

# Fully Automated and Explainable Liver Segmental Volume Ratio and Spleen Segmentation at CT for Diagnosing Cirrhosis

Sungwon Lee, MD, PhD • Daniel C. Elton, PhD • Alexander H. Yang, MD, MS • Christopher Koh, MD, MHSc • David E. Kleiner, MD, PhD • Meghan G. Lubner, MD • Perry J. Pickhardt, MD\* • Ronald M. Summers, MD, PhD\*

From the Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Department of Radiology and Imaging Sciences, National Institutes of Health Clinical Center, 10 Center Dr, Building 10, Room 1C224D, Bethesda, MD 20892-1182 (S.L., D.C.E., R.M.S.); Liver Diseases Branch, Intramural Research Program, National Institute of Diabetes and Digestive and Kidney Diseases (A.H.Y., C.K.), and Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, Md (D.E.K.); and Department of Radiology, University of Wisconsin School of Medicine & Public Health, Madison, Wis (M.G.L., P.J.P.). Received October 28, 2021; revision requested December 20; revision received July 27, 2022; accepted August 8. **Address correspondence to** R.M.S. (email: [rms@nih.gov](mailto:rms@nih.gov)).

Supported in part by the Intramural Research Program of the National Institutes of Health, Clinical Center, National Institute of Diabetes and Digestive and Kidney Diseases, and National Cancer Institute

\* P.J.P. and R.M.S. are co-senior authors.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2022; 4(5):e210268 • <https://doi.org/10.1148/ryai.210268> • Content codes: **AI** **GI**

**Purpose:** To evaluate the performance of a deep learning (DL) model that measures the liver segmental volume ratio (LSVR) (ie, the volumes of Couinaud segments I–III/IV–VIII) and spleen volumes from CT scans to predict cirrhosis and advanced fibrosis.

**Materials and Methods:** For this Health Insurance Portability and Accountability Act–compliant, retrospective study, two datasets were used. Dataset 1 consisted of patients with hepatitis C who underwent liver biopsy (METAVIR F0–F4, 2000–2016). Dataset 2 consisted of patients who had cirrhosis from other causes who underwent liver biopsy (Ishak 0–6, 2001–2021). Whole liver, LSVR, and spleen volumes were measured with contrast-enhanced CT by radiologists and the DL model. Areas under the receiver operating characteristic curve (AUCs) for diagnosing advanced fibrosis ( $\geq$  METAVIR F2 or Ishak 3) and cirrhosis ( $\geq$  METAVIR F4 or Ishak 5) were calculated. Multivariable models were built on dataset 1 and tested on datasets 1 (hold out) and 2.

**Results:** Datasets 1 and 2 consisted of 406 patients (median age, 50 years [IQR, 44–56 years]; 297 men) and 207 patients (median age, 50 years [IQR, 41–57 years]; 147 men), respectively. In dataset 1, the prediction of cirrhosis was similar between the manual versus automated measurements for spleen volume (AUC, 0.86 [95% CI: 0.82, 0.9] vs 0.85 [95% CI: 0.81, 0.89]; significantly noninferior,  $P < .001$ ) and LSVR (AUC, 0.83 [95% CI: 0.78, 0.87] vs 0.79 [95% CI: 0.74, 0.84];  $P < .001$ ). The best performing multivariable model achieved AUCs of 0.94 (95% CI: 0.89, 0.99) and 0.79 (95% CI: 0.71, 0.87) for cirrhosis and 0.8 (95% CI: 0.69, 0.91) and 0.71 (95% CI: 0.64, 0.78) for advanced fibrosis in datasets 1 and 2, respectively.

**Conclusion:** The CT-based DL model performed similarly to radiologists. LSVR and splenic volume were predictive of advanced fibrosis and cirrhosis.

Supplemental material is available for this article.

©RSNA, 2022

Chronic liver disease and cirrhosis are the ninth leading cause of death for men in the United States (1). Diagnosis is required to prevent complications and increase screening for hepatocellular carcinoma (2). Liver biopsy remains the reference standard in diagnosing cirrhosis, but because of the cost, invasiveness, and sampling errors (3,4), the use of noninvasive methods has become increasingly common in clinical practice.

Abdominal contrast-enhanced CT is a widely accepted method to evaluate not only liver cirrhosis but also the whole abdomen and pelvic region. CT findings of cirrhosis include increased nodularity of the liver surface, relative hypertrophy of the caudate and left lateral lobes, and portal hypertensive complications, such as splenomegaly (5).

There is a distinct change in the liver structure in cirrhosis, especially a decrease in the volumes of Couinaud

segments IV–VIII, and a compensatory enlargement in segments I–III (6), which is sometimes called “segmental redistribution.” The liver segmental volume ratio (LSVR) is a metric designed to measure this change of shape, calculated as the sum of the volumes of segments I–III divided by that of segments IV–VIII (6,7). Manually or semiautomatically measured LSVR and spleen volumes are good parameters for predicting cirrhosis and advanced fibrosis (areas under the receiver operating characteristic curve [AUCs] are 0.904 for LSVR and 0.920 for spleen volume in predicting cirrhosis), outperforming other two-dimensional parameters, such as caudate-to-right lobe ratio and spleen length (7,8). Although semiautomated software can speed up manual measurements, the measurements of the liver segments and spleen are impractical for daily practice because they are time-consuming (9) and prone to interreader errors (6,10). A fully automated method of

## Abbreviations

AUC = area under the receiver operating characteristic curve, DL = deep learning, HAI = histologic activity index, HCV = hepatitis C virus, LSVR = liver segmental volume ratio

## Summary

A deep learning–based model measuring the volume of liver Couinaud segments and spleen at contrast-enhanced CT performed similar to that of manual measurements in predicting histopathologic cirrhosis and advanced fibrosis.

## Key Points

- Fully automated measurements of splenic volume (area under the receiver operating characteristic curve [AUC], 0.85 [95% CI: 0.81, 0.89] vs 0.86 [95% CI: 0.82, 0.9]; significantly noninferior with  $P < .001$ ) and liver segmental volume ratio (LSVR) (AUC, 0.79 [95% CI: 0.74, 0.84] vs 0.83 [95% CI: 0.78, 0.87];  $P < .001$ ) performed similarly to that of manual measurements in predicting cirrhosis at contrast-enhanced CT.
- A multivariable model using the LSVR, splenic volume, and attenuation of the liver segments had an AUC of 0.94 (95% CI: 0.89, 0.97) for predicting cirrhosis in a hepatitis C cohort (a hold-out portion of the training data, dataset 1).
- In a group with multiple causes of cirrhosis (dataset 2), the automatic liver and spleen measurements had lower performance in predicting cirrhosis (AUC, 0.79 [95% CI: 0.71, 0.87] in dataset 2 vs 0.94 [95% CI: 0.89, 0.97] in dataset 1 for best performing multivariable model, which uses the spleen volume, LSVR, volumes, and attenuation of the liver Couinaud segments).

## Keywords

CT, Liver, Cirrhosis, Computer Applications–Detection/Diagnosis

measuring the liver segments and spleen volume that performs similarly to the manual measurements in predicting liver cirrhosis would be a quick, objective, and explainable method for diagnosing liver cirrhosis at CT.

Automated Couinaud (11) segmentations have previously been approached in a step-by-step manner (ie, whole liver segmentation, vessel segmentation, then Couinaud segmentation) (12). In this study, we take a deep learning (DL)–based approach to fully automate the process. The purpose of this study is to (a) obtain and evaluate a fully automated measurement of liver Couinaud segments and spleen volumes with use of abdominal contrast-enhanced CT and (b) predict the degree of liver cirrhosis or advanced fibrosis with use of automated measurements.

## Materials and Methods

### Patient Sample

This multi-institutional retrospective cohort study was compliant with the Health Insurance Portability and Accountability Act and was approved by the institutional review board of each institution. The need for additional signed informed consent was waived.

A summary of the study plan is illustrated in Figure 1. First, we developed a DL-based model that automatically segments the eight liver Couinaud segments and spleen. We then tested our DL model on two datasets from different institutions. Dataset 1 included 406 patients with positive results on antibody

testing for hepatitis C virus (HCV) who underwent abdominal contrast-enhanced CT between 2000 and 2016 and a liver biopsy within 1 year of CT at the University of Wisconsin Hospitals and Clinics. Liver biopsy in dataset 1 used the META-VIR scoring system (F0–F4) (13). Of note, dataset 1 is a subset of a previously published article that uses the data for manual measurements of liver and spleen volume in patients with HCV infection (469 patients) (14). However, in this study, the data are used to show automated measurements of liver Couinaud segments with use of a DL-based algorithm.

Dataset 2 included 207 patients who underwent abdominal contrast-enhanced CT between 2001 and 2021 and a liver biopsy within 1 year of CT at the National Institutes of Health Clinical Center. These patients had several underlying diseases, including viral hepatitis, steatohepatitis, and other cirrhosis-related diseases (full list in Table 1). Liver biopsy findings in dataset 2 were graded with both the Knodell histologic activity index (HAI) (0, 1, 3, and 4) (15) and Ishak staging system (modified Knodell system, 0–6) (16). CT scans that were not obtained with contrast material, did not include the full range of the liver, or were from patients who had undergone hepatectomy or splenectomy were excluded from datasets 1 and 2.

### Multidetector CT Technique

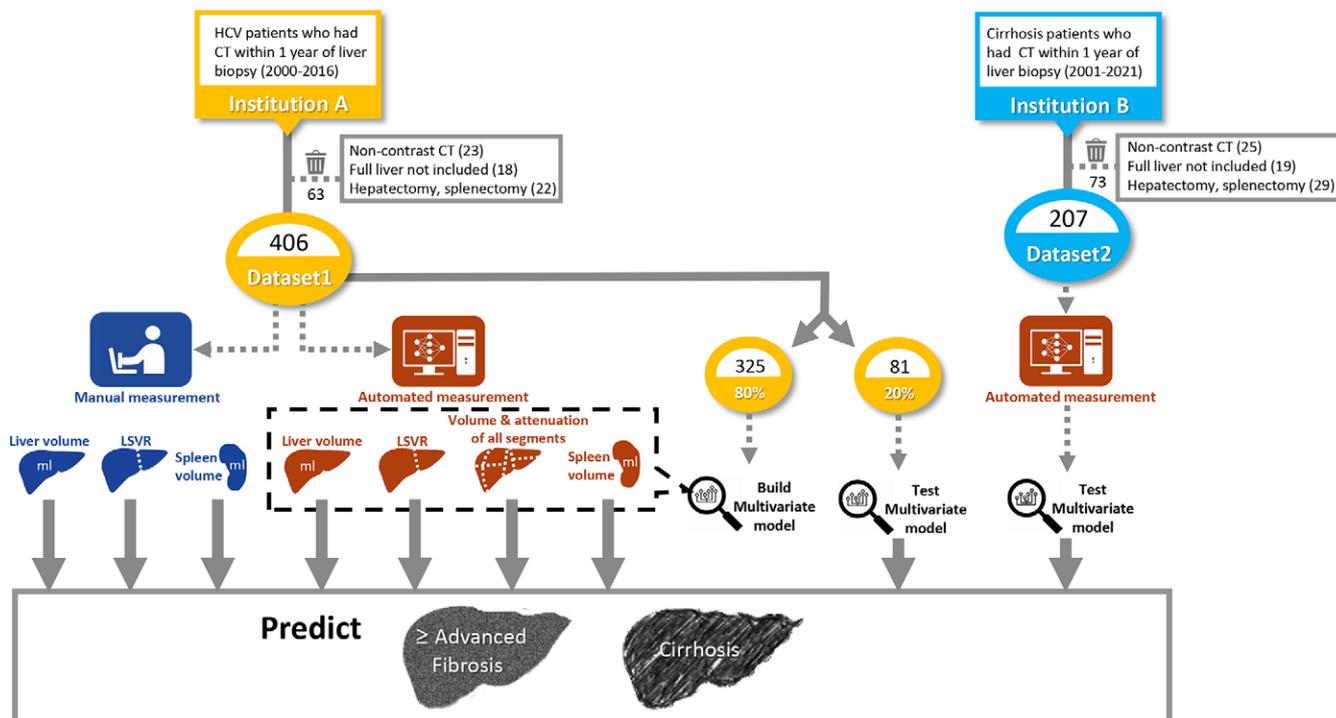
Because the datasets consisted of CT scans collected for more than a decade, they represent scanners and software of various manufacturers. We selected portal venous phase scans for all measurements, which were obtained at approximately 70 seconds from the start of the injection, based on a time-density graph or 45–55 seconds after aortic threshold enhancement. Details of CT scanner types and protocols can be found in Appendix E1 (supplement).

### Manual Measurements of Liver Segments and Spleen Volume

The manually measured liver volumes, LSVR, and spleen volumes in dataset 1 were provided from Pickhardt et al (14) (hereafter, reader 1). Measurements were performed with semiautomated software by several coauthors of Pickhardt et al (14) with CT research experience ranging from 2 to more than 20 years, confirmed by an experienced radiologist. Liver and spleen segmentations were initially performed with use of CT software (liver analysis application, Philips IntelliSpace Portal 11), after which manual adjustments were performed with digital brush-and-eraser tools. At the same time, Couinaud segments I–III were isolated from segments IV–VIII to derive the LSVR (sum of the volumes of segments I–III divided by that of segments IV–VIII).

### Fully Automated Measurements of Liver Segments and Spleen Volume

Two DL models developed in-house were used to automatically segment the eight liver Couinaud segments and spleen from a CT volume. Details of the training data and model development are provided in Appendix E1 (supplement). The outputs of the models include the segmentation, volume (in



**Figure 1:** Diagram of study design. Liver segmental volume ratio (LSVR) was calculated as the volume ratio of Couinaud segments I–III to segments IV–VIII. HCV = hepatitis C virus, Institution A = University of Wisconsin Hospitals and Clinics, Institution B = National Institutes of Health Clinical Center.

**Table 1: Patient Characteristics of the Two Datasets**

Characteristic	Dataset 1	Dataset 2
No. of patients	406	207
M:F	297:109	147:60
Median age (y)*	50 (44–56)	50 (41–57)
Underlying disease (%) <sup>†</sup>	100	...
HCV infection	...	38
Steatohepatitis	...	27
HBV infection	...	12
CVID	...	4
HDV plus HBV infection	...	3
Sickle cell disease	...	3
Hemochromatosis	...	3
PBC	...	2
Autoimmune hepatitis	...	1
GVHD	...	1
Drug-induced liver injury	...	1
Other <sup>‡</sup>	...	2
Median interval between CT and biopsy (mo)*	1 (0–4)	0 (0–2)

Note.—Data are numbers of patients unless otherwise stated. CVID = common variable immunodeficiency, GVHD = graft versus host disease, HBV = hepatitis B virus, HCV = hepatitis C virus, HDV = hepatitis D virus, PBC = primary biliary cholangitis.

\* Data are medians, with IQRs in parentheses.

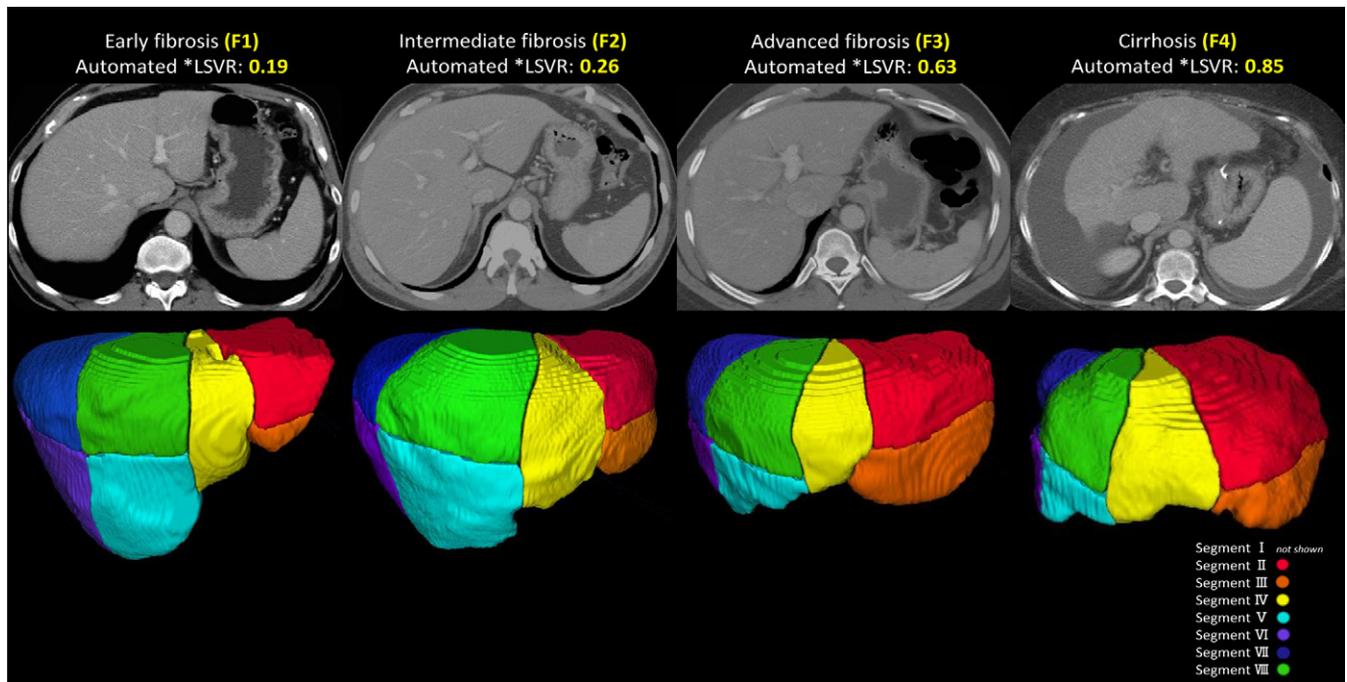
<sup>†</sup> Some patients had several underlying diseases.

<sup>‡</sup> Hyper eosinophilia,  $\alpha$ -1 antitrypsin, and Sjögren disease.

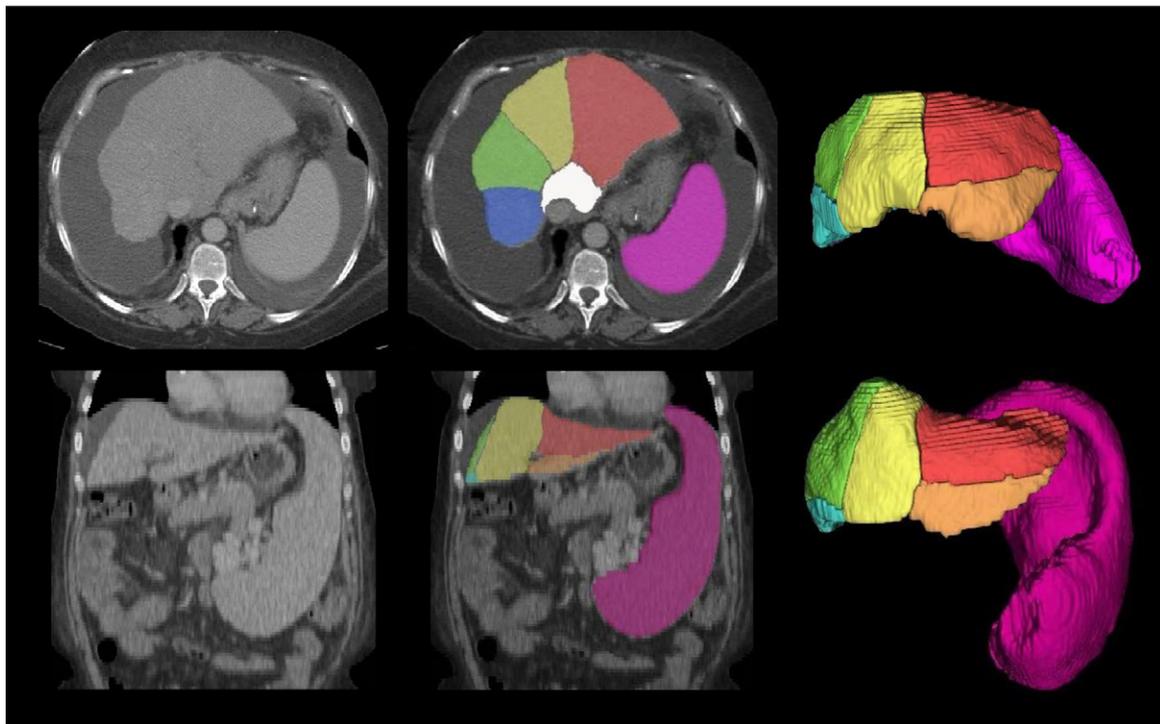
milliliters), and attenuation (mean and median Hounsfield unit with SD) of each of the eight liver Couinaud segments and the spleen. Automated LSVR was calculated in the same manner as the manual LSVR. A “volume proportion” of each Couinaud segment (the volume of each Couinaud segment divided by the entire liver volume) was also calculated. Example images of the automated segmentation and measurements are shown in Figures 2, 3, and E3 (supplement).

### Evaluating DL Model Performance

To compare the model performance on both dataset 1 and dataset 2, manual measurements of the whole liver, spleen, and Couinaud segments I, II, and III and segments IV, V, VI, VII, and VIII were additionally completed by a radiologist (S.L., with 12 years of experience; subsequently referred to as reader 2) on 70 randomly selected scans (35 from dataset 1 and 35 from dataset 2). When performing manual segmentation, the reader was blinded to the biopsy results. All manual segmentation, measurements, Dice similarity coefficient, and the absolute Hausdorff distance calculations were performed with use of Segment Comparison module in 3D Slicer (version 4.10, RRID: SCR\_005619).



**Figure 2:** CT scans show examples of automated liver Couinaud segment measurements. Automated segmentation of the eight liver Couinaud segments (I–VIII) in four different stages of liver fibrosis according to the METAVIR scale from dataset 1, shown on axial contrast-enhanced CT scans: F1 (54-year-old man with chronic hepatitis C virus [HCV] infection), F2 (51-year-old man with HCV infection), F3 (46-year-old man with chronic HCV infection), and F4 (55-year-old woman with chronic HCV infection). The liver segmental volume ratio (LSVR) was larger in higher grades. \*LSVR = a ratio of the volume of segments I–III to segments IV–VIII.



**Figure 3:** CT scans show examples of automated liver segments and spleen volume measurements in challenging cases. Automated segmentation of the eight liver Couinaud segments (six segments visible: white, red, orange, yellow, green, and blue) and spleen (pink) in two challenging cases from dataset 1, shown in axial (top row, first case) and coronal (bottom row, second case) contrast-enhanced CT scans: a 55-year-old woman with chronic hepatitis C virus (HCV) infection (METAVIR F4) and abundant ascites (top row) and a 49-year-old woman with chronic HCV infection (METAVIR F4) and prominent splenomegaly (bottom row). Challenging cases were defined as cases where the primitive deep learning model failed to segment the liver and spleen because of ascites and splenomegaly.

**Table 2: Patient Characteristics per Biopsy Staging System**

Parameter	METAVIR	Knodell HAI*	Ishak Score
No. of patients per stage	47/62/90/59/148 (F0/F1/F2/F3/F4)	58/56/52/41 (0/1/3/4)	58/32/24/29/22/10/32 (0/1/2/3/4/5/6)
Patients with cirrhosis (%)	37 (F4)	20 (4)	20 (5–6)
Patients with advanced fibrosis (%) <sup>†</sup>	73 (F2–4)	45 (3–4)	45 (3–6)

Note.—Data in parentheses are the biopsy staging system scores. HAI = histologic activity index.  
 \* The Knodell HAI staging system does not have a score of 2.  
 † Advanced fibrosis includes patients with cirrhosis.

### Statistical Analysis

To compare the model performance between datasets 1 and 2, Bland-Altman plot, Dice similarity coefficient, and Hausdorff distances between manual and automated measurements were calculated in 70 samples. Linear regression and Bland-Altman plots were used to compare the manual and automated measurements (whole liver volume, spleen volume, and LSVR) in the entire dataset 1 (by reader 1) and 35 samples of dataset 2 (by reader 2).

For power analysis, the coefficient of variation was calculated from a previous study of manually measured LSVR (coefficient of variation, 0.53) (7). The PowerTOST package, version 1.5–4, in R software (version 4.1.0) was used to calculate the sample size for the noninferiority test (R Project for Statistical Computing). We used the sampleN.noninf function with the following parameters: significant  $\alpha$  of .025, target power of 0.8, logscale = false, T/R difference of 0.05, margin of 0.2, and  $2 \times 2$  design. To compare the noninferiority of LSVR measurements of paired (manual vs automated) data, a sample size of 198 people would be required.

The measured values of the whole liver, spleen, and LSVR were compared across fibrosis stages, and the Kruskal-Wallis test was used to assess the differences. The stages were then categorized as cirrhosis (METAVIR F4 in dataset 1 and Knodell HAI 4 or Ishak 5–6 in dataset 2) and advanced fibrosis (METAVIR F2–4 in dataset 1 and Knodell HAI 3–4 or Ishak 3–6 in dataset 2), according to Goodman (17). The advanced fibrosis group includes the cirrhosis group. The medians and IQRs of the automatically measured parameters (whole liver volume, spleen volume, LSVR, volume proportion, median Hounsfield units, and SDs of each Couinaud segment) were calculated for each biopsy stage, and the Kruskal-Wallis test was used to assess the differences.

Discriminatory performance of the parameters in predicting cirrhosis (eg, METAVIR F4 vs F3–0 in dataset 1) and advanced fibrosis (eg, METAVIR F4–3 vs F2–0 in dataset 1) was examined by obtaining the areas under the receiver operating characteristic curve (AUCs). We considered AUC less than 0.6 as representing an ineffective predictor. The AUCs of manual and automated measurements were compared for noninferiority (margin of 0.15,  $\alpha$  of .05) with use of R codes (18,19).  $P < .05$  indicated statistically significant noninferiority.

Through use of combinations of the automated measurements, several multivariable models were built with use of multivariable logistic regression. We also calculated performance for patients with and without HCV infection, separately. Detailed methods and results for the multivariable models and performance in patients with and without HCV infection can be found in Appendix E1 (supplement).

### Results

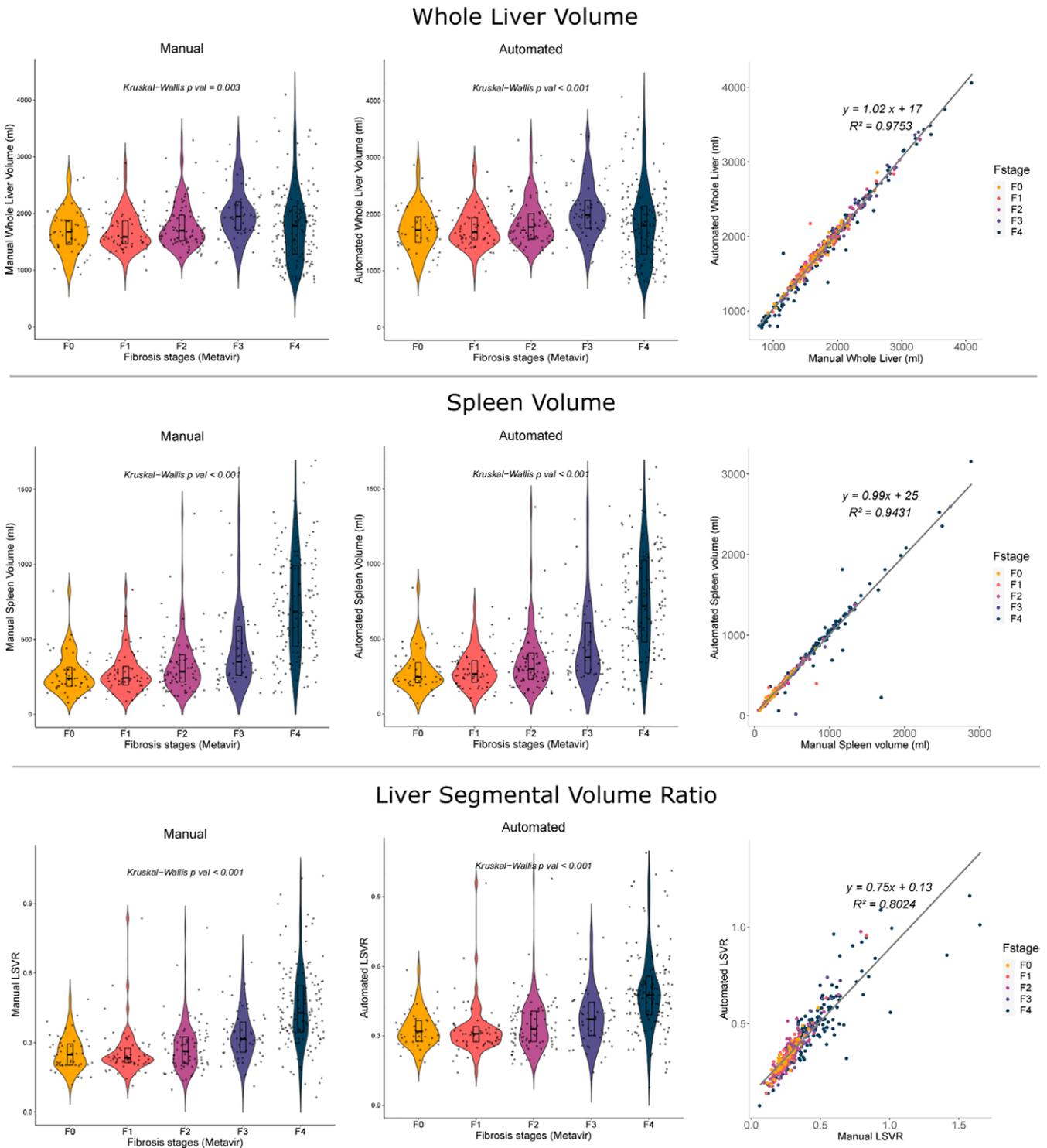
#### Demographic Characteristics in Datasets

Dataset 1 consisted of 406 adults (297 men and 109 women; median age, 50 years [IQR, 44–56 years]); 148 patients (37%) had cirrhosis (METAVIR F4) (Table 2). Dataset 2 consisted of 207 adults (147 men and 60 women; median age, 50 years [IQR, 41–57 years]); 41 patients (20%) had cirrhosis according to the Knodell HAI, and 42 patients (20.3%) had cirrhosis according to the Ishak staging system.

#### Comparison between Manual and Automated Measurements

In the 70 sample cases, Dice similarity coefficient exceeded 0.91 in the whole liver, spleen volume, and Couinaud segments I, II, and III and segments IV, V, VI, VII, and VIII, and the difference in Dice similarity coefficient between datasets 1 and 2 was less than 0.009 (Table E4, Fig E4 [supplement]). When the manual measurements of readers 1 and 2 were compared, reader 1 generally had a larger gap with the automated measurements compared with reader 2 in measuring the whole liver volume (–2.2% vs 0.6% difference for reader 1 vs reader 2), spleen volume (–7.4% vs –1.5% difference for reader 1 vs reader 2), and LSVR (–15.2% vs –6.3% difference for reader 1 vs reader 2) (Bland-Altman plot in Fig E5 [supplement]).

The linear regression line between the manual (reader 1) and automated measurements in dataset 1 had slopes of 1.02, 0.99, and 0.75, with  $R^2$  values of 0.98, 0.94, and 0.80, for whole liver volumes, spleen volumes, and LSVR, respectively (Fig 4, Bland-Altman plot in Fig E6 [supplement]). The linear regression line between the manual (reader 2) and automated measurement in dataset 2 had slopes of 1.01, 0.98, and 0.99, with  $R^2$  values of 0.99, 0.99, and 0.96, for whole liver volumes, spleen volumes, and LSVR, respectively (Fig E7, Bland-Altman plot in Fig E6 [supplement]).



**Figure 4:** Graphs show manual and automated liver and spleen volume measurements in dataset 1. Manual measurements (left column) and automated measurements (center column) of dataset 1. The whole liver volume (first row), spleen volume (second row), and liver segmental volume ratio (LSVR) (third row) across different METAVIR fibrosis stages are shown. The slopes of the regression line between the manual and automated values are close to 1 in the whole liver and spleen volume (right column). The values across the fibrosis stages show a similar pattern between the manual measurements and automated measurements. Although all values were significantly different between the fibrosis stages ( $P < .01$ ), only spleen volume and LSVR show a gradual increasing pattern in higher fibrosis grades. Manual and automated measurements were performed in all 406 patients of dataset 1. Box plot inside the violin plot represents the median and first and third interquartile values. LSVR was calculated as the volume ratio of Couinaud segments I–III to segments IV–VIII.

### Comparison of Automated Measurements between Datasets 1 and 2

We found no evidence of a difference in the automated measurements of the whole liver between datasets 1 and 2.

For spleen volume, measurements were higher in dataset 1 than dataset 2, especially in patients with cirrhosis (median, 736 vs 360 mL;  $P < .001$ ). Dataset 1 had a generally higher LSVR compared with dataset 2 (0.37 vs 0.34;  $P = .01$ ) but

**Table 3: AUCs for Univariable Liver and Spleen Measurements for Predicting Cirrhosis and Advanced Fibrosis**

Variable	Dataset 1: METAVIR System		Dataset 2			
	Advanced Fibrosis (F0–1 vs F2–4)*	Cirrhosis (F0–3 vs F4)	Knodell HAI System		Ishak Staging System	
			Advanced Fibrosis (0–1 vs 3–4)*	Cirrhosis (0–3 vs 4)	Advanced Fibrosis (0–2 vs 3–6)*	Cirrhosis (0–4 vs 5–6)
<b>Manual measurement</b>						
Whole liver volume (mL)	0.58 (0.52, 0.64)	0.48 (0.41, 0.54)	...	...	...	...
Spleen volume (mL)	0.78 (0.73, 0.82)	0.86 (0.82, 0.9)	...	...	...	...
LSVR	0.76 (0.71, 0.81)	0.83 (0.78, 0.87)	...	...	...	...
<b>Automated measurement</b>						
Whole liver volume (mL)	0.57 (0.51, 0.62)	0.46 (0.4, 0.52)	0.49 (0.41, 0.57)	0.45 (0.46, 0.65)	0.49 (0.41, 0.57)	0.46 (0.44, 0.64)
Spleen volume (mL)	0.77 (0.72, 0.81)	0.85 (0.81, 0.89)	0.66 (0.58, 0.73)	0.62 (0.53, 0.72)	0.66 (0.58, 0.73)	0.65 (0.55, 0.74)
LSVR	0.72 (0.67, 0.77)	0.79 (0.74, 0.84)	0.63 (0.55, 0.71)	0.76 (0.67, 0.85)	0.63 (0.55, 0.71)	0.75 (0.66, 0.85)
<b>P value between manual and automated measurement AUCs (noninferiority†)</b>						
Whole liver volume (mL)	<.001	<.001	...	...	...	...
Spleen volume (mL)	<.001	<.001	...	...	...	...
LSVR	<.001	<.001	...	...	...	...

Note.—Data are the areas under the receiver operating characteristic curve (AUCs), with DeLong 95% CIs in parentheses. Dataset 1 used the METAVIR biopsy staging system, and dataset 2 used both the Knodell histologic activity index (HAI) and Ishak staging system. LSVR = liver segmental volume ratio.

\* Advanced fibrosis includes patients with cirrhosis.

†  $P < .05$  indicates significant noninferiority between two methods.

not in patients with cirrhosis or advanced fibrosis (Table E5 [supplement]).

#### Volume and Attenuation Parameters across Fibrosis Stages

Whole liver volume, spleen volume, and LSVR differed significantly ( $P < .05$ ) among the fibrosis stages in both manual and automated measurements of dataset 1 (Fig 4) and dataset 2 (Fig E7 [supplement]), except for the whole liver volume in the Ishak staging system in dataset 2 ( $P = .19$ ). The volume proportions of segments II and III were significantly greater ( $P < .01$ ), whereas those of segment VIII were significantly smaller ( $P < .001$ ) in higher fibrosis stages in both datasets 1 and 2 and all types of biopsy systems (Tables E1–E3 [supplement]). Although only in a subset of the datasets, the median Hounsfield units and SDs of several segments were significantly lower in higher fibrosis stages (median Hounsfield units were significantly lower in segments I–III and VI–VIII of dataset 1 and segments I–VIII of dataset 2 with the Ishak staging system). The SD was significantly lower in segments I–VIII of dataset 1 ( $P < .05$ ).

#### Univariable Parameters in Predicting Cirrhosis and Advanced Fibrosis

The prediction performance of the spleen volume and LSVR were similar between the manual and automated measurements, with AUC differences less than 0.03. For spleen volume, the AUCs for predicting cirrhosis were 0.86 (95% CI: 0.82, 0.9) versus 0.85 (95% CI: 0.81, 0.89) (significantly noninferior with  $P < .001$ ); for LSVR, the AUCs were 0.83 (95% CI: 0.78, 0.87) versus 0.79 (95% CI: 0.74, 0.84) ( $P < .001$ ) (Table 3).

The performance of the automated measurements was lower in dataset 2 than in dataset 1. For spleen volume, the AUCs for predicting cirrhosis were 0.85 (95% CI: 0.81, 0.89) versus 0.65 (95% CI: 0.55, 0.74) in dataset 1 vs 2; for LSVR, the AUCs were 0.79 (95% CI: 0.74, 0.84) versus 0.75 (95% CI: 0.66, 0.85). Within dataset 2, performance was similar between different biopsy staging systems (for spleen volume, Ishak staging system vs Knodell HAI system AUCs for predicting cirrhosis, 0.65 [95% CI: 0.55, 0.74] vs 0.62 [95% CI: 0.53, 0.72]; for LSVR, AUCs were 0.75 [95% CI:

0.66, 0.85] vs 0.76 [95% CI: 0.67, 0.85], respectively). However, the HCV-only subset of dataset 2 had a generally higher performance than the whole dataset (for LSVR, AUCs for predicting cirrhosis were 0.79 [95% CI: 0.67, 0.9] vs 0.75 [95% CI: 0.66, 0.85]), and the non-HCV subset of dataset 2 had a generally lower performance than the whole dataset (for LSVR, AUCs for predicting cirrhosis were 0.69 [95% CI: 0.52, 0.87] vs 0.75 [95% CI: 0.66, 0.85]) in predicting advanced fibrosis and cirrhosis for all measurements (Table E7 [supplement]).

Whole liver volume was not a useful parameter in both manual (AUC, 0.48 in dataset 1 for predicting cirrhosis) and automated (AUCs, 0.46 in datasets 1 and 2 for predicting cirrhosis) measurements in all the datasets and biopsy staging systems.

### Multivariable Model for Predicting Cirrhosis and Advanced Fibrosis

The best performing multivariable model (S + L + V + D model: with use of a combination of automated spleen, LSVR, volume proportions, and SD of the attenuation in all the liver Couinaud segments) had an AUC of 0.94 (95% CI: 0.89, 0.99) that was significantly noninferior ( $P < .001$ ) to the best performance of the manual multivariable model (manual S + L model: AUC, 0.93; 95% CI: 0.88, 0.98) in dataset 1. Dataset 2 had a similar pattern but with lower performance. However, the HCV-only subset of dataset 2 had a generally higher performance than the whole dataset (AUCs for predicting cirrhosis, 0.82 [95% CI: 0.72, 0.91] vs 0.79 [95% CI: 0.71, 0.87] for S + L + V + D), and the non-HCV subset of dataset 2 had a generally lower performance than the whole dataset. Details can be found in Appendix E1 (supplement).

### Discussion

We found that the performance of automated measurements in predicting cirrhosis was similar to that of the manual measurements in spleen volume (AUCs, 0.85 vs 0.86; significantly noninferior with  $P < .001$ ), LSVR (AUCs, 0.79 vs 0.83;  $P < .001$ ), and multivariable models with use of both (S + L model: AUCs, 0.90 vs 0.93;  $P < .001$ ). However, with the automatic method, more measurements could easily be put into a multivariable model to bring the AUC as high as 0.94 (S + L + V + D model). The performance showed a similar pattern in predictions for advanced fibrosis but varied in an external dataset that used a different pathologic grading system and included different disease entities.

This study compared fully automated Couinaud segmentations with manual segmentation in patients with cirrhosis. Tian et al (20) reported a Dice score of 92.46% for automated Couinaud segmentations in patients with normal liver. Yang et al (21) reported a 45.2 mL  $\pm$  20.9 difference between the fully automatically estimated right lobe volume with intraoperatively calculated right lobe volume in 43 liver donors. However, in reality, the Couinaud segment measurements are much more relevant in abnormal livers than normal livers.

One reason for the discrepancy between the manual and automated measurements in this study was the mis-segmentations of the caudate lobe in the automated measurements. This was

mainly due to mis-segmentations of the caudate lobe in the ground truth training data (20), causing a similar mis-segmentation in the DL model (Fig E8 [supplement]). Some other less frequent mis-segmentations included undersegmentation of the left lateral liver segments in cases where the liver was wrapped around the spleen, undersegmentation of heterogeneously enhancing spleen, and oversegmentation of the adjacent stomach (Fig E9 [supplement]). Errors or variability in the manual segmentations can also be a possible reason for the discrepancy between manual and automated LSVR, as manual segmentations of the Couinaud segments are known to have variabilities even with semiautomated software (6,10). This was found in our study because the whole liver, spleen volume, and LSVR measurements differed between reader 1 and reader 2 when the same scans were assessed (Fig E5 [supplement]).

Although the spleen volume and LSVR were the most powerful univariable predictors of cirrhosis and advanced fibrosis, we also found that the volume proportions were higher in segments II and III and lower in segment VIII in cirrhosis and advanced fibrosis. These volume proportions were as powerful as the LSVR in predicting cirrhosis and advanced fibrosis. We also found that the median Hounsfield units and SDs of the liver attenuation were lower in higher fibrosis stages. The liver attenuations were a measurement of the liver parenchyma and intrahepatic vessels, which were all included in the automated segmentations. Therefore, this correlates well with the current literature because the cirrhotic liver is known to show a smaller diameter of intrahepatic veins (22) and decreased hepatic microperfusion (23), which are all factors that could lower the Hounsfield units and SDs of the liver attenuation in an intravenous contrast-enhanced CT scan.

We found that the spleen measurements had less diagnostic value in predicting cirrhosis in dataset 2 compared with dataset 1. This is understandable because splenomegaly is not a direct sign of liver cirrhosis but rather a result of portal hypertension, and the diagnostic value of spleen volume depends on the number of patients with portal hypertension or other causes of splenomegaly in the dataset. Patients with cirrhosis in dataset 2 had significantly smaller volumes of spleen compared with those in dataset 1 ( $P < .001$ ; Table E5 [supplement]), leading to lower diagnostic performance of spleen volume in predicting cirrhosis.

Liver measurements also had slightly different diagnostic values between dataset 1 and 2. One explanation can be found in the cause of cirrhosis, especially because dataset 1 (HCV dominant) had higher performance than dataset 2 (multiple causes) and an HCV-only subset of dataset 2 had higher performance than the non-HCV subset of dataset 2. The structure of the liver and spleen may differ according to the cause. For instance, hypertrophy of the left liver lobe is reported to be more prominent in hepatitis B virus compared with HCV ( $P = .038$ ) (24) and more prominent in viral hepatitis and alcoholic cirrhosis compared with nonalcoholic steatohepatitis ( $P < .001$ ) (25). Patients with Wilson disease–related cirrhosis are reported to have a higher risk for splenomegaly compared with patients with hepatitis B virus (odds ratio, 4.15) (26). Second, datasets 1 and 2 used different biopsy staging systems. Each biopsy staging system considers different histologic findings,

possibly contributing to a different CT finding. Whereas the Knodell HAI is a complex weighted system, the relatively simplified METAVIR system uses only piecemeal necrosis and lobular necrosis to determine the grade of activity, and Ishak includes portal infiltrate and confluent necrosis with the two previous parameters (27). Furthermore, the stages in a biopsy staging system do not represent measurements of continuous variables. They simply represent different categories of severity (28). Thus, direct conversion between different biopsy staging systems would be less accurate.

In this article, we used measurements such as LSVR and spleen volume to diagnose cirrhosis stages rather than directly predict them. We believe the strength of this lies in the explainability. DL algorithms using the whole CT image to directly predict cirrhosis have been reported to have performances as high as 0.97 and 0.95 for predicting advanced fibrosis and cirrhosis in an HBV-dominant dataset (29). However, in this kind of study, it would be difficult to explain the cause of a case of failure or interpret a difference of performance in an external dataset. By using quantification methods (LSVR quantifies segmental redistribution) that have been proven manually, our automated method has the advantage of giving a better explanation and helping us better understand the disease. Automated measurement also has the advantage of giving us a constant measurement with less variability compared with human measurements.

Some limitations should be noted. First, most patients in this study were diagnosed with HCV infection (406 of 406 in dataset 1; 79 of 207 in dataset 2). Performance was lower in the non-HCV group of dataset 2. Second, although reader 1 and reader 2 were supervised by experienced radiologists, they were not trained with a common reference, and the public data (20) used to train the DL model does not mention this either. This resulted in observer variability, especially in the caudate lobe. However, we believe this represents the real-world setting where manual measurements are made with errors between readers. The large error between the manual readers is the most important reason we need an objective measurement, such as the DL model. Although the manual measurements have interreader and intrareader variability, the DL model always produces the same value. Third, another limitation comes from the relatively long recruitment period of the two datasets (16–20 years), leading to variability in CT scanner types, protocol, and contrast media parameters within a dataset. Finally, needle liver biopsies may result in sampling errors between different regions of the sampled liver and intraobserver variation by different pathologists, leading to underdiagnosing cirrhosis by as much as 14.5% (4). We are planning to improve the DL model by adding automated surface nodularity measurements, considering that surface nodularity scores have high diagnostic performance with semiautomatic programs (AUC, 0.929–0.959) (30,31).

In this study, our DL model predicted biopsy-proven cirrhosis and advanced fibrosis with fully automated, CT-based measurements of the liver and spleen, showing performances similar to that of the manual measurements. The DL model had the advantage of using volume proportions and attenuation SDs of each Couinaud segment, along with the previously defined use of spleen volume and LSVR. We also demonstrate that the

performance of predicting cirrhosis and advanced fibrosis with use of liver and spleen volumes was higher in patients with HCV compared with patients with other causes.

**Acknowledgments:** The research used the high-performance computing facilities of the National Institutes of Health Biowulf cluster. We are also indebted to Seon-Pil Jin, MD, PhD, for help with statistical support.

**Author contributions:** Guarantors of integrity of entire study, **S.L., A.H.Y.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **S.L., A.H.Y., C.K., M.G.L., P.J.P.**; clinical studies, **S.L., A.H.Y., C.K., D.E.K., M.G.L., P.J.P.**; experimental studies, **S.L., C.K., P.J.P.**; statistical analysis, **S.L.**; and manuscript editing, all authors

**Disclosures of conflicts of interest:** **S.L.** No relevant relationships. **D.C.E.** No relevant relationships. **A.H.Y.** No relevant relationships. **C.K.** No relevant relationships. **D.E.K.** No relevant relationships. **M.G.L.** Prior grant funding from Philips, Ethicon. **P.J.P.** Consulting fees from Bracco; stock/stock options in SHINE and Elucen; royalties from Elsevier. **R.M.S.** Royalties for patent and software licenses (iCAD, PingAn, Philips, ScanMed, Translation Holdings); PingAn has cooperative research and development agreement with author's institution; associate editor of *Radiology: Artificial Intelligence*.

## References

- Heron M. Deaths: leading causes for 2016. *Natl Vital Stat Rep* 2018;67(6):1–77.
- Smith A, Baumgartner K, Bositis C. Cirrhosis: diagnosis and management. *Am Fam Physician* 2019;100(12):759–770.
- Afdhal NH. Diagnosing fibrosis in hepatitis C: is the pendulum swinging from biopsy to blood tests? *Hepatology* 2003;37(5):972–974.
- Regev A, Berho M, Jeffers LJ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol* 2002;97(10):2614–2618.
- Brancatelli G, Federle MP, Ambrosini R, et al. Cirrhosis: CT and MR imaging evaluation. *Eur J Radiol* 2007;61(1):57–69.
- Furusato Hunt OM, Lubner MG, Ziemlewicz TJ, Muñoz Del Rio A, Pickhardt PJ. The liver segmental volume ratio for noninvasive detection of cirrhosis: comparison with established linear and volumetric measures. *J Comput Assist Tomogr* 2016;40(3):478–484.
- Pickhardt PJ, Malecki K, Hunt OF, et al. Hepatosplenic volumetric assessment at MDCT for staging liver fibrosis. *Eur Radiol* 2017;27(7):3060–3068.
- Bezerra AS, D'Ippolito G, Faintuch S, Szejnfeld J, Ahmed M. Determination of splenomegaly by CT: is there a place for a single measurement? *AJR Am J Roentgenol* 2005;184(5):1510–1513.
- Lodewick TM, Arnoldussen CWKP, Lahaye MJ, et al. Fast and accurate liver volumetry prior to hepatectomy. *HPB (Oxford)* 2016;18(9):764–772.
- Hermoye L, Laamari-Azjal I, Cao Z, et al. Liver segmentation in living liver transplant donors: comparison of semiautomatic and manual methods. *Radiology* 2005;234(1):171–178.
- Couinaud C. Liver anatomy: portal (and suprahepatic) or biliary segmentation. *Dig Surg* 1999;16(6):459–467.
- Lebre MA, Vacavant A, Grand-Brochier M, et al. Automatic segmentation methods for liver and hepatic vessels from CT and MRI volumes, applied to the Couinaud scheme. *Comput Biol Med* 2019;110:42–51.
- Bedossa P, Poinard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;24(2):289–293.
- Pickhardt PJ, Gaffey 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.
- 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.
- Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. *J Hepatol* 1995;22(6):696–699.
- Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. *J Hepatol* 2007;47(4):598–607.
- Liu JP, Ma MC, Wu CY, Tai JY. Tests of equivalence and non-inferiority for diagnostic accuracy based on the paired areas under ROC curves. *Stat Med* 2006;25(7):1219–1238.

19. Non-inferiority test for paired ROC curves. <https://www.bioinfo-scrounger.com/archives/non-inferiority-test-roc/>. Published May 6, 2022. Accessed June 22, 2022.
20. Tian J, Liu L, Shi Z, Xu F. Automatic Couinaud segmentation from CT volumes on liver using GLC-UNet. In: Suk HI, Liu M, Yan P, Lian C, eds. *Machine Learning in Medical Imaging. MLMI 2019. Lecture Notes in Computer Science*, vol 11861. Cham, Switzerland: Springer, 2019; 274–282.
21. Yang X, Yang JD, Hwang HP, et al. Segmentation of liver and vessels from CT images and classification of liver segments for preoperative liver surgical planning in living donor liver transplantation. *Comput Methods Programs Biomed* 2018;158:41–52.
22. Zhang Y, Zhang XM, Prowda JC, et al. Changes in hepatic venous morphology with cirrhosis on MRI. *J Magn Reson Imaging* 2009;29(5):1085–1092.
23. Van Beers BE, Leconte I, Materne R, Smith AM, Jamart J, Horsmans Y. Hepatic perfusion parameters in chronic liver disease: dynamic CT measurements correlated with disease severity. *AJR Am J Roentgenol* 2001;176(3):667–673.
24. 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.
25. 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.
26. Zhong HJ, Sun HH, Xue LF, McGowan EM, Chen Y. Differential hepatic features presenting in Wilson disease-associated cirrhosis and hepatitis B-associated cirrhosis. *World J Gastroenterol* 2019;25(3):378–387.
27. Shiha G, Zalata K. Ishak versus METAVIR: terminology, convertibility and correlation with laboratory changes in chronic hepatitis C. *Liver Biopsy* 2011;10:155–170.
28. Takahashi H. Liver biopsy. BoD–Books on Demand. Published September 6, 2011. Accessed July 14, 2021.
29. 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.
30. Smith AD, Branch CR, Zand K, et al. Liver surface nodularity quantification from routine CT images as a biomarker for detection and evaluation of cirrhosis. *Radiology* 2016;280(3):771–781.
31. Pickhardt PJ, Malecki K, Kloke J, Lubner MG. Accuracy of liver surface nodularity quantification on MDCT as a noninvasive biomarker for staging hepatic fibrosis. *AJR Am J Roentgenol* 2016;207(6):1194–1199.