🌀 Thieme

# Artificial intelligence for the classification of focal liver lesions in ultrasound – a systematic review

## Künstliche Intelligenz zur Klassifikation fokaler Leberläsionen im Ultraschall – eine systematische Übersichtsarbeit

Authors
Marcel Vetter[1] 🆔, Maximilian J Waldner[1], Sebastian Zundler[1], Daniel Klett[1], Thomas Bocklitz[2, 3], Markus F Neurath[1], Werner Adler[4], Daniel Jesper[1]

Affiliations
1 Department of Internal Medicine 1, Erlangen University Hospital Department of Medicine 1 Gastroenterology Endocrinology and Pneumology, Erlangen, Germany
2 Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich-Schiller-Universitat Jena, Jena, Germany
3 Leibniz-Institute of Photonic Technology, Friedrich Schiller University Jena, Jena, Germany
4 Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

Correspondence
Dr. Daniel Jesper
Department of Internal Medicine 1, Erlangen University Hospital Department of Medicine 1 Gastroenterology Endocrinology and Pneumology, Ulmenweg 18, 91054 Erlangen, Germany
Daniel.jesper@uk-erlangen.de

🌐 Additional material is available at https://doi.org/10.1055/a-2066-9372.

## ABSTRACT

Focal liver lesions are detected in about 15 % of abdominal ultrasound examinations. The diagnosis of frequent benign lesions can be determined reliably based on the characteristic B-mode appearance of cysts, hemangiomas, or typical focal fatty changes. In the case of focal liver lesions which remain unclear on B-mode ultrasound, contrast-enhanced ultrasound (CEUS) increases diagnostic accuracy for the distinction between benign and malignant liver lesions. Artificial intelligence describes applications that try to emulate human intelligence, at least in subfields such as the classification of images. Since ultrasound is considered to be a particularly examiner-dependent technique, the application of artificial intelligence could be an interesting approach for an objective and accurate diagnosis. In this systematic review we analyzed how artificial intelligence can be used to classify the benign or malignant nature and entity of focal liver lesions on the basis of B-mode or CEUS data. In a structured search on Scopus, Web of Science, PubMed, and IEEE, we found 52 studies that met the inclusion criteria. Studies showed good diagnostic performance for both the classification as benign or malignant and the differentiation of individual tumor entities. The results could be improved by inclusion of clinical parameters and were comparable to those of experienced investigators in terms of diagnostic accuracy. However, due to the limited spectrum of lesions included in the studies and a lack of independent validation cohorts, the transfer of the results into clinical practice is limited.

## ZUSAMMENFASSUNG

Raumforderungen der Leber finden sich in ca. 15 % abdominaler Ultraschalluntersuchungen. Die Diagnose der meisten benignen Läsionen kann bei charakteristischem Befund, z. B. bei Zysten, Hämangiomen und typischen fokalen Fettverteilungsvarianten, oft bereits in der B-Bild-Sonographie zuverlässig gestellt werden. Bei unklaren Befunden erhöht der Einsatz der Kontrastmittelsonographie (CEUS) die diagnostische Treffsicherheit bezüglich der Unterscheidung von benignen und malignen Leberläsionen. Künstliche Intelligenz beschreibt Anwendungen, die versuchen, in Bereichen wie der Klassifikation von Bildern humane Intelligenz nachzubilden. Da die Sonographie als besonders untersucherabhängig gilt, könnte die

Anwendung künstlicher Intelligenz ein interessanter Ansatz für eine objektive und treffsichere Diagnose sein. Im Rahmen dieser systematischen Übersichtsarbeit haben wir analysiert, wie gut künstliche Intelligenz die Dignität und Entität von Leberläsionen auf Basis von B-Bild- oder CEUS-Daten bestimmen kann. Basierend auf einer strukturierten Recherche in Scopus, Web of Science, PubMed und IEEE fanden wir 52 Studien, welche die Einschlusskriterien erfüllten. Es zeigte sich eine gute diagnostische Genauigkeit sowohl für die Differenzier- ung der Dignität als auch der verschiedener Tumorentitäten. Die Ergebnisse ließen sich durch die zusätzliche Berücksichti- gung klinischer Parameter verbessern und waren bezüglich der diagnostischen Genauigkeit mit der erfahrener Untersu- cher vergleichbar. Aufgrund des begrenzten Spektrums un- tersuchter Läsionen und häufig fehlender unabhängiger Vali- dierungskohorten sind die Ergebnisse jedoch nur begrenzt auf die klinische Anwendung übertragbar.

## Introduction

Focal liver lesions (FLLs) are a common finding in abdominal ultra- sound examinations with a prevalence of approximately 15 % [1]. As incidental findings, focal liver lesions are frequently benign (e. g., cysts, hemangiomas). Sonography is particularly well suited for the classification of focal liver lesions (FLLs) due to its wide availability, lack of invasiveness, and low cost [2]. B-mode, in com- bination with Doppler sonography, is sufficient for a definitive di- agnosis in lesions like cysts, hemangiomas, and focal fatty chang- es in the non-cirrhotic liver. In unclear cases, contrast-enhanced ultrasound (CEUS) has a high diagnostic value (diagnostic accura- cy: 90 %, sensitivity: 92–95 % specificity: 83–90 %) to correctly classify tumor dignity[2, 3, 4]. Nevertheless, in some cases, the assessment of malignancy or specific tumor entity is not possible. In unclear cases biopsy is carried out to provide a final diagnosis based on histology. Although imaged-guided biopsy is a low-risk tool, as an invasive procedure it might be accompanied by pain, bleeding, infection, or injury to other organs [5].

Artificial intelligence (AI) generally describes computational methods that emulate human intelligence, at least in partial areas, such as decision-making. Machine learning is a subfield of AI in which a program is designed to learn from experience using train- ing data. On a research level, this process has already been eval- uated in a variety of medical fields (e. g., detection of polyps dur- ing colonoscopy) [6, 7]. Support vector machines (SVM) and artificial neural networks (ANN) are machine learning methods that can be applied to evaluate image data. In detail, an SVM is a mathematical method for dividing a set of objects into classes by maximizing margins between groups. ANNs use a structure that is similar to biological neural networks to classify data. Deep learn- ing (DL) represents a subfield of ANN-based machine learning with complex neural network architectures with multiple layers of artificial neurons. Large amounts of data are used to train DL- algorithms and in the case of image data, feature extraction is of- ten implicitly done by the DL-network. DL is regarded as the cur- rent state-of-the-art approach for AI-based image analysis.

AI could potentially improve the assessment of FLL dignity and entity by sonography. In such a scenario, the investigator could benefit from a more objective AI assessment and the comparabil- ity of results might be better. AI algorithms with good discrimina- tory ability regarding entities that are difficult to distinguish for humans would be particularly helpful.

In recent years, several papers have been published on this to- pic [6]. In this systematic review, we summarize the available data on the assessment of dignity and entity of FLLs by B-mode and CEUS using AI methods. The diagnostic value of those methods is discussed from a clinician's perspective. A special focus was placed on diagnostic accuracies and whether these can be im- proved by adding clinical parameters. In addition, comparative data between AI-based approaches and physicians was collected and assessed. Although individual reviews on this topic already ex- ist, our study represents the first systematic and most compre- hensive review [6, 7].

## Methods

### Search strategy

For this systematic review, articles on the characterization of FLLs by sonography were selected in the Scopus, Web of Science, PubMed, and IEEE databases. The literature search was conducted on 12/31/2021 according to the a priori defined search criteria. The following search terms were used: "artificial intelligence", "ma- chine learning", "neural network", "deep learning", "computer-as- sisted", "computer-aided", "ultrasound", "sonograph", "ultrasono- graphy", "liver", "hepat", "lesion", "tumor", "carcinoma", "mass", "focal." The detailed and complete search terms can be found in the supplemental data. The inclusion and exclusion criteria were determined a priori. Only articles from the years 2000 to 2021 were considered, as older articles mostly used algorithms that are outdated from today's perspective. Only articles that addressed liv- er tumor classification and/or diagnosis of a specific liver tumor entity either by B-mode and/or CEUS using artificial intelligence in humans were considered. Articles that did not report the diagnos- tic accuracy of AI-based classification of images were excluded. In addition, only original English language full-text articles or congress contributions with sufficient information were included.

### Data extraction

Two authors (MV and DJ) independently performed the data ex- traction and quality assessment. Any disagreements were dis- cussed and clarified in consensus with a third author. The extrac- ted data included authors, title, year of publication, study design (mono- or multicentric), number of cases and mode of ultra- sound (B-mode and/or CEUS). Diagnostic accuracy, sensitivity, specificity, and AUC (area under the curve) for lesion dignity

and/or specific tumor entities were recorded. Regarding AI, information on the extracted image features and algorithm that was used was collected. If multiple AI algorithms were used in one study, only the one with the best performance with respect to overall diagnostic accuracy was considered.

## Quality Assessment

The quality of the studies was evaluated using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [8]. Four scenarios were assessed by QUADAS-2: 1) Differentiation of benign and malignant liver lesions (tumor dignity) using AI on B-mode ultrasound; 2) Diagnosis of specific tumor entities using AI on B-mode ultrasound 3) Differentiation of benign and malignant liver lesions (tumor dignity) using AI on CEUS 4) Diagnosis of specific tumor entities using AI on CEUS. The focus of the QUADAS-2 tool is the detection of a possible bias and the applicability of studies. For the assessment of a potential bias, the articles were examined using 12 signaling questions regarding patient selection, index test, reference standard, and flow and timing [8]. According to the recommendations of QUADAS-2, these signaling questions were adapted with regard to our research question. For concerns of applicability, the following criteria were added: 1) Are at least the diagnoses of cyst (B-mode only), FNH, hemangioma, HCC, and metastasis included? (Patient selection) 2) Has the AI been verified using an independent data set? (Index test) 3) Were diagnoses based on pathology or CT/MRI or clinical follow-up for more than 6 months? (Reference standard). The full adjustments and detailed results can be found in the supplemental data (**supplemental Tab. 1–4**).

## Results

### Literature search

A total of 660 articles were found during the literature search in PubMed, Web of Science, Scopus, and IEEE, with 152 duplicates. During the screening of abstracts, 184 articles were excluded because they were not original works (▶ **Fig. 1**) (e. g., reviews). An additional 260 studies were not included because our research question was not addressed. During the full-text analysis of 64 studies, 7 further articles were removed because of the imaging modality used (computed tomography, endosonography, or shear wave elastography). One study investigated the detection of tumors only, one article investigated splenic lesions in dogs, and one study did not report the accuracy of AI-based image classification alone (only in combination with clinical data). Additionally, two duplicate studies were removed. Finally, 52 articles remained for the final analysis. Of these, 32 studies investigated FLLs using B-mode ultrasound (10x dignity, 25x diagnosis) and 21 studies using CE ultrasound (8x dignity, 13x diagnosis).

### General approach of the identified studies using artificial intelligence

All studies followed a similar pattern (▶ **Fig. 2**). The first step comprised image optimization followed by manual or automated segmentation. Subsequently, while some studies used raw image
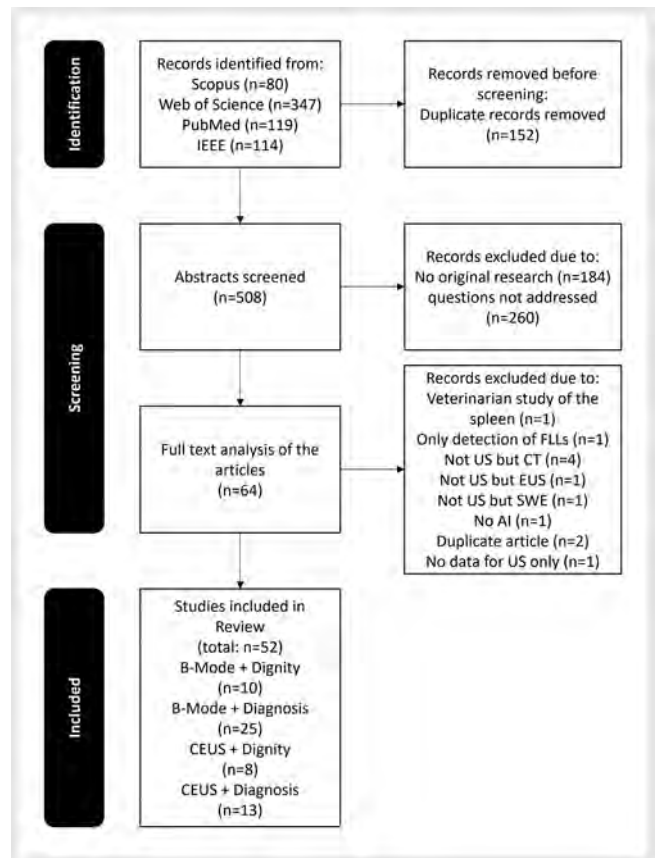


▶ **Fig. 1** Flowchart of the identification and selection process of studies. IEEE = Institute of Electrical and Electronics Engineers, US = ultrasound, CT = computed tomography, EUS = endoscopic ultrasound, SWE = shear wave elastography, CEUS: contrast-enhanced US.

data, others extracted specific image features to be analyzed by the AI algorithm. In the case of CEUS, some studies extracted time intensity curves (TIC). Examples of extracted B-mode data are contour properties and gray level features. Most often, a whole array of different features was extracted automatically by specific algorithms. Afterwards, feature selection was performed to reduce the number of collected features (in some studies several thousands) to a level the AI algorithm could work with efficiently. Few studies (B-mode only) considered additional clinical data for the classification process. Finally, the actual classification algorithm was applied, whereby ANNs and SVMs were used most often. AI was usually trained with the majority of images (about 80 %), followed by validation and testing with the remaining images from a database or cohort of patients. Many studies used x-fold cross-validation, a method in which the data are split into different training and validation sets repeatedly. External testing cohorts were rarely used.

### Artificial intelligence for the differentiation of benign and malignant liver lesions on B-mode ultrasound

We found ten studies using AI classification of B-mode ultrasound images to differentiate benign from malignant FLLs [9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. Studies were published between 2003
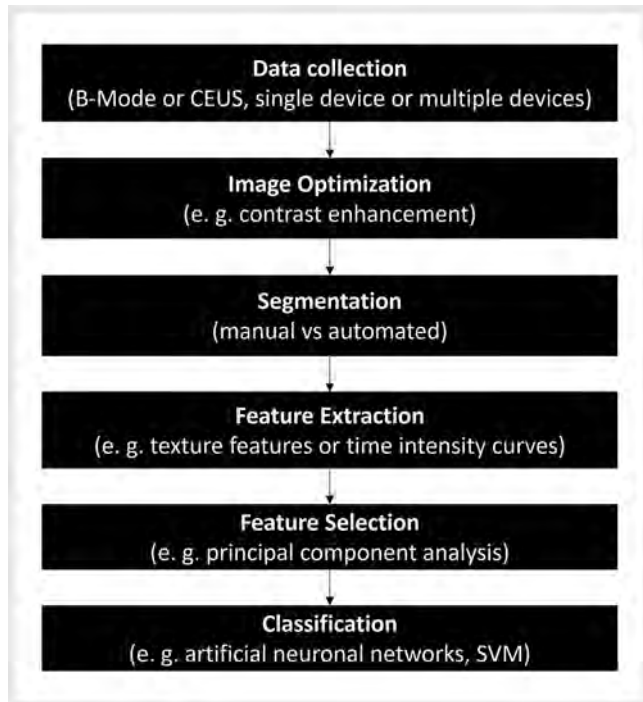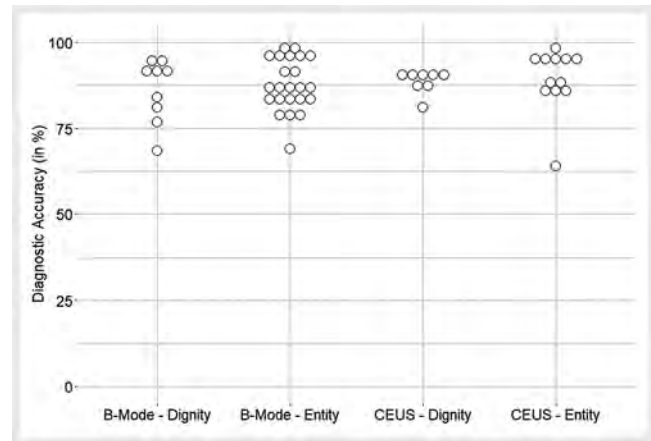
► **Fig. 2** General schematic of studies investigating AI-based classi-fication of FLLs.



► **Fig. 3** Overview of diagnostic accuracies. Each dot represents the reported diagnostic accuracy of a single study. For the study by Sri-tunyarat et al. (B-mode – entity), only values of external testing were available. Therefore, these are not shown here.

and 2021 (90 % from 2015 or later). Two studies were multicentric [15, 17]. Case numbers ranged from 101 to 23,756. Most studies (9 out of 10) included patients with hemangiomas and HCCs. FNHs were only considered in three studies. Where indicated, multiple different ultrasound machines were employed. Four studies extracted image features. All ten studies used an ANN to classify data (mainly convolutional neural networks (CNN), n = 8). The diagnostic accuracy for the assessment of tumor dignity ranged from 68.5 % to 94.8 % (► **Fig. 3**). Yang et al. exclusively conducted external testing with an independent patient cohort, the results of which did not substantially differ from internal test-ing [15]. Data are summarized in ► **Table 1**.

### Artificial intelligence for the differentiation of specific tumor entities in B-mode ultrasound

The database search revealed 25 studies using AI on B-mode ima-ges to diagnose specific tumor entities [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 11, 30, 31, 32, 13, 33, 34, 35, 36, 37, 16, 17, 38, 39]. Studies were published between 2003 and 2021, including six with a multicentric design [17, 20, 21, 29, 31, 32]. Case num-bers ranged from 51 to 3,873. Tumor entities differed substantial-ly between the studies. Regarding ultrasound devices, a wide range from a single machine to multiple devices from different manufacturers was found. In 16 studies texture features were ex-tracted, and in 9 AI obtained raw images. The most common algo-rithm used was an ANN (n = 17, mainly CNN (n = 8)). Furthermore, SVMs (n = 7) and logistic regression (n = 1) were applied to classify data. Diagnostic accuracies ranged from 69.0 % to 98.6 % (► **Fig. 3**). Two studies used an internal and external testing co-hort: Tiyarattanachai et al. observed hardly any differences be-

tween the two [32], whereas Ren et al. saw a noticeable decrease in diagnostic accuracy (internal testing: 79 %, external testing: 69 %) [29]. Data are summarized in ► **Table 2**.

### Artificial intelligence for the differentiation between benign and malignant liver lesions on CEUS

Eight studies using AI classification of CEUS data to differentiate be-nign from malignant FLLs published between 2014 and 2021 were found [40, 41, 42, 43, 44, 45, 46, 47]. Only one study was multicen-tric [45]. Most had a small sample size with a range from 26 to 363 cases and all but one performed their examinations with a single ul-trasound device. The remaining study used two machines from the same manufacturer [43]. Feature extraction was applied in all but one study and TIC data was used exclusively in two. Half of the stud-ies employed an SVM and three an ANN to classify lesions. The re-ported overall diagnostic accuracy ranged from 81.1 % to 91.6 % (► **Fig. 3**). Data are summarized in ► **Table 3**.

### Artificial intelligence for the differentiation of specific tumor entities on CEUS

Thirteen studies evaluated AI-based classification of different FLL entities with CEUS data [48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60]. The number of cases ranged from 37 to 527, the majority of studies (9/13) included more than 100. Only three studies in-cluded FNHs, hemangiomas, HCCs, and metastases in their analy-sis. Six studies used a single, three used multiple ultrasound devi-ces with the remaining four not disclosing this information. CEUS features were extracted in all studies, two of them analyzed TICs only. ANNs were most commonly used for FLL classification (58 %), followed by SVMs (30 %). Diagnostic accuracy ranged from 64.0 % to 98.3 % (► **Fig. 3**). Data are summarized in ► **Table 4**.

### Impact of the inclusion of clinical data on the diagnostic ac-curacy of artificial intelligence

Four studies investigated whether the additional consideration of clinical parameters to B-mode images is able to increase the diag-

► **Table 1** Summary of B-mode studies on lesion dignity. Studies are sorted alphabetically. Only the best diagnostic accuracy within one study without the consideration of clinical parameters is shown. Diagnostic accuracies are only comparable to a limited extent due to different testing measures and selection of diagnoses. [1]: When the number of patients was not available, the number of images was used. [2]: Value was estimated from a graph. [3]: Values for external testing. [4]: Retrospectively calculated diagnostic accuracy from sensitivity, specificity, and prevalence or positive/negative predictive values. ABS = abscess, AML = angiomyolipoma, ANN = artificial neural network, BEN = benign lesions, CCC = cholangiocarcinoma, CINO = cirrhotic nodule, FFD = focal fat deposition, FFS = focal fatty sparing, FNH = focal nodular hyperplasia, HCC = hepatocellular carcinoma, HEM = hemangioma, MAL = malignant lesions, MET = metastasis, N/A = not available, OBL = other benign lesions, OML = other malignant lesions.

| Author | Cases | Diagnoses | Feature extraction | AI | Accuracy (in %) | Sensitivity (in %) | Specificity (in %) | AUC |
|---|---|---|---|---|---|---|---|---|
| Acharya et al. 2018 | 101 | ABS, CYST, HCC, HEM, MET | yes | ANN | 93.0 | 90.8 | 97.4 | N/A |
| Hassan et al. 2021 | 352[1] | BEN, MAL | yes | ANN | ~92[2] | N/A | N/A | N/A |
| Ryu et al. 2021 | 3873 | CYST, HCC, HEM, MET | no | ANN | 90.4 | 95.0 | 86.0 | 0.970 |
| Sato et al. 2021 | 1080 | ABS, AML, CCC, CYST, FFD, FFL, FNH, HCC, HEM, MET, OBL | no | ANN | 68.5 | 67.3 | 69.8 | 0.721 |
| Tiyarattanachai et al. 2019 | 683 | CYST, HCC, HEM, FFD, FFS | no | ANN | 81 | 76 | 85 | 0.890 |
| Xi et al. 2021 | 596 | ABS, ADEN, CINO, CYST, FFD, FNH, HCC, HEM, OBL, OML | no | ANN | 84 | N/A | N/A | 0.830 |
| Yang et al. 2020 | 20625[1] | ABS, AML, CCC, ECH, FFS, FNH, HCC, HEM, MET, OBL, OML | yes | ANN | 76.7[4] (75.1)[3][4] | 80.5 (77.4)[3] | 60.1 (67.4)[3] | 0.779 (0.805)[3] |
| Yamakawa et al. 2019 | 324 | CYST, HCC, HEM, MET | no | ANN | 94.8 | 93.8 | 95.2 | N/A |
| Yamakawa et al. 2021 | 23756 | CYST, HCC, HEM, MET | no | ANN | 94.3 | 82.9 | 96.7 | N/A |
| Yoshida et al. 2003 | 44 | HCC, HEM, MET | yes | ANN | N/A | N/A | N/A | 0.92 |

nostic accuracy of AI-based classification [12, 15, 29, 39]. There was no data on this for CEUS. In all studies diagnostic accuracy could be improved. The effect was particularly pronounced in the study by Sato et al., in which the diagnostic accuracy increased from 68.5 % to 96.3 % [12]. Yang et al. were able to show that knowledge of the presence of hepatitis or tumor disease significantly improves the differentiation between benign and malignant lesions [15]. Zhou et al. were able to show that the consideration of CA-19–9 (OR 24.85) enhances the differentiation between HCC and other malignant processes of the liver almost as well as the AI algorithm itself (OR 29.52) [39]. Data are summarized in ► **Table 5**.

### Diagnostic performance of artificial intelligence in comparison to ultrasound professionals

A total of seven B-mode and CEUS studies compared the diagnostic accuracy of AI algorithms to radiologists interpreting the same cases [51, 42, 53, 57, 45, 14, 15]. Additional clinical information was available to radiologists in some of the studies. AI matched the diagnostic performance of experts in five studies and significantly outperformed beginners in two studies and experts in one study. Hu et al. reported that the diagnostic accuracy of less experienced examiners improved when combined with AI, while the diagnostic accuracy of experts worsened [51]. Data are summarized in ► **Table 6**.

### Quality assessment using QUADAS-2

All studies were reviewed for potential bias and applicability concerns using QUADAS-2. In general, most studies did not provide all the information needed to assess the risk of bias. For example, the domain "patient selection" remained unclear for most studies, as it was not evident from the articles whether patients were recruited consecutively or not. Using all available information, the risk of bias was considered to be low (► **Fig. 4a**). In contrast, applicability was a concern for most studies (► **Fig. 4b**). In the domain "patient selection", it was noticeable that the majority of studies did not include FNHs in their analysis. Furthermore, only a few studies validated the diagnostic accuracy of their AI algorithm with an independent data set, which decreases the applicability of the index test. There were similar results concerning

▶ **Table 2** Summary of B-mode studies on the assessment of tumor entity. Studies are sorted alphabetically. Only the best diagnostic accuracy within one study without the consideration of clinical parameters is shown. Diagnostic accuracies are only comparable to a limited extent due to different testing measures and selection of diagnoses. [1]: When the number of patients was not available, the number of images was used. [2]: Values for external testing. [3]: Retrospectively calculated diagnostic accuracy from sensitivity, specificity, and prevalence or positive/negative predictive values. ABS = abscess, AML = angiomyolipoma, ANN = artificial neural network, BEN = benign lesions, CCC = cholangiocarcinoma, CINO = cirrhotic nodule, FFD = focal fat deposition, FFS = focal fatty sparing, FNH = focal nodular hyperplasia, HCC = hepatocellular carcinoma, HEM = hemangioma, ICC = intrahepatic cholangiocarcinoma, MAL = malignant lesions, MET = metastasis, N/A = not available, OBL = other benign lesions, OML = other malignant lesions, SVM: support vector machine.

| Author | Cases | Diagnoses | Feature extraction | AI | Accuracy (in %) | Sensitivity (in %) | Specificity (in %) | AUC |
|---|---|---|---|---|---|---|---|---|
| Balasubramanian et al. 2017 | 160[1] | N/A | yes | ANN | 84.6[3] | N/A | N/A | N/A |
| Hassan et al. 2015[4] | 110[1] | CYST,HCC, HEM | yes | SVM | 96.5 | 97.6 | 92.5 | N/A |
| Hassan et al. 2017[4] | 110 | CYST, HCC, HEM | no | ANN | 97.2 | 98 | 95.7 | N/A |
| Hwang et al. 2015 | 115 | CYST, HEM, MAL | yes | ANN | 98.1[3] | N/A | N/A | N/A |
| Lee et al. 2011 | 102 | CYST, HEM, MAL | yes | SVM | 83.3 | 66.7 | 83.3 | 0.77 |
| Mao et al. 2021 | 114 | HCC, ICC, MET | yes | Other | 84.3 | 76.8 | 88.0 | 0.816 |
| Mitrea et al. 2019 | 300 | HCC, HEM | yes | ANN | 85.4 | 78.0 | 82.9 | 0.805 |
| Mittal et al. 2011[4] | 176 | CYST, HCC, HEM, MET | yes | ANN | 86.4 | N/A | N/A | N/A |
| Peng et al. 2022 | 589 | INF, MAL | yes | SVM | 79.1 | 86.3 | 45.2 | 0.745 |
| Qiu et al. 2011[4] | 256[1] | HCC, HEM | yes | SVM | 96.9[3] | N/A | N/A | N/A |
| Ren et al. 2021 | 188 | CCC, HCC | yes | SVM | 79.0 (69.2)[2] | 90.0 (66.7)[2] | 75.0 (70.0)[2] | 0.843 (0.730)[2] |
| Ryu et al. 2021 | 3873 | CYST, HCC, HEM, MET | no | ANN | 82.2 | 86.7 | 89.7 | 0.947 |
| Schmauch et al. 2019[4] | 544 | CYST, FNH, HCC, HEM, MET | no | ANN | N/A | N/A | N/A | (0.891)[2] |
| Sritunyarat et al. 2020 | 157 | CYST, HCC, HEM, FFD, FFS | no | ANN | (95.0)[2] | (87.0)[2] | (97.0)[2] | N/A |
| Tiyarattanachai et al. 2019 | 683 | CYST, HCC, HEM, FFD, FFS | no | ANN | 69 | N/A | N/A | N/A |
| Tiyarattanachai et al. 2021 | 3872 | CYST, HCC, HEM, FFD, FFS | no | ANN | 95.4 (95.3)[2] | 83.9 (84.9)[2] | 97.1 (97.1)[2] | N/A |
| Virmani et al. 2013[4] | 108[1] | CYST, HCC, HEM, MET | yes | ANN | 87.7 | N/A | N/A | N/A |
| Virmani et al. 2013 | 51 | HCC, MET | yes | SVM | 91.6 | N/A | N/A | N/A |
| Virmani et al. 2013[4] | 108[1] | CYST, HCC, HEM, MET | yes | SVM | 87.2 | N/A | N/A | N/A |
| Virmani et al. 2014 | 108[1] | CYST, HCC, HEM, MET | yes | ANN | 95.0 | N/A | N/A | N/A |
| Xu et al. 2020 | 79 | ABS, HCC | yes | ANN | 83.8 | N/A | N/A | N/A |
| Yamakawa et al. 2019 | 324 | CYST, HCC, HEM, MET | no | ANN | 88.0 | 80.4 | 96.0 | N/A |
| Yamakawa et al. 2021 | 23756 | CYST, HCC, HEM, MET | no | ANN | 91.1 | N/A | N/A | N/A |

▶ **Table 2** (Continuation)

| Author | Cases | Diagnoses | Feature extraction | AI | Accuracy (in %) | Sensitivity (in %) | Specificity (in %) | AUC |
|---|---|---|---|---|---|---|---|---|
| Zhang et al. 2010[4] | 280[1] | CYST, HCC, HEM | yes | ANN | 98.6[3] | N/A | N/A | N/A |
| Zhou et al. 2021 | 172 | HCC, OML | no | ANN | 78.4[3] | 57.1 | 91.3 | 0.74 |

▶ **Table 3** Summary of CEUS studies on tumor dignity. Studies are sorted alphabetically. Only the best diagnostic accuracy within one study without the consideration of clinical parameters is shown (values for sensitivity, specificity, and AUC are reported for the method with the best diagnostic accuracy). Diagnostic accuracies are only comparable to a limited extent due to different testing measures and selection of diagnoses. [1]: Including clips which could not be analyzed by AI. ABS = abscess, ANN = artificial neural network, BEN = benign lesions, FFS = focal fatty sparing, FNH = focal nodular hyperplasia, HCC = hepatocellular carcinoma, HEM = hemangioma, MAL = malignant lesions, MET = metastasis, N/A = not available, SVM = support vector machine, TIC = time intensity curve.

| Author | Cases | Diagnoses | Feature extraction | AI | Accuracy in % | Sensitivity in % | Specificity in % | AUC |
|---|---|---|---|---|---|---|---|---|
| Guo et al. 2017 | 93 | BEN, MAL | yes | Other | 90.4 | 93.6 | 89.3 | 0.95 |
| Guo et al. 2018 | 83 | CCC, FNH, HCC, HEM, MET, | yes | Other | 90.4 | 93.6 | 86.9 | 0.97 |
| Hu et al. 2021 | 363 | BEN, MAL | no | ANN | 91.0 | 92.7 | 85.1 | 0.93 |
| Kondo et al. 2017 | 94 | FNH, HCC, HEM, MET | Yes (TIC) | SVM | 91.6 | 94.0 | 90.3 | N/A |
| Qian et al. 2017 | 93 | BEN, MAL | yes | SVM | 89.4 | 89.7 | 89.8 | 0.96 |
| Ta et al. 2018 | 105 | BEN, MAL | Yes (+TIC) | ANN&SVM | 81.1 (73.3)[1] | 90.0 (83.3)[1] | 71.1 (62.7)[1] | 0.88 |
| Wu et al. 2014 | 26 | ABS, FFS, HCC, HEM, MET | Yes (TIC) | ANN | 86.4 | 83.3 | 87.5 | N/A |
| Zhang et al. 2021 | 153 | BEN, MAL | yes | SVM | 88.2 | 86.9 | 89.4 | 0.9 |

bias and applicability for the subgroups B-mode, CEUS, tumor dignity and tumor entity. More detailed information and the assessments of individual studies are included in the supplemental data (**Supplemental Fig. 1** and **Supplemental Tab. 1–4**).

## Discussion

Sonography can be used to reliably determine the dignity and entity of many focal liver lesions. However, even with the use of CEUS, not every lesion can be classified correctly. Since AI-based applications have found their way into many scientific fields, there is reasonable hope, that AI could also help to improve ultrasound-based diagnosis of FLLs and potentially avoid the need for additional imaging and invasive procedures. The aim of this systematic review was to analyze studies in which the dignity or entity of FLLs was assessed by AI, using B-mode or CEUS data. For this purpose, 52 articles found using a structured literature search approach were analyzed systematically in order to answer the following questions:

## How powerful is artificial intelligence for the classification of liver tumors?

Diagnostic accuracy describes the fraction of cases which are assigned the correct diagnosis based on the test procedure. Typically, diagnostic accuracy of more than 80 % is considered good and more than 90 % excellent [61].

Half of the B-mode studies assessing FLL dignity reported excellent diagnostic accuracy, and a further 20 % of the studies showed good performance (range: 68.5 % to 94.8 %). The impact of lesion size on diagnostic accuracy, sensitivity, specificity, and AUC was investigated in one study with no significant differences between sizes 1.1–2.0 cm, 2.1–5.0 cm, and > 5.0 cm [15]. Yamakawa et al. reported higher accuracies for cysts (99.0 %) and hemangiomas (91.0 %) in comparison to HCCs (67.5 %) and liver metastases (62.8 %) on B-mode [17]. A different study did not observe differences between these entities [31]. Studies that analyzed CEUS data to classify FLL-dignity all showed good (50 %) or excellent (50 %) diagnostic performance.

When assessing specific tumor entities based on the B-mode image, accuracies ranged from 69.0 % to 98.6 %. 40 % of studies reported good and a further 40 % reported excellent diagnostic

▶ **Table 4** Summary of CEUS studies on the assessment of tumor entity. Studies are sorted alphabetically. Only the best diagnostic accuracy within one study without the consideration of clinical parameters is shown (values for sensitivity, specificity, and AUC are reported for the method with the best diagnostic accuracy). Diagnostic accuracies are only comparable to a limited extent due to different testing measures and selection of diagnoses. [1]: Values for external testing; ADEN = adenoma, ANN = artificial neural network, BEN = benign lesions, FFC = focal fatty change, FNH = focal nodular hyperplasia, HCC = hepatocellular carcinoma, HEM = hemangioma, MAL = malignant lesions, MET = metastasis, N/A = not available, SVM: support vector machine, TIC = time intensity curve.

| Author | Cases | Diagnoses | Feature extraction | AI | Accuracy in % | Sensitivity in % | Specificity in % | AUC |
|---|---|---|---|---|---|---|---|---|
| Căleanu et al. 2014 | 37 | FNH, HCC, HEM, MET | yes | SVM | 64.0 | N/A | N/A | N/A |
| Căleanu et al. 2021 | 91 | FNH, HCC, HEM, MET | yes (TIC) | ANN | 95.7 | N/A | N/A | N/A |
| De Senneville et al. 2020 | 47 | ADEN, FNH | yes | Other | 95.9 | 93.4 | 97.6 | 0.97 |
| Hu et al. 2019 | 527 | N/A | yes | ANN | 85–88 | 89–94 | 67–70 | 0.89 |
| Huang et al. 2020 | 342 | FNH, HCC | yes | SVM | 94.4 | 94.8 | 93.6 | N/A |
| Li et al. 2021 | 226 | FNH, HCC | yes | SVM | N/A | 76.6 | 80.5 | 0.86 |
| Liang et al. 2016 | 353 | FNH, HCC, HEM | yes | Other | 84.8 | N/A | N/A | N/A |
| Shiraishi et al. 2008 | 103 | HCC, HEM, MET | yes | ANN | 88.3 | N/A | N/A | N/A |
| Sîrbu et al. 2020 | 95 | FNH, HCC, HEM, MET | N/A | ANN | 95.7 | N/A | N/A | N/A |
| Streba et al. 2012 | 112 | FFC, HCC, HEM, MET | Yes (TIC) | ANN | 87.1 | 93.2 | 89.7 | N/A |
| Sugimoto et al. 2009 | 137 | HCC, HEM, MET | yes | ANN | 94.2 | N/A | N/A | N/A |
| Sugimoto et al. 2010 | 137 | HCC, HEM, MET | yes | ANN | 88.3 | N/A | N/A | N/A |
| Zhou et al. 2021 | 186 | FNH, HCC | yes | SVM | 98.3 (96.7)[1] | 98.1 (98.7)[1] | 98.6 (94.7)[1] | N/A |

accuracy. In CEUS studies regarding the differentiation of FLL entities, all but one (92%) showed at least good performance with six reporting excellent accuracy.

In order to measure diagnostic accuracy as exactly as possible, AI-based classification algorithms should ideally be evaluated by means of external validation. This requires the use of an independent test set of patients, which the AI has not been trained on (even partially). Only five B-mode studies and one CEUS-based study performed external validation. Two of these studies compared the diagnostic accuracies with their internal set (the AI had also been trained on) and found no significant differences [15, 32]. The other two studies found a deterioration of diagnostic accuracy when using an external set [29, 60]. The remaining two B-mode studies did not test on the internal set, and, therefore, a comparison was not possible [30, 31]. These differing results for the external validation cohort might be due to a considerable variation in case numbers (3872 and 20,625 [32, 15] vs. 188 and 186 cases [29, 60]). Alternatively, the conflicting results could originate from unknown random or systematic differences between the internal and the validation data set.

## Can the potency of artificial intelligence be improved by adding clinical parameters?

Clinical data indicate pre-test probability and should, therefore, always be considered by physicians when making a diagnosis.

Somewhat surprisingly, only four B-mode studies considered this approach for their AI algorithms. All of them were able to show that the diagnostic accuracy of AI-based FLL classification can be improved by adding clinical parameters. Among other things, gender, age, and a positive history of hepatitis or cancer had a significant impact on diagnostic accuracy. In the multivariate analysis, some parameters (e. g., CA19–9) were almost as relevant for the correct classification as the interpretation of image data itself [39]. Sato et al. achieved the highest diagnostic accuracy among the aforementioned studies with the combination of B-mode image data and clinical parameters [12].

## Artificial intelligence vs. human intelligence – which is better?

A total of seven studies (2x B-mode, 5x CEUS) compared physicians' diagnostic performance with that of their AI algorithms. According to the results, AI performed as well as experienced radiologists in five studies and better in one study. However, in the latter study the human diagnostic accuracy was low (ACC 69.5%) [15]. Another study reported that the availability of AI-based classification improved the diagnostic accuracy of less experienced examiners but was a setback for experts [51]. These results are remarkable, even more so knowing that physicians had an advantage by having insight into the clinical parameters in three of the studies.

▶ **Table 5** Summary of B-mode studies adding clinical data to AI analysis. 1: retrospectively calculated diagnostic accuracy from sensitivity, specificity, and prevalence or positive/negative predictive values.

| Study | Mode | ACC without clinical data | ACC with clinical data | Clinical and sonographic parameters (odds ratio) |
|---|---|---|---|---|
| Sato et al. 2022 | B-mode / Dignity | 68.5 % | 96.3 % | Clinical parameters<br>Age, gender, AST, ALT, platelet count, albumin |
| Yang et al. 2020 | B-mode / Dignity | 76.7 %[1] | 87.0 %[1] | OR for malignant lesions:<br>Hypoechoic halo<br>(18.389 [9.921–34.084])<br>History of extrahepatic tumor (16.17 [9.311–28.065])<br>History of hepatitis<br>(11.736 [7.857–17.529])<br>Age > 65y<br>(3.323 [2.096–5.269])<br>Male gender<br>(2.303 [1.629–3.256])<br>Intratumoral vascularity<br>(1.911 [1.344–2.717]) |
| Ren et al. 2021 | B-mode / entity | 78.95 % | 86.8 % | Clinical parameters<br>Age, gender, history of hepatitis, AFP, ALT, AST, TB, CB, UCB, size of lesion |
| Zhou et al. 2021 | B-mode / entity (HCC vs. other malignancies) | 57.1 % | 78.6 % | OR for non-HCC malignancies<br>CA19–9<br>(24.85 [6.10–101.25])<br>Female gender<br>(3.72 [1.17–11.9]) |

▶ **Table 6** Summary of studies comparing the performance of AI with physician-based decisions. [1]: TIC analysis. [2]: Additional clinical information. N/A = not available.

| Author | Mode | ACC (expert) | p | ACC (beginner) | p | ACC (AI) | Conclusion |
|---|---|---|---|---|---|---|---|
| Hu et al. 2019 | CEUS | N/A | N/A | N/A | N/A | 85–88 | AI was a setback for experts |
| Hu et al. 2021[2] | CEUS | 87.5 | 0.256 | 83.0 | 0.021 | 91.0 | AI matched experts |
| Li et al. 2021 | CEUS | 0.84 (AUC) | N/A | N/A | N/A | 0.86 (AUC) | AI matched experts |
| Streba et al. 2012[1] | CEUS | N/A | 0.225 | N/A | N/A | 87.1 | AI matched experts |
| Ta et al. 2018 | CEUS | 81.4 | N/A | 72.0 | N/A | 81.1 | AI matched experts, better than beginners |
| Xi et al. 2021[2] | B-mode | 80.0 (1x)<br>73.0 (1x) | 0.18 | N/A | N/A | 84.0 | AI matched experts |
| Yang et al. 2020[2] | B-mode | 69.5 | < 0.01 | 64.7 | < 0.01 | 84.7 | AI better than experts |

Ta et al. observed that radiologists were able to successfully analyze (not classify) CEUS data more often (inexperienced: 95.2 %, experienced 97.1 %) than their AI algorithms (90.5 %) [45]. An inability to analyze cases was the result of poor image quality, contrast agent enhancement, or small size of the FLL (< 1 cm). When taking these unclassifiable lesions into account, the diagnostic accuracy of the AI-based approach dropped from 81.1 % to 73.3 % (for radiologists: inexperienced: 68.6 %, experienced 79.0 %). Whether the accuracies reported in other studies were calculated with this consideration in mind is doubtful.

It can be concluded that AI classification of FLLs is able to achieve diagnostic accuracies comparable to experienced human observers under rather artificial study conditions. There is not enough data to make reasonable conclusions about the differences in diagnostic performance between AI and humans in a real-world setting.
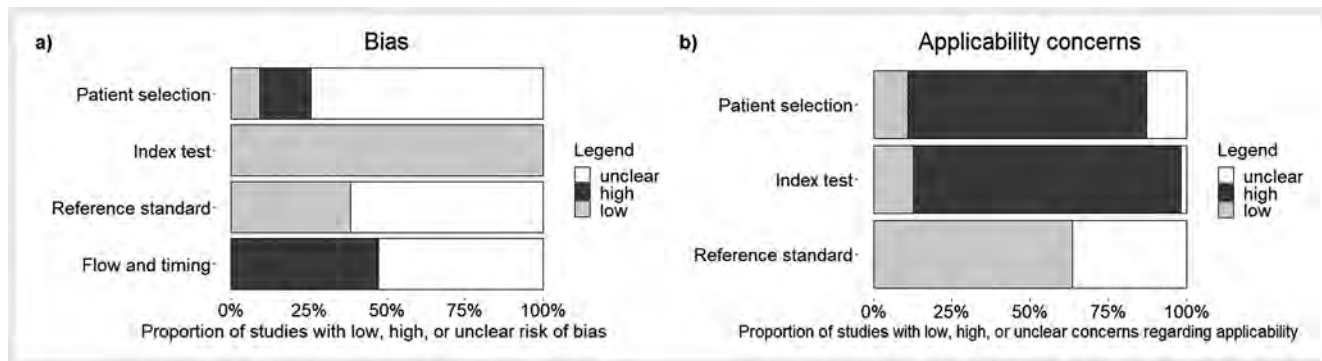
► **Fig. 4** QUADAS-2 overview. **a)** Risk of bias for all studies (light gray: low risk, dark gray: high risk, white: unclear risk). **b)** Applicability concerns for all studies (light gray: low level of concerns, dark gray: high level of concerns, white: unclear level of concerns).

## What are the concerns and limitations for the use of artificial intelligence to classify FLLs?

General concerns about the use of AI in the medical field include the protection of patients' individual rights and personal information, especially if data are not being analyzed on site. Another critical aspect of the methods discussed in this article is their black-box nature. There is often no easy way to interpret or explain the produced results. Providers of AI-based classification systems will need to ensure that their technological approach is as transparent and reliable as possible. A recent research topic called explainable AI is trying to resolve this issue [62]. Liability concerns will probably be the biggest obstacle keeping AI from implementation in clinical practice.

A limitation of the studies included in this systematic review is that image acquisition was often performed on only one type of ultrasound machine, raising doubts about a possible transfer to general clinical usage. Furthermore, most studies included a limited spectrum of different FLL entities in their analysis, which reduces the applicability for clinical practice. For example, FNHs were included only in a minority of B-mode studies (11 %), even though they are one of the most common FLLs. While this might seem understandable, since the diagnosis of FNH is not based on B-mode ultrasound, but rather is a domain of CEUS, it certainly leads to a selection bias and raises doubts about the significance of the reported diagnostic accuracies.

A major limitation of almost all studies we reviewed is the lack of a sufficiently large database. The number of images an AI method is trained on directly affects its diagnostic performance. Most studies, therefore, used augmentation techniques, such as mirroring or rotation of images, which cannot fully compensate for a lack of real data. In addition, these small data sets lead to limitations concerning the testing process. As mentioned above, an independent patient cohort was not used for testing in the majority of studies. Testing can be performed by splitting all images into a training/validation and test data set. This can lead to images from one patient ending up in both data sets, therefore resulting in an overestimation of testing accuracy. The issue can be addressed by splitting patients (and not images) into groups. A CEUS-based study, which compared the two approaches, observed a drop in diagnostic accuracy from 95.7 % to 56 % [49]. Although this pro-

nounced deterioration of accuracy can certainly not be generalized, it must be assumed that some of the reported results are overrated. This is especially true for small studies with a homogenous set of data or patients.

Another key issue is that all studies needed intervention by healthcare professionals not only to perform ultrasound scans, but also to process the collected data further (e. g., demarcation of the regions of interest (ROI)). This puts a significant part of diagnostic ultrasound, i. e., differentiating the FLL from the liver parenchyma, back into human hands. Some studies have tried to solve this problem by developing algorithms that are able to identify the ROI. Liang et al. trained an AI algorithm to track FLLs and their corresponding ROIs in CEUS clips automatically [54]. Nonetheless, they needed a physician with CEUS experience to identify the ROI at the start of the clip. With this approach, they were able to achieve diagnostic accuracies similar to studies with manual ROI placement (84.8 % for the differentiation of FLL entities and 92.7 % to distinguish benign from malignant FLLs). Also, there are studies that are solely focused on the detection of FLLs and not their classification (and were therefore disregarded for the purpose of this review). They have shown promising results, indicating that solutions addressing this issue seem possible [63]. For the implementation of AI techniques in the clinical routine, a combination of both techniques (detection and classification) would be ideal, as this would eliminate possible bias introduced by the examiner.

In summary, although the diagnostic capabilities of AI for the diagnosis of FLLs are almost all reported to be good or excellent, among other concerns, the lack of independent test sets and the exclusion of common FLL entities in almost all studies severely limit the real-world applicability of these data. Therefore, the pathway towards the implementation of AI in clinical ultrasound of the liver has many hurdles to overcome. User-friendly AI-based tools, which are built into ultrasound devices for specific questions such as "is this a malignant liver lesion?" could be a starting point. Ideally, real-world data from the application of these tools would be used to further improve AI performance in a continuous learning approach. Data protection concerns will limit this kind of feedback loop to clinical trials. Therefore, large multicenter cohorts will be necessary to improve AI-based ultrasound techniques before a significant impact on clinical practice seems feasi-

ble. In the long term, AI-based approaches will need to integrate data from multiple sources such as ultrasound, radiology, histopathology, laboratory tests, and clinical information to make a diagnosis [64]. As for now and the near future, the only viable field of use for AI in clinical ultrasound seems to be to support (especially inexperienced) physicians in their decision making.

A limitation of our review is the heterogeneity of the studies. Heterogeneity was observed in all study parts, starting with the selection of patients or image databases. Differences continued with respect to the pre-processing of images, extracted image features, and types of AI that were used (e. g., CNN or SVM). Finally, as outlined above, testing of the diagnostic performance varied significantly. These differences severely limit the comparability of studies included in this systematic review.

## Conclusion and Outlook

Data on the AI-based classification of ultrasound imaging of FLLs are promising. The diagnostic performance of AI-based classification should be improved by adding clinical data. AI could serve as a supportive system for ultrasound examinations of the liver, especially for inexperienced examiners. The main weaknesses of the available studies are the limited spectrum of FLL entities and the lack of external validation. Moreover, in addition to technical hurdles, regulatory hurdles must be overcome for a successful transfer of the technology to clinical practice. Large, cross-center ultrasound image databases could help to improve the diagnostic capabilities of AI-based classification systems.

## Conflict of Interest

The authors declare that they have no conflict of interest.

## References

[1] Kaltenbach TEM, Engler P, Kratzer W et al. Prevalence of benign focal liver lesions: ultrasound investigation of 45,319 hospital patients. Abdom Radiol (NY) 2016; 41: 25–32. doi:10.1007/s00261-015-0605-7

[2] Strobel D, Seitz K, Blank W et al. Contrast-enhanced ultrasound for the characterization of focal liver lesions–diagnostic accuracy in clinical practice (DEGUM multicenter trial). Ultraschall in Med 2008; 29: 499–505. doi:10.1055/s-2008-1027806

[3] Wu M, Li L, Wang J et al. Contrast-enhanced US for characterization of focal liver lesions: a comprehensive meta-analysis. Eur Radiol 2018; 28: 2077–2088. doi:10.1007/s00330-017-5152-x

[4] Friedrich-Rust M, Klopffleisch T, Nierhoff J et al. Contrast-Enhanced Ultrasound for the differentiation of benign and malignant focal liver lesions: a meta-analysis. Liver Int 2013; 33: 739–755. doi:10.1111/liv.12115

[5] Vetter M, Kremer AE, Agaimy A et al. The amount of liver tissue is essential for accurate histological staging in patients with autoimmune hepatitis. J Physiol Pharmacol 2021; 72. doi:10.26402/jpp.2021.1.13

[6] Nishida N, Yamakawa M, Shiina T et al. Current status and perspectives for computer-aided ultrasonic diagnosis of liver lesions using deep learning technology. Hepatol Int 2019; 13: 416–421. doi:10.1007/s12072-019-09937-4

[7] Survarachakan S, Prasad PJR, Naseem R et al. Deep learning for image-based liver analysis – A comprehensive review focusing on malignant lesions. Artif Intell Med 2022; 130: 102331. doi:10.1016/j.artmed.2022.102331

[8] Whiting PF, Rutjes AW, Westwood ME et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine 2011; 155. doi:10.7326/0003-4819-155-8-201110180-00009

[9] Acharya UR, Koh JEW, Hagiwara Y et al. Automated diagnosis of focal liver lesions using bidirectional empirical mode decomposition features. Computers in Biology and Medicine 2018; 94: 11–18. doi:10.1016/j.compbiomed.2017.12.024

[10] Hassan T, Alzoubi A, Du H et al. Towards optimal cropping: breast and liver tumor classification using ultrasound images. In: Agaian SS, (eds.) Multimodal Image Exploitation and Learning 2021: 12–16 April 2021, online only, United States. volume 11734. Proceedings of SPIE. Bellingham, Washington, USA: SPIE; 2021: 15. doi:10.1117/12.2589038

[11] Ryu H, Shin SY, Lee JY et al. Joint segmentation and classification of hepatic lesions in ultrasound images using deep learning. Eur Radiol 2021; 31: 8733–8742. doi:10.1007/s00330-021-07850-9

[12] Sato M, Kobayashi T, Soroida Y et al. Development of novel deep multimodal representation learning-based model for the differentiation of liver tumors on B-mode ultrasound images. J Gastroenterol Hepatol 2022; 37: 678–684. doi:10.1111/jgh.15763

[13] Tiyarattanachai T, Chaiteerakij R, Marukatat S et al. 685 – Computer-Assisted Ultrasonographic Image Analysis for Differentiation Between Hepatocellular Carcinoma (HCC) and Benign Focal Liver Lesions. Gastroenterology 2019; 156: S–1211. doi:10.1016/S0016-5085(19)40011-5

[14] Xi IL, Wu J, Guan J et al. Deep learning for differentiation of benign and malignant solid liver lesions on ultrasonography. Abdom Radiol (NY) 2021; 46: 534–543. doi:10.1007/s00261-020-02564-w

[15] 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

[16] Yamakawa M, Shiina T, Nishida N et al. Computer aided diagnosis system developed for ultrasound diagnosis of liver lesions using deep learning. In: 2019 IEEE International Ultrasonics Symposium (IUS). Piscataway, NJ: IEEE; 2019: 2330–2333. doi:10.1109/ULTSYM.2019.8925698

[17] Yamakawa M, Shiina T, Tsugawa K et al. Deep-learning framework based on a large ultrasound image database to realize computer-aided diagnosis for liver and breast tumors. In: IEEE IUS 2021: International Ultrasonics Symposium : virtual symposium, September 11–16, 2021 (September 11–12: short courses/September 12–16: technical program) : 2021 symposium proceedings. Piscataway, NJ, USA: IEEE; 2021: 1–4. doi:10.1109/IUS52206.2021.9593518

[18] Yoshida H, Casalino DD, Keserci B et al. Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images. Phys Med Biol 2003; 48: 3735–3753. doi:10.1088/0031-9155/48/22/008

[19] Balasubramanian D, Srinivasan P, Gurupatham R. Automatic classification of focal lesions in ultrasound liver images using principal component analysis and neural networks. Annu Int Conf IEEE Eng Med Biol Soc 2007; 2007: 2134–2137. doi:10.1109/IEMBS.2007.4352744

[20] Hassan TM, Elmogy M, Sallam E. A classification framework for diagnosis of focal liver diseases. In: Abdelaal WGA, (eds.) ICCES: 2015 Tenth International Conference on Computer Engineering & Systems (ICCES): Ain Shams University Guest House, Cairo, Egypt, December 23rd-24th, 2015: proceedings. Piscataway, NJ: IEEE; 2015: 395–401. doi:10.1109/ICCES.2015.7393083

[21] Hassan TM, Elmogy M, Sallam ES. Diagnosis of Focal Liver Diseases Based on Deep Learning Technique for Ultrasound Images. Arab J Sci Eng 2017; 42: 3127–3140. doi:10.1007/s13369-016-2387-9

[22] Hwang YN, Lee JH, Kim GY et al. Classification of focal liver lesions on ultrasound images by extracting hybrid textural features and using an artificial neural network. Biomed Mater Eng 2015; 26: S1599–611. doi:10.3233/BME-151459

[23] Lee S, Jo IA, Kim KW et al. Enhanced classification of focal hepatic lesions in ultrasound images using novel texture features. In: 2011 18th IEEE International Conference on Image Processing (ICIP 2011): Brussels, Belgium, 11 – 14 September 2011. Piscataway, NJ: IEEE; 2011: 2025–2028. doi:10.1109/ICIP.2011.6115876

[24] Mao B, Ma J, Duan S et al. Preoperative classification of primary and metastatic liver cancer via machine learning-based ultrasound radiomics. Eur Radiol 2021; 31: 4576–4586. doi:10.1007/s00330-020-07562-6

[25] Mitrea D, Nedevschi S, Mitrea P et al. HCC Recognition Within Ultrasound Images Employing Advanced Textural Features with Deep Learning Techniques. In: Li Q, Wang L, (eds.) Proceedings 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics: CISP-BMEI2019 : 19–21 October 2019, Huaqiao, China. Piscataway, NJ: IEEE; 2019: 1–6. doi:10.1109/CISP-BMEI48845.2019.8965874

[26] Mittal D, Kumar V, Saxena SC et al. Neural network based focal liver lesion diagnosis using ultrasound images. Comput Med Imaging Graph 2011; 35: 315–323. doi:10.1016/j.compmedimag.2011.01.007

[27] Peng JB, Peng YT, Lin P et al. Differentiating infected focal liver lesions from malignant mimickers: value of ultrasound-based radiomics. Clin Radiol 2022; 77: 104–113. doi:10.1016/j.crad.2021.10.009

[28] Qiu W, Wang R, Xiao F et al. Research on Fuzzy Enhancement in the Diagnosis of Liver Tumor from B-mode Ultrasound Images. In: Staff I, (eds.) 2011 International Conference on Intelligent Computation and Bio-Medical Instrumentation. Place of publication not identified: IEEE; 2011: 74–80. doi:10.1109/ICBMI.2011.17

[29] Ren S, Li Q, Liu S et al. Clinical Value of Machine Learning-Based Ultrasomics in Preoperative Differentiation Between Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma: A Multicenter Study. Front Oncol 2021; 11: 749137. doi:10.3389/fonc.2021.749137

[30] Sritunyarat Y, Chaiteerakij R, Tiyarattanachai T et al. 456 PERFORMANCE OF ARTIFICIAL INTELLIGENCE IN DIAGNOSING FOCAL LIVER LESIONS DETECTED BY VARIOUS TRANS-ABDOMINAL ULTRASONOGRAPHIC MACHINES: A VALIDATION STUDY. Gastroenterology 2020; 158: S–1278. doi:10.1016/S0016-5085(20)33857-9

[31] Schmauch B, Herent P, Jehanno P et al. Diagnosis of focal liver lesions from ultrasound using deep learning. Diagn Interv Imaging 2019; 100: 227–233. doi:10.1016/j.diii.2019.02.009

[32] Tiyarattanachai T, Apiparakoon T, Marukatat S et al. Development and validation of artificial intelligence to detect and diagnose liver lesions from ultrasound images. PLoS One 2021; 16: e0252882. doi:10.1371/journal.pone.0252882

[33] Virmani J, Kumar V, Kalra N et al. Characterization of primary and secondary malignant liver lesions from B-mode ultrasound. J Digit Imaging 2013; 26: 1058–1070. doi:10.1007/s10278-013-9578-7

[34] Virmani J, Kumar V, Kalra N et al. A comparative study of computer-aided classification systems for focal hepatic lesions from B-mode ultrasound. J Med Eng Technol 2013; 37: 292–306. doi:10.3109/03091902.2013.794869

[35] Virmani J, Kumar V, Kalra N et al. Neural network ensemble-based CAD system for focal liver lesions from B-mode ultrasound. J Digit Imaging 2014; 27: 520–537. doi:10.1007/s10278-014-9685-0

[36] Virmani J, Vinod V, Kalra N et al. PCA-SVM based CAD System for Focal Liver Lesions using B-Mode Ultrasound Images. DSJ 2013; 63: 478–486. doi:10.14429/dsj.63.3951

[37] Xu SSD, Chang CC, Su CT et al. Classification of Hepatocellular Carcinoma and Liver Abscess by Applying Neural Network to Ultrasound Images. Sensors and Materials 2020; 32: 2659. doi:10.18494/SAM.2020.2801

[38] Zhang XY, Diao XF, Wang TF et al. Study on Feature Extraction for Ultrasonic Differentiation of Liver Space-Occupying Lesions. In: 2010 4th International Conference on Bioinformatics and Biomedical Engineering: iCBBE 2010; Chengdu, China, 18 – 20 June 2010. Piscataway, NJ: IEEE; 2010: 1–4. doi:10.1109/ICBBE.2010.5517018

[39] Zhou H, Jiang T, Li Q et al. US-Based Deep Learning Model for Differentiating Hepatocellular Carcinoma (HCC) From Other Malignancy in Cirrhotic Patients. Front Oncol 2021; 11: 672055. doi:10.3389/fonc.2021.672055

[40] Lehang G, Dan W, Huixiong X et al. CEUS-based classification of liver tumors with deep canonical correlation analysis and multi-kernel learning. Annu Int Conf IEEE Eng Med Biol Soc 2017; 2017: 1748–1751. doi:10.1109/EMBC.2017.8037181

[41] Guo L-H, Wang D, Qian YY et al. A two-stage multi-view learning framework-based computer-aided diagnosis of liver tumors with contrast enhanced ultrasound images. Clin Hemorheol Microcirc 2018; 69: 343–354. doi:10.3233/CH-170275

[42] Hu HT, Wang W, Chen LD et al. Artificial intelligence assists identifying malignant versus benign liver lesions using contrast-enhanced ultrasound. J Gastroenterol Hepatol 2021; 36: 2875–2883. doi:10.1111/jgh.15522

[43] Kondo S, Takagi K, Nishida M et al. Computer-Aided Diagnosis of Focal Liver Lesions Using Contrast-Enhanced Ultrasonography With Perflubutane Microbubbles. IEEE Trans Med Imaging 2017; 36: 1427–1437. doi:10.1109/TMI.2017.2659734

[44] Yiyi Q, Jun S, Xiao Z et al. Multimodal Ultrasound imaging-based diagnosis of liver cancers with a two-stage multi-view learning framework. Annu Int Conf IEEE Eng Med Biol Soc 2017; 2017: 3232–3235. doi:10.1109/EMBC.2017.8037545

[45] Ta CN, Kono Y, Eghtedari M et al. Focal Liver Lesions: Computer-aided Diagnosis by Using Contrast-enhanced US Cine Recordings. Radiology 2018; 286: 1062–1071. doi:10.1148/radiol.2017170365

[46] Wu K, Chen X, Ding M. Deep learning-based classification of focal liver lesions with contrast-enhanced ultrasound. Optik 2014; 125: 4057–4063. doi:10.1016/j.ijleo.2014.01.114

[47] Zhang H, Guo L, Wang D et al. Multi-Source Transfer Learning Via Multi-Kernel Support Vector Machine Plus for B-Mode Ultrasound-Based Computer-Aided Diagnosis of Liver Cancers. IEEE J Biomed Health Inform 2021; 25: 3874–3885. doi:10.1109/JBHI.2021.3073812

[48] Caleanu CD, Simion G, David C et al. A study over the importance of arterial phase temporal parameters in focal liver lesions CEUS based diagnosis. In: 2014 11th International Symposium on Electronics and Telecommunications (ISETC 2014): Timişoara, România, 14 – 15 November 2014 ; [conference proceedings. Piscataway, NJ: IEEE; 2014: 1–4. doi:10.1109/ISETC.2014.7010799

[49] Căleanu CD, Sîrbu CL, Simion G. Deep Neural Architectures for Contrast Enhanced Ultrasound (CEUS) Focal Liver Lesions Automated Diagnosis. Sensors (Basel) 2021; 21. doi:10.3390/s21124126

[50] Denis de Senneville B, Frulio N, Laumonier H et al. Liver contrast-enhanced sonography: computer-assisted differentiation between focal nodular hyperplasia and inflammatory hepatocellular adenoma by reference to microbubble transport patterns. Eur Radiol 2020; 30: 2995–3003. doi:10.1007/s00330-019-06566-1

[51] Hu HT, Kuang M, Lu MD et al. IDDF2019-ABS-0148 Focal liver lesion classification using a convolutional neural network-based transfer-learning algorithm on tri-phase images of contrast-enhanced ultrasound. In: Clinical Hepatology. BMJ Publishing Group Ltd and British Society of Gastroenterology 062019: A140.1–A140. doi:10.1136/gutjnl-2019-IDDFAbstracts.274

[52] Huang Q, Pan F, Li W et al. Differential Diagnosis of Atypical Hepatocellular Carcinoma in Contrast-Enhanced Ultrasound Using Spatio-Temporal Diagnostic Semantics. IEEE J. Biomed. Health Inform 2020; 24: 2860–2869. doi:10.1109/JBHI.2020.2977937

[53] Li W, Lv XZ, Zheng X et al. Machine Learning-Based Ultrasomics Improves the Diagnostic Performance in Differentiating Focal Nodular Hyperplasia and Atypical Hepatocellular Carcinoma. Front Oncol 2021; 11: 544979. doi:10.3389/fonc.2021.544979

[54] Liang X, Lin L, Cao Q et al. Recognizing Focal Liver Lesions in CEUS With Dynamically Trained Latent Structured Models. IEEE Trans Med Imaging 2016; 35: 713–727. doi:10.1109/TMI.2015.2492618

[55] Shiraishi J, Sugimoto K, Moriyasu F et al. Computer-aided diagnosis for the classification of focal liver lesions by use of contrast-enhanced ultrasonography. Med Phys 2008; 35: 1734–1746. doi:10.1118/1.2900109

[56] Sirbu CL, Simion G, Caleanu CD. Deep CNN for Contrast-Enhanced Ultrasound Focal Liver Lesions Diagnosis. In: 2020 14th International Symposium on Electronics and Telecommunications (ISETC): November 05–06, 2020, Timişoara, România : conference proceedings. Piscataway, NJ: IEEE; 2020: 1–4. doi:10.1109/ISETC50328.2020.9301116

[57] Streba CT, Ionescu M, Gheonea DI et al. Contrast-enhanced ultrasonography parameters in neural network diagnosis of liver tumors. World J Gastroenterol 2012; 18: 4427–4434. doi:10.3748/wjg.v18.i32.4427

[58] Sugimoto K, Shiraishi J, Moriyasu F et al. Computer-aided diagnosis of focal liver lesions by use of physicians' subjective classification of echogenic patterns in baseline and contrast-enhanced ultrasonography. Acad Radiol 2009; 16: 401–411. doi:10.1016/j.acra.2008.09.018

[59] Sugimoto K, Shiraishi J, Moriyasu F et al. Computer-aided diagnosis for contrast-enhanced ultrasound in the liver. World J Radiol 2010; 2: 215–223. doi:10.4329/wjr.v2.i6.215

[60] Zhou J, Pan F, Li W et al. Feature Fusion for Diagnosis of Atypical Hepatocellular Carcinoma in Contrast- Enhanced Ultrasound. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 2022; 69: 114–123. doi:10.1109/TUFFC.2021.3110590

[61] Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC 2009; 19: 203–211

[62] Wojciech Samek, Grégoire Montavon, Andrea Vedaldi et al. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning; Springer; 2019

[63] Tiyarattanachai T, Apiparakoon T, Marukatat S et al. The feasibility to use artificial intelligence to aid detecting focal liver lesions in real-time ultrasound: a preliminary study based on videos. Sci Rep 2022; 12: 7749. doi:10.1038/s41598-022-11506-z

[64] Nam D, Chapiro J, Paradis V et al. Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction. JHEP Rep 2022; 4: 100443. doi:10.1016/j.jhepr.2022.100443