

Radiologe 2020 · 60:56–63
<https://doi.org/10.1007/s00117-019-00615-y>
 Online publiziert: 6. Dezember 2019
 © Der/die Autor(en) 2019



B. Bennani-Baiti · P. A. T. Baltzer

Universitätsklinik für Radiologie und Nuklearmedizin, Medizinische Universität Wien, Wien, Österreich

Künstliche Intelligenz in der Mammadiagnostik

Gestern, heute und morgen

Künstliche Intelligenz (KI) ist allgegenwärtig und wird gerade bei Radiologen teils kontrovers diskutiert. Dabei begleitet uns die KI gerade in der Mammadiagnostik schon seit fast 25 Jahren. Damals wurden die ersten computergestützten Diagnostiksysteme („computer-aided diagnostics“, CAD) für die Mammographie entwickelt. Diese fanden primär aus Refundierungsgründen in den USA eine breitere Anwendung als in Europa. Was ist heute anders als vor 25 Jahren? Was hat sich in den letzten Jahren verändert, und wohin steuert diese Entwicklung in der bildgebenden Diagnostik der Mamma?

Künstliche Intelligenz ist ein großes Gebiet, das unterschiedliche Algorithmen für unterschiedliche Aufgabenstellungen bietet. Akademische Forschung legt wünschenswerterweise diese Algorithmen offen, während industrielle Forschung naturgemäß ihre Methodik zwecks Vermarktung schützt. Ähnlich wie bei selbstfahrenden Transportsystemen fällt es schwer, einer Methode zu vertrauen, deren innere Gesetzmäßigkeiten man nicht kennt, selbst wenn große Datensätze den Nutzen dieser Methode belegen. Wichtig erscheint in diesem Zusammenhang die Darstellung der unterschiedlichen Herangehensweisen für computergestützte Diagnostik und ihrer Sinnhaftigkeit in der Praxis. Das Wohl der Patienten ist dabei oberstes Ziel: Wenn uns künstliche Intelligenz eine Verbreitung von Präzisionsmedizin bis in entlegene ländliche Regionen ermöglicht, ist dies begrüßenswert und kein Grund zur

Sorge („in 10 Jahren sind alle Radiologen arbeitslos“; [10]). Insbesondere bei präventiven Screening-Untersuchungen ist dies von großem Interesse. Letztlich liegt es an den Anwendern, hier Radiologen und Zuweisern gleichermaßen, die von KI profitierenden Bereiche zu definieren. Diese Übersichtsarbeit bietet einen Überblick über die Anwendungsgebiete von KI nach Bildgebungsmodalität, immer in Zusammenschau mit deren Implikationen für die klinische Praxis. Begriffsdefinitionen sind zur besseren Lesbarkeit in [Infobox 1](#) in den Übersichtsartikel integriert. Weiterführende Links zu öffentlichen Bilddatenbanken etc. finden sich am Ende des Artikels.

Künstliche Intelligenz in der Mammographie

In den mittlerweile nahezu nahtlos in der westlichen Welt eingesetzten Screening-Programmen spielt die Mammographie eine zentrale Rolle [21]. Die Mammographie bietet insbesondere bei geringer Brustdichte eine hervorragende diagnostische Genauigkeit zur Detektion von Brustkrebs. Im Rahmen dieser Screening-Programme fallen jährlich unzählige Mammographie-Untersuchungen an, von denen nur ein Bruchteil einen suspekten Befund aufweist und eine weiterführende diagnostische Abklärung benötigt. Die diagnostische Herausforderung ist hier die Ermüdung des die Untersuchung lesenden Arztes mit der Folge von Perzeptions- und Interpretationsfehlern.

KI-gestützte Systeme könnten hier vor allem folgende zwei Aufgaben erleichtern: 1) Selektion der abnormalen Befunde mit Blick auf Kosteneffizienz und

Reduktion einer stark monotonen Arbeitslast und 2) eine möglicherweise verbesserte diagnostische Genauigkeit. Im Gegensatz zum Menschen ermüdet der Computer insbesondere bei monotonen Aufgaben nicht. Darüber hinaus wünschenswert wären: 3) automatisierte standardisierte Dichtemessungen und 4) pathologisch-klinische Korrelation mit radiologischen Merkmalen für eine Risikoabschätzung oder zur Prädiktion von Therapieansprechen und zur Selektion einer entsprechend angepassten Therapie. Ob Letzteres (4) anhand mammographischen Bildmaterials möglich ist, bleibt fraglich. Hierfür liefern Modalitäten wie die Magnetresonanz und oder Positronen-Emissions-Tomographie [1, 24] oder der multiparametrische Ultraschall [8] wesentlich mehr biologische und tumorbiologische Informationen. Letztlich würden 5) auch automatische Integration von Bildmaterial aus anderen Modalitäten (z. B. Brustsonographie, Magnetresonanztomographie oder Tomosynthese) mit der Mammographie eine wesentliche Arbeitserleichterung mit sich bringen. Simpel und doch im Arbeitsalltag hilfreich wäre 6) die automatische Erkennung und Erfassung von technischen Parametern, Daten und Modalität von Voruntersuchungen sowie deren Integration in einen Befund. Insbesondere bei Tumorverlaufskontrollen wären automatisierte Größenmessungen der Zielläsionen im Verlauf hilfreich.

Historisch betrachtet, wurde das erste kommerziell eingesetzte CAD-System (Image Checker M1000; R2 Technology, Los Altos, Kalifornien, USA) bereits 1998 von der Food and Drug Administration (FDA) zugelassen [27]. Weitere gängige

Infobox 1 Unterstützende Begriffsdefinitionen

Maschinelles Lernen ist eine Form von künstlicher Intelligenz (KI), bei der verschiedene Algorithmen und Methoden zum Einsatz kommen, die ein statistisches Modell anhand eines Trainingsdatensatzes erstellen und im Eingangsdatensatz versteckte Muster detektieren, um Vorhersagen und Entscheidungen treffen zu können [14].

Beim überwachten maschinellen Lernen („supervised machine learning“) erhält der Algorithmus bekannte Daten als Eingangsvariablen (z. B. Bild Features, Risikofaktoren) sowie das erwartete Ergebnis (z. B. histologisches Ergebnis). Der Algorithmus lernt, indem er die Eingangsvariablen mit den tatsächlichen Ergebnissen vergleicht und so lange anpasst, bis dieser die erwartete Ausgangsvariable wiedergibt. Nach verifizierter Umsetzung des Lernprozesses werden verlässliche Vorhersagen für neue, dem Algorithmus unbekanntes Daten getroffen.

Beim unüberwachten maschinellen Lernen („unsupervised machine learning“) erhält der Algorithmus nur Eingangsdaten, aber nicht das erwartete Ergebnis. Es ist somit nicht dafür ausgelegt eine Vorhersage für eine bekannte Zielvariable (z. B. Diagnose) zu berechnen. Vielmehr suchen diese Algorithmen versteckte Muster und Gruppierungen in den Daten (z. B.: sog. Clusteringverfahren, welche u. a. in der genomischen Analytik weit verbreitet sind), anhand welcher diese kategorisiert werden. Das System findet dabei die Wahrheit in den Daten, welche mit der Zielstellung (z. B. Differenzierung verschiedener Tumortypen) in Einklang gebracht werden muss. Ein bekanntes Beispiel ist das Auffinden der intrinsischen genomischen Subtypen von Brustkrebs, welche nach Entdeckung relativ akkurat mit einfachen immunhistochemischen Markern diagnostiziert werden können. Die Klassifikation ist aber nicht 100% akkurat, weshalb ein luminal A Typ Tumor (Östrogenrezeptorpositiv, her2neu negativ, niedriger Proliferationsindex) nicht generell mit einem luminal A Tumor (durch ein genomisches Essay diagnostiziert) gleichzusetzen ist.

Deep Learning, *per definitionem* unter dem Überbegriff von maschinellem Lernen einzuordnen (Abb. 2), findet aufgrund höherer Rechenleistung von Grafikkarten seit 2012 breite Anwendung. Von Deep Learning spricht man ab einer höhergradigen Anzahl von zwischengeschalteten Layern in neuronalen Netzwerken, wobei es hier keinen definierten unteren Grenzwert gibt. Deep Learning nutzt vor allem Convolutional Neural Networks (CNN) zur Mustererkennung. Dabei können auch Pooling Layers Anwendung finden, die wesentlich zur Vereinfachung des Datensatzes beitragen, indem überflüssige Informationen verworfen werden. R-CNN (Region-based Convolutional Neuronal Networks) bieten zusätzlich eine Annotation der Lokalisation des auffälligen Befundes im Bild an.

KI/CAD gestützte Befundung beschreibt den Einsatz künstlicher Intelligenz in der Befundung. CAD und KI werden hier oft als Synonym verwendet. Auch rezenter entwickelte Software auf diesem Sektor ist definitionsgemäß ein CAD-System und fällt unter den Überbegriff KI-gestützter Befundung.

CAD-Systeme sind Parascript AccuDetect und VuCOMP M-Vu CAD. Eingangsbestandene neuronale Netzwerke (NNs) noch aus weitaus weniger verborgenen Schichten („hidden layers“) als heutzutage (ursprünglich bis zu 3, während heute durchaus 20 verborgene Schichten üblich sind; [7, 9]). Die schematische Darstellung eines neuronalen Netzwerks ist in Abb. 1 wiedergegeben, weitere Begriffsdefinitionen sind in Infobox 1 zusammengefasst. Das R2-Programm war darauf ausgelegt, Muster zu erkennen und basierte auf überwachtem maschinellen Lernen („supervised machine-learning“). Verkalkungen wurden anhand einer Ansammlung heller Punkte und spikulierter Rundherde mittels radiär angeordneter strahlenförmiger Linien in einem Radius von 3 bis 16 mm eines konzentrischen Kreises definiert. Daraus errechnete das R2-Programm eine

Malignitätswahrscheinlichkeit sowohl für suspekte Verkalkungen als auch für suspekte Rundherde und kennzeichnete sie, so ein vordefinierter Schwellenwert an Malignitätswahrscheinlichkeit überschritten wurde, für Radiologen im Bild. Die Anzahl falsch-positiver Markierungen pro Brust war nach heutigem Gesichtspunkt extrem hoch [12]. Somit war das R2-Programm als Zweitebender konzipiert, jedoch nicht als alleinestehendes diagnostisches Programm [27]. Auch die übrigen älteren CAD-Systeme setzten ähnlich diesem Vorgehen vordefinierte bekannte diagnostische Merkmale ein, um beispielsweise Architekturunterschiede/Asymmetrien durch Gegenüberstellung der beidseitigen Mammographien zu detektieren, und konnten im Vergleich zu heutigen Algorithmen nicht dazulernen [14]. Notabene wurde und wird das einem

Infobox 2 Weiterführende Links

- DREAM Challenge:
https://www.synapse.org/Digital_Mammography_DREAM_challenge
- Source code, Algorithmus und OsiriX Plugin der Ribli-Gruppe:
https://github.com/riblidezso/frcnn_cad.
- INBreast Datensatz: 115 Full Field Digital Mammographien, mit Pixel-Level Annotierung und histologisch verifizierten Karzinomen
http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database
- DDSM Datensatz: 2620 digitalisierte Film-Screen basierte Screening Mammographien, mit Pixel-Level Annotierung der Läsionen
<http://marathon.csee.usf.edu/Mammography/Database.html>
- ImageNet Datensatz: mehr als 14 Mio. nicht-radiologische, natürliche Bilder. Jedem Bild ist ein Substantiv zugeordnet. Zu jedem Substantiv gibt es mehr als 500 Bilder. Nutzbar beispielsweise für Transfer Learning.
<http://www.image-net.org>

neuronalen Netzwerk innewohnende Potenzial des kontinuierlichen Dazulernens bei kommerziellen Produkten nach der Trainings- und Validierungsphase *abgeschaltet*, damit sich der Algorithmus nach der Freigabe der Software nicht während des Betriebs durch die Anwender und in eine ungewollte Richtung weiterentwickelt, ohne dass der Hersteller in die Software korrigierend eingzugreifen könnte [7, 9].

Im Gegensatz zu Europa wurde in den USA bereits seit 2001 die CAD-unterstützte Befundung von den Krankenkassen erstattet. Im Jahr 2008 wurden in den USA 74% der der Medicare Population zugeordneten Mammographie Befunde CAD-unterstützt ausgewertet. Die jährlichen Kosten betragen dabei 400 Mio. Dollar bzw. ungefähr 10 Dollar pro Befund für CAD [12]. Dementsprechend folgte in den USA eine wesentlich breitere Anwendung von CAD-gestützten Befunden als in Europa. Im Jahr 2015 publizierten Lehman et al. die bis heute größte Studie zu diesem Thema mit Daten von knapp 500.000 Mammographien mit CAD vs. knapp 130.000 Mammographien ohne CAD zwischen 2003 und 2009. Befunde mit CAD boten mit einer

Sensitivität von 85,3 % und einer Spezifität von 91,6 % keine bessere diagnostische Genauigkeit im Vergleich zu Befunden ohne CAD, die eine höhere Sensitivität von 87,3 % und nur eine minimal geringere Spezifität von 91,4 % aufwiesen [12].

Zahlreiche andere Studien wurden zu traditionellen, auf überwachtem maschinellem Lernen basierten CAD-Systemen publiziert, die einerseits vielversprechend und andererseits ernüchternd hinsichtlich der Performance diverser CAD-Programme in der Mammographie waren. Wie bei technischen Neuentwicklungen zu erwarten, folgte auf eine anfängliche Euphorie die Ernüchterung. Das Fehlen öffentlicher repräsentativer Bilddatenbanken sowie die oft fehlende Offenlegung der zugrundeliegenden Algorithmen erschwert einen Vergleich zwischen den verschiedenen Methoden oder macht ihn gar unmöglich. Schließlich fand 2012 mit der breiten Anwendung von Deep Learning (■ **Infobox 1**) eine beschleunigte Entwicklung in der praktischen Anwendung von KI statt. Grund war die durch parallelisierte Berechnungen und die Einbindung von eigenständigen Grafikprozessoren deutlich gesteigerte Rechenleistung, welche u. a. die Vermehrung der verborgenen Schichten, bzw. der *Tiefe* des lernenden Netzwerks gestattete.

Dementsprechend interessant sind die Ergebnisse der „Digital Mammography DREAM Challenge (DM Challenge)“ aus den Jahren 2016–2017 (■ **Infobox 2** Weiterführende Links). Im Rahmen der DM Challenge [26] sollte ein Algorithmus zur verlässlichen Bewertung von Screening-Mammographien als maligne oder benigne entwickelt werden. Dafür wurde den Teilnehmern ein Datenset mit 640.000 Bildern von über 86.000 Patientinnen ohne Pixelannotation, aber mit Bewertung (maligne/benigne; Diagnose erfolgte binnen 12 Monaten nach der Untersuchung) zur Verfügung gestellt. Dabei wurde jede Brust separat bewertet. Ziel war die Identifikation biologisch aggressiver, invasiver Brusttumoren, die mit einer ungünstigeren Sterblichkeitsprognose assoziiert waren. Die Daten wurden den Teilnehmern auf einer cloudbasierten Plattform zu-

Radiologe 2020 · 60:56–63 <https://doi.org/10.1007/s00117-019-00615-y>
© Der/die Autor(en) 2019

B. Bennani-Baiti · P. A. T. Baltzer

Künstliche Intelligenz in der Mammadiagnostik. Gestern, heute und morgen

Zusammenfassung

Hintergrund. Künstliche Intelligenz (KI) wird zunehmend in der bildgebenden Mammadiagnostik eingesetzt.

Fragestellung. Was sind die wesentlichen Anwendungsgebiete und bereits verfügbaren KI-gestützten Systeme auf dem Gebiet der senologischen Radiologie?

Material und Methode. Zusammenfassung, übersichtliche Darstellung sowie Diskussion relevanter Daten aus Grundlagenarbeiten und Herstellerinformationen.

Ergebnisse. KI-gestützte Befundsysteme sind bereits seit fast 25 Jahren in der Mammadiagnostik im Einsatz, werden jedoch in den USA weitaus häufiger angewandt als in Europa. Initiale CAD-Systeme (computer-gestützte Diagnostik) zeichneten sich durch eine limitierte diagnostische Aussagekraft mit überproportionalen Raten falsch-positiver Befunde aus. Seit 2012 kommen

Deep-Learning-Algorithmen zum Einsatz und erweitern die Einsatzmöglichkeiten der KI. **Schlussfolgerung.** Derzeit gibt es keinen Algorithmus, der – nachweislich validiert – die Doppelbefundung durch zertifizierte Mammadiagnostiker übertrifft. In der mittelbaren Zukunft ist jedoch die Übernahme von Aufgaben durch KI absehbar: So kann die Vorselektion abnormer Studien die Arbeitslast gerade in der Früherkennung substantiell verringern. Auch die Erfassung von komplexen radiopathogenomischen Zusammenhängen und besonders ihre Translation in die klinische Praxis ist ohne KI kaum denkbar.

Schlüsselwörter

Brustkrebs · Mammographie · Computergestützte Diagnostik · Screening · Früherkennung

Artificial intelligence in the diagnosis of breast cancer. Yesterday, today and tomorrow

Abstract

Background. Artificial intelligence (AI) is increasingly applied in the field of breast imaging.

Objectives. What are the main areas where AI is applied in breast imaging and what AI and computer-aided diagnosis (CAD) systems are already available?

Materials and methods. Basic literature and vendor-supplied information are screened for relevant information, which is then pooled, structured and discussed from the perspective of breast imaging.

Results. Original CAD systems in mammography date almost 25 years back. They are much more widely applied in the United States than in Europe. The initial CAD systems exhibited limited diagnostic abilities and disproportionately high rates of false positive results. Since 2012, deep learning

mechanisms have been applied and expand the application possibilities of AI.

Conclusion. To date there is no algorithm that has beyond doubt been proven to outperform double reporting by two certified breast radiologists. AI could, however, in the foreseeable future, take over the following tasks: preselection of abnormal examinations to substantially reduce workload of the radiologists by either excluding normal findings from human review or by replacing the double reader in screening. Furthermore, the establishment of radio-patho-genomic correlations and their translation into clinical practice is hardly conceivable without AI.

Keywords

Breast neoplasms · Mammography · Computer-aided diagnosis · Screening · Early detection of cancer

gänglich gemacht. Dieses Projekt war insofern beispielhaft, als erstmals verschiedene Lösungsansätze aufgrund des identen Trainings- und Bewertungsdatensatzes verglichen werden konnten. Gewinner der DM Challenge waren die Ann-Arbor-Gruppe (Department of

Computational Medicine and Bioinformatics at the University of Michigan) und Therapixel (Paris, Frankreich). Beide verwendeten Deep-Learning-Strategien. Therapixel erreichte im DM-DREAM-Challenge-Datensatz eine anhand der Fläche unter der ROC-Kurve gemessene

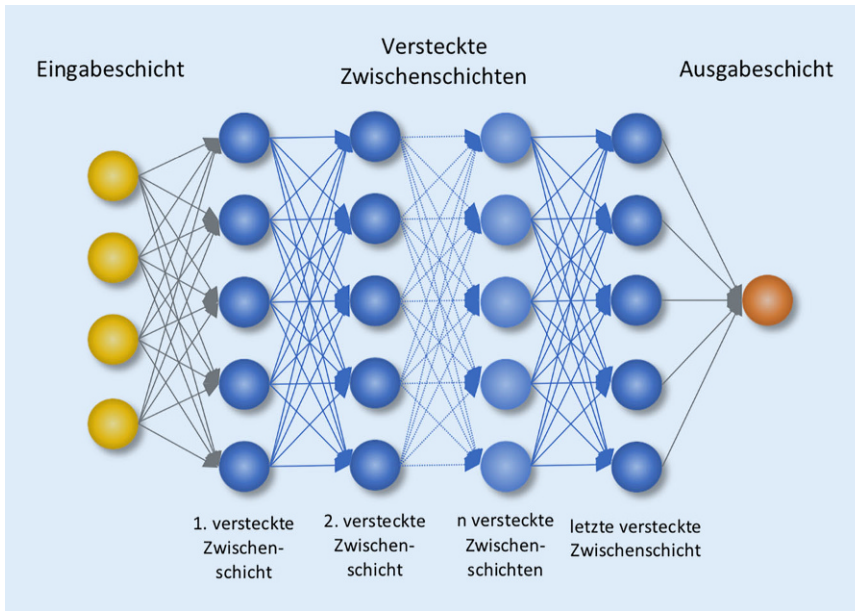


Abb. 1 ▲ Schematische Darstellung eines neuronalen Netzwerks (hier ein Feed-forward neuronales Netzwerk, da sich die Information ausschließlich nach vorne ausbreitet). Das einfachste neuronale Netzwerk (Perzeptron) besteht aus einer Eingangs- und einer Ausgangsschicht. Darüber hinaus sind der Komplexität neuronaler Netzwerke kaum Grenzen gesetzt

Genauigkeit von 87%. Das entspricht einer diagnostischen Genauigkeit von 80,4%. Die Ann-Arbor-Gruppe erreichte eine diagnostische Genauigkeit von 77,5%, hatte jedoch hinsichtlich anderer Parameter eine ähnliche Qualifizierung wie Therapixel, sodass beide Wettbewerber gemeinsam erstplatziert wurden.

Die drittplatzierte Gruppe um Dr. Ribli lag weitere 2% unter der von der Ann Arbor erreichten diagnostischen Genauigkeit. Therapixel hat dafür das in der Bildanalyse relativ neue *Transfer Learning* angewandt, wobei der Algorithmus anhand von natürlichen Bildern gewisse Features (z. B. Kontur etc.) lernt und dieses Wissen dann bei einem radiologischen Bilddatensatz anwendet. Derzeit arbeiten Therapixel und die Ann-Arbor-Gruppe im finalen Stadium der DM Challenge an einem durch das gemeinsame Wissen optimiertem Algorithmus, der noch bessere Resultate erzielen sollte. Bereits publiziert und öffentlich zugänglich ist der Algorithmus der zweitplatzierten Gruppe um Ribli et al. [18]. Auch diese Gruppe hat Deep Learning mittels einer Convolutional-Neuronal-Network(CNN)-Architektur eingesetzt. Deren Algorithmus bietet über R-CNN (■ **Infobox 1**), eine Markie-

rung der suspekten Areale im Bild, was damit sowohl eine Klassifizierung des Gesamtbildes (maligne vs. benigne) als auch eine Möglichkeit zur Verwendung als Unterstützung bei der Befundung ermöglicht. Dabei werden auch diese gekennzeichneten Areale vom Algorithmus auf einer Skala bewertet, die dem Befunder die Wahrscheinlichkeit der Malignität dieser Areale wiedergibt.

Die DM DREAM Challenge zeigt die durch Deep Learning im Vergleich zu ursprünglichen CAD-Systemen erreichbare Steigerung der verarbeitbaren Datenmenge und die damit implizierten Möglichkeiten in der klinischen Anwendung. Kürzlich hat eine Multireader-Studie mit 101 Radiologen die Performance eines kommerziell verfügbaren CAD-Systems (Transpara 1.4.0, Screenpoint Medical BV, Nijmegen, Niederlande) anhand von 2652 Untersuchungen (davon 653 maligne Tumoren) getestet. Transpara verwendet ebenso Deep-Learning-CNN-Algorithmen, um Verkalkungen sowie Weichteilläsionen zu detektieren. Das System klassifiziert das Untersuchungsergebnis auf einer Skala von 1 bis 10, wobei 10 eine hohe Wahrscheinlichkeit eines malignen Befunds indiziert. Es basiert seine *Diagnose* auf MLO und

CC Mammographien, von denen beide gemeinsam untersucht und verwertet werden. Kritisch anzumerken ist bei Transpara die fehlende Berücksichtigung von Vorbefunden für das Scoring. Auf die Performance aller Radiologen gemeinsam bezogen, war die Performance des Transpara-Algorithmus non-inferior (AUC Transpara 0,84 vs. AUC 0,81 Radiologen); das System erreichte einen höheren AUC-Wert als 61,4% der teilnehmenden Radiologen. Allerdings war Transpara stets den besten Radiologen unterlegen, zudem wurde die Studie im Sinne eines Interessenkonflikts nicht unabhängig von der Herstellerfirma durchgeführt. Trainiert wurde der Algorithmus ursprünglich an einer Datenbank mit über 9000 Karzinomaufweisenden und 180.000 unauffälligen Mammographien [20]. Die Vorversion dieser Software (Transpara 1.3.0) hat als unterstützendes Tool die durchschnittliche Befundperformance von 14 zertifizierten Brustradiologen von einer AUC von 0,87 (ohne KI) auf 0,89 (Befundung mit CAD) gesteigert ($p=0,002$). Die Befundzeit wurde durch die KI-gestützte Befundung nicht wesentlich beeinflusst (146 s ohne KI und 149 s mit KI; $p=0,15$). Letztere Performance-Daten wurden anhand eines wesentlich kleineren Datensatzes erhoben mit insgesamt 240 Untersuchungen, davon 100 unauffällige Mammographien, 100 mit malignem Befund und 40, die zu einem falsch-positiven Ergebnis geführt hatten [19]. Allein der Vergleich dieser beiden Studien verdeutlicht, dass AUC-Werte immer in Zusammenschau mit den zugrundeliegenden untersuchten Daten(banken) bewertet werden müssen und derzeit ein Vergleich der Performance von Radiologen mit der jeweiligen KI-gestützten Befundung/oder KI-assistierten Befundung in unabhängigen Datenbanken obligat ist, bevor an einen Einsatz in der klinischen Routine gedacht werden kann [13]. Weiterhin sind insbesondere von Firmen vollständig unabhängige Studien zur objektiven Bewertung von CAD/KI-Systemen notwendig. Eine kontinuierliche Qualitätskontrolle solcher Systeme ist vom ärztlichen Standpunkt im Sinne der Patientensicherheit unabdingbar.

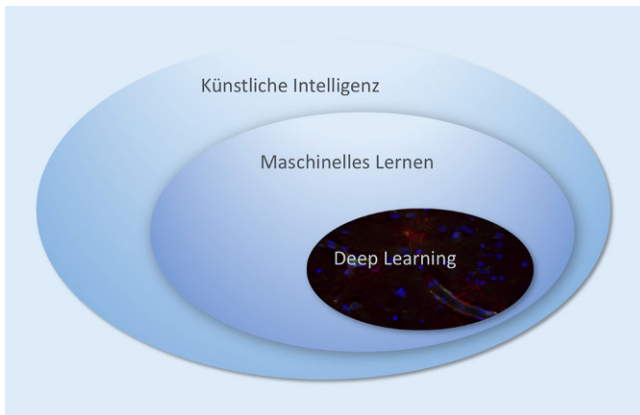


Abb. 2 ◀ Schematische Einordnung der Überbegriffe von künstlicher Intelligenz (KI)

Aus den beschriebenen Studien wird ersichtlich, dass die Forschungsbestrebungen derzeit in erster Linie schon aufgrund der hohen Relevanz und des damit verbundenen finanziellen Potenzials auf Screening-Mammographien und hier die primär die Reduktion der Befundlast abzielen. Indem ein Algorithmus Malignitätswahrscheinlichkeiten errechnet, müssten unterhalb eines validierten Schwellenwerts liegende Untersuchungen Befunde nicht weiter ärztlich begutachtet werden. Dazu gehört auch die Anwendung von CAD/KI-Systemen als Ersatz für den Zweitbefunder der Doppelbefundung. Im Gegensatz dazu gibt es derzeit keine Daten, welche die Überlegenheit von KI-gestützten Systemen gegenüber menschlich erstellten Befunden eindeutig bewiesen haben. Abgesehen von den oben beschriebenen Anwendungsmöglichkeiten von KI in der Mammographie, wäre eine andere Herangehensweise zwar finanziell weniger interessant, aber im Sinn der Patientensicherheit von großem Interesse: Wenn man davon ausgeht, dass ein nicht unwesentlicher Bestandteil der Mammographien zu falsch-negativen Ergebnissen führt und bei einer Sensitivität der Mammographie im Screening-Setting bei etwa 77–87% Brustkrebs unterdiagnostiziert [2, 12], wäre eine Verringerung falsch-negativ befundeter Mammographien mittels KI-Methoden wünschenswert.

Ebenso wäre die potenzielle Frühdetektion von Intervallkarzinomen anhand von minimalen Zeichen („minimal signs“) ein potenzielles Anwendungsgebiet für KI. All diese Bestrebungen sind immer durch die begrenzt vorhandene

Datenmenge limitiert; – ein Umstand, der sich durch die europäische Datenschutzgrundverordnung (DSGVO) nicht gebessert hat. Letztlich benötigt man für Deep Learning riesige, adäquat referenzierte und kalibrierte öffentlich zugängliche Datensätze, um die nötige Qualität zu gewährleisten [23]. Dafür bedarf es enger internationaler und interdisziplinärer Kollaboration, einheitliche Befundschemata und Definitionen und fächerübergreifenden Konsens. Gerade die komplementäre Mammadiagnostik, in der man aufgrund der engen Patienteninteraktion Zugang zu Anamnese und verschiedenen Untersuchungsmodalitäten hat, könnte hier eine der Vorreiterrollen in der KI der Bildgebung einnehmen – insbesondere auch unter dem Aspekt, dass man hier im Vergleich zu anderen Körperregionen oft histologische Verifikation sowie Follow-up-Untersuchungen zur Verfügung hat. Noch problematisch sind nicht nur die derzeit ungenügenden Datenbanken, sondern auch teilweise die großen Datenmengen – um auch die Detektion von Verkalkungen und subtilen Befunden zu ermöglichen, benötigt die Mammographie eine weitaus größere Menge an Bildinformationen (ca. 3000×3000 Pixel) als natürliches, nicht-radiologisches Bildmaterial (ca. 300×300 Pixel; [11]). Allerdings ist diese Limitation nur eine sehr bedingte, und eine, die mit der Weiterentwicklung von Grafikkarten und technischem Fortschritt, nur eine Limitation auf Zeit ist.

KI-gestützte Dichtemessungen

Erhöhte Brustdichte in der Mammographie geht mit einem erhöhten Risiko für Brustkrebs einher und bedingt eine erniedrigte Sensitivität der Untersuchung, da dichtes Brustgewebe suspektere Areale maskieren kann [15, 16]. Die Bewertung der Brustdichte erfolgt derzeit primär nach subjektiver Bewertung des Befunders anhand der BI-RADS-Kriterien und wird in 4 Kategorien (ACR A–D) unterteilt [5]. Brustdichte beschreibt das prozentuelle Verhältnis zwischen glandulärem Gewebe und Fettgewebe der Brust. Gerade zwischen ACR B und C besteht eine nicht unwesentliche Interobservervarianz. Auch auf diesem Gebiet gibt es bereits seit längerem Bestrebungen, diese Befundung mittels KI zu erleichtern und durch gleichmäßig arbeitende Algorithmen zu vereinheitlichen. Eine verbreitete kommerzielle Anwendung dafür ist VolparaDensity (Volpara Solutions, Wellington, New Zealand), die das glanduläre Brustgewebe automatisch als Prozentsatz des gesamten Brustvolumens berechnet und dann sowohl diesen Wert wiedergibt als auch eine BI-RADS-Klassifikation vornimmt. Die neue Definition der BI-RADS-C-Kategorie umfasst aber auch Befunde, in denen einige Areale von derart erhöhter Dichte sind, dass etwaige Herdbefunde maskiert werden könnten, selbst wenn das Brustgewebe insgesamt eher einer ACR-Dichte B oder A entspricht. Diese kann vom Algorithmus naturgemäß nicht berücksichtigt werden und sollte bei der Verwendung von Volpara bedacht werden [5].

Quantra 2,2 Breast Density Assessment Software (Hologic, Marlborough, Massachusetts, USA) ist darauf ausgelegt, konform mit der 5. BI-RADS-Edition zu bewerten, und erreicht 97,1% Sensitivität und 83,1% Spezifität, wenn BI-RADS A–B und BI-RADS C–D gepoolt untersucht werden. Nachdem erhöhte Brustdichte (ab BI-RADS C) von höherer klinischer Relevanz ist, scheinen diese Daten vielversprechend. Wäre eine individuelle Bewertung aller Brustdichtegrade essenziell, würde hier noch Verbesserungsbedarf bestehen [6]. Densitas (Emergo Group Inc.) ist das jüngste von der FDA zugelassene Produkt auf dem

Markt (seit 2018) und verwendet ebenfalls einen Algorithmus der die prozentuelle Brustdicke sowie die BI-RADS-Klassifikation ermittelt [28]. Nachdem erhöhte Dichtegrade insbesondere wegen der erhöhten Brustkrebswahrscheinlichkeit und der erniedrigten Sensitivität von diagnostischem Wert sind, wäre es hier in weiterer Zukunft besonders sinnvoll, einen Algorithmus zu entwickeln, der einerseits anzeigt, wann eine Untersuchung durch eine weitere diagnostische Modalität (z. B. Ultraschall, MRT, Tomosynthese) ergänzt werden sollte und wann aufgrund eventuell zusätzlich erhobener klinischer Parameter (Alter, hormoneller Status etc.) ein Risikoprofil zusammen mit der ermittelten Brustdicke personalisierte Screening-Intervalle und Untersuchungsmodalitäten ergibt.

Kontrastmittelgestützte Mammographie

Zu dem relativ jungen Verfahren der kontrastmittelgestützten Mammographie gibt bislang kaum Daten zur Anwendung von KI-Algorithmen. Potenziell sind die Anwendungsgebiete jedoch analog den bereits diskutierten.

Anwendungen in der Digitalen Brust-Tomosynthese

Die Tomosynthese als Weiterentwicklung der Mammographie bietet dank Pseudo-3-D-Schichtaufnahmen eine überlagerungsfreie Darstellung des Brustgewebes. In der Screening-Anwendung werden erhöhte Tumordetektionsraten und niedrigere Recall-Raten berichtet; der Wert zur Reduktion von Intervallkarzinomen als Surrogat für die Effizienz des Screenings wird noch kontrovers diskutiert. Bei erhöhter Brustdicke kann die digitale Tomosynthese (DBT) eine sinnvolle ergänzende Untersuchung darstellen [22]. Methodeninhärent bietet die dreidimensionale DBT mehr Datenpunkte als die klassische zweidimensionale Mammographie. Folglich resultieren erhöhte Befundzeiten (50–200 %; [25]).

Zusätzlich zu den für die Mammographie beschriebenen möglichen Einsatzgebieten kann KI insbesondere

in der Reduktion der Befundzeiten eine entscheidende Rolle spielen, wenn es um einen zukünftig breiteren Einsatz von DBT geht. Transpara bietet neben den oben beschriebenen 2-D-Mammographie-Lösungen nun auch Lösungsansätze für Tomosynthese basierend auf dem gesamten 3-D-Datensatz an. Auch dafür werden Algorithmen mittels Deep Learning trainiert, die Kalzifizierungen und Weichteilläsionen quantifizieren und im Bild markieren. Analog zum 2-D-Lösungsansatz ist auch hier angedacht, bei geringen Malignitätswahrscheinlichkeiten eine Triagefunktion zu erfüllen. Alternativ dazu bietet das ebenfalls Deep Learning-gestützte PowerLook Tomo Detection System (iCAD, Nashua, NH, USA) mit einem Fokus auf Verringerung der Befundungszeit, eine Reduktion des 3-D-Datensatzes auf ein 2-D-Bild an. Auf diesem markiert das System auffällige Regionen, um die Evaluierung und das Auffinden dieser Areale im 3-D-Datensatz zu beschleunigen [4]. Des Weiteren berichten Conant et al., dass PowerLook Tomo Detection gestützte Befundung sowohl AUC, Sensitivität und Spezifität gegenüber ausschließlich menschlicher Befundung erhöht und Rückrufraten reduziert [4]. Im Vergleich zu den 2-D-Lösungen gibt es für DBT noch deutlich weniger Lösungsansätze. KI-Ansätze könnten aber die Akzeptanz und Anwendung von DBT im Screening wesentlich beeinflussen. Dann kommen dieselben Überlegungen für weitere KI-Applikation wie in der Mammographie zu tragen. In diesem Szenario wären diese jedoch durch die vergleichsweise gering verfügbaren entsprechend annotierten Datensätze derzeit noch stark limitiert. Zudem benötigt man durch die pro Untersuchung um eine Dimension höhere Datenmenge auch entsprechend deutlich mehr Referenzdatensätze um ein Overfitting der Modelle zu vermeiden.

Künstliche Intelligenz in der Brustsonographie

Brustsonographie ist eine breit eingesetzte Methode insbesondere zur weiterführenden diagnostischen Abklärung klinisch oder bildgebend auffälliger Befunde inklusive der ultraschallgezielten

Biopsie abklärungsbedürftiger Areale. Dabei ist die diagnostische Leistung des Brustultraschalls stark vom individuellen Untersucher abhängig. Ultraschallbilder sind demnach nicht einfach mit künstlicher Intelligenz analysierbar, da es an einer einheitlichen Bildakquisition mangelt.

Damit konzentrieren sich die kommerziell verfügbaren Methoden derzeit primär auf automatisch erstellte Ultraschallbilder, die einheitliche Datensätze liefern („automated breast ultrasound“, ABUS). Beispielsweise bietet QView Medical mit QView CAD ein solches System seit 2016 an. Es basiert auf Deep-Learning-Mechanismen, die den Algorithmus anhand von 20.000 ABUS-Datensätzen (3000 davon bioptisch gesicherte Malignome) erstellt haben. Suspekte Areale werden ab einer Größe von 5 mm detektiert und für den Untersucher mit einem Malignitätsscore bewertet [17].

Koios Medical bietet mit Koios DS Software einen alternativen Ansatz. Die Software arbeitet integriert im PACS und analysiert die vom Untersucher auf dem Bild indizierte Region auf mittlerweile über 170.000 Features, um anhand derer eine entsprechende Bewertung zu erstellen. Diese Features wurden mittels supervidiertem Lernen erstellt. Im Gegensatz zu älteren rigiden CAD-Systemen aus der Mammographie wird diese Software durch fortwährendes Einspielen neuer Bilder mit assoziierter pathologischer Korrelation dynamisch korrigiert und weiterentwickelt, was jedoch neue Herausforderungen an die Qualitätskontrolle stellt [3].

Magnetresonanztomographie der Brust

Aufgrund der erheblichen Bilddatenmenge in der Magnetresonanztomographie (MRT) ist KI auf diesem Gebiet anders einzusetzen als in den vorab beschriebenen Modalitäten. Einen MR-Datensatz auf ein zweidimensionales Bild zu reduzieren, ist wenig sinnvoll und würde die Überlegenheit der Vielfalt an Informationen, die durch die multiparametrische Bildgebung gegeben ist, nivellieren. Gerade dem unerfahrenen Untersucher kann ein Großteil der

potenziell in der MRT-Untersuchung enthaltenen Daten entgehen. Computerunterstützung anhand von KI-Algorithmen kann diese für uns visualisieren und quantifizieren. Die beiden bekanntesten Softwareapplikationen im Brust-MRT-Bereich sind hierfür CADstream (Merge Healthcare) und DynaCAD (Invivo). Subjektiv betrachtet bietet die MRT die breiteste Einsatzmöglichkeit für KI-gestützte Befundung und Wissenschaft. Alle für die Mammographie beschriebenen Einsatzgebiete sind auch für die MRT denk- und anwendbar. Risikoabschätzung, Läsionsklassifizierung, Therapieansprechen, Vorliegen von Metastasen, hormonelle Aktivität etc. sind alles Themenbereiche die teils bereits erfolgreich beforscht, teils jedoch zeitnah mittels KI sicherlich besser beantwortet werden können, als wir das mit freiem Auge aus MRT-Bildern ablesen könnten. Auf diesem Gebiet können mittels Radiomics und Texturanalyse uns verborgene Features entdeckt werden und zusammen mit proteinomischen, metabolomischen und genomischen Ergebnissen Zusammenhänge erstellt werden, was nun aufgrund der Rechenleistung moderner Hardware möglich ist und die Kompetenz des einzelnen Individuums überschreiten würde.

Zusammenfassend ist KI bereits integraler Bestandteil in der Befundung. Dies insbesondere in den USA, primär da dort CAD-gestützte Befundung von den Kassen refundiert wird. Es gibt eine Reihe von sinnvollen Einsatzgebieten von KI in der Brustbildgebung, wobei festzuhalten ist, dass für die meisten dieser, radiologisches Bildmaterial als Grundlage alleine nicht ausreichend sein kann. Als Radiologe in der Mammadiagnostik sind Indikation und Fragestellung sowie Anamnese und klinischer Befund integraler Bestandteil in der Bewertung des Bildmaterials. Wenn man KI ausschließlich basierend auf Bildmaterial Diagnosen und Risikoabschätzungen erstellen lassen würde, kämen vermutlich sehr irreführende Resultate zustande. Dementsprechend muss die Erstellung von Algorithmen für KI in der Mammadiagnostik immer ein multidisziplinäres Projekt aller beteiligten Disziplinen sein, unter Einbindung von Gynäkologen, Chirurgen,

Onkologen und Pathologen und sollte nicht ausschließlich zwischen Radiologen und Softwareentwicklern stattfinden. KI kann wie in fast jeder Sparte auch in der Mammadiagnostik sinnvoll eingesetzt werden, um einerseits ganz einfache Aufgaben im Sinne der Reduktion von monotoner Arbeitslast oder andererseits hochkomplexe Aufgaben im Sinne von Durchführung einer erheblichen Rechenleistung und Integration z. B. von genetischem Datenmaterial, das im klinischen Alltag zu weit führen würde, zu erfüllen.

Es liegt auch an uns Klinikern, Forschern und Anwendern zu definieren, welche Aufgaben dies sind. Insbesondere sollten wir diese Entscheidung nicht primär profitorientierten Unternehmen überlassen. Zudem sind publizierte Ergebnisse auf diesem Gebiet immer kritisch zu hinterfragen, da hier aufgrund des kommerziellen Werts wirtschaftliche Interessen zu tragen kommen. Darüber hinaus wurde rezent gezeigt, dass nur ein Bruchteil der bislang publizierten Studien auf diesem Gebiet den dafür erforderlichen qualitativen Anforderungen genügt. Unabhängige Datensätze zur Validierung eines Modells für den direkten Vergleich von menschlicher Befundperformance mit KI-Befundperformance sind unerlässlich [13].

Fazit für die Praxis

- KI ist bereits seit nahezu 25 Jahren integraler Bestandteil in der senologischen Radiologie.
- KI ist abhängig von der Qualität der Datenbank, an welcher ein Algorithmus trainiert wird.
- Unabhängige Studien zur Evaluierung von Performance und klinischer Anwendbarkeit eines Algorithmus sind unerlässlich.

Korrespondenzadresse



Mag. DDr. B. Bennani-Baiti
 Universitätsklinik
 für Radiologie und
 Nuklearmedizin, Medizinische
 Universität Wien
 Währinger Gürtel 18–20,
 1090 Wien, Österreich
 barbara.bennani-baiti@
 meduniwien.ac.at

Funding. Open access funding provided by Medical University of Vienna.

Einhaltung ethischer Richtlinien

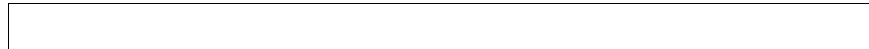
Interessenkonflikt. B. Bennani-Baiti und P.A.T. Baltzer geben an, dass kein Interessenkonflikt besteht.

Dieser Beitrag beinhaltet keine Studien an Menschen oder Tieren. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

1. Andrzejewski P, Wengert G, Helbich TH et al (2019) Sequential [18 F]FDG-[18 F]FMISO PET and Multiparametric MRI at 3T for Insights into Breast Cancer Heterogeneity and Correlation with Patient Outcomes: First Clinical Experience. *Contrast Media Mol Imaging*. <https://doi.org/10.1155/2019/1307247>
2. Banks E, Reeves G, Beral V et al (2004) Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: cohort study. *BMJ* 329:477. <https://doi.org/10.1136/bmj.329.7464.477>
3. Barinov L, Jairaj A, Becker M et al (2019) Impact of data presentation on physician performance utilizing artificial intelligence-based computer-aided diagnosis and decision support systems. *J Digit Imaging* 32:408–416. <https://doi.org/10.1007/s10278-018-0132-5>
4. Conant EF, Toledano AY, Periaswamy S et al (2019) Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis. *Radiology* 1:e180096. <https://doi.org/10.1148/ryai.2019180096>
5. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA et al (2013) ACR BI-RADS[®] Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology
6. Ekpo EU, Mello-Thoms C, Rickard M et al (2016) Breast density (BD) assessment with digital breast tomosynthesis (DBT): Agreement between



- Quantra™ and 5th edition BI-RADS®. Breast 30:185–190. <https://doi.org/10.1016/j.breast.2016.10.003>
7. Giger ML (2018) Machine learning in medical imaging. *J Am Coll Radiol* 15:512–520. <https://doi.org/10.1016/j.jacr.2017.12.028>
 8. Kapetas P, Clauser P, Woitek R et al (2019) Quantitative multiparametric breast ultrasound: application of contrast-enhanced ultrasound and elastography leads to an improved differentiation of benign and malignant lesions. *Invest Radiol* 54:257–264. <https://doi.org/10.1097/RLI.0000000000000543>
 9. Kegelmeyer WP, Pruneda JM, Bourland PD et al (1994) Computer-aided mammographic screening for spiculated lesions. *Radiology* 191:331–337. <https://doi.org/10.1148/radiology.191.2.8153302>
 10. Langs G, Röhrich S, Hofmanninger J et al (2018) Machine learning: from radiomics to discovery and routine. *Radiologe* 58:1–6. <https://doi.org/10.1007/s00117-018-0407-3>
 11. Le EPV, Wang Y, Huang Y et al (2019) Artificial intelligence in breast imaging. *Clin Radiol* 74:357–366. <https://doi.org/10.1016/j.crad.2019.02.006>
 12. Lehman CD, Wellman RD, Buist DSM et al (2015) Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med* 175:1828. <https://doi.org/10.1001/jamainternmed.2015.5231>
 13. Liu X, Faes L, Kale AU et al (2019) A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 1:e271–e297. [https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2)
 14. Mohammed M, Khan MB, Bashier EBM (2017) Machine learning: algorithms and applications. CRC Press, Taylor & Francis Group, Boca Raton
 15. Moshina N, Sebuodegård S, Lee CI et al (2018) Automated volumetric analysis of mammographic density in a screening setting: worse outcomes for women with dense breasts. *Radiology* 288:343–352. <https://doi.org/10.1148/radiol.2018172972>
 16. Pettersson A, Graff RE, Ursin G et al (2014) Mammographic density phenotypes and risk of breast cancer: a meta-analysis. *J Natl Cancer Inst*. <https://doi.org/10.1093/jnci/dju078>
 17. QView Medical QView Whitepaper. <https://www.qviewmedical.com>. Zugegriffen: 10. Okt. 2019
 18. Ribli D, Horváth A, Unger Z et al (2018) Detecting and classifying lesions in mammograms with deep learning. *Sci Rep* 8:4165. <https://doi.org/10.1038/s41598-018-22437-z>
 19. Rodríguez-Ruiz A, Krupinski E, Mordang J-J et al (2018) Detection of breast cancer with mammography: effect of an artificial intelligence support system. *Radiology* 290:305–314. <https://doi.org/10.1148/radiol.2018181371>
 20. Rodríguez-Ruiz A, Lång K, Gubern-Merida A et al (2019) Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. *J Natl Cancer Inst* 111:916–922. <https://doi.org/10.1093/jnci/djy222>
 21. Sardanelli F, Aase HS, Álvarez M et al (2017) Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey. *Eur Radiol* 27:2737–2743. <https://doi.org/10.1007/s00330-016-4612-z>
 22. Sardanelli F, Fallenberg EM, Clauser P et al (2017) Mammography: an update of the EUSOBI recommendations on information for women. *Insights Imaging* 8:11–18. <https://doi.org/10.1007/s13244-016-0531-4>
 23. Schütze B, Schlieter H (2019) Künstliche Intelligenz: Ein nützliches Werkzeug für den Radiologen? *Radiologe*. <https://doi.org/10.1007/s00117-019-00599-9>
 24. Stadlbauer A, Zimmermann M, Bannani-Baiti B et al (2018) Development of a non-invasive assessment of hypoxia and neovascularization with magnetic resonance imaging in benign and malignant breast tumors: initial results. *Mol Imaging Biol*. <https://doi.org/10.1007/s11307-018-1298-4>
 25. Tagliafico AS, Calabrese M, Bignotti B et al (2017) Accuracy and reading time for six strategies using digital breast tomosynthesis in women with mammographically negative dense breasts. *Eur Radiol* 27:5179–5184. <https://doi.org/10.1007/s00330-017-4918-5>
 26. Trister AD, Buist DSM, Lee CI (2017) Will machine learning tip the balance in breast cancer screening? *JAMA Oncol* 3:1463. <https://doi.org/10.1001/jamaoncol.2017.0473>
 27. U.S. Food and Drug Administration (1998) Summary of safety and effectiveness data: R2 technologies (P970058)
 28. U.S. Food and Drug Administration (2018) DM-Density 510(k) K170540 clearance letter

Hier steht eine Anzeige.

