Radiology: Artificial Intelligence

Fully Automated and Explainable Liver Segmental Volume Ratio and Spleen Segmentation in 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 (D.E.K.), National Institutes of Health, Bethesda, Md; and Department of Radiology, University of Wisconsin School of Medicine & Public Health, Madison, Wis (M.G.L., P.J.P.). Received XXX; revision requested XXX; revision received XXX; accepted XXX. Supported in part by the Intramural Research Program of the National Institutes of Health, Clinical Center. **Address correspondence to** R.M. (email: *rms@nih.gov*).

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

https://doi.org/10.1148/ryai.210268

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

Materials and Methods: For this retrospective study, two datasets were used. Dataset1 consisted of patients with hepatitis C with liver biopsy (Metavir F0-F4, 2000–2016). Dataset2 consisted of patients of multiple etiologies with liver biopsy (Ishak 0–6, 2001–2021). Whole liver, LSVR, and spleen volumes were measured from contrast-enhanced CT by radiologists and the DL model. Area under the receiver operating characteristic curve (AUC) values for diagnosing advanced fibrosis (≥ Metavir F2/Ishak 3) and cirrhosis (≥ Metavir F4/Ishak 5) were calculated. Multivariable models were built on Dataset1 and tested on Dataset1 (hold out) and 2.

Results: Datasets1 and 2 consisted of 406 patients (median age, 50 [IQR, 44–56] years; 297 men) and 207 patients (median age, 50 [41–57] years; 147 men). The prediction of cirrhosis was comparable between the manual versus automated measurements for spleen volume (AUCs, 0.86; 95% CI: 0.82, 0.9 versus 0.85; CI: 0.81, 0.89; significantly noninferior, $P < .001$) and LSVR (AUCs, 0.83; CI: 0.78, 0.87 versus 0.79; 0.74, 0.84; $P < .001$) in Dataset1. The best-performing multivariable model achieved AUCs of 0.94 (CI: 0.89, 0.99) and 0.79 (0.71, 0.87) for cirrhosis and 0.8 (0.69, 0.91) and 0.71 (0.64, 0.78) for advanced fibrosis in Datasets1 and 2, respectively.

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

A deep learning-based model measuring the volume of liver Couinaud segments and spleen on contrast-enhanced CT had performance comparable to manual measurements in predicting histopathologic cirrhosis and advanced fibrosis.

Abbreviations

AUC = area under the receiver operating characteristic curve, CECT = contrast-enhanced CT, DL = deep learning, HAI = histologic activity index, HBV = Hepatitis B virus, HCV = Hepatitis C virus, LSVR = liver segmental volume ratio (the ratio between the volume of Couinaud segment I-III and IV-VIII)

Key Points

Fully automated measurements of splenic volume (areas under the receiver operating characteristic curve [AUCs], 0.85; 95% CI: 0.81, 0.89 versus 0.86; 95% CI: 0.82, 0.9; significantly noninferior with $P < .001$) and liver segmental volume ratio (LSVR) (AUCs, 0.79; 95% CI: 0.74, 0.84 versus 0.83; 95% CI: 0.78, 0.87; $P < .001$) had comparable performance with manual measurements in predicting cirrhosis on 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, Dataset1).

In a group with multiple cirrhotic etiologies (Dataset2), the automatic liver and spleen measurements had lower performance in predicting cirrhosis (AUCs, 0.79; 95% CI: 0.71, 0.87 in Dataset2 versus 0.94; 95% CI: 0.89, 0.97 in Dataset1 for best performing multivariable model, which uses the spleen volume, LSVR, volumes and attenuation of the liver Couinaud segments).

Author contributions:

Guarantor 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

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

Chronic liver disease and cirrhosis is 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 due to the cost, invasiveness, and sampling errors (3,4), the use of noninvasive methods has become increasingly common in clinical practice.

Abdominal contrast-enhanced CT (CECT) is a widely accepted method to evaluate not only liver cirrhosis but 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 morphology in cirrhosis, especially a decrease in the volumes of Couinaud segments IV-VIII, and a compensatory enlargement in segments I-III (6), which is referred to by some as "segmental redistribution." The liver segmental volume ratio (LSVR) is a metric that is designed to measure this change of shape and is 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 have proven to be good parameters for predicting cirrhosis and advanced fibrosis (area under the receiver operating characteristic curve [AUC], 0.904, 0.920 for LSVR and 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 as they are time-consuming (9) and prone to interreader errors (6,10). A fully automated method of measuring the liver segments and spleen volume with comparable performance to the manual measurements in predicting liver cirrhosis would be a quick, objective, and explainable method in diagnosing liver cirrhosis in CT.

Automated Couinaud (11) segmentations have previously been approached in a step-by-step manner (ie, whole liver segmentation-vessel segmentation-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 [1] to obtain and evaluate a fully automated measurement of liver Couinaud segments and spleen volumes using abdominal CECT and [2] to predict the degree of liver cirrhosis or advanced fibrosis using automated measurements.

Materials and Methods

Patient Sample

This multi-institution retrospective cohort study was Health Insurance Portability and Accountability Act-compliant 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. Dataset1 included 406 patients with positive antibody testing for hepatitis C virus (HCV) who had abdominal CECT between 2000 and 2016 and a liver biopsy within 1 year of CT at the University of Wisconsin Hospitals and Clinics. Liver biopsy in Dataset1 used the Metavir scoring system (F0 to F4) (13). Of note, Dataset1 is a subset of a previously published paper that uses the data for manual measurements of liver and spleen volume in patients with HCV (469 patients) (14). However, in this paper, the data are used to show automated measurements of liver Couinaud segments using a deep learning-based algorithm.

Dataset2 included 207 patients who had abdominal CECT between 2001 to 2021 and a liver biopsy within 1 year of CT at the National Institutes of Health Clinical Center. These

patients had multiple underlying diseases including viral hepatitis, steatohepatitis, and other cirrhosis-related diseases (full list in Table 1). Liver biopsy in Dataset2 was 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 with only noncontrast images, full range of the liver not included, hepatectomy or splenectomy were excluded from both Dataset 1 and 2.

Multidetector CT Technique

As the datasets consist of CT scans collected for over a decade, they contain scanners and software of various manufactures. We selected portal venous phase scans for all measurements, which was scanned 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 the first section of the Supplementary Materials.

Manual Measurements of the Liver Segments and Spleen Volume

The manually measured liver volumes, LSVR, and spleen volumes in Dataset1 were provided from (14) (from now on described as Reader1). Measurements were done with semiautomated software by multiple coauthors (PMG, DJ, and BW) with CT research experience ranging from 2 to over 20 years, confirmed by an experienced radiologist. Liver and spleen segmentations were initially done using CT software (Liver Analysis application, Philips IntelliSpace Portal 11), after which manual adjustments were done 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 the Liver Segments and Spleen Volume

Two in-house developed DL models were used to automatically segment the eight liver Couinaud segments and spleen from a CT volume. Details of the training data and model development can be found in the second section of the Supplementary Materials. The outputs of the models include the segmentation, volume (mL), and attenuation (mean, median Hounsfield Unit [HU] and standard deviation) 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 Figure E3.

Evaluating Deep Learning Model Performance

To compare the model performance on both Dataset1 and Dataset2, manual measurements of the whole liver, spleen, Couinaud segments I+II+III, and segments IV+V+VI+VII+VIII were additionally done by a radiologist (SL, 12 years of experience; from now on described as Reader2) on 70 randomly selected scans (35 from Dataset1, 35 from Dataset2). When performing manual segmentation, the reader was blinded of the biopsy results. All manual segmentation, measurements, Dice similarity coefficient (DSC), and the absolute Hausdorff distance calculations were done using Segment Comparison module in 3D Slicer (version 4.10 RRID: SCR_005619).

Statistical Analysis

To compare the model performance between Dataset1 and 2, Bland-Altman plot, DSC 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 Dataset1 (by Reader1) and 35 samples of Dataset2 (by Reader2).

For power analysis, the coefficient of variation (CV) was calculated from a prior study of manually measured LSVR (CV = 0.53) (7). The 'PowerTOST package ver. 1.5–4' in R (ver. 4.1.0) was used to calculate the sample size for the noninferiority test. We used the 'sampleN.noninf' function with the following parameters: significant alpha of 0.025, target power of 0.8, logscale = false, T/R difference-0.05, margin-0.2, design "2 × 2." To compare the noninferiority of LSVR measurements of a paired (manual versus automated) data a sample 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 grouped into cirrhosis (Metavir F4 on Dataset1 and Knodell HAI 4 or Ishak 5–6 on Dataset2), and advanced fibrosis (Metavir F2–4 on Dataset1 and Knodell HAI 3–4 or Ishak 3–6 on Dataset2), according to reference (17). The advanced fibrosis group includes the cirrhosis group. The median and IQR values of the automatically measured parameters (whole liver volume, spleen volume, LSVR, volume proportion, median HU, and SD 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 versus F3–0 in Dataset1) and advanced fibrosis (eg, Metavir F4–3 versus F2–0 in Dataset1) was examined by obtaining the area under the receiver operating characteristic (ROC) curve (AUC) values. We considered AUC values of less than 0.6 as ineffective predictors. The AUC values of manual and automated measurements were compared for noninferiority (margin-0.15, alpha 0.05), using R codes from (18,19). $P < .05$ indicated a significant noninferiority.

Using combinations of the automated measurements, several multivariable models were built using multivariable logistic regression. We also calculated a performance for patients with and without HCV, separately. Detailed methods and results for the multivariable models and HCV/non-HCV patient's performance can be found in the third section of the Supplementary Materials.

Results

Demographics of Datasets

Dataset1 consisted of 406 adults (297 men and 109 women; median age, 50 [IQR 44–56] years), with 148 (37%, Metavir F4) patients with cirrhosis (Table 1). Dataset2 consisted of 207 adults (147 men and 60 women; median age, 50 [IQR 41–57] years), with 41 (20%) patients with cirrhosis according to the Knodell HAI and 42 (20.3%) according to the Ishak staging system.

Comparison between Manual and Automated Measurements

In the 70 sample cases, DSC exceeded 0.91 in the whole liver, spleen volume, Couinaud segments I+II+III, and segments IV+V+VI+VII+VIII, and the difference in DSC between Dataset1 and 2 was less than 0.009 (Table E4, example figures in Fig E4). When comparing the

manual measurements of Readers 1 and 2, Reader1 generally had a larger gap with the automated measurements compared with Reader2 in measuring the whole liver volume (-2.2% versus 0.6% difference for Reader1 versus 2), spleen volume (-7.4% versus-1.5% difference for Reader1 versus 2), and LSVR (-15.2% versus-6.3% difference for Reader1 versus 2) (Bland-Altman plot in Fig E5).

The linear regression line between the manual (Reader1) and automated measurements in Dataset1 had a slope of 1.02, 0.99, 0.75 with an R^2 value of 0.98, 0.94, 0.80 for whole liver volumes, spleen volumes, and LSVR, respectively (Fig 4, Bland-Altman plot in Fig E6). The linear regression line between the manual (Reader2) and automated measurement in Dataset2 had a slope of 1.01, 0.98, 0.99 with an R^2 value of 0.99, 0.99, 0.96 for whole liver volumes, spleen volumes, and LSVR, respectively (Fig E7, Bland-Altman plot in Fig E6).

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 Dataset 1 and 2. As for the spleen volume, measurements were higher in Dataset 1 compared with Dataset 2, especially in patients with cirrhosis (median, 736 versus 360 mL, $P < .001$). Dataset 1 had a generally higher LSVR compared with Dataset 2 (0.37 versus 0.34, $P = .01$), but not in patients with cirrhosis or advanced fibrosis (Table E5).

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 Dataset2 (Fig E7), except for the whole liver volume in the Ishak staging system in Dataset2 ($P = .19$). The volume proportions of segments II and III were significantly greater ($P < .01$), while 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). Although only in a subset of the datasets, the median HU and SD of several segments were significantly lower in higher fibrosis stages (Median HU was significantly lower in segments I-III, VI-VIII of Dataset 1, and segments I-VIII of Dataset 2 with Ishak staging system. 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 (AUCs, 0.86; 95% CI: 0.82, 0.9 versus 0.85; CI: 0.81, 0.89; significantly noninferior with $P < .001$ for spleen volume and AUCs, 0.83; CI: 0.78, 0.87 versus 0.79; CI: 0.74, 0.84; $P < .001$ for LSVR in predicting cirrhosis) (Table 2).

The performance of the automated measurements was lower in Dataset2 compared with Dataset1 (AUCs, 0.85; CI: 0.81, 0.89 versus 0.65; CI: 0.55, 0.737 in Dataset1 versus 2 for spleen volume and AUCs, 0.79; CI: 0.74, 0.84 versus 0.75; CI: 0.66, 0.85 for LSVR in predicting cirrhosis). Within Dataset2, performance was similar between different biopsy staging systems (AUCs, 0.65; CI: 0.55, 0.74 versus 0.62; CI: 0.53, 0.72 in Ishak staging system versus Knodell HAI system for spleen volume and AUCs, 0.75; CI: 0.66, 0.85 versus 0.76; CI: 0.67, 0.85 for LSVR in predicting cirrhosis). However, the HCV-only subset of Dataset2 had a generally higher performance than the whole dataset (AUCs, 0.79; CI: 0.67, 0.9 versus 0.75; CI: 0.66, 0.85

for LSVR in predicting cirrhosis), and the non-HCV subset of Dataset2 had a generally lower performance than the whole dataset (AUCs, 0.69; CI: 0.52, 0.87 versus 0.75 CI: 0.66, 0.85 for LSVR in predicting cirrhosis) in predicting advanced fibrosis and cirrhosis for all measurements (Table E6).

Whole liver volume was not a useful parameter in both manual (AUC, 0.48 in Dataset1 for predicting cirrhosis) and automated measurements (AUCs, 0.46, 0.46 in Datasets1 and 2 for predicting cirrhosis) in all the datasets and biopsy staging systems.

Multivariable Model in Predicting Cirrhosis and Advanced Fibrosis

The best performing multivariable model (S+L+V+D model: using a combination of automated spleen, LSVR, volume proportions, and standard deviation of the attenuation in all the liver Couinaud segments) had an AUC of 0.94 (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; CI: 0.88, 0.98) in Dataset1. Dataset 2 had a similar pattern but with lower performance. However, the HCV-only subset of Dataset2 had a generally higher performance than the whole dataset (AUCs, 0.82; CI: 0.72, 0.91 versus 0.79; CI: 0.71, 0.87 for S+L+V+D in predicting cirrhosis), and the non-HCV subset of Dataset2 had a generally lower performance than the whole dataset. Details can be found in the third section of the Supplementary Materials.

Discussion

We found that the performance of automated measurements in predicting cirrhosis were comparable to that of the manual measurements in spleen volume (AUCs, 0.85 versus 0.86, significantly noninferior with $P <$ .001), LSVR (AUCs, 0.79 versus 0.83, $P <$ .001), and multivariable models using both (S+L model, AUCs 0.90 versus 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 which used a different pathology 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 on patients with normal liver. Yang et al (21) reported a 45.2 ± 20.9 mL 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 pathologic 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). Some other less frequent mis-segmentations included under-segmentation of the left lateral liver segments in cases where the liver was wrapped around the spleen, under-segmentation of heterogeneously enhancing spleen, and over-segmentation of the adjacent stomach (Fig E9). 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 as there was a difference in the

whole liver, spleen volume, and LSVR measurements between Reader1 and Reader2 when assessing the same scans (Fig E5).

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 HU and SD 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 as 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 HU and SD 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 as splenomegaly is not a direct sign of liver cirrhosis but rather a result of portal hypertension, and the diagnostic value of spleen volume will depend 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), leading to lower diagnostic performance of spleen volume in predicting cirrhosis.

Liver measurements also had slightly different diagnostic values between Dataset1 and 2. One explanation can be found in the etiology of cirrhosis, especially because Dataset1 (HCV dominant) had higher performance than Dataset2 (multiple etiologies) and an HCV-only subset of Dataset2 had higher performance than the non-HCV subset of Dataset2. The morphology of the liver and spleen may differ according to the etiology. For instance, hypertrophy of the left liver lobe is reported to be more prominent in hepatitis B virus (HBV) 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 of splenomegaly compared with HBV patients (odds ratio = 4.15) (26). Secondly, Dataset1 and 2 used different biopsy staging systems. Each biopsy staging system considers different histologic findings, possibly contributing to a different CT finding. While the Knodell HAI is a complex weighted system, the relatively simplified Metavir system only uses 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 paper, 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. Deep learning 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. The majority of the patients in this study were diagnosed with HCV (406/406 in Dataset1, 79/207 in Dataset2). Performance was lower in the non-HCV group of Dataset2. Secondly, although Reader1 and Reader2 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. While the manual measurements have inter, intrareader variability, the DL model always produces the same value. Another limitation comes from the relatively long recruitment period of the two datasets (16 to 20 years), leading to variability in CT scanner types, protocol, and contrast media parameters within a dataset. Lastly, needle liver biopsies have a possibility of sampling errors between different regions of the sampled liver and 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, we show that our DL model was capable of predicting biopsy-proven cirrhosis and advanced fibrosis with fully automated, CT-based measurements of the liver and spleen, showing performances comparable to the manual measurements. The DL model had the advantage of using volume proportions and attenuation standard deviations 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 using liver and spleen volumes was higher in patients with HCV compared with patients with other etiologies.

Acknowledgments: The research used the high-performance computing facilities of the NIH Biowulf cluster. We are also indebted to Seon-Pil Jin for help with statistical support.

Disclosures of conflicts of interest: S.L. No relevant relationships. **D.C.E.** No relevant relationships. **A.H.Y.** One stock in Gilead. **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 Elucent. **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

1. Heron M. Deaths: Leading Causes for 2016. Natl Vital Stat Rep 2018;67(6):1–77.

2. Smith A, Baumgartner K, Bositis C. Cirrhosis: Diagnosis and Management. Am Fam Physician 2019;100(12):759–770.

3. Afdhal NH. Diagnosing fibrosis in hepatitis C: is the pendulum swinging from biopsy to blood tests? Hepatology 2003;37(5):972–974.

4. 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.

5. Brancatelli G, Federle MP, Ambrosini R, et al. Cirrhosis: CT and MR imaging evaluation. Eur J Radiol 2007;61(1):57–69.

6. 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.

7. Pickhardt PJ, Malecki K, Hunt OF, et al. Hepatosplenic volumetric assessment at MDCT for staging liver fibrosis. Eur Radiol 2017;27(7):3060–3068.

8. 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.

9. Lodewick TM, Arnoldussen CWKP, Lahaye MJ, et al. Fast and accurate liver volumetry prior to hepatectomy. HPB (Oxford) 2016;18(9):764–772.

10. 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.

11. Couinaud C. Liver anatomy: portal (and suprahepatic) or biliary segmentation. Dig Surg 1999;16(6):459–467.

12. 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.

13. 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.

14. 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.

15. 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.

16. Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. J Hepatol 1995;22(6):696–699.

17. Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. J Hepatol 2007;47(4):598–607.

18. 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 DATE.

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.

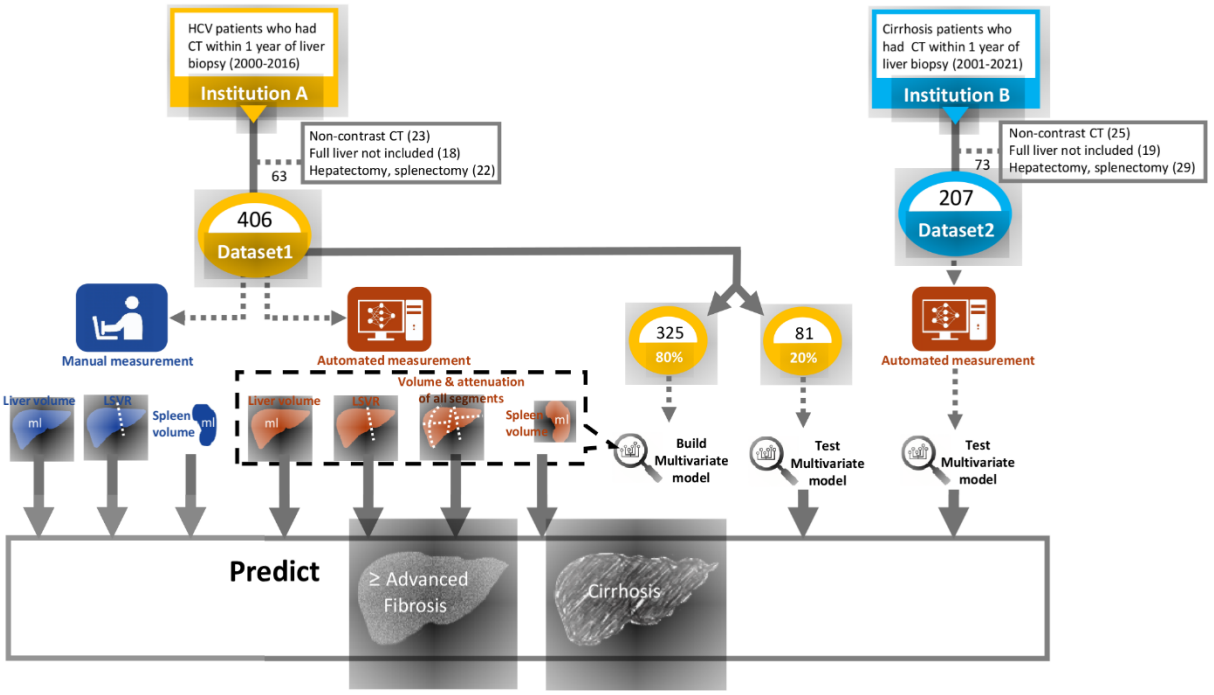


Figure 1: Diagram of study design. Liver segmental volume ratio (LSVR) was calculated as the volume ratio of Couinaud segments I-III to segments IV-VIII. HCV = hepatitis C virus. Institution A = University of Wisconsin Hospitals and Clinics, Institution B = National Institutes of Health Clinical Center.

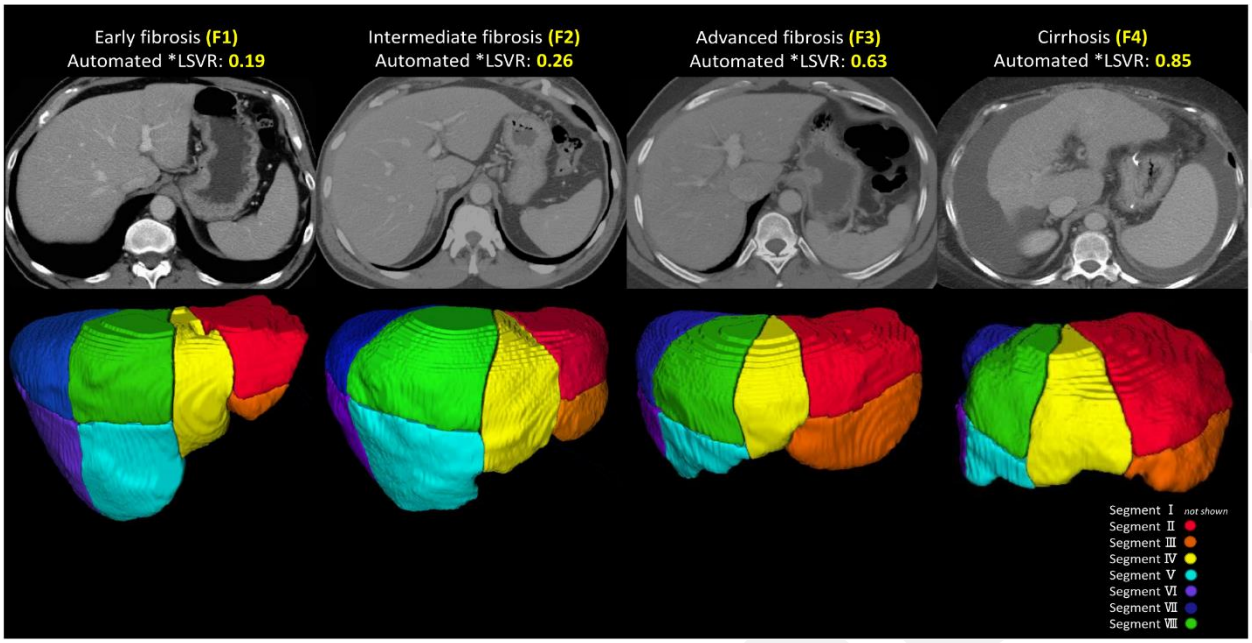


Figure 2: Examples of automated liver Couinaud segment measurements. Automated segmentation of the eight liver Couinaud segments (I-VIII) in four different stages of liver fibrosis according to the Metavir from Dataset1, shown on axial contrast-enhanced CT scans: F1 (54-year-old man with chronic hepatitis C virus [HCV]), F2 (51-year-old man with HCV), F3 (46-year-old man with chronic HCV), and F4 (55-year-old woman with chronic HCV). The liver segmental volume ratio (∗LSVR, a ratio of the volume of segments I-III to segments IV-VIII) was larger in higher grades.

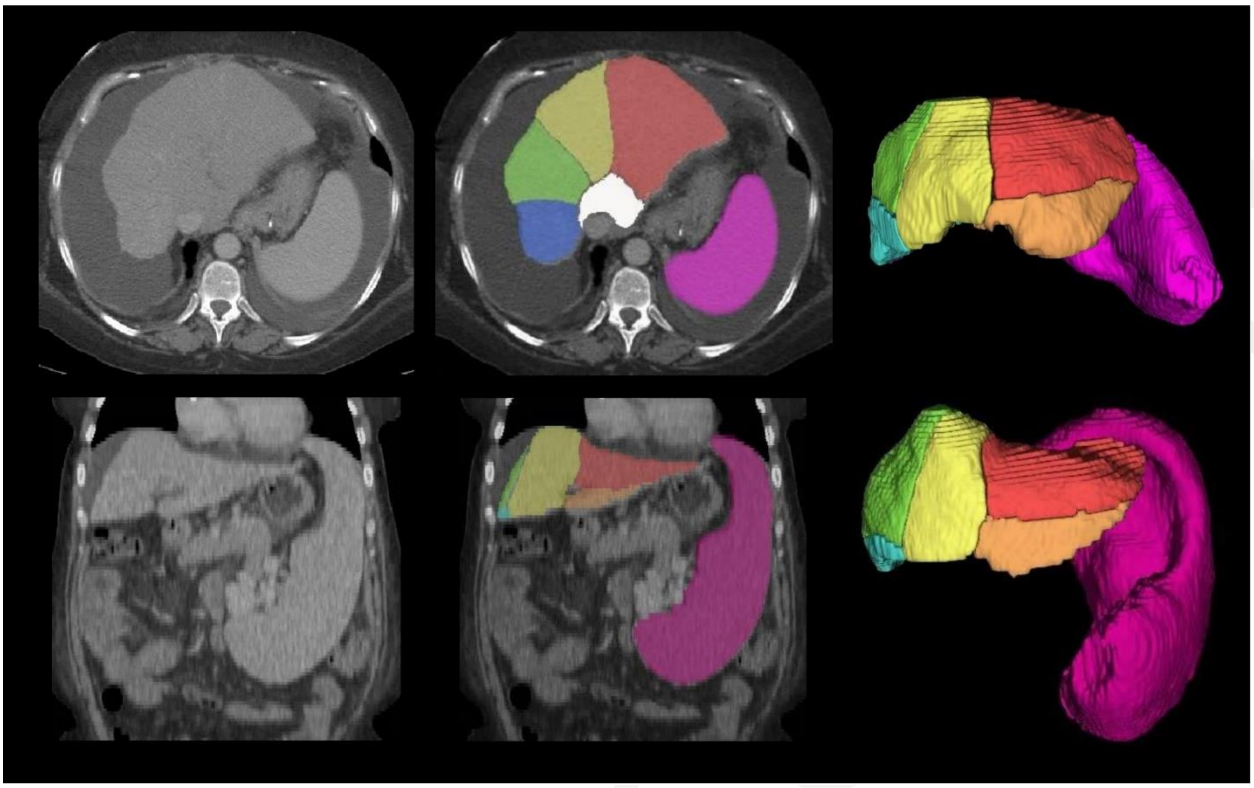


Figure 3: Examples of automated liver segments and spleen volume measurements in challenging cases. Automated segmentation of the eight liver Couinaud segments (six segments visible: white, red, orange, yellow, green, and blue) and spleen (pink) in two challenging cases from Dataset1, shown in axial (top row, first case) and sagittal (bottom row, second case) contrast-enhanced CT scans: 55-year-old woman with chronic hepatitis C virus (HCV) (Metavir F4) and abundant ascites (top row) and a 49-year-old woman with chronic HCV (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 due to ascites and splenomegaly.

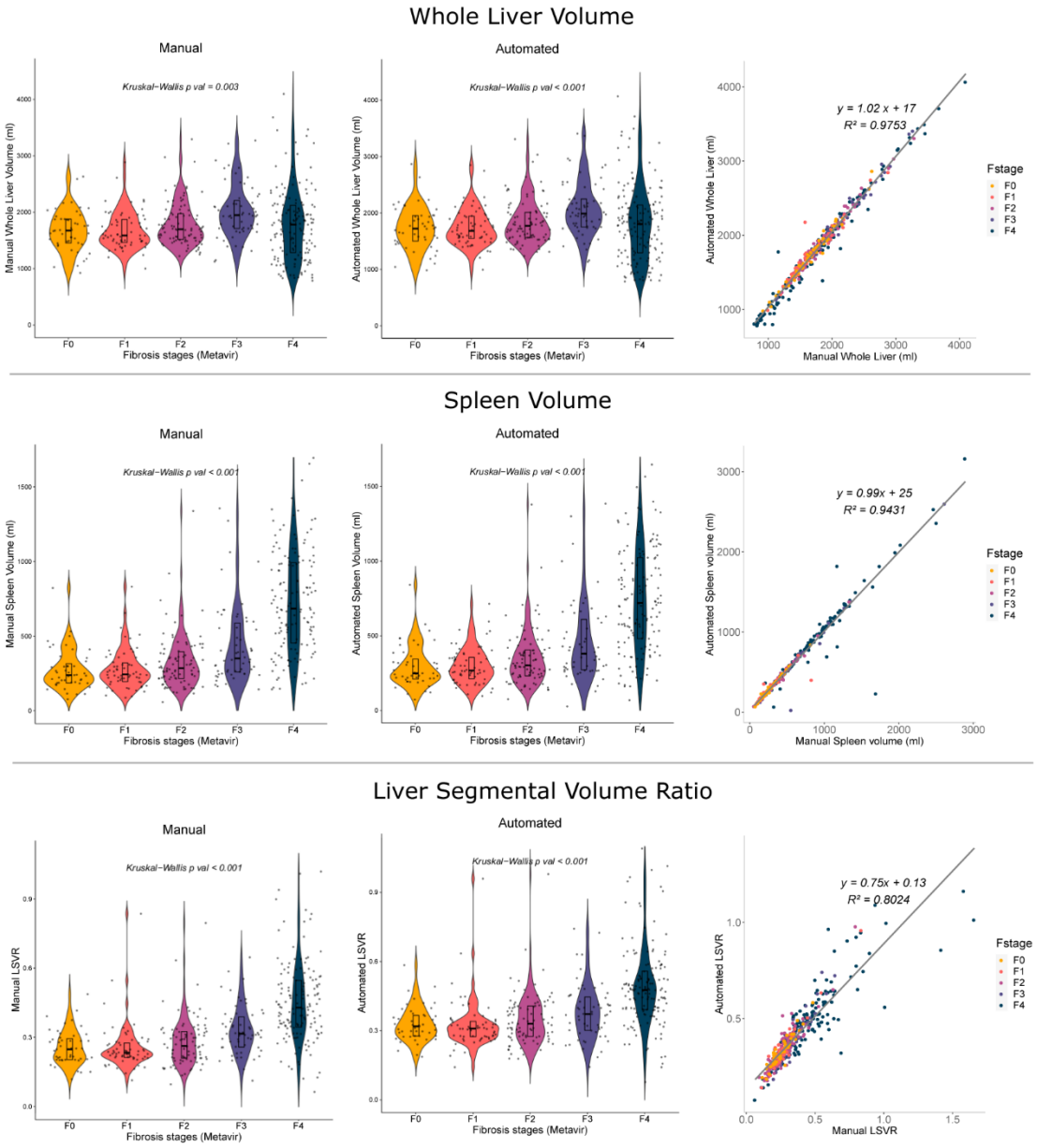


Figure 4: Manual and automated liver and spleen volume measurements in Dataset1. Manual measurements (left column) and automated measurements (center column) of Dataset1. The whole liver volume (1st row), spleen volume (2nd row), and liver segmental volume ratio (LSVR) (3rd row) across different Metavir 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 done in all 406 patients of Dataset1. Boxplot inside the violin plot represents the median and 1st

and 3rd interquartile values. LSVR was calculated as the volume ratio of Couinaud segments I-III to segments IV-VIII.

Table 1

Patient Characteristics of the Two Datasets

	Dataset1	Dataset2	
Total Patients	406	207	
Men:Women	297:109	147:60	
Age (y) (Median, IQR)	50 (44–56)	50 (41–57)	
Underlying disease* (% of patients) HCV Steatohepatitis HBV CVID HDV+HBV Sickle cell disease Hemochromatosis PBC Autoimmune hepatitis GVHD Drug-induced liver injury Other†	100%	 38% 27% 12% 4% 3% 3% 3% 2% 1% 1% 1% 2%	
Biopsy staging system	Metavir	Knodell HAI	Ishak score
Number 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% (Metavir F4)	20% (Knodell HAI 4)	20% (Ishak score 5–6)
Patients with advanced fibrosis (%)¶	73% (Metavir F2–4)	45% (Knodell HAI 3–4)	45% (Ishak score 3–6)
Interval (months) between CT and biopsy (median, IQR)	1 (0–4)	0 (0–2)	

Note.—HCV = hepatitis C virus, HBV = hepatitis B, CVID = common variable immunodeficiency, HDV+HBV = hepatitis D with hepatitis B, PBC = primary biliary cholangitis, GVHD = graft versus host disease, HAI = histologic activity index.

* Some patients had multiple underlying diseases.

† Hypereosinophilia, Alpha 1 Antitrypsin, Sjogren disease.

‡ The Knodell HAI staging system does not have a score of 2.

¶ Advanced fibrosis includes patients with cirrhosis.

Table 2

AUC Values of Univariable Liver and Spleen Measurements for Predicting Cirrhosis and Advanced Fibrosis

	Dataset1		Dataset2			
	Metavir System		Knodell HAI System		Ishak Staging System	
	Advanced Fibrosis (F0–1 vs F2–4) *	Cirrhosis (F0–3 vs F4)	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)
Manual measurements						
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)	—	—	—	—
Automated measurements						
Whole liver volume (mL)	0.57	0.46	0.49	0.45	0.49	0.46

	(0.51–0.62)	(0.4–0.52)	(0.41–0.57)	(0.46–0.65)	(0.41–0.57)	(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)
P value between manual and automated measurement AUCs (noninferiority†)						
Whole liver volume (mL)	< 0.001	< 0.001	—	—	—	—
Spleen volume (mL)	< 0.001	< 0.001	—	—	—	—
LSVR	< 0.001	< 0.001	—	—	—	—

Note.—Dataset1 uses the Metavir biopsy staging system and Dataset2 uses both the Knodell histologic activity index (HAI) and Ishak staging system. The values in the cell represent the area under the receiver operating characteristic curve (AUC) and DeLong 95% confidence intervals (CI). LSVR = liver segmental volume ratio.

* Advanced fibrosis includes patients with cirrhosis.

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

Supplementary Materials

Multidetector CT Technique

As the datasets consist of CT scans collected for over a decade, they contain scanners and software of various manufactures. Dataset1 includes 16-or 64-multidetector CT scanners from GE Medical systems (LightSpeed and subsequent models), and Dataset2 includes 4-to 192-multidetector CT scanners from GE (LightSpeed, HiSpeed, and other models), Siemens (Definition, SOMATOM, and other models), and Philips (Brilliance, Mx8000 and other models) with various versions of soft tissue kernels, such as ‘Soft’, ‘Br40’, and ‘B’. The parameters were 100–120 kVp, using a 2.5-mm to 5-mm section thickness (Dataset1, 8/406 scans were 2.5-mm and 398/406 were 5-mm; Dataset2, all scans were 5-mm), and patient-specific tube current settings. We selected portal venous phase scans for all measurements, which was scanned at approximately 70 seconds from the start of the injection, based on a time/density graph or 45–55 seconds after aortic threshold enhancement. The typical contrast media protocol for Dataset1 included 50–200 mL (Omni 300) or 60–100 mL (Omni 350) of Omnipaque (GE, iohexol injection) at a rate of 3 mL/sec. Dataset2 used 120–130 mL of Oxilan-300 (Guerbet, ioxilan) or Isovue-300 (Bracco, iopamidol injection) at a rate of 2 mL/sec.

Deep Learning Model Development

Training Data Used for DL Model Development

The training data used to develop the model included both public and in-house data, which contained cases of normal liver and spleen, liver tumors, liver cirrhosis, and splenomegaly. For training the first stage liver-only model we used 131 ground truth segmentations from the liver tumors challenge of the Medical Data Decathlon (1,2) and 443 ground truth segmentations for Data Decathlon hepatic vessels CT provided by Tian et al (3). These were supplemented with 24 ascites cases and 10 splenomegaly cases that were downloaded from our institute’s (BLINDED) PACS and labeled in-house by a board-certified radiologist with 12 years of experience. A large number of cases with large liver tumors were removed. In total, we used 268 cases for training, 7 for validation, and 14 for testing. The spleen segmentations for this dataset were generated using

a previously developed spleen model (4). For training the second stage liver Couinaud model we used 193 ground truth Couinaud segmentations which have been made available by Tian et al (3). The ground truth labels are associated with 193 CT scans from the hepatic vessels challenge of the Medical Data Decathlon (1,2). They were supplemented by 7 splenomegaly cases and 10 ascites cases. Altogether we used 188 images for training, 7 for validation, and 15 for testing.

For training the spleen deep learning (DL) model we used 40 noncontrast CT colonography scans from (BLINDED) with segmentations generated in-house for a prior work (4), 41 CECT from the Data Decathlon spleen challenge (1,2), 50 contrast-enhanced CT (CECT) from the "Beyond the Cranial Vault" challenge (5) with labels provided by Gibson et al (6), and 45 CECT from the Cancer Imaging Archive Pancreas-CT dataset (7) from the Cancer Imaging Archive (8) with labels provided by Gibson et al (6). The CECT were supplemented with synthetic noncontrast versions generated using the UNIT image translation technique described previously (9,10). This dataset was supplemented with 7 cases with ascites and 13 splenomegaly cases downloaded from our institute's PACS and labeled by a radiologist with 12 years of experience. Altogether we used 8 cases for validation and 20 for testing.

Details on DL Model Pipeline

The liver DL model consists of two stages-the first stage which segments the liver and spleen and a second stage which does the Couinaud segmentation using a box tightly cropped around the liver segmentation from the first stage, with nonliver pixels set to -1000 HU (See Fig E1 for illustration). The reason for segmenting both the liver and spleen in the first stage was to help the model not confuse the two organs, which have symmetrical shapes and appearance, especially in splenomegaly cases. Both stages used the same 3D U-Net (11) architecture with dropout and skip connections. The spleen DL model is a one-step 3D U-Net (See Fig E2 for illustration). The spleen DL model achieved an average Dice of 0.94 ± 0.02 on the hold-out test set of 20 cases, with a relative average volume difference (RAVD) of 0.056 ± 0.04. In a separate test set of 10 challenging cases (where the primitive model failed to segment the liver and spleen due to abnormal conditions such as ascites and splenomegaly) from the hold-out test set, it achieved a Dice of 0.95 ± 0.01 and RAVD of 0.03 ± 0.03. The runtime for inference was an average of 12 seconds per CT for the spleen segmentation and 33 seconds per CT for the Couinaud segmentation on a P100 GPU (16 Gb of memory). The standard deviations of the DL models performed on the same subject 20 times are close to zero ($< 1 \times 10^{-13}$).

The CT is first put into canonical orientation and then preprocessed by clipping the HU to a soft tissue window of -150–240 HU and rescaled by subtracting the mean and dividing by the standard deviation of the HU. After the liver is segmented by the first stage 3D U-Net, the largest connected component is extracted and the height of the liver is measured and the scan is cropped in the Z-direction from 15% of the liver height below to 15% above. The cropped-z box is then fed back into the liver segmentation model to get a refined output. This helps the model work for scans with a wide variety of fields of view. A connected components analysis is then done and the largest connected component in the segmentation is used as the liver. A tightly cropped box (with ± 1% border in the x, y, and z directions) is taken around the liver segmentation and pixels outside the segmentation are set to -1000. The image is then reclipped to -150–240 HU and renormalized and fed into the second stage U-Net. The second U-Net outputs the Couinaud segmentations. Any stray segmentation lying outside the liver segmentation from the first stage is removed.

For the spleen model, the architecture and training procedure are the same as the first stage of the liver model, but with only one class (the spleen) as output. We did not use a tightly cropped box, for the spleen. The images were resampled to 192 × 192 × 128 and a "temporal ensemble" of two models (saved at 14000 and 25000 iterations of training) was used.

For training both U-Net models, we utilized the Generalized Dice loss (12), which has been shown to outperform standard Dice loss for multiclass segmentation. We trained the model using the Rectified Adam optimizer (13) with a batch size of 2. For data augmentation, we used random rotations between ± 10 degrees around one of the XYZ axes and random elastic deformations using a B spline. The ascites and splenomegaly cases were reweighted in the training sampler to comprise about 50% of the training iterations. Temporal ensembling was used in both the first and second stages where 3–5 models taken from between 20,000–30,000 iterations of training are utilized for prediction and their results averaged.

Multivariable Models in Predicting Cirrhosis and Advanced Fibrosis

Materials and Methods

Using combinations of the automated measurements (whole liver volume, spleen volume, LSVR, volume proportions, attenuations), the following multivariable models were built using multivariable logistic regression.

Multivariate model using the [S]pleen volume and [L]SVR will be marked as the 'S+L model', [W]hole liver volume, [S]pleen, and [L]SVR as the 'W+S+L model'. Model using the [V]olume proportions (the volume of each Couinaud segment divided by the entire liver volume) of all the liver Couinaud segments will be marked as the 'V model', model using the [M]edian HU attenuation of all the liver Couinaud segments as the 'M model', model using the standard [D]eviation of the attenuation in all the liver Couinaud segments as the 'D model', model using the [S]pleen volume and [V]olume proportions (compared with the entire liver) of all the liver Couinaud segments as the 'S+V model', model using the [S]pleen, [L]SVR, and standard [D]eviation of the attenuation in all the liver Couinaud segments as the 'S+L+D model', and model using the [S]pleen, [L]SVR, [V]olume proportions (compared with the entire liver) of all the liver Couinaud segments, and standard [D]eviation of the attenuation in all the liver Couinaud segments as the 'S+L+V+D model'.

Before building the multivariable models, Dataset1 was first randomly split into a ratio of 80:20. The multivariable models were built on the 80% (325/406) split of Dataset1, then tested on the remaining 20% (81/406) of Dataset1. The same models were also tested on Dataset2 (See Fig 1 for illustration). AUCs between the multivariable models were also compared for noninferiority.

Since Dataset1 consisted solely of patients with HCV, while Dataset2 had multiple etiologies including viral hepatitis and steatohepatitis, we divided Dataset2 into patients with HCV-only (n = 79, patients with HCV in Dataset2) and patients with non-HCV (n = 128, all other etiologies) and calculated the performance of the automated measurements in predicting advanced fibrosis and cirrhosis.

Results

In predicting cirrhosis in Dataset1, the automated S+L model was comparable to the manual S+L model (AUCs, 0.90; CI: 0.84, 0.97 versus 0.93; CI: 0.88, 0.98 for automated versus manual S+L model, significantly noninferior with $P < .001$) (Table E6). The V model had an AUC of 0.79 (CI: 0.68, 0.89). The S+V model had similar performance with the S+L model (AUCs, 0.93; CI: 0.88, 0.98 versus 0.90; CI: 0.84, 0.97 for S+V model versus S+L model, significantly noninferior with $P < .001$). The M model and D model had AUCs lower than 0.75. However, the S+L+V+D model had an AUC of 0.94 (CI: 0.89, 0.99) that was significantly noninferior ($P < .001$) to the best performance of the manual model (manual S+L model: AUC, 0.93; CI: 0.88, 0.98). A similar pattern was observed in the prediction of advanced fibrosis.

In Dataset2, the multivariable models had a generally lower performance compared with Dataset 1, but had a similar pattern, with S+L+V+D as the best performing model (AUCs, 0.79; CI: 0.71, 0.87 in the Ishak staging system and 0.78; CI: 0.70, 0.86 in the Knodell HAI system for predicting cirrhosis, Table E6). However, the HCV-only subset of Dataset2 had a generally higher performance than the whole dataset (AUCs, 0.82; CI: 0.72, 0.91 versus 0.79; CI: 0.71, 0.87 for S+L+V+D in predicting cirrhosis), and the non-HCV subset of Dataset2 had a generally lower performance than the whole dataset (AUCs, 0.73; CI: 0.57, 0.90 versus 0.79; CI: 0.71, 0.87 for S+L+V+D in predicting cirrhosis) in predicting advanced fibrosis and cirrhosis in all multivariable models (Table E7). Receiver operating characteristic curves of the multivariable models in Dataset 1 and 2 can be found in Figure E10.

References

1. Medical Segmentation Decathlon. http://medicaldecathlon.com/. Accessed May 29, 2021.

2. Simpson AL, Antonelli M, Bakas S, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063. https://arxiv.org/abs/1902.09063. Posted February 25, 2019. Accessed DATE.

3. 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.

4. Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci Rep 2019;9(1):16884.

5. Landman B, Xu Z, Iglesias J, Styner M, Langerak T, Klein A. 2015 MICCAI Multi-Atlas Labeling Beyond the Cranial Vault - Workshop and Challenge. . Published February 13, 2015. Accessed DATE.

6. Gibson E, Giganti F, Hu Y, et al. Automatic multi-organ segmentation on abdominal CT with dense v-networks. IEEE Trans Med Imaging 2018;37(8):1822–1834.

7. Roth HR, Lu L, Farag A, et al. DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9349. Cham, Switzerland: Springer, 2015; 556–564.

8. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013;26(6):1045–1057.

9. Zhu Y, Elton DC, Lee S, Pickhardt PJ, Summers RM. Image translation by latent union of subspaces for cross-domain plaque detection. arXiv preprint arXiv:2005.11384. arXiv preprint arXiv:2005.11384. https://arxiv.org/abs/2005.11384. Posted May 22, 2020. Accessed DATE.

10. Zhu Y, Tang Y, Tang Y, et al. Cross-domain Medical Image Translation by Shared Latent Gaussian Mixture Model. In: Martel AL, Abolmaesumi P, Stoyanov D, et al, eds. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020. Lecture Notes in Computer Science, vol 12262. Cham, Switzerland: Springer, 2020; 379–389.

11. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Cham, Switzerland: Springer, 2016; 424–432.

12. Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In: Cardoso MJ, Arbel T, Carneiro G, et al, eds. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2017 2017. Lecture Notes in Computer Science, vol 10553. Cham, Switzerland: Springer, 2017; 240–248.

13. Liu L, Jiang H, He P, et al. On the Variance of the Adaptive Learning Rate and Beyond. arXiv preprint arXiv:1908.03265. arXiv preprint arXiv:1908.03265. https://arxiv.org/abs/1908.03265. Posted August 8, 2019. Accessed DATE.

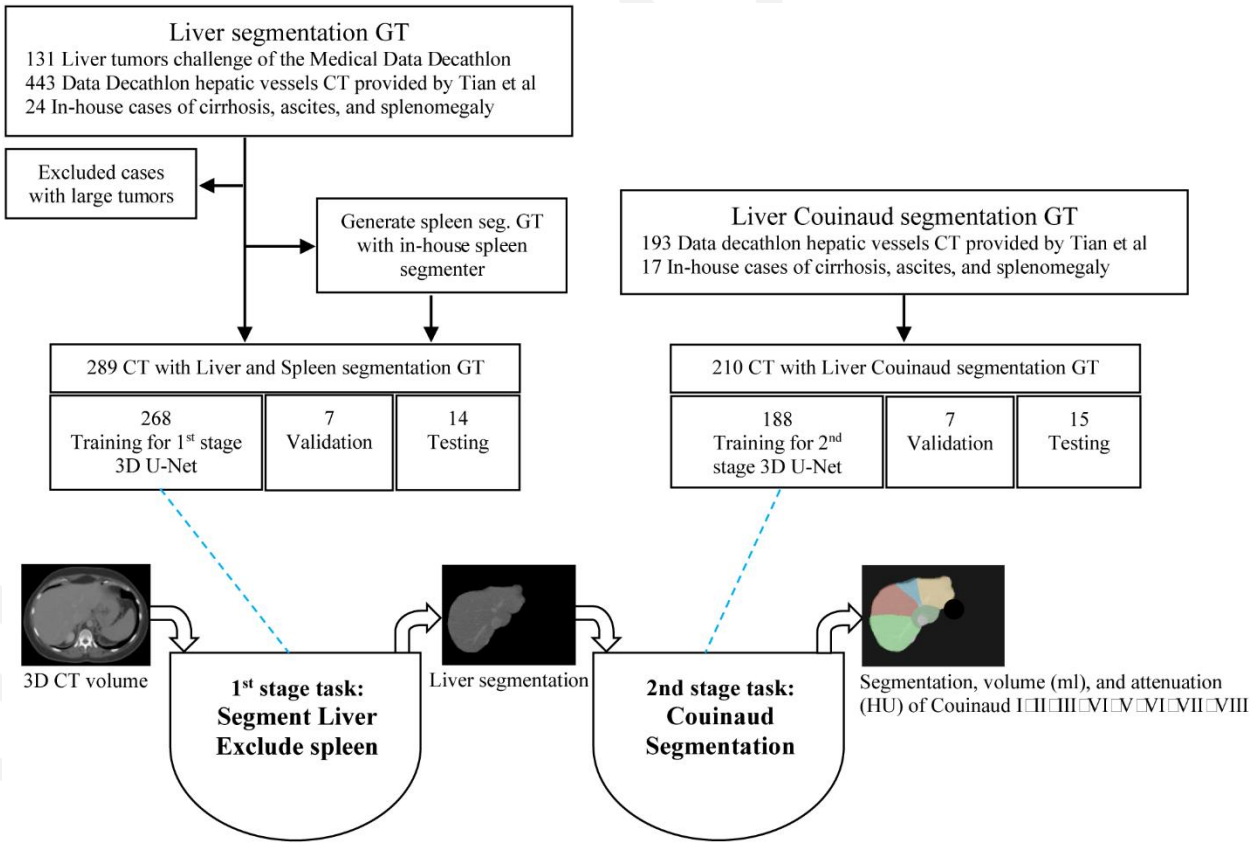


Figure E1: Development of Automated Liver Couinaud Segment Deep Learning Model.

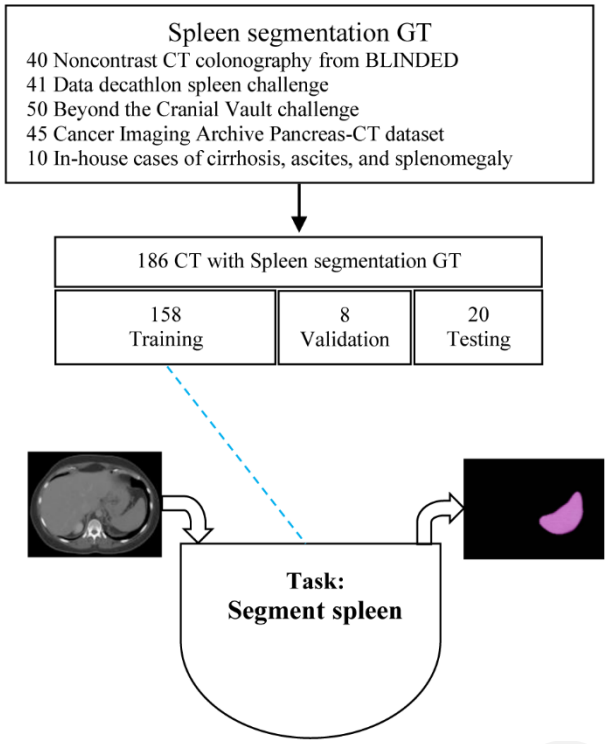


Figure E2: Development of Automated Spleen Segment Deep Learning Model.

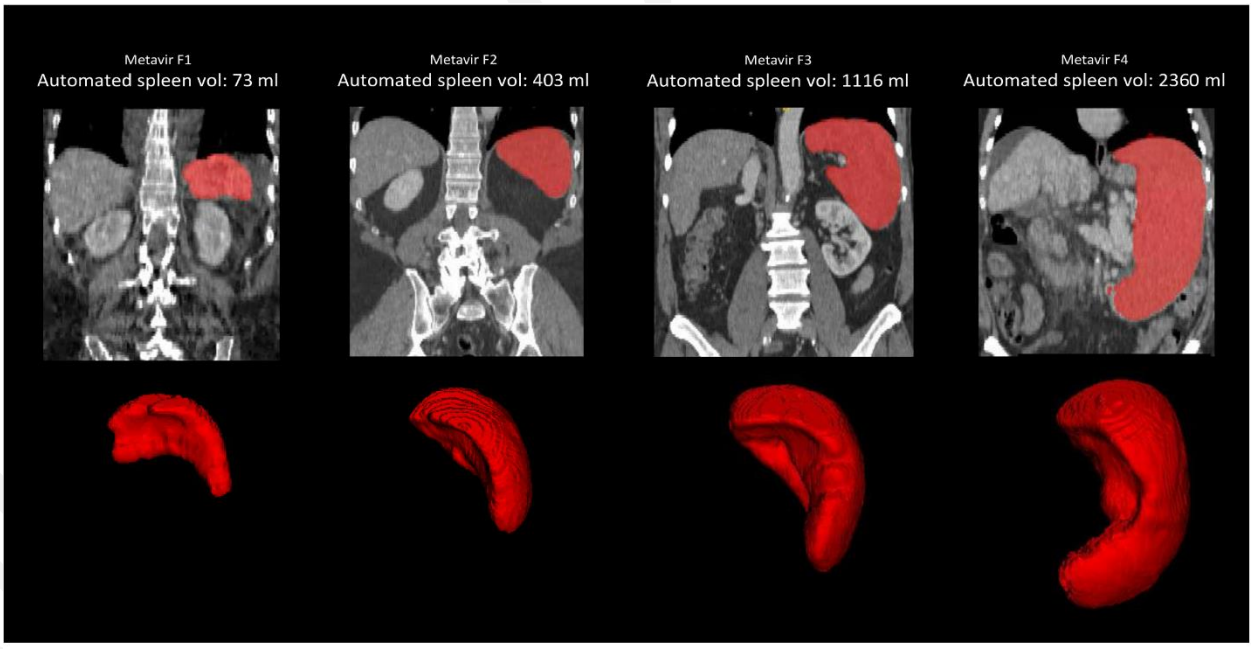


Figure E3: Examples of Automated Spleen Segment Measurements. Automated spleen segmentation in four different stages of liver fibrosis according to the Metavir

staging system in Dataset1, shown in sagittal contrast-enhanced CT scans: F1 (46-year-old female with chronic HCV), F2 (60-year-old male with chronic HCV), F3 (54-year-old male with chronic HCV), F4 (49-year-old female with chronic HCV). Spleen volume increased in higher fibrosis grades.

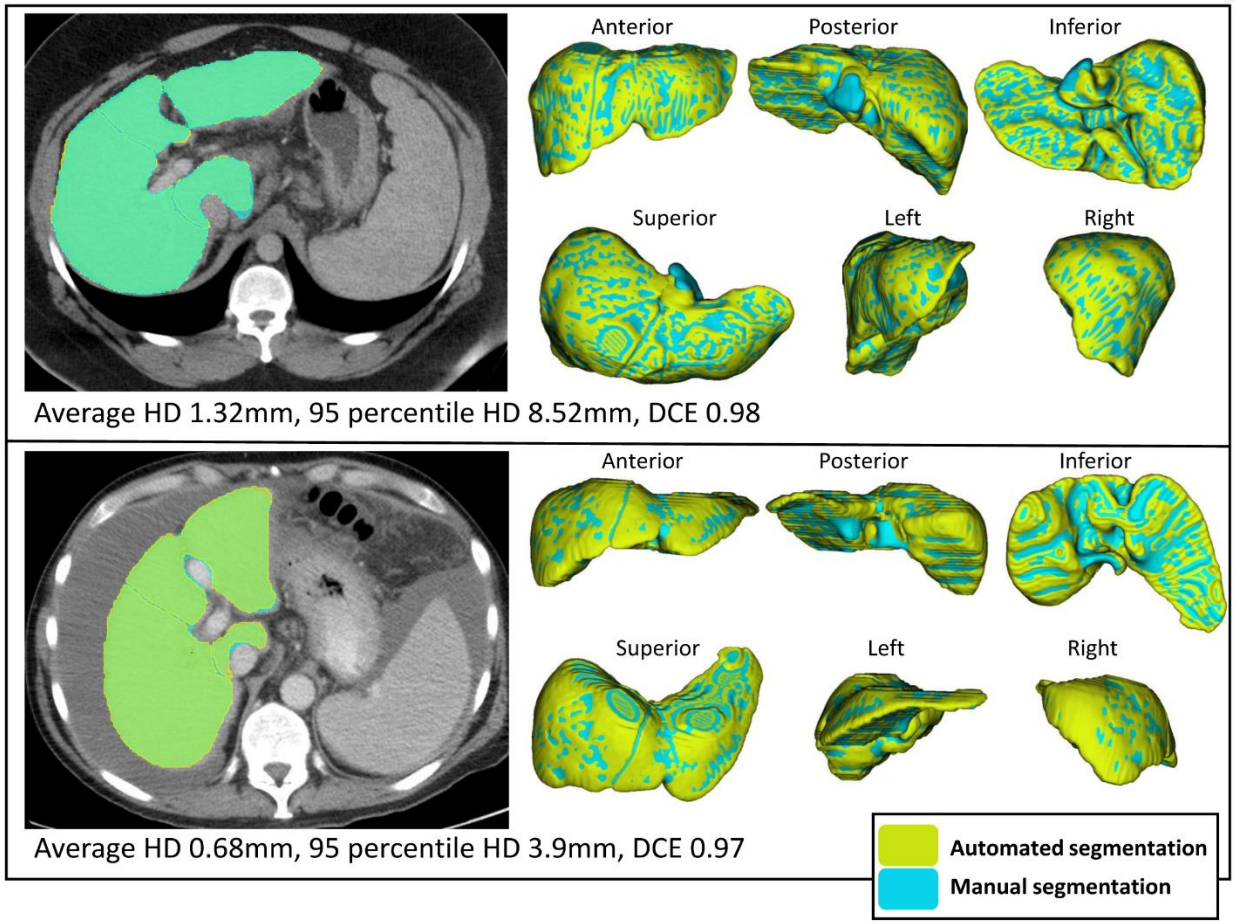


Figure E4: Sample Images of Manual and Automated Segmentation in Comparison. Two cases (top: 34-year-old male with chronic HCV, bottom: 56-year-old male with chronic HCV) from the 70 samples that were both segmented manually (Reader2) and automatically. The manual (sky blue) and automated (olive green) segmentations are shown overlapping each other.

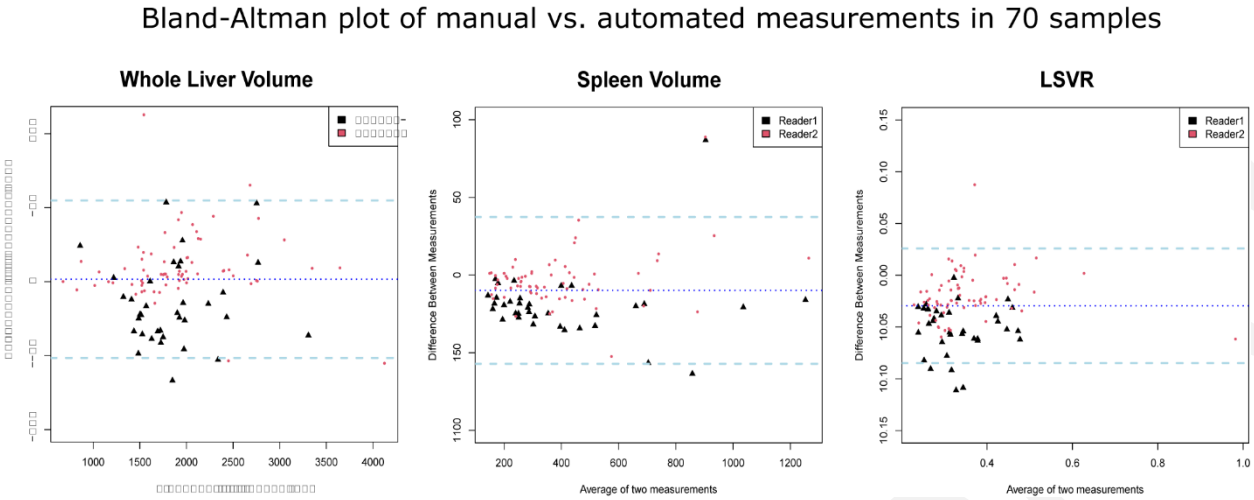


Figure E5: Bland-Altman Plot of the Manual and Automated Measurements in 70 Samples of Dataset 1 and 2. Difference between the manual and automatically measured whole liver volume (left), spleen volume (middle), and liver segmental volume ratio (LSVR) (right). Dotted lines represent the median difference and its 95% confidence intervals. Reader1 represents multiple readers (2 to more than 20 years in CT research experience), confirmed by an experienced radiologist. Reader2 represents a single radiologist with 12 years of experience. Reader1 (black triangles) generally has a larger gap with the automated measurements compared with Reader2 (red dots) in measuring the whole liver volume (median difference -44.2 versus 10.8 mL or -2.2% versus 0.6% difference in percentage for Reader1 versus 2), spleen volume (-20.7 versus -5.4 mL or -1.5% versus -7.4% difference in percentage for Reader1 versus 2), and LSVR (-0.052 versus -0.023 or -15.2% versus -6.3% difference in percentage for Reader1 versus 2). Percentages were calculated by the difference in the manual and automated measurements divided by the average of the manual and automated measurements.

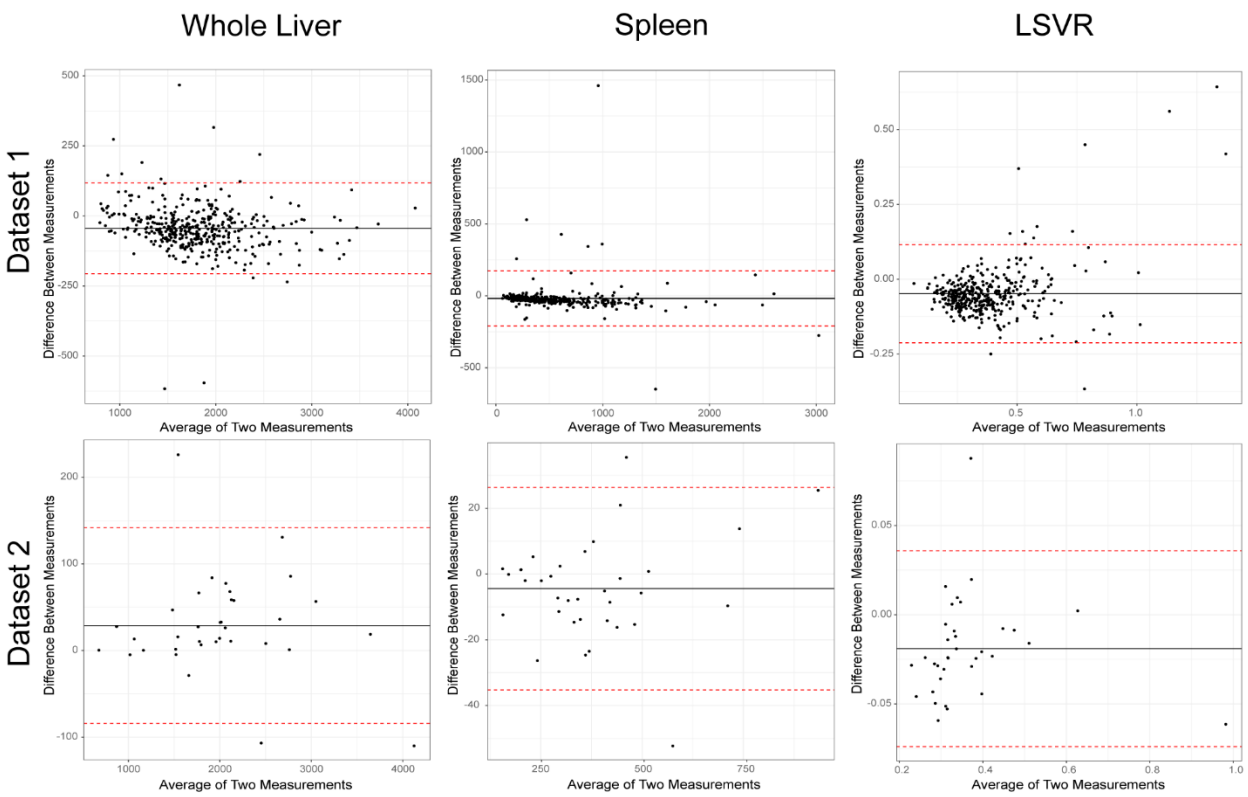


Figure E6: Bland-Altman Plot of the Manual and Automated Measurements. Difference between the manual and automatically measured whole liver volume (left), spleen volume (middle), and liver segmental volume ratio (LSVR) (right). For Dataset1, manual measurements of Reader 1 were compared with the automatically measured values in 591 subjects. For Dataset2, manual measurements of Reader 2 were compared with the automatically measured values in 35 subjects. Lines represent the median difference and its 95% confidence intervals.

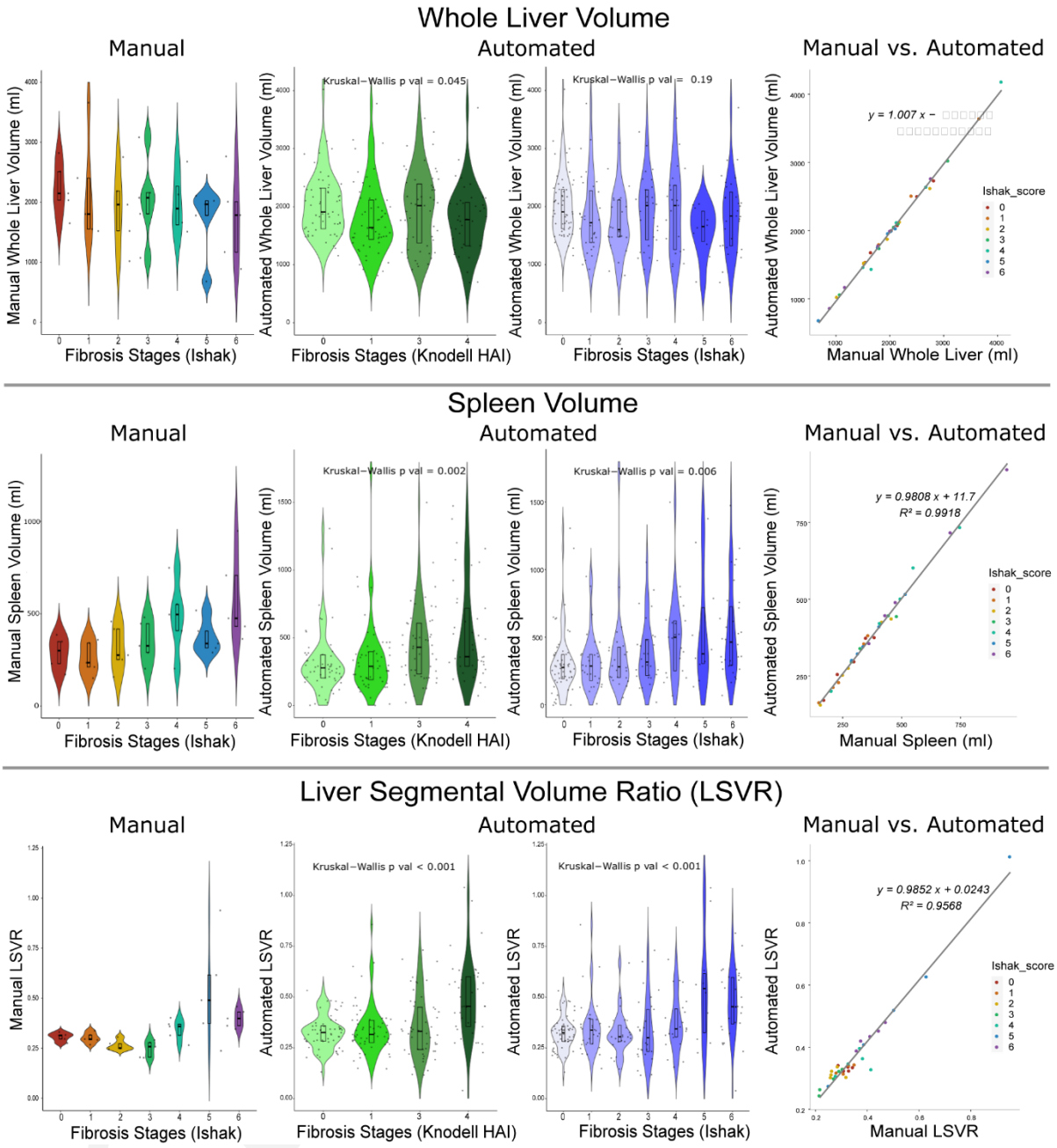


Figure E7: Manual and Automated Liver and Spleen Volume Measurements in Dataset2. Manual measurements (left column) and automated measurements (2nd and 3rd column) of Dataset2. The whole liver volume (1st row), spleen volume (2nd row), and liver segmental volume ratio (LSVR) (3rd row) across different fibrosis stages (Knodell HAI/Ishak) are shown. The slopes of the regression line between the manual and automated values are close to 1 in the whole liver, spleen volume, and LSVR (right column). Manual measurements were done in 35 samples of Dataset2. Automated measurements were done in all 207 subjects of Dataset2. Boxplot inside the violin plot represents the median and 1st and 3rd interquartile values. Liver segmental volume

ratio (LSVR) was calculated as the volume ratio of Couinaud segments I-III to segments IV-VIII.

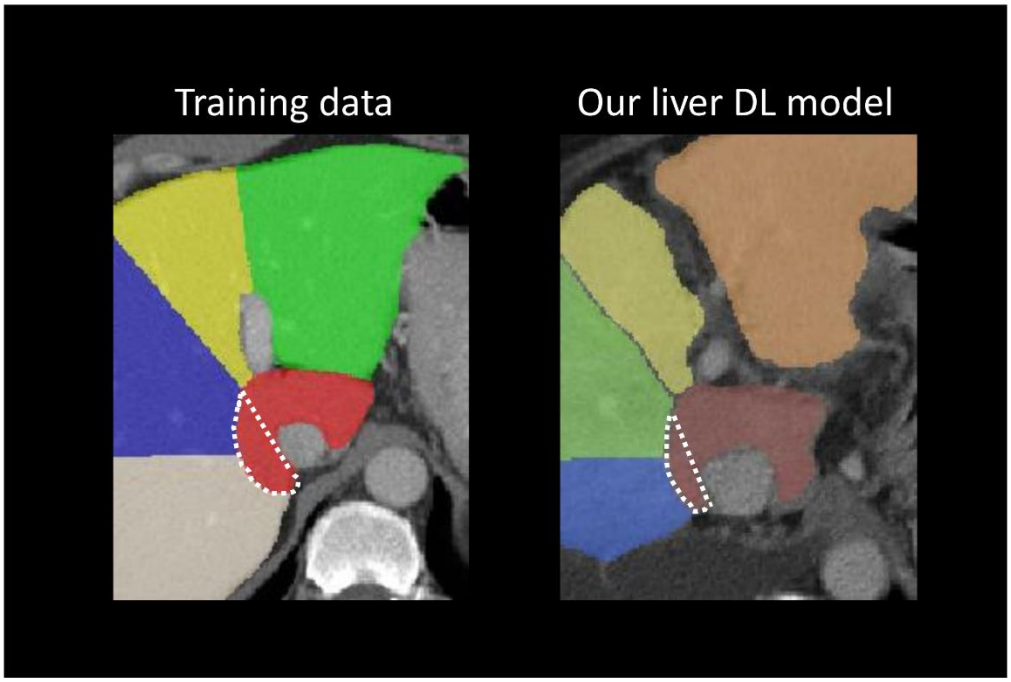


Figure E8: Most Common Mis-segmentations of the Liver DL Model. The training data that we used to train our liver DL model had a tendency to overestimate the caudate lobe (dotted area in the left image), leading to over-segmentation of the caudate lobe in the outputs of our liver DL model (dotted area in the right image).

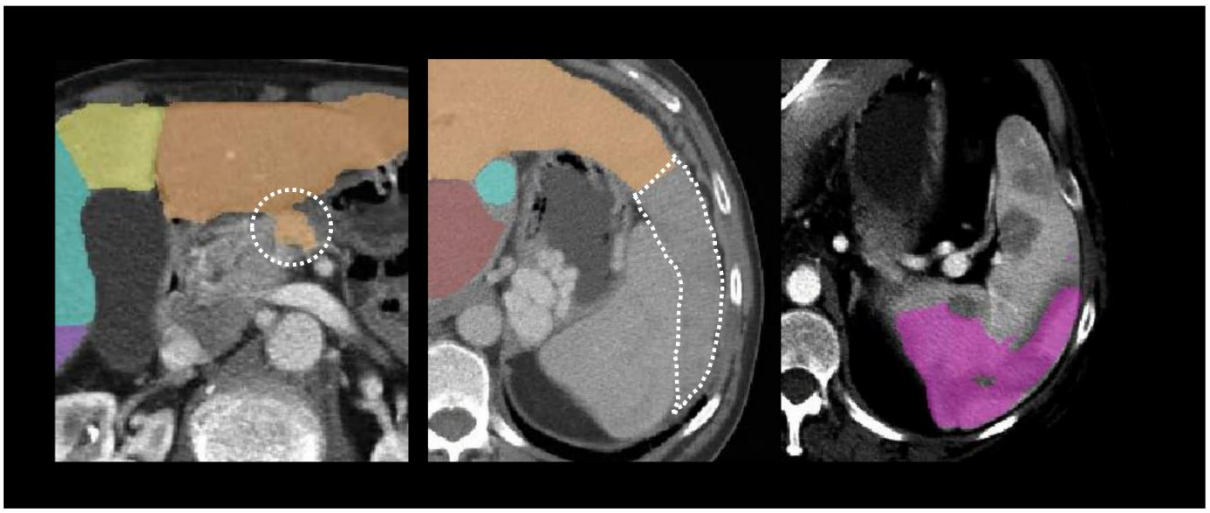


Figure E9: Other Mis-segmentations of the Liver and Spleen DL Model. The DL model produced over-segmentation in some cases of the stomach adjacent to the left lobe (dotted circle on left image), under-segmentation in some cases with the left lateral lobe wrapped around the spleen (dotted area on middle image), and irregular mis-segmentations in some cases of heterogeneously enhancing spleen (right image).

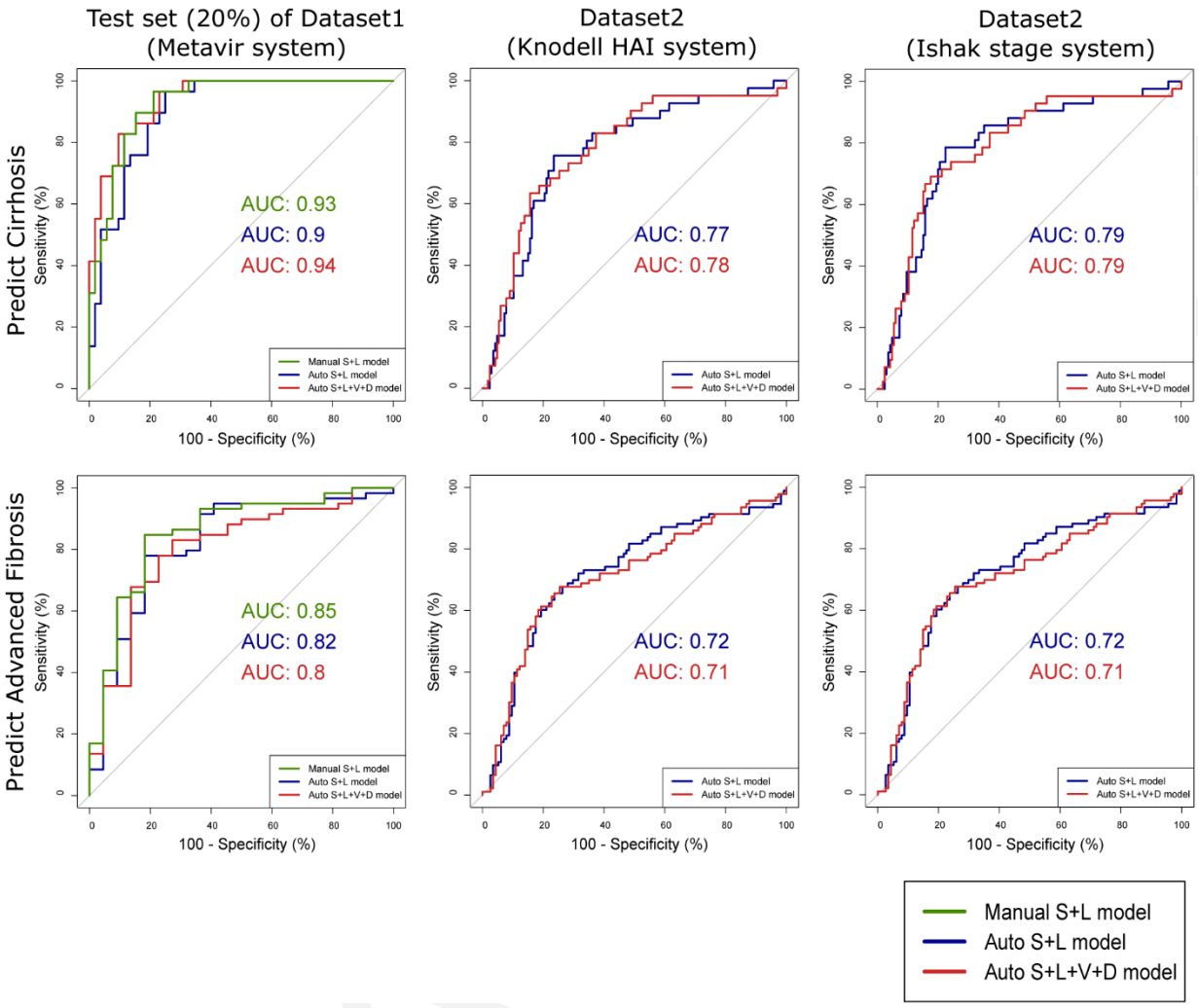


Figure E10: Receiver operating characteristic curves of multivariable models. Performance of multivariable models in a test set (20% split of Dataset1, left column) and Dataset2 (center and right column) in predicting cirrhosis (first row) and advanced fibrosis (second row). Green lines represent multivariable models built with manual measured spleen volume and liver segmental volume ratio (manual S+L model). Blue lines represent multivariable models built with automated spleen volume and liver segmental volume ratio (automated S+L model). Red lines represent multivariable models built with automated spleen volume, liver segmental volume ratio, volume proportions, and attenuation standard deviation of all liver Couinaud segments (Automated S+L+V+D model). Advanced fibrosis includes patients with cirrhosis.

Table E1

Automatically Measured Parameter Values of Each Fibrosis Stage (Metavir System) in Dataset1

		Dataset1					
		Metavir System					
		F0 (*n* = 47)	F1 (*n* = 62)	F2 (*n* = 90)	F3 (*n* = 59)	F4 (*n* = 148)	Kruskal-Wallis *P* Value

Whole liver volume (mL)		1722 (1501–1955)	1682 (1548–1942)	1772 (1566–2010)	1989 (1752–2247)	1803 (1294–2141)	< 0.001
Spleen volume (mL)		249 (210–344)	268 (213–355)	301 (231–406)	380 (274–623)	736 (500–1049)	< 0.001
LSVR		0.32 (0.28–0.37)	0.31 (0.27–0.34)	0.33 (0.28–0.41)	0.37 (0.3–0.45)	0.48 (0.39–0.56)	< 0.001
Volume proportions of segments compared with entire liver (%)*	SI	4.83 (4.13–5.49)	4.54 (4.03–5.09)	4.39 (3.96–4.88)	4.61 (4.02–5.26)	5.12 (4.12–6.44)	< 0.001
	SII	12.71 (11.73–14.63)	13.16 (11.18–14.63)	13.68 (12.12–15.10)	14.34 (11.14–16.56)	16.24 (13.67–19.16)	< 0.001
	SIII	5.82 (3.91–8.09)	5.84 (4.27–7.46)	6.56 (4.26–9.57)	8.14 (5.56–10.6)	9.49 (6.57–12.61)	< 0.001
	SIV	11.32 (9.61–12.56)	10.85 (9.18–12.24)	10.84 (9.27–12.65)	9.79 (8.31–11.23)	9.70 (8.03–11.99)	0.003
	SV	13.89 (10.36–15.7)	13.80 (11.14–16.06)	12.95 (11.25–14.95)	13.69 (11.79–16.00)	10.59 (8.58–13.88)	< 0.001
	SVI	11.30 (10.03–14.17)	12.09 (10.25–14.61)	12.39 (10.31–15.5)	12.21 (10.13–14.21)	10.88 (7.76–13.66)	0.002
	SVII	18.37 (15.85–21.16)	17.95 (15.61–21.23)	17.67 (14.86–19.80)	17.67 (15.65–21.26)	17.52 (15.07–19.64)	0.56
	SVIII	20.29 (18.80–23.23)	20.69 (18.00–22.98)	19.74 (16.90–21.95)	18.97 (16.38–20.53)	17.84 (15.53–19.91)	< 0.001
Attenuation of segments (Median HU)	SI	27.47 (23–32.13)	30.05 (25.75–36.12)	29.49 (26.15–34.13)	26.64 (24.4–30.11)	27.83 (23.81–33.34)	0.01
	SII	33.81 (28.84–39.52)	36.71 (30.83–41.7)	35.83 (31.14–40.8)	33.7 (30.42–39.12)	29.72 (24.76–36.58)	< 0.001
	SIII	37.2 (32.34–44.18)	38.91 (32.35–45.54)	40.63 (35.18–45.12)	38.25 (32.68–42.55)	31.94 (26.15–38.1)	< 0.001
	SIV	32.89 (28.58–40.54)	35.55 (30.78–43.96)	36.67 (30.62–42.6)	36.22 (31.42–43)	33.79 (27.56–43.32)	0.19
	SV	29.18 (25.58–33.84)	29.63 (26.14–36.05)	31.56 (26.1–36.53)	28.91 (25.25–35.07)	28.34 (23.07–34.18)	0.19
	SVI	23.29 (20.08–27.97)	24.17 (21.08–27.43)	25.01 (20.65–30.52)	23.18 (20.73–26.01)	21.7 (18.04–25.67)	< 0.001
	SVII	30 (25.19–37.59)	31.17 (26.58–43.57)	32.89 (28.97–40.67)	35.24 (27.58–44.64)	28.64 (21.94–40.01)	0.002
	SVIII	39.05 (33.71–50.58)	42.56 (37.13–56.78)	44.99 (38.48–54.3)	47.68 (38.02–56.9)	40.59 (28.7–50.89)	0.003
Attenuation of segments (Standard deviation)	SI	112 (97.5–124)	115.5 (101–127)	110 (99.25–123)	107 (93.5–115)	92 (80–110)	< 0.001
	SII	118 (103.5–129)	117 (101.25–131.75)	113.5 (105.25–127)	107 (94–119.5)	93.5 (82–108.25)	< 0.001
	SIII	113 (102.5–129)	114.5 (98.25–130)	112.5 (102–125.75)	105 (92.5–119)	94 (82.75–109.25)	< 0.001
	SIV	118 (104–133)	118.5 (103–132.75)	115 (105.25–129.75)	108 (95–119)	93 (80.75–110)	< 0.001
	SV	116 (106–135)	123.5 (105.5–135.75)	118 (106–129.25)	112 (96.5–124)	95.5 (82.75–111)	< 0.001
	SVI	118 (108–137)	122.5 (105.5–136.75)	118 (107–131.5)	113 (95.5–123.5)	99 (84–113)	< 0.001
	SVII	117 (102.5–133)	119.5 (100.5–133.25)	115 (103–126.75)	107 (91–118)	94.5 (80–111.25)	< 0.001
	SVIII	118 (103.5–136)	121 (104.5–135.75)	116 (104.25–128.25)	109 (92–121)	94 (82.75–110)	< 0.001

Note.—The values in the cell represent the median and 1st–3rd interquartile values of each parameter.

* Volume proportion was calculated by the volume of each Couinaud segment divided by the entire liver volume.

Table E2

Automatically Measured Parameter Values of Each Fibrosis Stage (Knodell HAI System) in Dataset2

		Dataset2

		Knodell HAI				
		0 (*n* = 58)	1 (*n* = 56)	3 (*n* = 52)	4 (*n* = 41)	Kruskal-Wallis *P* Value
Whole liver volume (mL)		1970 (1627–2328)	1643 (1438–2118)	2032 (1402–2429)	1787 (1331–2109)	0.045
Spleen volume (mL)		284 (205–385)	290 (194–409)	441 (247–625)	360 (290–717)	0.002
LSVR		0.33 (0.28–0.36)	0.32 (0.28–0.39)	0.33 (0.24–0.45)	0.45 (0.36–0.6)	< 0.001
Volume proportions of segments compared with entire liver (%)*	SI	4.10 (3.52–4.63)	4.31 (3.77–5.04)	4.26 (3.78–5.05)	4.55 (3.84–5.96)	0.06
	SII	13.39 (11.46–15.28)	13.23 (10.87–15.05)	12.76 (11.06–15.14)	15.78 (13.32–18.42)	< 0.001
	SIII	6.50 (4.37–9.45)	7.07 (4.10–9.20)	6.26 (3.77–10.05)	10.43 (7.73–13.79)	< 0.001
	SIV	11.55 (10.49–13.10)	10.62 (9.09–12.95)	9.16 (7.19–11.36)	10.19 (8.14–11.92)	< 0.001
	SV	13.49 (11.42–15.55)	14.53 (12.10–16.21)	13.58 (10.90–16.41)	11.57 (8.41–14.54)	0.006
	SVI	12.13 (9.85–14.19)	12.55 (8.58–15.54)	14.14 (10.00–17.18)	10.88 (8.30–13.60)	0.12
	SVII	17.05 (15.23–18.81)	17.75 (14.79–20.39)	17.67 (14.68–21.67)	16.63 (14.4–18.91)	0.36
	SVIII	20.60 (18.57–23.80)	19.88 (18.10–22.26)	18.71 (16.29–20.38)	17.67 (15.70–19.45)	< 0.001
Attenuation of segments (Median HU)	SI	28.09 (23.28–35.14)	26.8 (24.11–31.82)	27.13 (24.82–32.21)	29.48 (23.65–37.47)	0.87
	SII	33.4 (30.12–41)	33.29 (26.76–38.2)	32.23 (30–38.8)	29.41 (24.73–38.06)	0.29
	SIII	36.19 (31.52–40.33)	34.16 (27.88–39.31)	35.72 (30.39–39.1)	32.42 (27.41–38.94)	0.33
	SIV	33.86 (28.17–37.92)	32.37 (27.87–40.19)	34.14 (28.9–46.46)	35.46 (29.11–42.85)	0.7
	SV	27.82 (23.04–33.79)	27.56 (23.48–31.96)	30.06 (22.97–33.34)	27.52 (23.06–34.84)	0.92
	SVI	23.47 (20.26–27.45)	21.79 (17.49–27.42)	22.07 (18.57–25.14)	21.37 (16.42–25.22)	0.31
	SVII	27.52 (24.75–52.24)	28.05 (23.73–82.03)	27.12 (24.43–38.13)	28.21 (21.89–46.53)	0.88
	SVIII	36.94 (31.45–79.9)	36.69 (32.98–95.68)	37.55 (33.13–50.14)	39.12 (29.65–61.58)	0.97
Attenuation of segments (Standard deviation)	SI	100 (83.75–108.5)	100 (83.75–109)	88.5 (77.5–103)	91 (79–103)	0.11
	SII	95.5 (83.25–110.75)	101.5 (84–113)	88 (75.75–108.25)	91 (81–104)	0.21
	SIII	95.5 (82.25–108.5)	100 (82.75–114.25)	87 (74.75–104)	94 (81–104)	0.14
	SIV	96 (82.25–110)	99.5 (81.5–114)	87 (73–104.75)	92 (80–104)	0.15
	SV	96.5 (82.25–111.75)	102 (84.25–116)	88.5 (76.75–111.5)	92 (83–106)	0.14
	SVI	99.5 (83.25–111)	100.5 (83.75–118.25)	90 (77.75–113.25)	94 (85–107)	0.19
	SVII	93 (80–107)	96 (83.25–110.25)	84 (74.5–108.25)	90 (80–102)	0.23
	SVIII	95.5 (79.5–109)	99.5 (83.25–112.5)	87 (75.75–110)	91 (83–105)	0.28

Note.—The values in the cell represent the median and 1st–3rd interquartile values of each parameter.

* Volume proportion was calculated by the volume of each Couinaud segment divided by the entire liver volume.

Table E3

Automatically Measured Parameter Values of Each Fibrosis Stage (Ishak Scoring System) in Dataset2

		Dataset2							
		Ishak Scoring System							
		0 (*n* = 58)	1 (*n* = 32)	2 (*n* = 24)	3 (*n* = 29)	4 (*n* = 22)	5 (*n* = 10)	6 (*n* = 32)	Kruskal-Wallis *P* Value
Whole liver volume (mL)		1922 (1627–2315)	1724 (1382–2279)	1603 (1475–2118)	2021 (1427–2291)	2065 (1308–2459)	1658 (1399–1925)	1883 (1329–2353)	0.19
Spleen volume (mL)		284 (200–385)	292 (189–380)	283 (204–429)	341 (220–506)	511 (277–628)	379 (307–720)	466 (294–728)	0.006
LSVR		0.33 (0.28–0.36)	0.34 (0.27–0.39)	0.31 (0.28–0.36)	0.3 (0.23–0.44)	0.34 (0.3–0.44)	0.54 (0.33–0.62)	0.45 (0.37–0.6)	< 0.001
Volume proportions of segments compared with entire liver (%)*	SI	4.10 (3.52–4.62)	4.17 (3.76–4.60)	4.52 (4.15–5.21)	4.13 (3.75–4.71)	4.61 (3.90–5.50)	4.83 (3.94–5.91)	4.54 (3.89–5.81)	0.08
	SII	13.43 (11.46–15.37)	13.62 (11.26–15.05)	12.08 (10.12–14.04)	11.86 (10.54–14.5)	14.06 (12.53–15.74)	17.10 (13.89–21.54)	15.83 (13.22–18.18)	< 0.001
	SIII	6.50 (4.37–9.43)	6.91 (5.09–8.96)	7.56 (3.49–9.30)	6.14 (3.57–10.14)	6.40 (5.16–9.05)	8.14 (7.72–16.96)	10.67 (8.86–13.86)	0.003
	SIV	11.64 (10.74–13.10)	11.05 (9.3–13.67)	9.88 (9.09–11.42)	8.84 (7.11–10.93)	9.31 (7.97–11.67)	9.59 (6.96–12.84)	10.24 (8.11–11.92)	0.002
	SV	13.55 (11.42–15.61)	14.47 (12.13–15.78)	14.54 (12.10–16.05)	14.75 (10.91–16.52)	13.29 (10.23–15.69)	9.81 (6.75–14.03)	11.97 (9.7–14.67)	0.07
	SVI	12.11 (9.79–14.18)	12.94 (8.93–14.62)	12.50 (9.08–15.84)	14.46 (10.00–17.44)	14.14 (10.72–16.23)	11.41 (8.63–15.37)	10.87 (7.8–13.36)	0.27
	SVII	17.01 (15.22–18.81)	16.66 (13.52–19.61)	18.32 (16.69–21.35)	18.03 (14.92–22.03)	17.65 (14.37–21.41)	16.75 (16.03–19.18)	16.38 (14.38–18.89)	0.15
	SVIII	20.60 (18.57–23.8)	19.71 (18.76–21.8)	20.50 (17.37–22.39)	18.96 (16.9–20.43)	18.29 (15.49–20.45)	17.73 (14.28–18.46)	17.37 (16.12–19.46)	< 0.001
Attenuation of segments (Median HU)	SI	28.09 (23.28–35.14)	26.52 (24.14–30.38)	28.74 (24.08–36.5)	27.27 (24.89–30.69)	27.23 (25.09–32.3)	27.73 (23.82–64.71)	28.93 (23.52–36.28)	0.91
	SII	33.4 (30.12–41)	34 (25.9–37.08)	32.75 (29.38–40.12)	32.69 (30.96–38.89)	31.37 (27.26–38.32)	29.08 (26.38–50.01)	29.86 (24.38–37.83)	0.52
	SIII	36.19 (31.87–40.33)	33.66 (27.88–38.92)	34.8 (27.45–40.09)	36.31 (31.57–39.49)	35.62 (30.51–40.5)	33.78 (28.5–35.9)	29.82 (26.6–36.5)	0.33
	SIV	33.86 (28.94–37.92)	31.48 (26.69–37.24)	35.18 (29.43–43.26)	40.52 (30.67–46.57)	34.54 (29.22–43.57)	33.77 (30.58–38.84)	34.98 (27.58–43.27)	0.73
	SV	28.13 (23.21–33.79)	27.56 (23.18–32.27)	25.76 (23.42–31.53)	30.53 (23.85–33.59)	28.96 (23.6–33.35)	28.68 (25.74–36.54)	25.52 (21.56–32.21)	0.76
	SVI	23.64 (20.26–28.24)	21.49 (17.1–27.57)	22.14 (18.13–25.31)	22.37 (19.18–25.4)	21.84 (18.66–28.9)	21.94 (17.07–29.39)	20.51 (15.87–24.92)	0.38
	SVII	27.52 (24.75–52.24)	27.33 (23.22–50.11)	28.3 (25.42–86.64)	27.23 (24.64–46.75)	27.22 (23.27–40.75)	28.27 (24.08–37.62)	26.83 (21.57–52.1)	0.81
	SVIII	36.94 (31.45–79.9)	35.72 (32.12–57.95)	37.97 (34.28–103.86)	40.93 (34.46–76.58)	34.54 (29.25–40.02)	36 (33.01–52.34)	39.09 (28.66–65.62)	0.52
Attenuation of segments (Standard deviation)	SI	100.5 (83.75–110.5)	91 (80.5–103.5)	107.5 (87.75–115.75)	94 (83–108)	85.5 (70–94.75)	83 (79.25–93)	93 (79–103.25)	0.04
	SII	97.5 (83.25–111)	93.5 (79–104.5)	108 (85.5–117.25)	95 (81–109)	79 (68.25–100.5)	81.5 (79.5–97.5)	92.5 (81.75–104.5)	0.06
	SIII	97 (82.25–109)	95 (80–104)	105 (86.75–119.75)	95 (83–107)	80.5 (66.5–97.25)	85 (79–97.25)	92.5 (81–104.25)	0.06
	SIV	96 (82.25–110.75)	92.5 (78.25–105)	110 (84.25–117.5)	96 (78–111)	78.5 (65.25–92.75)	83 (80–95.5)	93 (80.75–104.5)	0.04
	SV	98.5 (82.25–112)	96.5 (80.75–106)	114 (87.75–125.25)	96 (84–116)	81 (69.5–93.5)	89 (84–96.5)	92 (81–106.25)	0.03
	SVI	101 (83.25–112.5)	96 (80.5–103.75)	114.5 (88–125.5)	97 (85–116)	84 (66.25–93.25)	89.5 (85.5–99)	93.5 (81.75–107.5)	0.02
	SVII	94.5 (80–107.75)	91 (78.75–100)	107.5 (84.75–116)	93 (81–112)	80.5 (61–85.75)	84.5 (80–93.75)	91 (80.5–104.25)	0.03
	SVIII	96 (79.5–110.5)	93.5 (79.25–101.25)	109.5 (84.75–118.75)	93 (83–112)	80.5 (65.5–87.5)	86.5 (81.5–94.5)	92 (83.5–105.25)	0.04

Note.—The values in the cell represent the median and 1st–3rd interquartile values of each parameter.

* Volume proportion was calculated by the volume of each Couinaud segment divided by the entire liver volume.

Table E4

Evaluation of Automated Segmentation Compared with Manual Segmentation (Reader2) Between Dataset 1 and Dataset 2

	Sample of Dataset 1 (*n* = 35)	Sample of Dataset 2 (*n* = 35)	Wilcoxon *P* Value
Dice similarity coefficient			
Whole liver volume	0.981 (0.978–0.985)	0.979 (0.971–0.981)	0.02
Spleen volume	0.957 (0.952–0.962)	0.950 (0.937–0.960)	0.05
Segments I+II+III	0.916 (0.906–0.932)	0.925 (0.909–0.935)	0.54
Segments IV+V+VI+VII+VIII	0.968 (0.963–0.971)	0.965 (0.951–0.971)	0.07
Segment I	0.637 (0.595–0.703)	0.656 (0.590–0.692)	0.99
Segment II	0.910 (0.900–0.929)	0.917 (0.867–0.938)	0.71
Segment III	0.875 (0.823–0.901)	0.877 (0.812–0.902)	0.83
Segment IV	0.822 (0.803–0.862)	0.854 (0.793–0.894)	0.29
Segment V	0.839 (0.806–0.88)	0.820 (0.766–0.869)	0.41
Segment VI	0.847 (0.788–0.875)	0.799 (0.751–0.855)	0.12
Segment VII	0.887 (0.834–0.916)	0.877 (0.854–0.921)	0.58
Segment VIII	0.856 (0.805–0.872)	0.872 (0.825–0.895)	0.16
Mean Hausdorff distance (mm)			
Whole liver volume	1.1 (0.9–1.2)	1.2 (0.9–1.6)	0.23
Spleen volume	0.7 (0.5–0.9)	0.7 (0.6–0.9)	0.44
Segments I+II+III	1.1 (0.9–1.3)	1.1 (0.9–1.7)	0.27
Segments IV+V+VI+VII+VIII	1.5 (1.3–1.7)	1.6 (1.2–1.9)	0.29
Segment I	4.7 (3.5–6.0)	3.9 (3.4–5.1)	0.11
Segment II	1.3 (1.1–1.7)	1.3 (1.0–2.1)	0.76
Segment III	2.0 (1.3–2.5)	1.7 (1.3–3.2)	0.70
Segment IV	2.3 (1.8–2.7)	1.7 (1.4–2.9)	0.21
Segment V	3.1 (2.3–4.5)	3.4 (2.3–4.2)	0.92
Segment VI	3.0 (2.0–4.1)	3.1 (2.3–4.5)	0.46
Segment VII	2.2 (1.5–3.4)	2.3 (1.4–3.3)	0.99
Segment VIII	3.0 (2.4–3.9)	2.7 (2.1–3.8)	0.46

Note.—The values in the cell represent the median and 1st–3rd interquartile values of each parameter.

Table E5

Automatically Measured Parameter Values Between Dataset 1 and Dataset 2

		Dataset 1		Dataset 2		Wilcoxen *P* Value
		Median (IQR)	*n*	Median (IQR)	*n*	
Whole liver volume (mL)	All stages	1788 (1528–2090)	406	1855 (1482–2291)	207	0.17
	Cirrhosis	1803 (1294–2141)	148	1787 (1331–2109)	41	0.75
	Advanced fibrosis*	1825 (1531–2144)	207	1927 (1331–2291)	93	0.58
Spleen volume (mL)	All stages	383 (260–668)	406	328 (209–522)	207	< 0.001
	Cirrhosis	736 (499–1049)	148	360 (290–717)	41	< 0.001
	Advanced fibrosis*	479 (294–794)	207	422 (262–640)	93	0.06
LSVR	All stages	0.37 (0.29–0.46)	406	0.34 (0.28–0.45)	207	0.01
	Cirrhosis	0.48 (0.39–0.56)	148	0.45 (0.36–0.60)	41	0.97
	Advanced fibrosis*	0.40 (0.31–0.50)	207	0.38 (0.27–0.55)	93	0.44

Note.—The values in the cell represent the median and 1st–3rd interquartile values of each parameter.

* Advanced fibrosis includes cirrhosis patients.

Table E6

AUC Values of Multivariable Models for Predicting Cirrhosis and Advanced Fibrosis

	Test set (20%) of Dataset1		Dataset2			
	Metavir System		Knodell HAI System		Ishak Staging System	
	Advanced Fibrosis (F0–1 vs F2–4) *	Cirrhosis (F0–3 vs F4)	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)
Manual measurements						
S+L	0.85 (0.75–0.95)	0.93 (0.88–0.98)	—	—	—	—
W+S+L	0.85 (0.74–0.95)	0.93 (0.88–0.99)	—	—	—	—
Automated measurements						
S+L	0.82 (0.71–0.93)	0.9 (0.84–0.97)	0.72 (0.65–0.8)	0.77 (0.69–0.85)	0.72 (0.65–0.8)	0.79 (0.71–0.86)
W+S+L	0.81 (0.7–0.92)	0.91 (0.84–0.98)	0.73 (0.65–0.8)	0.75 (0.67–0.84)	0.73 (0.65–0.8)	0.77 (0.69–0.85)
V	0.73 (0.61–0.86)	0.79 (0.68–0.89)	0.71 (0.63–0.78)	0.76 (0.66–0.86)	0.71 (0.63–0.78)	0.76 (0.66–0.85)
M	0.68 (0.55–0.81)	0.7 (0.58–0.83)	0.53 (0.45–0.61)	0.59 (0.49–0.69)	0.53 (0.45–0.61)	0.61 (0.51–0.71)
D	0.75 (0.63–0.87)	0.74 (0.62–0.85)	0.59 (0.51–0.67)	0.61 (0.51–0.7)	0.59 (0.51–0.67)	0.6 (0.51–0.69)
S+V	0.79 (0.68–0.91)	0.93 (0.88–0.98)	0.74 (0.67–0.81)	0.77 (0.69–0.85)	0.74 (0.67–0.81)	0.78 (0.71–0.86)
S+L+D	0.8 (0.69–0.91)	0.91 (0.85–0.97)	0.7 (0.63–0.78)	0.75 (0.67–0.84)	0.7 (0.63–0.78)	0.76 (0.68–0.85)
S+L+V+D	0.8 (0.69–0.91)	0.94 (0.89–0.99)	0.71 (0.64–0.78)	0.78 (0.7–0.86)	0.71 (0.64–0.78)	0.79 (0.71–0.87)
P value between multivariable model AUCs (noninferiority†)						
Manual S+L vs Automated S+L	< 0.001	< 0.001	—	—	—	—
Manual S+L vs Automated S+L+V+D	< 0.001	< 0.001	—	—	—	—
Automated S+L vs Automated W+S+L	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Automated S+L vs Automated S+V	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Automated S+L vs Automated S+L+D	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Automated S+L vs Automated S+L+V+D	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Note.—Dataset1 uses the Metavir biopsy staging system and Dataset2 uses both the Knodell histologic activity index (HAI) and Ishak staging system. The values in the cell represent the area under the receiver operating characteristic curve (AUC) and DeLong 95% confidence intervals (CI). Volume proportion was calculated by the volume of each Couinaud segment divided by the entire liver volume. LSVR = liver segmental volume ratio. Multivariable model names – [S+L]: Multivariable model using [S]pleen volume and [L]SVR, [W+S+L]; Multivariable model using [W]hole liver volume, [S]pleen volume, and [L]SVR; [V]: Multivariable model using [V]olume proportions of all the liver Couinaud segments; [M]: Multivariable model using [M]edian Hounsfield Unit attenuation of all the liver Couinaud segments; [D]: Multivariable model using Standard [D]eviation of the attenuation in all the liver Couinaud segments; [S+V]: Multivariable model using [S]pleen volume and [V]olume proportions of all the liver Couinaud segments; [S+L+D]: Multivariable model using [S]pleen volume, [L]SVR, and Standard [D]eviation of the attenuation in all the liver Couinaud segments; [S+L+V+D]: Multivariable model using [S]pleen volume, [L]SVR, [V]olume proportions, and Standard [D]eviation of the attenuation in all the liver Couinaud segments.

* Advanced fibrosis includes patients with cirrhosis.

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

Table E7

AUC Values of Dataset2 Divided into HCV and Non-HCV Groups for Predicting Cirrhosis and Advanced Fibrosis

		Entire Dataset2 (*n* = 207)		HCV-only patients in Dataset2 (*n* = 79)		Non-HCV patients in Dataset2 (*n* = 128)	
		Advanced Fibrosis (Ishak 0–2 vs 3–6)*	Cirrhosis (Ishak 0–4 vs 5–6)	Advanced Fibrosis (Ishak 0–2 vs 3–6)*	Cirrhosis (Ishak 0–4 vs 5–6)	Advanced Fibrosis (Ishak 0–2 vs 3–6)*	Cirrhosis (Ishak 0–4 vs 5–6)
Univariable	Whole liver volume (mL)	0.49 (0.41–0.57)	0.46 (0.44–0.64)	0.7 (0.58–0.82)	0.61 (0.47–0.74)	0.4 (0.49–0.72)	0.31 (0.53–0.85)
	Spleen volume (mL)	0.66 (0.58–0.73)	0.65 (0.55–0.74)	0.79 (0.69–0.89)	0.68 (0.56–0.8)	0.59 (0.48–0.71)	0.61 (0.45–0.77)
	LSVR	0.63 (0.55–0.71)	0.75 (0.66–0.85)	0.65 (0.53–0.77)	0.79 (0.67–0.9)	0.59 (0.48–0.71)	0.69 (0.52–0.87)
Multivariable models	S+L	0.72 (0.65–0.8)	0.79 (0.71–0.86)	0.85 (0.76–0.94)	0.82 (0.72–0.91)	0.64 (0.53–0.75)	0.75 (0.6–0.9)
	W+S+L	0.73 (0.65–0.8)	0.77 (0.69–0.85)	0.85 (0.76–0.93)	0.76 (0.65–0.88)	0.65 (0.54–0.76)	0.76 (0.62–0.91)
	V	0.71 (0.63–0.78)	0.76 (0.66–0.85)	0.67 (0.55–0.79)	0.75 (0.63–0.87)	0.70 (0.59–0.81)	0.72 (0.54–0.91)
	M	0.53 (0.45–0.61)	0.61 (0.51–0.71)	0.48 (0.33–0.62)	0.65 (0.52–0.77)	0.56 (0.45–0.67)	0.56 (0.37–0.74)
	D	0.59 (0.51–0.67)	0.6 (0.51–0.7)	0.61 (0.47–0.75)	0.66 (0.54–0.78)	0.59 (0.48–0.7)	0.55 (0.39–0.72)
	S+V	0.74 (0.67–0.81)	0.78 (0.71–0.86)	0.79 (0.69–0.89)	0.78 (0.67–0.88)	0.72 (0.61–0.82)	0.77 (0.63–0.91)
	S+L+D	0.7 (0.63–0.78)	0.76 (0.68–0.85)	0.82 (0.72–0.91)	0.82 (0.72–0.91)	0.65 (0.54–0.76)	0.72 (0.55–0.88)
	S+L+V+D	0.71 (0.64–0.78)	0.79 (0.71–0.87)	0.77 (0.67–0.88)	0.82 (0.72–0.91)	0.68 (0.58–0.79)	0.73 (0.57–0.9)

Note.—This table shows the performance of automated measurements in predicting advanced fibrosis or cirrhosis in all subjects of Dataset2 (left column, *n* = 207, median age 50), a subset of Dataset2 including only hepatitis C (HCV) patients (center column, *n* = 79, median age 47), and a subset of Dataset2 including all other etiologies including other viral hepatitis and steatohepatitis (right column, *n* = 128, median age 53). Multivariable model names – [S+L]: Multivariable model using [S]pleen volume and [L]SVR; [W+S+L]: Multivariable model using [W]hole liver volume, [S]pleen volume, and [L]SVR; [V]: Multivariable model using [V]olume proportions of all the liver Couinaud segments; [M]: Multivariable model using [M]edian Hounsfield Unit attenuation of all the liver Couinaud segments; [D]: Multivariable model using Standard [D]eviation of the attenuation in all the liver Couinaud segments; [S+V]: Multivariable model using [S]pleen volume and [V]olume proportions of all the liver Couinaud segments; [S+L+D]: Multivariable model using [S]pleen volume, [L]SVR, and Standard [D]eviation of the attenuation in all the liver Couinaud segments; [S+L+V+D]: Multivariable model using [S]pleen volume, [L]SVR, [V]olume proportions, and Standard [D]eviation of the attenuation in all the liver Couinaud segments.

* Advanced fibrosis includes cirrhosis patients.

RSNA

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

Key Result

The deep learning (DL) model performed comparably with radiologists in predicting cirrhosis on contrast-enhanced CT.

Patients:

Dataset 1: 406 with hepatitis C
- 148 with cirrhosis

Dataset 2: 207 with multiple cirrhotic etiologies
- 41 with cirrhosis

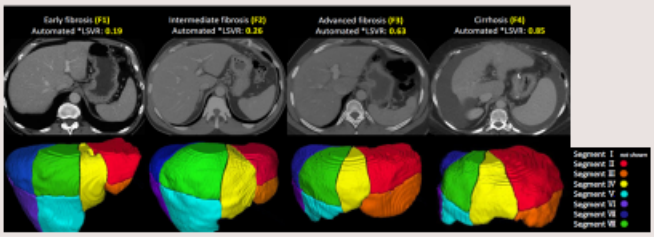
Methods:

- Whole liver, liver segmental ratio (LSVR), and splenic volumes were measured by the DL model and two radiologists.
- Multivariable models were built using combinations of automated measurements.

Results:

Measurement Type	AUC	95% CI
Automated		
Spleen volume	0.85	0.81, 0.89
LSVR	0.79	0.74, 0.84
Manual		
Spleen volume	0.86	0.82, 0.90
LSVR	0.83	0.78, 0.87

- A multivariable model using LSVR, splenic volume and attenuation of liver segments achieved the highest performance in predicting cirrhosis (AUC, 0.94).



Lee S et al. Published Online: August 24, 2022
https://doi.org/10.1148/ryai.210268

Radiology: Artificial Intelligence