Deep-learning-based hepatic fat assessment (DeHFt) on non-contrast chest CT and its association with disease severity in COVID-19 infections: A multi-site retrospective study

Gourav Modanwal,[a,*,i] Sadeer Al-Kindi,[b,i] Jonathan Walker,[b] Rohan Dhamdhere,[a] Lei Yuan,[c] Mengyao Ji,[d] Cheng Lu,[e,f] Pingfu Fu,[g] Sanjay Rajagopalan,[b] and Anant Madabhushi[a,h]

[a]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA
[b]Department of Medicine, Case Western Reserve University, Cleveland, OH, USA
[c]Department of Information Center, Renmin Hospital of Wuhan University, Wuhan, Hubei, China
[d]Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, Hubei, China
[e]Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, 510080, China
[f]Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, 510080, China
[g]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, 44106, USA
[h]Atlanta Veterans Administration Medical Center, Atlanta, GA, USA

Summary

Background Hepatic steatosis (HS) identified on CT may provide an integrated cardiometabolic and COVID-19 risk assessment. This study presents a deep-learning-based hepatic fat assessment (DeHFt) pipeline for (a) more standardised measurements and (b) investigating the association between HS (liver-to-spleen attenuation ratio <1 in CT) and COVID-19 infections severity, wherein severity is defined as requiring invasive mechanical ventilation, extracorporeal membrane oxygenation, death.

Methods DeHFt comprises two steps. First, a deep-learning-based segmentation model (3D residual-UNet) is trained (N = 80) to segment the liver and spleen. Second, CT attenuation is estimated using slice-based and volumetric-based methods. DeHFt-based mean liver and liver-to-spleen attenuation are compared with an expert's ROI-based measurements. We further obtained the liver-to-spleen attenuation ratio in a large multi-site cohort of patients with COVID-19 infections (D1, N = 805; D2, N = 1917; D3, N = 169) using the DeHFt pipeline and investigated the association between HS and COVID-19 infections severity.

Findings The DeHFt pipeline achieved a dice coefficient of 0.95, 95% CI [0.93–0.96] on the independent validation cohort (N = 49). The automated slice-based and volumetric-based liver and liver-to-spleen attenuation estimations strongly correlated with expert's measurement. In the COVID-19 cohorts, severe infections had a higher proportion of patients with HS than non-severe infections (pooled OR = 1.50, 95% CI [1.20–1.88], P < .001).

Interpretation The DeHFt pipeline enabled accurate segmentation of liver and spleen on non-contrast CTs and automated estimation of liver and liver-to-spleen attenuation ratio. In three cohorts of patients with COVID-19 infections (N = 2891), HS was associated with disease severity. Pending validation, DeHFt provides an automated CT-based metabolic risk assessment.

Funding For a full list of funding bodies, please see the Acknowledgements.

Keywords: COVID-19; NAFLD; Hepatic steatosis

Research in context

Evidence before this study

Deep-learning has shown great promise in its ability to facilitate automated liver segmentation and attenuation assessment, resulting in more standardised and reproducible measurements. Hepatic attenuation on computed tomography (CT) is a surrogate marker for cardiometabolic risk, including Type 2 diabetes and its progression. Despite its clinical importance, hepatic fat is not routinely assessed on coronary artery calcium (CAC) CTs owing to the difficulty in the accurate manual delineation of regions of interest (ROI), inter-reader variability in measurements, lack of consensus on measurement methodology, and spatial heterogeneity in liver attenuation.

To overcome these problems, we present a deep-learning-based hepatic fat assessment (DeHFt) pipeline for more standardised measurement. We utilise our pipeline to investigate further the association between HS (defined as liver-to-spleen attenuation ratio <1 in CT) and COVID-19 infection severity in a large multi-site cohort of patients with COVID-19 infections (N = 2891), wherein severity is defined as requiring invasive mechanical ventilation, extracorporeal membrane oxygenation, or death.

Added-value of this study

In this study, we showed that DeHFt could facilitate automated liver segmentation and attenuation assessment, resulting in more standardised and reproducible measurements of Hepatic steatosis (HS). The result demonstrated that DeHFt-based liver and liver-to-spleen attenuation estimations strongly correlated with manual ROI-based measurement performed by the expert. In multi-site COVID-19 cohorts (N = 2891), the severe infections had a higher proportion of patients with HS than non-severe infections (pooled OR 1.50, 95% CI [1.20–1.88])

Implications of all the available evidence

DeHFt pipeline may provide an integrated cardiometabolic and COVID-19 risk assessment by identifying HS on computed tomography (CT). The approach can potentially enable automation of HS assessment in clinical practice and thus help identify high-risk metabolic patients. It can also be applied to large datasets to understand the association between liver steatosis and cardiometabolic outcomes. DeHFt can be used as an aid for comprehensive risk assessment of patients with COVID-19 infection to assign care level (e.g., inpatient vs outpatient) and decision for early use of therapeutics (e.g., antivirals).

Introduction

Obesity and prediabetes affect a significant proportion of the American population (estimated at 90 million individuals), placing them at risk for future Type 2 diabetes (T2D).[1] As a chronic condition posing the highest risk for cardiovascular disease, T2D presents both a challenge and an opportunity for preventive approaches. Hepatic steatosis (HS, defined as liver-to-spleen attenuation ratio <1 in CT)[2] enhances the identification of patients with marked insulin resistance at risk for future progression to T2D and represents a risk predictor for future CV disease, thus representing a marker of cardiometabolic risk.[3]

Liver biopsy is used as the reference gold standard for assessing liver fat content and diagnosis of HS.[4] However, it is not widely used due to its invasive nature and high cost. Noninvasive techniques include ultrasound (US),[5] magnetic resonance imaging (MRI),[6] and computed tomography (CT). Among the noninvasive imaging techniques to quantify liver fat, MRI proton density fat fraction (PDFF)[6] is used as a reference standard for estimating the presence and grading of HS. Some studies[7,8] have recently found the correlation between PDFF and liver fat, allowing for the estimation of liver fat content using non-contrast CT. Thus, providing an opportunistic imaging-based risk assessment.[9] HS identified on CT may provide additional metabolic risk assessment in patients undergoing coronary artery calcium scoring, facilitating an integrated cardiometabolic risk assessment.[10]

Despite its clinical importance, hepatic fat is not routinely assessed on CTs (including those done for coronary artery calcium (CAC) scoring) owing to the difficulty in the accurate manual delineation of regions of interest (ROI), inter-reader variability in measurements, lack of consensus on measurement methodology, and spatial heterogeneity in liver attenuation. Convolutional Neural Networks (CNN) have shown great promise in their ability to facilitate automated liver segmentation and attenuation assessment, resulting in more standardised and reproducible measurements.[11] Prior studies[12–17] have suggested that deep-learning-based methods provide accurate liver segmentation on full-dose CTs with and without contrast. Liver segmentation on low-dose non-contrast CTs poses a challenge given the signal heterogeneity, lack of border contrast, and partial liver imaging. The manual ROI-based measurement limits the opportunistic screening for HS in a population-based study. In this work, we present a deep-learning-based hepatic fat assessment (DeHFt) pipeline for (a) liver and spleen segmentation and (b) CT attenuation estimation on non-contrast CT scans performed for CAC screening.

Further, HS is one of the most common diseases resulting from obesity's vast range of systemic

metabolic dysfunctions. Recently, few studies have shown that obesity has also been linked to more hospitalisations, mechanical ventilation due to acute respiratory failure, extensive coagulopathy, and, finally, death.[18,19] Recent studies[20,21] showed that HS was higher in the RT-PCR positive group than in controls. Among the COVID-19 positive group, hepatic manifestation was independently linked to the severity of coronavirus disease.[22] However, not many studies have investigated the association of HS with the severity of COVID-19 infections in a large multi-site cohort of patients. Consequently, we utilised the DeHFt pipeline to investigate the association of HS with disease severity in a large multi-site cohort of patients with COVID-19 infections (D1, N = 805; D2, N = 1917; D3, N = 169), wherein severity is defined as requiring invasive mechanical ventilation, extracorporeal membrane oxygenation, or death.

Methods
Inclusion and exclusion criteria for cohort construction
CAC CTs
We obtained a convenience sample of 129 anonymised CTs done for CAC assessment as part of clinical care at University Hospitals in Cleveland, OH, between January 2014 and December 2020. Participants were at least 45 years of age with no known coronary artery disease and at least one cardiovascular risk factor (smoking, hyperlipidemia, hypertension, diabetes, family history of heart disease, or chronic inflammatory conditions), refer to Fig. 1a. CTs were acquired using a standardised protocol. Subjects were positioned within the gantry of the MDCT scanner in the supine position. During a single breath-hold, images of the heart, from the level of the tracheal bifurcation to below the base of the heart, were acquired using prospective ECG triggering at 50–80% of the RR-interval, depending on the heart rate. Scan parameters: 16 × 1.5 mm collimation, 205 mm field of view (FOV), variable rotation time (scanner specific), 120 kVp, and 40–70 mAs (based on weight). This study was approved by the Institutional Review Board at University Hospitals.

COVID-19 CTs
Following the inclusion and exclusion criteria (Fig. 1b), we additionally obtained non-contrast chest CTs from 805 patients with COVID-19 infections admitted to Renmin Hospital of Wuhan University, Hubei General Hospital (D1). Another cohort used for the study was the STOIC2021-COVID-19 AI Challenge dataset (D2),[23] containing non-contrast chest CTs from 1917 patients suspected of being infected with COVID-19 during the first wave of the pandemic in France. The third cohort used for the study was the Stony Brook University
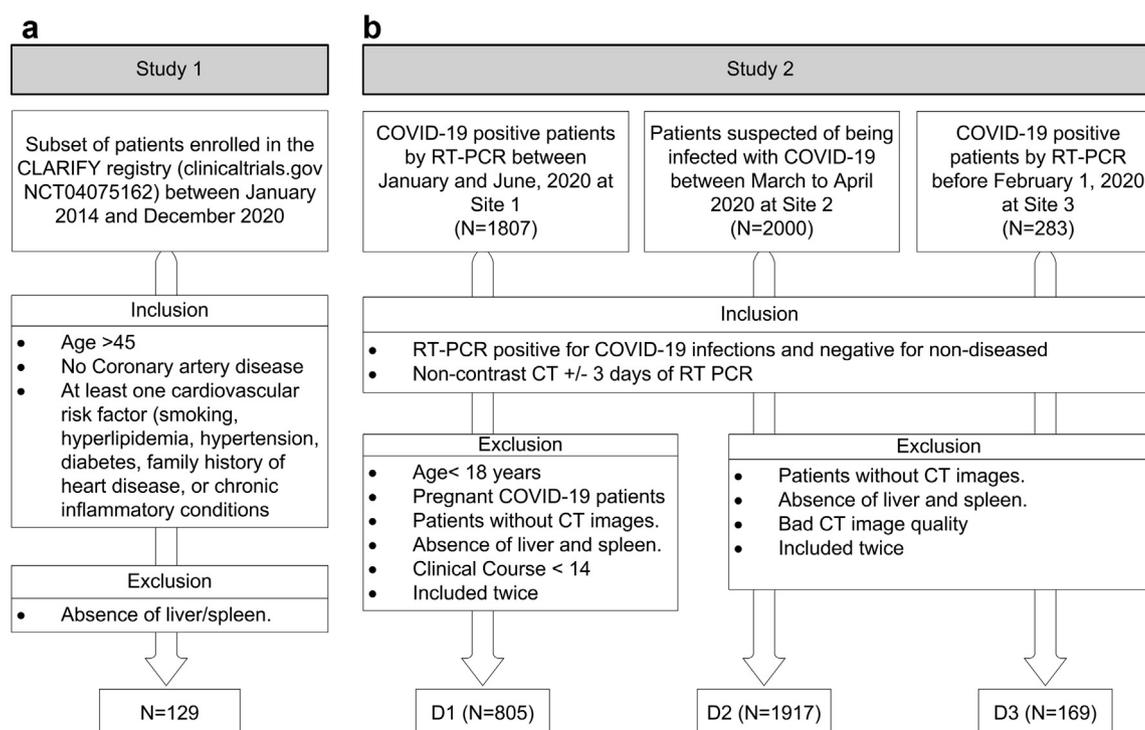
Fig. 1: Patient flow chart including patient enrollment, inclusion, and exclusion criteria.

COVID-19 Positive Cases dataset (D3),[24] which contains non-contrast chest CTs from 169 patients who tested positive for COVID-19. Patients were categorised into two groups based on the disease severity[25]: severe (requiring invasive mechanical ventilation, extracorporeal membrane oxygenation, or death) vs non-severe (no invasive ventilator support (no respiratory distress, oxygen supplementation, non-invasive ventilation). Altogether, three different cohorts of patients—D1, N = 805 (465 severe and 340 non-severe); D2, N = 1917 (288 severe, 887 non-severe, and 742 non-diseased); D3, N = 169 (47 severe and 122 non-severe)—were used to explore the association between automated HS quantification by the DeHFt pipeline with clinical severity of COVID-19 infections, given the known associations between HS and COVID-19 outcomes.[19]

Data annotation and ROI measurements
For each CT, a trained expert—with ten years of cardiac CT segmentation experience—performed manual segmentation and manual ROI-based CT attenuation measurements on CAC CTs. For segmentation, the expert performed manual demarcation of the liver and spleen on all slices using 3D Slicer software.[26] For the manual ROI-based CT attenuation measurements, circular ROI measuring 5–10 cm^2 were placed at different locations (three on the liver and one in the spleen), avoiding vasculature, cysts, blood vessels, bile ducts, calcification, and other heterogeneous areas (refer to Fig. S1). The liver-to-spleen attenuation ratio was obtained by dividing the mean Hounsfield units (HU) across the three ROIs in the liver by the mean HU of the spleen ROI as previously described.[2] The manual ROI-based measurements were performed on 49 anonymised validation scans. The same expert was asked to perform the repeated measurement on the same set of cases (after a washout period of a month) for intra-reader variability analysis. The human expert performed all these ROI-based measurements without knowing the automated results of the liver and liver-to-spleen attenuation ratio.

Deep-learning-based hepatic fat estimation (DeHFt)
The schematic for the DeHFt pipeline is illustrated in Fig. 2. It is a two-stage method; first, a deep-learning-based model is presented to segment the entire imaged liver and spleen. Second, CT attenuation is measured on the imaged organ using the DeHFt pipeline. Additional details are provided below.

Deep-learning-based liver and spleen segmentation
We utilised low-dose CAC CTs (N = 129) to train and validate the deep-learning-based segmentation models to segment the liver and spleen. We trained four state-of-the-art, convolutional neural network-based segmentation models[27–30] with random 80 low-dose CAC CT scans and evaluated their accuracy on the remaining 49

independent validation CAC CTs. The performance of these models was compared based on the dice similarity coefficient (DSC).[31] DSC is a spatial overlap index that ranges from 0 to 1, with 0 indicating no spatial overlap and 1 representing the total overlap between the predicted and ground truth segmentations. The best performing model (3D residual-UNet architecture from nnUnet framework) was trained for 1000 epochs using 3D image patches of size = (28, 256, 256) and batch = 2. The model was trained de novo (i.e., without any pre-training) with five-fold cross-validation. A schematic representation of the model architecture can be found in Fig. 3. We employed an Nvidia Tesla V100 GPU with 32 GB memory to train the model. The model training took a total of 43 h of computation on the GPU. The final trained model took ~1 min to segment the liver and spleen for each patient.

CT attenuation measurement
DeHFt pipeline employed two methods of measuring the CT attenuation automatically: slice-based and volumetric-based. In a slice-based estimation, the DeHFt pipeline identifies the slice with the maximum area and obtains the mean attenuation and standard deviation across that slice. In a volumetric-based estimation of liver and liver-to-spleen attenuation, the DeHFt pipeline considers the whole liver and spleen for the measurement, and the ratio of mean HU for liver and spleen is obtained. The mean liver and mean liver-to-spleen CT attenuation are estimated using both methods and compared with manual ROI-based measurement performed by the expert. We considered the average of the two intra-measurements done by the expert as the reference measurement.

Ethics
All ethical procedures for the study were approved by the Institutional Review Board committee of record at University Hospitals, Cleveland (STUDY20200213) and the Ethics committee of the Renmin Hospital of Wuhan University (2020KS02010), and the need for written consent was waived.

Statistical analysis
The Pearson correlation coefficient (ρ) was used to measure the linear association of DeHFt-based liver/liver-to-spleen attenuation estimation with manual ROI-based measurement by the human expert. Bland–Altman plots were presented to evaluate the agreement between the two measurement methods. We used a t-test to determine if there is a statistically significant difference between the mean liver/liver-to-spleen attenuation of severe vs non-severe COVID-19 patients' group. Further, a two-proportions z-test was used to compare the proportions of patients with HS (liver-to-spleen attenuation ratio <1) in the severe and non-severe
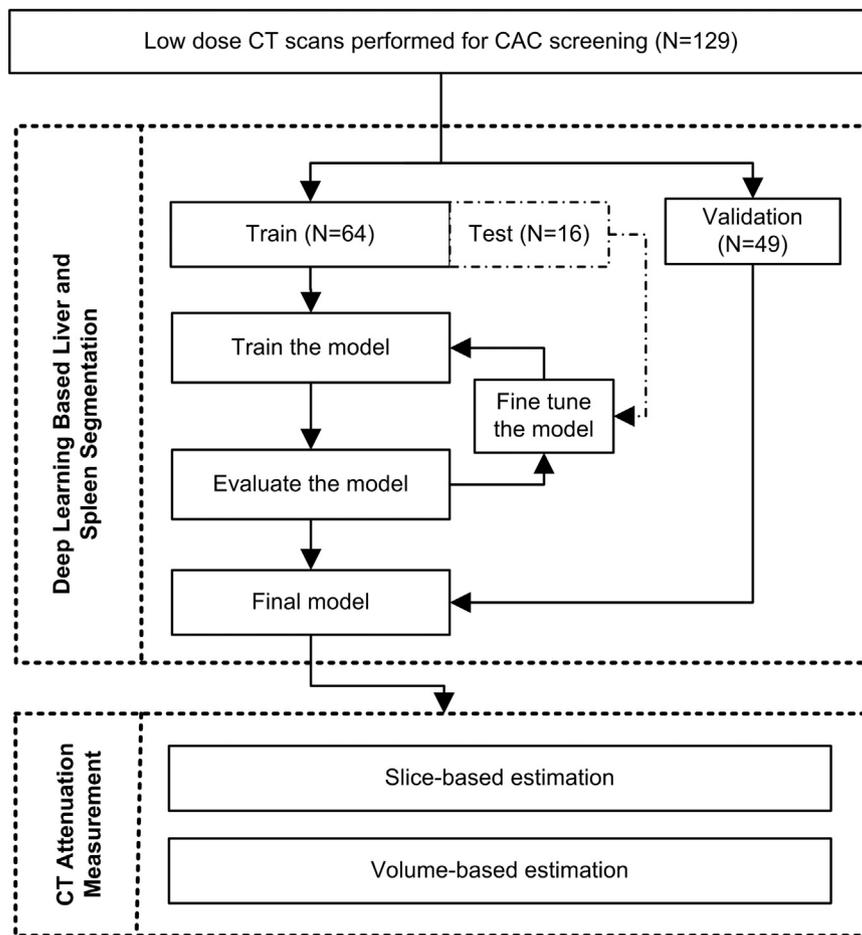
Fig. 2: Block diagram of deep-learning-based hepatic fat assessment (DeHFt) pipeline.

COVID-19 infections group. A meta-analysis was performed to estimate the pooled association between HS and severe COVID19 infection. We used Python (3.7.0) with the SciPy (1.5.2) package for statistical analyses. A P-value of less than 0·05 indicated statistical significance. P-values less than 0.05, 0.01, and 0.001 are marked with one, two, and three asterisks, respectively.

Role of the funding source
The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results
In the hold-out validation CAC cohort (N = 49), the attenuation (mean ± standard deviation) of the liver and spleen was 48.46 ± 11.30 and 42.96 ± 7.33, respectively. The mean liver-to-spleen attenuation was 1.17 ± 0.38. Fig. S2 shows the scatter plot and Bland–Altman graphs for intra-reader measurements by the expert reader.

There was a strong correlation for intra-reader variability of the liver, spleen, and liver-to-spleen attenuation measurements (0.94, 0.80, and 0.85, respectively). The mean of the liver attenuation was 48.14 ± 11.40 and 48.78 ± 11.52 (P = .12) during the intra-reader measurements by the human expert. The intra-reader mean of the spleen attenuation was 43.79 ± 8.01 and 42.12 ± 7.43 (P = .01), respectively. Similarly, the liver-to-spleen attenuation was 1.15 ± 0.41 and 1.20 ± 0.38 (P = .06), respectively. Bland–Altman graphs revealed that the bias (i.e., the mean difference between two measurements by the expert) for liver, spleen and liver-to-spleen attenuation were −0.64, 1.67, and −0.05, whereas the 95% limits of agreement range were −8.03 to 6.76, −7.81 to 11.15, and −0.49 to 0.38, respectively. The scatter plots suggest the lack of a consistent bias in the two measurements (intra-reader) by the expert.

All trained deep-learning-based segmentation models had excellent performance with DSC between 0.94 and 0.95 (refer to Table S1). The best performing model (3D residual-UNet architecture from nnUnet
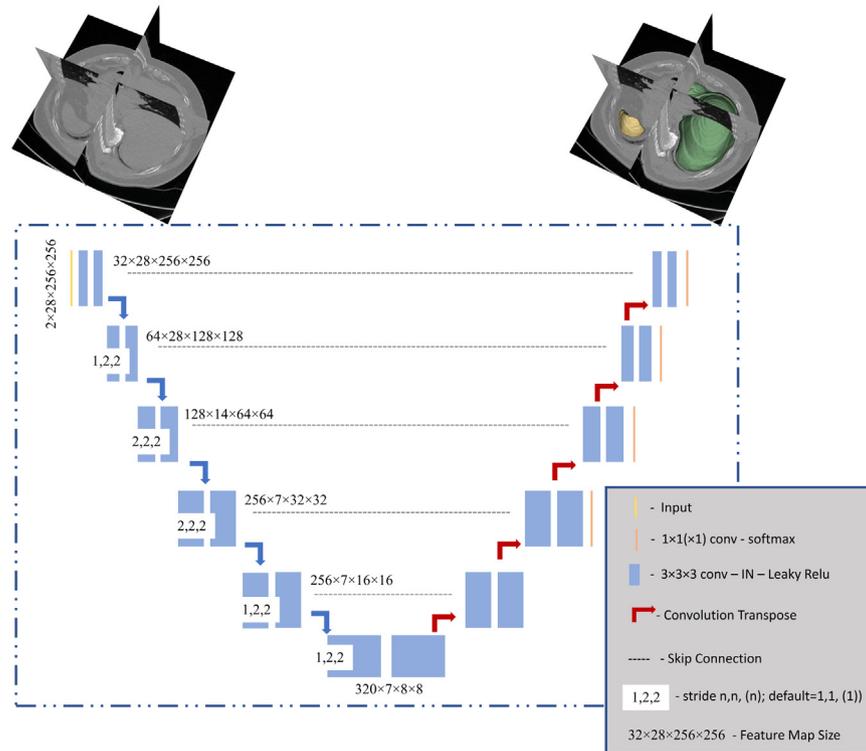
Fig. 3: Schematic representation of the 3D residual-UNet architecture.

framework) achieved a DSC of 0.95, 95% CI [0.93–0.96] on the hold-out validation set and was used for measuring CT attenuation. Fig. 4 shows an example of the automated and manual segmentation for the best, median, and worst-case according to the DSC score. The performance of the CT attenuation measurement methods is illustrated in Table 1. With ground truth segmentation (performed manually by the expert), a strong correlation was found between slice-based liver and liver-to-spleen attenuation estimation with manual ROI-based measurement (Pearson correlation coefficient $\rho = 0.97$ and $\rho = 0.94$, respectively). Similarly, volume-based liver and liver-to-spleen attenuation estimation achieved a Pearson correlation coefficient of $\rho = 0.93$ and $\rho = 0.89$ with manual ROI-based measurement, respectively. With the segmentation obtained using the DeHFt pipeline, slice-based automatic liver and liver-to-spleen attenuation estimation had a significant association with manual ROI-based measurement (Pearson correlation coefficient $\rho = 0.98$ and rho = 0.95, respectively, refer to Fig. 5a and b). With manual ROI-based liver and liver-to-spleen attenuation estimations, volume-based automatic liver and liver-to-spleen attenuation estimations achieved a Pearson correlation coefficient $\rho = 0.96$ and $\rho = 0.92$, respectively (refer to Fig. 5c and d). When compared to ground truth segmentation, the mean liver and mean liver-to-spleen CT attenuation

estimation with nnUnet[30] segmentation had a better correlation with manual ROI-based measurement.

Out of the 805 patients in the dataset (D1) with COVID-19 infections, 465 had severe infections, and 340 had non-severe infections. DeHFt-based liver-to-spleen attenuation ratio (1.19 ± 0.17 vs. 1.23 ± 0.18, P = .001) were lower in severe vs. non-severe COVID-19 patients. When categorised using traditional definitions of HS (liver-to-spleen attenuation ratio <1), patients with severe COVID-19 had a higher percentage of HS compared with non-severe COVID-19 infections (11.61% vs 8.24%, P = .12), refer to Fig. 6a. The odds ratio (OR) was 1.46, 95% CI [0.9–2.36] (refer to Table 2). Similar observations were obtained with the dataset (D2, N = 1917) and (D3, N = 169). In the D2 dataset, DeHFt-based-liver-to-spleen attenuation ratio (1.01 ± 0.31 vs. 1.05 ± 0.27, P = .03) was lower in severe (N = 288) vs. non-severe (N = 887) COVID-19 patients (refer to Fig. 6b). We also compared them with non-diseases control patients (N = 742). We found that the liver-to-spleen attenuation ratio (1.15 ± 0.30) was higher in the non-diseases control group compared to the severe (P < .001) and non-severe COVID-19 (P < .001) groups. We also found that patients with severe COVID-19 had a significantly higher percentage of HS compared with non-severe COVID-19 infections (48.96% vs 38.22%, P = .001) with OR = 1.55, 95% CI [1.19–2.02]). The non-

a 2D Axial Slices **b** 3D Volumes

Best (DSC = 0.9782)

Median (DSC=0.9589)

Worst (DSC=0.9136)

—— / 🟩 - Liver Manual Segmentation
—— / 🟨 - Spleen Manual Segmentation
----- / 🟥 - Liver Predicted Segmentation
----- / 🟦 - Spleen Predicted Segmentation
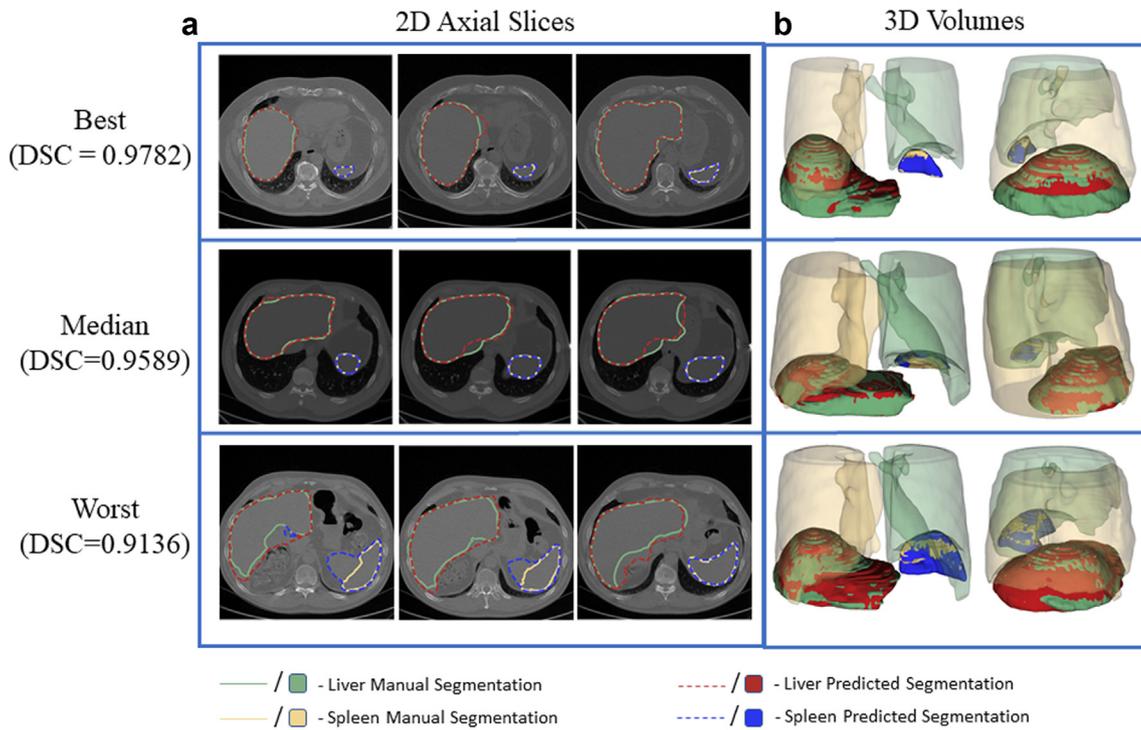
Fig. 4: Example of liver and spleen segmentation for the cases with the highest, median, and lowest DSC score (a) various 2D slices (b) 3D volume.

diseased control patients had a lower percentage of HS, with non-severe (26.95% vs. 38.22%, $P < .001$, $OR = 2.60$, 95% CI [1.96–3.45]) and severe (26.95% vs. 48.96%, $P < .001$, $OR = 1.67$, 95% CI [1.35–2.07]) COVID-19 infections. In D3, the DeHFt-based liver-to-spleen attenuation ratio was (0.93 ± 0.27 vs. 0.93 ± 0.32, $P = .88$) in severe (N = 47) vs. non-severe (N = 122) COVID-19 patients. As can be appreciated from Fig. 6c, similar findings hold for D3, where severe COVID-19 infections had a higher percentage of HS compared to non-severe COVID-19 infections (65.96% vs 59.84%, $P = .46$, $OR = 1.30$, 95% CI [0.64–2.63]). In a pooled fixed effect model including the 3 cohorts, HS was associated with increased odds for severe COVID19 infection (pooled OR 1.50, 95% CI [1.20–1.88], $P < .001$).

Discussion

In this work, we presented a fully automatic deep-learning-based model (DeHFt) to segment the liver and spleen on non-contrast computed tomography (CT) performed for coronary artery calcium (CAC) in an attempt to evaluate whole-volume liver and spleen attenuation and HS. The DeHFt model was also applied to large multi-site cohorts of COVID-19 patients to investigate if the severe COVID-19 infections were associated with lower hepatic attenuation values, in turn reflecting HS.

Liver segmentation is an essential task in medical image analysis, and various benchmarks/challenges for liver segmentation[14–16] have been organised thus far. The best liver segmentation algorithm[32] has reported a DSC

Segmentation methods	Acc.	Prec.	Recall	DSC	CT attenuation measurement methods	Liver		Liver-to-spleen	
						ρ	P-value	ρ	P-value
Manual segmentation	–	–	–	–	Slice-based	0.97	P < 0.001	0.94	P < 0.001
					Volumetric-based	0.93	P < 0.001	0.89	P < 0.001
nnUnet[30]	0.97	0.95	0.95	0.95	Slice-based	0.98	P < 0.001	0.95	P < 0.001
					Volumetric-based	0.96	P < 0.001	0.92	P < 0.001
DSC, Dice similarity coefficient; ρ, Pearson correlation coefficient.									

Table 1: Performance comparison of the segmentation methods & CT attenuation estimation.
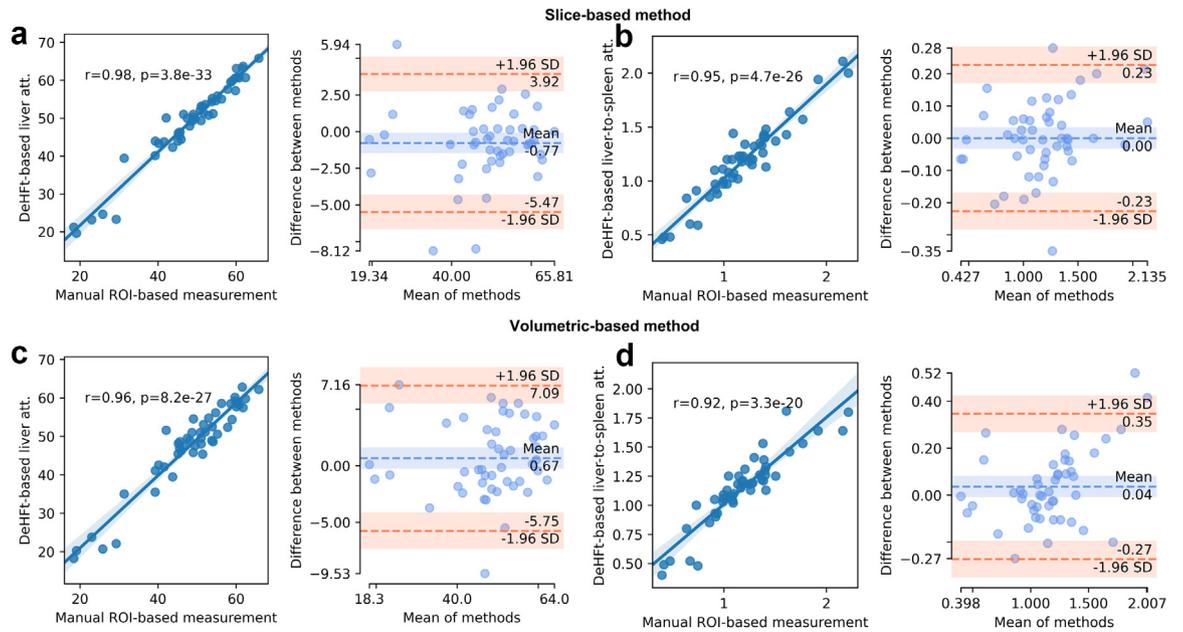
Fig. 5: Scatter plot and Bland–Altman graph showing variability in (a) Slice-based Liver attenuation estimation (b) Slice-based Liver-to-Spleen estimation (c) volumetric-based Liver attenuation estimation (d) volumetric-based Liver-to-Spleen. The dotted horizontal blue line shows the mean of the differences (=bias) between the two methods, and the doted red horizontal lines show the upper and lower 95% limits of agreement (= bias ± 1.96 × SD).

*** Significant at 0.001 level;
** Significant at 0.01 level;
* Significant at .05 level
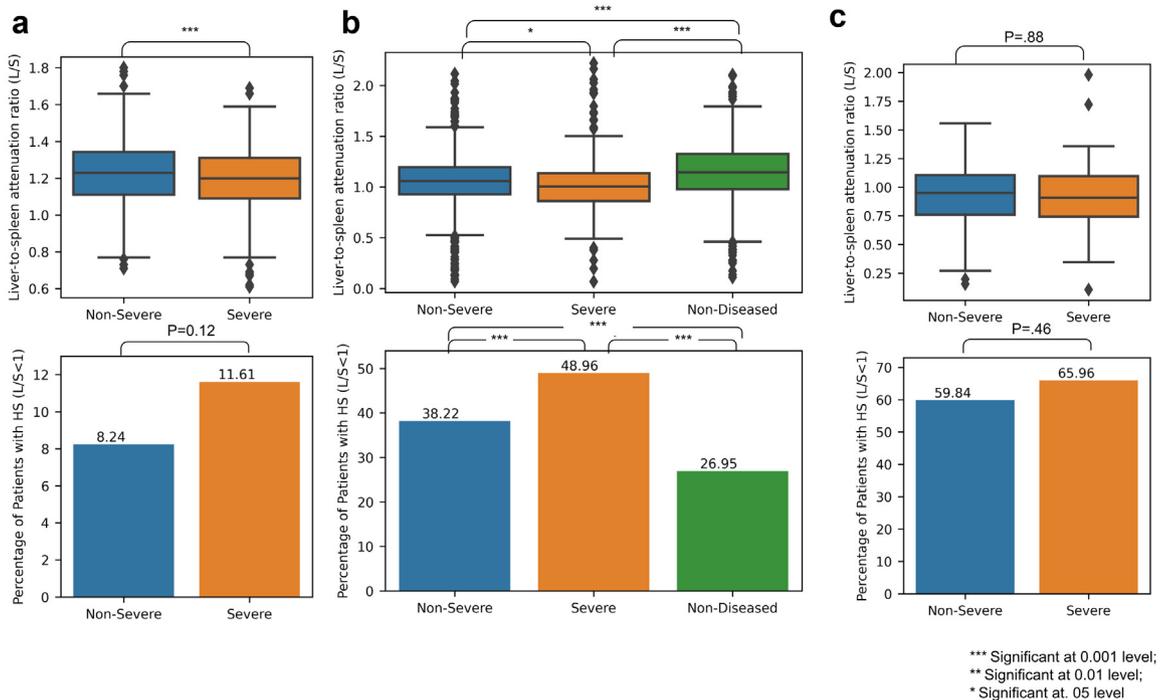
Fig. 6: Association between liver steatosis and COVID-19 infections severity in large multi-site COVID-19 cohorts. The figure shows the mean liver-to-spleen (L/S) attenuation ratio between non-severe and severe patients, and the percentage of patients with hepatic steatosis (L/S < 1) in non-severe vs. severe COVID-19 patients in (a) Renmin hospital of Wuhan University, Hubei General Hospital (D1, N = 805) (b) STOIC2021-COVID-19 AI Challenge dataset (D2, N = 1917) (c) Stony Brook University COVID-19 Positive Cases dataset (D3, N = 169). HS, hepatic steatosis.

Dataset	Outcome	Comparator	OR [95% CI]	P-value
Wuhan; D1 (N = 805)	Severe COVID19 (N = 465)	Non-severe COVID19 (N = 340)	1.46 [0.9–2.36]	P = .12
STOIC; D2 (N = 1917)	Severe COVID19 (N = 288)	Non-Severe COVID19 (N = 887)	1.55 [1.19–2.02]	P = .001
	Non-Severe COVID19 (N = 887)	Non-diseased COVID19 (N = 742)	1.67 [1.35–2.07]	P < .001
	Severe COVID19 (N = 288)	Non-diseased (N = 742)	2.60 [1.96–3.45]	P < .001
SBU: D3 (N = 169)	Severe COVID19 (N = 47)	Non-Severe COVID19 (N = 122)	1.30 [0.64–2.63]	P = .46
Total (fixed effects)	Severe COVID19 (N = 800)	Non-severe COVID19 (N = 1349)	1.50 [1.20–1.88]	P < .001

HS, hepatic steatosis; OR, odds ratio; CI, confidence interval.

Table 2: Association between HS (liver-to-spleen attenuation ratio <1) and COVID-19 infections severity in COVID-19 cohorts.

of 0.96 (in preprint). The public/challenge CT datasets for liver and spleen segmentations are near-universally performed on contrast-enhanced CT scans, while CAC scans are low-dose non-contrast CTs (thick slabs of 2.5 mm). The performance of these models is relatively poor on these low-dose non-contrast CTs.[17] In our study, we trained state-of-the-art deep-learning-based segmentation models on low-dose non-contrast CTs and achieved a DSC of 0.95, 95% CI [0.93–0.96] on the independent validation set (N = 49).

In a multi-stage algorithm, the effectiveness of the preceding stage affects the performance of the subsequent downstream analyses. Therefore, we evaluated the performance of these CT attenuation measurement methods on the segmentation obtained by the DeHFt pipeline and compared it with ground truth segmentation (performed manually by the expert). Our results suggest that the CT attenuation estimations with the DeHFt pipeline yielded a strong correlation with manual ROI-based measurement similar to the ground truth segmentation. Overall, the results suggest that slice and volumetric-based estimations via DeHFt were highly correlated with human reader assessment. Our results are also in-line with previous findings.[33] However, unlike their approach,[33] DeHFt does not need an expert to manually segment the liver or estimate liver fat.

We additionally utilised the DeHFt model to investigate the prevalence of HS (liver-to-spleen attenuation ratio <1 in CT) in patients with severe COVID-19 infections compared to non-severe COVID-19 infections. We found that the patients with severe COVID-19 had a higher percentage of HS compared with non-severe COVID-19 infections across three different cohorts from three different institutions. These differences were statistically significant for D2, where P < 0.05. The P-values were affected by sample size, effect size and other factors. As the sample size increases (D3, N = 169; D1, N = 805; D2, N = 1917), the P-value decreases (D3, 0.46; D1, 0.12, D3, <0.001) due to reduction in the impact of random error. However, the effect sizes across three datasets were comparable (OR: D1 = 1.46, D2 = 1.55, and D3 = 1.30), with pooled OR 1.50, 95% CI [1.20–1.88] suggesting a higher odd of HS among patients with severe COVID-19 than in non-severe COVID-19 infections. We also observed a few outliers,

as can be seen in Fig. 6. On further investigation, we found that these outliers were either due to the image artifacts or due to the inclusion of a portion of the heart causing incorrect estimation of liver attenuation.

Prior studies[34,35] have shown that obesity and visceral adiposity are linked with COVID-19 infections and poor outcomes. For example, Tahtabasi et al.[36] showed that liver steatosis as assessed by CT (liver attenuation <40 HU) was more frequent in patients with COVID-19 infections compared with controls (40.9% vs. 19.4%, P < .001).[36] Palomar-Lever et al.[35] showed that liver steatosis (liver-to-spleen attenuation ratio ≤ 0.9) was more common in patients with severe COVID-19 infections compared with non-severe infections (69.9 vs 29%). Our study utilises a deep-learning-based pipeline to investigate the association between HS and the severity of COVID-19 infections in a large cohort of patients. Our study further adds to the literature suggesting that liver steatosis assessment using the DeHFt pipeline can be utilised as an automated, reproducible method to assess risk in COVID-19 infections.

This study also has a few limitations that need to be acknowledged. The scans used in training did not include the entire liver as they were intended for CAC screening, and thus it is not clear how our model might perform in scans that include the entirety of the liver/spleen. Secondly, volume-based liver attenuation can be biased by the presence of large cysts, masses, or calcifications that were not present in our training datasets. Another limitation is that the study is retrospective as opposed to prospective. Also the sample size of the validation cohort is limited. Further research is needed to investigate the effect of ethnic characteristics on HS prevalence and its potential role in treatment response.

Conclusions

The deep-learning-based hepatic fat assessment (DeHFt) pipeline enabled accurate segmentation of liver and spleen on non-contrast CT scans and enabled automated estimation of liver and liver-to-spleen attenuation ratio. It has been demonstrated that the approaches could be applied to large epidemiologic studies, including COVID-19 studies, to assess HS and its role in treating and managing patients.

The presented pipeline has several potential applications. First, it can potentially enable automation of HS assessment as part of CAC and other non-contrast CTs and thus can facilitate its reporting to identify high-risk metabolic patients. Second, the automated method can be applied to large datasets to understand the association between liver steatosis and cardiometabolic outcomes. Third, volumetric segmentation allows understanding of radiomic biomarkers that are associated with several diseases (e.g., non-alcoholic fatty liver disease, cirrhosis, viral hepatitis) in addition to metabolic risk markers. Fourth, the new method should be investigated as an aid for comprehensive risk assessment of patients with COVID-19 infection to assign care level (e.g., inpatient vs outpatient) and decision for early use of therapeutics (e.g., antivirals).

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2022.104315.

References
1 National Diabetes Statistics Report. *National Diabetes Statistics Report, 2020.* CDC; 2020:2.
2 Zeb I, Li D, Nasir K, Katz R, Larijani VN, Budoff MJ. Computed tomography scans in the evaluation of fatty liver disease in a population based study. The multi-ethnic study of atherosclerosis. *Acad Radiol [Internet].* 2012;19(7):811–818. Available from: https://doi.org/10.1016/j.acra.2012.02.022.
3 Matsuzaka T, Shimano H. Molecular mechanisms involved in hepatic steatosis and insulin resistance. *J Diabetes Investig.* 2011;2(3):170–175.
4 Berger D, Desai V, Janardhan S. Con: liver biopsy remains the gold standard to evaluate fibrosis in patients with nonalcoholic fatty liver disease. *Clin Liver Dis.* 2019;13(4):114.
5 Khov N, Sharma A, Riley TR. Bedside ultrasound in the diagnosis of nonalcoholic fatty liver disease. *World J Gastroenterol WJG.* 2014;20(22):6821.
6 Ajmera V, Park CC, Caussy C, et al. Magnetic resonance imaging proton density fat fraction associates with progression of fibrosis in patients with nonalcoholic fatty liver disease. *Gastroenterology.* 2018;155(2):307–310.
7 Pickhardt PJ, Graffy PM, Reeder SB, Hernando D, Li K. Quantification of liver fat content with non-contrast MDCT:: phantom and clinical correlation with MRI proton density fat fraction. *AJR Am J Roentgenol.* 2018;211(3):W151.
8 Guo Z, Blake GM, Li K, et al. Liver fat content measurement with quantitative CT validated against MRI proton density fat fraction: a prospective study of 400 healthy volunteers. *Radiology.* 2020;294(1):89–97.
9 Bohte AE, van Werven JR, Bipat S, Stoker J. The diagnostic accuracy of US, CT, MRI and 1 H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: a meta-analysis. *Eur Radiol.* 2011;21(1):87–97.
10 Lin A, Wong ND, Razipour A, et al. Metabolic syndrome, fatty liver, and artificial intelligence-based epicardial adipose tissue measures predict long-term risk of cardiac events: a prospective study. *Cardiovasc Diabetol.* 2021;20(1):1–11.
11 Modanwal G, Walker JR, Al-Kindi S, Rajagopalan S, Madabhushi A. Machine learning-based hepatic fat assessment in low-dose coronary artery calcium scans is correlated with human reader assessment. *Circulation.* 2020;142(Suppl_3):A16796.
12 Zhou L, Wang L, Li W, Liang S, Zhang Q. Automatic segmentation of liver from CT scans with CCP–TSPM algorithm. *Int J Pattern Recogn Artif Intell.* 2019;33(14):1957005.
13 Jin X, Ye H, Li L, Xia Q. Image segmentation of liver CT based on fully convolutional network. In: *2017 10th International Symposium on Computational Intelligence and Design.* ISCID; 2017:210–213.
14 Bilic P, Christ PF, Vorontsov E, et al. The liver tumor segmentation benchmark (lits). *arXiv* Prepr arXiv190104056. 2019. https://doi.org/10.48550/arXiv.1901.04056.
15 Van Ginneken B, Heimann T, Styner M. 3D segmentation in the clinic: a grand challenge. In: *MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge.* 2007:7–15.
16 Deng X, Du G. 3D segmentation in the clinic: a grand challenge II-liver tumor segmentation. In: *MICCAI workshop.* 2008.
17 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):1–9.
18 Lavie CJ, Sanchis-Gomar F, Henry BM, Lippi G. *COVID-19 and obesity: links and risks.* Taylor & Francis; 2020.

19 Sachdeva S, Khandait H, Kopel J, Aloysius MM, Desai R, Goyal H. NAFLD and COVID-19: a pooled analysis. *SN Compr Clin Med*. 2020;2(12):2726–2729.

20 Medeiros AK, Barbisan CC, Cruz IR, et al. Higher frequency of hepatic steatosis at CT among COVID-19-positive patients. *Abdom Radiol*. 2020;45(9):2748–2754.

21 Ziaee A, Azarkar G, Ziaee M. Role of fatty liver in coronavirus disease 2019 patients' disease severity and hospitalization length: a case–control study. *Eur J Med Res*. 2021;26(1):1–8.

22 Mahamid M, Nseir W, Khoury T, et al. Nonalcoholic fatty liver disease is associated with COVID-19 severity independently of metabolic syndrome: a retrospective case-control study. *Eur J Gastroenterol Hepatol*. 2021;33:1578–1581.

23 Revel M-P, Boussouar S, de Margerie-Mellon C, et al. Study of thoracic CT in COVID-19: the STOIC Project. *Radiology*. 2021;301(1):E361–E370.

24 Saltz J, Saltz M, Prasanna P, et al. *Stony Brook University COVID-19 positive cases [data set]*. Cancer Imaging Arch; 2021.

25 Hiremath A, Bera K, Yuan L, et al. Integrated clinical and CT based artificial intelligence nomogram for predicting severity and need for ventilator support in COVID-19 patients: a multi-site study. *IEEE J Biomed Heal Informatics*. 2021;25(11):4110–4118.

26 Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*. 2012;30(9):1323–1341.

27 Chaurasia A, Culurciello E. Linknet: exploiting encoder representations for efficient semantic segmentation. In: *2017 IEEE Visual Communications and Image Processing*. VCIP); 2017:1–4.

28 Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017:2117–2125.

29 Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. 2015:234–241.

30 Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203–211.

31 Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26(3):297–302.

32 Yuan Y. Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation. *CoRR [Internet]*; 2017. abs/1710.0. Available from: http://arxiv.org/abs/1710.04540.

33 Lakshmanan S, Holda M, Joshi T, et al. Measurement of liver fat by A novel 3d segmentation method of liver on non contrast cardiac ct in evaporate cohort: methods and reproducibility. *J Cardiovasc Comput Tomogr*. 2021;15(4):S37–S38.

34 Chen VL, Hawa F, Berinstein JA, et al. Hepatic steatosis is associated with increased disease severity and liver injury in coronavirus disease-19. *Dig Dis Sci*. 2020:1–7.

35 Palomar-Lever A, Barraza G, Galicia-Alba J, et al. Hepatic steatosis as an independent risk factor for severe disease in patients with COVID-19: a computed tomography study. *JGH Open*. 2020;4(6):1102–1107.

36 Tahtabasi M, Hosbul T, Karaman E, et al. Frequency of hepatic steatosis and its association with the pneumonia severity score on chest computed tomography in adult COVID-19 patients. *World J Crit Care Med*. 2021;10(3):47.