Abbreviations

3D = three-dimensional, BSA = body surface area, CC = craniocaudal, CTC = CT colonography

Summary

Fully automated CT-based liver volume segmentation based on deep learning methods provided an objective and more accurate assessment of liver size than linear measures.

Key Results

- In a retrospective analysis of 3065 patients who underwent multidetector CT for colorectal screening (n = 1960, unenhanced) or renal donor evaluation (n = 1105, contrast-enhanced), patient weight was the major determinant of liver volume, allowing a weight-based upper limit of normal threshold for hepatomegaly: mL = 14.0 × (weight [kg]) + 979.
- Liver volumes measured using automated deep learning and manual methods were in close agreement, with a median difference of less than 3%.
- Linear estimates of liver size were inaccurate for determining the amount of liver tissue, as shown by the automated liver volume.

However, the purpose of these studies was to compare automated measurements with manual measurements, rather than characterize the liver size in large healthy populations.

The main purpose of our study was to establish the normal distribution of liver volumes in healthy adults using a fully automated CT-based artificial intelligence quantitative visualization tool and to suggest potential thresholds for hepatomegaly, considering patient-specific factors. We also assessed the performance of linear CT measurements for identifying cases of hepatomegaly according to these volume thresholds.

Materials and Methods

Study Patients

This investigation complied with the rules of the Health Insurance Portability and Accountability Act and was approved by the institutional review board at the University of Wisconsin and the Office of Human Subjects Research Protection at UW Health. The requirement for signed informed consent was waived for this retrospective assessment. The initial study cohort comprised consecutive generally healthy asymptomatic adult outpatients undergoing either unenhanced abdominal CT for colorectal cancer screening (CT colonography [CTC]) or postcontrast abdominal CT for potential renal (kidney) donation at a single medical center between April 2004 and December 2016. Patients who did not have the entire liver scanned at CT were excluded. Patients with missing data were excluded. Only one CT study was included per patient.

Basic demographic and clinical information (age, sex, weight, and height) was collected from the electronic health record. Body surface area (BSA) was calculated using the Mosteller method (16):

$$BSA\left(m^2\right) = \sqrt{\left(height\left[cm\right] \times weight\left[kg\right] / 3600\right)}.$$

CT Protocol

All CT studies were performed with eight- to 64-section multidetector–row scanners (GE Healthcare). The CT acquisition technique used has been previously described in detail (17,18). Briefly, for the study patients undergoing CTC, we used unenhanced supine CT performed with 120 kVp and modulated tube current (typically between 30 mA and 300 mA) to achieve a noise index of 50. For the renal donor group, we used the postcontrast parenchymal phase abdominal CT series, typically consisting of 120 kVp and modulated tube current (between 30 mA and 300 mA), with a noise index of 17–28, both based on patient size. The intravenous contrast agent was iohexol (Omnipaque, GE Healthcare). We also obtained a precontrast series for the renal donor protocol but typically did not include the entire liver. However, a subset of renal donor cases for which whole-liver pre- and postcontrast series were available was used to derive the small correction to convert unenhanced CTC liver volumes to a postcontrast equivalent (see following section). Before automated liver segmentation, we retrospectively reformatted all CT series to 3-mm-thick sections at 3-mm intervals.

Automated Liver Segmentation and Volume Assessment

The detailed description of the automated deep learning segmentation method for organ segmentation can be found elsewhere and includes a modified 3D U-Net and CycleGAN (19–21). The relevant code has been posted to *https://github.com/rsummers11/CADLab/tree/master/CT%20Liver%20Segmentation%20Software*. The deep learning algorithms used in this study have been previously developed, trained, and tested at the National Institutes of Health. No additional training, validation, or machine learning was required for this study, which precludes the need to repeat separate training and testing. This deep learning tool differs from an older tool that used traditional, manually designed image processing methods (5).

To enable processing of large 3D volumes with limited graphics processing unit memory, an initial strided convolution (step size 2) and a complementing final transposed convolution were added. Training data were obtained from the Medical Segmentation Decathlon project (22). Data augmentation was performed using 3D rotation, crop, elastic deformation, CycleGAN noncontrast image, and random flips. Model training and inference were performed on the National Institutes of Health Biowulf System high-performance computing cluster using four central processing unit threads and up to 48 graphics processing unit nodes (NVIDIA K80 or P100; graphics processing unit memory, 12 or 16 GB, respectively). The batch size was four, resolution was 256 × 256 × 192, and initial filters were 32. The initial learning rate was 0.0001, and training was performed for 10 000 iterations. All voxels designated as liver by the segmentation algorithm were analyzed, and the liver volume was computed. A mean Dice score of 0.887 ± 0.006 (standard deviation) was found in a prior validation report (21). To ensure the liver was completely included within the scanned range, at least one section without segmented liver was included, both superior and inferior to the segmented liver. To allow for quality assurance in individual scans, the tool provides a mask series that can be fused with the original imaging

This weight-dependent threshold for liver enlargement is displayed as a solid line in Figure 4.

The CC and maximal 3D linear measurements exhibited only moderate performance for identifying cases of volume-derived hepatomegaly, with many cases of under- or overestimation. A CC cutoff of 19 cm for detecting an enlarged liver (by volume definition) yielded a sensitivity of 71% (49 of 69 patients [95% CI: 60, 82]), specificity of 86% (887 of 1030 patients [95% CI: 84, 88]), positive predictive value of 26% (49 of 192 patients [95% CI: 19, 32]), and negative predictive value of 98% (887 of 907 patients [95% CI: 97, 99]). A maximal 3D line cutoff of 24 cm for detecting an enlarged liver yielded a sensitivity of 78% (54 of 69 patients [95% CI: 69, 88]), specificity of 66% (678 of 1030 patients [95% CI: 63, 69]), positive predictive value of 13% (54 of 406 patients [95% CI: 10, 17]), and negative predictive value of 98% (678 of 693 patients [95% CI: 97, 99]). Figure 1 shows an example in which a normal, more vertically oriented liver measurement was enlarged according to typical linear CC criteria but was within normal volume limits for the patient's weight. Similarly, Figure 2 depicts an enlarged liver based on our derived volume threshold, whereas the linear CC measurement did not indicate enlargement.

For the subset of 189 patients (mean age, 46 years ± 12; 77 men, 112 women) for whom we also obtained liver volume assessment with use of the semiautomated or manual software tool, the correlation between the liver volume and the deep learning tool was excellent (r^2 = 0.98; $P < .01$). The median difference in liver volume between the deep learning tool and the semiautomated or manual method was 2.3% (interquartile range, 0.7%–3.7%) (38 mL [interquartile range, 12–68 mL]). Figure 5 shows the overlap of volume distribution with each method.

Discussion

The deep learning CT tool used in this study segmented the liver for every patient evaluated, including both unenhanced and enhanced series. Our adult outpatient sample was generally healthy and asymptomatic and underwent CT imaging for either colorectal cancer screening or potential renal donation. The average liver volume ± standard deviation measured was 1533 mL ± 375, which correlated with patient weight ($P < .001$). In general, our results of liver volume assessment match well with those of smaller (n = 11–351) prior healthy study samples in investigations that used a variety of manual and semiautomated methods in terms of average values: 1493 mL (3), 1323 mL (8), 1510 mL (4), 1580 mL (7), 1419 mL (1), 1520 mL (9), 1510 mL (5), and 1671 mL (11). To establish the normal value, we decided to start with asymptomatic outpatients, but our threshold for hepatomegaly will need to be tested in symptomatic patients with known liver diseases expected to result in hepatomegaly. From the data obtained in regression analyses, we found that weight, rather than body surface area (BSA), was the major predictor of liver volume, and that sex was not an independent predictor. We also found that liver volume increased linearly with weight, which simplified the output. Furthermore, to establish the upper limit of normal, we found that elevated outliers beyond two standard deviations from the mean also increased linearly with weight. Thus, we decided that a simple linear threshold based on weight represents a good balance for characterizing true outliers (hepatomegaly). Previous research has used BSA (5), which requires additional computation and may not have the same linear relationships we found using patient weight.

Hepatomegaly is a challenging clinical and imaging diagnosis, partly because of the lack of an accepted in vivo radiologic definition. Evaluation is further complicated by acute, subacute, and chronic clinical presentations. Relevant acute causes include inflammation or hepatitis, which may be due to steatohepatitis from alcoholic-related liver disease and nonalcoholic fatty liver disease—or its more accurate name, metabolic dysfunction–associated fatty liver disease (24)—as well as viral, drug-induced, vascular, and autoimmune causes. More subacute and chronic

Table 2: Multivariable Linear Regression Analysis for Predicting Liver Volume

Variable	Standardized β Coefficient	*P* Value
Weight	0.62	<.001
Age	−0.27	<.001
Height	0.05	.002
Sex	Excluded	.46

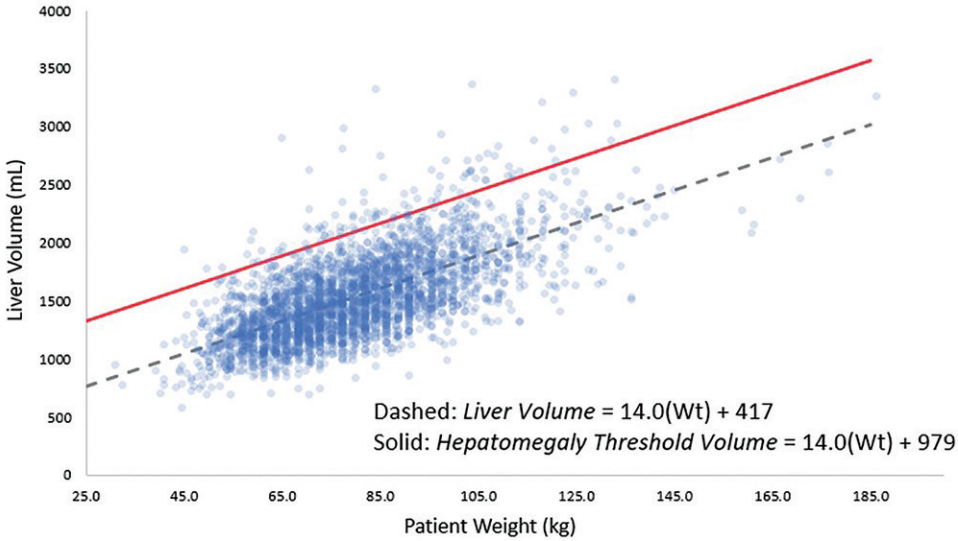

Figure 4: Graph shows automated CT-based liver volume according to patient weight (Wt). Weight was the dominant patient factor affecting liver volume. The solid red line represents the derived weight-based threshold for hepatomegaly based on two standard deviations above the mean (dashed line).

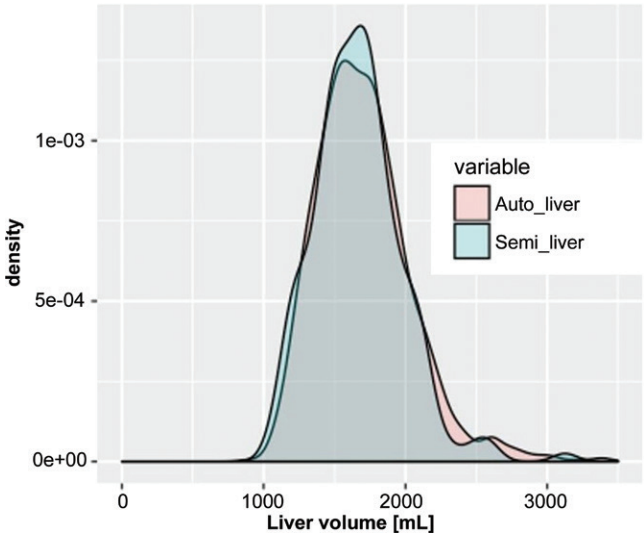

Figure 5: The density plot of the subanalysis comparing automated (Auto) liver volume (orange) and manual or semiautomated (Semi) liver volume (green) show good overlap in the distributions. The median volume difference was less than 3%.

causes include neoplastic disease (metastatic or primary tumor) as well as several metabolic, infiltrative, and congenital conditions. The correlation with liver enzymes can help distinguish hepatocellular from cholestatic causes, but these are generally nonspecific and often require additional clinical correlation and testing. Regardless of the cause, linear measurements at US, CT, or MRI provide only a crude assessment of liver size, reflected by the wide array of suggested measurements and thresholds, whereas volume estimation provides for the most logical and accurate assessment of the amount of liver tissue. Owing to technical limitations, volume assessment in cross-sectional imaging is largely limited to the research realm. However, with artificial intelligence deep learning approaches, organ volume assessment is now a feasible routine measurement that can be applied to other abdominal organs, such as the spleen.

Beyond the objective assessment of liver size by means of determination of volume, both CT and MRI can provide noninvasive assessment of liver fat content and liver fibrosis. Attenuation values obtained using unenhanced CT have a linear relationship with MRI-based proton density fat fraction (25,26), which is the current clinical standard. Automated mean whole-liver attenuation can be combined with volume assessment with use of artificial intelligence approaches (27,28). MRI elastography is a well-established tool to estimate liver fibrosis (29). However, several objective CT-based measurements can also be used to accurately predict the presence and degree of liver fibrosis, including liver surface nodularity, segmental volume changes, and parenchymal textural analysis (23,30–32). However, decreased total liver volume is actually a poor predictor of the liver fibrosis stage, primarily because of compensatory changes of Couinaud segments I–III relative to segments IV–VIII. This segmental redistribution, however, can be captured by the liver segmental volume ratio (11,23). We are currently working on a deep learning prototype for assessing segmental liver volume changes. In terms of hepatomegaly, our derived volume threshold now needs to be tested in mixed symptomatic cohorts to assess its utility.

Among the patients who underwent renal donor evaluation and had both pre- and postcontrast studies that included the entire liver, the postcontrast liver volume estimations were slightly but systematically larger than those in the precontrast equivalents, which could be described with a simple linear correlation. This minor (3.6%) discrepancy might be attributed in part to an actual physiologic increase as well as "pseudo-enlargement" related to the contrast enhancement effect that slightly expands the edges of liver segmentation. Because most abdominal CT scans are contrast enhanced, we decided to standardize all unenhanced volumes to contrast-enhanced equivalents.

We acknowledge some limitations to our study. First, based on our findings that patient weight was the major determinant, we chose to define hepatomegaly in terms of otherwise healthy outliers that were two standard deviations above the mean for a specific patient weight. This simplistic approach avoided the need for complex multivariate thresholding or nomograms, potentially allowing for easier clinical use if validated in symptomatic cohorts. Other researchers have used different approaches, including one– and 2.5–standard deviation thresholds using BSA-normalized livers (5). In future validation studies that include diseased populations, other volume thresholds may be explored. Additionally, more diverse populations from other centers and regions can be included for more generalizable results. Second, the lack of another reference standard definition of hepatomegaly beyond volume estimation may make further validation somewhat challenging. As we have demonstrated, linear measurements, such as CC assessment, cannot accurately predict liver size (ie, volume).

In summary, using a deep learning tool to measure liver volume, we derived a simple weight-based threshold for hepatomegaly that holds several advantages over the standard linear assessment used in routine clinical practice. Further validation of our simplified hepatomegaly approach is required, especially in terms of its application to more diverse symptomatic cohorts with diseased livers. Future studies on a deep learning prototype could provide additional value in assessing segmental liver volume changes. If this current approach is further validated in larger healthy and patient cohorts, then it could provide for an easier and more objective measurement of liver size with value in screening for patients with an abdominal CT scan or for diagnostic purposes.

Acknowledgment: We thank Oliver F. Hunt, MD, for his work in collecting semiautomated liver volumes for subanalysis comparison with automated volume tools.

Author contributions: Guarantor of integrity of entire study, A.A.P.; 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, A.A.P., V.N.K., M.G.L., P.M.G., P.J.P.; clinical studies, A.A.P., V.N.K., M.G.L., P.J.P.; statistical analysis, A.A.P., V.N.K., P.M.G.; and manuscript editing, A.A.P., M.G.L., P.M.G., J.W.G., D.C.E., P.J.P.

Disclosures of Conflicts of Interest: A.A.P. No relevant relationships. **V.N.K.** No relevant relationships. **M.G.L.** Prior grant funding from Philips and Ethicon; honorarium for a lecture from the International Society for Computed Tomography.

P.M.G. No relevant relationships. **J.W.G.** No relevant relationships. **D.C.E.** No relevant relationships. **R.M.S.** Cooperative Research and Development Agreement from PingAn; royalties or licenses for patents, software, or both from PingAn, iCAD, Philips, ScanMed, and Translation Holdings; graphics processing unit card donations from NVIDIA. **P.J.P.** Consulting fees from Zebra Medical Systems, GE Healthcare, and Bracco.

References

1. Andersen V, Sonne J, Sletting S, Prip A. The volume of the liver in patients correlates to body weight and alcohol consumption. Alcohol Alcohol 2000;35(5):531–532.
2. Feng LM, Wang PQ, Yu H, et al. New formula for predicting standard liver volume in Chinese adults. World J Gastroenterol 2017;23(27):4968–4977.
3. Henderson JM, Heymsfield SB, Horowitz J, Kutner MH. Measurement of liver and spleen volume by computed tomography. Assessment of reproducibility and changes found following a selective distal splenorenal shunt. Radiology 1981;141(2):525–527.
4. Kwo PY, Ramchandani VA, O'Connor S, et al. Gender differences in alcohol metabolism: relationship to liver volume and effect of adjusting for body mass. Gastroenterology 1998;115(6):1552–1557.
5. Linguraru MG, Sandberg JK, Jones EC, Petrick N, Summers RM. Assessing hepatomegaly: automated volumetric analysis of the liver. Acad Radiol 2012;19(5):588–598.
6. Poovathumkadavil A, Leung KF, Al Ghamdi HM, Othman IeH, Meshikhes AW. Standard formula for liver volume in Middle Eastern Arabic adults. Transplant Proc 2010;42(9):3600–3605.
7. Sandrasegaran K, Kwo PW, DiGirolamo D, Stockberger SM Jr, Cummings OW, Kopecky KK. Measurement of liver volume using spiral CT and the curved line and cubic spline algorithms: reproducibility and interobserver variation. Abdom Imaging 1999;24(1):61–65.
8. Stapakis J, Stamm E, Townsend R, Thickman D. Liver volume assessment by conventional vs. helical CT. Abdom Imaging 1995;20(3):209–210.
9. Suzuki K, Epstein ML, Kohlbrenner R, et al. Quantitative radiology: automated CT liver volumetry compared with interactive volumetry and manual volumetry. AJR Am J Roentgenol 2011;197(4):W706–W712.
10. Um EH, Hwang S, Song GW, et al. Calculation of standard liver volume in Korean adults with analysis of confounding variables. Korean J Hepatobiliary Pancreat Surg 2015;19(4):133–138.
11. 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.
12. Heymsfield SB, Fulenwider T, Nordlinger B, Barlow R, Sones P, Kutner M. Accurate measurement of liver, kidney, and spleen volume and mass by computerized axial tomography. Ann Intern Med 1979;90(2):185–187.
13. Cai W, He B, Fan Y, Fang C, Jia F. Comparison of liver volumetry on contrast-enhanced CT images: one semiautomatic and two automatic approaches. J Appl Clin Med Phys 2016;17(6):118–127.
14. Nakayama Y, Li Q, Katsuragawa S, et al. Automated hepatic volumetry for living related liver transplantation at multisection CT. Radiology 2006;240(3):743–748.
15. Wang K, Mamidipalli A, Retson T, et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. Radiol Artif Intell 2019;1(2):180022.
16. Mosteller RD. Simplified calculation of body-surface area. N Engl J Med 1987;317(17):1098.
17. Pickhardt PJ, Graffy PM, Zea R, et al. Automated abdominal CT imaging biomarkers for opportunistic prediction of future major osteoporotic fractures in asymptomatic adults. Radiology 2020;297(1):64–72.
18. Perez AA, Pickhardt PJ, Elton DC, Sandfort V, Summers RM. Fully automated CT imaging biomarkers of bone, muscle, and fat: correcting for the effect of intravenous contrast. Abdom Radiol (NY) 2021;46(3):1229–1235.
19. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. Medical Image Computing and Computer-Assisted Intervention 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. MICCAI 2016. Cham, Switzerland: Springer, 2016; 424–432.
20. Kayalibay B, Jensen G, van der Smagt P. CNN-based segmentation of medical imaging data. arXiv preprint arXiv:1701.03056. https://arxiv.org/abs/1701.03056. Posted January 11, 2017. Accessed February 25, 2021.
21. 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.
22. Medical Segmentation Decathlon. Generalisable 3D Semantic Segmentation. https://arxiv.org/abs/2106.05735. Accessed February 25, 2021.
23. Pickhardt PJ, Malecki K, Hunt OF, et al. Hepatosplenic volumetric assessment at MDCT for staging liver fibrosis. Eur Radiol 2017;27(7):3060–3068.
24. Fouad Y, Waked I, Bollipo S, Gomaa A, Ajlouni Y, Attia D. What's in a name? Renaming 'NAFLD' to 'MAFLD'. Liver Int 2020;40(6):1254–1261.
25. 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.
26. Pickhardt PJ, Graffy PM, Reeder SB, Hernando D, Li K. Quantification of liver fat content with unenhanced MDCT: phantom and clinical correlation with MRI proton density fat fraction. AJR Am J Roentgenol 2018;211(3):W151–W157.
27. Graffy PM, Sandfort V, Summers RM, Pickhardt PJ. Automated liver fat quantification at nonenhanced abdominal CT for population-based steatosis assessment. Radiology 2019;293(2):334–342.
28. Pickhardt PJ, Blake GM, Graffy PM, et al. Liver steatosis categorization on contrast-enhanced CT using a fully automated deep learning volumetric segmentation tool: evaluation in 1204 healthy adults using unenhanced CT as a reference standard. AJR Am J Roentgenol 2021;217(2):359–367.
29. Tang A, Cloutier G, Szeverenyi NM, Sirlin CB. Ultrasound elastography and MR elastography for assessing liver fibrosis: part 2, diagnostic performance, confounders, and future directions. AJR Am J Roentgenol 2015;205(1):33–40.
30. Lubner MG, Jones D, Kloke J, Said A, Pickhardt PJ. CT texture analysis of the liver for assessing hepatic fibrosis in patients with hepatitis C virus. Br J Radiol 2019;92(1093):20180153.
31. 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.
32. Roberts AS, Shetty AS, Mellnick VM, Pickhardt PJ, Bhalla S, Menias CO. Extramedullary haematopoiesis: radiological imaging features. Clin Radiol 2016;71(9):807–814.