



Strahlenschutzkommission

Geschäftsstelle der
Strahlenschutzkommission
Postfach 12 06 29
D-53048 Bonn
<http://www.ssk.de>

**Anwendung von künstlicher Intelligenz bei der
Bildrekonstruktion und -verarbeitung in Radiologie und
Nuklearmedizin**

Empfehlung der Strahlenschutzkommission

Verabschiedet in der 339. Sitzung der Strahlenschutzkommission am 27./28. Oktober 2025

Vorwort

Künstliche Intelligenz (KI) befasst sich mit der Entwicklung von Maschinen, die Aufgaben erledigen sollen, die bisher eine menschliche Intelligenz erfordern. Ein wichtiger Teilbereich der KI ist das maschinelle Lernen (ML), welches es Maschinen ermöglicht, Muster in Datenmengen zu erkennen. Unter deep learning (DL) wird dann das maschinelle Lernen anhand großer Datenmengen verstanden.

In der Medizin werden mehr und mehr Verfahren der künstlichen Intelligenz in verschiedenen Anwendungsbereichen entwickelt und zum Teil bereits eingesetzt. Dabei spielt die Verwendung von KI-basierten Verfahren auch für den Einsatz in der medizinischen Bildgebung eine wichtige Rolle.

Neben der Unterstützung bei der Bildevaluierung und der Befundung werden viele Ansätze verfolgt, die Bildrekonstruktion dreidimensionaler medizinischer Bilddatensätze und-/oder die Bildverarbeitung zur besseren Visualisierung durchzuführen, z. B. mittels Rauschunterdrückung mit Hilfe von KI-basierten Verfahren. Dies geschieht auch für Bilddatensätze, deren Ausgangsdaten unter Verwendung ionisierender Strahlung gewonnen wurden.

Eine Optimierung der medizinischen Anwendung ionisierender Strahlung ist im Sinne des Strahlenschutzes geboten. Daher wird die Anwendung von KI-basierten Ansätzen verwendet, um aus den Bildern bessere diagnostische Informationen bei gleicher Patientenexposition zu erhalten oder bei gleicher diagnostischer Aussagekraft der Röntgenbildgebung oder nuklearmedizinischer Bildgebung die Patientenexposition zu verringern.

Somit ist, wenn möglich in naher Zukunft, eine Nutzung der Möglichkeiten KI-basierter Verfahren, insbesondere solcher, die maschinelles Lernen verwenden, anzustreben. Andererseits ist zu bewerten, inwieweit mögliche, mit dem Einsatz maschinellen Lernens verbundene Probleme relevant sind. Dies gilt insbesondere auch im Sinne des Strahlenschutzes. Es besteht beispielsweise bei Verwendung von KI-assistierten Verfahren zur Rekonstruktion oder Rauschreduktion die Möglichkeit, dass für die Befundung relevante Information verloren geht, verfälscht wird oder falsche Information hinzugefügt wird. Dies kann möglicherweise in den zu befundenden Bildern nicht erkannt und anhand der abgespeicherten Daten häufig sogar nicht reproduziert werden. In diesem Fall würde dann ionisierende Strahlung am Menschen angewandt, ohne den medizinischen, rechtfertigenden Zweck zu erfüllen. Dies muss vor dem Hintergrund des Strahlenschutzrechts (insbesondere § 83 StrlSchG (StrlSchG 2017)) beurteilt werden.

Vor diesem Hintergrund hat das Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (BMUV) die Strahlenschutzkommission (SSK) mit Schreiben vom 07. Oktober 2020 um eine Stellungnahme zur Anwendung von Methoden der künstlichen Intelligenz (KI) bei der Bildrekonstruktion in Radiologie und Nuklearmedizin, im Hinblick auf die eingesetzte ionisierende Strahlung aus Sicht des Strahlenschutzes, zu betrachten.

Das vorliegende Beratungsergebnis beantwortet die Fragen des Beratungsauftrages mit besonderem Fokus auf die Computertomographie, Bildgebung bei Interventionen, Positronen-Emissions-Tomographie (PET), Einzelphotonen-Emissions-Computertomographie (SPECT), die Cone-Beam-Computertomographie und die digitale Tomosynthese. Die Betrachtung der automatisierten Befundung von Bilddaten ist nicht Teil des Beratungsauftrages.

Zur Erarbeitung der vorliegenden Empfehlung richtete die SSK im Jahr 2021 eine Arbeitsgruppe ein, der folgende Mitglieder angehörten:

Herr Prof. Dr. Christoph Hoeschen (Universität Magdeburg), Vorsitzender der Arbeitsgruppe

Herr Prof. Dr. Stefan Delorme (Deutsches Krebsforschungszentrum)

Herr Prof. Dr. Jörg van den Hoff (Technische Universität Dresden)

Herr Prof. Dr. Marc Kachelrieß (Deutsches Krebsforschungszentrum)

Herr Prof. Dr. Andreas Maier (Universität Erlangen)

Herr Prof. Dr. Georg Schramm (Universität Leuven, Belgien)

Frau Dr. Sarina Thomas (Universität Oslo, Norwegen)

Herr Prof. Dr. Michael Uder (Uniklinikum Erlangen)

Die SSK dankt dem BfS für die Bereitstellung umfangreicher Fachinformationen und fachlicher Unterstützung.

Bonn, im Dezember 2025

Prof. Dr. Christoph Hoeschen

Vorsitzender der Arbeitsgruppe „KI in der
Medizin“

Prof. Dr. Ursula Nestle

Vorsitzende der Strahlenschutzkommission

INHALT

1	Einleitung	6
1.1	Einführung	6
1.2	Begriffserklärungen	7
1.3	Situation der Anwendung von KI in der Medizinischen Bildgebung	9
1.4	Beratungsauftrag	12
2	Empfehlungen und Stellungnahme	14
	Wissenschaftliche Begründung	18
3	Anwendung von KI-Verfahren in der Medizin	18
3.1	Bildgebende Diagnostik in Radiologie und Nuklearmedizin sowie Interventioneller Bildgebung	18
3.1.1	Rauschreduktion.....	18
3.1.2	Bildrekonstruktion	22
3.1.3	Auflösungserhöhung (Superresolution).....	24
3.1.4	Artefaktreduktion bei CT und Röntgendiagnostik.....	24
3.1.4.1	Fehlende Daten	24
3.1.4.2	Fehlerbehaftete Daten	25
3.1.4.3	Bewegungskorrektur und -kompensation in der Röntgendiagnostik.....	26
3.1.5	Bewegungskorrektur in der Nuklearmedizin.....	26
3.1.6	Weitere Verfahren zur Reduktion der Strahlenexposition in der Diagnostik.....	27
3.1.7	Segmentierung	27
3.1.8	Klassifizierung	28
3.1.9	Radiomics.....	28
3.1.10	Dosisberechnung.....	29
3.2	Strahlentherapie	30
3.2.1	Pseudo-CT	30
3.2.2	Dosisberechnung.....	30
3.2.3	Decision-Support-Systeme	31
4	Rahmenbedingungen	31
4.1	Rechtliche Rahmenbedingungen.....	31
4.1.1	Rahmenbedingungen aufgrund der rechtfertigenden Indikation.....	32
4.1.2	Rahmenbedingungen aufgrund des Optimierungsgebots	33
4.1.3	Rechtliche Fragen aufgrund potenziell verursachter Fehlbehandlungen ...	33
4.1.4	Datenschutzrechtliche Aspekte.....	34
4.2	Technische Rahmenbedingungen	34

4.2.1	Training von tiefen neuronalen Netzen und Backpropagation-Algorithmus	35
4.2.2	Datensammlung und Datenschutz in Europa	35
4.2.3	Rechenkapazität für das Training von tiefen Netzen	35
4.2.4	Software	36
4.2.5	Qualitätsmanagement und Messverfahren	36
4.3	Anwenderspezifische Rahmenbedingungen	38
5	Potenzielle Probleme und Limitierungen	40
5.1	Garbage in, Garbage out	40
5.2	Probleme erkennen	41
5.3	Prinzipielle Problematik der Überprüfbarkeit von KI-Verfahren	41
6	Qualitätsgesicherte Eingangsdaten	43
7	Mögliche Testverfahren zur Qualitätssicherung	44
7.1	Variable Phantome	44
7.2	Vorgaben für Daten – Datentest	45
7.3	Statistische Analysen	45
7.4	Virtuelle Phantome und Qualitätssicherungsmethoden	47
7.5	Offenlegung der Trainings-, Validierungs- und Testdatenverteilung	48
8	Eingang in Regelwerke	49
9	Hinweise für die Entwicklung KI-unterstützter Rekonstruktions- und Weiterverarbeitungsalgorithmen in Radiologie und Nuklearmedizin	50
10	Schlussbetrachtungen	51
11	Literaturverzeichnis	53
	Begriffsbestimmungen	59
	Abkürzungsverzeichnis	63

1 Einleitung

1.1 Einführung

Künstliche Intelligenz (KI) ist ein Bereich der Informatik, der sich mit der Entwicklung von Maschinen befasst, die in der Lage sind, Aufgaben zu erledigen, die bisher eine menschliche Intelligenz erfordern. Ein wichtiger Teilbereich der KI ist das maschinelle Lernen (ML), das es Maschinen ermöglicht, Muster in Datenmengen zu erkennen. Unter deep learning (DL) wird dann das maschinelle Lernen anhand großer Datenmengen verstanden, d. h. DL ist ein Teil des maschinellen Lernens, welches wiederum ein Teil der KI darstellt. Dies korrespondiert mit der Definition der Europäischen Kommission im Artikel 3 (1) des EU Artificial Intelligence Act (EU 2024). Dort wird ein „KI-System“ definiert als „machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments“.

Der grundlegende Ablauf zur Erstellung eines KI-basierten Verfahrens gliedert sich wie folgt:

Zunächst muss eine Problemstellung, die durch das KI-basierte Verfahren gelöst werden soll, definiert werden. Für diese wird ein so genanntes Modell entwickelt oder ausgesucht. Die Modelle sind neuronalen Netzen nachempfunden, imitieren also ein Netz von Neuronen verschiedener Ebenen, die miteinander verknüpft sind. Das gewählte Modell beschreibt die Struktur der Datenverarbeitung. In diesen Modellen gibt es je nach Komplexität unterschiedlich viele Parameter, also interne Variablen, die die Anpassung des Modells an die gegebene Aufgabe ermöglichen. Diese Parameter werden erlernt. Zum Erlernen der Parameter wird das Modell trainiert, z. B. durch Vergleich zwischen einem gewünschten Ergebnis und einem Eingangssignal oder aber auch durch das Erkennen bestimmter Muster in großen Datenmengen. Im Anschluss wird das Ergebnis des Trainings validiert. Am Ende der Erstellung des KI-basierten Verfahrens wird das Modell mit den aus dem Training und der Validierung hervorgegangenen Parametern getestet.

In der Medizin werden mehr und mehr Verfahren der künstlichen Intelligenz in verschiedenen Anwendungsbereichen entwickelt und zum Teil bereits eingesetzt. Dabei spielt die Verwendung von KI-basierten Verfahren auch für den Einsatz in der medizinischen Bildgebung eine wichtige Rolle. Unter Verfahren der medizinischen Bildgebung sind verschiedene Methoden zusammenzufassen: Mikroskopische Bildgebung vor allem für pathologische, ophthalmologische oder dermatologische Fragestellungen, Ultraschallbildgebung, Projektionsbildgebung mittels Röntgenstrahlung oder Erstellung von Szintigrammen mit Hilfe einer planaren Gammakamera, tomographische Verfahren wie Röntgen-Computertomographie (CT), Einzelphotonen-Emissions-Computertomographie (SPECT) und Positronen-Emissions-Tomographie (PET), Magnetresonanztomographie (MRT) oder auch optische Kohärenztomographie etc. Neben der Unterstützung bei der Bildevaluierung und der Befundung werden viele Ansätze verfolgt, die Bildrekonstruktion dreidimensionaler medizinischer unter Einsatz ionisierender Strahlung erzeugter Bilddatensätze und-/oder die Bildverarbeitung zur besseren Visualisierung durchzuführen, z. B. mittels Rauschunterdrückung mit Hilfe von KI-basierten Verfahren.

Eine Optimierung der medizinischen Anwendung ionisierender Strahlung ist im Sinne des Strahlenschutzes geboten. Daher wird die Anwendung von KI-basierten Ansätzen verwendet, um aus den Bildern besser erkennbare diagnostische Informationen bei gleicher Patientenexposition zu erhalten oder bei gleicher diagnostischer Aussagekraft der Röntgenbildgebung oder nuklearmedizinischer Bildgebung die Patientenexposition zu verringern. Somit ist, wenn

möglich in naher Zukunft, eine Nutzung der Möglichkeiten KI-basierter Verfahren, insbesondere solcher, die maschinelles Lernen verwenden, anzustreben. Andererseits ist zu bewerten, inwieweit mögliche, mit dem Einsatz maschinellen Lernens verbundene Probleme relevant sind. Dies gilt insbesondere auch im Sinne des Strahlenschutzes. Es besteht beispielsweise bei Verwendung von KI-assistierten Verfahren zur Rekonstruktion oder Rauschreduktion die Möglichkeit, dass für die Befundung relevante Information verloren geht, verfälscht wird oder falsche Information hinzugefügt wird. Dies kann möglicherweise in den zu befundenden Bildern nicht erkannt und anhand der abgespeicherten Daten häufig sogar nicht reproduziert werden. In diesem Fall würde dann ionisierende Strahlung am Menschen angewandt, ohne den medizinischen rechtfertigenden Zweck zu erfüllen. Dies muss vor dem Hintergrund des Strahlenschutzrechts (insbesondere § 83 StrlSchG (StrlSchG 2017)) beurteilt werden.

1.2 Begriffserklärungen

Basierend auf der generellen Definition der künstlichen Intelligenz werden im Folgenden weitere Begriffe erläutert, die zum Verständnis des Textes erforderlich sind:

Bei algorithmischer Modellierung im Kontext des maschinellen Lernens unterscheidet man im Wesentlichen zwischen White-Box-, Gray-Box- und Black-Box-Modellen. Die Wahl des Verfahrens hängt von der Art des Lernproblems, also der zu lösenden Aufgabe, und den verfügbaren Daten ab:

- White-Box-Modelle sind keine lernbasierten Verfahren, da die Struktur des Modells und die zugehörigen Parameter – beispielsweise aus der Physik – bekannt sind. Diese Modelle werden häufig in der Bildgebung eingesetzt, da sie transparent sind und die Entscheidungsfindung nachvollziehbar machen. Im Wesentlichen können alle klassischen Verfahren zur tomographischen Bildrekonstruktion, von der gefilterten Rückprojektion bis hin zur iterativen Rekonstruktion, zu diesen Modellen gezählt werden.
- Black-Box-Modelle hingegen sind Verfahren, bei denen die innere Struktur des Modells und die zugehörigen Parameter nicht oder nicht in hinreichendem Umfang bekannt sind. Hier werden die Modelle durch die Verwendung geeigneter Trainingsdaten angelernt. Diese Modelle werden in der Bildgebung unter anderem im medizinischen Bereich eingesetzt, sofern sie in anderen Gebieten wie z. B. der Mustererkennung oder der Datenaufbereitung leistungsstarke Ergebnisse geliefert haben und sich einfach auf die Bildgebung anwenden lassen. Beispiele für Black-Box-Modelle sind U-Net-Rekonstruktionen sowie AUTOMAP (siehe Abschnitt 3.1.2).
- Gray-Box-Modelle sind eine Mischung aus White-Box- und Black-Box-Modellen. Bei diesen Verfahren ist die Struktur des Modells bekannt, aber nicht alle Parameter. Gray-Box-Modelle können von Vorwissen profitieren, um das Lernproblem zu vereinfachen, das Ergebnis zu verbessern und die Anzahl der benötigten Parameter zu reduzieren. Ein Beispiel für ein Gray-Box-Modell ist das so genannte Known-Operator-Learning, bei dem ein Teil des Modells bereits bekannt ist und das Modell nur die verbliebenen Parameter lernen muss. Methoden wie so genannte Variational Networks, FBP-Net, oder andere Hybridverfahren (Maier et al. 2022a) erlauben es so, klassische Theorie mit maschinellem Lernen zu kombinieren.

In dieser Empfehlung liegt der Fokus der Überlegungen zur KI auf Black-Box- und Gray-Box-Modellen, da diese Ansätze in der Literatur erstaunliche Ergebnisse erzielt haben und aktuell besonders diskutiert werden. Neben den Black-Box- und Gray-Box-Modellen sind natürlich auch White-Box-Modelle wichtige Verfahren der KI. Sie sollten aber getrennt betrachtet werden, da hierfür die klassischen Analyse- und Qualitätssicherungsverfahren im Wesentlichen

übernommen werden können. Bei Black-Box-Modellen gelingt dies nicht, weshalb sie in der vorliegenden Empfehlung im besonderen Fokus stehen. Viele der Black-Box-Modelle beruhen auf so genannten convolutional neural networks (CNNs), die über Inputdaten trainiert werden. Bei diesen faltungs-basierten Netzwerken handelt es sich um zahlreiche nichtlineare Filteroperationen, die gleitend auf die Bilddaten bzw. deren Zwischenergebnisse angewendet werden. Im Gegensatz zu so genannten Fully-Connected-Networks, bei denen jedes Voxel eines Zwischenergebnisses mit jedem Voxel des folgenden Zwischenergebnisses verknüpft ist, werden beim CNN nur eng benachbarte Voxel verknüpft. Somit bleiben a) die Nachbarschaftsbeziehungen der Bildinhalte erhalten und b) die Anzahl der durch das Training zu bestimmenden Unbekannten ist um etliche Größenordnungen kleiner.

Wie ein natürliches besteht auch ein künstliches neuronales Netz aus vielen Verarbeitungsschichten (Layern). Jeder Layer bearbeitet die Daten nichtlinear und reicht sie an den nächsten Layer weiter. Nach typischerweise einigen Dutzend bis Hundert Schichten gibt das neuronale Netz die bearbeiteten Daten, das bearbeitete Bild oder den bearbeiteten Bilderstapel aus. Konkret wird bei bildverarbeitenden neuronalen Netzen in den Layern meist eine Faltung¹ durchgeführt, d. h. das Bild wird mit einem 3×3-Faltungskern (oder mehreren) gefiltert und das Ergebnis der Faltungsoperation für jedes Pixel durch eine nichtlineare skalare Funktion geschickt (z. B. indem negative Werte auf Null gesetzt werden). Jeder Faltungskern hat z. B. bei zweidimensionalen Kernen 9 und bei dreidimensionalen Kernen 27 unbekannte Koeffizienten pro Dimension eines so genannten Features. Durch die zahlreichen Faltungskerne pro Schicht und die zahlreichen Layer beläuft sich die Anzahl der im Training zu bestimmenden Koeffizienten auf typischerweise mehrere Millionen bis einige Milliarden. Die Koeffizienten werden im Training bestimmt – dafür bedarf es der o. g. zahlreichen Trainingsdaten.

Beschreibung/Klarstellung verschiedener Datenklassen

Der grundlegende Ablauf der Entwicklung eines KI-basierten Verfahrens benötigt verschiedene Datensätze, die auch als Datenklassen bezeichnet werden:

- Trainingsdaten beschreiben die Eingabedaten, die zur Optimierung der neuronalen Netzparameter mittels Lernens herangezogen werden.
- Validierungsdaten beschreiben die Daten, die während der Entwicklung zur Beurteilung der Güte des Netzes herangezogen werden. Ist das Netz noch nicht ausreichend gut, so werden dessen Parameter (Netzlayout, Anzahl der Trainingsepochen, Lernrate etc.) geändert und erneut mit den Trainingsdaten trainiert. Somit werden die Validierungsdaten indirekt in den Trainingsprozess einbezogen, nämlich um übergeordnete Modellparameter zu optimieren.

¹ Die Faltung von zwei Funktionen ist eine mathematische Operation, bei der der Wert der ersten Funktion f durch einen Funktionswert ersetzt wird, welcher entsteht als das Integral des Werts der ersten Funktion multipliziert mit einer verschobenen und gespiegelten zweiten Funktion g über alle Verschiebungen. Das heißt, es wird die Überlagerung der Funktionen betrachtet und f durch g quasi „verschmiert“. Die mathematische Formulierung für zwei Funktionen f und g ist $(f * g)(x) = \int_{-\infty}^{+\infty} f(x') \cdot g(x - x') dx'$. Die Faltung kann auf Grund ihrer Funktionsweise z. B. sehr gut verwendet werden, um Filterfunktionen darzustellen. Man kann Faltungen auch mehrdimensional durchführen und als diskrete Operation anwenden, wie dies z. B. für die erwähnten CNNs verwendet wird.

- Die Testdaten sind ein unabhängiger Datensatz, der verwendet wird, um das entwickelte KI-Verfahren nach der Validierung zu testen. Sie sind möglichst repräsentativ für die relevante Population in der klinischen Anwendung des Verfahrens, und werden entweder extern akquiriert oder von der zur Entwicklung herangezogenen Gesamtdatenmenge entnommen.

Ist die Entwicklung abgeschlossen, das Netzwerk und dessen Parameter finalisiert und evaluiert worden, sollte das KI-basierte Verfahren noch auf einem weiteren (zweiten) Testdatensatz – auch als Verifikationsdatensatz bezeichnet – evaluiert werden. Dieser Verifikationsdatensatz sollte vollkommen isoliert von der Entwicklung behandelt werden und bestmöglich die Gesamtvariabilität der Eingabedaten der finalen Anwendung abdecken, d. h. multizentrisch (in verschiedenen Zentren gewonnen) und repräsentativ sein. Dabei ist im Falle medizinischer Daten insbesondere auf Einflussfaktoren, wie zum Beispiel Geschlecht, Gewicht, Herkunft, Pathologie und Alter zu achten, sofern Hersteller bzw. Entwickelnde des Verfahrens Zugang zu diesen Daten haben.

1.3 Situation der Anwendung von KI in der Medizinischen Bildgebung

Besonders bedeutsam ist der Einsatz von elektronischer Datenverarbeitung in allen Bereichen der Medizin, die von jeher mit großen Datenmengen umgehen. Das gilt insbesondere dann, wenn von Ärzten oder Ärztinnen ein besonderes Maß an Aufmerksamkeit gefordert wird und die Gefahr besteht, dass wichtige Befunde dem Nachweis entgehen. Dies betrifft somit alle Disziplinen, die mit Bildern umgehen, seien es radiologische oder endoskopische Bilder oder Fotografien, z. B. der Haut oder von mikroskopischen Gewebeschnitten, die ohne Weiteres in ein digitales Format umgewandelt werden können.

So fordert die Durchsicht mikroskopischer Gewebeschnitte hohe Aufmerksamkeit, indem z. B. bei der Untersuchung einer Biopsie einer tumorverdächtigen Läsion der Nachweis atypischer Zellen in Blut- oder Lymphgefäßen beweisend für die Malignität und Invasivität des Prozesses ist. Solche invasiven Herde können sehr klein sein, so dass die Anforderung an die Aufmerksamkeit bei der Untersuchung hoch ist.

In der Radiologie verhält es sich ähnlich. So fordert der Ausschluss von Lungenrundherden in einer Computertomographie des Thorax hohe Aufmerksamkeit, da alle angeschnittenen Gefäße (Lungenarterien und -venen) auf dem einzelnen Schnittbild meist eine ovale Form haben und sich erst beim „Blättern“ durch den Bildstapel ihre längliche Form erschließt. Von Untersuchenden wird die Aufgabe als sowohl anstrengend als auch stumpfsinnig empfunden, so dass die Gefahr übersehener Befunde nicht zu unterschätzen ist. Ähnlich verhält es sich bei Ganzkörperuntersuchungen mit der Durchmusterung der Knochen auf umschriebene Abnormitäten.

Etwas anders sind z. B. Ganzkörper-MRT-Untersuchungen zu werten, die zum einen eine kaum zu bewältigende Informationsfülle tragen und zum anderen artefaktanfällig sind. Hier kommt zur bloßen Bewältigung der Aufgabe noch die intellektuelle Leistung hinzu, echte Befunde von Artefakten zu unterscheiden.

Eine Reihe digitaler Hilfsmittel wird bereits eingesetzt: Die Einfachsten von ihnen ändern lediglich die Bildpräsentation, um den Untersuchenden ihre Aufgabe zu erleichtern. Hierzu zählt zum Beispiel die „Thick-Slab-MIP“ (MIP: Maximal Intensity Projection) bei der Rundherdsuche in Lungen-CT-Aufnahmen. Hierbei werden aus ca. 1 cm dicken Stapeln von Dünnschicht-CT-Aufnahmen (mit 1 mm Schichtdicke) der Lunge MIPs erzeugt, auf denen Gefäßanschnitte länglich und Rundherde weiterhin rund abgebildet sind – ohne Abstriche hinsichtlich der Auflösung, was das Auffinden selbst kleiner Rundherde wesentlich erleichtert. Andere Algorithmen überführen durch elastische Umformung die Darstellung der Rippen in ein Aufsichtsbild,

auf dem die Rippen gestreckt und parallel angeordnet sind, was das Auffinden osteolytischer oder osteosklerotischer Läsionen wesentlich erleichtert und beschleunigt.

Weitergehende Ansätze der Bildverarbeitung bestehen darin, pathologische Befunde automatisch aufzufinden, z. B. Herdbefunde der Lungen. Solche Lösungen werden teilweise bereits in der Software der Scanner integriert ausgeliefert, können als Option hinzugekauft oder als separate Stand-Alone-Software erworben werden. Viele dieser Lösungen beinhalten zusätzlich Werkzeuge zur Segmentierung der gefundenen Läsionen, z. B. für Volumenmessungen, wie sie für die Beurteilung des Therapieansprechens von Tumoren anhand vorgegebener Kriterien wie „RECIST“ (Response Evaluation Criteria In Solid Tumors, (Therasse et al. 2006)) erforderlich sind. Allen genannten Lösungen gemeinsam ist ein hohes Maß an Interaktion mit der auswertenden Person, die entweder nach entsprechender Aufbereitung der Bilder die Befunde selbst identifizieren und annotieren muss oder eine von der Software vorgeschlagene Annotation korrigieren oder verwerfen kann. Entsprechend hoch sind die Anforderungen an die Expertise der auswertenden Person. Die meisten dieser Ansätze beruhen auf algorithmischen Ansätzen ohne Einsatz von Deep Learning; inzwischen werden erste Ansätze zum Einsatz von auf Deep Learning basierenden Methoden verfolgt.

Wissenschaftlich erprobt werden Ansätze, mit Hilfe von künstlicher Intelligenz und so genannter Radiomics-Verfahren, bei denen aus lokalen Bildeigenschaften von Befunden (wie der Signalintensität, der Textur oder der Kontur) versucht wird, z. B. Zusammenhänge mit der Artdiagnose (z. B. benigne oder maligne Läsion) abzuleiten (Lambin et al. 2012). Auch Korrelationen mit biologischen Eigenschaften von Befunden (z. B. Wahrscheinlichkeit eines Therapieansprechens) oder der allgemeinen Krankheitsprognose werden untersucht. Diese Verfahren befinden sich ausnahmslos in der wissenschaftlichen Erprobung und stehen nicht als Medizinprodukt oder als Teil davon zur Verfügung. Auch steht es der anwendenden Person frei, die Resultate solcher Prozesse zu beurteilen, zu modifizieren oder zu verwerfen, denn ihr steht jederzeit die gesamte Bildinformation zur Verfügung.

Von diesen Ansätzen zur Unterstützung der Diagnose und zur Evaluation gegebener Bilddaten zu klinischen oder wissenschaftlichen Zwecken sind durch die KI selbst vorgenommene Änderungen der Bildinformation zu unterscheiden. Letztere wirken sich möglicherweise darauf aus, ob die originäre Information in den bearbeiteten Bildern überhaupt noch in gleicher Weise erhalten ist. Zudem hat die befundende Person, falls die originären Daten stark durch Rauschen oder andere Artefakte beeinträchtigt sind, keine Möglichkeit, sich von der Wirklichkeitstreue der erzeugten Bilder zu überzeugen. Derartige Verfahren sind in der Entwicklung, je nach Anwendung schon näher an der klinischen Erprobung oder weiter entfernt. Genauere Beschreibungen hierzu finden sich in der wissenschaftlichen Begründung.

Der im letzten Absatz dargestellte Einsatz von KI-Verfahren in der Bearbeitung von Daten aus bildgebenden Untersuchungen – sowohl im Bild- als auch im Rohdatenraum – kann in vielfältiger Weise genutzt werden: So können die KI-Verfahren darauf zielen, die Strahlenexposition ohne Abstriche an die Bildqualität zu reduzieren (Priorisierung der Strahlenreduktion) oder bei gegebener Strahlenexposition bessere Bilder zu erzeugen (Priorisierung der Bildqualität). Ein KI-basiertes Verfahren, das im Falle der Bildverarbeitung typischerweise durch ein neuronales Netz realisiert wird, lernt anhand von Trainingsdaten vorbereitend zur eigentlichen Bearbeitung der Daten aus einer konkreten Untersuchung, was zu tun ist. Bei den genannten Beispielen, in denen stärker verrauschte Bilder in weniger verrauschte Bilder umzurechnen sind, kann dies im einfachsten Fall dadurch geschehen, dass man einem neuronalen Netz sehr viele Datenpaare aus stark verrauschten und wenig verrauschten Patientenbildern (oder Bilderstapeln) präsentiert. Die stark verrauschten Bilder oder Bilderstapel lassen sich beispielsweise durch Addition von realistischem Rauschen auf den wenig verrauschten Bildern oder Bilderstapeln erzeugen.

Das neuronale Netz wird damit dahingehend trainiert, dass es die stärker verrauschten Bilder so entrauscht, dass sie den wenig verrauschten Bildern möglichst ähnlich werden.

Ein KI-Verfahren kann lediglich eine Schätzung vornehmen, um die fehlende Information zu ergänzen. Diese Information stammt dann aus der Verteilung von Trainingsdaten aus einer größeren Zahl anderer Untersuchungen und nicht aus dem untersuchten Patienten bzw. der untersuchten Patientin. Oft handelt es sich um gepaarte Trainingsdaten (überwachtes Training, im Englischen supervised learning, (Mohri et al. 2012)), so dass beispielsweise im Falle der Rauschreduktion je ein CT-Volumen mit viel Rauschen und ein zugehöriges CT-Volumen mit wenig Rauschen vorliegt. Somit besteht grundsätzlich die Möglichkeit, dass Information, die aufgrund einer angestrebten Bildqualitätsverbesserung im Bild angezeigt wird, möglicherweise nicht der aktuellen Messung entstammt. Auch hier gilt, dass KI-Verfahren lediglich die Wahrscheinlichkeitsverteilung der Trainingsdaten lernen konnten und somit auch nur die dort enthaltene Information in das Bild des eigentlichen Patienten oder der eigentlichen Patientin transferieren können.

Das wohl sichtbarste und typischste Anwendungsfeld von KI in der Bildgebung, welches auch bereits in einigen Fällen besonders weit in der klinischen Erprobung ist, ist die Verarbeitung von Bildern mit dem Ziel der Rauschreduktion bei gleichzeitiger Erhaltung der Bildschärfe. Dies entspricht einer Erhöhung des Kontrast- bzw. Signal-zu-Rauschverhältnisses. Mit einem derartigen Verfahren lässt sich auch eine Dosisreduktion erreichen, wenn die Untersuchungsprotokolle entsprechend angepasst werden, z. B. durch eine Reduktion der Belichtungszeit oder des Röhrenstroms am CT. Diese KI-Rauschreduktionsverfahren finden nicht nur in der Computertomographie Anwendung, sondern auch bei Projektionsaufnahmen, wie beispielsweise dem Röntgen oder der Durchleuchtung, und auch in nuklearmedizinischen Verfahren.

Ein weiteres Anwendungsfeld von KI-Verfahren in der Bildgebung mit ionisierender Strahlung ist die Verbesserung des Bildeindrucks bei Röntgenaufnahmen inklusive Durchleuchtung und Mammographie. Im Gegensatz zu tomographischen Bildern, bei denen die angezeigten Grauwerte auch quantitativ bedeutsam sind und daher von den KI-Verfahren möglichst nicht verändert werden dürfen, ist bei Projektionsaufnahmen lediglich meist der qualitative Bildeindruck wichtig. Bei dem Grauwert eines Pixels in der Projektionsaufnahme handelt es sich um eine Überlagerung der Schwächungswerte im Strahlengang hintereinander liegender Strukturen, und der Grauwert des Pixels lässt keine quantitativen Rückschlüsse auf die Verhältnisse in den einzelnen Strukturen zu. Nicht KI-basierte Veränderungen der quantitativen physikalischen Information, wie nichtlineare Kontrastanhebungen oder Kontrastabsenkungen, auch solche, die abhängig vom Bildinhalt sind, und auch solche, die irreversibel sind und/oder informationsreduzierend wirken, werden seit langem erfolgreich eingesetzt und sind klinischer Standard, sofern die befundrelevante Information nachweislich erhalten bleibt. Einfache konkrete Beispiele hierfür, deren Arbeitsweise gut verstanden und nachvollziehbar sind, sind die Gammakurven oder die Biasfilter. Für die unverfälschten Rohdaten besteht in der Strahlenschutzgesetzgebung keine Pflicht zur Archivierung und somit werden sie typischerweise nicht gespeichert. Die zur Befundung verwendeten Bilddaten sollen aber in jedem Fall ohne Qualitätsverlust archiviert werden, wie die SSK bereits 2016 empfohlen hat (SSK 2016). Mit Hilfe von KI-Verfahren lassen sich nun deutlich raffiniertere Effekte erzielen, wie beispielsweise das selektive Hervorheben oder Unterdrücken von Knochenstrukturen in Röntgenbildern, das Herausarbeiten von Gefäßen (virtuelle DSA) (Duan et al. 2024) oder der Ausgleich von Bewegungen in Durchleuchtungsbildern. Allerdings lässt sich die Arbeitsweise von KI-Verfahren nicht einfach nachvollziehen, geschweige denn verstehen, insbesondere beim Einsatz großer Datenmengen. Dies gilt z. B. für

Foundation Models², die Ergebnisse aufgrund von großen Bild- und Textdatenbanken, welche Patientenbefunde mit radiologischen Bildern koppeln, beeinflussen. In jedem Fall ist darauf zu achten, dass die Anwendung von KI keine relevante Information aus den Röntgenaufnahmen entfernt. Dies gilt für alle medizinisch bildgebenden Verfahren, auf Grund der strahlenschutzrechtlichen Regelungen insbesondere für solche, die ionisierende Strahlung verwenden.

KI-Verfahren können nicht nur auf Bilder, sondern auch auf Rohdaten angewendet werden, um Artefakte zu reduzieren oder Unzulänglichkeiten der bildgebenden Systeme auszugleichen (Magonov et al. 2024). Diese Prozesse sind für Anwendende möglicherweise nicht sichtbar und laufen ohne deren Kenntnis ab.

1.4 Beratungsauftrag

Vom Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (jetzt: Bundesministerium für Umwelt, Klimaschutz, Naturschutz und nukleare Sicherheit, BMUKN) wurde ein Beratungsauftrag formuliert, der mit Datum vom 07. Oktober 2020 die SSK bittet, die Problematiken im Zusammenhang mit dem Einsatz von KI in der Medizin zu untersuchen, aufzuzeigen und auch Wege zur Lösung der Problematiken vorzuschlagen. Dieser Beratungsauftrag fokussiert auf folgende Fragestellungen:

- Welche Anwendungsansätze von KI können für die Bildoptimierung (z. B. Streustrahlenreduktion, Dosisreduktion, Artefaktreduktion) sinnvoll sein, welche Risiken birgt der Einsatz in der medizinischen Bildgebung? Können dem Anwendenden Grenzen und potenziell nachteilige Auswirkungen einer KI-Methode verdeutlicht werden, und können zur Bewertung der Eignung der eingesetzten Methoden Kriterien abgeleitet werden?
- Welche Rahmenbedingungen bestehen derzeit für den Einsatz dieser Verfahren und welche müssen zukünftig gegeben sein, damit der Einsatz ionisierender Strahlung sicher ist und die tatsächlichen Befundkennzeichen hinreichend sicher gezeigt werden?

Bei den Betrachtungen entsprechender Anforderungen an die KI-Algorithmen und ihre Qualitätssicherung sollten die Trainingsdaten, die Anforderungen an die Validierung und Testung der Algorithmen sowie die Testung KI-unterstützter Bildrekonstruktionsverfahren begleitend zur klinischen Anwendung einbezogen werden.

- Welchen Beitrag hinsichtlich der Qualitätssicherung von KI-Methoden kann die internationale Normung und das Medizinproduktrecht von bildgebenden Systemen im Hinblick auf die Zulassung leisten?
- Methoden der KI lassen sich auch in Situationen mit niedrigem Rauschanteil oder im Hochkontrastbereich anwenden, um Verbesserungen der Bildqualität zu erreichen. Die anzuwendenden Methoden können sich daher aufgrund der Fragestellung und der Aufnahmesituation (Niedrig- und Hochkontrastaufnahmen) unterscheiden. Können Anforderungen an den Umgang mit unterschiedlichen verfügbaren Methoden der KI diesen Umstand berücksichtigen und wie könnte dies geschehen?
- Können Anforderungen für die Methodik eines KI-Algorithmus in der Bildoptimierung mit Bezug auf das vorgesehene Anwendungsfeld, die verwendbaren Geräte und Aufnahmeparameter, die benötigten Testdatensätze, die Validierung und zum anderen für

² Ein Foundation Model oder Basismodell ist ein großes maschinelles Lernmodell, das auf einer breiten Datenbasis vortrainiert wurde und durch seine Vielseitigkeit als Grundlage für eine Vielzahl spezifischer Anwendungen dient.

die Etablierung eines Testsystems, mit dem die informationssichere Anwendung von Verfahren überprüft werden kann, entwickelt werden? Wenn dies möglich ist, können mit diesen Anforderungen Vorgaben gemacht werden, was wie getestet und dokumentiert werden muss, um sicher sein zu können, dass die Rahmenbedingungen unter allen Umständen eingehalten werden? Dazu gehört auch, wie die Testverfahren selbst getestet werden können und wie damit sichergestellt werden kann, dass die vorgesehene Einsatzbandbreite der KI-Verfahren auch tatsächlich überprüft werden kann.

- Der Einsatz von Bildrekonstruktionsverfahren mit Methoden der KI beeinflusst auch messtechnische Überprüfungen, die z. B. im Rahmen von Abnahmeprüfungen oder auch zur Bestimmung der Bildqualität und der Dosiseffizienz durchgeführt werden. Die Anwendung von KI-Methoden auf Phantome in solchen Verfahren könnte zu Verzerrungen der Charakterisierung von Parametern wie Bildqualität und Dosiseffizienz führen. Die Anforderungen an die Validierung und Testung der Algorithmen sollten auch dies berücksichtigen und ggf. alternative Prüfverfahren vorschlagen.

2 Empfehlungen und Stellungnahme

Die SSK empfiehlt für die Verwendung KI-basierter Verfahren für die Bildrekonstruktion oder Bildverarbeitung bei Verfahren der medizinischen Bildgebung, die auf dem Einsatz ionisierender Strahlung basieren, dass

1. um bildverarbeitungsbedingte Fehldiagnosen auszuschließen, die den Nutzen der Untersuchung und damit die rechtfertigende Indikation in Frage stellen würden (siehe Abschnitt 4.1.1), bei Anwendung von KI-Verfahren zur Bildrekonstruktion oder Weiterverarbeitung von patientenbezogenen Bilddaten Sorge zu tragen ist, dass alle befundrelevanten Strukturen mit einer im Vergleich zu nicht KI-gestützten Verfahren mindestens im wesentlichen unveränderten Sensitivität und Spezifität in Bezug auf die diagnostische Sicherheit erhalten bleiben und ferner keine artifiziellen Strukturen generiert werden. Gleichmaßen dürfen für den Einsatz computertomographischer Bilddaten für die Bestrahlungsplanung in der Strahlentherapie die Hounsfield-Einheiten³ nicht signifikant verändert werden;
2. bei einer Dosisreduktion in Kombination mit der Anwendung von Verfahren der KI zur Generierung der zu befundenden Bilder alle befundrelevanten Strukturen mit vergleichbarer oder höherer Sensitivität und Spezifität identifiziert werden können sollen, wie sie ohne Dosisreduktion und KI-Einsatz gegeben wären, um der rechtfertigenden Indikation der Untersuchung gerecht zu werden (siehe Abschnitt 4.1.2);
3. zum Nachweis der Einhaltung der ersten beiden Empfehlungen Human-Observer-Studien durchgeführt werden sollen, es sei denn es liegen bereits wissenschaftlich validierte Ergebnisse vor, die eine äquivalente Aussagekraft anderer (messtechnischer) Verfahren bestätigen. In letzterem Fall ist darauf zu achten, dass die betreffenden Verfahren so gewählt werden, dass sie die zu überprüfenden KI-Verfahren in der Evaluation nicht bevorzugen und dass das Fehlen oder die Erzeugung subtiler Strukturen hinreichend berücksichtigt werden (siehe Abschnitte 4.1.2. und 4.2.5).

Voraussetzung dafür, dass der Qualität der Datenverarbeitung vertraut werden kann, ist die Transparenz der Prozesse (wie in Abschnitt 7.5 beschrieben). Daher empfiehlt die SSK weiterhin, dass

4. der gesamte Prozess der Erstellung und Verwendung von KI-basierten Verfahren für die Bildrekonstruktion oder Bildverarbeitung qualitätsgesichert erfolgen soll. Insbesondere sollen generative Netze aus einem Eingangsdatensatz immer einen eindeutigen, nicht durch Zufall beeinflussten Ausgangsdatensatz erzeugen.
5. eine zuständige Stelle benannt werden soll, welche die Leistungsfähigkeit eines Algorithmus vor seinem klinischen Einsatz für die spezifizierten Anwendungsfälle auf unabhängigen, ausreichend charakterisierten Testdaten überprüft (vgl. Abschnitt 4.2.5, 5.3, 7.2, 7.3, 7.5). Diese zuständige Stelle kann entweder eine vorhandene Stelle sein, in die zusätzliche Kompetenzen integriert werden, oder sie kann neu geschaffen werden.

³ Die Hounsfield-Einheit beschreibt in einer computertomographischen Darstellung die Abschwächung (als monoenergetisch angenommener) Röntgenstrahlung durch das Gewebe. Die Hounsfield-Skala wurde von Godfrey Hounsfield vorgeschlagen, die Schwächung der Röntgenstrahlung durch ein Gewebe wird dabei durch seinen Schwächungskoeffizienten μ beschrieben und der jeweilige Wert wird in Beziehung gesetzt zum Schwächungskoeffizienten von Wasser. Der CT-Wert bzw. die Hounsfield Einheit berechnet sich dann wie folgt:

$$\text{CT - Wert } [\mu_{\text{Gewebe}}] = \frac{\mu_{\text{Gewebe}} - \mu_{\text{Wasser}}}{\mu_{\text{Wasser}} - \mu_{\text{Luft}}} * 1000 \text{ HU}$$

Die durch die zuständige Stelle zu prüfenden Punkte sollten insbesondere für die Zulassung und Beurteilung KI-basierter Rekonstruktions- und Weiterverarbeitungsalgorithmen in der Radiologie und Nuklearmedizin die folgenden Aspekte beinhalten:

- Validität der von den Entwickelnden verwendeten Metriken und Methoden zur Beurteilung der Leistungsfähigkeit des KI-Algorithmus für spezifische klinische Anwendungen.
- Leistungsfähigkeit des zu beurteilenden KI-Algorithmus für die spezifizierte Anwendung anhand der offengelegten Teilmenge der Testdaten und Metriken der Entwickelnden.
- Externe und unabhängige Beurteilung der Leistungsfähigkeit des zu beurteilenden KI-Algorithmus für die spezifizierte Anwendung anhand von unabhängigen repräsentativen Testdaten, die den Entwickelnden **nicht** zur Verfügung standen.
- Angemessene Einbindung von erfahrenen klinischen und Bildgebungs-Experten und -Expertinnen in die Beurteilung der Leistungsfähigkeit eines KI-Algorithmus nach den Empfehlungen 1 und 2, sofern Grund zur Annahme besteht, dass eine Beurteilung der Leistungsfähigkeit durch mathematisch-technische Metriken nicht gewährleistet werden kann.
- Umfang und Repräsentativität praktisch relevanter Eingangsdaten zur Evaluierung der Generalisierungsfähigkeit von KI-Algorithmen, d. h. bei welchen Anwendungen, für welche Geräte und Untersuchungsparameter und welche Patientenkohorten ein Algorithmus verwendet werden kann.
- Existenz und Plausibilität der Risikoabschätzung und Gefahrenbeurteilung für potenzielle Versagensszenarien sowie Fehlanwendungen des KI-Algorithmus. Dies bezieht sich auf eine potenziell mögliche fehlerhafte Funktion des KI-Verfahrens.

Entsprechende verbindliche Regelungen sollten getroffen werden.

Die SSK empfiehlt zur Umsetzung:

6. Die zuständige Stelle soll Zugang zu von den Trainingsdaten unabhängigen Verifikationsdaten erhalten, um die Güte des eingesetzten Systems zu überprüfen, ähnlich den Phantomen, die bei konventionellen Methoden eingesetzt werden (vergleiche Abschnitte 5.3, 7.5).
7. Die unabhängige Stelle soll für die von ihr zu erstellende Analyse eine ausreichende Menge an multizentrischen und repräsentativen Daten verschiedener Geräte und Hersteller inklusive notwendiger Metadaten und Annotierungen, also den Markierungen oder Klassifizierungen von Bilddaten, erhalten. Die Bilddaten müssen zur Überprüfung der Güte der Algorithmen geeignet sein. Die Daten sind dabei nicht zu veröffentlichen, so dass gewährleistet wird, dass die Verifikationsdaten unabhängig sind und keinen Einfluss auf das Training der Algorithmen genommen haben (siehe Abschnitte 4.2.2 und 7.5).
8. Zur Beurteilung der Datengüte hinsichtlich der Verwendbarkeit zur Entwicklung KI-basierter Verfahren, unabhängige insbesondere auch sollen klinische Experten und Expertinnen herangezogen werden (siehe Abschnitte 4.3, 5.1).
9. Zur qualitätsgesicherten Beurteilung der Annotierungen (siehe Empfehlung 7) sollten Standards definiert und überprüft werden (siehe Abschnitte 4.3, 5.1).

Die SSK empfiehlt zur Entwicklung, Einführung und klinischen Umsetzung von KI-basierten Verfahren nach Empfehlung 1 und 2:

10. Verbindliche Anforderungen sollen an die Hersteller und Entwickelnden eines KI-Verfahrens gestellt werden, als notwendige Voraussetzung für eine Verifizierung der Einhaltung der ersten beiden Empfehlungen:
 - Für die Entwicklung sollen Trainings-, Validierungs- und Testdaten als Stichproben einer gemeinsamen Grundgesamtheit randomisiert entnommen werden – falls dies nicht der Fall ist, sollte es begründet werden. Hersteller sollen den Anwendungsfall spezifizieren (siehe auch Abschnitte 4.2.5, 5.1, 6 und 7.5).
 - Eine umfassende Dokumentation der benutzten Trainings-, Validierungs- und Testdaten hinsichtlich der genutzten Patientenkohorte, des verwendeten Radiopharmakons, der Dosis oder zumindest der applizierten Aktivität, bzw. der Exposition durch Röntgenstrahlung sowie der Aufnahme- und Rekonstruktionsparameter soll vorgelegt werden. Dies sollte unter der Voraussetzung erfolgen, dass der Datenschutz gewährleistet ist, und sofern der Hersteller Zugang zu den zuvor benannten untersuchungsrelevanten Daten hat. Falls Verfahren zur Datenaugmentierung zur Anwendung kamen, so sollten diese ebenfalls beschrieben werden (siehe Abschnitte 4.1.4, 4.2.5, 5.1, 6 und 7.5).
 - Ein repräsentativer Teil der Trainings-, Validierungs- und Testdaten sollte offengelegt werden, zumindest gegenüber einer zuständigen Stelle (siehe Abschnitte 4.2.5, 5.1, 6 und 7.5).
 - Es soll dargelegt werden, anhand welcher Kriterien die genutzten Trainings-, Validierungs- und Testdaten ausgewählt wurden und warum gerade diese Vorgehensweise sicherstellt, dass Out-of-Distribution-Fälle nicht in einem relevanten Maß auftreten werden (siehe auch Abschnitte 4.2.5, 5.1, 6 und 7.5).
 - Es soll dargelegt werden, ob und ggf. welche Maßnahmen vom Hersteller getroffen werden, um Out-of-Distribution-Fälle zu identifizieren (siehe Abschnitt 7.5).
 - Die Hersteller sollen darlegen, an welcher Stelle in der Datenvorverarbeitung, Bildrekonstruktion oder Bildverarbeitung KI-Algorithmen eingesetzt werden und welchem Zweck sie dienen. Es soll eine Spezifikation der Ein- und Ausgangsdaten vorgelegt und dargestellt werden, ob generative Netze zur Anwendung kommen und ob vortrainierte Netze eingesetzt werden. Falls dies der Fall ist, soll ebenso dokumentiert und offengelegt werden, womit die Netze vortrainiert wurden und wie dies geschah (z. B. supervised, semi-supervised, unsupervised, transfer learning) (siehe Abschnitte 4.2.1, 4.2.5, 5.1 und 7.2).
 - Die Hersteller sollen eine Risikoabschätzung dafür angeben, was eine Fehlfunktion der KI, z. B. bei einem Out-of-Distribution-Patienten, für den Patienten bzw. die Patientin bedeuten kann. Dabei ist darzustellen, ob eine Fehlfunktion zu deutlich sichtbaren Artefakten führt und somit leicht für die Anwendenden erkennbar ist oder ob es eventuell zu subtilen Veränderungen kleiner Details führt, die kaum erkannt und leicht als fehlende oder neu entstandene Pathologie missinterpretiert werden können. Der Hersteller soll darlegen, wie er zu seinen Einschätzungen gelangt und somit eine Risikoanalyse vorlegen (siehe Abschnitte 5.2, 5.3, 7.3, 7.4).
 - Die Hersteller sollen zu jedem KI-Verfahren erläutern, in welcher Weise es seine Eingangsdaten modifiziert (siehe Abschnitt 4.2.1).

11. Der Hersteller soll das Anwendungsfeld des Algorithmus basierend auf den Verifikationsdaten abgrenzen. Er soll ferner mögliche Limitierungen klar benennen, die sich ungünstig auf den Patienten bzw. die Patientin auswirken könnten, z. B. eine signifikante Unterrepräsentation von bestimmten Merkmalen wie Altersklassen oder ggf. ethnischen Gruppen in den Trainingsdaten (vergleiche Abschnitt 7.5). Ggf. sind Einschränkungen der Anwendungsbereiche oder -klassen, also eine Einschränkung z. B. auf Bilddaten von Personengruppen oder für bestimmte Erkrankungen, festzulegen.

Die SSK empfiehlt zudem als allgemeine Aspekte für die sichere Einführung und die dauerhafte sichere Verwendung von KI-basierten Verfahren gemäß Empfehlungen 1 und 2:

12. Es sollen möglichst neben den KI-basierten Verfahren weiterhin auch bereits etablierte, konventionelle Verfahren bereitgestellt werden, damit die Anwendenden die Entscheidung über das genutzte Verfahren selbst treffen und im Zweifelsfall eine Referenz heranziehen können (vergleiche Abschnitte 4.3, 5.3).
13. Weitere Forschung soll gefördert werden, um Testverfahren zu entwickeln, die eine Charakterisierung und Qualitätssicherung inklusive der potenziellen Limitierungen und Schwierigkeiten medizinischer Bildrekonstruktionen und -verarbeitungen beim Einsatz von KI-Verfahren ermöglichen (vergleiche Abschnitte 4.2.5, 5.3 und 10).
14. Internationale Abstimmungen zu Zulassungs-, Test- und Qualitätssicherungsverfahren für den Einsatz von KI-Verfahren für die Bildrekonstruktion und Weiterverarbeitung medizinischer Bilddaten sollten vorangetrieben werden (vergleiche Abschnitte 4.2, 5.2, und 8).
15. Bei der Datenakquise zur Entwicklung von auf künstlicher Intelligenz basierenden Algorithmen sollte in einer Datenvereinbarung zwischen klinischen und industriellen Partnern geregelt werden, dass ein ausreichend großer Teil der Daten für eine unabhängige, externe Verifikation bereitgestellt wird (siehe Abschnitt 4.2.2). Diese Daten könnten dann potenziell der zuständigen Stelle zur Verfügung gestellt werden (siehe Empfehlung 6).

Stellungnahme

Die SSK hält es im Kontext der Bildrekonstruktion und Weiterverarbeitung unter Anwendung ionisierender Strahlung erzeugter medizinischer Bilddaten mit Hilfe von KI-Verfahren für hilfreich, die Möglichkeit der Normung und ihres Einsatzes zu betrachten, um die oben dargestellten Empfehlungen zu unterstützen bzw. leichter umsetzbar zu machen. Daher nimmt die SSK wie folgt Stellung:

- Eine Normung, zumindest aber eine Standardisierung der Anforderungen an Daten für Training und Validierung, erscheint der SSK möglich und für die Qualitätssicherung sinnvoll. Gleiches gilt auch für Verfahren der Entwicklung, Testung und Qualitätssicherung der KI-basierten Anwendungen (siehe Kapitel 8 und 10). Eine Normung der KI-basierten Verfahren zur Rekonstruktion oder Bildverarbeitung selbst erscheint der SSK nicht sinnvoll, da unterschiedliche Funktionsweisen der einzelnen Algorithmen vorliegen, auch klassische Verfahren zur Rekonstruktion nicht normiert wurden und die Normung einzelner Verfahren aus Sicht der SSK kaum möglich erscheint und keine zusätzliche Sicherheit verspricht. Zudem ist die Entwicklung der KI-basierten Verfahren so schnell, dass der Normungsprozess zu langsam wäre.

Wissenschaftliche Begründung

3 Anwendung von KI-Verfahren in der Medizin

3.1 Bildgebende Diagnostik in Radiologie und Nuklearmedizin sowie Interventioneller Bildgebung

3.1.1 Rauschreduktion

Die medizinische Diagnostik verwendet in großem Umfang bildliche Information für die Diagnosestellung. Wenn dabei Informationen aus dem Körperinneren gewonnen werden sollen, werden zur Bildgebung meist entweder ultraschallbasierte Methoden, Magnetresonanzverfahren, röntgenbasierte oder nuklearmedizinische Verfahren verwendet. Die Verfahren, welche dreidimensionale Informationen zur Verfügung stellen, werden immer häufiger verwendet.

Bei allen diesen Verfahren ist dabei das Bildrauschen ein zentraler Faktor in der zu erzielenden Bild- und der damit verbundenen Diagnosequalität. Bei den o. g. Bildgebungsverfahren korreliert das Rauschen entweder invers zur Strahlenexposition (röntgenbasierte- und nuklearmedizinische Verfahren) oder zur Messzeit (Nuklearmedizin, Ultraschall und Kernspin). Der zahlenmäßig überwiegende Anteil röntgenstrahlenbasierter Untersuchungen entfällt dabei auf Projektionsaufnahmen. Die Strahlenexposition ist vergleichsweise gering, so dass der Einsatz rauschreduzierender Bildverarbeitungstechniken von den Herstellern derzeit weniger vorangetrieben oder weniger in die Öffentlichkeit getragen wird, als bei den tomographischen Verfahren. Die Bildqualität der medizinischen tomographischen Bildgebungsmodalitäten in Radiologie (CT, MRT) und Nuklearmedizin (SPECT, PET) wird in unterschiedlichem, modalitätsspezifischem Umfang von Bildrauschen beeinträchtigt. Bei den auf dem Einsatz ionisierender Strahlung beruhenden tomographischen Verfahren (CT, SPECT, PET) besteht ein direkter oder mittelbarer Zusammenhang zwischen Bildrauschen und der mit der Untersuchung verbundenen Strahlendosis. Letztere ist in der CT apparativ durch das Zeitintegral des Röhrenstroms der betreffenden Aufnahme determiniert.

Bei nuklearmedizinischen Verfahren ist die Strahlenexposition hingegen allein vom genutzten Radiopharmakon und seiner Aktivität abhängig, nicht aber von der gewählten Aufnahmedauer. Dabei hängt die zu applizierende Aktivität auch von den physikalischen und technischen Gegebenheiten des bildgebenden Systems ab. Das Bildrauschen wird hier durch die Parameter Aktivität und Aufnahmedauer bestimmt. Im Gegensatz zur CT kann daher in gewissen Grenzen eine Verminderung des Rauschens durch längere Aufnahmezeiten bei unveränderter Aktivität des applizierten Radiopharmakons erreicht werden. Da aber einer Verlängerung der Aufnahmezeiten sowohl praktisch (z. B. Durchsatz, Compliance der Patientin bzw. des Patienten, Patientenbewegung) als auch physikalisch (z. B. Halbwertszeit des genutzten Nuklids) recht enge Grenzen gesetzt sind, besteht grundsätzlich auch für die nuklearmedizinischen Verfahren ein kausaler Zusammenhang zwischen einer Verminderung der Strahlenexposition durch Reduktion der injizierten Aktivität und der damit einhergehenden unerwünschten Erhöhung des Bildrauschens bzw. der Verschlechterung des Signal-zu-Rauschverhältnisses.

Die oft angestrebte Reduktion der Strahlenexposition führt zu erhöhtem Bildrauschen, sofern nicht Gegenmaßnahmen ergriffen werden. Wird eine Verbesserung der Bildqualität bei gleichbleibender Strahlenexposition angestrebt, so kann diese ebenfalls oft durch Rauschreduktionsverfahren erreicht werden.

Eine wesentliche Anforderung an Rauschreduktionsverfahren ist die Erhaltung der räumlichen Auflösung, damit die Bilder nicht unscharf werden. Die Rauschreduktion muss also kantenerhaltend durchgeführt werden und ist somit per Definition nicht translationsinvariant. Klassische Verfahren zur Rauschreduktion bestehen oft aus einer oder mehreren Komponenten, die Kanten erkennen oder bewerten, und Komponenten, die die Glättung der Bilder so durchführen, dass an den Kanten nicht geglättet oder zumindest dort nur parallel zur Kante geglättet wird. Meist sind die Verfahren zudem nichtlinear: In Regionen starken Rauschens wird stärker geglättet als in Bereichen geringen Rauschens. In vielen Fällen erfolgt die Rauschreduktion iterativ.

Neuronale Netze, die zum Teil bereits in der Praxis zur Rauschreduktion eingesetzt werden, haben die gleiche Zielsetzung. Jedoch ist die o. g. Trennung von Kantendetektion und Rauschreduktion aufgrund der Komplexität der Netze nicht vorhanden oder sie lässt sich nicht erkennen – zumindest nicht ohne erheblichen Aufwand.

Das Training der Netze erfolgt oft, indem vorhandene Daten (z. B. klinische Bilder oder Volumina) durch einen Algorithmus nachträglich verrauscht werden, um eine geringere Strahlenexposition zu simulieren. Diese Niedrigdosisdaten (Input) und die zugehörigen Normaldosisdaten (Label) dienen dem Netz dann als Datenpaar zum Training. Alternativ dazu werden Rauschreduktionsnetze auch unüberwacht trainiert, d. h. mit einem Datenpool von Niedrigdosisdaten und einem weiteren Datenpool von Normaldosisdaten, die aber nicht von denselben Patienten bzw. Patientinnen stammen müssen. Ferner ist es denkbar, rein synthetisch erzeugte Daten zum Training zu verwenden.

In der CT werden seit einiger Zeit KI-basierte Rauschreduktionsverfahren von drei der vier großen CT-Hersteller angeboten (AiCE von Canon, TrueFidelity von GE, Precise Image von Philips). Allerdings sind zu den Algorithmen nur spärliche Informationen bekannt. Letztendlich werden vermutlich alle drei Verfahren überwacht trainiert. Als Input stehen nachträglich verrauschte und mit gefilterter Rückprojektion (FBP) rekonstruierte Bilder zur Verfügung. Als Label scheinen Canon und GE iterativ rekonstruierte Bilder der Normaldosisdaten zu verwenden, wohingegen Philips FBP-Rekonstruktionen der Normaldosisdaten nutzt. Die Verfahren haben zudem gemein, dass sie sich nicht nachträglich auf bereits rekonstruierte Bilder anwenden lassen, sondern mit der eigentlichen FBP-Rekonstruktion gekoppelt sind, um sie als Bildrekonstruktionsverfahren vermarkten zu können. Der vierte Hersteller, Siemens, bietet derzeit offiziell kein KI-basiertes Rauschreduktions- (bzw. „Rekonstruktionsverfahren“) an.

Bei der CBCT (interventionelle C-Bogen-Systeme, dentale DVT-Systeme, bildgebende CBCTs der Strahlentherapiegeräte) und der Radiographie (Durchleuchtung, Mammographie, digitale Subtraktionsangiographie) sind derzeit keine in einem Medizinprodukt installierte Anwendungen von KI-Verfahren zur Rauschreduktion bekannt. Generell ist dies aber mittelfristig zu erwarten, insbesondere für dynamische Bildgebungssysteme oder CBCT-Systeme. Hierbei können die Ansätze zur Rauschreduktion denen der CT durchaus ähneln, insbesondere im Hinblick auf die verwandten Methoden. Auch in wissenschaftlichen Veröffentlichungen betrifft ein Großteil der Veröffentlichungen zu Rauschreduktionsverfahren mit KI-Verfahren die CT (Balogh und Janos Kis 2022, Pashazadeh und Hoeschen 2023).

In der Nuklearmedizin besteht ebenfalls ein hoher Bedarf an der Reduktion des Bildrauschens, das im Vergleich zum CT aus methodischen Gründen grundsätzlich hoch ist. Der derzeit übliche Weg besteht in der Filterung von nach wenigen Iterationen gestoppten iterativen Maximum Likelihood Expectation Maximization (MLEM) Bildrekonstruktionen mit ortsinvarianten Tiefpassfiltern (zum Beispiel Gaußfilter), welche hochfrequentes Rauschen selektiv unterdrücken. Mit dieser Filterung ist jedoch stets ein merklicher Verlust an räumlicher Auflösung verbunden, welcher insbesondere die Detektion und Quantifizierung von Strukturen nahe der inhärenten

Auflösungsgrenze des Tomographen nachteilig beeinflusst. Es existieren auch andere Verfahren, die lokal adaptive kantenerhaltende Filterverfahren nutzen. Diese werden aber aus verschiedenen Gründen (variable Qualität der Ergebnisse, zu hohe Abhängigkeit von verschiedenen wählbaren Parametern zur Justierung der Filtereigenschaften etc.) bisher nicht routinemäßig eingesetzt. Alternativ kann das Vorwissen, dass das untersuchte Objekt selbst in Regionen homogen ist und deshalb nicht stark verrauscht sein kann, direkt innerhalb der iterativen Rekonstruktion zur Rauschunterdrückung genutzt werden. Dieses Verfahren bezeichnet man als Maximum-a-posteriori-Rekonstruktion (MAP-Rekonstruktion) mit „Glättungsprior“ (von engl. „prior knowledge“ = Vorwissen) (Qi 2006). Letzterer ist in diesem Zusammenhang eine „Straffunktion“, welche die verrauschten Rekonstruktionen zurückweist.

Deep-learning-basierte Verfahren unter Nutzung tiefer neuronaler Netze, wie z. B. CNNs, bieten mittlerweile eine z. T. bereits im klinischen Alltag verfügbare alternative Option zur lokal adaptiven Filterung tomographischer Datensätze, welche prospektiv eine relevante Verbesserung der Rauschunterdrückung mit weniger Auflösungsverlust ermöglicht, als dies bei heute gängigen *white-box* Verfahren der Fall ist. Als besonders beliebt und erfolgreich hat sich in diesem Zusammenhang – wie generell im Bereich der Bildrestaurierung (neben Rauschunterdrückung im engeren Sinne betrifft dies etwa Kontrastverbesserung, Auflösungsrückgewinnung, Bewegungskorrektur usw.) – eine spezielle Netzwerkarchitektur erwiesen: das im Jahr 2015 vorgeschlagene so genannte U-Net (welches zu den so genannten Encoder-Decoder Networks gehört) (Ronneberger et al. 2015). Ursprünglich wurde diese Architektur für Segmentierungsaufgaben entwickelt, wird aber mittlerweile in einer breiten Palette von Bildrestaurierungsaufgaben erfolgreich eingesetzt.

Der am weitesten verbreitete Ansatz für das so genannte Netzwerktraining, also den Optimierungsprozess zur Bestimmung geeigneter Werte der in der Regel extrem großen Zahl freier Parameter des Netzwerks (z. B. Gewichte gelernter Faltungskerne) ist das so genannte *Supervised Learning*. Im Kontext der Rauschreduktion wird das Netzwerk hierbei darauf trainiert, aus geeigneten Trainingsdaten aus verrauschten Bildern solche mit deutlich verbessertem Signal-zu-Rausch-Verhältnis zu restaurieren. Soweit es die nuklearmedizinische Bildgebung betrifft, nutzt das Training, wo verfügbar, Datensätze, welche mit den klinisch derzeit üblichen applizierten Aktivitäten sowie möglichst langen Aufnahmezeiten generiert wurden, so dass hiermit eine qualitativ hochwertige Referenz verfügbar ist. Diese Referenzdatensätze werden oft mit dem englischen Begriff *Ground Truth* bezeichnet. Für das Training von Entrauschungsnetzwerken in der PET sind hoch qualitative rauscharme Bilder besonders hilfreich. Für die Erlangung solcher Bilddaten bieten sich mittlerweile auch Daten von „*Total-Body PET*“ Systemen an. Diese sind erst seit 2020 auf dem Markt, zeichnen sich durch eine äußerst hohe Sensitivität aus und können daher mit kürzeren Aufnahmezeiten oder geringeren Aktivitäten rauscharme und hochwertige Bilder liefern.

Aus nuklearmedizinischen Rohdaten, die üblicherweise im *Listmode*-Format (Liste von Photonenereignissen) aufgenommen werden, können durch stochastische Entnahme einer Datenteilmenge (*Poisson Thinning*) virtuelle Rohdatensätze simuliert werden, die mit Aufnahmen mit verkürzter Messzeit oder reduzierter applizierter Aktivität oder einem Scanner mit geringerer Sensitivität (kürzeres axiales Gesichtsfeld, d. h. Ausdehnung des Tomographen in Kopf-Fuß-Richtung) vergleichbar sind. Die Rekonstruktion der virtuellen Teildaten und die Rekonstruktion der originalen kompletten Rohdaten ermöglichen es, Paare von Bildern bestehend aus einem simulierten Eingangsbild mit erhöhtem Rauschniveau und dem originalen *Ground Truth*-Bild retrospektiv (stochastisch korrekt) zu generieren. Diese Bildpaare können anschließend zur Optimierung und Validierung der Netzwerkparameter verwendet werden.

Eine zentrale Anforderung an das Training von neuronalen Netzwerken ist die Verfügbarkeit einer beträchtlichen Zahl validierter Eingangs-Ausgangs Bildpaare, welche insbesondere die Bandbreite der in der Praxis auftretenden inter-individuellen Variabilität der Bildeigenschaften abdecken müssen. Im nuklearmedizinischen Kontext sind dies unter anderem unterschiedliche physikalische und biologische Tracereigenschaften, nicht-standardisierte applizierte Aktivitäten und Aufnahmezeiten und, daraus resultierend, unterschiedliche Signal-Rausch-Verhältnisse, (patho-)physiologische und anatomische Unterschiede zwischen verschiedenen Patienten bzw. Patientinnen und Unterschiede in der Bildqualität verschiedener Scanner. Eine hinreichende Abdeckung dieser Bandbreite ist wichtig, um Netzwerke zu erhalten, die auch im klinischen Alltag valide und robuste Ergebnisse liefern. Des Weiteren ist es essenziell, die Bandbreite der verwendeten Trainingsdaten zu kennen, um über die Eignung eines gegebenen Netzwerkes für eine gegebene Fragestellung Schlussfolgerungen treffen zu können wie etwa: Kann ein gegebenes, trainiertes Netzwerk für Bilder genutzt werden, die mit einem anderen Radiotracer (oder Gerät) generiert wurden, welche in den Trainingsdaten noch nicht berücksichtigt werden konnten?

Da die für das *Supervised* Training benötigten rauscharmen Referenzbilder nicht für alle Anwendungsfelder verfügbar oder generierbar sind, verfolgen andere Ansätze zur Erzeugung der Referenzbilder komplementäre Strategien. Diese umfassen z. B. (a) das Nutzen von Monte-Carlo-simulierten Daten, bei denen das rauschfreie Referenzbild bekannt ist, (b) aufwändigere Rekonstruktionen (MAP) oder Nachbearbeitung von klinischen Bildern mit kantenerhaltenden Filtern oder strukturellem Vorwissen (Prior).

Darüber hinaus gibt es eine Vielzahl von Trainingsmethoden, die darauf abzielen, die Anzahl der benötigten Trainingsdaten zu vermindern. Zu diesen gehören unter anderem das *Transfer Learning* und *unsupervised* Trainingsmethoden. Im *Transfer Learning* werden Netzwerke auf einem bestimmten Datensatz aus einer Domäne vortrainiert (z. B. ein großer Datensatz von oft genutzten Radiotracern). Anschließend wird ein Teil der Netzparameter mittels eines zusätzlichen kleinen Datensatzes aus einer anderen Domäne (z. B. ein neuer Radiotracer) weiter optimiert (Parameter fine-tuning). Methoden des *Unsupervised Learnings* verzichten sogar ganz auf das Vorhandensein von Referenzbildern. Dazu gehören unter anderem die Methode des „*Deep Image Priors*“ und „konditionierte generative Netzwerke“. Letztere sind aber im Vergleich zu Methoden des „*Supervised*“ oder „*Semi-supervised Learnings*“ schwieriger zu trainieren und beim Vorhandensein von guten Referenzbildern definitiv unterlegen.

Im Vergleich zu konventionellen Verfahren zeichnen sich die DL-Ansätze durch eine sehr einfache Implementierbarkeit und hohe Flexibilität aus und sind in der visuellen Beurteilung etablierten konventionellen Verfahren häufig überlegen oder vermeiden die Notwendigkeit häufiger Nutzerintervention, wodurch Interobservervariabilität und Zeitbedarf minimiert werden. Dem gegenüber steht – wie bei allen CNN-Anwendungen – der *Black-Box* Charakter des schlussendlich operativen Algorithmus, welcher sich einer mathematischen Analyse im Wesentlichen entzieht. Dementsprechend sind auch im Kontext der Rauschunterdrückung die wesentlichen Fragen der Verifizierbarkeit und Generalisierbarkeit der in Trainings- und Validierungskohorten erzielten Ergebnisse auf die prospektiv zu verarbeitenden Patientendaten im klinischen Einsatz als wesentlich und nicht abschließend gelöst zu betrachten.

Die Gefahr besteht, dass DL-basierte Rauschreduktionsverfahren unkritisch genutzt werden, unter der Fehlwahrnehmung, dass diese es ohne jeglichen Qualitäts- oder Informationsverlust erlaubten, bei nuklearmedizinischen Untersuchungen die Aufnahmezeiten und die injizierten Aktivitäten bzw. bei radiologischen Untersuchungen die Strahlenexposition zu reduzieren. Zu diesem Punkt existieren nur wenige multizentrische Studien mit Aussagekraft für klinische Fragestellungen.

3.1.2 Bildrekonstruktion

Unter Bildrekonstruktion versteht man in der medizinischen Bildgebung die Berechnung eines medizinischen Bildes oder Bilderstapels (Volumen) aus den gemessenen Rohdaten. Im Allgemeinen handelt es sich dabei um eine Funktion, die aus der Rohdatendomäne in die Bilddomäne abbildet. Die Rohdatendomäne ist abhängig von der Modalität. Beispielweise wird bei der MR-Bildgebung im Fourierraum gemessen, bei der CT-Bildgebung im Projektionsraum (Röntgen-Transformation) und bei der PET-Bildgebung handelt es sich um Linienintegrale aus allen möglichen Richtungen. Die Bildrekonstruktion vermittelt also zwischen zwei verschiedenen Domänen und ist somit eine Lösung eines inversen Problems. Bei algorithmischen Verfahren, die lediglich ein Bild oder einen Bilderstapel verarbeiten und daraus ein verändertes Bild oder einen veränderten Bilderstapel erzeugen, spricht man üblicherweise von Bildverarbeitung oder Bildrestauration, aber nicht von Bildrekonstruktion. Bei der Verarbeitung bereits rekonstruierter Bilder spricht man auch von Nachverarbeitung. Dieser Abschnitt behandelt Bildrekonstruktionsverfahren, die auf KI-Verfahren basieren. Nachverarbeitungsverfahren sind in den Abschnitten Rauschreduktion und Superresolution behandelt. Aus Marketinggründen verknüpfen manche CT-Hersteller die (KI-freie) Bildrekonstruktion mit einer KI-basierten Bildnachverarbeitung und sprechen dann fälschlicherweise von KI-basierter Rekonstruktion.

Neben der Anwendung zur Rauschreduktion auf bereits rekonstruierten Bildern können ML-Methoden auch direkt in den tomographischen Bildrekonstruktionsprozess, welcher aus gemessenen „Rohdaten“ (in diesem Kontext hier „Projektionsdaten“) mehrdimensionale Bilder berechnet, integriert werden. Ziel dieser direkten Integration von ML-Methoden ist eine Verbesserung der resultierenden Bildqualität im Vergleich zu konventionellen Rekonstruktionsverfahren hinsichtlich verschiedener Kenngrößen wie z. B. Rauschen, Auflösung oder dem Auftreten verschiedener Bildartefakte.

In verschiedenen wissenschaftlichen Vergleichsstudien (Muckley et al. 2021, Sidky und Pan 2022) konnte gezeigt werden, dass Methoden, die ML direkt in den (iterativen) Rekonstruktionsprozess integrieren, Post-Rekonstruktions-ML-Methoden überlegen sein können. Allerdings gilt zu beachten, dass das Training von ML-Rekonstruktionsmethoden oft sehr aufwändig und die Implementierung und Anwendung dieser Methoden im Vergleich zu Post-Rekonstruktions-ML-Methoden komplexer ist, welches einer breiten Nutzung der ersteren entgegensteht.

Die in den letzten Jahren publizierten Ansätze zur Integration von ML in die medizinische Bildrekonstruktion lassen sich grob wie folgt einteilen:

1. Beim „*End-to-end Learning*“ wird ein Netzwerk trainiert, welches aufgenommene Projektionsdaten direkt in Bilder transformiert, ohne vorhandenes Wissen über die Physik der Datenakquisition oder Mathematik tomographischer Verfahren zu nutzen. Beispiele hierfür sind „AUTOMAP“ (Zhu et al. 2018) oder „DeepPET“ (Hägström et al. 2019). Im Vergleich zu anderen Methoden hat „*End-to-end learning*“ jedoch den Nachteil, dass sehr große Netzwerke und sehr große Datenmengen zum Training benötigt werden. Darüber hinaus wurde gezeigt, dass diese Ansätze häufig unter fälschlich hinzugefügten Bildstrukturen leiden und anderen weniger komplexen ML-Methoden nicht überlegen sind.
2. Im „*Physics-driven Learning*“ werden konventionelle „*Bild-Updates*“ basierend auf der Physik und Statistik des Datenaufnahmeprozesses mit gelernten Regularisierungsschritten in Form von trainierbaren Netzwerken kombiniert. Beispiele hierfür sind die so genannten „*unrolled variational networks*“ (Hammernik et al. 2018, Mehranian et al. 2021, „*Learned primal-dual*“ Rekonstruktion (Adler und Oktem 2018), „*Plug-and-*

- play“ Methoden (Ebner und Haltmeier 2024) der gelernten gefilterten Rückprojektion (Würfl et al. 2016).
3. „Konditionierte generative Modelle“ zielen darauf ab, zunächst ein generatives Modell zu trainieren, welches die Mannigfaltigkeit der möglichen hochqualitativen Bilder erlernen soll. Nach dem Netzwerktraining wird der Generierungsprozess mit den gemessenen Projektionsdaten konditioniert, um ein hochqualitatives Bild zu erhalten, welches gleichzeitig mit den Projektionsdaten „vereinbar“ ist (Chung et al. 2022, Singh et al. 2023).
 4. Im „*Post-Reconstruction Learning*“ werden Netzwerke trainiert, welche die Qualität eines „konventionell“ rekonstruierten Bildes nach der Rekonstruktion verbessern sollen. Konzeptionell ist diese Methode ähnlich der Rauschunterdrückung aus Abschnitt 3.1.1, umfasst jedoch auch die Auflösungsverbesserung und Artefaktreduktion (Berker et al. 2018).
 5. Beim „*Erlernen der Physik des Datenaufnahmeprozesses (learning of the forward model)*“ werden Netzwerke darauf trainiert, bestimmte physikalische Prozesse, welche „klassisch“ schwierig zu modellieren sind, Daten-getrieben zu erlernen. Ein Beispiel hierfür ist die Modellierung von Streustrahlung in CT (Maier et al. 2018, Maier et al. 2019b) und die PET-Rekonstruktion (Schramm et al. 2021).

Die obigen Ansätze unterscheiden sich hinsichtlich der Komplexität der erforderlichen Netzwerke, der Generalisierbarkeit sowie der Anzahl der benötigten Trainingsdaten. Es lässt sich mathematisch zeigen, dass Methoden, die mehr bekanntes Wissen über die Physik und die Statistik des Datenaufnahmeprozesses sowie die Mathematik der Tomographie nutzen (beispielsweise mit bekannten Vorwärtsoperatoren), besser generalisieren und weniger Trainingsdaten erfordern (Maier et al. 2019a).

„*Bayesianisches Learning*“ ist ein weiterer vielversprechender ML-Rekonstruktionsansatz, der darauf abzielt, neben dem rekonstruierten Bild auch eine Konfidenzkarte zu erzeugen, welche die regionale Unsicherheit des ML-Rekonstruktionsmodells repräsentiert.

Wie bei allen Netzwerken sind sog. „adversarial Attacks“ – also Angriffe, bei denen unter gezielter Nutzung von bekannten Schwachstellen Eingangsdaten gezielt manipuliert werden, um KI-Modelle absichtlich in einem bereits trainierten Netz, in die Irre zu führen – ein Problem. Vorwissen kann jedoch helfen, das Netzwerk in möglichst vielen Teilen sicher zu fixieren, damit es mit klassischen Methoden analysiert werden kann (Gottschling 2022, Huang et al. 2018).

ML-Rekonstruktionsverfahren in der Radiologie

Die Idee, vorherige Operatoren zu verwenden, kann auch für Netzwerke verwendet werden, die von der klassischen gefilterten Rückprojektion inspiriert sind, um begrenzte Winkelgeometrien besser zu approximieren, die typischerweise nicht durch klassische analytische Inversionsmodelle gelöst werden können. Interessanterweise werden Fehler in der Diskretisierung oder Initialisierung der Filterschritte durch den Lernprozess intrinsisch korrigiert. Das Verfahren ist auch mit anderen DL-Ansätzen kompatibel, die in der Lage sind, eine zusätzliche Datengetriebene Regularisierung zu lernen, im Prinzip aber eine klassische iterative Rekonstruktion durchführen.

Im Bereich der CBCT-Bildgebung, z. B. für die interventionelle Bildgebung, finden sich grundsätzlich ähnliche Ansätze wie bei der DL-basierten Rekonstruktion. In der Forschung liegt ein besonderes Augenmerk auf der Bildrekonstruktion mit limitierten Projektionswinkeln, der Metallartefaktreduktion und der Berechnung von geeigneten Trajektorien. Die optimale

Planung hat eine zur Rekonstruktion verminderte bzw. veränderte Anzahl an Projektionen zur Folge, die sowohl die Strahlenexposition positiv als auch die Auswirkung von Störfaktoren positiv oder negativ beeinflussen kann

ML-Rekonstruktionsverfahren in der Nuklearmedizin

Eine maßgebliche Anforderung an die nuklearmedizinische Schnittbildgebung ist die Gewährleistung der quantitativen Genauigkeit der Bilddaten, da die nuklearmedizinische funktionelle Bildgebung prinzipiell auf der Ableitung quantitativer Kenngrößen aus den tomographischen Bilddaten beruht. Dies bedeutet, dass das betrachtete Rekonstruktionsverfahren inklusive der oben erwähnten Streustrahlungs- und Absorptionskorrekturen nicht nur verzerrungsfreie Bilder mit unverfälschtem lokalem Kontrast liefern, sondern gleichzeitig auch die korrekte globale Skalierung der Daten gewährleisten muss. Im Hinblick auf den Einsatz ML-basierter Verfahren erhöht sich hierdurch der Bedarf an geeigneten Qualitätskontrollinstrumenten. Zurzeit konzentrieren sich die meisten ML-Rekonstruktionsverfahren in der Nuklearmedizin darauf, die Bildqualität von „*low-count*“ Aufnahmen zu verbessern. Darüberhinaus gibt es auch ML-Ansätze, um Artefakte in PET-Aufnahmen mit begrenzter Winkelgeometrie zu reduzieren (Chin et al. 2022).

3.1.3 Auflösungserhöhung (Superresolution)

Ähnliche Konzepte, wie sie zur Rauschreduktion genutzt werden, können auch zur Verbesserung der räumlichen Auflösung der Bilder eingesetzt werden. Diese Superresolution-Lösungen sind vorrangig in der Verarbeitung photographischer Bilder oder Videos zu finden. Allerdings gibt es im Bereich der CT-Bildgebung eine Anwendung des Herstellers Canon: Canons PIQE (Precise IQ Engine) ist ein neuronales Netz, das darauf abzielt, die räumliche Auflösung der CT-Daten über das Auflösungslimit des jeweiligen CT-Geräts hinaus zu steigern (Canon Medical Systems o. J.). Der vorgesehene Anwendungsbereich ist die CT-Herzbildgebung. Mit dem Aquilion Precision hat Canon ein CT-Gerät auf dem Markt, das – verglichen mit allen anderen herkömmlichen CT-Systemen außer dem neuen CT der Firma Siemens mit Photonen-zählendem Detektor – etwa die doppelte räumliche Auflösung bietet. Die hochaufgelösten Patientenbilder des Precision dienen nun als Trainingslabel. Der Trainingsinput ist eine geglättete Version der hochaufgelösten Bilder.

Allerdings gilt bei derartigen Verfahren, dass sie nicht wirklich die Auflösung des abbildenden Systems verbessern, sondern lediglich Kanten steiler erscheinen lassen. Dies ist insbesondere in der CT wichtig, um Bloomingartefakte zu unterdrücken. Darunter versteht man den Effekt, dass Kalzifikationen in Gefäßen aufgrund der engen Graustufenfensterung, mit der die Gefäße typischerweise betrachtet werden, größer aussehen, als sie eigentlich sind. Dies führt dazu, dass die Radiologin oder der Radiologe tendenziell das Lumen des Gefäßes an der Stelle der Kalzifikation zu klein einschätzt. Mit Hilfe von Superresolution-KI-Verfahren würde dieser Effekt nicht auftreten und das Lumen korrekt eingeschätzt.

Was solche Verfahren nicht leisten können, ist die tatsächliche Erhöhung der räumlichen Auflösung: zwei dicht benachbarte Läsionen, die im CT-System (ohne KI) aufgrund der limitierten Auflösung nur als eine Läsion erscheinen, sind auch nach der Anwendung von auflösungserhöhenden KI-Verfahren als eine einzige Läsion sichtbar.

3.1.4 Artefaktreduktion bei CT und Röntgendiagnostik

3.1.4.1 Fehlende Daten

Ein Teil der CT-Artefakte resultiert aus fehlenden Daten. Beispielsweise entstehen Metallartefakte dadurch, dass bestimmte Objektbestandteile den Röntgenstrahl so stark schwächen, dass

im Metallschatten keine nutzbare Information gemessen werden kann. Die Projektionen haben also eines oder mehrere Löcher, d. h. Bereiche fehlender Daten, die beispielsweise durch ein Inpaintingverfahren geschlossen werden müssen.

Ein weiteres Beispiel sind Trunkierungsartefakte. Sie entstehen dadurch, dass das Objekt größer ist als der CT-Detektor. Trunkierungsartefakte kommen insbesondere bei CBCT vor, da heutige Flachdetektoren meist nur ca. 40 cm groß sind und somit der Messfelddurchmesser auf Grund der Strahlengeometrie bei Anwendung des Strahlensatzes ca. 20 cm beträgt. Um die Artefakte zu reduzieren, muss der Detektor virtuell vergrößert und mit sinnvollen Daten aufgefüllt werden, d.h. dass Daten um das tatsächliche Messfeld herum ergänzt werden.

In der Literatur finden sich zahlreiche Vorschläge, die Strahlenexposition dadurch zu reduzieren, dass weniger Projektionen pro Umlauf aufgenommen werden, bei gleichbleibender Dosis pro Projektion. Auch wenn es naheliegender und sinnvoller wäre, statt der Anzahl der Projektionen einfach die Dosis pro Projektion zu reduzieren, können die entstehenden Artefakte entweder durch ein Inpaintingverfahren, welches die fehlenden Projektionen oder Daten ergänzt, oder durch ein spezielles Rekonstruktionsverfahren reduziert werden.

Letztendlich lassen sich für die o. g. Anwendungsfälle, bei denen Daten oder Datenbereiche fehlen, beschädigt oder fehlerhaft und somit unbrauchbar sind, neuronale Netze einsetzen, die diese Datenlücken auffüllen. Die in der Literatur beschriebenen Lösungsansätze tun dies entweder direkt im Rohdatenraum, also in den Projektionsdaten oder sie führen eine Korrektur der entsprechenden Artefakte im Bildraum durch und nutzen dieses korrigierte Volumen für eine Vorwärtsprojektion. Aus den vorwärtsprojizierten Daten werden dann die in den Messdaten fehlenden Daten genommen und in die Lücken der Originaldaten kopiert. So wird garantiert, dass in Originaldaten lediglich die fehlenden Bereiche ergänzt werden. Korrekturalgorithmen, die lediglich aus dem artefaktbehafteten Volumen ein optisch ansprechenderes errechnen, sind als unsicher anzusehen, da sie auch Teile der guten Rohdaten verändern.

Beispiele für derartige Verfahren sind beschrieben (Byl et al. 2021) im Bereich der Metallartefaktkorrektur, (Fonseca et al. 2021, Kabelac et al. 2025, Ketola et al. 2021) für die Detrunkierung von CT-Daten mittels neuronaler Netze sowie (Han et al. 2017, Kofler et al. 2020) und (Ronneberger et al. 2015) im Bereich der schwach besetzten Projektionen.

3.1.4.2 Fehlerbehaftete Daten

Ein weiterer Teil der CT-Artefakte stammt von fehlerbehafteten Daten, wie zum Beispiel von nicht korrekt kalibrierten Detektoren, zeitlich variierenden Detektoreigenschaften, Streustrahlung oder Strahlaufhärtung.

CT-Ringartefakte beispielsweise entstehen durch Detektorpixel, deren Antwortverhalten während der Messung sich von dem während der Kalibrierung unterscheidet. Verantwortlich dafür können Temperaturschwankungen oder die Expositionshistorie des Pixels oder seiner Nachbarn sein. Die Korrektur kann auf klassische Art oder durch neuronale Netze entweder im Rohdatenraum oder im Bildraum oder einer Kombination aus beiden durchgeführt werden. Ein Beispiel für ein solches Ringartefaktkorrekturverfahren mittels neuronaler Netze ist beschrieben in (Trapp et al. 2022).

Streustrahlartefakte entstehen durch die Streustrahlung, die im Objekt oder in den Vorfiltern entsteht und den Detektor aus einer Richtung erreicht, die nicht der der Primärstrahlung entspricht. Auch beim Einsatz von Streustrahlrastern entstehen Streustrahlartefakte. In Röntgenbildern führen die Streustrahlartefakte zu einem Kontrastverlust, der aber, wie oben beschrieben, durch Bildverarbeitungsmaßnahmen korrigiert werden kann. Eine physikalisch richtige Korrektur ist dort nicht nötig. Um die Streustrahlartefakte im CT zu korrigieren, muss hingegen

der Streustrahlanteil möglichst korrekt geschätzt und von den Rohdaten abgezogen werden. Der Goldstandard für diese Schätzung ist der Einsatz von Monte-Carlo-Streustrahlberechnungen, die jedoch Rechenzeiten von Stunden oder Tagen benötigen und zudem das Vorhandensein eines 3D-Modells des Patienten oder der Patientin, also beispielsweise eines CT-Volumens, erfordern. Sie können daher nicht eingesetzt werden, um einzelne Projektionsbilder (z. B. Röntgenprojektionsbilder oder Fluoroskopiebilder) zu korrigieren. Im Bereich der KI ist die Deep Scatter Estimation (DSE) der Standard, um Streustrahlung zu schätzen (Maier et al. 2018, Maier et al. 2019b). Die DSE benötigt kein 3D-Modell und kann somit auch für Projektionsbilder eingesetzt werden. Zudem ist die Laufzeit mit ein bis zwei Millisekunden pro Projektion sehr kurz. Basierend auf den Ideen der DSE finden sich zahlreiche Folgeveröffentlichungen in der wissenschaftlichen Literatur (Laurent et al. 2023, Xiang et al. 2020).

3.1.4.3 Bewegungskorrektur und -kompensation in der Röntgendiagnostik

Patientenbewegungen während der Untersuchung führen zu Artefakten und Auflösungsverlust in den rekonstruierten tomographischen Bildern, wobei deren Art und Umfang je nach Modalität unterschiedlich ausgeprägt sind. DL-Verfahren kommen typischerweise zum Einsatz, um Bewegungsvektorfelder zu bestimmen (Displacement Vector Fields: DVF), welche die zeitabhängige dreidimensionale, willkürliche und unwillkürliche Bewegung und Deformation (Organbeweglichkeit, Atem- und Herz-Zyklus) beschreiben. Prinzipiell lassen sich Verfahren zur Bewegungskorrektur in zwei Kategorien aufteilen: solche, welche die der Bildrekonstruktion zugeführten Projektionsdaten vorab modifizieren, um die Patientenbewegung zu kompensieren und solche, welche auf den bewegungskompromittierten rekonstruierten tomographischen Bilddaten operieren. Im radiologischen Bereich ist der Begriff der Bewegungskorrektur für Postrekonstruktionsverfahren reserviert, die im Wesentlichen eine Verbesserung des optischen Bildeindrucks bewirken. Wo dieses möglich ist, kann die Korrektur der Projektionsdaten während der Bildrekonstruktion erfolgen. Dieses wird oft mit dem Begriff Bewegungskompensation bezeichnet. Diese begriffliche Unterscheidung ist im nuklearmedizinischen Umfeld jedoch unüblich, hier spricht man in beiden Fällen von Bewegungskorrektur.

Bei CT-Aufnahmen können Patientenbewegungen zu Bewegungsartefakten sowohl in Projektionsbildern als auch in den rekonstruierten CT-Volumina führen. Zahlreiche algorithmische Verfahren zur Reduktion oder Korrektur derartiger Artefakte sind in der Literatur zu finden. Im Bereich der KI sind diese Verfahren oft kosmetischer Natur und erzeugen somit ein gutaussehendes Ergebnis, dessen Wahrheitsgehalt sich jedoch kaum überprüfen lässt. Nach Möglichkeit sollten Verfahren zur Bewegungskompensation eingesetzt werden. Dort wird, im Gegensatz zur Bewegungskorrektur, die tatsächliche Bewegung eines Voxels geschätzt und fließt dann zur Berechnung der kompensierten Bilder ein, beispielsweise, indem sich das jeweilige Voxel während der Bildrekonstruktion entsprechend dem Patienten oder der Patientin bewegt, wodurch dann die Beiträge bei der Rückprojektion immer auf das richtige Voxel addiert werden. Bei der Bewegungskompensation ist die Aufgabe des neuronalen Netzes meist die der Bewegungsschätzung, also der Erzeugung eines Bewegungsvektors oder eines Bewegungsvektorfelds (Maier et al. 2021, Maier et al. 2025). Auch hier kann es zu einer fehlerhaften Schätzung kommen. Im Gegensatz zur reinen Korrektur lassen sich jedoch Plausibilitätsprüfungen durchführen, z. B. im Hinblick auf die Bewegungsgeschwindigkeit.

3.1.5 Bewegungskorrektur in der Nuklearmedizin

Auch in der Nuklearmedizin ist der Einfluss der unvermeidbaren willkürlichen und unwillkürlichen Patientenbewegung während der Akquisition ein für die resultierende Bildqualität relevanter Aspekt. Dies betrifft die nuklearmedizinische Bildgebung aufgrund der im Vergleich zur

CT weitaus längeren Aufnahmezeiten in besonderem Maß. Sowohl zyklische als auch stochastische Bewegung (Unruhe) führt zu Auflösungsverlust. Das in der Praxis wesentlichste Problem ist die atemzykluskorrelierte Bewegung bei Ganzkörperuntersuchungen von onkologischen Patientinnen oder Patienten, die typische Amplituden von 2 cm aufweist. Neben sensorgestützten stroboskopischen Messprotokollen („Atemgating“) gewinnen rein datengetriebene Ansätze zunehmend an Bedeutung. Gerade in diesem Kontext sind DL-basierte Verfahren aus ähnlichen Gründen wie bei der Rauschunterdrückung und Auflösungsrückgewinnung im Rahmen der Bildrekonstruktion attraktiv und zum Teil sehr ähnlich angelegt, da das Problem der Bewegungskorrektur formal eine Entfaltung des von Bewegungsunschärfe betroffenen Bildes erfordert.

3.1.6 Weitere Verfahren zur Reduktion der Strahlenexposition in der Diagnostik

Unter Methoden zur Reduktion der Strahlenexposition in der Diagnostik versteht man alle Verfahren, die darauf abzielen, die Strahlenexposition bei diagnostischen bildgebenden Verfahren zu verringern. Neben den bereits genannten Rekonstruktionsverfahren, die die Strahlenreduktion durch Reduzierung der Anzahl der Projektionen oder die Reduktion der benötigten Photonen pro Projektion bei gleichbleibender Bildqualität zum Ziel haben, gibt es Ansätze, die bereits bei der Projektionsaufnahme KI-Verfahren anwenden. Darunter fallen die Optimierung der Kollimation sowie die automatisierte Patientenpositionierung. Besonders für die 2D Röntgendiagnostik werden bereits Systeme angeboten, die auf KI-Methoden basieren. Als Beispiel sind hier das von Siemens eingesetzte Verfahren „YSIO X.pree“ (Siemens Healthineers AG o. J.-a) zu nennen, welches die Position des Thorax mit Hilfe einer 3D Kamera und KI-Methoden bestimmt und darauf basierend die Kollimation einstellt. Omega medical imaging bietet eine automatische Berechnung der Kollimatoreinstellung basierend auf einer Zielregions-Bestimmung an (Omega Medical Imaging 2023). Im Bereich der CT-Bildgebung bieten verschiedene Hersteller (Siemens FAST Integrated Workflow™ (Nikolaus 2018), Philips CT Precise Suite™ (Philips GmbH 2021), GE Revolution™ Maxima (GE Healthcare o. J.) eine automatische Patientenpositionierung an, um die Zielstruktur in das Isozentrum zu verschieben. Ein Kamerasystem mit Tiefeninformation wird für eine 3D Patientenrekonstruktion verwendet und daraus die Zielstruktur innerhalb des Patienten bzw. der Patientin für die Positionierung abgeleitet. Genauere Angaben zu KI-Methoden der Hersteller lassen sich allerdings nicht finden.

3.1.7 Segmentierung

Die jüngsten Entwicklungen im Bereich des Deep Learnings haben der Bildsegmentierung, bei der die Umrisse von Organen oder anatomischen Strukturen genau bestimmt werden, große Vorteile gebracht. CNNs haben dieses Feld dominiert, und mehrere Ansätze wurden unter Verwendung von CNNs entwickelt. Beispiele hierfür sind Deeporgan, Gehirn-MR-Segmentierung mit CNNs oder vollständig faltende Multi-Energie 3-D-U-Netze. Insbesondere das so genannte U-Net (Ronneberger et al. 2015) hat sich hier als wesentlicher Stand-der-Technik etabliert. Untersuchungen auf unterschiedlichsten Problemstellungen haben gezeigt, dass die ursprüngliche Variante aus dem Jahr 2015 mit automatischer Parametereinstellung Spezialisierungen für gewisse Problemstellungen im Allgemeinen überlegen ist. Diese sich selbstkonfigurierende Variante von U-Net firmiert unter dem von den Autoren eingeführten Namen nn-U-Net (Isensee et al. 2021).

Ansätze zu Gray-Box-Modellen existieren bereits seit 2004 und experimentierten beispielsweise mit der Fusion von neuronalen Netzen und aktiven Konturmodellen. Allerdings wurde dort noch komponentenweise vorgegangen (also das neuronale Netz nur für einen Teil der Aufgabe trainiert und das Konturmodell separat optimiert) und keine tiefen Netze oder End-to-End-

Trainings (also für die gesamte Aufgabe) eingesetzt. Es scheint vielversprechend zu sein, traditionelle Segmentierungsansätze zu überdenken und sie mit Deep Learning basierend auf einem End-to-End-Training zu verschmelzen. Dies ermöglicht es, die Zahl der Parameter deutlich zu reduzieren und trotzdem stabile Ergebnisse zu erhalten. Allerdings zeigen die hybriden Verfahren noch nicht die Leistungsfähigkeit von Black-Box Modellen.

Insgesamt sind Techniken der Segmentierung als wichtig einzuschätzen, da diese für die Konturierung von Organen und folgenden Dosisschätzungen sowie die Planung von Strahlentherapie grundlegend sind. Ein Bezug zur Strahlenschutzrelevanz im Sinne des oben beschriebenen Beratungsauftrags ist zunächst nicht zu sehen, auf Grund der Bedeutung für strahlentherapeutische Verfahren ist eine spätere genauere Betrachtung aber notwendig.

3.1.8 Klassifizierung

Basierend auf den segmentierten Bildern oder auch basierend auf der Erkennung von bestimmten Strukturen oder Strukturclustern wird in der medizinischen Bildgebung für verschiedene Fragestellungen an Methoden zur Klassifizierung gearbeitet oder auch bereits entsprechende Modelle in kommerziell verfügbaren Analyseprodukten eingesetzt. In besonderem Maße trifft dies auf Verfahren der bildgebenden Darstellung der Brust, also vor allem der Projektionsmammographie oder der Brusttomosynthese zu. Wie bereits erwähnt, werden anhand verschiedener Charakteristika entweder Bildbereiche oder ganze Bilder in Klassen eingeteilt. Diese Klassen reichen von der Zuordnung zu bestimmten Organen, über Risikogewebe oder suspekter Gewebereich bis hin zur Definition von Tumorgewebe oder erkranktem Gewebe insgesamt. Für die mammographische Bildgebung wird insbesondere im Screening die Klassifizierung in suspekte Bereiche hinsichtlich möglicher Tumorerkrankungen und potenziell normalen Gewebes vorgenommen, es gibt aber auch Klassifizierungen für andere Bereiche des Körpers, in denen andere Klassifizierungen eine Rolle spielen, zum Beispiel für die Erkennung von Perfusionsstörungen oder von Emphysemen. Die verwendeten Verfahren basieren heute vor allem auf der Verwendung von großen Datensätzen mit bekannten Erkrankungen und korrespondierenden Bilddatensätzen. Es werden sowohl CNN-Ansätze, häufig U-NET-basierte Verfahren oder Erweiterungen aber auch fast alle anderen DL-basierten Ansätze für derartige Klassifizierungsstrategien verwendet. Da im Allgemeinen diese Anwendungen die Bildinformation in heute verfügbaren Systemen nicht verändern, sondern genutzt werden, um auf potenziell diagnostisch relevante Information hinzuweisen, ist mit dieser Art der Klassifizierung zunächst kein strahlenschutzrelevanter Aspekt verbunden. Zum Teil wird aber an Klassifizierungssystemen gearbeitet, bei denen basierend auf der Klassifizierung Entscheidungen über den weiteren Behandlungsverlauf inklusive der Anwendung ionisierender Strahlung entschieden wird. Als Beispiel sei hier die Klassifizierung von gesundem, von Tumorgewebe und von zerstörtem Tumorgewebe in der minimalinvasiven Tumorthapie benannt (Mahmoodian et al. 2022). Da in diesem Anwendungsfall z. B. über die weitere Ablation unter röntgenbasierter Bildgebung entschieden wird, könnten sich strahlenschutzrelevante Aspekte ergeben. Diese Fragestellung ist aber nicht im Beratungsauftrag spezifiziert. Da zudem solche Methoden Gegenstand gerade noch in Entwicklung befindlicher Ansätze sind, wird dieser Aspekt hier nicht weiter untersucht. Er sollte aber für zukünftige Überlegungen nicht vergessen werden.

3.1.9 Radiomics

Noch einen Schritt weiter als die einfache Klassifizierung geht das mit „Radiomics“ bezeichnete Verfahren und alle ähnlichen Ansätze. Die Annahme von Radiomics ist, dass in den Bilddaten die mit dreidimensionalen oder vierdimensionalen Datensätzen gewonnenen Informationen derart vielfältig und unüberschaubar sind, dass DL-Methoden besser als eine rein mensch-

liche Evaluierung der großen Datenmengen in der Lage sind, subtile Bildinformation zu detektieren und Erkrankungen zuzuordnen. In der Erarbeitung von Radiomics-Verfahren werden daher verschiedene Parameter und Charakteristika von Bilddaten ganz allgemein bestimmt und anhand großer Fallzahlen geprüft, ob es Verbindungen zwischen solchen quantifizierbaren Charakteristika oder Parametern (im allgemeinen als Features bezeichnet) und bestimmten Erkrankungen gibt. Der Unterschied zu den vorher im Bereich der Klassifizierung beschriebenen Verfahren besteht im Wesentlichen darin, dass man nicht mehr Bereiche klassifiziert und Annahmen über die Ursache der Klassifizierung macht, sondern dass das Muster, welches zur Erkennung führen kann, aus vielen nicht direkt zugänglichen Informationen wie z. B. Form von Strukturen, Nachbarschaftsanalysen, Strukturgradienten und vielen anderen besteht (Lambin et al. 2012). In Bezug auf die strahlenschutzrelevante Bewertung der Radiomicsverfahren gelten die gleichen Betrachtungen wie für die Klassifizierungsansätze.

3.1.10 Dosisberechnung

Die Berechnung patientenspezifischer Dosisverteilungen bei der Röntgenbildgebung und der Strahlentherapie basiert auf der räumlichen Verteilung $\mu(x, y, z)$ der Schwächungskoeffizienten der Gewebe der Patientin bzw. des Patienten. Die derzeit beste Methode für die Dosisverteilungsberechnung ist die Monte-Carlo-Simulation aller Teilchenhistorien innerhalb der Patientin bzw. des Patienten. Derartige Monte-Carlo-Berechnungen sind jedoch rechen- und somit zeitaufwändig, so dass sie oft durch Näherungsverfahren, wie z. B. kernelbasierter Dosisberechnung, ersetzt werden müssen. In Zusammenhang mit der CT werden zurzeit Dosisverteilungen ausschließlich offline und nur zu wissenschaftlichen Zwecken berechnet. Weit verbreitete und teils routinemäßig erfasste Dosiswerte, wie der CTDI-Wert, das Dosislängenprodukt (DLP), die größenpezifischen Dosis-schätzungen (SSDE) oder die über den k-Faktor aus dem DLP gerechnete effektive Dosis sind nicht patientenspezifisch, sondern beziehen sich auf Dosiswerte eines Phantoms gefolgt von einigen einfachen Umrechnungen.

Zur Optimierung von Scanprotokollen kann es jedoch sinnvoll sein, die Dosisverteilung im Patienten bzw. der Patientin schon im Voraus zu bestimmen. Beispielsweise gibt es im CT die so genannte Röhrenstrommodulation (tube current modulation – TCM). Wären bereits vor dem Scan die Schwächungswertverteilung des Patienten bzw. der Patientin und somit die Lage der Organe grob bekannt, und könnte dann in sehr kurzer Zeit eine Dosisverteilung berechnet werden, so ließe sich die Röhrenstrommodulation mit deutlich verringerter Strahlenexposition und damit verringertem Strahlenrisiko für Patienten bzw. Patientinnen durchführen (Baader und Kachelrieß 2025, Klein et al. 2022). Die Voraussetzungen dafür lassen sich mit Hilfe von KI-Verfahren erreichen, z. B. durch die Berechnung eines CT-Volumens aus dem Topogramm (also dem zuerst durchgeführten Scan als Projektionsbild zur Einstellung des Scanbereichs) gefolgt von einer Dosisberechnung in Echtzeit. Es ist daher abzusehen, dass zukünftig KI-Verfahren eingesetzt werden, die direkten Einfluss auf die Scandurchführung und somit auf die Strahlenexposition haben.

Zur Dosisberechnung selbst kann die KI folgendermaßen (Maier et al. 2022b) eingesetzt werden: Basierend auf dem CT-Volumen des Patienten oder der Patientin und der Scangeometrie wird eine Dosisverteilung erster Ordnung berechnet, d. h. eine Dosisverteilung wie sie ohne Streustrahlung aussehen würde. Diese Rechnung entspricht einer Rückprojektion der Strahlenintensität in das Volumen und lässt sich im Subsekundenbereich durchführen. Basierend auf dem CT-Volumen und aus dem First-Order-Dosisvolumen rechnet dann ein neuronales Netz, das beispielsweise mit Monte-Carlo-Rechnungen trainiert wurde, die Dosisverteilung inklusive aller höheren Ordnungen aus.

3.2 Strahlentherapie

Der Einsatz der KI in der Strahlentherapie außer zum Einsatz in der Bildrekonstruktion und Rauschreduktion für diagnostische Anwendungen wie zur Bestrahlungsplanung und zum Staging ist nicht Gegenstand dieser Empfehlung. Es sei hier angemerkt, dass es für die Bildrekonstruktion und die Rauschreduktion mittels KI beim Einsatz der CT-Bildgebung für die Bestrahlungsplanung neben den Anforderungen, die für die zu Diagnosezwecken erstellten Bilder gelten, zudem besonders wichtig ist, dass sich die Hounsfield-Einheiten in den rekonstruierten Bildern durch die KI-Methoden nicht signifikant verändern. Aus Gründen der Vollständigkeit werden weitere potenziell relevante Einsatzbereiche der KI in der Strahlentherapie dennoch kurz beleuchtet.

3.2.1 Pseudo-CT

In der Strahlentherapie bedeutet der Begriff Pseudo-CT, dass ein künstliches CT-Bild auf Basis eines realen MR-Bildes erzeugt wird. Angewandt wird das Verfahren, falls für den Patienten bzw. die Patientin keine CT vorliegt oder diese aus strahlenhygienischen Gründen nicht zu häufig wiederholt werden soll. Somit kann die Pseudo-CT als Hilfsmittel für die Planung oder deren Optimierung dienen, ohne die Patientin oder den Patienten Strahlung auszusetzen. Mit Hilfe von registrierten MR-CT Bildpaaren kann ein neuronales Netz trainiert werden, welches ausgehend von den MR-Bildern entsprechende Hounsfield-Unit-Karten erzeugt. Es gibt ein paar bereits kommerziell eingesetzte Beispiele: Das System von Siemens, Synthetic CT™ (Siemens Healthineers AG o. J.-b), kombiniert ein U-Net zur Gewebesegmentierung und ein cGAN (conditional Generative Adversarial Network) zur CT-Erzeugung. Dieses Netzwerk besteht aus einem Generator, der Bilder erzeugt und einem konkurrierenden Diskriminator, der die Wirklichkeitstreue des Bildes evaluiert. Spectronic Medical AB (integriert in die Systeme von Canon und GE Healthcare (Spectronic Medical AB o. J.)) schätzt laut Hersteller basierend auf mehreren 3D CNNs eine lokale affine Transferfunktion. MRCAT™ Brain von Philips (Philips Austria GmbH o. J.) wurde für Anwendungen im Kopf- und Hals-Bereich entwickelt und erzeugt ein synthetisches CT-Bild; allerdings wurde die zugrundeliegende Methode nicht offengelegt. Neben klassischen U-Net-basierten Ansätzen finden sich in der jüngeren Literatur auch vermehrt Verfahren, die auf den so genannten Denoising Diffusion-Modellen basieren, einer weiteren Kategorie von generativen Modellen, die schrittweise Rauschen zu einem strukturierten Bild hinzufügen, um damit neue Datensätze zu erzeugen und damit zu lernen, Datensätze zu entrauschen.

3.2.2 Dosisberechnung

Ähnlich wie in Abschnitt 3.1.10 im Falle der CT beschrieben kann auch die Dosisberechnung in der Strahlentherapie durchgeführt werden (Martins et al. 2023). Basierend auf der Schwächungswertverteilung des Patienten bzw. der Patientin, also dem CT-Volumen, wird beispielsweise zunächst ein näherungsweise recheneffizientes Dosisberechnungsverfahren eingesetzt, um eine angenäherte Dosisverteilung zu erhalten. Ein neuronales Netz rechnet diese dann in eine realistische Dosisverteilung um. Andere Ansätze betrachten lediglich die Dichteverteilung des Patienten bzw. der Patientin (also ein CT-Bilddatensatz oder ein Pseudo-CT-Bilddatensatz) und die Information zum Therapiestrahle, also den Bestrahlungsplan, und ein neuronales Netz errechnet daraus die Dosisverteilung. Auf eine initiale Berechnung einer angenäherten Dosisverteilung wird dabei verzichtet.

3.2.3 Decision-Support-Systeme

Wie in den Abschnitten 3.1.8 zur Klassifizierung und 3.1.9 zu Radiomics bereits beschrieben werden bildbasierte Ansätze zusammen mit DL-Verfahren dazu verwendet, bestimmte Informationen aus den Bilddatensätzen zu extrahieren, die geeignet sind, die Diagnose durch Hinweise zu unterstützen. Diese Verfahren werden auch genutzt, um Vorschläge zur Therapieoptimierung zu generieren. Dies betrifft dann auch strahlentherapeutische Verfahren, sei es bei minimalinvasiven oder brachytherapeutischen Verfahren oder zur (Online-) Optimierung teletherapeutischer Verfahren – zum Beispiel durch Veränderung der Bestrahlungspläne, oder durch Bestimmung neuer Bestrahlungsparameter zur besseren Einhaltung von Bestrahlungsplänen. Die Systeme schlagen im Allgemeinen Diagnoseoptionen oder Behandlungsoptionen und oder Änderungen im Behandlungsschema vor und sollen so dem Arzt oder der Ärztin die Therapieentscheidung erleichtern. Zudem werden bereits automatisiert Bestrahlungspläne erstellt. Auch hier kommen praktisch alle gängigen Verfahren der DL-basierten Ansätze zum Tragen. Insbesondere für die adaptive Strahlentherapie spielen diese Ansätze auf Grund der wachsenden Möglichkeiten der online zur Verfügung stehenden Bildgebung und der zur Verfügung stehenden Algorithmen eine stetig wachsende Rolle. Auf eine Auflistung der aktuell kommerziell verfügbaren Algorithmen und Produkte wird auf Grund der rasanten Entwicklung in diesem Feld verzichtet. Die Decision-Support-Systeme, die bei strahlentherapeutischen Anwendungen eingesetzt werden, sind aus Sicht des Strahlenschutzes relevant, da sie die radioonkologische Behandlung verändern können. Allerdings liegt die endgültige Entscheidung beim fachkundigen Arzt oder bei der Ärztin, so dass zurzeit ein direkter strahlenschutzrechtlicher Aspekt nicht eindeutig gegeben ist. Eine erneute Bewertung wird erforderlich sein, sobald auf Basis der Änderungsvorschläge für die Behandlungsschemata die Entscheidungen so schnell getroffen werden sollen, dass eine echte Neubewertung durch den Arzt oder die Ärztin nicht mehr erfolgen kann, was für die Zukunft im Sinne eines besseren Patientenmanagements denkbar ist.

4 Rahmenbedingungen

4.1 Rechtliche Rahmenbedingungen

Bei der Betrachtung der Anwendung künstlicher Intelligenz in der strahlenbasierten Medizin und insbesondere für die Anwendung von Rauschreduktionsverfahren und Rekonstruktionsverfahren für die Röntgenbildgebung sind aus rechtlicher Sicht vier Aspekte von besonderer Bedeutung. Erstens ist eine Strahlenanwendung am Menschen nur dann zulässig, wenn sie mit einem erwarteten Nutzen verbunden ist. Nähere Einzelheiten in Bezug darauf werden im folgenden Abschnitt 4.1.1 zur „rechtfertigenden Indikation“ behandelt. Zweitens besteht ein Gebot zur Optimierung der Exposition der Patienten und Patientinnen bei der medizinischen Anwendung ionisierender Strahlung. Die Implikationen dieses Gebots werden im Abschnitt 4.1.2 „Optimierungsgebot“ behandelt. Drittens ist zu berücksichtigen, dass Haftungsfragen entstehen, wenn aufgrund einer hinzugefügten oder unterdrückten Bildinformation eine falsche klinische Entscheidung getroffen wird. Viertens ist zu bedenken, welche rechtlichen Gegebenheiten sich für die Entwicklung der Modelle und das Training zur Anwendung künstlicher Intelligenz aus den Datenschutzvorgaben der DSGVO (EU 2016) ergeben, da hier mit Patientendaten gearbeitet werden muss. Eine entsprechende Einschätzung findet sich in Abschnitt 4.1.4.

Neben den rechtlichen Rahmenbedingungen spielen möglicherweise auch ethische Aspekte der Anwendung künstlicher Intelligenz für die medizinische Bildgebung eine Rolle. Dies betrifft naturgemäß die Verwendung von Patientendaten für das Training der Modelle, die später zur

Bildverarbeitung verwendet werden sollen, und hat somit einen Bezug zu den Datenschutzvorgaben. Daneben sind Fragen zu betrachten, die mit der gleichen Verfügbarkeit von optimierten Verfahren und somit auch potenziell reduziertem Risiko durch Anwendung ionisierender Strahlung verbunden sind. Möglicherweise sind KI-basierte Verfahren, die eine Dosisreduktion bei der Bildgebung erlauben, nicht überall oder für alle Patientinnen oder Patienten verfügbar, weil sie teuer sind, nur auf bestimmten Geräten sicher eingesetzt werden können oder die Patientinnen oder Patienten zu so genannten Out-of-Distribution Fällen gehören.

Umgekehrt sind mögliche Fehldiagnosen und die Verantwortung dafür nicht nur mit rechtlichen Aspekten assoziiert, sondern auch mit ethischen. Für die Betrachtung strahlenschutzrelevanter Fragen spielen diese Aspekte jedoch eine untergeordnete Rolle. Dies gilt auch für Decision Support-Systeme, die ohnehin nicht Gegenstand dieser Empfehlung sind und im Abschnitt 3.2.3 nur der Vollständigkeit halber betrachtet wurden. Deren ethische Dimension dürfte sich nicht von bisherigen Verfahren unterscheiden, solange die letztendliche klinische Entscheidung von behandelndem Arzt bzw. behandelnder Ärztin zusammen mit der Patientin bzw. dem Patienten getroffen wird.

4.1.1 Rahmenbedingungen aufgrund der rechtfertigenden Indikation

Die Anwendung ionisierender Strahlung am Menschen im Rahmen der bildgebenden medizinischen Diagnostik ist nur aufgrund einer rechtfertigenden Indikation zulässig. Diese ist gegeben, wenn die potenziellen Risiken der Strahlenanwendung durch ihren erwarteten Nutzen aufgewogen werden (§ 83 Absatz 3 StrlSchG, (StrlSchG 2017)). Hierfür muss der fachkundige Arzt bzw. die fachkundige Ärztin, der bzw. die die rechtfertigende Indikation stellt, die Aussagekraft und mögliche Einschränkungen der jeweiligen Untersuchung einschätzen können. Dies ist bei bildgebenden Verfahren ohne Mitwirkung maschinellen Lernens möglich, weil hinreichend einschlägige Erfahrung damit vorliegt und die mathematischen Verfahren, die aus den Rohdaten die diagnostischen Bilder berechnen, definiert und bekannt sind.

Wenn hingegen die Erzeugung von Bild- aus Rohdaten auf Verfahren des maschinellen Lernens beruht, und der „Lösungsweg“ weder vorhersehbar ist noch im Nachhinein offengelegt wird, dann ist eine Einschätzung der Aussagekraft und Einschränkungen mit Unsicherheiten behaftet. Gerade wenn die Rohdaten mit einem starken Rauschanteil behaftet und die daraus konventionell rekonstruierten Bilder nur mit Einschränkungen interpretierbar sind, kann nicht ausgeschlossen werden, dass in den mit KI-Verfahren erzeugten Bildern diagnostisch wichtige Details nicht mehr enthalten sind oder dass sie Details enthalten, die in Wirklichkeit nicht vorhanden sind.

Für die Stellung der rechtfertigenden Indikation kann das bedeuten, dass bei Anwendung von KI-Verfahren zur Bildrekonstruktion aus Rohdaten mit hohem Rauschanteil die Einschätzung eines möglichen Nutzens mit Unsicherheit behaftet ist, sofern nicht Maßnahmen getroffen werden, um sich davon zu überzeugen, dass die Bildinformation nicht unwillentlich verfälscht worden ist.

Hieraus ergibt sich zusammen mit einem Hinweis zur Nutzung von Bilddaten für die Bestrahlungsplanung in Kapitel 3.2 die Empfehlung 1 der SSK in Kapitel 2:

Die SSK empfiehlt für die Verwendung KI-basierter Verfahren für die Bildrekonstruktion oder Bildverarbeitung bei Verfahren der medizinischen Bildgebung, die auf dem Einsatz ionisierender Strahlung basieren, dass

- bei Anwendung von KI-Verfahren zur Bildrekonstruktion oder Weiterverarbeitung von patientenbezogenen Bilddaten Sorge zu tragen ist, dass alle befundrelevanten Strukturen mit einer im Vergleich zu nicht KI-gestützten Verfahren im wesentlichen unveränderten

Sensitivität und Spezifität in Bezug auf die diagnostische Sicherheit erhalten bleiben und ferner keine artifiziellen Strukturen generiert werden, um bildverarbeitungsbedingte Fehldiagnosen auszuschließen, die den Nutzen der Untersuchung und damit die rechtfertigende Indikation in Frage stellen würden. Für den Einsatz computertomographischer Bilddaten für die Bestrahlungsplanung in der Strahlentherapie dürfen die Hounsfield-Einheiten nicht signifikant verändert werden.

4.1.2 Rahmenbedingungen aufgrund des Optimierungsgebots

Die aktuelle Strahlenschutzgesetzgebung sieht zwar bei der Anwendung ionisierender Strahlung zu diagnostischen oder therapeutischen Zwecken keine Grenzwerte für die Exposition von Patientinnen oder Patienten vor, sehr wohl aber ein Optimierungsgebot (§ 83 Absatz 5 StrlSchG, (StrlSchG 2017)). Dies heißt insbesondere auch, dass Verfahren der diagnostischen Bildgebung dahingehend zu optimieren sind, dass bei Einhaltung der diagnostischen Aussagekraft die Exposition mit ionisierender Strahlung so weit wie möglich reduziert wird. Diese Reduktion der Strahlenexposition ist durch physikalisch-technische Grenzen limitiert. Allerdings könnten KI-, insbesondere DL-basierte Verfahren geeignet sein, diese Grenzen auch wirklich auszuschöpfen. In diesem Sinne wäre, falls dies nicht durch andere Verfahren ebenfalls zu gewährleisten wäre, die Verwendung DL basierter Verfahren möglicherweise sogar geboten. Es muss aber dabei sichergestellt werden, dass

- die physikalischen Grenzen nicht überschritten werden (also ausreichend Patienten- und Situationsbezogene Information vorliegt, um reale Information darstellen zu können) und
- keine relevante Information bei der Optimierung verloren geht oder zusätzliche nicht gegebene Information hinzugefügt wird, weil sonst die diagnostische Aussagekraft nicht erhalten bleibt.

Hieraus ergeben sich die beiden folgenden Empfehlungen 2 und 3 der SSK:

- Bei einer Dosisreduktion in Kombination mit der Anwendung von Verfahren der KI zur Generierung der zu befundenden Bilder sollten alle befundrelevanten Strukturen mit vergleichbarer Sensitivität und Spezifität identifiziert werden können wie sie ohne Dosisreduktion und KI-Einsatz gegeben wären.
- Zur Verifizierung der Einhaltung dieser Empfehlung sollen human-observer Studien durchgeführt werden, es sei denn, es liegen bereits wissenschaftlich validierte Ergebnisse vor, die eine äquivalente Aussagekraft messtechnischer Verfahren bestätigen. Dabei ist darauf zu achten, dass die messtechnischen Verfahren so gewählt werden, dass nicht KI-Verfahren in der Evaluation bevorzugt werden, und dass tatsächlich das Fehlen oder die Erzeugung subtiler Strukturen hinreichend bei der Prüfung berücksichtigt wird (siehe Abschnitt 4.2.5).

4.1.3 Rechtliche Fragen aufgrund potenziell verursachter Fehlbehandlungen

Neben den eigentlichen strahlenschutzrechtlichen Fragen, die auf Grund fehlerhafter zusätzlicher oder unterdrückter Information, im Sinne der Anwendung ionisierender Strahlung am Menschen ohne den erwarteten Nutzen der sicheren Diagnosestellung (siehe Abschnitte 4.1.1 und 4.1.2) entstehen können, werden sicherlich auch andere rechtliche Fragen als relevant einzuschätzen sein. Dies trifft insbesondere darauf zu, dass die Möglichkeit besteht, dass auf Grund der fehlerhaften Information oder der fehlerhaften Darstellung der Information – was durchaus schwierig nachzuweisen sein kann – fehlerhafte Behandlungen oder Therapieentscheidungen

resultieren. Dies kann sehr weitreichend sein und zum Beispiel dazu führen, dass eine Tumorerkrankung oder eine andere schwerwiegende Erkrankung nicht erkannt wird oder als weniger schwerwiegend eingeschätzt wird und so eine notwendige Behandlung nicht stattfindet. Es kann aber auch das Gegenteil der Fall sein, dass eine Erkrankung aufgrund des KI-Anwendung für schwerwiegender gehalten wird, als sie ist, oder überhaupt nur vermeintlich vorliegt und so eine unnötige Behandlung die Folge ist. In beiden Fällen wird die Frage relevant, wer für die getroffene Therapieentscheidung oder Behandlungsstrategie verantwortlich ist. Diese Fragen sind allerdings nicht relevant im Sinne des Strahlenschutzrechts und werden daher hier nicht beleuchtet; es soll jedoch darauf hingewiesen werden, dass sie bei einer Einführung bzw. Verwendung der Verfahren mit betrachtet werden müssen.

Inwieweit bei KI-basierten Bildrekonstruktionsverfahren spezielle Medizinproduktrechtliche Aspekte zu berücksichtigen sind, ist derzeit schwer absehbar. Generell sollte ein Verfahren, welches dem Medizinproduktegesetz unterliegt, das leisten, was ausgewiesen ist – in diesem Fall Bilder zu erzeugen, die die Wirklichkeit korrekt wiedergeben. Bei „Black Box“-Verfahren, deren Arbeitsweise nicht überwacht und dokumentiert werden kann, kann dies grundsätzlich in Frage gestellt werden, sofern nicht anhand ausreichender empirischer experimenteller oder klinischer Anwendungsdaten belegt wird, dass das Medizinprodukt der ausgewiesenen Zweckbestimmung gerecht wird. Die Anwendung von Verfahren, die für den Anwender und sogar für den Hersteller eine Black-Box als Teil der Bilderzeugung darstellen, ist allerdings ein Novum. Dies wird vom Medizinprodukterecht auf seinem aktuellen Stand noch überhaupt nicht berücksichtigt.

4.1.4 Datenschutzrechtliche Aspekte

KI-Modelle, die aus Rohdaten mit hohem Rauschanteil diagnostisch verwertbare Bilder erzeugen sollen, müssen mit Datensätzen aus Untersuchungen bei Patienten bzw. Patientinnen trainiert werden, die sich den Untersuchungen aus klinischen Gründen unterzogen haben. Diese Daten werden zur Entwicklung der KI-Verfahren übermittelt, möglicherweise mit ergänzenden z. B. klinischen Daten. Empfänger sind überwiegend Dritte, also die Hersteller, deren Geräte in der Institution betrieben werden, und die beabsichtigen, entsprechende KI-Verfahren zu entwickeln.

Eine vollständige Anonymisierung der Datensätze ist praktisch nicht möglich, da mit Hilfe eines Pseudonyms zumindest festgehalten werden muss, welche Datensätze jeweils ein und derselben Person zuzuordnen sind. Ebenso müssen relevante klinische Informationen wie die Art der abgebildeten Pathologie sowie allgemeine Informationen z. B. zu Alter, Geschlecht und ggf. Ethnie erkenntlich sein. Grundsätzlich kann bei Daten aus bildgebenden Untersuchungen eine Rückverfolgung zur Person, z. B. anhand von Konturen des Gesichts oder Körperstamms, Merkmalen des Skeletts oder Veränderungen aufgrund früherer ärztlicher Maßnahmen, niemals vollständig ausgeschlossen werden, auch wenn Daten, die wie z. B. der Name eine direkte Identifizierung ermöglichen, entfernt wurden. Somit unterliegen die Speicherung und Weitergabe dieser Daten den Anforderungen des Datenschutzrechts. Auch muss die informierte Einwilligung der betroffenen Personen vorliegen, entweder unmittelbar in Bezug auf den Forschungszweck oder zumindest in Zusammenhang mit dem Behandlungs- und Untersuchungsvertrag mit der Institution.

4.2 Technische Rahmenbedingungen

In diesem Abschnitt werden die technischen Rahmenbedingungen für den Einsatz von KI zur Bildrekonstruktion und ihren möglichen Einfluss auf die Dosisreduktion erläutert. Dafür werden das Training von tiefen neuronalen Netzen (englisch deep neural networks – DNN), der

Backpropagation-Algorithmus und Varianten wie supervised Learning und unsupervised Learning (inkl. Autoencoder) erklärt. Schwierigkeiten ergeben sich bei der Beschaffung von ausreichend vielen Trainingsbeispielen, die datenschutzkonform gesammelt werden müssen, sowie bei der Bereitstellung ausreichender Rechenkapazitäten. Es fehlen auch effektive und objektive Messverfahren zur Beurteilung der Qualität von KI-Rekonstruktionsverfahren.

4.2.1 Training von tiefen neuronalen Netzen und Backpropagation-Algorithmus

KI-Verfahren haben das Potenzial, Bildrekonstruktion in medizinischen Anwendungen zu verbessern und gleichzeitig die Strahlenexposition für Patienten bzw. Patientinnen zu reduzieren. Die Grundlage dieser Technologie sind tiefe neuronale Netze (DNNs), die aus vielen Schichten von künstlichen Neuronen bestehen. Diese Netze werden mithilfe von Algorithmen wie Backpropagation trainiert, um komplexe Muster in den Daten zu erkennen und darauf basierend Vorhersagen oder Rekonstruktionen zu treffen.

Der Backpropagation-Algorithmus wird in künstlichen neuronalen Netzen verwendet, um die Gewichte der Neuronen während des Trainingsprozesses anzupassen. Dabei berechnet er den Gradienten der Verlustfunktion des Netzes durch wiederholte Anwendung der Kettenregel auf die einzelnen Schichten des Netzes. Dieser Gradient wird dann verwendet, um die Gewichte des Netzes durch Gradientenabstieg zu optimieren, um die Verlustfunktion zu minimieren, d. h. die Parameter des Modells werden iterativ so lange angepasst, bis es nahezu keine Veränderungen mehr gibt. Durch die Anpassung der Gewichte während des Trainingsprozesses kann das Netzwerk lernen, Muster in den Eingabedaten zu erkennen und präzise Vorhersagen zu treffen und sogar komplexe Probleme, wie die Bildrekonstruktion zu lernen.

Es gibt verschiedene Trainingsansätze für DNNs, wie beispielsweise das überwachte Lernen (supervised learning) und das unüberwachte Lernen (unsupervised learning). Beim überwachten Lernen werden den DNNs sowohl Eingabedaten als auch die korrekten Ausgabedaten (Labels) präsentiert, um das Netz anzupassen. Im Gegensatz dazu verwendet das unüberwachte Lernen keine Labels und versucht stattdessen, die Struktur oder Verteilung der Eingabedaten direkt zu modellieren. Autoencoder sind ein Beispiel für unüberwachtes Lernen, bei dem ein DNN versucht, die Eingabedaten auf effiziente Weise zu komprimieren und anschließend wieder zu rekonstruieren.

4.2.2 Datensammlung und Datenschutz in Europa

Ein zentrales Problem bei der Anwendung von KI in der Bildrekonstruktion ist die Notwendigkeit, viele tausende Trainingsbeispiele zu sammeln, um das neuronale Netz effektiv zu trainieren. In Europa stellt das Sammeln dieser Daten eine Herausforderung dar, da strenge Datenschutzgesetze die Verwendung von Patientendaten einschränken. Zudem ist es wichtig, eine ausreichende Variation von Pathologien, Patienten und Patientinnen verschiedener ethnischer Herkunft und Aufnahmeprotokollen über verschiedene klinische Standardvorgehensweisen hinweg abzudecken, um eine umfassende und zuverlässige KI-Lösung zu gewährleisten. Weltweit sind große industrielle Anbieter bereits dabei große Mengen solcher Daten anzuhäufen, wobei Qualität und Herkunft der Daten nicht immer zweifelsfrei geklärt werden kann. U. a. die Realisierung des europäischen Gesundheitsdatenraumes könnte genutzt werden um geeignete Daten in großer Breite zur Verfügung stellen zu können.

4.2.3 Rechenkapazität für das Training von tiefen Netzen

In der Entwicklungsphase von DNNs für die tomographische Bildgebung sind die Hardware-Anforderungen substanziell. Dies ist dem Umstand geschuldet, dass der iterative Optimierungs-

prozess inhärent rechenintensiv ist und auf einer großen Menge dreidimensionaler Trainingsdaten operiert. Hardwareseitig haben sich für diese Aufgabe Hochleistungsgraphikkarten (GPUs) durchgesetzt. Selbst unter Einsatz dieser dedizierten Hardware kann das Training des Netzwerkes in typischen Anwendungsfällen mehrere Tage Rechenzeit in Anspruch nehmen. Da überdies Design und Optimierung einer neuen Netzwerkarchitektur in aller Regel ein empirischer oder heuristischer trial-and-error-Prozess ist, welcher erfordert, nach jeder Modifikation das Training des Netzwerkes zu wiederholen und die Ergebnisse verschiedener Designvarianten sodann zu vergleichen, ist die Verfügbarkeit möglichst leistungsfähiger Hardware für die effiziente Entwicklung neuer DL-Ansätze von maßgeblichem Belang.

Ein spezielles Problem bei der Implementierung von KI-Lösungen insbesondere zur Bildrekonstruktion und Dosisreduktion ist die erforderliche Rechenkapazität für das Training von DNNs. Solche Rechenleistungen sind oft bei großen Unternehmen und globalen Akteuren verfügbar, während Universitäten, Kliniken und öffentliche Institutionen möglicherweise keinen ausreichenden Zugang zu solchen Ressourcen haben. Für kleinere Studien und stark spezialisierte Verfahren sind übliche Ausstattungen von forschenden Instituten typischerweise ausreichend. Für die Verarbeitung von großen Datenmengen, wie sie beispielsweise für Foundational Models notwendig sind, kann die Rechenausstattung unzureichend sein. Insbesondere setzen traditionelle Großrechenanlagen in Deutschland auf klassische zentrale Rechenkerne (CPUs), während eine Vielzahl von ML-Algorithmen „General Purpose Graphical Processing Units“ (GPUs) erfordern, was noch nicht in allen deutschen Höchstleistungsrechenzentren in ausreichendem Maße zur Verfügung steht. Dies stellt eine wesentliche Hürde für die Nachvollziehbarkeit und Verbreitung dieser Methoden dar. Weiterhin findet sich der führende Hersteller der GPU Rechenhardware (Nvidia) in Taiwan, was geopolitisch zu einem Risiko werden könnte.

Im Gegensatz hierzu sind die Hardware-Anforderungen für den Einsatz trainierter Netze weniger kritisch und übliche moderne Arbeitsplatzrechner in der bildgebenden Diagnostik in der Regel ausreichend.

4.2.4 Software

Softwareseitig hat sich im letzten Jahrzehnt eine geringe Anzahl umfänglicher, frei verfügbarer Bibliotheken herauskristallisiert, welche das Gebiet in weiten Teilen dominieren und als „Baukästen“ genutzt werden. Zu nennen sind hier insbesondere die folgenden Entwicklungsplattformen:

- Tensorflow (Google),
- PyTorch (Meta),
- MONAI (Nvidia),

welche derzeit in den weitaus meisten Anwendungen von DL-Methoden (auch in der medizinischen Bildgebung) genutzt werden. Es ist hierdurch vergleichsweise leicht geworden, Standardtechniken für spezifische Aufgabenstellungen anzuwenden. Gleichzeitig ist aber der Komplexitätsgrad der dabei ablaufenden Berechnungen derart, dass der Nutzer oder die Nutzerin selten in der Lage ist, ein tieferes Verständnis über das Verhalten von trainierten Netzwerken zu entwickeln. Dies ist insbesondere der Fall, wenn Voraussagen über das Netzwerkverhalten bei Anwendung auf Daten benötigt werden, die am Rand oder außerhalb der Mannigfaltigkeit der Trainingsdaten liegen.

4.2.5 Qualitätsmanagement und Messverfahren

Eine weitere Herausforderung im Bereich der KI-gestützten Bildrekonstruktion und in Teilen auch anderer, KI-basierter Bildverarbeitungsalgorithmen ist das Qualitätsmanagement. In der

Literatur finden sich häufig positive Ergebnisse, während negative Studienergebnisse seltener veröffentlicht werden. Es gibt derzeit keine anerkannte messtechnische Methode, die zweifelsfrei insbesondere die Qualität von KI-Rekonstruktionsverfahren messen kann. Traditionelle Qualitätssicherungsinstrumente wie z. B. Phantome sind anfällig für Manipulationen, wenn sie in die Trainingsdaten aufgenommen werden.

Um KI-Verfahren sicher für die Dosisreduktion einsetzen zu können und Halluzinationen (also nicht in der Realität vorhandene Bildinhalte, die das KI-basierte Verfahren erzeugt) zu vermeiden, sind effektive und objektive Messverfahren zur Beurteilung der Qualität unerlässlich. Je nach Ausmaß des KI-Einsatzes und der Anzahl der freien Parameter kommen dafür unterschiedliche Ansätze in Frage. Es ist wichtig, solche Methoden zu entwickeln und zu etablieren, um die Strahlenschutzgrundsätze der Rechtfertigung und der Optimierung einhalten zu können und das Vertrauen in die Anwendung von KI in der medizinischen Bildgebung zu fördern. Im Endeffekt ist es unumgänglich, dass eine Gleichwertigkeit der eingesetzten Verfahren im Hinblick auf die diagnostische Sicherheit mit bisherigen Verfahren oder Vorgehensweisen gezeigt werden kann. Aufgrund dessen sind so genannte Human-Observer Studien als Referenz anzusehen. In diesen können die Sensitivität und Spezifität eines Verfahrens für eine bestimmte diagnostische Fragestellung evaluiert werden. Diese Studien sind allerdings sehr aufwendig, und ihre Aussagekraft hängt von der Anzahl und Varianz der betrachteten Fälle, der Anzahl und Eignung der auslesenden Personen sowie der Komplexität und Schwierigkeit der betrachteten Fragestellung ab. Auf eine ausreichende statistische Untermauerung ist also für die angestrebten Vergleiche in jedem Fall zu achten. Dies muss dokumentiert werden. Wenn messtechnische Verfahren verwendet werden sollen, um die Vergleichbarkeit der KI-basierten Verfahren mit Standardverfahren darzustellen, muss sichergestellt sein, dass diese Verfahren gleichwertige Ergebnisse wie die Human-Observer Studien zeigen würden. Dies ist in Vergleichsstudien zu beweisen, die mit beiden Ansätzen – also Human-Observer Studie und messtechnischer Studie – durchgeführt werden und an repräsentativen Beispielen vergleichbare Ergebnisse zeigen. Auf jeden Fall sind Beispiele mit repräsentativen Unterschieden in diagnostischer Sicherheit bei Human-Observer-Studien und solche ohne derartige Unterschiede einzubeziehen, Entsprechende Verfahren müssen für KI-basierte Anwendungen zur Bildrekonstruktion und Rauschreduktion noch entwickelt und evaluiert werden.

Daher empfiehlt die SSK, dass

entsprechende Forschungsförderung als ein zentraler Schritt für die sichere Anwendung KI-basierter Verfahren für diese Bereiche der medizinischen Bildgebung, insbesondere bei Verwendung ionisierender Strahlung ermöglicht werden soll.

Ein wesentlicher Aspekt für die Beurteilung der Leistungsfähigkeit eines KI-Verfahrens ist, in der Anwendung des Verfahrens auf unabhängige, gut charakterisierte Testdaten zu sehen, welche nicht zuvor bei Entwicklung und initialen Evaluierung des Verfahrens genutzt wurden („unseen Data“). Eine solche Datenbasis sollte dementsprechend nicht frei zugänglich, sondern der Nutzung durch eine zuständige Stelle vorbehalten sein, um die bereits erwähnte Problematik einer möglichen manipulativen „Optimierung“ des Verfahrens durch Integration besagter Daten in ein Re-Training des betrachteten Verfahrens auszuschließen.

Die SSK empfiehlt dementsprechend in Empfehlung 5 in Kapitel 2, dass

- *eine zuständige Stelle benannt werden soll, welche die Leistungsfähigkeit des Algorithmus für die spezifizierten Anwendungsfälle auf unabhängigen, ausreichend charakterisierten Testdaten überprüft. Diese zuständige Stelle kann entweder eine vorhandene Stelle sein, in die zusätzliche Kompetenzen integriert werden, oder sie kann neu geschaffen werden.*

- *Entsprechende gesetzliche Regelungen oder untergesetzliche Bestimmungen sollten getroffen werden.*

Die einzelnen Aufgaben dieser Stelle ergeben sich aus den Ausführungen zu verschiedenen Kapiteln und sind in Empfehlung 5 zusammengefasst. In den folgenden Empfehlungen sind noch die möglichen Umsetzungen beschrieben, die sich ebenfalls aus den verschiedenen Ausführungen in den folgenden Kapiteln ergeben.

4.3 Anwenderspezifische Rahmenbedingungen

Von Anwenderseite bestehen Anforderungen an KI-basierte Systeme, die im Folgenden aufgeführt werden:

Geschwindigkeit: Die Abläufe in radiologischen und nuklearmedizinischen Einrichtungen sind eng getaktet und verzahnt, sodass nennenswerte Verzögerungen bis zur Bereitstellung der mit KI-Verfahren errechneten Bilder nicht hinnehmbar sein dürften. Gleichzeitig ist es grundsätzlich auch möglich, dass Verfahren und Darstellungen beschleunigt werden können, was sicher einen Vorteil im klinischen Alltag darstellen kann.

Ein *Zugewinn an Bildqualität* sollte bereits anhand des visuellen Eindrucks erkennbar sein, nicht *allein* anhand numerischer Kenngrößen wie z. B. dem Kontrast-Rauschverhältnis auf klinischen oder Phantomaufnahmen, oder einer Modulations-Transfer-Funktion. Ein Zugewinn an Bildqualität kann je nach Anwendungsfall bedeuten, dass mit KI-Verfahren errechnete Bilder trotz signifikant verringerter Exposition oder applizierter Aktivität den konventionell erzeugten an Qualität gleichwertig sind oder dass mit beibehaltener Dosis die Bilder erkennbar besser ausfallen. Die Bewertung der Bildqualität sollte sich nicht auf den bloßen subjektiven Bildeindruck beschränken, dessen Beurteilung fraglos von Vorlieben und Sehgewohnheiten beeinflusst wird: Bildrauschen, Schärfe oder die Prägnanz pathologischer Befunde und wichtiger anatomischer Bildmerkmale sind Beispiele für mögliche, objektivierbare Kriterien. Aus Sicht des Strahlenschutzes und des klinischen Anwenders ist insbesondere die diagnostische Aussagekraft das entscheidende Kriterium.

Die *anatomische Korrektheit* der Bilder ist unabdingbar. Dass vorhandene Strukturen nicht unterdrückt und nicht vorhandene Strukturen nicht erzeugt werden dürfen, ist trivial, aber die Anforderung geht weit darüber hinaus. Die scharfe oder unscharfe Begrenzung von Organen und Läsionen, die glatte oder unregelmäßige Beschaffenheit von Kanten, wie z. B. der Oberfläche von Organen sind wichtige diagnostische Merkmale, die keinesfalls verfälscht werden dürfen. Hierzu gehört auch die Unschärfe infolge des Partialvolumeneffekts von Grenzflächen, die schräg durch die Schicht verlaufen und eine wichtige Information über die anatomischen Verhältnisse darstellt. An bereits im klinischen Einsatz befindlichen Geräten konnte z. B. der Effekt beobachtet werden, dass die Leberoberfläche im Bild fälschlich kleinhöckrig verändert wiedergegeben wird, was ein Indiz für eine Leberzirrhose wäre. Ein weiterer, beobachteter, artifizieller Fehlbefund kann an der Grenze zwischen Leber und gefüllter Gallenblase auftreten, die in einem flachen Winkel durch die Schicht verläuft. Im Normalfall würde die Grenze infolge des Partialvolumeneffekts als Verlauf zwischen zwei verschiedenen Grauwerten korrekt unscharf dargestellt, errechnet wurde aber ein Übergang mit einem diskreten Grauwert, der eine Verdickung der Gallenblasenwand z. B. durch eine Entzündung vortäuschte. Die kritiklose Beseitigung diagnostisch wichtiger Artefakte oder Akzeptanz von Nebeneffekten bei der Bildglättung kann zu Fehldiagnosen führen.

Die *Verfügbarkeit konventionell berechneter Bilder* als Referenz ist zumindest so lange unabdingbar, wie die KI-basierte Bildrekonstruktion noch nicht in einer solchen Breite evaluiert wurde, dass auf eine anatomisch korrekte Darstellung vertraut werden kann. Bis dahin sollte

der regelmäßig begleitende Vergleich von mit KI- und konventionellen Verfahren rekonstruierten Bildern zur Implementierung gehören.

Aus demselben Grund ist die entsprechende *Kennzeichnung von KI-basierten Bildern* in den Overlays zu fordern, ebenso von ärztlichen Befunden, die mit diesen Bildern erstellt wurden.

Die *Integrität der Bilder auch in Hinblick auf Hard- und Softwareupdates* soll sicherstellen, dass nicht allein aufgrund der Rekonstruktion oder allgemein der Bildverarbeitung Befunde scheinbar hinzugekommen oder verschwunden sind, bzw. andere Bildmerkmale sich verändert haben. Gewiss erfahren die Bilder auch beim Umstieg auf eine neue Gerätegeneration Veränderungen, doch sind sie in Kenntnis der physikalischen und mathematischen Verfahren vorhersehbar und erklärbar. Bei Änderungen an der KI-basierten Rekonstruktion bzw. Bildverarbeitung ist die hierfür nötige Transparenz aber nicht unbedingt gegeben. In analoger Weise ist für ein Mindestmaß an *Vergleichbarkeit von Bildern verschiedener Hersteller* zu sorgen.

Es muss sichergestellt sein, dass der Arzt oder die Ärztin seine oder ihre Verantwortung bei der Anwendung ionisierender Strahlung wahrnimmt und die Diagnose aufgrund der verfügbaren Aufnahmen sichergestellt werden kann. Dies bedeutet, der Arzt oder die Ärztin muss sicher sein können, dass die Informationen, die dargestellt werden, echt sind, und dass keine Informationen verlorengegangen oder hinzugekommen sind.

In der praktischen Umsetzung ist jedoch zu beobachten, dass der Einsatz von KI-Verfahren das Risiko für einen Verlust an Fertigkeiten und Routine beim medizinischen Personal schafft. Hinzu kommt, dass Zeitdruck und Ressourcenknappheit in der klinischen Versorgung dazu führen können, dass Ergebnisse von KI-Verfahren tatsächlich nicht ausführlich vom medizinischen Personal geprüft werden. Es sind daher Maßnahmen zu treffen, um diese Entwicklung zu verhindern. Dabei sind jedoch Aspekte der praktischen Umsetzung zu bedenken, insbesondere der potentielle Verlust an Fertigkeiten und Routine, auch vor dem Hintergrund des klinischen Zeitdrucks und der Ressourcen-Knappheit.

Insbesondere die Daten, die zur Modellbildung oder zum Training der DL-basierten Verfahren verwendet werden, müssen bestimmte Qualitätsanforderungen erfüllen. Damit ist nicht gemeint, dass es ausschließlich ideale Bilder mit bester Bildqualität sein müssen, sondern dass die Rahmenbedingungen der Aufnahmen bekannt sind und bestimmten Vorgaben entsprechen und in diesem Sinne eine hohe Qualität aufweisen müssen. Dann lässt sich darstellen, wie genau die Verfahren arbeiten können, mit welchen Abweichungen und Unsicherheiten in den Bildrekonstruktionen oder Bilddarstellungen gerechnet werden muss und wie dies von Patientenparametern abhängt. Eine Kenntnis dieser Genauigkeit und der potenziellen Abweichungen ist für den Arzt oder die Ärztin ebenso wichtig wie die Kenntnis des Verfahrens als solches, um abschätzen zu können, wie sicher er bzw. sie mit ihrer Diagnose und/oder Therapieentscheidung sein kann. Es muss auch überprüfbar und nachvollziehbar sein, wie sich dies möglicherweise mit mehr verfügbaren Datensätzen verändert. Unterstützung bei der Einschätzung des Verfahrens und seiner Auswirkungen kann z. B. durch eine/n MPE erfolgen. Der bzw. die MPE in der Radiologie und der Nuklearmedizin ist für die Qualitätssicherung radiologischer bzw. nuklearmedizinischer Anlagen und die Auswahl der Geräte und Ausrüstungen zuständig, was die KI-Verfahren einschließt.

5 Potenzielle Probleme und Limitierungen

5.1 Garbage in, Garbage out

Im maschinellen Lernen ist die Qualität der Trainingsdaten von entscheidender Bedeutung für die Leistungsfähigkeit des trainierten Modells. Wenn die Daten von geringer Qualität sind, werden die Modelle nur suboptimale Ergebnisse liefern. Dieses Problem wird oft als „Garbage in, Garbage out“ bezeichnet, was bedeutet, dass, wenn schlechte Daten in ein Modell eingegeben werden, auch schlechte Ergebnisse herauskommen.

Die Qualitätssicherung der Trainingsdaten sollte daher einen hohen Stellenwert haben, um sicherzustellen, dass die Daten frei von Fehlern, konsistent und repräsentativ sind. Dazu gehört beispielsweise die Überprüfung auf fehlerhafte oder unvollständige Daten, die Entfernung von Ausreißern und das Ausbalancieren von Klassen, um sicherzustellen, dass das Modell nicht auf eine Klasse übertrainiert wird. Auch muss die Zahl der Trainingsbeispiele im Verhältnis zu der Zahl der aus ihnen gewonnenen Trainingsparameter passen.

Insgesamt ist es wichtig, sich bewusst zu sein, dass die Qualität der Trainingsdaten direkt die Leistungsfähigkeit eines Modells beeinflusst. Es ist daher von entscheidender Bedeutung, sorgfältig zu überprüfen und sicherzustellen, dass die Daten von hoher Qualität sind, um genaue und verlässliche Ergebnisse zu erzielen. Daten von hoher Qualität zu verwenden, bedeutet also nicht, dass ausschließlich ideale Bilder mit bester Bildqualität verwendet werden müssen, sondern dass die Rahmenbedingungen der Aufnahmen bekannt sind und bestimmten Vorgaben wie oben angegeben entsprechen.

Die Daten werden wie in der Einleitung (Kapitel 1.2) beschrieben in Trainings-, Validierungs- und Testdaten aufgeteilt. Wichtig ist, dass dies randomisiert geschieht. Wäre dies nicht der Fall, dann würden Trainings-, Validierungs- und Testdaten womöglich unterschiedliche statistische Verteilungen haben. Beispielsweise wäre es bedenklich, zum Training und Validieren nur Daten adipöser Patienten oder Patientinnen heranzuziehen und zum Testen dann nur die Daten von Patienten bzw. Patientinnen mit niedrigem BMI. Um beurteilen zu können, ob das Verfahren mit sinnvollen Daten trainiert wurde, ist eine umfassende Dokumentation der genutzten Daten wichtig. Diese Dokumentation beinhaltet eine Beschreibung der Patientenkohorte, der untersuchten Körperregion (z. B. Kopf, Hals, Thorax, Abdomen, Becken) sowie der Bildgebungsparameter. Zudem muss bekannt sein, ob ausschließlich oder zusätzlich Daten von Prüfkörpern in das Training mit aufgenommen wurden. Die Beschreibung sollte auch die Kriterien, nach denen die Daten ausgewählt wurden und eine Begründung dieser Auswahl beinhalten. Zudem sind statistische Parameter (Minimum, Median, Maximum etc.) zu den Patientendaten wie z. B. Alter, Größe, Gewicht zur Beurteilung der Datengüte hilfreich, ebenso die Angabe, ob die statistische Verteilung der Daten der Verteilung der Gesamtheit der Bevölkerung entspricht oder zumindest alle relevanten Gruppen hinreichend häufig vertreten sind. Mit Hilfe dieser Information können Rückschlüsse auf die Eignung eines KI-Verfahrens für einen konkreten Patienten oder eine konkrete Patientin gezogen werden. Würden beispielsweise im Training nur Patienten oder Patientinnen bis 80 kg vorkommen und der konkrete Patient oder die konkrete Patientin hätte 130 kg Gewicht, dann könnte die Anwendung des KI-Verfahrens möglicherweise suboptimale Ergebnisse liefern. Außerdem ist es wichtig zu wissen, ob Datenaugmentierungsverfahren angewendet wurden. Das sind Verfahren, die die initialen Daten verändern, um die Menge an Trainingsdaten zu vervielfachen.

Idealerweise werden die Daten offengelegt, so dass eine Beurteilung von unabhängiger Stelle aus erfolgen kann. Da die Offenlegung aller Trainingsdaten dem kommerziellen Interesse der Hersteller widerspricht, könnte alternativ eine Offenlegung nur gegenüber einer zuständigen

Prüf- oder Zertifizierungsstelle (zuständige Stelle) erfolgen, oder es könnte nur eine kleine Untermenge der Trainingsdaten offengelegt werden.

5.2 Probleme erkennen

Das prinzipielle Problem aller ML-basierten, insbesondere jedoch der DL-basierten Eingriffe in den Bildentstehungsprozess ist die Frage der Validierbarkeit generierter Ergebnisse im Einzelfall. Dies lässt sich zurückführen auf die grundlegende Frage der Fähigkeit des trainierten Netzwerks zur Generalisierbarkeit. Es ist zu untersuchen, ob im Rahmen des Trainings nicht verwendete Daten („unseen data“) korrekt verarbeitet werden können oder ob nicht erkannte spezifische Eigenheiten der Trainingsdaten dazu führen, dass das Netzwerk diese fälschlicherweise in der Anwendung auf neue Daten ergänzt, also Artefakte in den Bildern generiert. Empirisch wird sich diese Frage am ehesten durch regelmäßige Vergleiche zwischen konventionell generierten Bildern und solchen, welche einer DL-basierten Verarbeitung unterlagen, beantworten lassen.

Wesentliche mögliche Artefakte sind

1. Falsch positive Läsionen: Konversion eines Rauschartefakts durch das Netzwerk in eine vom Befundenden als solche wahrgenommene Läsion.
2. Falsch negative Befunde: Entfernung von als Rauschen interpretierte Niedrigkontraststrukturen, welche ansonsten vom Befundenden noch hätten identifiziert werden können.
3. Signalüberhöhung im Rahmen der Auflösungsrückgewinnung. Dieses Problem ist nicht spezifisch für DL-Ansätze, betrifft aber auch diese. Für die Nuklearmedizin ist dies besonders problematisch im Hinblick auf die hier häufig durchgeführte quantitative Bestimmung der Traceranreicherung etwa zur Bewertung des Therapieansprechens bei Tumorerkrankungen.
4. (Lokaler) systematischer Bias in der Bildintensität, der die absolute Quantifizierung beeinflusst.

Derzeit drängen alle Hersteller von CT-, SPECT- und PET-Geräten sowie auf Software spezialisierte Hersteller mit Produkten auf den Markt, welche als „KI-gestützt“ oder „KI-unterstützt“ beworben werden. Welche Aufgaben genau die KI-Module hierbei übernehmen, wird nicht näher spezifiziert. Besonders problematisch ist, dass die kommerzielle Entwicklung vermutlich dahingehen wird, dass der Nutzer die entsprechenden Eingriffe in den Bildentstehungsprozess nicht deaktivieren kann und nicht mehr ohne weiteres in der Lage sein wird, im Einzelfall den Umfang sowie die Güte und Validität der vorgenommenen Bildmodifikation selbst zu beurteilen. Auf absehbare Zeit wird das Problem der Verifizierbarkeit somit nur durch große klinische Validierungsstudien geklärt werden können.

5.3 Prinzipielle Problematik der Überprüfbarkeit von KI-Verfahren

Die Testung DL-basierter Verfahren weist eine fundamentale Problematik auf: Die etablierten konventionellen Methoden zur Beurteilung der Bildqualität und -integrität in der radiologischen und nuklearmedizinischen Bildgebung basieren auf Messungen mit dedizierten Phantomen einfacher Geometrie, die den jeweiligen Bildgebungs- und Bildverarbeitungsprozess durchlaufen. Aus den hieraus resultierenden Bilddatensätzen können sodann mittels geeigneter physikalischer Auswerteverfahren quantitative Parameter zur Beurteilung der Bildqualität abgeleitet werden. Überdies liefern derart standardisierte Messungen mit invarianten Phantomen die Grundlage für eine valide Observer-basierte quantitative Bewertung der Bildqualität.

Eine grundlegende Schwierigkeit ist nun, dass sich dieser Ansatz nur mit großen Einschränkungen und Vorbehalten auf die Überprüfung von KI-basierten Verfahren übertragen lässt, da physikalische Phantome a priori stets außerhalb des angestrebten Anwendungsbereichs des Verfahrens liegen, welches ja mit (und für die Anwendung auf) Patientendaten trainiert wurde. Insofern ließen die mit dem Phantom gewonnenen Resultate keine belastbaren Rückschlüsse auf die zu erwartende Qualität im Anwendungsfall zu (weder in positiver noch in negativer Hinsicht). Geeignete Phantomgeometrien, welche möglichst treu die anatomischen Verhältnisse reproduzieren und dennoch nicht vom Algorithmus 1 zu 1 wiedererkannt werden dürfen, sind denkbar, würden aber den Aufwand in jeder Hinsicht massiv erhöhen. Ideal wäre es zudem, wenn das Testverfahren einzeln den eigentlichen Bildaufnahmeprozess ebenso evaluieren kann wie die DL-basierte Bildverarbeitung.

Eine alternative Strategie, welche jedoch erst implementiert werden müsste, ist in der Nutzung von komplexen Softwarephantomen zu sehen, welche die Verhältnisse in realen Patientendaten hinreichend gut abbilden. Problematisch wäre aber in jedem Fall der Umstand, dass die Nutzung invarianter standardisierter Phantomgeometrien dem Entwickler des KI-Verfahrens prinzipiell die Möglichkeit bieten würde, sein Verfahren zielgerichtet auf gute Ergebnisse in entsprechenden phantomgestützten Qualitätskontrollmessungen zu trainieren, was diese natürlich entwerten oder wertlos machen würde. Variable, nur dem Tester bekannte Phantomgeometrien könnten helfen, dieses Problem zu entschärfen, würden aber den methodischen Aufwand noch weiter erhöhen. Solange es sich bei den komplexen Phantomgeometrien um reine Softwarephantome also ohne physische Repräsentation handelt, ist eine Testung des Gesamtsystems wie oben benannt, zudem nicht möglich.

Eine aktuell bereits zur Verfügung stehende Möglichkeit ist darin zu sehen, dass die Überprüfung von KI-Verfahren auf der Nutzung von realen Patientendaten beruht. Die offensichtliche konzeptionelle Schwierigkeit ist dabei, dass die inhärente interindividuelle Variabilität dieser Daten und die nicht gegebene Möglichkeit, Struktur und Kontrastverhältnisse wie bei Phantomen nach messtechnischen Erfordernissen frei zu wählen, zu neuen methodischen Problemen bei der quantitativen wie auch qualitativen Bewertung der resultierenden Bilder führt. Sammlungen gut charakterisierter Bilddatensätze von Untersuchungen an Patienten oder Patientinnen, welche der KI-gestützten Verarbeitung unterzogen werden können und einen gut definierten Soll-Ist-Vergleich erlauben, dürften daher eine wesentliche Bedeutung für eine objektive Bewertung erlangen. Diese Sammlungen dürfen dabei den Entwicklern von KI-Systemen nicht bekannt und nicht zugänglich sein. Zudem vermitteln DL-basierte Verfahren der Bildverarbeitung und -rekonstruktion insgesamt meist einen visuell sehr guten Bildeindruck, der nicht notwendigerweise zur diagnostischen Sensitivität und Spezifität korreliert.

Eine weitere Schwierigkeit ist, dass Standardtestverfahren im Sinne insbesondere statistischer Evaluierungen oder der Bestimmung von Abweichungen nur bedingt aussagekräftig sein können, da in diesen Verfahren häufig Abweichungen subtiler Strukturen nicht ausreichend bewertet werden, weil ein größeres Gewicht auf Abweichungen bei höheren Kontrasten oder großen Flächen gelegt wird. Diese können definitionsbedingt bei DL-basierten Verfahren häufig besser dargestellt werden als mit anderen, herkömmlichen Verfahren.

Um die objektive Bewertung von DL-basierten Verfahren zu ermöglichen und potenzielle Schwächen in KI-Algorithmen zu entdecken, ist es wünschenswert, unabhängigen Zulassungsbehörden Zugang zu einem repräsentativen Teil der Verifikationsdaten zu gewähren.

Diese Daten sollten idealerweise die gesamte Bandbreite der geplanten klinischen Anwendungen widerspiegeln. Dieser Ansatz gewährleistet eine höhere Transparenz und Glaubwürdigkeit, da die Leistungsfähigkeit des Systems unabhängig geprüft und validiert wird.

Es ist deutlich, dass die Evaluierung bildgebender Verfahren, welche KI-basierte Bildverarbeitungsverfahren verwenden, große Herausforderungen beinhaltet und daher ein erhöhter Forschungsbedarf besteht, um die Qualitätssicherung auf dem bekannten Niveau auch für neue Methoden etablieren zu können.

6 Qualitätsgesicherte Eingangsdaten

Wie in Abschnitt 5.1 beschrieben, ist es von zentraler Bedeutung, dass die Eingangsdaten für das Training und die Evaluation eines Netzwerks qualitätsgesichert sind. Dies gilt unabhängig von der konkreten betrachteten Fragestellung. Grundsätzlich sind bei der Dokumentation der Daten alle relevanten Aspekte bei ihrer Generierung inkl. ihrer Vorverarbeitung zu berücksichtigen. Die Qualität der Eingangsdaten, die für Training und Evaluierung eines neuronalen Netzwerks genutzt werden, ist entscheidend für seine objektive Leistungsfähigkeit, wobei deren korrekte Einschätzung wiederum von der Qualität der genutzten unabhängigen Testdaten abhängt. Als Nebenbemerkung sei hier angeführt, dass natürlich dann auch die Eingangsdaten bei der Verwendung der KI-basierten Verarbeitung denselben Qualitätsstandards genügen müssen. In diesem Kapitel werden die wesentlichen Gesichtspunkte bei der Qualitätsbeurteilung im Kontext der in dieser Empfehlung behandelten Verfahren der tomographischen Bildrekonstruktion, der Rauschunterdrückung und der Bewegungskorrektur dargestellt. Diese sind:

- Wie hoch ist die Zahl der genutzten unabhängigen Datensätze?

Eine große Anzahl separater Datensätze (siehe unten) ist eine notwendige (aber keine hinreichende) Voraussetzung dafür, dass die Trainingsdaten insgesamt als repräsentativ betrachtet werden könnten für die in der klinischen Anwendung später abzudeckende Spannbreite.

- Wie gut ist die Abdeckung der typischen interindividuellen (patho-)physiologischen Variabilität der regionalen Signalverteilung (globale Bilddynamik, lokaler Kontrast, Textur etc.)?

Dieser Aspekt spielt eine entscheidende Rolle, da das Netzwerk bei unzureichender Abdeckung auf die Reproduktion von Spezifika der zum Training benutzten (aber nicht repräsentativen) Daten trainiert wird, was in der Anwendung auf andere „unerwartete“ Daten zu suboptimaler Performance oder Artefakten führen kann. Z. B. ist darauf zu achten, dass betrachtet wird, inwieweit Aspekte der gegebenen Prävalenz in die Auswahl der Datensätze einzubeziehen sind.

- Wie gut ist die Abdeckung der relevanten interapparativen Variabilität und derjenigen Variabilität die in der Durchführung einer Untersuchung begründet liegt (verschiedene Tomographen mit verschiedenen Leistungsparametern wie intrinsischen Auflösungen und Sensitivitäten, verschiedene Akquisitions- und Rekonstruktionsparameter oder Radionuclidaktivitäten, verschiedene Rauschniveaus etc.) für den Fall, dass die Verfahren entsprechend flexibel verwendet werden sollen? Hiermit ist gemeint, dass mehr unterschiedliche Datensätze zum Training erforderlich sind, je System-unabhängiger und Anwendungsunabhängiger ein KI-Verfahren eingesetzt werden soll.
- Existiert eine qualifizierte *Ground-Truth*-Definition durch erfahrene Befunderinnen bzw. Befunder oder Bildgebungsexpertinnen oder -experten (sofern *Supervised-Learning* Ansätze verwendet werden)?

Hinsichtlich des zu fordernden notwendigen Minimums an separaten Trainingsdatensätzen können keine allgemein gültigen Vorgaben gemacht werden. Für einige der Verfahren und Anwendungen ist davon auszugehen, dass die Schwelle eher bei einigen 10-100 3D- Bildvolumina als bei einigen 1000 oder mehr liegt, bei anderen Verfahren zum Beispiel im Kontext von Screening-Verfahren mit Niedrigdosis-CT und entsprechend niedriger Prävalenz ist abhängig von der durch das Netzwerk zu lösenden Aufgabe eher mit einer höheren Zahl an benötigten Datensätzen zu rechnen.

Um die Verwendung qualitätsgesicherter Eingangsdaten sicherstellen zu können, ist zunächst eine klare Formulierung der Frage- bzw. Aufgabenstellung vorzunehmen. So sind die relevanten Kenngrößen für die Aufgabe hinsichtlich diagnostischer Verfahren, Aufnahmeparametern und aussagefähigen Qualitätskriterien zu identifizieren und zu spezifizieren. Im Falle von *Supervised-Learning* Ansätzen ist die Erstellung der Ground-Truth zu beschreiben. Die einzelnen Aspekte müssen klar dokumentiert werden. Die Dokumentation ist für alle Daten, die zur Modellgenerierung, Testung und Validierung verwendet werden, so durchzuführen, dass die Qualität der Daten jederzeit lückenlos darstellbar ist. Dazu gehören Hinweise zum Stand der KI-basierten Software, patientenbezogene Daten, Aufnahmeparameter, Geräteparameter, nachgewiesene Qualitätssicherung des Gerätes, Softwareversionen etc.

Standard Operating Procedures (SOPs) können dazu beitragen, die Qualitätssicherung der Eingangsdaten erheblich zu erleichtern. Aber auch bei Verwendung von SOPs ist auf eine entsprechende Dokumentation zu achten.

7 Mögliche Testverfahren zur Qualitätssicherung

7.1 Variable Phantome

Wie insbesondere im Abschnitt 5.3 dargestellt, ist eines der zentralen Probleme bei der Qualitätssicherung von KI-basierten Bildverarbeitungsverfahren, dass eine Testung, Prüfung und Qualitätssicherung als sehr schwierig anzusehen ist, da ein auf ML basierendes Verfahren in der Lage sein kann, auch die Testphantome anhand der zur Verfügung gestellten Roh- oder Bilddaten zu erkennen und so beliebig gute Bilder darzustellen. Dies ist insbesondere wichtig vor dem Hintergrund, dass die meisten der heutzutage verwendeten Testphantome für die Qualitätssicherung relativ einfache Strukturen enthalten und so leicht anhand weniger Daten zu erkennen sind.

Ein Ansatz, um eine Testung, Prüfung und eine anschließende Qualitätskontrolle auch von KI-basierten Bildverarbeitungsverfahren durchführen zu können, ist es daher, komplexe Strukturen enthaltende Basisphantome mit wechselnden, variablen Einsätzen zu erstellen, so dass sich die zu untersuchenden Phantome wenig, aber doch messbar und relevant unterscheiden und die Gesamtstruktur des Phantoms komplex ist. Wichtig ist dabei natürlich, dass die Inhalte und Anordnung der Phantome und ihrer Komponenten entweder von vornherein genau bekannt sind oder im Nachhinein durch eine CT- oder PET-Messung mit hoher Dosis und hoher Auflösung erfasst werden können, um so eine Grundwahrheit zu erhalten und eine echte Qualitätssicherung durchführen zu können.

Um die komplexe Grundstruktur zu entwerfen, bietet es sich an, Bilddatensätze interessierender Bereiche mit häufig kritischer Bildqualität zu untersuchen, als Dateien mit vorgegebener Information zu speichern und dann zum Beispiel davon 3D-Drucke anzufertigen. Sie sollten Bereiche enthalten, in die dann die variablen Einsätze mit z. B. Strukturen zur Auflösungsbestimmung, zur Kontrastevaluation und zur Rauschmessung eingebracht werden können. Die Phan-

tome müssen dabei so gestaltet werden, dass das System die Phantombilder als zur Aufgabengesamtheit zugehörig identifiziert und weder spezielle Algorithmen für die Verarbeitung der Phantomdaten entwickelt noch die Verarbeitung der Patientendaten verschlechtert wird. Letzteres könnte geschehen, indem das System lernt, Bilder der Phantome optimal zu rekonstruieren oder zu entauschen und dies nicht denselben Anforderungen wie für Patientenbilder entspricht. Für die nuklearmedizinische Bildgebung ergeben sich hier besondere Herausforderungen, weil typische Radionuklidverteilungen simuliert und entsprechende Auflösungs- und Kontrasttests mit radioaktivem Material bestückt eingebracht werden können müssen, die dennoch das Prinzip der variablen Phantome ermöglichen.

7.2 Vorgaben für Daten – Datentest

Integraler Bestandteil bei der Validierung trainierter Netzwerke ist eine sorgfältige Analyse der in unabhängigen (externen) Testdaten erzielten Ergebnisse. Solche Testdaten helfen, den Gültigkeitsbereich, in welchem ein gegebenes Netzwerk valide Resultate liefern kann (sowie den Anteil von manifestem Fehlverhalten) zu quantifizieren. Idealerweise sollten solche Testläufe regelmäßig mit jeweils unabhängigen Testkohorten wiederholt werden und Metriken verwenden, welche im engen Zusammenhang mit relevanten klinischen Aufgaben stehen. Wünschenswert erscheint es in diesem Zusammenhang ferner, dass von der benannten, unabhängigen zuständigen Stelle gut charakterisierte, anderweitig aber unzugängliche Testkohorten zur Verfügung stehen, um in die Lage versetzt zu werden, die herstellerseitig behauptete Leistungsfähigkeit des betrachteten Verfahrens objektiv überprüfen zu können. Hierbei kann es ratsam sein, auf eine angemessene Einbindung von erfahrenen Bildgebungs- und klinischen Expertinnen und Experten zu orientieren, sofern eine Beurteilung der Leistungsfähigkeit durch mathematisch-technische Metriken allein nicht gewährleistet werden kann. Weiterhin wäre es wünschenswert, wenn Befunde und Annotationen für die Charakterisierung der Test- und Validierungsdatensätze nicht nur von einer einzelnen Person erstellt wurden, sondern diese einen Konsens von mehreren Expertinnen und Experten abbilden.

Wünschenswert ist ferner die Möglichkeit, Eingriffe des Netzwerks in den Bildentstehungsprozess optional zu deaktivieren, um fallbezogen oder im Rahmen regelmäßiger Qualitätsüberprüfungen einen „vorher/nachher“ Vergleich zu ermöglichen, da hiermit dem erfahrenen Nutzer eine Einschätzung des Umfangs des Einflusses des Netzwerks auf die resultierenden Bilddaten ermöglicht wird, wie auch die erfahrungsbasierte Entscheidung, in welchen Fällen Resultate als nicht vertrauenswürdig oder valide angesehen werden müssen. Allerdings ist davon auszugehen, dass im zunehmenden Umfang eine tiefe Integration von DL-basierten Methoden in den Bildgenerierungsprozess seitens der Hersteller eine solche Deaktivierung oder Umgehung der DL-basierten Verarbeitungsschritte erschweren oder unmöglich machen wird.

7.3 Statistische Analysen

Zur Qualitätssicherung sind statistische Analysen essenziell, um die Resultate der KI-Methoden objektiv zu bewerten, zu optimieren und Hypothesen über die Nutzbarkeit und Robustheit aufzustellen. Im Vergleich zwischen den mittels KI-Methoden berechneten Bildern und Referenzbildern können zum einen quantitative, bildbasierte Metriken herangezogen werden, die aber, wie in Abschnitt 5.3 erläutert, nur bedingt aussagekräftig sind. Durch die Beurteilung durch klinische Experten und Expertinnen können qualitative Aussagen zur Güte des Bildeindrucks getroffen werden, die üblicherweise in Skalen unterteilt werden. Darüber hinaus können weitere relevante Eigenschaften wie zum Beispiel die Strahlenexposition oder die Durchführungszeit quantifiziert werden. Vor allem bei generativen KI-Methoden besteht die potenzielle Gefahr von Halluzinationen, die Informationen hinzufügen oder entfernen können, wie im Abschnitt

4.2 beschrieben. Dabei sollte unterschieden werden, ob je nach Eingangsdaten im Fehlerfall signifikante anatomisch inkorrekte Veränderungen im Bild auftreten oder schwieriger identifizierbare anatomisch korrekte, aber dennoch fehlerhafte Veränderungen wie z. B. interstitielle Lungenveränderungen, die potenziell die Befundung beeinflussen könnten. Eine Identifizierung und Quantifizierung dieser Bildfehler durch Experten und Expertinnen ist sinnvoll, um eine Risikoanalyse durchzuführen.

Diese Betrachtungen führen zur Empfehlung 4 der SSK in Kapitel 2:

Die SSK empfiehlt, dass

- der gesamte Prozess der Erstellung und Verwendung von KI-basierten Verfahren für die Bildrekonstruktion oder Bildverarbeitung qualitätsgesichert erfolgen soll. Insbesondere sollen generative Netze aus einem Eingangsdatensatz immer einen eindeutigen, nicht durch Zufall beeinflussten Ausgangsdatensatz erzeugen.

In der Entwicklung von KI-Methoden können die beschriebenen Metriken auf den gesamten Verifikationsdaten ermittelt werden und als Grundlage für statistische Analysen dienen. Durch die Berechnung von Parametern wie mittleren Fehlern, Quantilen und Standardabweichungen können Wahrscheinlichkeiten berechnet werden, unter welchen die KI-Methoden ein erwünschtes Ergebnis liefern. Zudem können Signifikanztests, wie zum Beispiel p-Tests, t-Tests oder Wilcoxon-Tests (Siegel 2001), auch oder Bayes-Verfahren eingesetzt werden, um die Überlegenheit eines Verfahrens gegenüber einem anderen KI- oder einem klassischen Verfahren zu bestimmen. Dabei hängt die Wahl der Methode von der Verteilung der Daten ab. Ferner kann eine multivariate Analyse darüber Aufschluss geben, ob die Ergebnisse von gewissen Einflussfaktoren oder einer Unterrepräsentation von Gruppen innerhalb der Daten abhängig sind. Dies ermöglicht eine differenzierte Bewertung, um Verzerrungen zu minimieren und zu identifizieren.

Im Patientenbetrieb ist die Durchführung solcher Analysen mittels repräsentativer Stichprobensammlungen oder im Rahmen klinischer Studien denkbar. Die Durchführung setzt dabei voraus, dass je nach Anwendung eine simultane Akquisition von Referenzdaten mit einem vertretbaren Aufwand möglich ist.

Ein weiterer zentraler Aspekt der Qualitätssicherung ist die Durchführung statistischer Analysen während des routinemäßigen Patientenbetriebs. Ziel derartiger Analysen ist es, kontinuierlich die Leistung der eingesetzten KI-Modelle zu überwachen und sicherzustellen, dass diese auch unter realen klinischen Bedingungen konsistente und valide Ergebnisse liefern. Hierbei können verschiedene Metriken genutzt werden, wie beispielsweise die Sensitivität, Spezifität, Genauigkeit oder der F1-Score von Diagnosen (Lang 2023), die auf Basis validierter Testdaten im Patientenbetrieb berechnet werden. Auch kommen klinische Parameter wie Quality of Life oder Behandlungserfolg in Betracht. Diese Metriken sollten im Hinblick auf die klinische Anwendung definiert, regelmäßig überprüft und mit den im Entwicklungsprozess des Modells erzielten Ergebnissen verglichen werden, um etwaige Abweichungen frühzeitig zu erkennen.

Ein zentraler Bestandteil dieser Analysen ist die Überwachung der Modellperformance über verschiedene Zeiträume hinweg. Änderungen in den zugrundeliegenden Daten, wie etwa durch den Einsatz neuer Geräte, geänderte Protokolle oder unterschiedliche Patientenkollektive, können die Leistungsfähigkeit des Modells beeinflussen (so genannter „Data Drift“). Dies kann beispielsweise durch Veränderungen im Patientenkollektiv im Hinblick auf Alter, Geschlecht, Gewicht, oder anderen Faktoren geschehen. Daher ist es essenziell, die Verteilung der Eingangsdaten und die Verteilung der Modellergebnisse kontinuierlich zu analysieren.

Zudem bietet sich der Einsatz von Methoden zur Detektion von „Out-of-Distribution“-Daten an, um sicherzustellen, dass das Modell nicht auf Daten angewendet wird, die außerhalb seines ursprünglich vorgesehenen Einsatzbereichs liegen. Im Idealfall können solche Fälle automatisch erkannt und entweder an den Anwender gemeldet oder zur weiteren Analyse und Nachschulung des Modells gespeichert werden.

Um die Ergebnisse aus den statistischen Analysen in die Praxis zurückzuführen, sollte ein Feedback-Mechanismus implementiert werden. Dieser könnte es dem klinischen Personal ermöglichen, auffällige oder potenziell fehlerhafte Ergebnisse an die Entwicklerteams zu melden. Ergänzend wäre eine zentrale Datenbank hilfreich, in der statistische Qualitätsparameter sowie annotierte und validierte Fehlerfälle gesammelt werden, um diese für die kontinuierliche Verbesserung der KI-Modelle zu nutzen. Weiterhin wäre es empfehlenswert, immer wieder Diagnostik mit klassischen Ansätzen (z. B. höherer Dosis), durchzuführen, um kontinuierlich Evidenz für die verbesserte Behandlung anhand von KI-Algorithmen zu sammeln.

Zusammenfassend stellen statistische Analysen ein unverzichtbares Instrument zur Qualitätssicherung dar, da sie es ermöglichen, die Leistung und Verlässlichkeit von KI-basierten Bildverarbeitungsverfahren im praktischen Einsatz zu überwachen, Abweichungen frühzeitig zu identifizieren und eine kontinuierliche Verbesserung zu gewährleisten.

7.4 Virtuelle Phantome und Qualitätssicherungsmethoden

Statt der Nutzung physikalischer Phantome, die aufwändig hergestellt werden müssen und deren Variabilität zumindest durch die Anzahl möglicher Qualitätssicherungsmessungen eingeschränkt ist, kann es von großem Vorteil sein, virtuelle Daten zu erzeugen und von den KI-basierten Rekonstruktions- oder Nachverarbeitungsverfahren bearbeiten zu lassen. Dies hilft allerdings immer nur zur Qualitätssicherung der Bildverarbeitungsverfahren, nicht zur Qualitätssicherung der bilderzeugenden Systeme oder des Zusammenspiels aus beiden Komponenten. Vorzugsweise basieren solche virtuellen Daten auf Patientendaten, denn nur so lassen sich Stichproben generieren, die der tatsächlich zu erwartenden statistischen Verteilung der Daten entsprechen. Eine ausreichend große Anzahl von klinischen Daten ist wichtig. In der Praxis lässt sich die Anzahl der Patientendaten durch die so genannte Datenaugmentierung, also einer Veränderung des Datensatzes, vervielfachen. Beispielsweise kann ein vorhandener CT-Datensatz durch Skalierung, Verschiebung, Deformierung und durch Graustufenänderung manipuliert werden und so als neuer Datensatz dem Training zugeführt werden. Weitere Beispiele für die Erzeugung zusätzlicher Datensätze sind:

- Vorhandene Bilddaten können virtuell mit zusätzlichem Rauschen – „Noise Injection“ beaufschlagt werden, um damit die KI-basierten Entrauschungsmethoden zu testen.
- Vorhandene Bilddaten können künstlich unschärfer gemacht werden, um KI-basierte Superresolutionsalgorithmen zu testen.
- Es können Simulationen des Streustrahlenanteils vorgenommen werden und dieser erhöht werden, um Kontrastveränderungen zu simulieren, ggf. kann das Volumen des Patienten oder der Patientin virtuell zu diesem Zweck angepasst werden.

Allerdings benötigen derartige Qualitätssicherungsmethoden geeignete Schnittstellen zum bildgebenden System bzw. dessen Software, die vom Hersteller zur Verfügung gestellt werden müssten. Zudem liefern diese Daten im Allgemeinen keine Möglichkeit der Überprüfung des Gesamtsystems, sondern nur der KI-basierten Bildverarbeitungsverfahren. Es ist zudem erforderlich, dass die Datenaugmentierung selber nach qualitätsgesicherten Prinzipien durchgeführt wird und sichergestellt bleibt, dass weder unsinnige Daten erzeugt werden noch die Daten eine im Sinne des Lernens zu große Ähnlichkeit haben.

7.5 Offenlegung der Trainings-, Validierungs- und Testdatenverteilung

Die Offenlegung wesentlicher Eigenschaften der verwendeten Trainings-, Validierungs- und Testdaten ist ein zentraler Bestandteil der Qualitätssicherung. Nur durch eine solche Offenlegung wird es zulassenden Behörden oder unabhängigen Stellen ermöglicht, die Eignung der zugrundeliegenden Daten für die angestrebten Anwendungsbereiche des KI-Algorithmus zu bewerten.

Eine sorgfältige Dokumentation der Daten trägt dazu bei, mögliche Verzerrungen (Bias) in den Daten zu identifizieren. Es sollte sichergestellt werden, dass die Trainings-, Validierungs- und Testdaten die gesamte Bandbreite der klinischen Realität widerspiegeln. Außerdem muss die Dokumentation der Daten so gestaltet sein, dass Datenschutzbestimmungen eingehalten werden.

Um eine fundierte Bewertung der Datenqualität und ihrer Eignung für den jeweiligen Anwendungsbereich zu ermöglichen, sollten folgende Aspekte der verwendeten Trainings-, Validierungs- und Testdaten dokumentiert werden:

- An wie vielen Patienten bzw. Patientinnen wurde trainiert?
- Welches Alter, Geschlecht, Gewicht und andere Merkmale hatten die den Daten zugehörigen Patientinnen und Patienten?
- Welches Aufnahme- und Rekonstruktionsprotokoll wurde genutzt? Welche Augmentierungstechniken wurden eingesetzt?
- Wurde auch an Qualitätssicherungsphantomen trainiert? Falls ja, warum?
- Hat das Training an Phantomen Einfluss auf die Ergebnisse an klinischen Daten und umgekehrt?
- Wodurch ist sichergestellt, dass die Trainingsverteilung der zu erwartenden Patientenverteilung entspricht?
- Wie wurde trainiert (supervised, unsupervised, ...)?
- Wie wurden die nötigen Datenpaare erzeugt (Messung, Simulation, ...)?
- Welche Maßnahmen wurden ergriffen, um Out-of-Distribution (OOD)-Fälle zu erkennen und abzufangen bzw. wie in das Modell zu integrieren, falls möglich und erforderlich?
- Gibt es fortlaufendes Training des verwendeten Modells anhand neuer klinischer Bilddaten? Wie werden in diesem Fall Veränderungen dokumentiert?

Basierend auf dieser Information könnten dann zur Qualitätssicherung spezielle echte oder virtuelle Phantome generiert werden, um darauf KI-Verfahren anzuwenden und eine Qualitätskontrolle zu erzeugen. Ferner ist eine solche Dokumentation erforderlich, um einer zuständigen Stelle eine Beurteilung der Validität der von den Entwicklern verwendeten Metriken und Methoden im Hinblick auf die angemessene Leistungsfähigkeit des KI-Algorithmus für spezifische klinische Anwendungen zu erlauben.

Out-of-Distribution-Fälle (OOD-Fälle) beziehen sich auf Eingabedaten, die außerhalb der Verteilung liegen, auf der der KI-Algorithmus trainiert wurde. Dies können beispielsweise außerordentlich adipöse Patienten bzw. Patientinnen oder Patienten und Patientinnen mit seltenen Anomalien sein, die in den Trainingsdaten unterrepräsentiert oder gar nicht enthalten waren. Das Erkennen solcher OOD-Fälle ist wichtig, da KI-Modelle auf Daten aus ihrer Trainingsverteilung optimiert sind. Bei Daten, die nicht Teil dieser Verteilung sind, können unvorhersehbare

Ergebnisse nicht ausgeschlossen werden, welches in der klinischen Praxis ein Risiko darstellen kann. Eine OOD-Erkennung (z. B. durch eine Analyse von Aufnahmeparametern) ist wünschenswert und für manche Modelle sogar erforderlich, um unvorhersehbare Ergebnisse zu minimieren und die Sicherheit sowie Zuverlässigkeit der klinischen Anwendung zu gewährleisten.

8 Eingang in Regelwerke

Die Empfehlungen der SSK zu der Verwendung von KI-Verfahren zur Bildrekonstruktion und -weiterverarbeitung medizinischer Bilder sollten in die relevanten Regelwerke Eingang finden, um deren stringente Umsetzung zu gewährleisten. Insbesondere sind hier das Medizinprodukte-durchführungsgesetz (MPDG 2020), die Medizinprodukte-Betreiberverordnung (MPBetreibV 2025) und die Strahlenschutzverordnung (StrlSchV 2018) und untergesetzliche Regelungen zu adressieren. Natürlich sind auch internationale Regelungen zu beachten.

Übergeordnet sicherzustellen ist, dass

- in medizinischen Bildern kein befundrelevanter Informationsgehalt verloren geht oder fehlinterpretierbare Bildinhalte neu generiert werden,
- alle zur Sicherstellung dieser Vorgaben notwendigen Informationen vorliegen,
- die Verwendung der KI-Verfahren klar beschrieben wird.

Aber gleichzeitig dürfen auch keine unnötigen Hürden für die Verwendung von KI-Verfahren für optimierte Bildgebungsverfahren bei gleichzeitiger Sicherung der Qualität und insbesondere des Strahlenschutzes entstehen. Daher ist bei der Umsetzung der Empfehlungen in das Regelwerk zu klären:

- Wie kann bzw. muss die Sicherstellung der relevanten Aspekte Eingang in Regelwerke finden?
- In welche Regelwerke ist dies am besten und zielführendsten möglich?

Eine Möglichkeit, Regelungen für Verfahren zu etablieren ist es, die Verfahren zu normieren. Für klassische Verfahren zur Bildrekonstruktion (Filtered Backprojection oder iterativer Verfahren) und zur Bildverarbeitung wie z. B. der Rauschreduktion (verschiedene Filter oder auch nichtlineare Verfahren) ist festzustellen, dass diese Verfahren bisher nicht Gegenstand von Normung waren oder sind.

Eine Normung der KI-basierten Verfahren zur Rekonstruktion oder Bildverarbeitung erscheint der SSK noch deutlich schwieriger, da unterschiedliche Funktionsweisen der einzelnen Algorithmen, die zu einem Teil auch nicht nachvollziehbar oder mathematisch beschreibbar sind, vorliegen, wie in den vorherigen Kapiteln deutlich wurde. Zudem erscheint die Normung einzelner Verfahren kaum möglich. Aufgrund der potenziell fehlerhaften Anwendung in nicht evaluierten Anwendungsbereichen würde die Normung einzelner Verfahren aus Sicht der SSK auch keine zusätzliche Sicherheit versprechen.

Alternativ zur Normung der KI-basierten Verfahren selbst ist für die Anwendung solcher Verfahren die Entwicklung geeigneter Standardarbeitsanweisungen (SOPs) relevanter im Sinne der Patientensicherheit und des Strahlenschutzes. Dies erleichtert die Umsetzung der Empfehlungen in Kapitel 2 hinsichtlich der Entwicklung, der Überprüfung und Verwendung der KI-basierten Verfahren. Eine Normung oder Standardisierung von Anforderungen an Daten für Training und Validierung oder Verfahren der Entwicklung, Testung und Qualitätssicherung von Verfahren erscheint hingegen aus Sicht der SSK möglich und ggf. sinnvoll.

9 Hinweise für die Entwicklung KI-unterstützter Rekonstruktions- und Weiterverarbeitungsalgorithmen in Radiologie und Nuklearmedizin

Um herstellerseitig KI-basierte Verfahren zu entwickeln und auf den Markt bringen zu können, die mit den oben gemachten Empfehlungen kongruent sind, gibt die SSK im Folgenden Hinweise, wie dies erreicht werden kann:

1. Die Aufgabe des KI-basierten Rekonstruktions- oder Weiterverarbeitungsalgorithmus ist klar zu spezifizieren
 - Positivbeispiel: „Ziel des entwickelten KI-Rekonstruktionsalgorithmus ist es:
 - diagnostische Bilder aus Aufnahmen mit im Vergleich zum bisherigen Standard reduzierter Strahlendosis (z. B. 25 % Reduktion) zu rekonstruieren, die der Standardrekonstruktion hinsichtlich der diagnostischen Qualität nicht unterlegen sind.
 - Bildartefakte in einem Organ/einer bestimmten Region/im gesamten Bild aufgrund der Atembewegung zu entfernen/reduzieren.
 - die räumliche Auflösung der rekonstruierten Bilder zu verbessern.“
 - Negativbeispiel: „Ziel des entwickelten KI-Rekonstruktionsalgorithmus ist es:
 - die Bildqualität zu verbessern (Problem: keine Spezifikation was bessere Bildqualität bedeutet)
 - die Bilder zu entrauschen (Problem: keine Angabe von Strahlendosis/Rauschniveaus in den Eingangs- und Referenzdaten).“
2. Eine strikte Trennung zwischen Trainings-, Validierungs- und Testdaten muss eingehalten werden. Letztere dürfen keinen direkten (Optimierung der Netzwerkgewichte) oder indirekten (z. B. Optimierung von Trainings- oder Netzwerkhyperparametern) Einfluss auf die Entwicklung des KI-Algorithmus haben, welcher der zuständigen Stelle zur Zulassung vorgelegt wird.
 - Positivbeispiel: „Aus den vorliegenden Daten wurden randomisiert 80 % der Daten für das Training (Optimierung der Netzwerkparameter), 10 % zur Validierung (Optimierung der Netzwerk- und Training Hyperparameter) und 10 % einmalig zum finalen Testen benutzt.“
 - Negativbeispiel: „Aus den vorliegenden Daten wurden randomisiert 80 % der Daten für das Training und 20 % zum Testen verwendet.“ Problem: Es bleibt unklar, ob die Testdaten einen Einfluss auf die Wahl der Hyperparameter hatten.
3. Eine ausführliche Charakterisierung der Trainings-, Validierungs- und Testdaten sowie eventuell genutzter Datenaugmentierungstechniken ist zu erarbeiten.
 - Positivbeispiel: „Tabellarische Auflistung von relevanten demographischen Daten (Alter, Geschlecht, Größe, Gewicht, ...), der Prävalenz von Pathologien und Untersuchungsindikationen, der Prävalenz von relevanten Artefakten, der verwendeten Radiotracer und Aufnahmezeit (sofern relevant: Nuklearmedizin). Zur Datenaugmentierung wurden folgende Techniken benutzt: ...“
4. Eine ausführliche Dokumentation von Geräte-, Protokollparametern, die zur Aufnahme der Rohdaten verwendet wurden, muss erstellt werden:

- Positivbeispiel: „Tabellarische Angabe von Scannermodell, Aufnahmeprotokoll, Rekonstruktionsparameter und -algorithmus.“
5. Die “Klasse” des verwendeten KI-Ansatzes muss spezifiziert werden.
 - Positivbeispiel: „Bei dem entwickelten Netzwerk handelt es sich um ein: „end-to-end“ Netzwerk / „unrolled variational Network“ / „konditioniertes Diffusionsmodell“ / „Entrauschungsnetzwerk“, welches im Bildraum nach der Rekonstruktion wirkt.“
 - Negativbeispiel: „Bei dem entwickelten Netzwerk handelt es sich um ein neuronales Netzwerk.“ Problem: Eine solche Angabe ist zu unspezifisch.
 6. Die verwendete übergeordnete Netzwerkkategorie und die Größenordnung der optimierten Netzwerkgewichte sollten spezifiziert werden.
 - Positivbeispiel: „Bei dem entwickelten Netzwerk handelt es sich um ein: convolutional neural CNNetwork mit U-net Architektur/Transformer Netzwerk - fully connected Network/FCN mit ca. 102, 103, 106 oder 109 trainierbaren Parametern.“
 - Negativbeispiel: „Bei dem entwickelten KI-Algorithmus handelt es sich um ein neuronales Netzwerk.“ Problem: Auch hier ist die Angabe zu unspezifisch.
 7. Die Kostenfunktion und der Optimierungsalgorithmus, welche im Trainingsprozess verwendet wurden, um die Netzwerkgewichte zu optimieren, sollten spezifiziert werden.
 - Positivbeispiel: „Für das Netzwerktraining wurde ein „supervised oder unsupervised“ Ansatz und der mittlere quadratische Fehler als Kostenfunktion genutzt. Für die Optimierung der Netzwerkparameter wurde der ADAM-Optimierer mit einer anfänglichen Schrittweite von 0,001 gewählt.“
 8. Die Metriken und Methoden, die benutzt wurden, um die Leistungsfähigkeit des entwickelten Algorithmus für spezifische Anwendungen anhand der gewählten Testdaten zu demonstrieren, müssen spezifiziert werden.
 - Positivbeispiel: „Die Leistungsfähigkeit des Netzwerkes wurde anhand der Sensitivität und Spezifität der Detektion von folgenden Läsionen/Strukturen/Pathologien evaluiert: ...“
 - Negativbeispiel: „Das Netzwerk wurde mit einer Referenzmethode verglichen.“ Problem: Es ist unspezifisch, wie der Vergleich durchgeführt wurde.
 9. Ein kleiner, aber repräsentativer Teil der Testdaten, welcher eine externe und unabhängige Verifikation der Leistungsfähigkeit des KI-Algorithmus ermöglicht, sollte offengelegt werden, zumindest gegenüber einer zulassenden oder beurteilenden Stelle.
 - Begründung: Eine transparente und nachvollziehbare Prüfung fördert das Vertrauen von Anwendenden und Regulierungsbehörden in die Sicherheit und Effektivität des KI-Algorithmus.

10 Schlussbetrachtungen

KI-basierte Verfahren gewinnen zunehmend in vielen Bereichen an Bedeutung. Dementsprechend gibt es auch sehr viele Ansätze, KI-basierte Verfahren für die Bildrekonstruktion und Rauschreduktion medizinischer Bilddaten zu entwickeln (siehe Kapitel 3). Prinzipiell können derartige Ansätze für eine bessere Verwendung der aufgenommenen bzw. gemessenen Bild(roh)daten geeignet sein. Wenn eine effizientere Ausnutzung der Daten gelingt, können

damit bessere Bilder im diagnostischen Sinn bei gleicher Strahlenexposition des Patienten oder der Patientin oder bei Bildern gleicher diagnostischer Eignung eine Reduktion der Strahlenexposition erreicht werden. Trotz den vielfältigen Ansätzen sind Produkte bisher nur sehr begrenzt in der klinischen Routine im Einsatz.

Aufgrund der in den meisten Fällen nicht nachvollziehbaren algorithmischen Verarbeitung der Bilddaten bei KI-basierten Verfahren und der häufig nicht mehr zu validierenden Originalinformation sowie der potenziellen Möglichkeit der Unterdrückung oder artifiziellen Einfügung von Bildinhalten muss die diagnostische Gleichwertigkeit zu Standardverfahren in jedem Fall sichergestellt werden. Dies ist eine zentrale Maßgabe auch aus Sicht der Strahlenschutzgesetzgebung (siehe Kapitel 4 und 8).

Aus diesen Aspekten ergeben sich die ersten drei grundlegenden Empfehlungen in Kapitel 2.

Um Verfahren zu entwickeln und im klinischen Betrieb sicher – auch im Sinne des Strahlenschutzes und der Patientensicherheit – einsetzen zu können, sind verschiedene Voraussetzungen zu erfüllen. Neben der Dokumentation, welche Modelle eingesetzt werden und wie sie trainiert werden, ist es insbesondere wichtig, die verwendeten Daten für Trainings-, Test-, und Validierungszwecke geeignet zu wählen (siehe Kapitel 6 und 7). Für die Anforderungen an die verschiedenen Datensätze ließen sich evtl. Standards mit Hilfe z. B. internationaler Normung definieren. Insgesamt ist eine internationale Abstimmung bzgl. der Rahmenbedingungen für den Einsatz KI-basierter Verfahren der Bildrekonstruktion und Rauschreduktion in medizinischen Bildern sicher sinnvoll (siehe Kapitel 8).

Außerdem müssen geeignete Verfahren verwendet und in großem Maße überhaupt erst entwickelt werden, um die KI-basierten Verfahren für die medizinische Bildgebung und auch für das gesamte Bildgebungsverfahren einschließlich der Datenaufnahme zu evaluieren und im Sinne der Qualitätssicherung zu überprüfen. Insbesondere in diesem Bereich ist aus strahlenschutztechnischer Sicht ein großer Forschungsbedarf zu konstatieren (siehe Kapitel 4, 5 und 7).

Für die Qualitätssicherung im Sinne des Strahlenschutzes spielen neben der Überprüfung der Systeme auch die korrekte Anwendung der Verfahren eine zentrale Rolle. Hierzu sind das Verständnis der medizinisch tätigen Nutzer und Nutzerinnen zu den Verfahren sowie Standardarbeitsanweisungen von zentraler Bedeutung (siehe Kapitel 4 und 8). Eine Normung einzelner Verfahren als solche erscheint hingegen der SSK weniger sinnvoll.

Entwicklungen in anderen Bereichen der KI-unterstützten Bildverarbeitung in der Medizin sollten aufmerksam beobachtet werden, um Implikationen der strahlenschutzspezifischen Anwendung zu prüfen.

Zusammenfassend bietet der Einsatz KI-basierter Verfahren viele Möglichkeiten, die bildgestützte Versorgung von Patientinnen und Patienten im Sinne diagnostischer Sicherheit und reduzierter Strahlenexposition zu optimieren, sofern die in dieser SSK-Empfehlung dargestellten Empfehlungen und Stellungnahmen umgesetzt werden. Um dies praktisch zu erleichtern wurden Empfehlungen und Hinweise zur Erarbeitung von KI-basierten Verfahren, ihrer Testung und Überprüfung für Entwickler, Hersteller und zuständige Stellen aufgezeigt (siehe Kapitel 2 und 9).

11 Literaturverzeichnis

- Adler und Oktem 2018 Adler J, Oktem O. Learned Primal-Dual Reconstruction. *IEEE Trans Med Imaging*. 2018 Jun;37(6):1322-32, doi: 10.1109/TMI.2018.2799231
- Baader und Kachelrieß 2025 Baader E, Kachelrieß M. Risk-minimizing tube current and tube voltage modulation for CT: A simulation study. *Med Phys*. 2025 Aug;52(8):e18047, doi: 10.1002/mp.18047
- Balogh und Janos Kis 2022 Balogh ZA, Janos Kis B. Comparison of CT noise reduction performances with deep learning-based, conventional, and combined denoising algorithms. *Med Eng Phys*. 2022 Nov;109:103897, doi: 10.1016/j.medengphy.2022.103897, Epub 20220924
- Berker et al. 2018 Berker Y, Maier J, Kachelrieß M. Deep Scatter Estimation in PET: Fast Scatter Correction Using a Convolutional Neural Network. 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), 10-17 Nov. 2018. p. 1-5
- Byl et al. 2021 Byl A, Klein L, Sawall S, Heinze S, Schlemmer HP, Kachelrieß M. Photon-counting normalized metal artifact reduction (NMAR) in diagnostic CT. *Med Phys*. 2021 Jul;48(7):3572-82, doi: 10.1002/mp.14931, Epub 20210528
- Canon Medical Systems o. J. Canon Medical Systems. Precise IQ Engine (PIQE): A New Concept in Clarity and Confidence in Cardiac Imaging. https://global.medical.canon/publication/ct/V38_CTEU220164, zuletzt aufgerufen am 12.11.2025
- Chin et al. 2022 Chin M, Chinn G, Innes D, Levin CS. Deep Learning-Based Limited Angle Tomography for a 1-Millimeter Resolution Dual-Panel Clinical PET System. 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 5-12 Nov. 2022. p. 1-3
- Chung et al. 2022 Chung H, Kim J, McCann MT, Klasky ML, Ye CJ. Diffusion Posterior Sampling for General Noisy Inverse Problems. 2022, doi: 10.48550/arXiv.2209.14687
- Duan et al. 2024 Duan L, Eulig E, Knaup M, Adamus R, Lell M, Kachelriess M. Training of a deep learning based digital subtraction angiography method using synthetic data. *Med Phys*. 2024 Jul;51(7):4793-810, doi: 10.1002/mp.16973, Epub 20240214
- Ebner und Haltmeier 2024 Ebner A, Haltmeier M. Plug-and-Play Image Reconstruction Is a Convergent Regularization Method. *IEEE Trans Image Process*. 2024;33:1476-86, doi: 10.1109/TIP.2024.3361218, Epub 20240221
- EU 2016 Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG. Amtsblatt L 119 vom 4.5.2016, S. 1, ber. Amtsblatt L 314 vom 22.11.2016, S. 72, Amtsblatt L 127 vom 23.5.2018, S. 2 (DSGVO)

- EU 2024 Europäisches Parlament und Rat der Europäischen Union. Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz). <http://data.europa.eu/eli/reg/2024/1689/oj>, zuletzt aufgerufen am 05.11.2025
- Fonseca et al. 2021 Fonseca GP, Baer-Beck M, Fournie E, Hofmann C, Rinaldi I, Ollers MC, van Elmpt WJC, Verhaegen F. Evaluation of novel AI-based extended field-of-view CT reconstructions. *Med Phys.* 2021 Jul;48(7):3583-94, doi: 10.1002/mp.14937, Epub 20210531
- GE Healthcare o. J. GE Healthcare. Revolution Maxima. <https://www.gehealthcare.de/products/computed-tomography/revolution-family/revolution-maxima>, zuletzt aufgerufen am 12.11.2025
- Gottschling 2022 Gottschling N. On existence, stability, accuracy and learning of approximate decoders for ill-posed inverse problems. doi: 10.17863/CAM.91560
- Häggström et al. 2019 Häggström I, Schmidlein CR, Campanella G, Fuchs TJ. DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. *Med Image Anal.* 2019 May;54:253-62, doi: 10.1016/j.media.2019.03.013, Epub 20190330
- Hammernik et al. 2018 Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, Knoll F. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med.* 2018 Jun;79(6):3055-71, doi: 10.1002/mrm.26977, Epub 20171108
- Han et al. 2017 Han Y, Gu J, Ye JC. Deep learning interior tomography for region-of-interest reconstruction. *arXiv.* 2017, doi: 10.48550/arXiv.1712.10248
- Huang et al. 2018 Huang Y, W\ rfl T, Breininger K, Liu L, Lauritsch Gn, Maier A. Some Investigations on Robustness of Deep Learning in Limited Angle Tomography. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018: Springer-Verlag.* p. 145–53
- Isensee et al. 2021 Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2021 Feb;18(2):203-11, doi: 10.1038/s41592-020-01008-z, Epub 20201207
- Kabelac et al. 2025 Kabelac A, Eulig E, Maier J, Hammermann M, Knaup M, Kachelrieß M. Latent space reconstruction for missing data problems in CT. *Med Phys.* 2025 Jul;52(7):e17910, doi: 10.1002/mp.17910, Epub 20250604
- Ketola et al. 2021 Ketola JHJ, Heino H, Juntunen MAK, Nieminen MT, Siltanen S, Inkinen SI. Generative adversarial networks improve interior computed tomography angiography reconstruction. *Biomed Phys Eng Express.* 2021 Oct 29;7(6), doi: 10.1088/2057-1976/ac31cb, Epub 20211029

- Klein et al. 2022 Klein L, Liu C, Steidel J, Enzmann L, Knaup M, Sawall S, Maier A, Lell M, Maier J, Kachelrieß M. Patient-specific radiation risk-based tube current modulation for diagnostic CT. *Med Phys.* 2022 Jul;49(7):4391-403, doi: 10.1002/mp.15673, Epub 20220429
- Kofler et al. 2020 Kofler A, Haltmeier M, Schaeffter T, Kachelriess M, Dewey M, Wald C, Kolbitsch C. Neural networks-based regularization for large-scale medical image reconstruction. *Phys Med Biol.* 2020 Jul 6;65(13):135003, doi: 10.1088/1361-6560/ab990e, Epub 20200706
- Lambin et al. 2012 Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012 Mar;48(4):441-6, doi: 10.1016/j.ejca.2011.11.036, Epub 20120116
- Lang 2023 Lang N. F1-Score – einfach erklärt! 9. August 2023 / Statistik. <https://databasecamp.de/statistik/f1-score>, zuletzt aufgerufen am 12.11.2025
- Laurent et al. 2023 Laurent B, Bousse A, Merlin T, Nekolla S, Visvikis D. PET scatter estimation using deep learning U-Net architecture. *Phys Med Biol.* 2023 Mar 10;68(6), doi: 10.1088/1361-6560/ac9a97, Epub 20230310
- Magonov et al. 2024 Magonov J, Maier J, Erath J, Sunnegardh J, Fournie E, Stierstorfer K, Kachelriess M. Reducing windmill artifacts in clinical spiral CT using a deep learning-based projection raw data upsampling: Method and robustness evaluation. *Med Phys.* 2024 Mar;51(3):1597-616, doi: 10.1002/mp.16938, Epub 20240116
- Mahmoodian et al. 2022 Mahmoodian N, Thadesar M, Sadeghi M, Georgiades M, Pech M, Hoeschen C. Liver Tumor Segmentation Using Deep Learning Method: RESLU-NET. *Physica Medica.* 2022 12/01;104:S132, doi: 10.1016/S1120-1797(22)02430-9
- Maier et al. 2022a Maier A, Köstler H, Heisig M, Krauss P, Yang SH. Known operator learning and hybrid machine learning in medical imaging—a review of the past, the present, and the future. *Progress in Biomedical Engineering.* 2022a;4(2), doi: 10.1088/2516-1091/ac5b13
- Maier et al. 2019a Maier AK, Syben C, Stimpel B, Würfl T, Hoffmann M, Schebesch F, Fu W, Mill L, Kling L, Christiansen S. Learning with Known Operators reduces Maximum Training Error Bounds. *Nat Mach Intell.* 2019a Aug;1(8):373-80, doi: 10.1038/s42256-019-0077-5, Epub 20190809
- Maier et al. 2018 Maier J, Sawall S, Knaup M, Kachelrieß M. Deep Scatter Estimation (DSE): Accurate Real-Time Scatter Estimation for X-Ray CT Using a Deep Convolutional Neural Network. *J Nondestruct Eval.* 2018 2018/07/10;37(3):57, doi: 10.1007/s10921-018-0507-z

- Maier et al. 2019b Maier J, Eulig E, Vöth T, Knaup M, Kuntz J, Sawall S, Kachelrieß M. Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation. *Med Phys.* 2019b Jan;46(1):238-49, doi: 10.1002/mp.13274, Epub 20181126
- Maier et al. 2021 Maier J, Lebedev S, Erath J, Eulig E, Sawall S, Fournié E, Stierstorfer K, Lell M, Kachelrieß M. Deep learning-based coronary artery motion estimation and compensation for short-scan cardiac CT. *Med Phys.* 2021 Jul;48(7):3559-71, doi: 10.1002/mp.14927, Epub 20210526
- Maier et al. 2022b Maier J, Klein L, Eulig E, Sawall S, Kachelrieß M. Real-time estimation of patient-specific dose distributions for medical CT using the deep dose estimation. *Med Phys.* 2022b Apr;49(4):2259-69, doi: 10.1002/mp.15488, Epub 20220214
- Maier et al. 2025 Maier J, Sawall S, Arbeit M, Paysan P, Kachelrieß M. Deep learning-based cone-beam CT motion compensation with single-view temporal resolution. *Med Phys.* 2025 Jul;52(7):e17911, doi: 10.1002/mp.17911, Epub 20250604
- Martins et al. 2023 Martins JC, Maier J, Gianoli C, Nepl S, Dedes G, Alhazmi A, Veloza S, Reiner M, Belka C, Kachelrieß M, Parodi K. Towards real-time EPID-based 3D in vivo dosimetry for IMRT with Deep Neural Networks: A feasibility study. *Phys Med.* 2023 Oct;114:103148, doi: 10.1016/j.ejmp.2023.103148, Epub 20231004
- Mehranian et al. 2021 Mehranian A, Wollenweber SD, Walker MD, Bradley KM, Fielding PA, Su K-H, Johnsen R, Kotasidis F, Jansen FP, McGowan DR. Image enhancement of whole-body oncology [18F]-FDG PET scans using deep neural networks to reduce noise. *Eur J Nucl Med Mol Imaging.* 2021;49(2):539-49, doi: 10.1007/s00259-021-05478-x
- Mohri et al. 2012 Mohri M, Rostamizadeh A, Talwalkar A. *Foundations of Machine Learning.* The MIT Press, 2012, ISBN 9780262018258
- MPBetreibV 2025 Medizinprodukte-Betreiberverordnung vom 14. Februar 2025 (BGBl. 2025 I Nr. 38), die zuletzt durch Artikel 1 der Verordnung vom 31. Oktober 2025 (BGBl. 2025 I Nr. 263) geändert worden ist
- MPDG 2020 Medizinprodukte-Durchführungsgesetz vom 28. April 2020 (BGBl. I S. 960), das zuletzt durch Artikel 3 des Gesetzes vom 23. Oktober 2024 (BGBl. 2024 I Nr. 324) geändert worden ist
- Muckley et al. 2021 Muckley MJ, Riemenschneider B, Radmanesh A, Kim S, Jeong G, Ko J, Jun Y, Shin H, Hwang D, Mostapha M, Arberet S, Nickel D, Ramzi Z, Ciuciu P, Starck JL, Teuwen J, Karkalousos D, Zhang C, Sriram A, Huang Z, Yakubova N, Lui YW, Knoll F. Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction. *IEEE Trans Med Imaging.* 2021 Sep;40(9):2306-17, doi: 10.1109/TMI.2021.3075856, Epub 20210831
- Nikolaus 2018 Nikolaus K. Verbesserte CT-Scans - dank Künstlicher Intelligenz. Siemens AG.
<https://www.siemens.com/de/de/unternehmen/stories/forschungstechnologien/kuenstliche-intelligenz/verbesserte-ct-scans-dank-kuenstlicher-intelligenz.html>, zuletzt aufgerufen am 12.11.2025

- Omega Medical Imaging 2023 Omega Medical Imaging. Innovations in Interventional Imaging. <https://www.omegamedicalimaging.com/innovations-in-interventional-imaging/>, zuletzt aufgerufen am 12.11.2025
- Pashazadeh und Hoeschen 2023 Pashazadeh A, Hoeschen C. Möglichkeiten der künstlichen Intelligenz im Strahlenschutz. *Die Radiologie*. 2023 Jul;63(7):530-8, doi: 10.1007/s00117-023-01167-y, Epub 20230622
- Philips Austria GmbH o. J. Philips Austria GmbH. MRCAT Brain. Klinische MR-RT-Anwendung. <https://www.philips.at/healthcare/product/HCNMRF320/mrcat-brain-klinische-mr-rt-anwendung>, zuletzt aufgerufen am 12.11.2025
- Philips GmbH 2021 Philips GmbH. Philips Incisive CT Precise Suite. KMA-Online, Georg Thieme Verlag KG. <https://www.kma-online.de/unternehmen/philips/produkte/detail/philips-incisive-ct-precise-suite-82>, zuletzt aufgerufen am 12.11.2025
- Qi 2006 Qi J. Comparison of Lesion Detection and Quantification in MAP Reconstruction with Gaussian and Non-Gaussian Priors. *Int J Biomed Imaging*. 2006;2006:87567, doi: 10.1155/IJBI/2006/87567, Epub 20060629
- Ronneberger et al. 2015 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015, doi: 10.48550/arXiv.1505.04597
- Schramm et al. 2021 Schramm G, Rigie D, Vahle T, Rezaei A, Van Laere K, Shepherd T, Nuyts J, Boada F. Approximating anatomically-guided PET reconstruction in image space using a convolutional neural network. *Neuroimage*. 2021 Jan 1;224:117399, doi: 10.1016/j.neuroimage.2020.117399, Epub 20200921
- Sidky und Pan 2022 Sidky EY, Pan X. Report on the AAPM deep-learning sparse-view CT grand challenge. *Med Phys*. 2022 Aug;49(8):4935-43, doi: 10.1002/mp.15489, Epub 20220209
- Siegel 2001 Siegel S. Nichtparametrische statistische Methoden. Klotz, 2001, ISBN 9783880741027
- Siemens Healthineers AG o. J.-a Siemens Healthineers AG, (Siemens Healthineers AG). YSIO X.pree. Intelligence for excellence. <https://www.siemens-healthineers.com/de/radiography/digital-x-ray/ysio-xpree>, zuletzt aufgerufen am 12.11.2025
- Siemens Healthineers AG o. J.-b Siemens Healthineers AG. MR-only Radiotherapy Planning. Powered by syngo.via RT Image Suite. <https://www.siemens-healthineers.com/magnetic-resonance-imaging/clinical-specialities/synthetic-ct>, zuletzt aufgerufen am 12.11.2025
- Singh et al. 2023 Singh IRD, Denker A, Jin B, Thielemans K, Arridge S. Investigating Intensity Normalisation for PET Reconstruction with Supervised Deep Learning. 2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD), 4-11 Nov. 2023
- Spectronic Medical AB o. J. Spectronic Medical AB. MRI only radiotherapy planning. <https://medical.spectronic.se/page-2/index.html>, zuletzt aufgerufen am 12.11.2025

- SSK 2016 Strahlenschutzkommission. Dosisdokumentation und Archivierung digitaler Bild- und Untersuchungsdaten in Radiologie und Nuklearmedizin, verabschiedet in der 282. Sitzung der Strahlenschutzkommission am 23./24. Juni 2016. urn:nbn:de:101:1-201703087447. Bekanntmachung im BAnz AT 20.03.2017 B3
- StrlSchG 2017 Gesetz zum Schutz vor der schädlichen Wirkung ionisierender Strahlung (Strahlenschutzgesetz - StrlSchG) vom 27. Juni 2017 (BGBl. I S. 1966), das zuletzt durch die Bekanntmachung vom 3. Januar 2022 (BGBl. I S. 15) geändert worden ist
- StrlSchV 2018 Verordnung zum Schutz vor der schädlichen Wirkung ionisierender Strahlung (Strahlenschutzverordnung - StrlSchV) vom 29. November 2018. BGBl. I S. 2034, 2036, die durch Artikel 1 der Verordnung vom 27. März 2020 (BGBl. I S. 748) geändert worden ist
- Therasse et al. 2006 Therasse P, Eisenhauer EA, Verweij J. RECIST revisited: a review of validation studies on tumour assessment. *Eur J Cancer*. 2006 May;42(8):1031-9, doi: 10.1016/j.ejca.2006.01.026, Epub 20060417
- Trapp et al. 2022 Trapp P, Amato C, Sawall S, Kachelrieß M, Vöth T. DeepRAR: a CNN-based approach for CT and CBCT ring artifact reduction. *SPIE Medical Imaging*, 2022: SPIE. p. 337-44
- Würfl et al. 2016 Würfl T, Ghesu FC, Christlein V, Maier A Hrsg. *Medical Image Computing and Computer-Assisted Intervention - MICCAI*. 2016 Cham
- Xiang et al. 2020 Xiang H, Lim H, Fessler JA, Dewaraja YK. A deep neural network for fast and accurate scatter estimation in quantitative SPECT/CT under challenging scatter conditions. *Eur J Nucl Med Mol Imaging*. 2020 Dec;47(13):2956-67, doi: 10.1007/s00259-020-04840-9, Epub 20200515
- Zhu et al. 2018 Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature*. 2018 Mar 21;555(7697):487-92, doi: 10.1038/nature25988

Begriffsbestimmungen

Adversarial Attacks	Gezielte Manipulationen der Eingabedaten, die für Menschen möglicherweise kaum sichtbar sind, aber ein KI-Modell zu extrem falschen Ergebnissen verleiten.
Annotierung	Das manuelle oder halbautomatische Markieren von Bilddaten (z. B. Umriss einer Läsion, oder Rauschstrukturen oder stark verrauschtes Bild etc.), damit diese für das Training von KI-Methoden genutzt werden können.
Autoencoder	Ein neuronales Netz, das Daten in eine komprimierte Darstellung überführt und daraus wieder rekonstruiert. Nützlich beispielsweise zur Rauschunterdrückung oder Merkmalsextraktion.
Backpropagation-Algorithmus	Das Standardverfahren zum Training neuronaler Netze, bei dem Fehler zwischen der Netzausgabe und dem Trainingslabel rückwärts durch das Netz propagiert werden, um die Gewichte anzupassen.
Bayesianisches Learning	Ein Verfahren, das Wahrscheinlichkeiten und Unsicherheiten im Modell explizit berücksichtigt und so auch Aussagen über die Verlässlichkeit von Vorhersagen ermöglicht.
Black-Box-Modell	Ein Modell, dessen innere Struktur und Parameter nicht vollständig bekannt oder nachvollziehbar sind. Die Funktionsweise erschließt sich nur aus den Eingabe-Ausgabe-Beziehungen. Es kann sich hierbei um geheim gehaltene Funktionen handeln oder wie im Falle der neuronalen Netze um Funktionen mit extrem vielen Parametern (z. B. Millionen bis Milliarden Parameter), die zwar nicht geheim gehalten werden, aber aufgrund der schier unendlichen Anzahl nicht zu verstehen sind.
Convolutional Neural Network (CNN)	Ein spezieller Typ neuronaler Netze, der Faltungsoperationen verwendet. Besonders geeignet für Bilddaten, da Nachbarschaftsbeziehungen erhalten bleiben. Für die Bildrekonstruktion spielen 2D und 3D CNNs besonders häufig eine wichtige Rolle.
Datenaugmentierung	Methoden zur künstlichen Erweiterung von Datensätzen (z. B. durch Drehen, Spiegeln oder Rauschen von Bildern), um die Robustheit von KI-Verfahren zu erhöhen. Durch die Augmentierung kann die Menge an Trainingsdaten um ein Vielfaches erhöht werden.
Deep Learning (DL)	Unter deep learning wird das maschinelle Lernen (siehe dort) anhand großer Datenmengen verstanden.
Deep Scatter Estimation	Ein KI-Verfahren zur extrem schnellen und genauen Schätzung von Streustrahlung in CT- und PET-Daten, auf Basis tiefer neuronaler Netze. Der Goldstandard der Streustrahlungsberechnung wäre die Monte-Carlo-Methode oder die Boltzmann-Transport-Methode. Beide sind aber sehr rechenaufwändig und benötigen daher deutlich längere Rechenzeiten.

End-to-End Learning	Ein Ansatz, bei dem ein neuronales Netz Rohdaten direkt in ein Ergebnis (z. B. ein Bild) überführt, ohne explizit physikalische oder mathematische Modelle einzubeziehen. Das Gegenteil von End-to-End-Learning wäre ein modularer oder komponentenbasierter Ansatz, bei dem der gesamte Verarbeitungsprozess nicht von einem einzigen Netz abgebildet wird, sondern in mehrere aufeinanderfolgende Schritte zerlegt ist, die einzeln trainiert oder optimiert werden.
Foundation Model	Ein Foundation Model oder Basismodell ist ein großes maschinelles Lernmodell, das auf einer breiten Datenbasis vortrainiert wurde und durch seine Vielseitigkeit als Grundlage für eine Vielzahl spezifischer Anwendungen dient.
Generative Netze	Neuronale Netze, die darauf trainiert werden, neue Daten zu erzeugen, die den Trainingsdaten möglichst ähnlich sind. In der Bildgebung können sie z. B. realistisch wirkende, aber künstlich erzeugte Bilder oder Bildanteile generieren. Generative Netze haben fast immer eine stochastische Komponente – also eine Zufallszahl oder einen Zufallsvektor als Eingabe. Dadurch liefern solche generative Netze mit stochastischer Komponente nicht deterministisch immer das gleiche Ergebnis, sondern können aus derselben Datenverteilung viele verschiedene, aber „realistisch wirkende“ Varianten erzeugen.
Gradientenabstieg	Ein Optimierungsverfahren, das Parameter iterativ in Richtung des steilsten Abfalls einer Fehlerfunktion verändert, um diese zu minimieren. Backpropagation (s. o.) ist die Technik, mit der die Gradienten berechnet werden.
Gray-Box-Modell	Ein Hybridmodell, bei dem die Struktur bekannt ist, aber nicht alle Parameter. Vorwissen wird genutzt, um die Zahl der zu lernenden Parameter zu reduzieren. Siehe auch Black-Box-Modell.
Ground Truth	Referenzwerte oder -bilder, die als „wahr“ angenommen werden und zum Vergleich mit den Ergebnissen von KI-Verfahren dienen, also für das Training herangezogen werden. In diesem Fall dient die Ground Truth als Label (s. u.) für das Training.
Grundgesamtheit	Die Gesamtheit aller Objekte oder Personen, über die in einer Untersuchung Aussagen getroffen werden sollen. Stichproben für Training, Validierung und Testung sollen daraus zufällig entnommen werden.
Hyperparameter	Hyperparameter bezeichnen Parameter eines KI-Modells, die verwendet werden, um den Trainingsalgorithmus zu steuern. Ihre Werte müssen vor dem Training des Modells festgelegt werden.
Inpaintingverfahren	Beim Inpaintingverfahren werden zerstörte oder nicht vorhandene Bildanteile oder Projektionsdaten ergänzt, um so ein überarbeitetes, möglichst artefaktfreies Bild zu erzeugen.

Künstliche Intelligenz (KI)	Bereich der Informatik, der sich mit der Entwicklung von Maschinen befasst, die menschenähnliche Intelligenz aufweisen.
Label	Die zu einem Eingabedatum gehörende Referenzinformation, die im überwachten Lernen als Ziel dient. Labels können z. B. Klassenzugehörigkeiten (Tumor vorhanden/nicht vorhanden), Bildmarkierungen (Segmentierungen) oder Referenzbilder (z. B. solche mit sehr wenig Rauschen oder wenig Artefakten) sein. Sie entstehen typischerweise durch Annotierung.
Layer	Verarbeitungsschicht im Neuronalen Netz (siehe dort)
Maschinelles Lernen (ML)	Maschinelles Lernen (ML) ist ein wichtiger Teilbereich der KI (siehe dort), welches es Maschinen ermöglicht, aus Erfahrung zu lernen, Muster in Datenmengen zu erkennen und so Aufgaben zu automatisieren, für die normalerweise menschliche Intelligenz erforderlich wäre.
Maximum-a-posteriori-Rekonstruktion (MAP)	Ein Rekonstruktionsverfahren, das Vorwissen („Priors“) in die Berechnung einbezieht, um robustere Ergebnisse bei verrauschten oder unvollständigen Daten zu erzielen.
Monte-Carlo-Simulation	Ein Verfahren, das mithilfe von Zufallszahlen physikalische Prozesse nachbildet. Es wird z. B. in der medizinischen Physik oft genutzt, um oder Dosis- oder Streustrahlverteilungen zu berechnen.
Neuronales Netz	Eine beliebige Anzahl miteinander verknüpfter Neuronen, bei künstlichen neuronalen Netzen digitale Neuronen, deren Aufgaben und Verknüpfungen von biologischen Neuronen inspiriert sind. Ein neuronales Netz besteht aus vielen Verarbeitungsschichten (Layern). Jeder Layer bearbeitet die Daten nichtlinear und reicht sie an den nächsten Layer weiter. Nach typischerweise einigen Dutzend bis Hundert Schichten gibt das neuronale Netz die bearbeiteten Daten, das bearbeitete Bild oder den bearbeiteten Bilderstapel aus.
Out-of-Distribution	Bezeichnet Daten, die sich deutlich von den Trainingsdaten unterscheiden. Modelle liefern hier oft unsichere oder falsche Ergebnisse. Out-of-Distribution-Fälle können durch vielfältige und repräsentative Trainingsdaten sowie geeignete Teststrategien reduziert, aber nicht vollständig ausgeschlossen werden. Es gibt auch Mechanismen zur Erkennung und Anzeige von derartigen Unsicherheiten.
Physics-driven Learning	Ein Verfahren, das physikalisches Vorwissen (z. B. Bildentstehungsprozesse) in den Lernprozess integriert, um Robustheit und Plausibilität der Ergebnisse zu erhöhen.
Poisson-Thinning	Ein Verfahren, um aus vollständigen Messdaten stochastisch korrekt Teilmengen zu erzeugen, die einer reduzierten Zählrate entsprechen, etwa bei kürzerer Messzeit oder geringerer Dosis.

Post-Reconstruction-Learning	KI-Verfahren, die auf bereits rekonstruierte Bilder angewendet werden, um deren Qualität zu verbessern (z. B. Rauschunterdrückung, Artefaktreduktion).
Prior	Ein Vorwissen oder eine Annahme, die vor der Datenanalyse in ein Modell eingebracht wird, z. B. Glattheit oder Homogenität bestimmter Strukturen.
Radiomics	Ein meistens KI-basiertes Verfahren, bei denen aus lokalen Bildeigenschaften von Befunden, wie der Signalintensität, der Textur oder der Kontur versucht wird, z. B. Zusammenhänge mit der Artdiagnose (z. B. benigne oder maligne Läsion) abzuleiten.
Regularisierung	Ein allgemeiner Ansatz, um Modelle zu stabilisieren und Überanpassung zu vermeiden. Dabei wird zusätzliches Wissen oder eine Nebenbedingung in die Optimierung eingebracht, z. B. Glattheit. Regularisierung ersetzt fehlende Informationen durch Vorwissen und stabilisiert so die Lösungsauswahl.
Regularisierungsschritt	Eine Maßnahme im Training, die Überanpassung (Overfitting) verhindert, indem Modellkomplexität reduziert oder geglättet wird.
Semi-supervised Learning	Eine Mischform, bei der sowohl annotierte als auch nicht annotierte Daten für das Training genutzt werden.
Supervised Learning	Überwachtes Lernen ist ein Lernverfahren, bei dem Eingabedaten mit bekannten Zielwerten (Labels, siehe dort) trainiert werden. Das Netz lernt, die Zuordnung von Eingabe zu Ausgabe nachzuvollziehen.
Total-Body PET	Ein PET-Scanner mit großem axialem Sichtfeld (oft 1 m bis 2 m), der einen großen anatomischen Bereich oder gar den ganzen Körper gleichzeitig erfassen kann und dadurch besonders hohe Sensitivität ermöglicht.
Transfer-Learning	Ein Ansatz, bei dem ein bereits vortrainiertes Modell auf eine neue, ähnliche Aufgabe angepasst wird, um weniger Trainingsdaten und Rechenzeit zu benötigen.
U-Netz	Eine spezielle Architektur von CNNs, die Eingabedaten in einer Encoder-Decoder-Struktur verarbeitet, wobei die Enkoder- und Dekoderebenen durch direkte Verbindungen, den so genannten Skip-Connections, verbunden sind. Der Encoder verarbeitet die Eingangsdaten und erzeugt einen Vektor, der die relevante Information in komprimierter Weise enthält. Der Decoder nutzt diese Information, kann sie filtern und aus den Daten dann eine Ausgabe erzeugen. Ursprünglich für Segmentierung entwickelt, heute auch in vielen Bildverbesserungsaufgaben eingesetzt.

Unsupervised Learning	Unüberwachtes Lernen ist ein Lernverfahren, das ohne Labels (siehe dort) auskommt. Das Netz erkennt selbstständig Strukturen oder Muster in den Daten.
Vorwärtsprojektion	Die rechnerische Simulation von Projektionen aus einem Bild oder Volumen (Bilderstapel) heraus, basierend auf den physikalischen Eigenschaften des Bildgebungsverfahrens. Die Vorwärtsprojektion eines CT-Volumens erzeugt eine Projektion, die wie ein Röntgenbild aussieht.
White-Box-Modell	Ein Modell mit vollständig bekannten Strukturen und Parametern, oft physikalisch fundiert und gut nachvollziehbar.
Wilcoxon-Test	Ein nichtparametrischer statistischer Test zum Vergleich zweier verbundener Stichproben, ohne dass eine Normalverteilung angenommen werden muss.

Abkürzungsverzeichnis

AUTOMAP	Software-Anwendung, die als Schnittstelle zwischen Hardware-Controller und anderen Software-Anwendungen fungiert.
BMUKN	Bundesministerium für Umwelt, Klimaschutz, Naturschutz und nukleare Sicherheit, früher BMUV: Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
CBCT	Kegelstrahl-Computertomographie (Cone Beam Computed Tomography)
cGAN	Konditionales generatives antagonistisches Netzwerk (conditional Generative Adversarial Network)
CNN	faltungsbasiertes neuronales Netzwerk (Convolutional Neural Networks)
CPU	Hauptprozessor, zentraler Rechenkern (Central Processing Unit)
CT	Computertomographie (Computed Tomography)
CTDI	Computer Tomographie Dosis Index (Computed Tomography Dose Index)
DL	Deep Learning
DLP	Dosislängenprodukt (Dose-Length Product)
DNN	Tiefes Neuronales Netzwerk (Deep Neural Network)
DSA	Digitale Subtraktionsangiographie (Digital Subtraction Angiography)
DSE	Deep Scatter Estimation
DVF	Displacement Vector Fields

DSGVO	Datenschutz-Grundverordnung
DVT	Digitale Volumentomografie (Digital Volume Tomography)
FBP	Gefilterte Rückprojektion (Filtered Back Projection)
GPGPU	Hochleistungsgrafikprozessor (General Purpose Computation on Graphics Processing Unit)
GPU	Grafikprozessor (Graphic Processing Unit)
HU	Hounsfield Einheit (Hounsfield Unit)
KI	Künstliche Intelligenz
MAP	Maximum-a-posteriori-Rekonstruktion
MIP	Maximumintensitätsprojektion (Maximal Intensity Projection)
ML	Maschinelles Lernen (Machine Learning)
MLEM	Ermittlung der maximal bedingten Auftretenswahrscheinlichkeit für ein gefundenes Ergebnis (Maximum Likelihood Expectation Maximization)
MRT	Magnetresonanztomographie
OOD	Out-of-Distribution
PET	Positronen Emissions Tomographie (Positron Emission Tomography)
RECIST	Kriterien für die Bewertung des Ansprechens der Behandlung bei soliden Tumoren (Response Evaluation Criteria In Solid Tumors)
SOP	Standardarbeitsanweisung (Standard Operating Procedure)
SPECT	Einzelphotonen- Emissions- Computertomographie (Single-Photon Emission Computed Tomography)
SSDE	Größenspezifische Dosisabschätzung (Size-Specific Dose Estimation)
SSK	Strahlenschutzkommission
StrlSchG	Strahlenschutzgesetz
StrlSchV	Strahlenschutzverordnung
TCM	Röhrenstrommodulation (Tube Current Modulation)
U-Net	CNN zur schnellen und präzisen Segmentierung von Bildern
Voxel	3D-Bildelement (Volume Picture Element)