

Arthroskopie
<https://doi.org/10.1007/s00142-024-00669-8>
 Angenommen: 13. März 2024

© The Author(s) 2024

Redaktion
 W. Petersen, Berlin
 A. Voss, Regensburg



Chance und Risiken von künstlicher Intelligenz und Machine Learning in der bildgebenden Diagnostik

Stefan Nehrer¹ · Kenneth Chen^{1,2} · Richard Ljuhar³ · Christoph Götz³

¹ Fakultät für Gesundheit und Medizin, Department für Gesundheitswissenschaften, Medizin und Forschung, Zentrum für Regenerative Medizin, Zentrum für Gesundheitswissenschaften und Medizin, Universität für Weiterbildung Krems Donau, Krems, Österreich; ² Abteilung für Orthopädie und Traumatologie, Landesklinikum Waidhofen/Ybbs, Waidhofen an der Ybbs, Österreich; ³ Research & AI Development Abteilung, ImageBiopsy Lab, Wien, Österreich

In diesem Beitrag

- Interrater- und Intrarater-Variabilität
- Künstliche Intelligenz und Machine Learning Einteilung
- Interne vs. externe Validierung
- Muskuloskeletale Erkrankungen und bildgebende Diagnostik
- Klassifizierung der Gonarthrose
- Automatisierte Vermessung der Hüfte
- Automatisierte Vermessung der Beinachse
- Status quo in der bildgebenden Diagnostik – Potenzial der KI für den Befundablauf

Bemerkung und Offenlegung

R. Ljuhar und C. Götz sind in der Geschäftsführung von IB Lab tätig. S. Nehrer und K. Chen sind im Rahmen eines Forschungsprojekts in Kooperation mit IB Lab tätig. Das Projekt wird von der GFF Niederösterreich (Projektnummer: LS200020) gefördert.

Während der Erstellung dieser Arbeit verwendeten die Autoren ChatGPT 3.5 von OpenAI, um den Lesefluss zu verbessern und Wortsynonyme zu finden. Die Autoren überprüften den Inhalt anschließend und haben diesen entsprechend überarbeitet.



QR-Code scannen & Beitrag online lesen

Zusammenfassung

Künstliche Intelligenz (KI) findet zunehmend Anwendung in der bildgebenden Diagnostik. Dieser umfassende Begriff bezieht sich auf Computerprogramme, die in der Lage sind, intelligente Aufgaben zu übernehmen und zu lösen. Die stetige Weiterentwicklung der KI-Architekturen ermöglicht es, auch anspruchsvolle Aufgaben wie die Erkennung und Quantifizierung von radiologischen Parametern auf einem höheren Niveau zu bewältigen. Gegenwärtig erfolgen die Beurteilung und Beschreibung solcher Parameter größtenteils auf manuelle Weise und in narrativer Form. Diese manuelle Vorgehensweise ist nicht nur zeitintensiv, sondern auch anfällig für Interrater- und Intrarater-Variabilität, da sie stark von der beurteilenden Person und äußeren Einflüssen beeinflusst wird. Mithilfe von KI-Algorithmen können standardisierte und reproduzierbare Ergebnisse entstehen, da sie unabhängig von externen Einflüssen Informationen in Bilddaten auf den Pixel genau auswerten kann. Ein entscheidender Vorteil besteht darin, dass die KI im Gegensatz zur manuellen Beurteilung auch auf umfangreiche Hintergrunddaten zurückgreifen kann, was zu einer weiteren Verbesserung der Genauigkeit führt. In der Rolle eines unterstützenden Tools trägt die KI dazu bei, die Qualität der Röntgenbildbeurteilung zu steigern, während gleichzeitig die Arbeitsbelastung reduziert wird.

Schlüsselwörter

Röntgenbildgebung · Radiologische Parameter · Machine Learning · Deep Learning · Validierung

In den letzten Jahren hat künstliche Intelligenz (KI) erheblich an Bedeutung gewonnen. Obwohl die Ursprünge der KI bereits im Jahr 1955 liegen [1], ist sie seit dem Auftreten von Large Language Models wie ChatGPT aktueller denn je. Im orthopädisch-traumatologischen Bereich sind bereits KI-Modelle im Einsatz, die diagnostische Funktionen übernehmen, beispielsweise in der Frakturerkennung oder bei der Vermessung bzw. Quantifizierung von radiologischen Parametern [2–7]. Trotz dieser Fortschritte werden Röntgenbilder im klinischen Alltag weiterhin manuell von Ärztinnen beurteilt,

vermessen und narrativ beschrieben. Diese herkömmliche Bewertungsmethode ist jedoch subjektiv, was zu erheblichen Unterschieden in den Ergebnissen führen kann.

Interrater- und Intrarater-Variabilität

Die manuelle Bewertung und Vermessung von Röntgenbildern unterliegen verschiedenen Einflussfaktoren. Stress im klinischen Alltag, Ermüdung oder Unachtsamkeit können das Ergebnis beeinträchtigen. Die Intrarater-Variabilität beschreibt

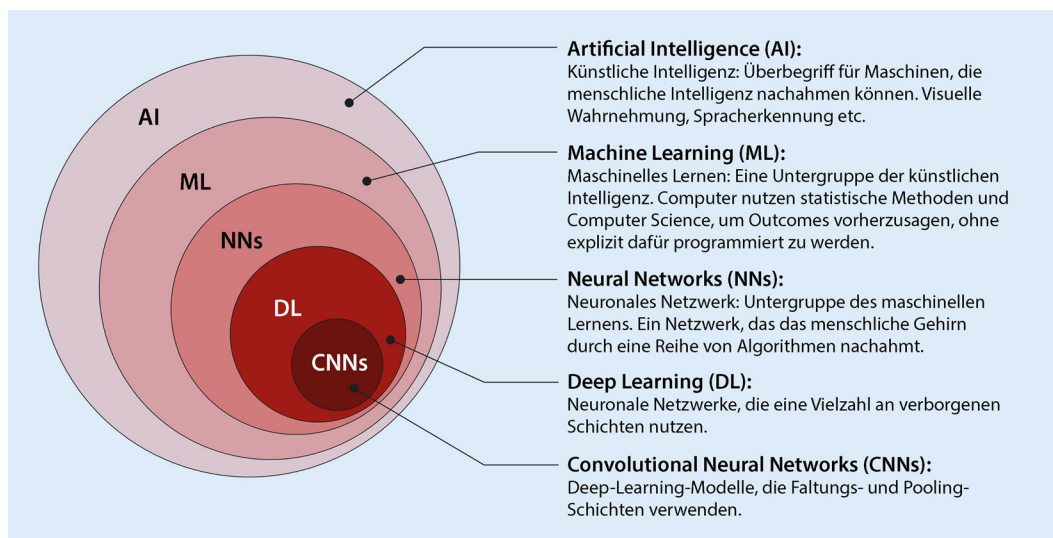


Abb. 1 ◀ Das Mengendiagramm zeigt die Untergruppen der künstlichen Intelligenz (KI) auf

Unterschiede zwischen Bewertungen desselben Befunders, während die Interrater-Variabilität Unterschiede zwischen zwei verschiedenen Befundern angibt. Trotz der Verfügbarkeit von 5 Megabyte an Daten in einem DICOM-Röntgenbild (Digital Imaging and Communications in Medicine) werden die Bilder nach wie vor subjektiv auf einem Bildschirm bewertet, ähnlich wie vor 100 Jahren. Die Variabilität in der Genauigkeit von Messergebnissen bei rein manuellen Auswertungen konnte bereits im Zuge früherer Studien gezeigt werden [8–10].

Ein Algorithmus basierend auf künstliche Intelligenz (KI) ist unabhängig von Erfahrung oder Tagesverfassung eines Befunders. Zusätzlich kann die Software die volle Dateninformation verwenden und beispielsweise Kontrastunterschiede auf den Pixel genau beurteilen. Dies führt zu einem präzisen und vor allem reproduzierbaren Ergebnis.

Künstliche Intelligenz und Machine Learning

Künstliche Intelligenz (KI) ist ein Sammelbegriff für Computerprogramme, die die Fähigkeit besitzen, intelligente Aufgaben zu übernehmen und zu lösen. Die KI erlebte eine neue Blütezeit mit dem Fortschritt der Rechenleistung, ermöglicht durch moderne Graphics Processing Units (GPU) sowie der Verfügbarkeit umfangreicher (Trainings- und Validierungs-)Datensätze. Diese Entwicklungen führten zur Entstehung

von Untergruppen innerhalb der KI durch die Implementierung komplexerer Algorithmen (▣ Abb. 1).

Machine Learning (ML) ist eine dieser Untergruppen. ML-Algorithmen können Zusammenhänge von Variablen erkennen und zuvor unbekannte Probleme lösen, ohne explizit dafür programmiert zu sein. ML-Algorithmen können eigenständig Muster und Hypothesen aus Trainingsdaten ableiten, ohne dass Menschen die zu erkennenden Features selbst vorgeben müssen. So löst ein ML-Algorithmus eine Aufgabe eigenständig und ist flexibel in der Findung des Lösungswegs. Einerseits sind diese Eigenständigkeit und Flexibilität ein Vorteil, andererseits können sie auch zum Nachteil werden. Dadurch, dass nicht jeder einzelne Schritt überwacht werden muss oder kann, können nicht alle Teilschritte nachvollzogen werden, und man spricht von einer *Black-Box*.

Einteilung

Machine Learning kann unterteilt werden in:

- **Überwachtes Lernen (Supervised Learning):** In diesem Ansatz wird ein Algorithmus anhand von Daten trainiert, die zu einem bestimmten Input (z. B. Knie-Röntgenbilder) ein gewünschtes Ergebnis (z. B. Kellgren-Lawrence-Grade) liefern. Das Ziel besteht darin, dass der Algorithmus nach erfolgreichem Training und Validie-

rung in der Lage ist, eigenständig neue Daten zu bewerten.

- **Unüberwachtes Lernen (Unsupervised Learning):** Im unüberwachten Lernen wird der Algorithmus nicht anhand vordefinierter Ergebnisse trainiert. Hier steht die Erzeugung von Gruppierungen von Merkmalen, auch durch sog. Cluster-Verfahren, im Vordergrund. Ziel ist es, dass der Algorithmus Muster erkennt, die Menschen schwer erfassen können. Bei großen Datenmengen können so verborgene Muster identifiziert werden, was wiederum Rückschlüsse auf Ursachen von Erkrankungen ermöglichen kann.
- **Bestärkendes Lernen (Reinforcement Learning):** Dieser Ansatz belohnt den Algorithmus für das Erreichen eines bestimmten Ziels. Dabei sucht der Algorithmus nach dem optimalen Weg zur Zielerreichung. Ein bekanntes Anwendungsbeispiel findet sich im Bereich von Spielen wie Schach, wo das Gewinnen als definiertes Ziel gilt und der Algorithmus den optimalen Weg dafür sucht.

Deep Learning (DL) ist eine spezielle Methode des ML. Es verwendet neuronale Netzwerke, die in ihrer Architektur dem menschlichen Gehirn ähneln, um komplexe Aufgaben wie Bilderkennung zu lösen. In der bildgebenden Diagnostik werden DL-Algorithmen verwendet, um (unter anderen) Anomalien auf Röntgenbildern zu detektieren. Eine entscheidende

Voraussetzung dafür ist, dass beim Training des Algorithmus qualitativ hochwertige und vielschichtige Daten zum Einsatz kommen. Das bedeutet, dass sämtliche Merkmale einer Erkrankung auf den Röntgenbildern, die für das Training verwendet werden, möglichst umfassend vertreten sind. Die Qualität und Vielfalt der Daten spielen eine zentrale Rolle. Bei einem Datensatz mit hoher Varianz (unterschiedliche Erscheinungsbilder bei derselben Erkrankung) sind größere Datensätze erforderlich, um robuste DL-Modelle zu entwickeln. Eine umfassende Repräsentation der Erkrankungsmerkmale auf den Trainingsbildern ist entscheidend, um die Fähigkeit des Algorithmus zur zuverlässigen Erkennung verschiedener Krankheitsbilder zu gewährleisten. Damit leisten die Auswahl und Qualität der Trainingsdaten einen maßgeblichen Beitrag zur Effektivität und Genauigkeit von DL in der medizinischen Bildgebung.

Interne vs. externe Validierung

Die Entwicklung eines DL-Algorithmus erfordert sowohl Training als auch Validierung, wobei Letztere die Bewertung der Genauigkeit eines Modells darstellt. Der Validierungsprozess kann in interne und externe Validierung unterteilt werden. Bei der internen Validierung erfolgt eine Aufteilung desselben Datensatzes in Trainings- und Validierungsdaten. Die Cross-Validierung ist ein Prozess, bei dem der Datensatz mehrmals in Trainings- und Validierungsdaten unterteilt wird, bis alle Daten für das Training verwendet wurden. Der Vorteil der internen Validierung liegt in der Notwendigkeit kleinerer Datensätze und der einfacheren Durchführung. Dennoch ermöglicht selbst eine hochgenaue Cross-Validierung keine Aussage über die Reproduzierbarkeit und Anwendbarkeit auf andere Datensätze. Es besteht auch die Möglichkeit eines zusätzlichen Bias, wenn der Trainingsdatensatz die gewünschte Zielgruppe nicht angemessen repräsentiert. Ein Beispiel dafür ist die Untersuchung von Zech et al., die zeigte, dass das Wort „portable“ als entscheidender Faktor für die Diagnose einer Lungenentzündung auf Thorax-Röntgenbildern fungierte, obwohl es nur als Kennzeichnung für das tragbare Rönt-

gerät auf den Bildern erschien [11]. Externe Validierung hingegen verwendet einen Datensatz für Training und einen zusätzlichen (externen) Datensatz für die Validierung. Der Validierungsdatensatz soll sich grundsätzlich vom Trainingsdatensatz unterscheiden, beispielsweise geografisch durch Röntgenbilder aus verschiedenen Ländern bzw. Modalitäten. Externe Validierung ist deswegen zwar aufwendiger und Daten schwieriger zu erheben, allerdings können hierdurch oben genannte Nachteile der internen Validierung verringert werden. In der Regel wird eine externe Validierung als Grundlage für die Zulassung als Medizinprodukt verlangt.

Muskuloskeletale Erkrankungen und bildgebende Diagnostik

Muskuloskeletale (MSK) Erkrankungen sind die häufigsten arbeitsbezogenen Erkrankungen in Europa und der häufigste Grund für das Aufsuchen der Notaufnahme [12–14]. Mit der wachsenden Anzahl an Röntgenbildern die befundet werden müssen, stehen befundende Ärzt:innen unter erhöhtem Druck eine hohe Qualität aufrecht zu erhalten. Die geschätzte radiologische Fehlerrate liegt bei 4–30%. Auf eine Milliarde Röntgenbilder im Jahr würden demzufolge etwa 40 Mio. radiologische Bilder fehlerhaft diagnostiziert [13, 15].

MSK-Bildgebung wird verwendet, um ein besseres Verständnis für die Anatomie zu gewinnen und mit diesem Verständnis diagnostische Verfahren für die Erkennung von MSK-Erkrankungen zu entwickeln. Im Kontext von KI kann in diesem Bereich in 2D und 3D unterteilt werden. Zur Bildererkennung werden meist DL-Modelle verwendet, um vor allem Auswertungen auf 2D-Bildern durchzuführen. Obwohl Computertomographie (CT) und Magnetresonanztomographie (MRT) als 3D-Verfahren angesehen werden, basiert die Auswertung auf einer Aneinanderreihung von 2D-Bildern. Spezielle DL-Architekturen können auch multidimensionale Daten klassifizieren anstelle von sequenziellen 2D-Daten.

Die Implementierung von KI in die MSK-Bildgebung kann aktuell so verstanden werden, dass bestimmte, klar definierte Aufgaben, die normalerweise von Radiologen oder Orthopäden durchgeführt wer-

den, übernommen werden können. Dies beinhaltet Aufgaben wie die Erkennung und Vermessung von anatomischen Strukturen, Frakturerkennung sowie komplexere Aufgaben wie die Beurteilung des anatomischen Knochenalters oder der Klassifizierung von unterschiedlichen Stadien der Arthrose.

Klassifizierung der Gonarthrose

Die Arthrose des Kniegelenks bzw. Osteoarthrose (OA) betrifft über 200 Mio. Menschen weltweit, mit einem Lebenszeitrisiko von bis zu 45% [16, 17]. Durch die Befundung nach dem semiquantitativen Kellgren-Lawrence (KL)-Score weist diese eine hohe Inter- und Intra-rater-Variabilität auf und ist dadurch ungeeignet für standardisierte Therapieentscheidungen und Studien, welche die Wirksamkeit für Gonarthrose untersuchen [8].

In Studien konnte bereits gezeigt werden, dass mithilfe von KI eine höhere Übereinstimmung zwischen befundenden Ärzt:innen erreicht wird [18, 19]. Als erste KI-Lösung für dieses Fachgebiet wurde IB Lab KOALA (■ Abb. 2) umgesetzt. Dabei handelt es sich um ein DL-Modell für die automatisierte Bestimmung des KL-Grads, welches an über 35.000 Knie-Röntgenbildern trainiert wurde. Dies ist eine verhältnismäßig hohe Zahl, da es auch andere KI-Modelle gibt, für welche unter 500 Trainingsbilder verwendet wurden [20–22].

Automatisierte Vermessung der Hüfte

Wie bereits erwähnt, unterliegt die manuelle Vermessung von Röntgenbildern einem gewissen Bias. Für die Erstellung von Diagnosen und Indikationsstellung für Therapien ist allerdings eine möglichst standardisierte und reproduzierbare Messung ausschlaggebend. Mit dem Hüftvermessungstool IB Lab HIPPO (■ Abb. 3) konnte gezeigt werden, dass die KI bei der Vermessung von Caput-Caput-Diaphysen-Winkel (CCD), Lateral-Center-Edge Winkel (LCE), Tönnis-Winkel, Sharp-Winkel und Femur-Kopf-Extrusion-Index im Vergleich zu menschlichen Readern gleich gut oder besser abschneidet [2].



Abb. 2 ▲ IB Lab KOALA Beispiel Report. Kellgren-Lawrence-Grade, OARSI-Grade und Gelenkspalthöhe sowie die standardisierte Gelenkspalthöhe werden vermessen

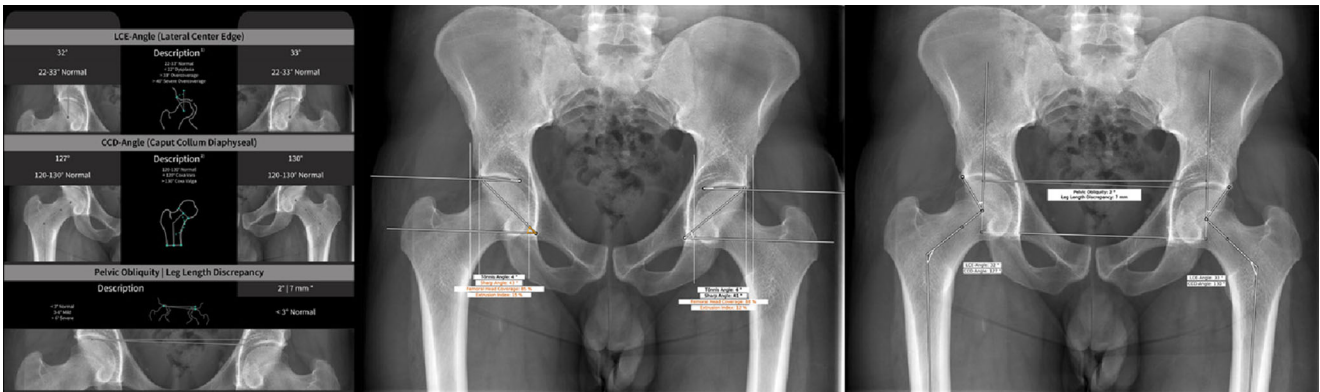
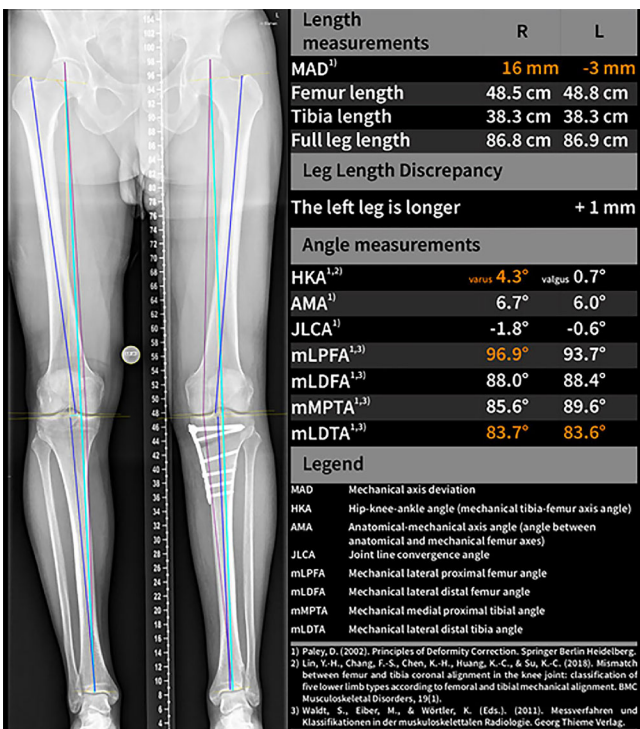


Abb. 3 ▲ IB Lab HIPPO Beispiel Report. Vollautomatisierte Vermessung der Hüfte



Automatisierte Vermessung der Beinachse

Die korrekte Vermessung und Beurteilung der Beinachse ist für (beispielsweise) Umstellungsosteotomien ein wichtiger Parameter. Das American College of Radiology Data Science Institute erkannte die automatische Vermessung von Ganzbein-Röntgenbildern als „AI Use Case“ an, da gezeigt werden konnte, dass die Reproduzierbarkeit durch KI verbessert und eine hohe Genauigkeit erzielt werden kann (Abb. 4; [3, 4, 23]).

Status quo in der bildgebenden Diagnostik – Potenzial der KI für den Befundablauf

Abb. 4 ◀ IB Lab LAMA Beispiel Report. Vollautomatisierte Vermessung der Ganzbeinachse

Wie eingangs erwähnt, wird die Befundung von Röntgenbildern nach wie vor manuell durchgeführt. Mess- bzw. Befundergebnisse erfordern ein hohes Maß an Ge-

naugigkeit, welche für die Erkennung von Erkrankungen und der Zuweisung adäquater Therapien entscheidend sind. Aktuelle Abläufe führen bei der Befundung teilweise zu Übereinstimmungsraten von nur 30 % [24].

Durch den Einsatz von KI-basierter Software können Ärzt:innen im klinischen Alltag unterstützt werden. Dies ermöglicht nicht nur eine Reduzierung zeitaufwändiger Aufgaben, sondern trägt auch zur Verringerung der Arbeitslast bei, gleichzeitig jedoch zur Verbesserung der Qualität. Es ist jedoch wichtig zu betonen, dass die Resultate der derzeit verfügbaren KI-Tools nicht ohne menschliche Validierung angemessen angewendet werden kann.

Des Weiteren könnte die Neudefinierung von Normparametern mithilfe von KI realisiert werden. Neben der bereits hohen Inter- und Intra-rater-Variabilität wurden viele Normparameter vor Jahrzehnten anhand kleiner Studienpopulationen definiert und seither nicht mehr aktualisiert, wie beispielsweise die Kellgren-Lawrence-Klassifizierung. In der Ära der Digitalisierung könnten sehr umfangreiche Datensätze unter Berücksichtigung von Faktoren wie ethnischen Unterschieden mit standardisierten und reproduzierbaren Methoden analysiert werden. Dies würde es ermöglichen, Normparameter neu zu definieren und an aktuelle Gegebenheiten anzupassen.

Fazit für die Praxis

- **Machine Learning (ML), als Teilgebiet der künstlichen Intelligenz (KI), ermöglicht durch Bilderkennungsalgorithmen die Beurteilung und Vermessung von Röntgenbildern.**
- **Manuelle Messungen sind anfällig für die individuellen Einflüsse der messenden Person und äußere Umwelteinflüsse, was zu Inter- und Intra-rater-Variabilität führt.**
- **Mithilfe von KI kann die Inter- und Intra-rater-Variabilität reduziert und die Reproduzierbarkeit erhöht werden.**
- **Externe Validierung, bei der der Validierungsdatensatz grundlegend vom Trainingsdatensatz abweicht, ist entscheidend für die Bewertung der Leistungsfähigkeit und Zuverlässigkeit eines KI-Modells.**
- **Als unterstützendes Werkzeug kann KI die Arbeitslast in der diagnostischen Praxis reduzieren, indem sie routineorientierte Aufgaben übernimmt, und gleichzeitig**

die Qualität durch Standardisierung der Beurteilung verbessern.

- **Menschliche Validierung ist nach wie vor unerlässlich, um die klinische Korrektheit und Sicherheit zu gewährleisten.**

Korrespondenzadresse

Univ.-Prof. Dr. Stefan Nehrer, MSc
Fakultät für Gesundheit und Medizin,
Department für Gesundheitswissenschaften,
Medizin und Forschung, Zentrum für
Regenerative Medizin, Zentrum für
Gesundheitswissenschaften und Medizin,
Universität für Weiterbildung Krems Donau
Dr.-Karl-Dorrek-Straße 30, 3500 Krems,
Österreich
stefan.nehrer@donau-uni.ac.at

Funding. Open access funding provided by Danube University Krems University for Continuing Education.

Einhaltung ethischer Richtlinien

Interessenkonflikt. S. Nehrer und K. Chen sind Projektmitarbeiter an einem Drittmittelgeförderten Projekt (GFF NÖ) der UWK. R. Ljuhar und C. Götz sind in der Geschäftsführung von IBL.

Für diesen Beitrag wurden von den Autor/-innen keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

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

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

1. McCarthy J, Minsky ML, Rochester N, Claude E Shannon, A Proposal for the Dartmouth Summer Research Project on. Artif Intell

2. Stotter C, Klestil T, Röder C, Reuter P, Chen K, Emprechtinger R, Hummer A, Salzlechner C, DiFranco M, Nehrer S (2023) Deep Learning for Fully Automated Radiographic Measurements of the Pelvis and Hip. *Diagnosics*. <https://doi.org/10.3390/diagnostics13030497>
3. Stotter C, Klestil T, Chen K, Hummer A, Salzlechner C, Angele P, Nehrer S (2023) Artificial intelligence-based analyses of varus leg alignment and after high tibial osteotomy show high accuracy and reproducibility. *Knee surgery, sports traumatology, arthroscopy official journal of the ESSKA*, Bd. 31, 5885–5895 <https://doi.org/10.1007/s00167-023-07644-0>
4. Schock J, Truhn D, Abrar DB, Merhof D, Conrad S, Post M, Mittelstrass F, Kuhl C, Nebelung S (2021) Automated Analysis of Alignment in Long-Leg Radiographs by Using a Fully Automated Support System Based on Artificial Intelligence. *Radiology. Artif Intell* 3:e200198. <https://doi.org/10.1148/ryai.2020200198>
5. Murata K, Endo K, Aihara T, Suzuki H, Sawaji Y, Matsuoka Y, Nishimura H, Takamatsu T, Konishi T, Maekawa A, Yamauchi H, Kanazawa K, Endo H, Tsuji H, Inoue S, Fukushima N, Kikuchi H, Sato H, Yamamoto K (2020) Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. *Sci Rep* 10:20031. <https://doi.org/10.1038/s41598-020-76866-w>
6. Kalmet PHS, Sanduleanu S, Primakov S, Wu G, Jochems A, Refaee T, Ibrahim A, Hulst LV, Lambin P, Poeze M (2020) Deep learning in fracture detection: a narrative review. *Acta O* 91:215–220. <https://doi.org/10.1080/17453674.2019.1711323>
7. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, Hanel D, Gardner M, Gupta A, Hotchkiss R, Potter H (2018) Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci U S A* 115:11591–11596. <https://doi.org/10.1073/pnas.1806905115>
8. R.W. Wright, Osteoarthritis Classification Scales: Interobserver Reliability and Arthroscopic Correlation, *The Journal of bone and joint surgery. American volume* 96 (2014) 1145–1151. <https://doi.org/10.2106/JBJS.M.00929>.
9. Engh CA, Sychterz CJ, Young AM, Pollock DC, Toomey SD, Engh CA (2002) Interobserver and intraobserver variability in radiographic assessment of osteolysis. *J Arthroplasty* 17:752–759. <https://doi.org/10.1054/arth.2002.33554>
10. M. Gstoettner, K. Sekyra, N. Walochnik, P. Winter, R. Wachter, C.M. Bach, Inter- and intraobserver reliability assessment of the Cobb angle: manual versus digital measurement tools, *European spine journal official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 16 (2007) 1587–1592. <https://doi.org/10.1007/s00586-007-0401-3>.
11. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK (2018) Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med* 15:e1002683. <https://doi.org/10.1371/journal.pmed.1002683>
12. Laur O, Wang B (2022) Musculoskeletal trauma and artificial intelligence: current trends and projections. *Skelet Radiol* 51:257–269. <https://doi.org/10.1007/s00256-021-03824-6>
13. Ajmera P, Kharat A, Botchu R, Gupta H, Kulkarni V (2021) Real-world analysis of artificial intelligence in musculoskeletal trauma. *J Clin Orthop Trauma* 22:101573. <https://doi.org/10.1016/j.jcot.2021.101573>

14. CDT, Work-related MSDs: prevalence, costs and demographics in the EU—Summary.
15. Bruno MA, Walker EA, Abujudeh HH (2015) Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction, Radiographics a review publication of the Radiological Society of North America. Inc, Bd. 35, S 1668–1676 <https://doi.org/10.1148/rg.2015150023>
16. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015, Lancet (London, England) 388 (2016) 1545–1602. [https://doi.org/10.1016/S0140-6736\(16\)31678-6](https://doi.org/10.1016/S0140-6736(16)31678-6).
17. L. MURPHY, T.A. SCHWARTZ, C.G. HELMICK, J.B. RENNER, G. TUDOR, G. KOCH, A. DRAGOMIR, W.D. KALSBECK, G. LUTA, J.M. JORDAN, Lifetime risk of symptomatic knee osteoarthritis, Arthritis and Rheumatism 59 (2008) 1207–1213. <https://doi.org/10.1002/art.24021>.
18. Nehrer S, Ljuhar R, Steindl P, Simon R, Maurer D, Ljuhar D, Bertalan Z, Dimai HP, Goetz C, Paixao T (2021) Automated Knee Osteoarthritis Assessment Increases Physicians' Agreement Rate and Accuracy: Data from the Osteoarthritis Initiative. CARTILAGE 13:957S–965S. <https://doi.org/10.1177/1947603519888793>
19. Maria Anna, Smolle Christoph, Goetz Dietmar, Maurer Ines, Vielgut Michael, Novak Gerhard, Zier Andreas, Leithner Stefan, Nehrer Tiago, Paixao Richard, Ljuhar Patrick, Sadoghi (2023) Artificial intelligence-based computer-aided system for knee osteoarthritis assessment increases experienced orthopaedic surgeons' agreement rate and accuracy Abstract Knee Surgery Sports Traumatology Arthroscopy 31(3):1053–1062. <https://doi.org/10.1007/s00167-022-07220-y>
20. Horng M-H, Kuok C-P, Fu M-J, Lin C-J, Sun Y-N (2019) Cobb Angle Measurement of Spine from X-Ray Images Using Convolutional Neural Network. Comput Math Methods Med 2019:6357171. <https://doi.org/10.1155/2019/6357171>
21. Y. Pan, Q. Chen, T. Chen, H. Wang, X. Zhu, Z. Fang, Y. Lu, Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays, European spine journal official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society 28 (2019) 3035–3043. <https://doi.org/10.1007/s00586-019-06115-w>.
22. Berlin C, Adomeit S, Grover P, Dreischarf M, Halm H, Dürr O, Obid P (2023) Novel AI-Based Algorithm for the Automated Computation of Coronal Parameters in Adolescent Idiopathic Scoliosis Patients: A Validation Study. on, Bd. 100. Preoperative Full Spine X-Rays, Global spine journal. <https://doi.org/10.1177/21925682231154543>
23. Simon S, Schwarz GM, Aichmair A, Frank BJH, Hummer A, DiFranco MD, Dominkus M, Hofstaetter JG (2022) Fully automated deep learning for knee alignment assessment in lower extremity radiographs: a cross-sectional diagnostic study. Skelet Radiol 51:1249–1259. <https://doi.org/10.1007/s00256-021-03948-9>
24. Lidströmer N, Ashrafian H (Hrsg) (2020) Artificial Intelligence in Medicine. Springer, Imprint Springer, Cham

Opportunities and risks of using artificial intelligence and machine learning in imaging diagnostics

Artificial intelligence (AI) is increasingly being employed in diagnostic imaging. This is a broad term encompassing computer programs capable of undertaking and solving intelligent tasks. The evolving complexity of AI architectures enables more demanding tasks, such as the recognition and quantification of radiological parameters to be overcome with more sophistication. Currently, in the majority of cases the assessment and description of such parameters are carried out manually and in a narrative form. This manual evaluation process is not only time-consuming but also susceptible to interrater and intrarater variability, as it is strongly influenced by the person doing the assessment and external factors. Using AI-algorithms, standardized and reproducible results can be generated as it can exactly evaluate information from imaging data down to the individual pixels, independent of external influences. A decisive advantage is that in contrast to manual assessment, AI has the capability to incorporate extensive background data into the evaluation, which leads to a further enhancement of the precision. Functioning as a supportive tool, AI can elevate the quality of X-ray image assessment while concurrently alleviating the workload.

Keywords

X-ray imaging · Radiological parameters · Machine Learning · Deep Learning · Validation

Hinweis des Verlags. Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.