹, Susanne A. Elsner ¹, Timo Gomille ⁴, Gerold Hecht ⁵, Sylvia Heywang-Köbrunner ⁶, Regine Rathmann ⁷, Katja Siegmann-Luz ⁸, Thilo Töllner ⁹, Toni Werner Vomweg ¹⁰, Christian Leibig ³, Alexander Katalinic ¹¹

Sprechernotizen – Folie 05: Von CAD zu Deep Learning

Die Mammographie ist historisch das Pionierfeld der radiologischen KI. 1998 wurde das erste CAD-System zugelassen – der R2 ImageChecker. Ab 2003 explodierten die Nutzungszahlen in den USA nach Einführung der Medicare-Vergütung.

Allerdings waren die Ergebnisse des klassischen CAD ernüchternd: hohe Falsch-positiv-Raten und ein fraglicher Nettonutzen. Der Durchbruch kam ab etwa 2016 mit Deep Learning – CNNs lernen Bildmerkmale direkt aus den Daten, ohne manuelle Feature-Definition. Die Ergebnisse sind qualitativ andere: deutlich höhere Sensitivität bei stabiler oder verbesserter Spezifität.

Seit 2021 haben wir erstmals prospektive, bevölkerungsbasierte Studien – MASAI, PRAIM, ScreenTrustCAD –, die diese Systeme unter Realbedingungen testen.

Aktuelle Evidenz – prospektive Screening-Studien

Studie	Design / n	Detektionsrate	Arbeitslast
PRAIM (Nat. Med. 2025)	Prospektiv 460.000 Frauen 12 Standorte (DE)	+18 % (6,7 vs. 5,7/1.000)	Potenzial: bis zu 50 % (modelliert)
MASAI (Lancet Oncol.)	RCT 80.033 Frauen 4 Standorte (SE)	+20 % (6,1 vs. 5,1/1.000) Zwischenauswertung 2023	-44 %
ScreenTrustCAD (ECR/RSNA)	Prospektiv, Real-World 55.000+ Frauen Stockholm (SE)	+4 % (non-inferior)	1 Radiologe statt 2

Kernbefund: Ca. 1 zusätzlicher Brustkrebs pro 1.000 Frauen bei stabilen Recall-Raten und erheblicher Arbeitsentlastung.

Sprechernotizen – Folie 06: Aktuelle Evidenz

Die aktuelle Evidenz kommt aus drei prospektiven Studien – die Details für Chemnitz vertieft Dr. Hamm.

Bevor ich die Zahlen zeige, drei Erläuterungen zur Lesart der Tabelle:

Detektionsrate: Wenn ich hier +18 % oder +20 % zeige, dann heißt das: Wir vergleichen mit vs. ohne KI. Praktisch geht es um entdeckte Karzinome pro 1.000 Frauen – zum Beispiel 5,7 vs. 6,7 pro 1.000, also etwa ein zusätzlicher Brustkrebs pro 1.000 gescreente Frauen.

Arbeitslast: Die –44 % Befundungsaufwand beziehen sich auf die Anzahl der Lesungen durch Radiologinnen und Radiologen. In den Studien lesen im Standardarm zwei Radiologen, im KI-Arm reicht ein Radiologe plus KI oder die KI triagiert unauffällige Fälle. Dadurch sinkt die Zahl der menschlichen Lesungen um rund 40–45 %.

Und wichtig zum Vergleichsarm: Die KI wird nicht gegen irgendwelche alten Forschungsdaten getestet, sondern gegen den aktuellen klinischen Standard – in Europa typischerweise die klassische Doppelbefundung.

Prospektiv heißt: Die KI war im laufenden Screeningbetrieb aktiv. Die Befunde wurden in Echtzeit angezeigt und haben die Recall-Entscheidungen beeinflusst – also keine nachträgliche Labor-Auswertung, sondern Alltagspraxis unter Studienbedingungen.

Nun zu den einzelnen Studien:

PRAIM, publiziert 2025 in Nature Medicine: über 460.000 Frauen an zwölf deutschen Standorten. Eingesetztes System: Vara von Merantix Healthcare, Berlin. Detektionsrate plus 18 Prozent bei stabiler Recall-Rate.

MASAI, ein randomisierter Trial aus Schweden mit 80.000 Frauen, publiziert in Lancet Oncology: Eingesetztes System: Transpara Version 1.7.0 von ScreenPoint Medical. Plus 20 Prozent Detektion, 44 Prozent weniger Befundungsaufwand.

ScreenTrustCAD aus Stockholm, publiziert in Lancet Digital Health: Eingesetztes System: Lunit INSIGHT MMG. KI als vollwertiger Zweitbefunder – ein Radiologe plus KI vs. zwei Radiologen ohne KI – non-inferior, leicht höhere Detektion.

Der gemeinsame Nenner: Etwa ein zusätzlicher Brustkrebs pro 1.000 gescreenter Frauen bei stabilen Recall-Raten. Die großen prospektiven Evidenzstudien nutzen also jeweils unterschiedliche KI-Systeme – und kommen dennoch zu konsistenten Ergebnissen.

Hintergrundwissen: Screening-Kennzahlen

Hintergrundwissen: Die wichtigsten Screening-Kennzahlen

Recall-Rate: Anteil der Frauen, die nach dem Screening zur Abklärung einbestellt werden – unabhängig vom Ergebnis. Frage: Wie viele Frauen müssen wir zurückrufen, wie hoch ist die Belastung durch Abklärungen? Im deutschen MSP liegt die Recall-Rate bei Folgeuntersuchungen bei 2,8 Prozent – der EU-Referenzwert fordert unter 5 Prozent.

Sensitivität: Anteil der Frauen mit Brustkrebs, die im Screening korrekt als verdächtig erkannt werden. Frage: Wie gut finden wir die vorhandenen Karzinome?

Spezifität: Anteil der Frauen ohne Brustkrebs, die korrekt als unauffällig klassifiziert werden. Frage: Wie gut vermeiden wir falsche Alarmer?

Positiver prädiktiver Wert (PPV): Anteil der auffälligen Befunde, bei denen tatsächlich Brustkrebs vorliegt. Frage: Wie groß ist die Trefferquote unter den zurückgerufenen Frauen? Im deutschen MSP liegt der PPV der Befundung bei 16 Prozent.

Negativer prädiktiver Wert (NPV): Anteil der unauffälligen Befunde, bei denen tatsächlich kein Brustkrebs vorliegt. Frage: Wie sicher kann ich mich auf einen unauffälligen Befund verlassen?

Diese Kennzahlen sind die Sprache der Qualitätskonferenz. Wenn wir über KI im Screening sprechen, sind es genau diese Parameter, an denen der Nutzen gemessen wird. Dr. Hamm wird Ihnen gleich zeigen, wie sich diese Kennzahlen in der Praxis mit und ohne KI verhalten.

Hintergrundwissen: Recall-Rate als Stellschraube

Hintergrundwissen: Recall-Rate als Stellschraube – und warum man sie nie isoliert betrachten darf

Die Recall-Rate ist die zentrale Stellschraube im Screening. Aber sie steht immer in einem Spannungsfeld:

Zu niedrige Recall-Rate: Wenige Rückrufe, hohe Spezifität – aber Risiko für übersehene Karzinome. Die Sensitivität sinkt, es drohen mehr Intervallkarzinome.

Sweet Spot: Ausgewogene Recall-Rate – hohe Sensitivität bei noch akzeptabler Spezifität und PPV. Das ist der Zielbereich der europäischen Leitlinien.

Zu hohe Recall-Rate: Die Sensitivität steigt kaum noch, aber die Zahl der Falschpositiven wächst. Der PPV sinkt, die Belastung für die Frauen und die Kosten steigen.

Entscheidend: Wenn KI die Detektionsrate steigert, muss man immer prüfen, ob die Recall-Rate stabil bleibt. Eine höhere Detektionsrate bei gleichzeitig steigender Recall-Rate wäre weniger beeindruckend als eine höhere Detektionsrate bei stabiler oder sogar sinkender Recall-Rate.

Genau das zeigen die Studienergebnisse: Die KI-Systeme in PRAIM, MASAI und ScreenTrustCAD steigern die Detektionsrate bei stabilen Recall-Raten – das ist der qualitative Unterschied zum klassischen CAD, das die Recall-Rate typischerweise erhöhte.

Wie KI im Screening-Workflow eingesetzt wird

Concurrent Reader

KI liest parallel zum Radiologen;
zeigt Befunde in Echtzeit

Zweiter Leser

KI ersetzt den zweiten Reader;
vergleichbare Qualität

Triage-System

Sortierung nach Dringlichkeit;
unauffällige Fälle anders geroutet

Dichtebewertung

Automatische Klassifikation;
reproduzierbarer als menschlich

Sprechernotizen – Folie 07: KI im Screening-Workflow

Wie wird KI konkret eingesetzt? Vier Hauptrollen.

Concurrent Reader: KI liest parallel mit und zeigt Befunde in Echtzeit – klassische Entscheidungsunterstützung.

Zweiter Leser: KI ersetzt den zweiten Befunder in der Doppelbefundung – das ist die Konfiguration, die ScreenTrustCAD in Stockholm mit Lunit INSIGHT MMG prospektiv getestet hat.

Triage: KI sortiert nach Dringlichkeit, unauffällige Fälle werden anders geroutet – MASAI hat hier mit Transpara 44 Prozent Reduktion der Lesungen gezeigt.

Dichtebewertung: Automatische, reproduzierbare Klassifikation der Brustdichte.

Bei allen Rollen gilt: KI unterstützt – sie ersetzt den Radiologen nicht.

Zur regulatorischen Einordnung in Deutschland: KI im Mammographie-Screening läuft derzeit im Rahmen von Implementationsstudien und lokalen Projekten. Sie ist nicht Bestandteil der G-BA-Pflichtvorgaben. Wie das praktisch in einer konkreten Screening-Einheit aussieht, zeigt Ihnen gleich Dr. Hamm für den Standort Chemnitz.

Chancen und Limitationen

Chancen

- Höhere Sensitivität bei stabiler Spezifität
- Reduktion der Leselast um bis zu 44 %
- Konsistenz ohne Ermüdung und Tagesformschwankung
- Fokussierung der Expertise auf komplexe Fälle
- Skalierbare Qualität auch bei geringer Facharztdichte
- Potenzial für frühere Detektion bei hoher Brustdichte

Limitationen & offene Fragen

- Bias: Trainingsdaten ggf. nicht repräsentativ
- Generalisierbarkeit zwischen Geräten/Standorten
- Überdiagnose: Mehr Detektion \neq gesicherter Nutzen
- Automation Bias: unkritische Übernahme von KI-Output
- Haftung bleibt nach aktueller Rechtslage beim befundenden Arzt
- Langzeitdaten (Intervallkarzinome, Mortalität) fehlen

Sprechernotizen – Folie 08: Chancen und Limitationen

Chancen: Höhere Sensitivität bei stabiler Spezifität. Bis zu 44 Prozent weniger Leselast. Konsistenz ohne Ermüdung. Fokussierung auf komplexe Fälle. Skalierbare Qualität auch in Regionen mit geringerer Facharztdichte – relevant für Sachsen.

Limitationen: Bias durch möglicherweise nicht repräsentative Trainingsdaten. Performance-Varianz zwischen Geräten und Standorten. Überdiagnose bleibt eine offene Frage. Automation Bias ist ein reales Risiko. Haftung bleibt beim Arzt. Und Langzeitdaten zu Intervallkarzinomraten und Mortalität stehen noch aus.

Zusammenfassung

- 1 KI im Mammascreeing ist bildanalytische Mustererkennung mit Deep Learning – kein Sprachmodell, kein Autopilot.
- 2 Das Mammographie-Screening ist das derzeit stärkste Evidenzfeld radiologischer KI – mit prospektiven Daten auf Bevölkerungsniveau.
- 3 KI ist Entscheidungsunterstützung. Die ärztliche Verantwortung bleibt beim Befunder.
- 4 Qualität entsteht nicht im Algorithmus, sondern in der Implementierung, Validierung und Überwachung.

→ Dr. Hamm zeigt Ihnen jetzt die konkreten Auswirkungen – mit realen Qualitätsindikatoren aus dem Chemnitzer Screening.

I Stopped Using ChatGPT for 30 Days. What Happened to My Brain Was Terrifying.

91% of you will abandon 2026 resolutions by January 10th. Here's how to be in the 9% who actually win.



Kusireddy

Follow

8 min read · Dec 28, 2025



13.1K



485

Medium



I was three sentences into an email when I caught myself reaching for ChatGPT.

Not for anything complex. Not for code. For *an email*.

That's when I realized: I couldn't think without AI anymore.

So I ran an experiment. 30 days. No ChatGPT. No AI shortcuts. Just my brain, the same one I'd spent years training and then promptly outsourced to a server farm.

Kontakt

Prof. Dr. Dieter Fedders

Institut für Radiologie und Neuroradiologie

Klinikum Chemnitz gGmbH

Flemmingstraße 2

09116 Chemnitz

d.fedders@skc.de