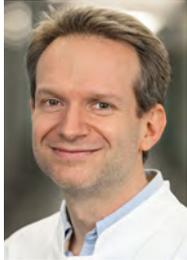


# Ultrasound Diagnosis of Hepatocellular Carcinoma: Is the Future Defined by Artificial Intelligence?

Ultraschalldiagnostik des hepatozellulären Karzinoms: In Zukunft nur mit künstlicher Intelligenz?



Prof. Maximilian J Waldner  
[rerif]



Prof. Deike Strobel

## Correspondence

Prof. M. Waldner  
Medical Clinic 1, Erlangen University Hospital, Ulmenweg 18,  
91054 Erlangen, Germany  
Tel.: +49/0 91 31/8 53 50 00  
maximilian.waldner@uk-erlangen.de

## Bibliography

Ultraschall in Med 2024; 45: 8–12  
DOI 10.1055/a-2171-2674  
ISSN 0172-4614  
© 2024, Thieme. All rights reserved.  
Georg Thieme Verlag KG, Rüdigerstraße 14,  
70469 Stuttgart, Germany

## Introduction

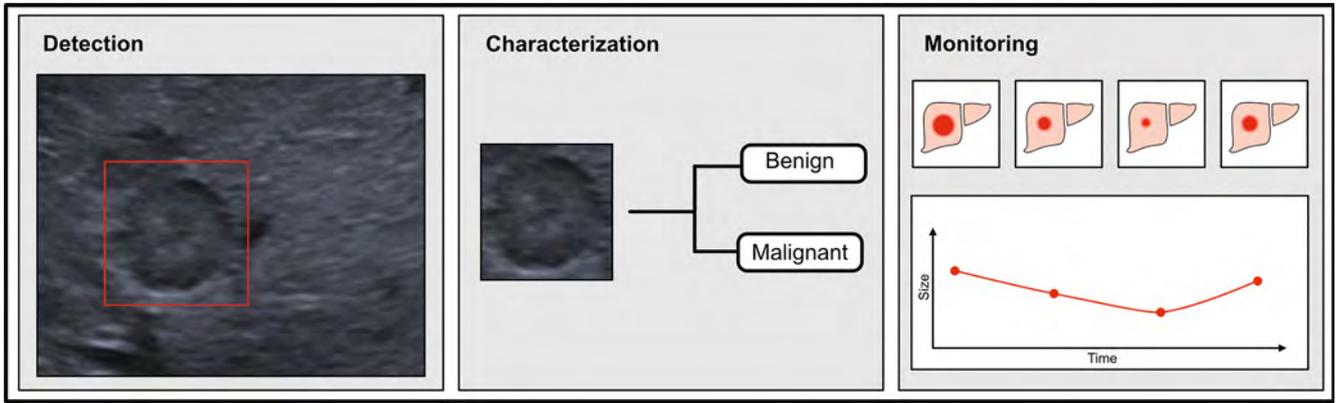
Technological progress and the development of complex mathematical models that allow the analysis of large and partially unstructured data have led to the rapid development of artificial intelligence (AI) since the 2010s [1]. New AI applications, such as the recent “chatbot” ChatGPT, regularly attract a great deal of media attention with headlines ranging from euphoric to critical. As a result, the population developed specific expectations of the benefits, but also concerns about the potential risks of AI. In a survey published in 2023 by the digital association Bitkom, 73% of the 1007 people surveyed saw AI as an opportunity [2]. Two thirds wanted AI to be used when it would bring specific benefits, for example in medicine or transportation. 14% and 10% of respondents saw AI rather or exclusively as a risk, respectively. The majority of respondents assumed that AI would noticeably change our society in the coming years. These survey results impressively show how much is already expected of AI. And indeed, AI accompanies us consciously or unconsciously in many everyday situations. There are also several examples in clinical medicine, e. g., in the automated evaluation of ECGs, differential blood counts, etc. [1]. Intensive research has been carried out into the possible use of AI in medical imaging since its beginnings over 80 years ago. In addition to approaches for image optimization, this primarily includes automated diagnosis for disease detection and classification as well as therapeutic monitoring. Enormous progress has been made in the field of imaging in recent years through the use of deep learning (DL) technologies. In contrast to classic forms of machine learning (ML), DL is based on neural networks in which several network levels are linked together [3]. Convolutional neural

networks (CNNs) are frequently used in the field of image recognition. These are characterized by a hierarchical recognition of image patterns by the different network levels [3]. If initial structures such as corners, edges, or simple shapes are recognized, the linking of these simple structures in the deeper network levels enables the classification of complex structures such as malignancies in clinical imaging. In addition to better predictability, CNNs are more flexible than traditional ML. Furthermore, the time-consuming extraction of diagnostically relevant image information (“feature extraction”), which is necessary with classic ML, is no longer required, as image features are recognized independently by CNNs.

Particularly in radiological imaging (including computed tomography and magnetic resonance imaging), CNNs have achieved outstanding results in clinical diagnostics and therapeutic monitoring. These have already led to various commercially available and approved AI applications in the field of oncological imaging, among others [4] (► Fig. 1).

Compared to radiological imaging, the use of AI in sonographic imaging involves particular challenges [6]. Due to the examiner dependency in sonographic image acquisition, possible image material for training AI generally has a higher variability. The representation of identical findings in different scanning planes can, for example, lead to considerable differences in image interpretation and thus make correct classification by the AI more difficult. Similar effects can be caused by differences in the devices or transducers used.

In the following, we will discuss the extent to which these effects impact the use of AI in sonography, where AI currently stands in sonography, what limitations are to be expected, and



► **Fig. 1** Use of artificial intelligence (AI) in oncological imaging. Previous studies have investigated the possible use of AI in the detection, characterization, and therapeutic monitoring of oncological diseases. Illustration modified according to El Naqa et al., Br J Radiol, 2020 [5].

► **Abb. 1** Einsatz der Künstlichen Intelligenz (KI) in der onkologischen Bildgebung. Bisherige Studien untersuchten einen möglichen Einsatz der KI bei der Erkennung (Detection), Charakterisierung (Characterization) und beim therapeutischen Monitoring onkologischer Erkrankungen. Abbildung modifiziert nach El Naqa et al., Br J Radiol, 2020 [5].

what steps are needed next to enable the successful translation of AI into clinical sonography, using the example of the sonographic diagnosis of hepatocellular carcinoma (HCC).

## AI for early screening and characterization of focal liver lesions in sonographic imaging

The incidence of malignant liver tumors has increased continuously in recent decades. Primary hepatocellular carcinoma (HCC) is the most common malignant liver tumor worldwide, in the western world mainly due to liver cirrhosis related to chronic HCV infection or alcohol and the increase in metabolic dysfunction-associated steatotic liver disease (MASLD) with advanced fibrosis or cirrhosis of the liver [7, 8]. For this reason, the various professional societies recommend that patients at increased risk of HCC participate in an early detection program based on biannual ultrasound examinations of the liver [8, 9].

The sensitivity of B-scan ultrasonography for the early screening of HCC is reported to be between 47 and 84% [10]. The reasons for the great variability in sensitivity are the experience of the examiner, inadequate acoustic conditions in very obese patients, or the inhomogeneous echotexture of cirrhosis of the liver. While sonography has a comparable detection rate to computed tomography (CT) and magnetic resonance imaging (MRI) for larger lesions, the sensitivity of sonography is significantly lower than MRI when it comes to detecting small lesions (<2 cm). The question arises as to whether the use of AI-supported procedures in sonography can lead to improved detection of early forms of HCC.

Studies published to date have primarily investigated the possibility of using AI to detect focal liver lesions in sonographic image data. For example, Tiyyarattanachai trained a CNN with more than 20 000 individual images from the B-scan ultrasonography of almost 3500 patients [11]. The authors achieved a sensitivity of 83.9% for the detection of focal liver lesions in an internal vali-

dation dataset and 84.9% in an external validation dataset. Yang et al. achieved a sensitivity of 86.5% and a specificity of 85.5% with a CNN that was trained using over 20 000 ultrasound images from more than 2000 patients in conjunction with clinical information (including age, gender, AFP value) [12]. In this study, the CNN was superior to experienced examiners and achieved similar results to contrast-enhanced computed tomography.

If a suspected HCC lesion is detected during surveillance, it should be characterized using a contrast media-based procedure in accordance with the guideline recommendations [8]. In addition to contrast media-enhanced magnetic resonance imaging (MRI) and computed tomography (CT), contrast medium-enhanced ultrasound (CEUS) can also be used for HCC diagnosis [13]. As the data of the prospective DEGUM multicenter study show, a typical perfusion pattern in contrast media ultrasonography (arterial hypervascularization and washout in the portal venous and venous phase) allows the diagnosis of HCC with a sensitivity of 94% and a specificity of 65% (or 79% when using standardized CEUS algorithms) [8, 14, 15]. However, similar results in everyday clinical practice require experienced examiners. It is therefore not surprising that the use of AI to characterize focal liver lesions and to diagnose HCC has also been investigated in various studies. In a recently published summary of the review, the AI-supported characterization of focal liver lesions in some of the studies published to date was based only on B-scan data, in some cases with CEUS data [16]. The diagnostic accuracy of B-scan-trained AI was between 69 and 98.6%, while the diagnostic accuracy of CEUS-trained AI was between 64 and 98.3%. Only a small proportion of the studies published to date have compared the results of AI-assisted sonography with medical assessment [16]. Primarily, however, the studies to date indicate that AI-based classification is comparable to experienced examiners, but can achieve better results than inexperienced examiners. Despite this positive assessment, however, the data to date must also be critically scrutinized.

## Limitations of the clinical use of AI in HCC sonography

The systematic assessment of the scientific quality of a total of 52 studies on the characterization of focal liver lesions in sonographic datasets using the QUADAS-2 criteria showed that the transferability of the results of many studies is limited [16]. This is primarily due to the fact that no independent datasets were used in the studies for final testing (so-called test dataset) of the fully trained CNN. In addition, not all of the most common types of focal liver lesions were included in some studies, which further limits the use of the AI algorithms used in everyday clinical practice.

A known risk when training neural networks is “bias”, i. e., data distortion, e. g., due to the lack of inclusion of certain patient groups. In the aforementioned QUADAS-2 analysis, it was not possible to make a statement on the risk due to bias in many studies, as relevant information was missing in the description of the methodology used in individual studies [16]. In order to check the bias of the AI methods, an understanding of the underlying criteria for image classification would be necessary. As it is generally not possible to check the algorithms of the CNNs due to the high complexity of the systems, careful planning of the data used during training is essential to avoid bias. This requires not only appropriate expertise in conducting AI-based studies, but also among the reviewers and readers in order to be able to adequately classify the significance and quality of the studies. Kuang et al. summarized various quality criteria for the use of AI in sonography, which should also help readers who are inexperienced with AI to evaluate the relevant studies [17]. Here, too, the urgent need for independent test datasets (ideally external datasets) was pointed out. In addition to other points, care should be taken to ensure that the AI algorithms are freely accessible and thus visible, that the performance of the AI is compared with the results of experienced examiners, and that the results are compared with data from comparable published studies.

## Conclusion

Even if the majority of the data published to date on the use of AI in the sonographic diagnosis of HCC is based on a retrospective analysis of previously acquired image data and cannot easily be transferred to a real-time assessment of the liver in everyday clinical practice, the data to date are already promising.

Similar to ultrasound of HCC, data on the use of AI in sonography are already available for numerous applications, albeit often with similar limitations to the studies discussed here. A frequent problem is the transferability to the very heterogeneous situation of everyday clinical diagnostics (e. g., possible influence of different devices, presets, transducers, examiners, etc.). Due to the high number of possible influencing factors in sonography, correspondingly larger datasets from different centers are required to counteract the heterogeneity of the image data. However, a targeted and critically questioned use of AI with regard to its added value, together with careful planning and a multicenter collection of training, validation, and test datasets, could provide valuable support in the future, especially for inexperienced examiners.

## Ultraschalldiagnostik des hepatozellulären Karzinoms: In Zukunft nur mit künstlicher Intelligenz?

### Einleitung

Technologischer Fortschritt und die Entwicklung komplexer mathematischer Modelle, welche die Analyse großer und teilweise unstrukturierter Daten erlauben, führten seit den 2010er-Jahren zu einer rasanten Weiterentwicklung der künstlichen Intelligenz (KI) [1]. Regelmäßig bekommen neue KI-Anwendungen, wie zuletzt der „Chatbot“ ChatGPT, eine hohe mediale Aufmerksamkeit – mit euphorischen bis hin zu kritischen Schlagzeilen. Entsprechend entwickelten sich in der Bevölkerung konkrete Erwartungen an den Nutzen, aber auch Bedenken aufgrund möglicher Gefahren der KI. In einer 2023 veröffentlichten Umfrage des Digitalverbands Bitkom sahen 73 % der 1007 befragten Personen KI als Chance [2]. Zwei Drittel wünschten sich den Einsatz der KI, wenn er einen konkreten Nutzen bringt, etwa in der Medizin oder im Verkehr. Jeweils 14 % und 10 % der Befragten sahen KI eher oder ausschließlich als Gefahr. Überwiegend gingen die Befragten davon aus, dass KI unsere Gesellschaft in den kommenden Jahren spürbar verändern werde. Diese Umfragewerte zeigen eindrucksvoll, wieviel der KI bereits zugetraut wird. Und tatsächlich begleitet uns KI bereits bewusst oder unbewusst in vielen Alltagssituationen. Entsprechende Beispiele gibt es auch in der klinischen Medizin, z. B. bei der automatisierten Auswertung von EKGs, Differentialblutbildern etc. [1]. Auch an einem möglichen Einsatz der KI in der medizinischen Bildgebung wurde seit deren Anfängen vor über 80 Jahren intensiv geforscht. Dazu zählt neben Ansätzen zur Bildoptimierung in erster Linie die automatisierte Befunderhebung zur Krankheitserkennung und -klassifizierung sowie zum therapeutischen Monitoring. Durch den Einsatz sogenannter Deep-Learning-Technologien (kurz: DL-Technologien) wurden im Bereich der Bildgebung in den vergangenen Jahren enorme Fortschritte erzielt. Im Gegensatz zu klassischen Formen des maschinellen Lernens (ML) basiert DL auf neuronalen Netzen, bei denen mehrere Netzwerkebenen miteinander verknüpft werden [3]. Für einen Einsatz bei der Bilderkennung werden häufig Convolutional Neuronal Networks (CNNs) eingesetzt. Diese zeichnen sich durch eine hierarchische Erkennung von Bildmustern durch die verschiedenen Netzwerkebenen aus [3]. Werden initial Strukturen wie Ecken, Kanten oder einfache Formen erkannt, ermöglicht die Verknüpfung dieser einfachen Strukturen in den tieferen Netzwerkebenen die Klassifizierung komplexer Strukturen, wie z. B. von Malignomen, in der klinischen Bildgebung. Neben einer besseren Vorhersagekraft besitzen CNNs im Vergleich zum klassischen ML eine höhere Flexibilität. Zusätzlich entfällt die beim klassischen ML notwendige und zeitaufwendige Extrahierung diagnostisch relevanter Bildinformationen („feature extraction“), da diese Bildmerkmale durch CNNs selbstständig erkannt werden.

Besonders in der radiologischen Bildgebung (u. a. in der Computertomografie und der Magnetresonanztomografie) konnten mithilfe von CNNs herausragende Ergebnisse in der klinischen

Diagnostik und beim Therapie-Monitoring erzielt werden. Diese führten bereits zu verschiedenen kommerziell verfügbaren und zugelassenen KI-Applikationen, u. a. im Bereich der onkologischen Bildgebung [4].

Im Vergleich zur radiologischen Bildgebung bringt der Einsatz der KI in der sonografischen Bildgebung besondere Herausforderungen mit sich [6]. Aufgrund der Untersucherabhängigkeit bei der Bilderfassung in der Sonografie verfügt mögliches Bildmaterial für das Training einer KI in der Regel über eine höhere Variabilität. Die Darstellung des identischen Befundes in unterschiedlichen Schnittebenen kann z. B. bereits zu erheblichen Unterschieden bei der Bildinterpretation führen und damit eine korrekte Klassifizierung durch die KI erschweren. Ähnliche Effekte können durch Unterschiede der verwendeten Geräte oder Schallköpfe hervorgerufen werden.

Inwiefern sich diese Effekte auf den Einsatz der KI in der Sonografie auswirken, wo die KI in der Sonografie aktuell steht, mit welchen Einschränkungen zu rechnen ist und welche Schritte als nächstes nötig sind, um eine erfolgreiche Translation der KI in die klinische Sonografie zu ermöglichen, werden wir im Folgenden am Beispiel der sonografischen Diagnostik des hepatozellulären Karzinoms (HCC) diskutieren.

## KI zur Früherkennung und Charakterisierung fokaler Leberläsionen in der sonografischen Bildgebung

Die Inzidenz maligner Lebertumore hat in den vergangenen Jahrzehnten kontinuierlich zugenommen. Das primäre Leberzellkarzinom (HCC) ist hierbei der weltweit häufigste maligne Lebertumor, in der westlichen Welt vor allem aufgrund chronischer HCV-Infektion-bedingter oder alkoholbedingter Leberzirrhosen und der Zunahme der mit einer metabolischen Dysfunktion assoziierten Fettleber-Erkrankungen (MASLD), mit fortgeschrittener Fibrose oder Zirrhose [7, 8]. Aus diesem Grund empfehlen die verschiedenen Fachgesellschaften Patienten mit erhöhtem HCC-Risiko die Teilnahme an einem Früherkennungsprogramm, basierend auf halbjährlichen Ultraschall-Untersuchungen der Leber [8, 9].

Die Sensitivität der B-Bild-Sonografie für die HCC-Früherkennung wird mit Werten zwischen 47 und 84% angegeben [10]. Als Gründe für die große Variabilität der Sensitivität gelten die Erfahrung des Untersuchers, unzureichende Schallbedingungen bei sehr adipösen Patienten oder die inhomogene Echotextur der Leberzirrhose. Während die Sonografie bei größeren Läsionen eine vergleichbare Detektionsrate wie die Computertomografie (CT) und die Magnetresonanztomografie (MRT) aufweist, liegt die Sensitivität der Sonografie deutlich unter der MRT, wenn es um die Erkennung kleiner Läsionen (< 2 cm) geht. Es stellt sich die Frage, ob der Einsatz KI-gestützter Verfahren in der Sonografie zu einer verbesserten Detektion von HCC-Frühhformen führen kann.

Bislang veröffentlichte Studien untersuchten in erster Linie die Möglichkeit, mithilfe der KI fokale Leberläsionen in sonografischen Bilddaten zu erkennen. So trainierten Tiyyarattanachai et al. ein CNN mit mehr als 20 000 Einzelbildern aus den B-Bild-Sonografien von knapp 3500 Patienten [11]. Die Autoren erzielten dabei eine Sensitivität von 83,9% für die Erkennung fokaler Leber-

läsionen in einem internen und 84,9% in einem externen Validierungsdatensatz. Yang et al. erzielten mit einem CNN, welches Anhand von über 20 000 Ultraschallbildern von mehr als 2000 Patienten, in Verbindung mit klinischen Informationen (u. a. Alter, Geschlecht, AFP-Wert) trainiert wurde, eine Sensitivität von 86,5% und eine Spezifität von 85,5% [12]. Das CNN war in dieser Studie gegenüber erfahrenen Untersuchern überlegen und erzielte ähnliche Ergebnisse wie eine kontrastmittelverstärkte Computertomografie.

Wird eine HCC-suspekte Läsion im Rahmen der Surveillance erkannt, sollte diese nach den Leitlinien-Empfehlungen durch ein kontrastmittelgestütztes Verfahren charakterisiert werden [8]. Zur HCC-Diagnostik kann neben der kontrastmittelverstärkten Magnetresonanztomografie (MRT) und der Computertomografie (CT) auch die kontrastmittelverstärkte Sonografie (CEUS) eingesetzt werden [13]. Wie die Daten der prospektiven DEGUM-Multi-center-Studie zeigen, erlaubt ein typisches Perfusionsmuster in der Kontrastmittel-Sonografie (arterielle Hypervaskularisation und Auswaschen in der portalvenösen und venösen Phase) die Diagnose eines HCCs mit einer Sensitivität von 94% und einer Spezifität von 65% (bzw. 79% bei Verwendung standardisierter CEUS-Algorithmen), [8, 14, 15]. Ähnliche Ergebnisse im klinischen Alltag setzen allerdings erfahrene Untersucher voraus. Somit ist es nicht verwunderlich, dass auch der Einsatz der KI zur Charakterisierung fokaler Leberläsionen und zur Diagnose eines HCCs in verschiedenen Studien untersucht wurde. Wie in einer kürzlich veröffentlichten, systematischen Übersichtsarbeit zusammengefasst, erfolgte die KI-gestützte Charakterisierung fokaler Leberläsionen in einem Teil der bislang publizierten Studien nur mithilfe von B-Bild-Daten, teilweise mit CEUS-Daten [16]. Die diagnostische Genauigkeit einer B-Bild-trainierten KI lag zwischen 69 und 98,6%, bei einer CEUS-trainierten KI lag die diagnostische Genauigkeit zwischen 64 und 98,3%. Nur ein kleiner Teil der bisher veröffentlichten Studien verglich die Ergebnisse der KI-gestützten Sonografie mit der ärztlichen Beurteilung [16]. In erster Linie weisen die bisherigen Studien jedoch darauf hin, dass eine KI-basierte Klassifizierung mit erfahrenen Untersuchern vergleichbar ist, aber bessere Ergebnisse als unerfahrene Untersucher erzielen kann. Trotz dieser positiven Einschätzung müssen die bisherigen Daten jedoch auch kritisch hinterfragt werden.

## Limitationen eines klinischen Einsatzes der KI bei der Sonografie des HCCs

Die systematische Beurteilung der wissenschaftlichen Qualität von insgesamt 52 Studien zur Charakterisierung fokaler Leberläsionen in sonografischen Datensätzen mithilfe der QUADAS-2-Kriterien zeigte, dass bei vielen Studien nur eine eingeschränkte Übertragbarkeit der Ergebnisse gegeben ist [16]. In erster Linie liegt das daran, dass in den Arbeiten keine unabhängigen Datensätze zur finalen Testung (sog. Test-Datensatz) des fertig trainierten CNNs verwendet wurden. Zusätzlich wurden in einigen Studien nicht alle typischen Vertreter fokaler Leberläsionen eingeschlossen, wodurch der Einsatz der verwendeten KI-Algorithmen im klinischen Alltag weiter eingeschränkt wird.

Eine bekannte Gefahr beim Training neuronaler Netzwerke ist der „Bias“, also eine Datenverzerrung, z. B. durch den fehlenden Einschluss bestimmter Patientengruppen. In der genannten QUADAS-2-Analyse ließ sich bei vielen Studien keine Aussage zum Risiko durch Bias treffen, da relevante Informationen in der Beschreibung der verwendeten Methodik in einzelnen Arbeiten fehlen [16]. Um ein Bias der KI-Methoden zu überprüfen, wäre grundsätzlich ein Verständnis der zugrunde liegenden Kriterien für die Bild-Klassifizierung nötig. Da eine Überprüfung der Algorithmen der CNNs aufgrund der hohen Komplexität der Systeme in der Regel nicht möglich ist, ist zur Vermeidung eines Bias eine sorgfältige Planung der beim Training eingesetzten Daten unabdingbar. Dies setzt nicht nur eine entsprechende Expertise bei der Durchführung von KI-basierten Studien, sondern auch bei den Gutachtern und Lesern voraus, um die Aussagekraft und Qualität der Arbeiten adäquat einordnen zu können. Für den Einsatz der KI in der Sonografie fassten Kuang et al. verschiedene Aspekte zusammen, die auch KI-unerfahrenen Lesern eine Hilfestellung bei der Bewertung entsprechender Studien geben sollen [17]. Auch hier wurde auf die dringende Notwendigkeit unabhängiger Test-Datensätze (optimalerweise externer Datensätze) hingewiesen. Neben weiteren Punkten sollte darauf geachtet werden, dass die KI-Algorithmen frei zugänglich und damit einsehbar sind, dass die Leistung der KI mit dem Untersuchungsergebnis erfahrener Untersucher verglichen wird, und dass die Ergebnisse mit Daten aus vergleichbaren, publizierten Studien verglichen werden.

## Schlussfolgerung

Auch wenn der Großteil der bisher veröffentlichten Daten zum Einsatz der KI bei der sonografischen Diagnostik des HCCs auf einer retrospektiven Analyse zuvor erhobener Bilddaten basiert und nicht ohne Weiteres auf eine Echtzeit-Beurteilung der Leber im klinischen Alltag übertragbar ist, sind die bisherigen Daten bereits vielversprechend.

Ähnliche Daten zum Einsatz der KI in der Sonografie gibt es bereits für zahlreiche Anwendungen, allerdings mit ähnlichen Einschränkungen wie bei den hier diskutierten Studien. Eine häufige Problematik stellt die Übertragbarkeit auf die sehr heterogene Situation der klinischen Alltagsdiagnostik dar (u. a. ein möglicher Einfluss unterschiedlicher Geräte, Presets, Schallköpfe, Untersucher etc.). Aufgrund der hohen Zahl möglicher Einflussfaktoren bei der Sonografie sind entsprechend größere Datensätze von verschiedenen Zentren nötig, um der Heterogenität der Bilddaten entgegenzuwirken. Ein gezielter und bzgl. des Mehrwerts kritisch hinterfragter Einsatz der KI, zusammen mit einer sorgfältigen Planung und einer multizentrischen Erhebung von Trainings-, Validierungs- und Testdatensätzen, könnte in Zukunft jedoch – besonders für unerfahrene Untersucher – eine wertvolle Unterstützung darstellen.

## Conflict of Interest

The authors declare that they have no conflict of interest.

## References

- [1] Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine. *N Engl J Med* 2023; 388 (13): 1201–1208. doi:10.1056/NEJMr2302038
- [2] Streim A, Beerlink KP. Drei Viertel sehen Künstliche Intelligenz als Chance. Presseinformation [Web Page] 2023 [cited 2023 11/12/2023]; Available from: <https://www.bitkom.org/Presse/Presseinformation/Drei-Viertel-sehen-Kuenstliche-Intelligenz-Chance>
- [3] Kim M, Yun J, Cho Y et al. Deep Learning in Medical Imaging. *Neurospine* 2019; 16 (4): 657–668. doi:10.14245/ns.1938396.198
- [4] Koh DM, Papanikolaou N, Bick U et al. Artificial intelligence and machine learning in cancer imaging. *Commun Med (Lond)* 2022; 2: 133. doi:10.1038/s43856-022-00199-0
- [5] El Naqa I, Haider MA, Giger ML et al. Artificial Intelligence: reshaping the practice of radiological sciences in the 21st century. *Br J Radiol* 2020; 93: 20190855. doi:10.1259/bjr.20190855
- [6] Park SH. Artificial intelligence for ultrasonography: unique opportunities and challenges. *Ultrasonography* 2021; 40 (1): 3–6. doi:10.14366/usg.20078
- [7] Rinella ME, Lazarus JV, Ratziu V et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Ann Hepatol* 2023; 29 (1): 101133. doi:10.1016/j.aohp.2023.101133
- [8] Gross S, Bitzer M, Blodt S et al. Leitlinienreport der S3-Leitlinie Diagnostik und Therapie des Hepatozellulären Karzinoms und biliärer Karzinome – Version 3.0 – Juli 2022, AWMF-Registernummer: 032/053OL. *Z Gastroenterol* 2023; 61 (4): e157–e171. doi:10.1055/a-1975-0570
- [9] European Association for the Study of the Liver. Electronic address, e.e.e. and L. European Association for the Study of the, EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* 2018; 69 (1): 182–236. doi:10.1016/j.jhep.2018.03.019
- [10] Dietrich CF, Teufel A, Sirlin CB et al. Surveillance of hepatocellular carcinoma by medical imaging. *Quant Imaging Med Surg* 2019; 9 (11): 1904–1910. doi:10.21037/qims.2019.10.04
- [11] Tiyyarattanachai T, Apiparakoon T, Marukat S et al. Development and validation of artificial intelligence to detect and diagnose liver lesions from ultrasound images. *PLoS One* 2021; 16 (6): e0252882. doi:10.1371/journal.pone.0252882
- [12] Yang Q, Wei J, Hao X et al. Improving B-mode ultrasound diagnostic performance for focal liver lesions using deep learning: A multicentre study. *EBioMedicine* 2020; 56: 102777. doi:10.1016/j.ebiom.2020.102777
- [13] Strobel D. Using contrast-enhanced ultrasound in the diagnosis of hepatocellular carcinoma – what we have and have not achieved. *Ultraschall in Med* 2021; 42 (2): 120–124. doi:10.1055/a-1347-1388
- [14] Schellhaas B, Bernatik T, Dirks K et al. Contrast-Enhanced Ultrasound Patterns for the Non-invasive Diagnosis of Hepatocellular Carcinoma: A Prospective Multicenter Study in Histologically Proven Liver Lesions in a Real-Life Setting Demonstrating the Benefit of Extended Late Phase Observation. *Ultrasound Med Biol* 2021; 47 (11): 3170–3180. doi:10.1016/j.ultrasmedbio.2021.07.010
- [15] Schellhaas B, Bernatik T, Bohle W et al. Contrast-Enhanced Ultrasound Algorithms (CEUS-LIRADS/ESCALAP) for the Noninvasive Diagnosis of Hepatocellular Carcinoma – A Prospective Multicenter DEGUM Study. *Ultraschall in Med* 2021; 42 (2): 178–186. doi:10.1055/a-1198-4874
- [16] Vetter M, Waldner MJ, Zundler S et al. Artificial intelligence for the classification of focal liver lesions in ultrasound – a systematic review. *Ultraschall in Med* 2023. doi:10.1055/a-2066-9372
- [17] Kuang M, Hu HT, Li W et al. Articles That Use Artificial Intelligence for Ultrasound: A Reader’s Guide. *Front Oncol* 2021; 11: 631813. doi:10.3389/fonc.2021.631813