REVIEW

MAYO CLINIC PROCEEDINGS: DIGITAL HEALTH

MAYO CLINIC

Deep Learning for Computed Tomography Assessment of Hepatic Fibrosis and Cirrhosis: A Systematic Review

Numan Kutaiba, MBChB; Ariel Dahan, MD; Mark Goodwin, BMBCh; Adam Testro, PhD; Gary Egan, PhD; and Ruth Lim, DMedSc

Abstract

Studies were identified using deep learning artificial intelligence methods for the analysis of computed tomography images in the assessment of hepatic fibrosis and cirrhosis. A systematic review was conducted in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies protocol to evaluate the accuracy of deep learning algorithms for this objective (PROSPERO CRD 42023366201). A literature search was conducted on Medline, Embase, Web of Science, and IEEE Xplore databases. The search was conducted with a timeline from January 1, 2000,through November 13, 2022. Our search resulted in 3877 studies for screening, which yielded 6 studies meeting our inclusion criteria. All studies were retrospective. Three studies performed internal validation, and 2 studies performed external validation. Four studies used image classification algorithms, whereas 2 studies used image segmentation algorithms to derive volumetric measurements of the liver and spleen. Accuracy of the algorithms was variable in diagnosing significant and advanced fibrosis and cirrhosis, with the area under the curve ranging from 0.63 to 0.97. Deep learning algorithms using computed tomography images have the potential to classify fibrosis stages. Heterogeneity in cohorts and methodologies limits the generalizability of these studies.

 ■ Mayo Clin Proc Digital Health 2023;1(4):574-585

From the Department of Radiology (N.K., A.D., M.G., R.L.) and Department of Gastroenterology (A.T.), Austin Health, Heidelberg VIC, Australia; The University of Melbourne, Parkville, Australia (N.K., M.G., A.T., R.L.); and Monash Biomedical Imaging, Monash University, Clayton VIC, Australia (G.E.).

Chronic liver disease is responsible for considerable morbidity and mortality worldwide.[1] The dynamic and overlapping nature of various etiologies such as viral hepatitis, alcoholic liver disease, and nonalcoholic fatty liver disease creates further challenges in risk stratification, treatment, and surveillance.[1,2] The spectrum of liver disease starts with a continuous process of injury, inflammation, destruction, and regeneration of liver parenchyma leading to fibrosis, which can progress to cirrhosis. Data on the global prevalence of cirrhosis are limited, but estimates of its prevalence range from 0.7% to 1.3%.[3,4] Liver fibrosis estimates range from 3.6% to 13.0% with a high prevalence of silent liver disease among individuals without known liver disease.[5] Liver fibrosis stage represents an important predictor of clinical outcomes and mortality in chronic liver disease. For example, in patients with nonalcoholic fatty liver disease, the risk of liver-related mortality increases exponentially with each fibrosis stage, reaching a hazard ratio of 15:1 for end-stage fibrosis (cirrhosis).[6] Although the gold standard for determining liver fibrosis stage is biopsy, noninvasive tests and serological biomarkers are often used instead for risk stratification and monitoring.[7,8] Transient elastography (FibroScan), a technique using ultrasound waves to measure liver stiffness, has been widely applied for hepatic fibrosis staging with high accuracy for the detection of cirrhosis. However, its accuracy is still somewhat limited in assessing intermediate fibrosis stages (precirrhosis) and may be influenced by factors such as body habitus and serological biomarkers.[9,10]

Imaging represents an essential component of the screening, diagnosis, and monitoring of patients with suspected or known liver disease. Ultrasound is the most commonly used imaging modality given its availability and affordability, whereas

Mayo Clin Proc Digital Health ■ December 2023;1(4):574-585 ■ https://doi.org/10.1016/j.mcpdig.2023.08.008
www.mcpdigitalhealth.org

computed tomography (CT) and magnetic resonance imaging (MRI) are reserved for more specific indications.[11] Advanced imaging techniques such as MR elastography provide high accuracy in fibrosis staging but are limited to academic settings owing to the requirement of additional equipment and expertise.[7,8,12] The widespread availability of CT has led to its increased use in answering specific questions regarding decompensation, focal liver lesions, and sudden presentations.[13,14] Computed tomography scans are usually performed with intravenous iodinated contrast and provide findings on the status of liver parenchyma, such as morphologic changes indicating cirrhosis, characterization of focal liver lesions, and the status of portal hypertension. The latter may include portal venous thrombosis, portosystemic shunts, splenomegaly, and ascites. However, subjective interpretation of CT in assessment of cirrhosis is of limited accuracy and suffers from interobserver variability.[15] Furthermore, subjective assessment of fibrosis (precirrhosis) is less accurate given that the radiological changes are more subtle.[16,17] To overcome this limitation, quantitative morphology-based methods have been suggested, including manual and semiautomated measurements acquired from CT images. These include 2D measurements (caudate-to-right lobe ratios, hepatic vein diameters and ratios), 3D volumetric assessment (total liver and spleen volumes, liver segment volumes), and quantitative semiautomated assessment of liver surface nodularity. Such methods require additional measurements by experienced readers but are still subject to measurement bias, are often time consuming, and may require additional software processing.[18,19]

Artificial intelligence (AI) is a broad field that includes various automated methods to analyze imaging and nonimaging data. Deep learning is a subfield of AI that involves the use of neural networks to perform such tasks as image segmentation and classification. Such methods have been applied to CT and MRI images to perform volumetric segmentation of the liver and spleen and to assess and classify hepatic fibrosis.[20-22] Previous systematic reviews on the use of AI in hepatology evaluated the literature from a broad perspective covering diffuse and focal liver disease and included all imaging modalities (ultrasound, CT, and MRI).[23-26] Although such general reviews are important, the technical details related to each imaging modality were not specifically explored. Such details are important in AI methods and can considerably affect the training, performance, and applicability of developed algorithms. In this systematic review, we aim to assess the performance of deep learning methods in staging hepatic fibrosis and cirrhosis using CT images.

METHODS

The study was conducted on the basis of the Preferred Reporting Items for Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies checklist.[27] A meta-analysis was not performed because of the expected high heterogeneity of studies.

Search Strategy

A literature search was conducted on Medline, Embase, Web of Science, and IEEE Xplore databases. The search was conducted with a timeline from January 1, 2000, through November 13, 2022. The search keywords were "liver diseases," "liver fibrosis," "hepatic fibrosis," "artificial intelligence," "algorithms," "neural networks," "deep learning," and "computed tomography." A detailed literature search methodology is provided in the Supplementary File (available online at https://www.mcpdigitalhealth.org/).

Inclusion criteria were as follows: (1) studies that evaluated deep learning algorithm(s) for volumetric assessment of liver and spleen or for image classification for the purpose of hepatic fibrosis staging; (2) used CT images; (3) included histopathology or elastography as reference standards; (4) published full-text in English; and (5) provided an accuracy measure of the algorithm(s). Abstracts and nonhuman studies were excluded.

Two reviewers (NK and AD, anonymized for manuscript review) independently screened the results: NK, a hepatobiliary subspecialist radiologist, 5 years post fellowship, and AD, a final year radiology trainee. Results were screened using the Covidence systematic review web-based platform (www.covidence.org). Discrepant results were resolved by consensus. The study was registered with the

international prospective register of systematic reviews (PROSPERO CRD 42023366201).

Data Extraction

Extracted data from included studies were collated into a standard data sheet. Data included first author, year of publication, country, details of study population, number of participants, size of training and validation datasets, use of internal and external validation datasets, augmentation methods, reference standard, outcome measures, and results. Technical details related to CT imaging were extracted where available and included acquisition parameters, reconstructed slice thickness, use of intravenous iodinated contrast, and phase of contrast imaging. Details on preprocessing of imaging before deep learning development, such as segmentation, cropping, use of single slices or volumes, and labeling were also extracted. Finally, the availability of developed algorithms and datasets was assessed.

Outcome Measures

Outcome measures included area under the receiver operating characteristic (AUROC) curve. For studies using volumetric assessment of the liver and spleen to stage hepatic fibrosis, accuracy measures for segmentation, including Dice similarity coefficient were also assessed.[28]

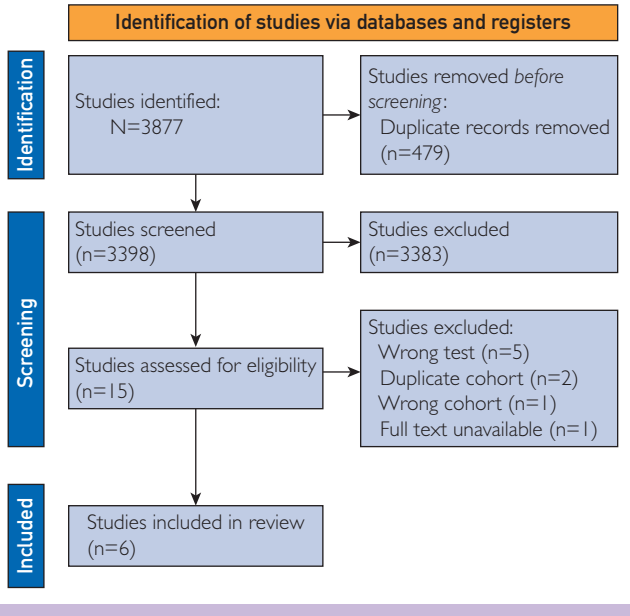

FIGURE 1. Preferred Reporting Items for Systematic Review and Meta-Analysis of Diagnostic Test Accuracy flowchart of search and inclusion process.

Reference Standard

Studies using histopathology and elastography as reference standards for hepatic fibrosis were included. Binary classifications of significant vs nonsignificant fibrosis, advanced vs nonadvanced fibrosis, and cirrhosis vs noncirrhosis were evaluated. For histopathology, where the meta-analysis of histological data in viral hepatitis score or the New Inuyama classification were used, significant fibrosis was regarded as F2 or higher, advanced fibrosis as F3 or higher, and cirrhosis as F4.[29,30] Where Ishak score was used, significant fibrosis was regarded as score 3 or higher, advanced fibrosis as score 4 or higher, and cirrhosis as score 6.[31] Other histopathology reporting systems, such as Knodell hepatic activity index (HAI) were reported as presented in the corresponding studies.[32] If elastography was used, thresholds for significant fibrosis, advanced fibrosis, and cirrhosis were reported as presented.

Quality Assessment

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was used to assess the risk of bias for the included studies. Modifications of questions to allow assessment of AI studies was made on the basis of previously suggested QUADAS-AI.[33,34] Questions were related to 4 domains and are as follows: (1) subject selection; (2) index test; (3) reference standard; and (4) work flow. Studies were rated as low, high, or unclear, reflecting the risk of bias and concerns regarding applicability.

RESULTS

We screened 3,877 studies from our search results and identified 6 studies meeting our inclusion criteria (Figure 1). The PRISMA-DTA checklist is presented in Supplementary Materials.

Characteristics of Selected Studies

Six studies met our inclusion criteria and included 10,408 individuals, including 7331 men (70.4%), with mean ages of 44-66

years.[16,35-39] Four studies were from Asia (2 from Korea, 1 from Japan, and 1 from China); 1 was from Europe (Netherlands); and 1 from North America (United States). Table 1 summarizes the study characteristics, and Table 2 summarizes the technical details, including CT parameters, image analysis, and processing. Tables 3 and 4 summarize performance scores. All studies were retrospective, and 5 studies evaluated algorithm performance on validation datasets, including 2 studies with external validation.[16,35] The study by Li et al[36] did not report on algorithm performance in a separate validation dataset. Yasaka et al[38] used 496 CT scans from 286 patients and Yin et al[39] included 2 patients twice as they underwent repeat transplantation. The remaining studies used 1 CT scan per patient. None of the datasets in the included studies were derived from public datasets and none of them have been made publicly available.

Included studies fall into 2 types of approaches in fibrosis staging: (1) studies with classification algorithms;[16,36,38,39] and (2) studies with liver and spleen segmentation and volume measurements.[35,37]

Studies With Classification Algorithms

Four studies used an image classification approach. Choi et al[16] performed the largest multicenter study with 8352 patients from 4 centers and evaluated their algorithm on external datasets. They randomly selected 50 CT scans to develop an initial liver segmentation algorithm, which was used to segment the liver for the entire dataset as a first step. The Dice similarity coefficient for this segmentation algorithm was 0.92. A classification algorithm for fibrosis staging was then developed and applied to the dataset as a second step.[16] Yin et al[39] also used volumetric data but applied the soft tissue window density range (−150 to +225 Hounsfield units) to the CT slices as a first step rather than creating a segmentation algorithm to exclude irrelevant densities from the imaging field. The processed CT slices were then used for classification of fibrosis stages. They provided heat maps for their algorithm, which highlighted liver and spleen parenchyma as contributors to their neural network predictions. Limitations of their study included a lack of external validation and a relatively small sample size of 252 patients, including 133 CT scans from a trauma setting considered as normal livers (F0) without a liver biopsy.[39] The study by Yasaka et al[38] used a manually selected single slice of the liver, which was then manually cropped as a JPEG image of the anterior left lobe of the liver. The images were used to develop and test a classification algorithm. In the study by Li et al,[36] 3 adjacent axial slices were manually selected at the expected level of liver biopsy in the eighth or ninth intercostal spaces, which were then manually contoured around the liver. The stacks of images were then used for a classification algorithm. Their study was the only one without a validation dataset (internal or external)—an important limitation. Unlike other studies, their study used noncontrast CT scans, whereas others exclusively used portal venous phase contrast-enhanced CT images.[36]

Most of these studies described data augmentation. Choi et al[16] used multiple techniques including rotating the original CT images (−15 to +15°) or adding Gaussian noise on a random basis. They applied these techniques to fibrosis stages F1, F2, and F3 only instead of the entire cohort acknowledging a class imbalance in these subgroups compared with F0 and F4. Yasaka et al[38] augmented their entire cohort by changing the location of cropping and by rotation (through 5, 90, 180, 270, and 355°) and adding Gaussian noise. Li et al[36] augmented their data with random horizontal and vertical flipping, rotation (within 10°), image magnification, and horizontal and vertical shift. Yin et al[39] did not describe any augmentation of their imaging data before neural network training.

Studies With Segmentation Algorithms

Two studies used a segmentation approach of the liver and spleen to derive volume measurements. These measurements were then used to derive numeric indices and ratios to classify fibrosis stages. Son et al[37] used CT scans of 558 patients from 3 centers, which were part of the study by Choi et al.[16] They used the segmentation algorithm developed in the study by Choi et al and reported on manual checking and editing of segmentation results that required additional reporting time (9.1±5.0 min, range 2-43 min). The liver

TABLE 1. General Characteristics of Included Studies[a]

Reference, Year	Country	Design	Site(s)	Recruitment Period	No. of Patients	Age (y), Mean ± SD	Men	F0/F1/F2/F3/F4	Etiology of Liver Disease	Reference	Nonimaging Data Used
Choi et al,[16] 2018	Korea	Retrospective	Multicenter	2007-2016	8352	44.2±14.7[b] 51.5±13.3[c]	5896 (70.6%)	3475/222/445/633/3577	40.1% HBV and 45.2% no disease[b]	Histopathology	No
Yasaka et al,[38] 2018	Japan	Retrospective	Single	2014-2015	286 (496 CTs)	66.2±11.6[b] 66.1 ±11.6[c]	354 (71.4%)	142/45/70/82/157[d]	N/A	Histopathology	Yes: age and sex
Li et al,[36] 2020	China	Retrospective	Single	2017-2019	347	45 (35-54)[e]	213 (61.4%)	74/59/35/20/159	57% HBV and 21.1% other	Histopathology	No
Yin et al,[39] 2021	Netherlands	Retrospective	Single	2006-2018	252 (254 CTs)	59 (48-65)[e]	140 (55.6%)	134/8/10/18/82	2.5% HBV, 53% no liver disease and 22% EtOH	Histopathology F0 group (53%) were trauma patients (no histopathology)	No
Son et al,[37] 2020	Korea	Retrospective	Multicenter	2007-2017	558	48.7 (13.1)	284 (50.9%)	78/67/109/86/218	Mixed; viral hepatitis 41.2% and no liver disease 8.1%	Histopathology	Height and weight[f]
Lee et al,[35] 2022	United States	Retrospective	Multicenter	Dataset1 2000 to 2016 Dataset2 2001-2021	613 Dataset 1—406 Dataset 2—207	Dataset1 50 (44-56)[e] Dataset2 50 (41-57)[e]	444 (72.4%)	Dataset1 47/62/90/59/148 Dataset2[g]	HCV for Dataset1 Mixed for dataset2, including 38% HCV, 12% HBV, and 3% HDV/HBV	Histopathology	No

[a]CT, computed tomography; HBV, hepatitis B virus, HCV, hepatitis C virus; HDV, hepatitis D virus; EtOH, alcohol; N/A, not available
[b]For development dataset
[c]For validation dataset
[d]Number of CT scans
[e]Median (IQR)
[f]Height and weight used for calculation of body surface area to standardize liver and spleen volumes.
[g]Dataset2 used a different fibrosis staging system.

TABLE 2. Technical Characteristics of Included Studies

Reference, Year	Acquisition Details Provided	Slice Thickness	Phase of CT Imaging	Single Slice vs Volumetric	Preprocessing	Augmentation
Choi et al,[16] 2018	Yes	5mm in 71.2%	Portal venous	Volumetric	Normalization and resizing Automated liver segmentation before classification algorithm	Yes for F1, F2 and F3 subgroups
Yasaka et al,[38] 2018	Yes	0.5 mm or 0.625 mm	Portal venous	Single axial slice	Manually cropped image of anterior left lobe, resized JPEG	Yes
Li et al,[36] 2020	Yes	3 mm	Noncontrast	Three adjacent axial slices	Manually selected slices Manually segmented Density width 350 HU and window 40 HU	Yes
Yin et al,[39] 2021	Not mentioned	Not mentioned	Portal venous	Volumetric	Automated volumetric segmentation of liver Density range applied −125 to +225 HU	Not described
Son et al,[37] 2020	Yes	3 mm in 27% 5 mm in 72%	Portal venous	Volumetric	Same as automated segmentation in study by Choi et al[16]	N/A
Lee et al,[35] 2022	Yes	Dataset1: 5mm in 98% Dataset2: all 5 mm.	Portal venous	Volumetric	Soft tissue density range −150 to 240 HU	Yes

CT, computed tomography; N/A, not available.

and spleen volumes were standardized to body surface area to derive standardized liver and spleen volumes rather than absolute volumes. They also calculated a liver-to-spleen volume ratio using absolute volumes. These standardized indices were then correlated with fibrosis stage.[37] Lee et al[35] used 2 segmentation algorithms to segment the 8 liver Couinaud segments and the spleen. From these segmentations, total liver volume, spleen volume, and liver segmental volume ratio (segments I-III divided by segments IV-VIII) were derived and correlated with fibrosis stages.

Reference Standard

Histopathology was used as a reference standard in the studies (Supplemental Table 1, available online at https://www.mcpdigitalhealth.org/). In addition to histopathology, Yin et al[39] used 133 CT scans from patients from a trauma setting as controls (F0) without histopathology, representing 53% of their cohort.[39] Most studies used

TABLE 3. Performance of Classification Algorithm Studies[a]

Reference, Year	Training or Validation %	Validation	Validation dataset AUROC Sensitivity Specificity Significant Fibrosis	Advanced Fibrosis	Cirrhosis
Choi et al,[16] 2018	90/10	Internal and external	0.96 85.50% 89.90%	0.97 94.60% 95.40%	0.95 84.60% 96.60%
Yasaka et al,[38] 2018	80/20	Internal	0.74 76% 68%	0.76 75% 65%	0.73 75% 57%
Li et al,[36] 2020[b]	80/20	N/A	0.9 83% 59%	0.94 79% 79%	0.97 81% 89%
Yin et al,[39] 2021	Not described	Internal	0.92 83.00% 91.70%	0.89 79.50% 88.20%	0.88 75.10% 86.50%

[a]AUROC, area under the receiver operating characteristic curve; N/A, not available
[b]Li et al[36] used 5-fold cross validation across the cohort with 80/20 split. The mean AUROC, sensitivity and specificity as reported for 5-fold cross validation.

TABLE 4. Performance (AUROC) of Segmentation Studies[a,b]

Reference, Year	Volumetric Parameter	Significant Fibrosis	Advanced Fibrosis	Cirrhosis
Son et al,[37] 2020[c]	-		Whole dataset	
	sVolL	-	0.63	0.68
	sVolS	-	0.82	0.84
	sVolL / sVolS	-	0.82	0.85
Lee et al,[35] 2020[d]	-		Dataset1	
	Liver	0.57	-	0.46
	Spleen	0.77	-	0.85
	LSVR	0.72	-	0.79
	-		Dataset2	
	Liver	0.49	-	0.46
	Spleen	0.66	-	0.65
	LSVR	0.63	-	0.75

[a]AUROC, area under receiver operating characteristic curve; LSVR, liver segmental volume ratio; N/A, not available
[b]sVolL and sVolS represent liver and spleen volumes standardized to body surface area, respectively.
[c]Did not report on segmentation accuracy.
[d]Reported Dice similarity coefficients of 0.98 and 0.95, respectively for whole liver and spleen.

the meta-analysis of histological data in viral hepatitis score for fibrosis staging with F2-F4, F3-F4, and F4 indicating significant fibrosis, advanced fibrosis, and cirrhosis, respectively. Son et al[37] further divided F4 into compensated (F4C) and decompensated (F4D) subgroups. Lee et al[35] used meta-analysis of histological data in viral hepatitis scores for their training Dataset1 and reported on Ishak and Knodell HAI scores in their validation Dataset2 with cut-offs for significant fibrosis and cirrhosis. In their study, they referred to stages F2-F4 as advanced fibrosis rather than significant fibrosis.[35] This is presented accordingly in Table 4.

Subgroups

All studies included patients with different fibrosis stages. Yasaka et al[38] did not report on aetiologies of liver disease in their cohort. Other studies from Asia had higher proportions of patients with hepatitis B virus compared with the studies from the Netherlands and the United States. The Netherlands cohort reported by Yin et al[39] comprised 134 healthy controls and 119 patients with liver disease, of whom 38% had unknown etiology, 22% were alcohol-related, and 11% reported viral hepatitis.[39] The training dataset in the study by Lee et al[35] exclusively comprised patients with hepatitis C virus, but their validation dataset was of mixed etiologies, including hepatitis C virus (38%), steatohepatitis (27%), and hepatitis B virus (12%).[35] In all studies, the proportions of patients within each fibrosis stage were weighted heavily at the ends of the distribution (F0 and F4) (Table 1 and Figure 2).

Performance in Fibrosis Staging

Three of the 4 classification algorithm studies reported high AUROC (0.88-0.97) for significant and advanced fibrosis and for cirrhosis.[16,36,39] Yasaka et al[38] reported moderate AUROC (0.73-0.76).[38] For volumetric indices in the 2 relevant studies, splenic volume (absolute and standardized to body surface area) revealed moderate AUROC (0.65-0.85). Liver-to-spleen volume ratio and liver segmental volume ratio also revealed moderate results, whereas liver volume (absolute and standardized to body surface area) revealed poor results.[35,37] Tables 3 and 4 summarize performance results.

Quality Assessment

All included studies reported ethical approval statements. All studies included conflict of interest disclosures and funding declaration statements.

Mayo Clin Proc Digital Health ■ December 2023;1(4):574-585 ■ https://doi.org/10.1016/j.mcpdig.2023.08.008
www.mcpdigitalhealth.org

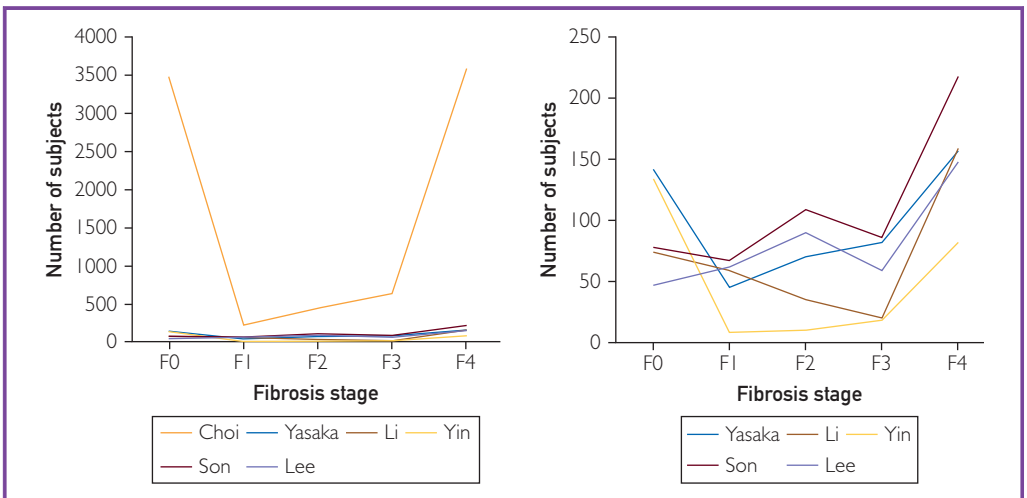

FIGURE 2. Left—distribution of patients across fibrosis stages for all studies. Right—distribution of patients across fibrosis stages after removing, Choi et al.[16]

Bias

Three studies that did not perform external validation[37-39] received unclear risk of bias scores, and 1 study with no internal or external validation[36] received a high risk of bias for the index test. A high risk of bias regarding patient flow and index test timing was attributed to 1 study owing to the aforementioned use of control patients without previous histopathological liver data as more than half the studied cohort.[39]

Applicability

Three studies received an unclear risk of bias regarding the index test[36,38,39] and 1 study received an unclear risk of bias regarding the reference standard.[39] A quality assessment summary and scores for each study are provided in Supplemental Figure and Supplemental Tables 2 and 3, available online at https://www.mcpdigitalhealth.org/.

DISCUSSION

Accurate staging of hepatic fibrosis requires the use of an invasive procedure (liver biopsy) or advanced imaging (MR elastography).[7,8] Such tools are relatively expensive, not widely available, and, in the case of liver biopsy, carry a risk of morbidity and mortality.[40] Therefore, noninvasive tools have been increasingly used in clinical practice for staging hepatic fibrosis. With the increasing availability of CT, harnessing information within images beyond routine radiologist interpretation has been a subject of AI research throughout the past decade. Our review demonstrates that AI algorithms using CT images can help in staging hepatic fibrosis in 2 domains: (1) image classification; and (2) volumetric segmentation of the liver and spleen.

Image classification algorithms require processing before use in training neural networks. In the study by Choi et al,[16] the first step was segmenting the liver from volumetric CT data.[16] Yin et al[39] used volumetric CT data but did not segment the liver. Instead, they applied a density range to preprocess the images to remove unnecessary data (eg, air, fat, and bone) before training their network. This approach used structures, in addition to the liver, within the imaging field (specifically the spleen) in their algorithm predictions.[39] Studies by Yasaka et al[38] and Li et al[36] used manually selected slices. Yasaka et al[38] cropped a single slice to include the anterior margin of the liver (without segmenting the liver in the image) to focus the algorithm on the liver capsule because of the known association of liver capsule nodularity with fibrosis stage. Li et al[36] selected 3 adjacent slices of the CT stack and then manually contoured the liver on these slices. In the latter 2 studies, manual handling of images required experienced radiologists. It is not known whether

including the anterior left lobe of the liver and adjacent fat in segmentation of the liver would result in improved prediction of fibrosis stages. Quantifying liver surface nodularity with semiautomated methods for fibrosis staging has shown good accuracy in a few studies.[41] In a study of hepatitis C virus patients, liver surface nodularity and a clinical marker (fibrosis-4 index) were similar to more advanced quantification methods such as volumetric measurements and texture analysis.[42] The cohort of this study is the original cohort from the study by Lee et al.[35] Training algorithms on all components of the CT scan, such as the approach by Yin et al,[39] may result in improved accuracy of classification algorithms, particularly in more advanced stages of liver disease (splenomegaly, portosystemic shunts, and ascites). However, when there is already obvious portal hypertension or decompensation, fibrosis staging is of dubious clinical value.

Two studies evaluated segmentation of the entire liver and spleen to provide volume measurements that could then be used for hepatic fibrosis staging.[35,37] Previous studies on this topic using manual and semiautomated approaches found that liver-to-spleen ratio, liver segmental volume ratio, and spleen volume alone can be used for staging hepatic fibrosis.[42-44] Son et al [37]did not report on segmentation accuracy in their study but reported on additional time for review of segmentation masks generated by the deep learning algorithm (average 9.1±5.0 min per case). Subsequently, their group improved their algorithm in 2 separate studies (<5% error) with Dice similarity coefficients of 0.973 and 0.974, respectively for the liver and spleen, and a processing time of 33 seconds.[20,45] Lee et al,[35] in a subset of 70 patients with manual measurements, reported Dice similarity coefficients of 0.98 and 0.95, respectively for the whole liver and spleen, and 0.91 and 0.96, respectively for left and right lobes of liver (Couinaud segments I, II, and III and segments IV, V, VI, VII, and VIII).

The studies by Yasaka et al[38] and Li et al[36] recruited patients over 2 and 3 years respectively. However, the remainder of studies recruited patients over periods of 10 years or more. Such long periods of recruitment result in a variety of patients with different liver disease etiologies undergoing CT imaging using various software and hardware technologies. The indications for liver biopsies and the distribution of liver disease etiologies have changed considerably over the past 2 decades, with fewer biopsies performed for viral hepatitis indications currently compared with those in the past.[46] There is also a growing proportion of nonalcoholic steatohepatitis patients in different countries,[47,48] which are not necessarily appropriately represented in the studies included in this review. Furthermore, developments in CT technology have resulted in a change in how the images are reconstructed and their quality.[49] The effect of such developments on the accuracy of image classification algorithms trained on older technology is unknown. Using volumetric measurements of the liver and spleen in fibrosis staging means that changes in image quality should not greatly affect volume measurements. One advantage of image segmentation and volumetric measurements over image classification approaches is the improvement of fibrosis classification on the basis of known morphological changes in liver disease.[35] However, different etiologies of liver disease may lead to different morphological changes of the liver in different stages of chronic liver disease.[50,51] An additional benefit of volumetric measurements in chronic liver disease is their value in predicting clinical outcomes in patients with established cirrhosis.[52,53] Image classification approaches have been described as "black boxes" and are generally considered less trustworthy compared with segmentation approaches.[54] However, attempts to provide explainable or interpretable AI approaches (eg, providing heat maps) to show which pixels or regions of images were used by a classification algorithm to reach a specific decision (class) may overcome this trustworthiness issue.[39,54]

Additional information available on CT images has not been fully explored using deep learning methods for the purpose of liver fibrosis staging. For example, using manual measurements of liver enhancement and semiautomated volumetry on multiphase CT, Tago et al[55] have shown good accuracy in fibrosis staging. Applying deep learning methods to segmentation of the liver on multiphase CT may improve fibrosis staging accuracy

compared with using noncontrast or portal venous phase images only. Furthermore, applying deep learning methods for segmentation of other regions in the imaging field, such as the skeletal muscle area, may result in improved accuracy in fibrosis staging and detection of cirrhosis given its association with sarcopenia.[56] A deep learning approach for the identification of ascites or large portosystemic shunts on CT images could also prove beneficial, but is yet to be tested in this setting.

Incorporating nonimaging data into imaging deep learning algorithms can result in better accuracy.[57] Yasaka et al[38] incorporated age and sex into their algorithm. However, the accuracy of their algorithm using only imaging data was not reported. Son et al[37] standardized liver and spleen volumes to body surface area using a height and weight formula but did not include other data such as age or sex.[37] The remainder of studies did not incorporate nonimaging data. Previous studies incorporating imaging and nonimaging data for fibrosis staging found promising results.[42] However, real-world application of these systems is challenged by the not infrequent absence of some relevant clinical data. To assess hepatic fibrosis of the liver using available clinical data (ie, no imaging data), Blanes-Vidal et al[58] built 6 separate AI models reporting a relatively high performance for excluding significant fibrosis in a screening population. Therefore, incorporation of a variety of inputs including imaging, demographic characteristics, and laboratory results should ideally be explored in future studies. Choi et al[16] and Son et al[37] found that their models were superior to serum fibrosis tests (aspartate aminotransferase to platelet ratio index and fibrosis-4 index). However, other studies did not perform such comparisons, and none of the included studies performed comparisons to other modalities such as transient elastography.

None of the datasets in the included studies are open-source. In addition, there are no publicly available CT datasets for chronic liver disease without cancer. The HCC-TACE-Seg dataset from the Cancer Imaging Archive includes cirrhotic and noncirrhotic patients, but all cases have hepatocellular carcinoma.[59] Other CT datasets include healthy individuals (eg, CHAOS)[60] or liver lesions (eg, CT-ORG).[61] The lack of availability of such datasets limits independent validation. Furthermore, multimodality imaging approaches such as incorporating ultrasound or MRI with CT may improve the accuracy of algorithms using a single modality.

Our systematic review has some limitations. The heterogeneity of the included studies limited our ability to perform a meaningful meta-analysis of performance results. Our review focused on CT, which resulted in a small number of included studies but allowed for more in-depth exploration of important cohort characteristics and technical details across studies.

CONCLUSION

Deep learning algorithms using CT images have the potential to classify fibrosis stages. High heterogeneity, the retrospective nature of the studies, and the lack of external validation limit the generalizability of these studies. Sharing datasets that include cohorts of patients representative of the current chronic liver disease population would greatly enhance and expedite advancements in this field.

POTENTIAL COMPETING INTERESTS

The authors declare no financial or ethical conflicts of interest regarding the contents of this submission.

SUPPLEMENTAL ONLINE MATERIAL

Supplemental material can be found online at https://www.mcpdigitalhealth.org/. Supplemental material attached to journal articles has not been edited, and the authors take responsibility for the accuracy of all data.

Abbreviations and Acronyms: **AI,** artificial intelligence; **AUROC,** area under receiver operating characteristic curve; **CT,** computed tomography; **MRI,** magnetic resonance imaging

Correspondence: Address to Numan Kutaiba, MBChB, Department of Radiology, Austin Health, 145 Studley Road, Heidelberg VIC 3084, Australia (nkutaiba@gmail.com).

ORCID

Numan Kutaiba: https://orcid.org/0000-0003-4627-9847

REFERENCES

1. Cheemerla S, Balakrishnan M. Global epidemiology of chronic liver disease. *Clin Liver Dis*. 2021;17(5):365-370.
2. Paik JM, Golabi P, Younossi Y, Mishra A, Younossi ZM. Changes in the global burden of chronic liver diseases from 2012 to 2017: the growing impact of NAFLD. *Hepatology*. 2020;72(5):1605-1616.
3. Huang DQ, Terrault NA, Tacke F, et al. Global epidemiology of cirrhosis—aetiology, trends and predictions. *Nat Rev Gastroenterol Hepatol*. 2023;20(6):388-398.
4. Vento S, Cainelli F. Chronic liver diseases must be reduced worldwide: it is time to act. *Lancet Glob Health*. 2022;10(4):e471-e472.
5. Caballería L, Pera G, Arteaga I, et al. High prevalence of liver fibrosis among European adults with unknown liver disease: a population-based study. *Clin Gastroenterol Hepatol*. 2018;16(7):1138-1145.e5.
6. Ng CH, Lim WH, Hui Lim GE, et al. Mortality outcomes by fibrosis stage in nonalcoholic fatty liver disease: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2023;21(4):931-939.e5. https://doi.org/10.1016/j.cgh.2022.04.014.
7. Schambeck JPL, Forte GC, Gonçalves LM, et al. Diagnostic accuracy of magnetic resonance elastography and point-shear wave elastography for significant hepatic fibrosis screening: systematic review and meta-analysis. *PLOS ONE*. 2023;18(2):e0271572.
8. Selvaraj EA, Mózes FE, Jayaswal ANA, et al. Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: a systematic review and meta-analysis. *J Hepatol*. 2021;75(4):770-785.
9. Jiang W, Huang S, Teng H, et al. Diagnostic accuracy of point shear wave elastography and transient elastography for staging hepatic fibrosis in patients with non-alcoholic fatty liver disease: a meta-analysis. *BMJ Open*. 2018;8(8):e021787.
10. Nguyen-Khac E, Thiele M, Voican C, et al. Non-invasive diagnosis of liver fibrosis in patients with alcohol-related liver disease by transient elastography: an individual patient data meta-analysis. *Lancet Gastroenterol Hepatol*. 2018;3(9):614-625.
11. Gerstenmaier JF, Gibson RN. Ultrasound in chronic liver disease. *Insights Imaging*. 2014;5(4):441-455.
12. Kennedy P, Wagner M, Castéra L, et al. Quantitative elastography methods in liver disease: current evidence and future directions. *Radiology*. 2018;286(3):738-763.
13. Angeli P, Bernardi M, Villanueva C, et al. EASL Clinical Practice Guidelines for the management of patients with decompensated cirrhosis. *J Hepatol*. 2018;69(2):406-460.
14. Nadarevic T, Giljaca V, Colli A, et al. Computed tomography for the diagnosis of hepatocellular carcinoma in chronic advanced liver disease. *Cochrane Database Syst Rev*. 2019;2019(6):CD013362.
15. Rubino JG, Nasirzadeh AR, van der Pol CB, Dhindsa K, Chung AD. Quantitative and qualitative liver CT: imaging feature association with histopathologically confirmed hepatic cirrhosis. *Abdom Radiol (NY)*. 2022;47(7):2314-2324.
16. Choi KJ, Jang JK, Lee SS, et al. Development and validation of a deep learning system for staging liver fibrosis by using contrast agent—enhanced CT images in the liver. *Radiology*. 2018;289(3):688-697.
17. Venkatesh SK, Yin M, Takahashi N, Glockner JF, Talwalkar JA, Ehman RL. Non-invasive detection of liver fibrosis: MR imaging features vs. MR elastography. *Abdom Imaging*. 2015;40(4):766-775.
18. Huber A, Ebner L, Heverhagen JT, Christe A. State-of-the-art imaging of liver fibrosis and cirrhosis: A comprehensive review of current applications and future perspectives. *Eur J Radiol Open*. 2015;2:90-100.
19. Smith AD, Porter KK, Elkassem AA, Sanyal R, Lockhart ME. Current imaging techniques for noninvasive staging of hepatic fibrosis. *AJR Am J Roentgenol*. 2019;213(1):77-89.
20. Kim DW, Ha J, Lee SS, et al. Population-based and personalized reference intervals for liver and spleen volumes in healthy individuals and those with viral hepatitis. *Radiology*. 2021;301(2):339-347.
21. Wang K, Mamidipalli A, Retson T, et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. *Radiol Artif Intell*. 2019;1(2):180022.
22. Yasaka K, Akai H, Kunimatsu A, Abe O, Kiryu S. Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid—enhanced hepatobiliary phase MR images. *Radiology*. 2018;287(1):146-155.
23. Anteby R, Klang E, Horesh N, et al. Deep learning for noninvasive liver fibrosis classification: a systematic review. *Liver Int*. 2021;41(10):2269-2278.
24. Decharatanachart P, Chaiteerakij R, Tiyarattanachai T, Treeprasertsuk S. Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis. *BMC Gastroenterol*. 2021;21(1):10.
25. Wang L, Zhang L, Jiang B, Zhao K, Zhang Y, Xie X. Clinical application of deep learning and radiomics in hepatic disease imaging: a systematic scoping review. *Br JRadiol*. 2022;95(1136):20211136.
26. Radiya K, Joakimsen HL, Mikalsen KØ, Aahlin EK, Lindsetmo RO, Mortensen KE. Performance and clinical applicability of machine learning in liver computed tomography imaging: a systematic review. *Eur Radiol*. 2023;33(10):6689-6717.
27. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. 2018;319(4):388-396.
28. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index1. *Acad Radiol*. 2004;11(2):178-189.
29. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology*. 1996;24(2):289-293.
30. Ichida F, Tsuji T, Omata M, et al. New Inuyama classification; new criteria for histological assessment of chronic hepatitis. *Int Hepatol Commun*. 1996;6(2):112-119.
31. Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. *J Hepatol*. 1995;22(6):696-699.
32. Knodell RG, Ishak KG, Black WC, et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology*. 1981;1(5):431-435.
33. Sounderajah V, Ashrafian H, Rose S, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nat Med*. 2021;27(10):1663-1665.
34. Xu HL, Gong TT, Liu FH, et al. Artificial intelligence performance in image-based ovarian cancer identification: a systematic review and meta-analysis. *EClinicalmedicine*. 2022;53:101662.
35. Lee S, Elton DC, Yang AH, et al. Fully automated and explainable liver segmental volume ratio and spleen segmentation at CT for diagnosing cirrhosis. *Radiol Artif Intell*. 2022;4(5):e210268.
36. Li Q, Yu B, Tian X, Cui X, Zhang R, Guo Q. Deep residual nets model for staging liver fibrosis on plain CT images. *Int J Comput Assist Radiol Surg*. 2020;15(8):1399-1406.
37. Son JH, Lee SS, Lee Y, et al. Assessment of liver fibrosis severity using computed tomography—based liver and spleen volumetric indices in patients with chronic liver disease. *Eur Radiol*. 2020;30(6):3486-3496.
38. Yasaka K, Akai H, Kunimatsu A, Abe O, Kiryu S. Deep learning for staging liver fibrosis on CT: a pilot study. *Eur Radiol*. 2018;28(11):4578-4585.
39. Yin Y, Yakar D, Dierckx RAJO, Mouridsen KB, Kwee TC, de Haas RJ. Liver fibrosis staging by deep learning: a visual-based

explanation of diagnostic decisions of the model. *Eur Radiol*. 2021;31(12):9620-9627.
40. Howlett DC, Drinkwater KJ, Lawrence D, Barter S, Nicholson T. Findings of the UK national audit evaluating image-guided or image-assisted liver biopsy. Part II. Minor and major complications and procedure-related mortality. *Radiology*. 2013;266(1):226-235.
41. Heo S, Kim DW, Choi SH, Kim SW, Jang JK. Diagnostic performance of liver fibrosis assessment by quantification of liver surface nodularity on computed tomography and magnetic resonance imaging: systematic review and meta-analysis. *Eur Radiol*. 2022;32(5):3377-3387.
42. Pickhardt PJ, Graffy PM, Said A, et al. Multiparametric CT for noninvasive staging of hepatitis C virus–related liver fibrosis: correlation with the histopathologic fibrosis score. *AJR Am J Roentgenol*. 2019;212(3):547-553.
43. Obmann VC, Marx C, Hrycyk J, et al. Liver segmental volume and attenuation ratio (LSVAR) on portal venous CT scans improves the detection of clinically significant liver fibrosis compared to liver segmental volume ratio (LSVR). *Abdom Radiol (NY)*. 2021;46(5):1912-1921.
44. Pickhardt PJ, Malecki K, Hunt OF, et al. Hepatosplenic volumetric assessment at MDCT for staging liver fibrosis. *Eur Radiol*. 2017;27(7):3060-3068.
45. Ahn Y, Yoon JS, Lee SS, et al. Deep learning algorithm for automated segmentation and volume measurement of the liver and spleen using portal venous phase computed tomography images. *Korean J Radiol*. 2020;21(8):987-997.
46. Jain D, Torres R, Celli R, Koelmel J, Charkoftaki G, Vasiliou V. Evolution of the liver biopsy and its future. *Transl Gastroenterol Hepatol*. 2021;6:20.
47. Adams LA, Roberts SK, Strasser SI, et al. Nonalcoholic fatty liver disease burden: Australia, 2019-2030. *J Gastroenterol Hepatol*. 2020;35(9):1628-1635.
48. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol*. 2018;15(1):11-20.
49. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. *Jpn J Radiol*. 2019;37(1):73-80.
50. Kim I, Jang YJ, Ryeom H, et al. Variation in hepatic segmental volume distribution according to different causes of liver cirrhosis: CT volumetric evaluation. *J Comput Assist Tomogr*. 2012;36(2):220-225.
51. Ozaki K, Matsui O, Kobayashi S, Minami T, Kitao A, Gabata T. Morphometric changes in liver cirrhosis: aetiological differences correlated with progression. *Br J Radiol*. 2016;89(1059):20150896.
52. Kwon JH, Lee SS, Yoon JS, et al. Liver-to-spleen volume ratio automatically measured on CT predicts decompensation in patients with B viral compensated cirrhosis. *Korean J Radiol*. 2021;22(12):1985-1995.
53. Patel M, Tann M, Liangpunsakul S. CT-scan based liver and spleen volume measurement as a prognostic indicator for patients with cirrhosis. *Am J Med Sci*. 2021;362(3):252-259.
54. Reyes M, Meier R, Pereira S, et al. On the interpretability of artificial intelligence in radiology: challenges and opportunities. *Radiol Artif Intell*. 2020;2(3):e190043.
55. Tago K, Tsukada J, Sudo N, et al. Comparison between CT volumetry and extracellular volume fraction using liver dynamic CT for the predictive ability of liver fibrosis in patients with hepatocellular carcinoma. *Eur Radiol*. 2022;32(11):7555-7565.
56. Petta S, Ciminnisi S, Di Marco V, et al. Sarcopenia is associated with severe liver fibrosis in patients with non-alcoholic fatty liver disease. *Aliment Pharmacol Ther*. 2017;45(4):510-518.
57. Huang S-C, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *npj Digit Med*. 2020;3(1):136.
58. Blanes-Vidal V, Lindvig KP, Thiele M, Nadimi ES, Krag A. Artificial intelligence outperforms standard blood-based scores in identifying liver fibrosis patients in primary care. *Sci Rep*. 2022;12(1):2914.
59. Moawad AW, Morshid A, Khalaf AM, et al. Multimodality annotated hepatocellular carcinoma data set including pre- and post-TACE with imaging segmentation. *Sci Data*. 2023;10(1):33.
60. Kavur AE, Gezer NS, Barış M, et al. CHAOS Challenge – combined (CT-MR) healthy abdominal organ segmentation. *Med Image Anal*. 2021;69:101950.
61. Rister B, Yi D, Shivakumar K, Nobashi T, Rubin DL. CT-ORG, a new dataset for multiple organ segmentation in computed tomography. *Sci Data*. 2020;7(1):381.