

ICS 35.240.80

**Leitfaden für die Entwicklung von Deep-Learning-Bilderkennungssystemen
in der Medizin;
Text Deutsch und Englisch**

Guideline for the development of deep learning image recognition systems in medicine;
Text in German and English

Guide pour le développement de systèmes de reconnaissance d'images d'apprentissage
profond en médecine;
Texte en allemand et anglais

Gesamtumfang 35 Seiten

Dieses Dokument wurde durch die im Vorwort genannten Verfasser erarbeitet und verabschiedet.

Inhalt

	Seite
Vorwort	3
Einleitung	4
1 Anwendungsbereich.....	6
2 Normative Verweisungen	6
3 Begriffe	6
4 Allgemeine Anforderungen.....	9
4.1 Allgemeines	9
4.2 Qualitätsanforderungen.....	9
4.3 Repräsentativer Datensatz zur Abbildung des Problems.....	9
4.4 Datensammlung	10
4.5 Einbettung in den medizinischen Arbeitsablauf.....	11
4.6 Methodik.....	12
4.7 Erklärbarkeit und Nutzervertrauen	12
5 Technische Umsetzung	13
5.1 Allgemeines	13
5.2 Formalisierung und Problemdefinition	13
5.3 Datenakquisitionsprozess und Erstellung des Datensatzes	13
5.4 Software-Entwicklung.....	14
5.5 Trainieren des DL-Systems	14
5.6 Modell-Deployment	14
5.7 Qualitätssicherung des Deep-Learning-Bilderkennungssystems	15
5.8 Kontinuierliche Verbesserung und Nachlernen	15
Literaturhinweise.....	16

Vorwort

Diese DIN SPEC wurde nach dem PAS-Verfahren erarbeitet. Die Erarbeitung von DIN SPEC nach dem PAS-Verfahren erfolgt in DIN SPEC (PAS)-Konsortien und nicht zwingend unter Einbeziehung aller interessierten Kreise.

Die Erarbeitung und Verabschiedung des Dokuments erfolgten durch die nachfolgend genannten Initiatoren und Verfasser:

- MindPeak GmbH: Felix Faber, Nora Bartels, Marc Päpper und Dr. Tobias Lang
- PSIORI GmbH: Dr. Sascha Lange, Dr. Christian Kaul und Lars Eickmeier
- FUSE-AI GmbH: Matthias Steffen und Dr. Sabrina Reimers-Kipping
- Hochschule Düsseldorf: Prof. Dr.-Ing. Thomas Zielke
- Quality Match GmbH: Dr. Daniel Kondermann
- IABG mbH: Bastian Bernhardt

Für dieses Thema bestehen derzeit keine Normen im Deutschen Normenwerk.

DIN SPEC (PAS) sind nicht Teil des Deutschen Normenwerks.

Für diese DIN SPEC wurde kein Entwurf veröffentlicht.

Trotz großer Anstrengungen zur Sicherstellung der Korrektheit, Verlässlichkeit und Präzision technischer und nicht-technischer Beschreibungen kann das DIN SPEC (PAS)-Konsortium weder eine explizite noch eine implizite Gewährleistung für die Korrektheit des Dokuments übernehmen. Die Anwendung dieses Dokuments geschieht in dem Bewusstsein, dass das DIN SPEC (PAS)-Konsortium für Schäden oder Verluste jeglicher Art nicht haftbar gemacht werden kann. Die Anwendung der vorliegenden DIN SPEC (PAS) entbindet den Nutzer nicht von der Verantwortung für eigenes Handeln und geschieht damit auf eigene Gefahr.

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. DIN ist nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren.

Die kostenfreie Bereitstellung dieses Dokuments als PDF-Version über den Beuth WebShop wurde im Vorfeld finanziert.

Aktuelle Informationen zu diesem Dokument können über die Internetseiten von DIN (www.din.de) durch eine Suche nach der Dokumentennummer aufgerufen werden.

Einleitung

Bilderkennung spielt in vielen Feldern der Medizin eine zentrale Rolle, wie etwa in der Radiologie und der Histopathologie. Diese Bilderkennung wird heute von menschlichen Experten geleistet. Gerade repetitive Standard-Bilderkennungsaufgaben, die eine hohe Genauigkeit und komplexe Auswertungen erfordern, nehmen dabei einen erheblichen Teil des Arbeitstages von Ärzten ein.

Automatisierte Bilderkennung mit Deep Learning bietet enormes Potenzial für Effizienz- und Qualitätsgewinne in der Medizin [1]. Deep Learning bildet den neuen State-of-the-Art der Bilderkennung und wird bereits mit Erfolg in verschiedenen Feldern von automatisiertem Fahren bis zur Gesichtserkennung angewendet [2]. Deep Learning ist eine Methode der Künstlichen Intelligenz (KI) und basiert auf tiefen künstlichen neuronalen Netzen. Anstatt mit viel Aufwand fest codierte Regeln durch einen menschlichen Experten festzulegen, werden mit einem Deep-Learning-System (DL-System) statistische Muster aus Beispieldaten gelernt. Diese Technik erlaubt es, auch in der Medizin bisher automatisiert nicht-lösbare komplexe Aufgaben in der Bilderkennung künftig mit Computerunterstützung zu lösen.

Bilderkennungssysteme mit Deep Learning können potenziell in folgenden medizinischen Feldern angewendet werden:

- Radiologie;
- Histopathologie;
- Augenheilkunde;
- Hämatologie;
- Mikrobiologie;
- Dermatologie;
- Tiermedizin;
- Pharmaindustrie;
- Forschung (z. B. Alzheimer-Krankheit);
- Chirurgie (z. B. operierende Roboter müssen auch „sehen“ können).

Bilderkennungsaufgaben in der Medizin können verschiedene Zweckbestimmungen haben:

- a) diagnostisch zur Feststellung von Krankheiten;
- b) prognostisch zur Vorhersage von Krankheitsverläufen;
- c) therapeutisch zur Empfehlung von Therapieentscheidungen.

Dabei nimmt der Einfluss des DL-Systems auf den Patienten von diagnostisch über prognostisch zu therapeutisch zu.

Bilderkennung kann in verschiedenen Automatisierungsgraden eingesetzt werden. Bei voll automatisiertem Einsatz wird das Ergebnis nicht mehr von einem menschlichen Experten überprüft. Ein voll automatisierter Einsatz liegt auch dann vor, wenn bei einer hohen Anzahl von Auswertungen nur die mit Unsicherheiten behafteten Ergebnisse von einem menschlichen Experten überprüft werden. Alternativ kann ein DL-System unterstützend eingesetzt werden: ein menschlicher Experte überprüft jedes Ergebnis und ist für die finale Entscheidungsfindung allein verantwortlich. Diese Vorgehensweise hat insbesondere bei Screening-Programmen eine große Bedeutung. Bei Mammographie-Screening konnte gezeigt werden, dass der Einsatz eines KI-Systems den Aufwand für eine Überprüfung durch menschliche Experten um über 80 % reduzieren kann, ohne dass die Erkennungsraten verringert wurden [3]. Außerdem ist es möglich, dass ein DL-System parallel und unabhängig vom Experten im Hintergrund eine automatisierte KI-Zweitmeinung erstellt.

Die Nutzer von Bilderkennungssystemen können sowohl Ärzte als auch Patienten sein. Hämatologen können zum Beispiel bei der visuellen Bestimmung des Blutbilds durch ein DL-System unterstützt werden. Patienten können Hautveränderungen durch eine Mobiltelefon-App screenen lassen oder sich für Röntgenbilder online automatisierte Zweitmeinungen einholen.

Für den praktischen Einsatz von DL-Systemen sind verschiedene Szenarien denkbar: zur Diagnose; zur Triagierung, bei der das DL-System Fälle entsprechend der Analyse-Ergebnisse für den Menschen priorisiert; zum Ausfiltern von eindeutig negativen Fällen; für eine automatisierte Zweitmeinung, die als Sicherheitsnetz die Erstmeinung des menschlichen Experten überprüft und gegebenenfalls einen Warnhinweis gibt.

Die konkreten Bilderkennungsaufgaben lassen sich technisch in folgende Problemarten unterteilen:

- a) Klassifikation von Bildern (z. B. Ist auf dem Röntgenbild eine Knochenfraktur zu erkennen?);
- b) Segmentierung (z. B. Wo befindet sich Tumorgewebe?);
- c) Objekterkennung (z. B. Gibt es Malaria-Erreger im Blutaussstrich?);
- d) Objektlokalisierung (z. B. Wo befinden sich Malaria-Erreger im Blutaussstrich?).

1 Anwendungsbereich

Dieses Dokument gibt die Anforderungen an, unter denen Bilderkennungsprobleme in der Medizin mit Hilfe eines Deep-Learning-Bilderkennungssystems bearbeitet werden können. Es erlaubt Entscheidungsträgern, Kenntnisse über die Anwendungsmöglichkeiten eines Deep-Learning-Bilderkennungssystems in der Medizin und seine Struktur zu gewinnen.

Mit Hilfe dieses Dokumentes kann die Schätzung des Aufwandes und des Nutzens eines Deep-Learning-Bilderkennungssystems unterstützt werden sowie eine genauere Erfolgsprognose erstellt werden.

Dieses Dokument gibt Leitlinien zur praktischen Entwicklung eines Deep-Learning-Bilderkennungssystems in der Medizin vom Vorgehen bei der Datensammlung über die Strukturierung der Daten zum Lernen der KI-Bilderkennung bis zur Ablaufstruktur von Lern-Experimenten, insbesondere mit Rücksicht auf die erhöhten Qualitätsmaßstäbe und regulativen Vorgaben in der Medizin.

Dieses Dokument ist insbesondere für Hersteller von DL-Systemen und die Beteiligten an Forschungs- und Entwicklungsprojekten zum Einsatz von Deep-Learning-Bilderkennungssystemen in der Medizin.

Dieses Dokument legt keine spezifischen Angaben zu aktivem Lernen, mentalem Lernen, automatischem kontinuierlichem Lernen und dem bestimmungskonformen Einsatz des DL-Systems in der Praxis fest.

2 Normative Verweisungen

Die folgenden Dokumente werden im Text in solcher Weise in Bezug genommen, dass einige Teile davon oder ihr gesamter Inhalt Anforderungen des vorliegenden Dokuments darstellen. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

DIN SPEC 13266:2020-04, *Leitfaden für die Entwicklung von Deep-Learning-Bilderkennungssystemen*

3 Begriffe

Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

DIN und DKE stellen terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen bereit:

- DIN-TERMinologieportal: verfügbar unter <https://www.din.de/go/din-term>
- DKE-IEV: verfügbar unter <http://www.dke.de/DKE-IEV>

ISO und IEC stellen terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen bereit:

- ISO Online Browsing Platform: verfügbar unter <https://www.iso.org/obp>
- IEC Electropedia: verfügbar unter <http://www.electropedia.org/>

3.1 Bilderkennung
Erfassung und Analyse eines Bildes, der Objekte, aus denen es besteht, sowie deren Eigenschaften und räumlichen Beziehungen durch eine Funktionseinheit

Anmerkung 1 zum Begriff: Bilderkennung schließt Szenenanalyse ein.

[QUELLE: DIN ISO/IEC 2382-28:1998-04, 28.01.14]

3.2**Bilderkennungsproblem**

Problem ein bestimmtes Ziel der Bilderkennung zu lösen, z. B. ein Bild zu klassifizieren

[QUELLE: DIN SPEC 13266:2020-04, 3.2]

3.3**Deep Learning****DL**

Lernen statistischer Muster durch tief verschachtelte neuronale Netze

[QUELLE: DIN SPEC 13266:2020-04, 3.4, modifiziert — Ein zusätzlicher Begriff wurde ergänzt.]

3.4**Deep-Learning-Architektur****DL-Architektur**

Beschreibung der verschiedenen genutzten Recheneinheiten des neuronalen Netzes sowie deren Verbindung

Anmerkung 1 zum Begriff: Aus der Architektur ergibt sich die Anzahl der lernbaren Gewichte im Netzwerk. Je höher diese Anzahl, desto schwierigere Probleme sind lösbar; aber desto mehr Daten sind dazu notwendig.

[QUELLE: DIN SPEC 13266:2020-04, 3.5]

3.5**Deep-Learning-Bilderkennungssystem****DL-System**

das zu entwickelnde Computermodell, welches das Bilderkennungsproblem löst

[QUELLE: DIN SPEC 13266:2020-04, 3.6]

3.6**Domäne**

spezifischer Wissensbereich oder bereichsspezifisches Expertenwissen; genauer der Bereich, in dem das DL-Modell eingesetzt wird

[QUELLE: DIN SPEC 13266:2020-04, 3.7]

3.7**Grundwahrheiten**

zu den Daten zugehörige ideale Ausgabewerte

[QUELLE: DIN SPEC 13266:2020-04, 3.8]

Anmerkung 1 zum Begriff Grundwahrheiten sind z. B. Klassifikationen (Bild 1 — Hund, Bild 2 — Katze).

3.8**Initialdaten**

Gesamtmenge der Daten, die zum Start der Entwicklung des DL-Systems zur Verfügung stehen oder gesammelt werden, und die in Trainings-, Test- und Validierungsdaten aufgeteilt werden

Anmerkung 1 zum Begriff: In der Regel bestehen Initialdaten aus Bildern und den zugehörigen Grundwahrheiten/Annotationen.

[QUELLE: DIN SPEC 13266:2020-04, 3.9]

3.9

künstliche Intelligenz

KI

Fähigkeit einer Funktionseinheit, solche Funktionen auszuführen, die im allgemeinen menschlicher Intelligenz zugeordnet werden, wie z. B. Schlussfolgern und Lernen

[QUELLE: DIN ISO/IEC 2382-28:1998-04, 28.01.02]

3.10

Modell

DL-Modell

neuronaes Netz, das mit den Trainingsdaten trainiert wird, um das Bilderkennungproblem zu lösen

[QUELLE: DIN SPEC 13266:2020-04, 3.11]

3.11

neuronaes Netz

künstliches neuronaes Netz

Computermodell unter Verwendung verteilter und paralleler Verarbeitung vor Ort, das aus einem Netzwerk einfacher als künstliche Neuronen bezeichneter Verarbeitungselemente, die ein komplexes umfassendes Verhalten darstellen können, besteht

[QUELLE: DIN ISO 18115-1:2017-07, 8.1, modifiziert — Begriff „ANN“ und alle Anmerkungen zum Begriff wurden nicht übernommen.]

3.12

Qualitätssicherungsdaten

Datensatz bestehend aus einer Teilmenge des Testdatensatzes sowie hinzukommender Daten während des Live Betriebs des Systems; der Datensatz kann sich mit der Zeit verändern

[QUELLE: DIN SPEC 13266:2020-04, 3.13]

3.13

Trainingsdaten

Teilmenge der Initialdaten, die dazu dient unterschiedliche Modelle zu trainieren

[QUELLE: DIN SPEC 13266:2020-04, 3.14]

3.14

Testdaten

Teilmenge der Initialdaten, mit deren Hilfe die Analysequalität des finalen Modells evaluiert werden kann

[QUELLE: DIN SPEC 13266:2020-04, 3.15, modifiziert — „das finale Modell ausgewählt“ wurde durch „die Analysequalität des finalen Modells evaluiert“ ersetzt.]

3.15

Validierungsdaten

Teilmenge der Initialdaten, die dazu dient unterschiedliche auf den Trainingsdaten trainierte Modellvarianten zu validieren

[QUELLE: DIN SPEC 13266:2020-04, 3.16]

4 Allgemeine Anforderungen

4.1 Allgemeines

DIN SPEC 13266:2020-04, Abschnitt 5 bis Abschnitt 7 müssen angewendet werden.

ANMERKUNG 1 Anwendungen in der Medizin zeichnen sich durch eine Vielzahl an Besonderheiten aus.

DL-Systeme in der Medizin können Medizinprodukte oder Teile davon sein.

ANMERKUNG 2 Aktuell gibt es weder in der EU noch in den USA allgemeingültige Gesetze und Normen, die DL-Systeme regulieren [4].

Für Medizinprodukte muss eine Zweckbestimmung festgelegt werden (z. B. diagnostisch oder prognostisch).

Die Produktentwicklung sollte auf etablierten Managementsystemstandards beruhen (z. B. DIN EN ISO 9001, DIN EN ISO/IEC 27001, DIN EN ISO/IEC 27701, DIN EN ISO 13485); allerdings decken solche Standards, soweit sie zur Zeit veröffentlicht sind, lediglich Teilbereiche der Entwicklung und Nutzung von KI ab (Qualität, Sicherheit, Datenschutz, usw.).

Das Risikomanagement sollte nach DIN EN ISO 14971 erfolgen und muss sowohl Risiken in der Produktentwicklung als auch Risiken hinsichtlich falscher Analyseergebnisse beachten. Die vorgesehenen Nutzer sowie der genaue Anwendungskontext müssen bestimmt werden. Es muss festgelegt werden, für die Bilder welcher Hardware das DL-System ausgelegt ist und bei welchen Eingabebildern keine Erkennungsgüte des DL-Systems garantiert werden kann.

4.2 Qualitätsanforderungen

Die Qualitätsanforderungen sind in der Medizin besonders hoch, da das Wohl von Patienten sichergestellt werden muss. Falsche Vorhersagen wie zum Beispiel Falsch-Negative bei der Diagnose von Krebsfällen haben außerordentlich hohe Kosten.

Um Nutzerakzeptanz zu finden, muss die Erkennungsgüte mindestens auf dem Niveau menschlicher Experten sein. Diese Erkennungsgüte muss in klinischen Studien validiert worden sein (in retrospektiven oder prospektiven Studien) [5].

Die Analyseergebnisse von DL-Systemen müssen reproduzierbar sein.

DL-Systeme müssen verlässlich sein. Die Systeme dürfen im Klinikalltag nicht ausfallen, da sonst der darauf angepasste klinische Arbeitsablauf (en: workflow) bedroht sein kann.

Die Nutzerakzeptanz ist in der Medizin besonders entscheidend; dazu muss die KI-Bilderkennung sich nahtlos in medizinische Arbeitsabläufe einbinden. Außerdem müssen die Vorhersagen nachvollziehbar und erklärbar sein, damit sich menschliche Experten auf die Ergebnisse verlassen können.

Das Qualitätsmanagement sowie Prozessprüfungen sollten auf etablierten Managementsystemstandards beruhen (z. B. DIN EN ISO 9001, DIN EN ISO/IEC 27001, DIN EN ISO/IEC 27701, DIN EN ISO 13485).

4.3 Repräsentativer Datensatz zur Abbildung des Problems

Eine große Herausforderung in der Bilderkennung in der Medizin ist die Heterogenität der Bilddaten. Bilddaten werden mit Geräten und Softwares verschiedener Hersteller erstellt.

ANMERKUNG 1 Beispielsweise gibt es in der Pathologie verschiedene Scanning-Systeme für pathologische Proben von verschiedenen Herstellern. Die verschiedenen Scanning-Systeme und Softwares führen zu verschiedenartigen Bildcharakteristiken, etwa in Kontrast und Schärfe. Selbst bei Systemen desselben Modelltyps gibt es häufig Varianzen. Weiterhin gibt je nach Krankenhaus oder Labor unterschiedliche Bedingungen, unter denen die Bilder entstehen:

beispielsweise können die Lichtverhältnisse variieren oder die chemischen Färbemittel bei pathologischen Proben haben unterschiedliche Zusammensetzungen.

Krankenhäuser und Labore haben oft unterschiedliche Zusammensetzungen von Patientengruppen, die sie versorgen. Labore können auf bestimmte Krankheiten spezialisiert sein. Dies führt zu unterschiedlichen Datenverteilungen an Krankheitsbildern. Bei der Datensatzerstellung muss berücksichtigt werden, dass durch die Laborauswahl kein Bias in die Datenerhebung kommt.

ANMERKUNG 2 Allgemein spielt in der Medizin die Heterogenität der Patienten eine Rolle, so dass die resultierenden Bilder für die Bilderkennung häufig starke individuelle Varianzen aufweisen.

Bei der DL-Entwicklung muss analysiert werden, welchen Einfluss Hardware- und Software-Beschaffenheiten auf die Datengüte und die Analyseergebnisse haben.

Der Hersteller des DL-Produkts zur Bildverarbeitung muss festlegen, für welche Typen von Eingabebildern und Bildquellen Vorhersagen gemacht werden können. Für eine große Anwendungsbreite ist in der Regel ein entsprechend breiter und repräsentativer Trainingsdatensatz notwendig.

4.4 Datensammlung

Die Datengewinnung für die KI-Systementwicklung ist in der Medizin eine besonders große Herausforderung, da hierfür medizinisches Fachwissen und hochqualitative Daten notwendig sind und sich medizinische Daten durch eine große Heterogenität auszeichnen. Die Qualität der gewonnenen Daten ist von herausragender Bedeutung, da nur mit solchen Daten eine hohe Erkennungsleistung erreicht werden kann.

Zur Erzeugung eines geeigneten Datensatzes muss sichergestellt werden, dass die Auswahl der Daten repräsentativ für die Anwendung sind, dass die Qualität sowohl der Daten als auch der Annotationen hinreichend sind, und dass die gewählten Annotationsregeln möglichst eindeutig sind. Eine detaillierte Diskussion des Themas Datensatzqualität befindet sich in DIN SPEC 13266.

Ein DL-System wird typischerweise mit überwachtem Lernen trainiert. Auch andere DL-Trainingsmethoden sind anwendbar.

ANMERKUNG 1 Es ist möglich, durch den zusätzlichen Einsatz von unüberwachtem Lernen geeignete Feature-Repräsentationen für das überwachte Lernen zu lernen.

Das Training mit überwachtem Lernen erfordert Annotationen (Labels) der Daten. Im Gegensatz zu anderen Domänen müssen Annotationen in der Medizin mit Expertenwissen erstellt werden.

ANMERKUNG 2 Somit ist es schwierig und teuer, Daten zu erhalten, da hierfür eine Vielzahl an Spezialisten erforderlich ist.

Die Kompetenz von Annotatoren muss durch Eignungstests sichergestellt werden. Dies kann mittels eines vorgehaltenen Testdatensatzes geschehen, für den von Experten die Grundwahrheit eindeutig bestimmt wurde.

Der Annotationsprozess muss genau definiert und überwacht werden, um eine konsistente Güte der Daten zu erreichen. Dazu müssen Annotatoren entsprechend geschult werden. Die Arbeit der Annotatoren muss konstant überwacht werden, um die Qualität zu sichern. Dies kann beispielsweise dadurch geschehen, dass in den Annotationsprozess bereits bekannte Bilder eingestreut werden, für die man die Analyse des Annotators mit der bereits bekannten Grundwahrheit vergleicht. Für eine hohe Qualität ist es wichtig, dass die Incentivierung der Annotatoren nicht falsche Anreize setzt.

BEISPIEL Ein falscher Anreiz kann eine Vergütung pro abgeschlossenem Fall sein.

Eine weitere Schwierigkeit in der Medizin ist, dass häufig eine eindeutige Grundwahrheit fehlt. Die Datensammlung muss solche Ungenauigkeiten abbilden.

ANMERKUNG 3 Beispielsweise können in der Histopathologie oftmals Zellen nicht mit eindeutiger Gewissheit klassifiziert werden, da diese Information in den Bilddaten fehlt.

Um eine hohe Datenqualität sicherzustellen, ist es häufig notwendig, mehrere Annotatoren dieselben Daten annotieren zu lassen. Es sollten dann nur solche Daten zum Trainieren und Testen verwendet werden, bei denen die Annotatoren übereinstimmen.

Für die Definition des Annotationsprozesses müssen Abschätzungen zur Inter- und Intraobserver-Variabilität gewonnen werden. Diese sagen aus, wie sehr die Meinungen verschiedener Annotatoren (inter) bzw. wie sehr die Meinung desselben Annotators zu verschiedenen Zeitpunkten (intra) abweicht. Damit erhält man eine Abschätzung der Schwierigkeit des Annotationsproblems und damit der Analyseaufgabe. Außerdem kann man abschätzen, wie viele Experten man zur Annotation braucht.

Der Annotationsprozess inklusive des Qualitätssicherungsprozesses muss detailliert dokumentiert werden. Eine Dokumentation muss insbesondere enthalten:

- a) Informationen zu Ausbildung, relevanten Qualifikationen und Hintergrundwissen des Annotators;
- b) Informationen zu Datenschutzmaßnahmen und IT-Sicherheit;
- c) angewendete Incentivierungsmaßnahmen;
- d) Instruktionen für die Annotatoren in der Form von mindestens Tool- und Annotations-Instruktionen);
- e) Dokumentation des Annotationstools (z. B. Screenshots, Bildschirmaufnahmen);
- f) Methodik zur Überprüfung der Korrektheit der Annotationen (z. B. Clustering, Wiederholungen);
- g) Methoden zur Qualitätssicherung, z. B. Methode der Aggregation mehrfach wiederholter Annotationen durch unterschiedliche Annotatoren und Überwachung der Annotatoren.

Datenschutz und Anonymisierung bzw. Pseudonymisierung spielen in der Medizin eine herausragende Rolle. Hierfür muss definiert und dokumentiert werden, ob für das DL-System Meta-Daten in Form persönlicher Patientendaten vonnöten sind.

Bei der Datensammlung muss definiert werden, für welche Eingabedaten (Formate, Bildgrößen, Farbkodierungen usw.) das DL-System ausgelegt sein muss.

Bei der Datensammlung muss darauf geachtet werden, einen systematischen Bias zu vermeiden sowie so genannte Datenleckagen auszuschließen, bei denen korrelierte, aber nicht kausale Daten vom DL-System fälschlicherweise als Grundlage der Bildanalyse gelernt werden (z. B. Bilder mit der Klasse "Krebs", bei denen zusätzlich immer ein Lineal zu sehen ist).

4.5 Einbettung in den medizinischen Arbeitsablauf

Veränderungen in medizinischen Arbeitsprozessen sind besonders aufwendig. Vor der Entwicklung eines DL-Systems muss geklärt werden, ob die Veränderungen zum Einsatz eines DL-Systems und die Eingliederung in klinische Prozesse realistisch sind.

Die Resultate der KI-Bildererkennung müssen für die Anwender leicht verständlich sein. Es dürfen keine Unklarheiten entstehen, wie mit den Ergebnissen umgegangen wird.

ANMERKUNG Beispielsweise führen Wahrscheinlichkeitsangaben etwa in Form von Wahrscheinlichkeits-Heatmaps häufig zu Unklarheiten für Anwender.

Der Anbieter des DL-Systems muss sicherstellen, dass das System konstant verfügbar im medizinischen Arbeitsablauf ist.