

Bildererkennung in der Krebsdiagnostik mit KI

Dr. Tobias Lang, Dr. Ralf Banisch, Sebastian Springenberg

Mindpeak GmbH, Hamburg, Germany

2022

Zusammenfassung

Bildererkennung ist ein wichtiger Bestandteil der Medizin. Histopathologische Gewebeschnitte werden in der Krebsdiagnostik visuell beurteilt. Der Einsatz von Künstlicher Intelligenz bietet großes Potenzial, um Analysen präzise, reproduzierbar und effizient durchzuführen sowie neuartige diagnostische Gewebemuster zu entdecken. Im Vergleich zum herkömmlichen KI-Forschungskontext gelten für die Anwendung von KI zur Bildererkennung in der Pathologie verschiedenartige Herausforderungen. Das zeigt sich darin, dass erst vor kurzem das erste KI-System in den USA in den praktischen Einsatz in der pathologischen Primärdiagnostik gelangt ist.

Dieser Beitrag zeigt den Nutzen und die möglichen Anwendungsfälle von Bildererkennung mit KI in der histopathologischen Krebsdiagnostik auf. Methodische Grundlagen und besondere Herausforderungen werden diskutiert sowie Lösungsansätze und Handlungsempfehlungen für die Entwicklung praxistauglicher KI-Systeme dargestellt.

Einführung

Bildererkennung spielt in der medizinischen Diagnostik eine zentrale Rolle. Dies gilt neben der Radiologie insbesondere für die Krebsdiagnostik: Sie basiert zu einem erheblichen Teil auf der visuellen Beurteilung von histopathologischen Gewebeschnitten. Heutzutage wird diese Aufgabe von menschlichen Expertinnen gelöst. Dabei nehmen repetitive Standardanalysen, die trotz ihrer Häufigkeit ein hohes Maß an Genauigkeit und Konzentration erfordern, einen großen Teil des Arbeitsalltags ein.

Künftig werden die zunehmenden Anforderungen an die histopathologische Krebsdiagnostik nicht von Menschen alleine zu lösen sein: Weltweit steigen die Zahlen von Krebserkrankungen infolge alternder Gesellschaften, während die Anzahl an Expertinnen stagniert. Gleichzeitig bieten kommende individualisierte Therapieformen bessere Erfolgsaussichten, verlangen aber mehr diagnostische Beurteilungen pro Patient.

Der Einsatz von KI ist ein wichtiger Schlüssel, um diesen Herausforderungen zu begegnen. Bilderkennung mit KI kann Expertinnen in der klinischen Diagnostik maßgeblich unterstützen: Beispielsweise können mit KI Tausende Gewebezellen genau lokalisiert und klassifiziert (vgl. Abbildung 1), Metastasen in Lymphknoten automatisch gefunden und Krankheitsverläufe etwa durch KI-gestützte Quantifizierung von Lymphozyteninfiltration besser prognostiziert werden.

Darüber hinaus können mittels KI neuartige diagnostische Muster in Gewebeschnitten gefunden werden, die bisher noch nicht bekannt sind oder für das menschliche Auge heutzutage häufig erst mithilfe aufwendiger Spezialfärbungen sichtbar werden. Solche neuartigen Biomarker könnten die Krebstherapie deutlich verbessern.

Aufgrund seines Potenzials beherrscht das Thema KI in der Pathologie Konferenzen, Fachjournale und Unternehmensselbstdarstellungen. Fachfremden kann es erscheinen, als hätte die KI bereits breiten Einzug in die medizinische Praxis erhalten. Dies ist nicht der Fall. Bei der KI-Bilderkennung in der Histopathologie gelten, im Vergleich zum sonstigen Forschungskontext, besondere anspruchsvolle Herausforderungen im Hinblick auf die Methodik. So ist es einzuordnen, dass beispielsweise erst im Dezember 2021 in den USA das erste KI-Produkt für die Primärdiagnostik in der Routinebefundung in den Einsatz gelangte. Auch in Europa ist die Zahl der tatsächlich in der Praxis eingesetzten KI-Bilderkennungssysteme noch sehr begrenzt. In diesem Artikel werden die damit verbundenen Herausforderungen und davon abgeleitete Handlungsempfehlungen beleuchtet.

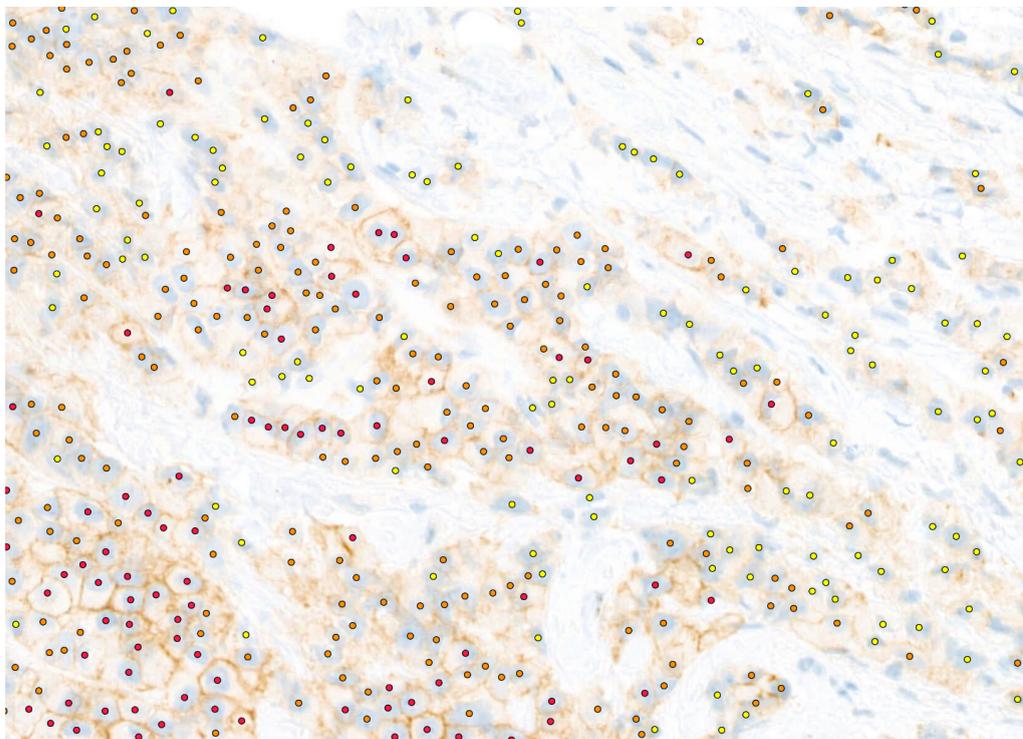


Abbildung 1: Mit KI-Bilderkennung können große Anzahlen an Tumorzellen präzise und schnell bestimmt werden: Hier erkennt die KI fünf Tumorzellklassen (markiert mit Punkten unterschiedlicher Farben) und unterscheidet sie von gefundenen

Nicht-Tumorzellen (nicht markiert) in einem Gewebeschnitt von Brustkrebs (invasivem Mammakarzinom). Verschiedene Anzahlen des Rezeptors HER2 auf den Zellmembranen sind immunhistochemisch hervorgehoben, zu erkennen anhand der Brauntöne. (Quelle: Mindpeak)

Bildererkennung mit KI

Methoden des KI-Teilgebiets Maschinelles Lernen (ML) haben in den vergangenen zehn Jahren die computergestützte Bildererkennung revolutioniert. In traditionellen Verfahren der Bildererkennung geben menschliche Experten dem Computer Regeln vor. Beispielsweise definiert ein Programmierer, welche Größe Tumorzellen haben. Dies funktioniert in der Praxis jedoch selten in ausreichender Güte: Für menschliche Experten ist es meist unmöglich, ihre eigene visuelle Erkennungsleistung in vollständigen und präzisen Regelsystemen abzubilden. Händisch definierte Regeln sind zudem meist nicht adaptiv genug, um Robustheit gegenüber dem Abwechslungsreichtum von Bildern in der Praxis zu erlangen.

Im Gegensatz dazu erlernt ein ML-System Regeln eigenständig ohne menschliche Hilfe anhand von Beispielen. Es erkennt statistisch komplexe diagnostische Muster und kann so anspruchsvolle Bildererkennungsaufgaben lösen. Die wichtigste moderne ML-Methode für die Bildererkennung ist das Deep Learning, das zumeist auf sogenannten tiefen neuronalen Netzen basiert. Dies sind mehrschichtige Berechnungseinheiten, sehr vage inspiriert vom menschlichen Gehirn, die in einer Kaskade von mathematischen Operationen zunehmend abstraktere Bildmerkmale und schließlich das finale Analyseergebnis ermitteln.

Die aktuelle Forschung wird dominiert von Problemstellungen der Bildererkennung auf Alltagsfotos von Menschen, Produkten oder Verkehrsszenen, wie sie sich aus den Anwendungsfeldern im Onlinemarketing etwa der großen Internetwerkeunternehmen wie Google und Meta und aus dem autonomen Fahren ergeben. Bilder von Gewebeschnitten mit vielen Tausend Zellen haben dagegen oftmals andersartige Herausforderungen, wie wir im nächsten Kapitel aufzeigen werden.

Ein ML-System wird für eine bestimmte Anwendung trainiert, indem man im Rahmen von überwachtem Lernen (supervised learning) in Beispielbildern gewünschte Zielstrukturen vorgibt. Zur Entwicklung eines Zellerkennungssystems kann man beispielsweise Gewebebilder mit markierten Beispielzellen bereitstellen. ML hat dabei eine herausfordernde Eigenschaft: Es ist sehr datenhungrig und braucht je nach Anwendung viele Tausend annotierte Beispielbilder. Dies ist nicht unproblematisch, da Datenannotation zeitaufwendig ist und gerade in der Pathologie oftmals nur von Experten umgesetzt werden kann.

An alternativen Lernansätzen, die mit weniger Beispielannotationen auskommen, wird in den letzten Jahren daher verstärkt geforscht. Im so genannten schwachen überwachten Lernen wie dem Multiple

Instance Learning wird ein gröberes Lernsignal bereitgestellt: Für ein Gewebebild wird etwa nur global festgelegt, ob sich darauf Tumor befindet oder nicht; einzelne Tumorzellen werden hingegen nicht markiert. Für diese Lernform ist ein Vielfaches an Beispielen im Vergleich zum normalen überwachten Lernen erforderlich.

Ein vielversprechender weiterer Ansatz ist das selbstüberwachte Lernen. Hier wird gänzlich auf ein vorgegebenes Lernsignal verzichtet. Die Bilderkennungs-KI lernt stattdessen allgemeine Regelmäßigkeiten in Beispielen, etwa typische Strukturen von Zellen. Technisch wird dies durch allgemeingültige Stellvertreteraufgaben erreicht, beispielsweise durch die Aufgabe, zu erkennen, dass ein Bild und seine Rotation zusammengehören. Darauf basierend genügen im Anschluss einige wenige Beispiele im überwachten Lernen, um eine Zielaufgabe zu lösen.

Zunehmende Aufmerksamkeit erfährt die Erklärbarkeit von KI („Explainable AI“). Zwar sind neuronale Netze nicht vollends interpretierbar, doch können Entscheidungen und Analysen nachvollziehbar gemacht und plausibilisiert werden. In Bildern können hierfür die Bereiche markiert werden, die für die KI-Entscheidung maßgeblich sind, oder ähnliche Trainingsbilder zum Vergleich herangezogen werden.

Herausforderungen in der Histopathologie

Bilderkennungsarbeiten in der Medizin können verschiedene Zweckbestimmungen haben: diagnostisch zur Feststellung von Krankheiten, prognostisch zur Vorhersage von Krankheitsverläufen sowie therapeutisch zur Empfehlung von Therapieentscheidungen. Die zugehörigen konkreten Analysen lassen sich methodisch in verschiedene Formalisierungen unterteilen: Segmentierung („Wo auf dem Gewebeschnitt verläuft die Grenze zwischen Tumor und gesundem Gewebe?“); Klassifikation („Welchen Schweregrad hat ein Prostatatumor?“); Objekterkennung („Wie viele Krebszellen gibt es auf diesem Gewebeschnitt?“); Objektlokalisierung („Wie groß ist ein Prostatatumor?“).

Grundlegend in der Krebsdiagnostik ist dabei das Erkennen von Tumorzellen (vgl. Abbildung 2). Dafür gibt es zwei verschiedene Informationsquellen: zum einen die morphologischen Merkmale einzelner Zellen wie Größe, Färbung und Form, zum anderen den Zellkontext. In gesundem Gewebe bilden Zellgruppen funktional ausdifferenzierte Strukturen wie Brustdrüsen, während in krebsbefallenem Gewebe diese Ausdifferenzierung verloren gehen und Zellen durcheinandergewürfelt angeordnet sein können. Zur Bestimmung der morphologischen Zellmerkmale ist eine hohe Bildvergrößerungsstufe notwendig. Dagegen können Gewebeverbände besser auf einer mittleren Vergrößerungsstufe beurteilt werden. Daher ist ein KI-Forschungsschwerpunkt in der Krebsdiagnostik das systematische Zusammenführen von Analysen auf verschiedenen Vergrößerungsstufen in einem Modell.

Eine grundlegende Herausforderung bildet die Bildgröße histopathologischer Gewebeschnitte (sogenannter Whole-Slide-Images). Diese sind häufig 10.000-mal größer als in der sonstigen Bilderkennungs-forschung und enthalten Zehntausende von kleinen Zielstrukturen, nämlich die Zellen. Das Gesamtbild in viele kleine Teilausschnitte ("Patches") zu zerlegen, kann rechenintensiv sein und erfordert Überlegungen hinsichtlich dessen, wie die Teilergebnisse bestmöglich zusammengeführt werden, insbesondere wenn sich Zellverbände über mehrere Ausschnitte erstrecken. Daher werden Methoden zur Vorauswahl geeigneter Teilgebiete für die rechenintensiven Bildanalyseschritte erforscht.

Ein Problem, das in der Krebsdiagnostik lange Zeit unterschätzt wurde, ist die große Vielfalt an Bilddaten. Sie prägt die reale Welt der pathologischen Praxis aufgrund unterschiedlicher Gewebeaufbereitungsschritte, verschiedener Färbereagenzien und Scanning-Hardwaresysteme. Tatsächlich zeigt sich oft, dass ein KI-System aus wissenschaftlichen Studien in der Breite der praktischen Anwendung nur selten eine hinreichende Erkennungsgüte erreicht. Vergleichbar ist das etwa mit dem Unterschied zwischen autonomem Fahren auf einem Parkplatz oder autonomem Fahren in unbekanntem Offroad-Gelände. Neuartige Architekturen von neuronalen Netzen, eine heterogene hochqualitative Datenannotation und die Übertragung neuester Forschungserkenntnisse – wie z. B. aus dem Bereich des selbstüberwachten Lernens – auf die Pathologie können helfen, robuste ML-Systeme für die Breite zu entwickeln.

Für das Training der KI müssen Gewebebilder annotiert werden, etwa indem Tumorregionen eingezeichnet werden. Im Gegensatz zu anderen Domänen erfordern Annotationen in der Medizin Expertenwissen. Somit ist es schwierig und teuer, Daten zu erhalten, da hierfür eine Vielzahl an Spezialisten erforderlich ist. Die Kompetenz der Annotateure muss durch Eignungstests gewährleistet werden, etwa mittels eines vorgehaltenen Test-Datensatzes, für den von Experten die Grundwahrheit eindeutig bestimmt wurde. Der Annotationsprozess muss eindeutig definiert und überwacht werden, um eine dauerhafte konsistente Güte der Daten zu erzielen.

Eine grundlegende Herausforderung in der Krebsdiagnostik ist das oftmalige Fehlen einer eindeutigen Grundwahrheit. Nicht immer ist es eindeutig zu bestimmen, ob eine Zelle eine Tumorzelle ist. Inter- und Intraobserver-Varianzen, die aussagen, wie sehr die Meinungen verschiedener Experten (inter) bzw. desselben Experten über die Zeit (intra) voneinander abweichen, sind nicht gering. Die Datensammlung muss solche Ungenauigkeiten abbilden.

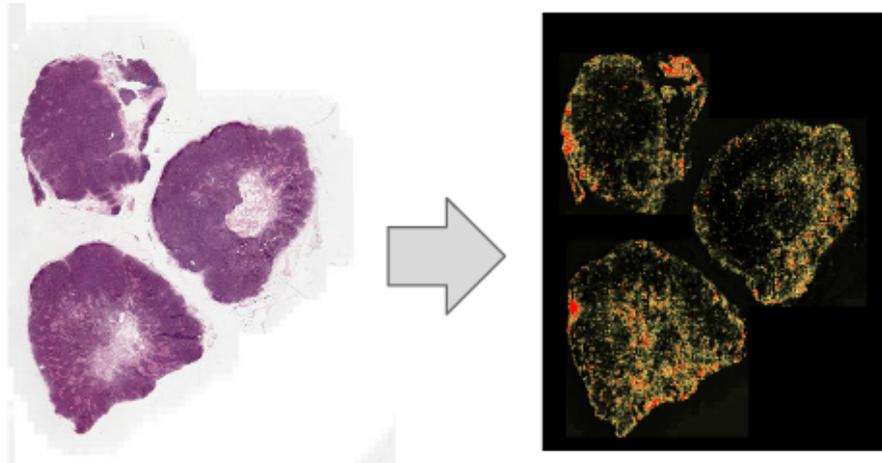


Abbildung 2: Mit KI-Unterstützung können Tumorzellen in großflächigen Gewebebildern effizient gefunden werden: In diesem Beispiel werden auf einem H&E-Lymphknoten-Schnitt (links) in einer Wahrscheinlichkeitskarte (rechts) die Gewebebereiche hervorgehoben, in denen die KI mit hoher Konfidenz eine große Anzahl an Tumorzellen gefunden hat. Die KI-Bildererkennung findet auf einem erheblich vergrößerten Bild statt, das ausgedruckt mehreren Metern entspräche. (Quelle: Mindpeak)

Handlungsempfehlungen

Vor Aufnahme eines KI-Projekts zur Lösung einer medizinischen Diagnostikaufgabe gilt es zu klären, ob Deep Learning (DL) ein geeigneter Ansatz ist. Zuvorderst müssen Daten ausreichend verfügbar und repräsentativ für die spätere Anwendung sein. In der Pathologie ist DL deshalb besonders geeignet für die Analyse häufiger Tumorarten. Es muss möglich sein, die Zieldiagnose mit der erwünschten Genauigkeit allein aus den Bilddaten zu erstellen, ohne dass Kontextwissen mitzubersichtigen ist. Für den praktischen KI-Einsatz muss die notwendige technische Infrastruktur im medizinischen Kontext umsetzbar sein (Rechenleistung, Austausch von Daten, Schnittstellen zu Laborsystemen).

Zu Projektbeginn ist eine klare quantitative Zieldefinition für das KI-System notwendig. Dafür sind Metriken wie Genauigkeit oder Sensitivität zur Beurteilung der Erkennungsleistung festzulegen. In der Praxis gibt es hierbei oftmals subtile Effekte: Bestimmte Diagnosefehler wie falsch-negative Ergebnisse können aus Anwendersicht unerwünschter sein als andere; oder nicht alle Tumorzellklassen haben die gleiche Wichtigkeit in der späteren Anwendung. Ein hochqualitativer Testdatensatz ist unverzichtbar, um die Erkennungsleistung verschiedener Entwicklungsversionen des KI-Systems verlässlich zu beurteilen. Insgesamt bilden Testdaten und Evaluationsmetriken die Grundlage, von der sich die gesamte weitere KI-Entwicklung ableitet.

Messungen von Inter- und Intraobserver-Varianzen von menschlichen Expertinnen helfen, die im praktischen Einsatz zu erzielende Mindestgenauigkeit des KI-Systems festzulegen und allgemein eine Abschätzung der Schwierigkeit des Lernproblems und der zugehörigen Datensammlung zu erhalten.

Zu Projektbeginn ist es ratsam, mit einer allgemein etablierten Wahl der Problemformalisierung und der Architektur des neuronalen Netzes zu starten. Eine solche "Baseline" hilft als Referenzmodell bei der späteren Exploration komplexerer Verfahren. Die Baseline-Entwicklung erweist sich oftmals als anspruchsvoller als erwartet. Sie zeigt erste Herausforderungen in Form einer mangelnden Datenqualität und von Lücken im Problemverständnis. Es ist ein häufiger Fehler, bei anfänglich schlechten Ergebnissen zu voreilig auf einen komplexeren Ansatz zu wechseln, ohne ein hinreichendes Problemverständnis entwickelt zu haben.

Bei der Wahl eines geeigneten komplexeren Ansatzes kann die große Anzahl an KI-Publikationen in der Krebsdiagnostik verwirren. Dabei ist davon auszugehen, dass infolge der hohen Datenvarianzen ein großer Teil der in der Literatur beschriebenen Ansätze mit den eigenen Daten nicht ohne Weiteres funktionieren wird. Für Anfänger kann es hilfreich sein, verstärkt Ansätze und Ideen in den Blick zu nehmen, für die mehrere unabhängige Autorengruppen erfolgreiche Ergebnisse erzielen konnten.

Erfolgreiche KI-Entwicklung ist „hypothesengetrieben“ (Andrew Ng). Sie bedarf der „Scientist“-Fähigkeiten, die zum Berufsbild des Data Scientists gehören. Unerfahrene KI-Entwicklerinnen experimentieren häufig eher zufällig mit einer Vielzahl unterschiedlicher Architekturen in der Hoffnung, dass einer der Ansätze ihr Problem lösen wird, oder sie setzen auf eine Vergrößerung des Datensatzes ohne Systematik in der Sammlung. Stattdessen ist es für eine erfolgreiche KI-Entwicklung unerlässlich, sich zielgerichtet dem Problemverständnis im Kern zu nähern. Dazu muss man Hypothesen über die eigenen Annahmen zum Lernproblem und seine Lösung aufstellen und diese systematisch validieren. Beispielsweise kann man untersuchen, weshalb bestimmte Zellklassen besonders schlecht erkannt werden, welche Auswirkungen das Weglassen der Bilder einer Bildquelle hat oder wie Lernratenparameter die Optimierung einzelner Module des neuronalen Netzes beeinflussen.

Eine datenzentrische KI-Entwicklung („Data-centric AI“) ist in der Krebsdiagnostik von grundlegender Bedeutung. Dabei werden die Bilddatensätze systematisch und kontinuierlich weiterentwickelt, anstatt unsystematisch nur die Bildsammlung stets weiter zu vergrößern. Die Qualität von Annotationen wird regelmäßig überprüft und Lücken im Datenbestand etwa für bestimmte Tumorarten werden geschlossen. Der Aufwand für diese Arbeit wird in der Regel unterschätzt. Bei vielen erfolgreichen Projekten ändern sich die Architekturen von neuronalen Netzen eher selten, die Daten aber kontinuierlich. Um mit der Herausforderung einer fehlenden Grundwahrheit umzugehen, kann es empfehlenswert sein, Bilder von mehreren Experten annotieren zu lassen und nur solche mit Übereinstimmungen zu verwenden. Um die Annotationsqualität regelmäßig zu überprüfen, können in den Annotationsprozess Bilder mit bereits bekannter Grundwahrheit eingestreut werden.

Ein wichtiges Werkzeug für eine erfolgreiche KI-Entwicklung sind Tools und Softwares zur Versionsverwaltung von Datenbeständen, Reproduzierbarkeit von Experimenten und Veröffentlichung der KI-Modelle als Produkte. Während die Versionsverwaltung für Softwarecode etabliert ist, entwickeln sich sogenannte "MLOps"-Tools erst seit einigen wenigen Jahren. Die Auswahl und Einrichtung wirksamer Werkzeuge ist aufwendig, aber essenziell für zielgerichtete und effiziente KI-Entwicklung.

Fazit und Ausblick

Der Einsatz von KI-Unterstützung in der Bilderkennung kann die histopathologische Krebsdiagnostik effizienter und reproduzierbar werden lassen und somit die Patientenversorgung vor dem Hintergrund der stark steigenden Zahl von Diagnosen verbessern und beschleunigen.

Interessanterweise verfügen Mensch und KI über komplementäre Stärken. Während der Mensch über Kontextwissen verfügt, kann die KI klar abgegrenzte Arbeiten ohne Konzentrationsverlust in großer Zahl ausführen. Die Verbindung von Pathologin und KI kann hierbei mehr Nutzen schaffen, als für jede allein möglich wäre. Gleichmaßen kann die Grundlagenforschung in der Pathologie von KI durch die Entdeckung neuartiger klinisch relevanter histopathologischer Muster profitieren.

Bei der Entwicklung von robuster und leistungsstarker KI gibt es besondere Herausforderungen, die in der allgemeinen KI-Forschung nur wenig Berücksichtigung finden. In den nächsten Jahren sind hier durch vermehrte Forschungsanstrengungen große Fortschritte zu erwarten. Für die KI-Anwenderin ist dabei immer ein gesundes Maß an Skepsis gegenüber publizierten Ergebnissen geboten. Letztlich werden umfangreiche Validierungsstudien und der praktische Einsatz über die Geeignetheit von KI-Systemen in der Krebsdiagnostik entscheiden.