Performance of an automated deep learning algorithm to identify hepatic steatosis within noncontrast computed tomography scans among people with and without HIV

Jessie Torgersen, MD, MHS, MSCE[1,2,3], Scott Akers, MD[3], Yuankai Huo, PhD[4], James G. Terry[5], J. Jeffrey Carr, MD, MSc[5], Alexander T. Ruutiainen, MD[3], Melissa Skanderson[6,7], Woody Levin[6,7], Joseph K. Lim, MD[6,7], Tamar H. Taddei, MD[6,7], Kaku So-Armah, PhD[8], Debika Bhattacharya, MD[9], Christopher T. Rentsch, PhD, MPH[6,7,10], Li Shen, PhD[2], Rotonya Carr, MD[11], Russell T. Shinohara, PhD[2,12,13], Michele McClain[14], Matthew Freiberg, MD, MSc[15], Amy C. Justice, MD, PhD[6,7,16], Vincent Lo Re III, MD, MSCE[1,2]

Affiliations:

1. Department of Medicine, Penn Center for AIDS Research, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
2. Department of Biostatistics, Epidemiology, and Informatics, Center for Clinical Epidemiology and Biostatistics, Center for Real World Effectiveness and Safety of Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
3. Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA
4. Department of Computer Science, Vanderbilt University, Nashville, TN, USA
5. Department of Radiology and Radiological Sciences, Vanderbilt University School of Medicine, Nashville, TN, USA
6. Department of Medicine, Yale School of Medicine, New Haven, CT, USA
7. VA Connecticut Healthcare System, West Haven, CT, USA

8. Department of Medicine, Boston University School of Medicine, Boston, MA, USA
9. VA Greater Los Angeles Healthcare System and David Geffen School of Medicine at UCLA, Los Angeles, CA, USA
10. Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK
11. Department of Medicine, Division of Gastroenterology, University of Washington, Seattle, WA, USA
12. Center for Biomedical Image Computing and Analysis (CBICA), Department of Radiology, University of Pennsylvania, Philadelphia, PA, 19104
13. Penn Statistics in Imaging and Visualization Endeavor (PennSIVE), Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA, 19104
14. VA Office of Information and Technology, Frederick, MD, USA
15. Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA
16. Division of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA

Corresponding Author: Jessie Torgersen, MD, MHS, MSCE
Penn Presbyterian Medical Center
51 N. 39th Street, MAB Ste, 103A
Philadelphia, PA 19104
E-mail: jessie.torgersen@pennmedicine.upenn.edu
Tel: 215-662-8344; Fax: 215-243-4658

Running Title: ALARM Identification of Hepatic Steatosis by HIV Status

Funding Statement: This study was funded by a pilot grant from the Penn Center for AIDS Research (P30AI045008; PI: J.T.), the National Institute of Diabetes and Digestive and Kidney Diseases (K08DK132977; PI: J.T.), the National Cancer Institute (R01CA206465; PI: A.C.J and V.L.R.), and the National Institutes on Alcohol Abuse and Alcoholism (R01AA026302 [PI: R.C.] and U01AA013566, U24AA020794, and U01AA020790 [PI: A.C.J]). This work uses data provided by patients and collected by the VA as part of their care and support.

Conflict of Interest: YH has received research support from IBM and Merative Co. DB has received research grant support, paid to the institution, from Gilead Sciences. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs. The funding sources had no role in the conduct of the research and/or preparation of the manuscript.

Prior Presentation: Part of this work has been accepted for presentation as an abstract at The 30th Conference on Retroviruses and Opportunistic Infections, February 19-22, 2023.

Data Availability: Due to US Department of Veterans Affairs (VA) regulations and our ethics agreements, the analytic data sets used for this study are not permitted to leave the VA firewall without a Data Use Agreement. This limitation is consistent with other studies based on VA data. However, VA data are made freely available to researchers

with an approved VA study protocol. For more information, please visit https://www.virec.research.va.gov or contact the VA Information Resource Center (VIReC) at VIReC@va.gov.

Plain Language Summary

Hepatic steatosis (fatty liver disease) is very common, particularly in people living with HIV. Yet studies evaluating medications associated with developing hepatic steatosis are limited due to lack of tools to identify hepatic steatosis within clinical images. We compared the performance of the Automatic Liver Attenuation Region-of-Interest-based Measurement (ALARM) program to identify hepatic steatosis within computed tomography images to manual radiologist review. ALARM demonstrated excellent accuracy for identifying moderate-to-severe hepatic steatosis among people with and without HIV. By validating ALARM's ability to accurately identify hepatic steatosis, this tool can be applied to clinical images within electronic medical record databases, allowing for large studies to identify medications and other factors associated with hepatic steatosis and assess differences by HIV status.

Key Points

- Hepatic steatosis (fatty liver disease), defined by liver triglyceride content >5%, affects 25% of the adult population globally, particularly people living with HIV.
- Pharmacoepidemiologic studies to examine the medications associated with hepatic steatosis have not been conducted since liver biopsy has traditionally been used for diagnosis and because methods to evaluate the presence of liver fat within digitized images stored in electronic medical record databases have not been developed or validated.
- We determined the performance characteristics of a deep learning algorithm called the Automatic Liver Attenuation Region-of-Interest-based Measurement (ALARM) program to identify hepatic steatosis within clinically-obtained noncontrast abdominal CT images compared to manual radiologist review and evaluated its performance among people with and without HIV infection.
- Sensitivity, specificity, positive predictive value, and negative predictive value of ALARM compared to manual radiologist review were 91.7% (95%CI, 51.5-99.8%), 96.3% (95%CI, 90.8-99.0%), 73.3% (95%CI, 44.9-92.2%), and 99.0% (95%CI, 94.8-100%), respectively. No differences in performance were observed by HIV status.
- Application of ALARM to radiographic repositories could facilitate real-world studies to evaluate medications associated with steatosis and assess differences by HIV status.

Abstract

Purpose: Hepatic steatosis (fatty liver disease) affects 25% of the world's population, particularly people with HIV (PWH). Pharmacoepidemiologic studies to identify medications associated with steatosis have not been conducted because methods to evaluate liver fat within digitized images have not been developed. We determined the accuracy of a deep learning algorithm (Automatic Liver Attenuation Region-of-Interest-based Measurement [ALARM]) to identify steatosis within clinically-obtained noncontrast abdominal CT images compared to manual radiologist review and evaluated its performance by HIV status.

Methods: We performed a cross-sectional study to evaluate the performance of ALARM within noncontrast abdominal CT images from a sample of patients with and without HIV in the US Veterans Health Administration. We evaluated the ability of ALARM to identify moderate-to-severe hepatic steatosis, defined by mean absolute liver attenuation <40 Hounsfield units (HU), compared to manual radiologist assessment.

Results: Among 120 patients (51 PWH) who underwent noncontrast abdominal CT, moderate-to-severe hepatic steatosis was identified in 15 (12.5%) persons via ALARM and 12 (10%) by radiologist assessment. Percent agreement between ALARM and radiologist assessment of absolute liver attenuation <40 HU was 95.8%. Sensitivity, specificity, positive predictive value, and negative predictive value of ALARM were 91.7% (95%CI, 51.5-99.8%), 96.3% (95%CI, 90.8-99.0%), 73.3% (95%CI, 44.9-92.2%), and 99.0% (95%CI, 94.8-100%), respectively. No differences in performance were observed by HIV status.

Conclusions: ALARM demonstrated excellent accuracy for moderate-to-severe hepatic steatosis regardless of HIV status. Application of ALARM to radiographic repositories could facilitate real-world studies to evaluate medications associated with steatosis and assess differences by HIV status.

Keywords: Hepatic steatosis, fatty liver disease, machine learning, validation

1. Introduction

Hepatic steatosis, also referred to as fatty liver disease, is defined by liver triglyceride content >5%.[1] This condition is highly prevalent, affecting 25% of the adult population globally,[2] and is a leading indication for liver transplantation.[3] People with HIV (PWH) are at particular risk for hepatic steatosis due to use of obesogenic antiretroviral drugs, concomitant alcohol/substance use, and because infection promotes liver fat deposition by enhancing translocation of gastrointestinal bacteria.[4-6] Moreover, hepatic steatosis can lead to liver inflammation, liver fibrosis, and liver complications such as decompensated cirrhosis and hepatocellular carcinoma.[7]

Despite the potential impact of hepatic steatosis, pharmacoepidemiologic studies to examine the medications associated with its development have not been conducted, largely because liver biopsy has traditionally been used for diagnosis. Radiographic methods to identify hepatic steatosis, particular non-contrast computed tomography (CT), have now supplanted use of liver biopsy and are routinely employed in clinical settings. However, automated methods to identify hepatic steatosis within digitized noncontrast CT scan images stored in radiographic repositories of electronic medical record databases have not been developed or validated, precluding large-scale analyses of the factors associated with steatosis and how they might differ by HIV status.

Recent advances in artificial intelligence have allowed for the development of computer-aided diagnostics in which deep learning algorithms can standardize measurements of CT images of the liver, offering the potential to characterize hepatic steatosis within clinically-obtained noncontrast abdominal CT scans in large samples of patients. The Automatic Liver Attenuation Region-of-Interest-based Measurement

(ALARM) program is a recently developed, open-source, deep learning tool that offers this potential. However, before ALARM can be used to facilitate the study of the medications associated with development of hepatic steatosis, the ability of this algorithm to identify liver fat validly within digital liver images must be determined. Moreover, given the important contribution of HIV infection to steatosis, evaluating the accuracy of ALARM by HIV status is important. To address this issue, we determined the performance characteristics of ALARM to identify hepatic steatosis among PWH and people without HIV (PWOH) within clinically-obtained noncontrast abdominal CT images compared to manual radiologist review.

2. Materials and Methods

2.1. Study Design and Data Source

We conducted a cross-sectional study among a sample of PWH and PWOH in the Veterans Health Administration (VA) at the Corporal Michael J. Crescenz Philadelphia VA Medical Center (Crescenz VAMC) who were included in the Veterans Aging Cohort Study (VACS) between January 1, 2010, and September 30, 2017. The VACS collects electronic medical record data from PWH receiving care at >1,200 VA medical facilities across the US. At each VA center, PWH are matched on age, sex, and race/ethnicity to two PWOH. The VACS includes >40,000 PWH and >80,000 PWOH across the VA system with available electronic health record data, including digital images, from as early as 1997.[8] Data collected by the VACS include hospital and outpatient diagnoses (recorded using International Classification of Diseases, Ninth and Tenth Revision codes), procedures (recorded using Current Procedural Terminology

[CPT] codes), laboratory results, dispensed medications, and radiographic data. Digital Imaging and Communications in Medicine (DICOM) image files are stored in the local Picture Archiving and Communication Systems (PACS) and are available from the VA's national imaging repository system (VistA, the Veterans Health Information System Technology Architecture), one of the world's largest clinical imaging repositories.

2.2. Study Patients

Patients were included if they: 1) underwent a noncontrast abdominal CT scan as part of clinical care between January 1, 2010, and September 30, 2017, at the Crescenz VAMC in Philadelphia, Pennsylvania, 2) had CT imaging files stored on the PACS, and 3) were ≥18 years of age at the time of the CT scan. If patients had multiple noncontrast abdominal CT scans performed, only the first was selected for analysis. All eligible patients were selected for inclusion in the analysis.

2.3. Main Study Outcome

The primary outcome was moderate-to-severe hepatic steatosis (i.e., ≥30% triglyceride content within the liver), defined by absolute liver attenuation <40 Hounsfield units (HU) on noncontrast abdominal CT scan. Liver attenuation <40 HU on noncontrast abdominal CT is an established threshold identifying moderate-to-severe hepatic steatosis and has been validated for the diagnosis of hepatic steatosis compared to liver biopsy, with a sensitivity of up to 81.7% and specificity of up to 97.7%.[9-11] Moreover, this threshold of steatosis has also been associated with increased morbidity and mortality

in PWOH.[12-14] There is a linear, inverse association between decreasing absolute liver attenuation and increasing severity of steatosis.[9]

The reference standard assessment was defined by manual measurement of absolute liver attenuation by a board-certified radiologist blinded to the results of ALARM. The summary measurement of absolute liver attenuation was derived from the mean of three periphery regions of interest (ROI) within the liver. These ROIs were measured at the level of portal vein entry into the liver and comprised approximately 100 mm^2 in area, while taking care to exclude regions of non-uniform parenchyma attenuation, including hepatic vessels.

The fully automated liver attenuation assessment generated by ALARM analysis similarly included the mean liver attenuation of three peripheral ROIs, as previously described.[15] Briefly, ALARM consists of several different imaging processing algorithms, including image preprocessing, deep learning-based liver segmentation, ROI extraction, and visualization, as previously described.[15] This segmentation process employs deep convolutional neural networks to segment the liver into discrete fields for analysis. The subsequent morphological operations generate ROIs for the center of the liver (reflecting anatomy in which vascular structures are located) and the three locations in the liver periphery (reflecting more homogeneous hepatic parenchyma with representative hepatic fat content). Three periphery ROIs are obtained to allow for a robust estimation of absolute liver attenuation. These three periphery ROIs include the posterior, lateral, and anterior locations relative to the center, and are identified at the points two-thirds of the radius from the center to the boundary of the liver segmentation. The ALARM output readings are generated within five minutes, including quantitative

values of mean absolute liver attenuation from three periphery ROIs and waist circumference.

2.4. Data Collection

Clinically obtained noncontrast CT scans of the abdomen were identified through query of CPT code 74150 within the VACS. Eligible CT scans were downloaded from the PACS system at the Crescenz VAMC as de-identified DICOM files. The de-identified files were transferred to Vanderbilt University via encrypted USB, where the ALARM program was deployed by the developers (Y.H., J.J.C., J.G.T.) to analyze the liver images within each DICOM file. All imaging data were reviewed in a semi-automated quality assurance step, including signal-to-noise analysis, imaging artifacts, protocol validation (ensuring noncontrast CT), data integrity and special distortions.[15,16]

Demographic and clinical data were collected from VA electronic health records in the VACS within six months prior to the noncontrast abdominal CT scan and included: age at scan, sex, race/ethnicity, body mass index, HIV status,[17] diabetes mellitus (defined by random glucose ≥200 mg/dL, hemoglobin A1c ≥6.5%, or anti-diabetic drug use),[18] hypertension (defined as systolic blood pressure ≥140 mmHg, diastolic blood pressure ≥90 mm Hg, or antihypertensive drug use),[19] and previously validated diagnoses of alcohol dependence/abuse.[20] Among PWH, use of “obesogenic” antiretroviral therapy (ART; i.e., medications associated with ≥10% increase in weight) was abstracted.[21] Use of obesogenic ART, including tenofovir alafenamide and integrase inhibitors (i.e., dolutegravir, elvitegravir, raltegravir), was defined by a prescription dispensed for at least one of these medications at the time of the CT scan. When

available, the free text clinical indication for CT scan was recorded.

Laboratory results collected included: alanine aminotransferase, aspartate aminotransferase, serum albumin, platelet count, total cholesterol, low-density lipoprotein, high-density lipoprotein, and triglyceride level. If more than one laboratory result was recorded within the six months prior to the CT scan, we used the result closest, but prior, to the scan. We determined hepatitis B virus (HBV) status (ever positive HBV surface antigen) and hepatitis C virus (HCV) status (ever detectable HCV antibody, RNA, or genotype). Advanced hepatic fibrosis/cirrhosis was defined by Fibrosis-4 Index for Hepatic Fibrosis (FIB-4) >3.25.[22] This FIB-4 cut-off identifies advanced fibrosis/cirrhosis with an area under the receiver operating characteristic curve (AUROC) of 0.81 [23] for PWH and 0.80 for PWOH.[22,24]

2.5. Statistical Analysis

Differences in characteristics by HIV status were assessed by Chi-square tests for categorical data and Wilcoxon rank-sum for continuous data. We evaluated the correlation between absolute liver attenuation measured by the ALARM program and manual radiologist review with Pearson's correlation coefficient. Bland Altman plots were created to define the mean difference and limits of agreement between ALARM and manual radiologist measurements. We evaluated the ALARM program's sensitivity (i.e., ability of ALARM to correctly identify those who have moderate-to-severe hepatic steatosis), specificity (i.e., ability of ALARM to correctly identify those who do not have moderate-to-severe steatosis), positive predictive value (i.e., proportion with steatosis by ALARM confirmed to have the condition by radiologist review), and negative

predictive value (i.e., proportion without steatosis by ALARM who do not have the condition by radiologist review) and calculated the AUROC curve at the moderate-to-severe hepatic steatosis threshold of 40 HU.[25] We additionally performed sensitivity analyses evaluating the performance characteristics of ALARM at higher absolute liver attenuation thresholds of 48 HU and 51 HU. The threshold of 48 HU has previously been utilized to maximize specificity for the identification of moderate-to-severe hepatic steatosis, while the threshold of 51 HU has been utilized to identify mild hepatic steatosis (i.e., 6-29% triglyceride content in the liver).[10,26] We calculated exact binomial 95% confidence intervals for various estimates of performance characteristics of ALARM; based on the inclusion of 51 PWH and 69 PWOH, we anticipated sufficient precision (+/-16% around point estimate) to determine the performance characteristics of ALARM, particularly for the higher (>90%) estimates. Statistical analyses were performed with STATA 14.1 (Stata Corporation; College Station, TX).

3. Results

3.1. Patient Characteristics

We identified 120 patients (PWH=51 [42.5%]) within the VACS at the Crescenz VAMC who underwent a noncontrast abdominal CT scan between January 1, 2010, and September 30, 2017. The most common indication for the scan was abdominal pain; indications did not differ significantly by HIV status (p=0.57) (**Table 1**). In the overall sample, patients were predominantly black, male, and had a high prevalence of chronic HCV, alcohol abuse/dependence, and metabolic comorbidities, including diabetes, dyslipidemia, and hypertension (**Table 1**). PWH were more likely to have chronic HBV

or HCV, lower cholesterol and low-density lipoprotein, and FIB-4 >3.25, and less likely to have obesity, diabetes, or hypertension, compared to PWOH. Among the 51 PWH, 43 (84.3%) received antiretroviral therapy; 30 (58.8%) were dispensed obesogenic ART as part of their regimen. Among 40 PWH who had available HIV RNA and CD4 cell count, 32 (80.0%) had HIV RNA <200 copies/mL and 6 (15.0%) had CD4 cell percentage <14%.

3.2. Absolute Liver Attenuation

Mean absolute liver attenuation in the overall sample was 48.9 HU by ALARM and 51.8 HU by manual radiology assessment. Measurements of liver attenuation by ALARM and manual review were highly correlated (Pearson's correlation coefficient, 0.93; p<0.001) (**Figure 1**). When compared to ALARM assessment, the mean difference in absolute liver attenuation as measured by the radiologist was 1.54 HU higher, a clinically insignificant difference. The Bland Altman 95% limits of agreement between these measurements were -5.80 to 8.89 HU, similar to the magnitude of attenuation variability that have been observed on repeated measurements of liver attenuation (**Figure 2**).[27]

Overall, moderate-to-severe hepatic steatosis, as defined by mean absolute liver attenuation <40 HU, was confirmed in 15 (12.5%) people by ALARM and 12 (10.0%) by manual radiologist assessment (**Table 2**). Percent agreement between the ALARM and radiologist assessment of mean absolute liver attenuation using a threshold of 40 HU was 95.8% and did not differ by HIV status (96.1% and 95.6% among PWH and PWOH, respectively). At the <40 HU threshold, ALARM achieved a sensitivity of 91.7% (95% CI:

51.5-99.8%), specificity of 96.3% (95% CI: 90.8-99.0%), and negative predictive value of 99.0% (95% CI: 94.8-100%) (**Table 3**). Positive predictive value of ALARM was lower at 73.3% (95% CI: 44.9-92.2%); the AUROC was 0.94 (95% CI: 0.86-1.00). The performance characteristics of ALARM did not differ by HIV status (**Tables S1 and S2**) or with absolute liver attenuation thresholds of 48 HU or 51 HU (**Table S3 and S4**).

Five (4.2%) of the 120 noncontrast abdominal CT scans did not achieve agreement between ALARM and radiologist absolute liver attenuation assessment at the threshold of 40 HU. Among these five, the absolute liver attenuation measurements were within 5 HU from this *a priori* threshold (range: 36.6 to 44.3 HU). The difference in mean absolute liver attenuation measurements ranged between -2.8 to 4.7 HU (**Table 4**).

4. Discussion

This study is the first to apply ALARM, a fully automated, open-source, deep learning algorithm, for the assessment of moderate-to-severe hepatic steatosis in people with and without HIV infection. We found that ALARM demonstrated high correlation with manual radiology assessment for hepatic steatosis when applied to images obtained in clinical care. Moreover, ALARM demonstrated excellent sensitivity, specificity, and agreement compared to manual radiologist classification, and results did not differ by HIV status. These results suggest that the fully automated ALARM program can be applied to large repositories of clinically-obtained CT images for accurate assessment of moderate-to-severe hepatic steatosis in real-world cohorts of people with

and without HIV, which could facilitate evaluation of the medications associated with steatosis in these groups.

Fully automated assessment of liver attenuation offers the potential to extract accurate objective, quantitative data from CT images for observational investigations if such tools are externally validated.[28] While the accuracy of ALARM was not expected to be impacted by the presence of HIV due to objective value of liver attenuation, confirming the validity of this novel tool was considered necessary when applying to clinically-obtained DICOM files. We found that ALARM produced results that correlated well with manual measurement of absolute liver attenuation and accurately identified moderate-to-severe steatosis with 95.8% agreement. The high level of agreement is on par with prior validation work among PWOH using ALARM,[15] highlighting the robust accuracy across different populations and CT scanners. Moreover, the performance of ALARM for the assessment of liver attenuation is similar to other fully automated CT-based tools. A recent study by Graffy et al., evaluating hepatic steatosis within 5,265 CT scans, showed that their fully automated algorithm achieved 97.9% agreement with manual assessment for categorization of moderate-to-severe hepatic steatosis.[29] However, unique to the open-source nature, ALARM offers the potential for independent investigators to reliably identify hepatic steatosis in diverse patient populations, further demonstrating external validity.

ALARM identified moderate-to-severe steatosis in 15.7% of PWH and 10.1% of PWOH, similar to prevalence estimates determined by noncontrast abdominal CT in other populations with and without HIV.[30] While the present study was not powered to evaluate determinants of moderate-to-severe hepatic steatosis by HIV status, it further

highlights the critical need for investigations employing large-scale image analysis to overcome limitations of small sample sizes and limited generalizability in the majority of studies of steatosis to date.[31] Application of ALARM in real-world observational cohorts can allow for rapid, objective, and accurate identification of hepatic steatosis. This method would permit pharmacoepidemiologic studies to evaluate the medications associated with hepatic steatosis and would allow assessment of differences by HIV status. ALARM would be of particular use to evaluate the impact of obesogenic ART use on development of hepatic steatosis among PWH. Studies could also be conducted to determine the risk of clinical outcomes (e.g., decompensated cirrhosis or hepatocellular carcinoma) associated with hepatic steatosis. PWH are at increased risk of liver-related mortality and, given the expected increase in prevalence of hepatic steatosis in the years to come, it will be critical to define the clinical consequences of hepatic steatosis to inform mitigation strategies by HIV status.[32]

Among the five CT scans that did not achieve agreement for classification at the 40 HU threshold, the measurement variability between ALARM and radiologist was low, ranging from -2.8 to 4.7 HU. This magnitude of variability is similar to that observed between other automated liver assessment tools when compared to manual assessment of absolute liver attenuation.[29,33] Moreover, variability of absolute liver attenuation within an individual noncontrast CT scan has been reported to range between 1.8 to 3.1 HU in abdominal scans and 3.9 to 6.7 HU on chest scans, thus reflecting the expected variability in the measurement of liver attenuation.[27]

We observed that sensitivity analyses evaluating the performance of ALARM at alternative absolute liver attenuation thresholds of 48 HU and 51 HU demonstrated

similar AUROCs. The performance characteristics of ALARM at these thresholds were also not significantly different from the primary analysis. Future larger studies employing predetermined attenuation thresholds will warrant sensitivity analyses to assess the robustness of measures of associations near the threshold value.

This study has several potential limitations. While ALARM offers objective assessment of liver attenuation to define presence and absence of moderate-to-severe hepatic steatosis, inclusion in this analysis is limited to those patients who previously underwent noncontrast CT imaging of the abdomen. Such inclusion criterion may limit generalizability; however, it facilitates enrichment of a patient population at greatest risk of hepatic steatosis. The threshold employed for the identification of hepatic steatosis has been validated for the identification of moderate-to-severe steatosis; however, patients with mild steatosis (i.e., 6-29% hepatic triglyceride content) may not be captured. While performance of ALARM had high AUROC at the liver attenuation threshold of 51 HU, correlating with mild steatosis,[10] noncontrast CT has poor performance of identification of hepatic steatosis encompassing <30% hepatic triglyceride content.[11,34]

This study has several strengths. The successful deployment of this open-source deep learning tool confirms the accuracy of ALARM, allowing for reproducibility in diverse patient cohorts by any investigator. Additionally, the use of ALARM for the assessment of clinical images within the VA healthcare system demonstrates the implementation of a novel tool within the largest integrated health system in the US that possesses one of the largest imaging repositories in the world. Lastly, the application of

ALARM in VACS is the first study to apply artificial intelligence for quantitative imaging analysis among PWH, a patient population at high risk for liver disease.

5. Conclusion

The open source, deep learning ALARM algorithm demonstrated excellent accuracy for moderate-to-severe hepatic steatosis among people with and without HIV. Application of ALARM to radiographic repositories within electronic medical record databases could facilitate the conduct of large-scale real-world studies to evaluate medications and other factors associated with steatosis and assess differences by HIV status.

Ethics Statement: The study was approved by the Institutional Review Boards of the Crescenz VAMC, VA Connecticut Healthcare System, and Yale University, and was conducted under a waiver of informed consent per 45 CFR §46.117(c).

References

1. Brunt EM. Pathology of fatty liver disease. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2007;20 Suppl 1:S40-48.
2. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology (Baltimore, Md).* 2016;64(1):73-84.
3. Cholankeril G, Wong RJ, Hu M, et al. Liver Transplantation for Nonalcoholic Steatohepatitis in the US: Temporal Trends and Outcomes. *Digestive diseases and sciences.* 2017;62(10):2915-2922.
4. Agarwal N, Iyer D, Gabbi C, et al. HIV-1 viral protein R (Vpr) induces fatty liver in mice via LXRα and PPARα dysregulation: implications for HIV-specific pathogenesis of NAFLD. *Scientific reports.* 2017;7(1):13362.
5. Pérez-Matute P, Pérez-Martínez L, Blanco JR, Oteo JA. Role of mitochondria in HIV infection and associated metabolic disorders: focus on nonalcoholic fatty liver disease and lipodystrophy syndrome. *Oxidative medicine and cellular longevity.* 2013;2013:493413.
6. Kaspar MB, Sterling RK. Mechanisms of liver disease in patients infected with HIV. *BMJ Open Gastroenterol.* 2017;4(1):e000166.
7. Cholankeril G, Perumpail RB, Pham EA, Ahmed A, Harrison SA. Nonalcoholic Fatty Liver Disease: Epidemiology, Natural History, and Diagnostic Challenges. *Hepatology (Baltimore, Md).* 2016;64(3):954.

8. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans Aging Cohort Study (VACS): Overview and description. *Medical care.* 2006;44(8 Suppl 2):S13-24.
9. Kodama Y, Ng CS, Wu TT, et al. Comparison of CT methods for determining the fat content of the liver. *AJR Am J Roentgenol.* 2007;188(5):1307-1312.
10. 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. *Academic radiology.* 2012;19(7):811-818.
11. Bohte AE, van Werven JR, Bipat S, Stoker J. The diagnostic accuracy of US, CT, MRI and 1H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: a meta-analysis. *Eur Radiol.* 2011;21(1):87-97.
12. Behrns KE, Tsiotos GG, DeSouza NF, Krishna MK, Ludwig J, Nagorney DM. Hepatic steatosis as a potential risk factor for major hepatic resection. *J Gastrointest Surg.* 1998;2(3):292-298.
13. Kooby DA, Fong Y, Suriawinata A, et al. Impact of steatosis on perioperative outcome following hepatic resection. *J Gastrointest Surg.* 2003;7(8):1034-1044.
14. Belghiti J, Hiramatsu K, Benoist S, Massault P, Sauvanet A, Farges O. Seven hundred forty-seven hepatectomies in the 1990s: an update to evaluate the actual risk of liver resection. *J Am Coll Surg.* 2000;191(1):38-46.
15. Huo Y, Terry JG, Wang J, et al. Fully automatic liver attenuation estimation combing CNN segmentation and morphological operations. *Medical physics.* 2019;46(8):3508-3519.
16. Tang Y, Gao R, Lee HH, et al. Phase identification for dynamic CT enhancements with generative adversarial network. *Medical physics.* 2021;48(3):1276-1285.

17. Fultz SL, Skanderson M, Mole LA, et al. Development and verification of a "virtual" cohort using the National VA Health Information System. *Medical care.* 2006;44(8 Suppl 2):S25-30.
18. Butt AA, McGinnis K, Rodriguez-Barradas MC, et al. HIV infection and the risk of diabetes mellitus. *AIDS (London, England).* 2009;23(10):1227-1234.
19. Armah KA, Chang CC, Baker JV, et al. Prehypertension, hypertension, and the risk of acute myocardial infarction in HIV-infected and -uninfected veterans. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2014;58(1):121-129.
20. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Archives of internal medicine.* 1998;158(16):1789-1795.
21. Sax PE, Erlandson KM, Lake JE, et al. Weight Gain Following Initiation of Antiretroviral Therapy: Risk Factors in Randomized Comparative Clinical Trials. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2020;71(6):1379-1389.
22. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology (Baltimore, Md).* 2006;43(6):1317-1325.
23. Lemoine M, Assoumou L, De Wit S, et al. Diagnostic Accuracy of Noninvasive Markers of Steatosis, NASH, and Liver Fibrosis in HIV-Monoinfected Individuals at

562 Risk of Nonalcoholic Fatty Liver Disease (NAFLD): Results From the ECHAM
563 Study. *Journal of acquired immune deficiency syndromes (1999).* 2019;80(4):e86-
564 e94.
565 24. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ. Comparison of
566 noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease.
567 *Clinical gastroenterology and hepatology : the official clinical practice journal of the*
568 *American Gastroenterological Association.* 2009;7(10):1104-1112.
569 25. Rothman KJ GS. Modern Epidemiology, Second ed. Philadelphia: Lippincott
570 Williams and Wilkins. 1998.
571 26. Pickhardt PJ, Park SH, Hahn L, Lee SG, Bae KT, Yu ES. Specificity of unenhanced
572 CT for non-invasive diagnosis of hepatic steatosis: implications for the
573 investigation of the natural history of incidental steatosis. *European radiology.*
574 2012;22(5):1075-1082.
575 27. Speliotes EK, Massaro JM, Hoffmann U, et al. Liver fat is reproducibly measured
576 using computed tomography in the Framingham Heart Study. *Journal of*
577 *gastroenterology and hepatology.* 2008;23(6):894-899.
578 28. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians:
579 systematic review of design, reporting standards, and claims of deep learning
580 studies. *BMJ (Clinical research ed).* 2020;368:m689.
581 29. Graffy PM, Sandfort V, Summers RM, Pickhardt PJ. Automated Liver Fat
582 Quantification at Nonenhanced Abdominal CT for Population-based Steatosis
583 Assessment. *Radiology.* 2019;293(2):334-342.

30. Price JC, Seaberg EC, Latanich R, et al. Risk factors for fatty liver in the Multicenter AIDS Cohort Study. *The American journal of gastroenterology.* 2014;109(5):695-704.

31. Lake JE, Overton T, Naggie S, et al. Expert Panel Review on Non-Alcoholic Fatty Liver Disease in Persons With HIV. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association.* 2020.

32. Estes C, Anstee QM, Arias-Loste MT, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016-2030. *Journal of hepatology.* 2018;69(4):896-904.

33. Kullberg J, Hedström A, Brandberg J, et al. Automated analysis of liver fat, muscle and adipose tissue distribution from CT suitable for large-scale studies. *Scientific reports.* 2017;7(1):10425.

34. Schwenzer NF, Springer F, Schraml C, Stefan N, Machann J, Schick F. Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography and magnetic resonance. *Journal of hepatology.* 2009;51(3):433-445.

Table 1. Characteristics of patients in the Veterans Aging Cohort Study with noncontrast abdominal computed tomography scan who underwent assessment with the Automatic Liver Attenuation Region-Of-Interest-based Measurement (ALARM), by human immunodeficiency virus (HIV) infection status.

	Overall (n=120)	People With HIV (n=51)	People Without HIV (n=69)	*P*
Demographics				
Median (IQR) age, years	61.1 (55.3-64.6)	61.1 (54.6-65.9)	61.2 (55.6-64.0)	0.76
Male sex, n (%)	118 (98.3%)	50 (98%)	68 (98.5%)	0.83
Race, n (%)				<0.01
White	11 (9.2%)	10 (19.6%)	1 (1.4%)	
Black	107 (89.2%)	41 (80.4%)	66 (95.6%)	
Other	2 (1.7%)	--	2 (2.9%)	
Comorbidities, n (%)				
BMI ≥30 kg/m^2	38 (31.7%)	11 (21.6%)	27 (39.1%)	0.04
Diabetes	62 (51.7%)	21 (41.2%)	41 (59.4%)	0.05
Hypertension	101 (84.2%)	38 (74.5%)	63 (91.3%)	0.01
Dyslipidemia[a]	59 (49.2%)	29 (56.9%)	30 (43.5%)	0.15
Alcohol abuse/dependence	56 (46.7%)	24 (47.1%)	32 (46.4%)	0.94
Hepatitis B virus infection	4 (3.3%)	4 (7.8%)	--	0.02
Hepatitis C virus infection	53 (44.2%)	29 (56.9%)	41 (34.9%)	0.02
Median FIB-4	1.78 (1.30-3.02)	2.19 (1.45-4.34)	1.56 (1.24-2.20)	<0.01
FIB-4 >3.25	29 (24.2%)	17 (33.3%)	12 (17.4%)	0.05
Laboratory Values				
Median (IQR) CD4 cell count, cells/mm^3[b]	--	364 (213-557.5)	--	
CD4 <14%[b], n (%)	--	6 (15%)	--	
Median (IQR) HIV RNA, copies/mL[b]	--	50 (48-152.5)	--	
HIV RNA >200 copies/mL[b], n (%)	--	8 (20%)	--	
Median (IQR) albumin, gm/dL	3.9 (3.4-4.3)	3.7 (2.8-4.2)	4 (3.6-4.3)	0.07
Median (IQR) AST, U/L	29.5 (23-48)	34.5 (28-53)	26.5 (22-41.5)	0.03
Median (IQR) ALT, U/L	28.0 (19-44)	27 (18-49)	28 (19-37.5)	0.63
Median (IQR) platelets, x 10^6/L	202.5 (158.5-252.5)	194 (140-249)	207 (170-257)	0.19
Median (IQR) cholesterol, mg/dL	167 (142-191)	151.5 (135-179.5)	182 (151-201)	<0.01
Median (IQR) HDL mg/dL	42 (34-50)	39.5 (29.5-49.5)	43.5 (36-50.5)	0.13
Median (IQR) triglycerides, mg/dL	117 (86-172)	130.5 (98-187.5)	111.5 (82-158)	0.06
Median (IQR) LDL, mg/dL	93 (72-116)	80.5 (61.5-104.5)	107 (87-128)	<0.01
CT Characteristics				
Indication for CT				0.57
Pain	60 (50.0%)	38 (55.1%)	22 (43.1%)	
Disease Screening/Staging	19 (15.8%)	11 (15.9%)	8 (15.7%)	
Bleeding	9 (7.5%)	6 (8.7%)	3 (5.9%)	
Mass	8 (6.7%)	4 (5.8%)	4 (7.8%)	
Infection	8 (6.7%)	3 (4.3%)	5 (9.8%)	
Other	16 (13.3%)	7 (10.1%)	9 (17.6%)	

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CT, computed tomography; FIB-4, Fibrosis-4 Index; HIV, human immunodeficiency virus; IQR, interquartile range.
[a]Defined by triglyceride level >150 mg/dL or HDL <40 mg/dL in males or <50 mg/dL in females.
[b]HIV and CD4 cell count assessed in 40 patients.

Table 2. Absolute liver attenuation measurements based on assessment from the Automatic Liver Attenuation Region-Of-Interest-based Measurement (ALARM) and radiologist assessment, by human immunodeficiency virus (HIV) infection status.

	Overall (n=120)	**People with HIV (n=51)**	**People Without HIV (N=69)**	***P***
ALARM Metrics				
Mean (SD) Hounsfield units	49.8 (9.3)	49.7 (9.6)	50.0 (9.2)	0.88
Liver attenuation <40 HU	15 (12.5%)	8 (15.7%)	7 (10.1%)	0.36
Radiology Metrics				
Mean (SD) Hounsfield units	51.8 (11.4)	52.1 (13.5)	51.1 (9.7)	0.78
Liver attenuation <40 HU	12 (10%)	6 (11.8%)	6 (8.7%)	0.58

Abbreviations: HU, Hounsfield units; SD, standard deviation.

Table 3. Performance characteristics of the Automatic Liver Attenuation Region-Of-Interest-based Measurement (ALARM) for identification of moderate-to-severe hepatic steatosis compared to radiologist review.

	Radiologist Review		
ALARM	≥40 HU	<40 HU	
Liver attenuation ≥40 HU	104	1	105
Liver attenuation <40 HU	4	11	15
	108	12	120
Sensitivity	91.7% (95% CI: 61.5-99.8%)		
Specificity	96.3% (95% CI: 90.8-99.0%)		
Positive Predictive Value	73.3% (95% CI: 44.9-92.2%)		
Negative Predictive Value	99.0% (95% CI: 94.8-100.0%)		

Abbreviations: CI, confidence interval; HU, Hounsfield units

Table 4. Mean absolute liver attenuation measurements based on the Automatic Liver Attenuation Region-Of-Interest-based Measurement (ALARM) and radiologist assessment for CT scans with disagreement in classification of moderate-to-severe steatosis.

CT Scan Number	ALARM	Radiology Assessment	Difference
7	38.6 HU	40.5 HU	1.9
32	39.6 HU	41.4 HU	1.8
66	41.8 HU	39.0 HU	-2.8
78	39.6 HU	44.3 HU	4.7
110	36.6 HU	40.8 HU	4.2

Abbreviations: CT, computed tomography; HU, Hounsfield units

Figure 1. **Correlation of mean absolute liver attenuation between mean radiologist measurements and mean ALARM measurements, displayed in Hounsfield Unit (HU).** Overall Pearson's correlation coefficient r = 0.93.

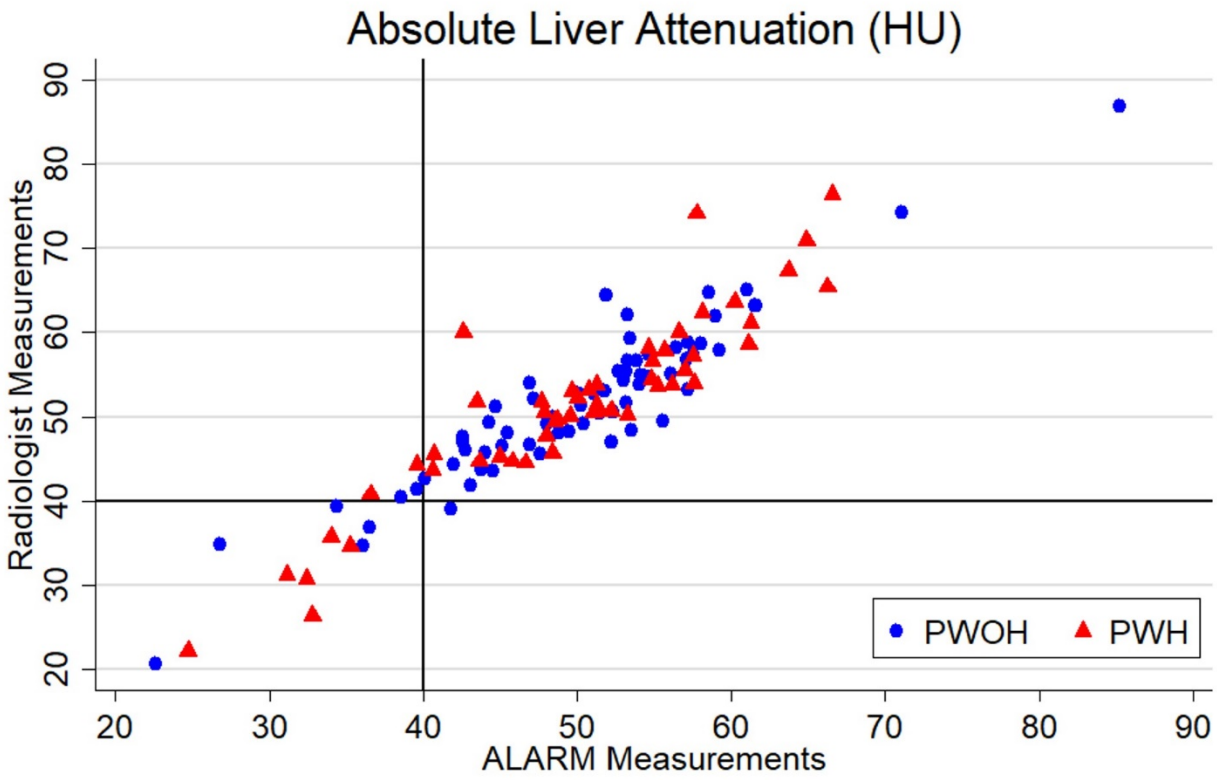
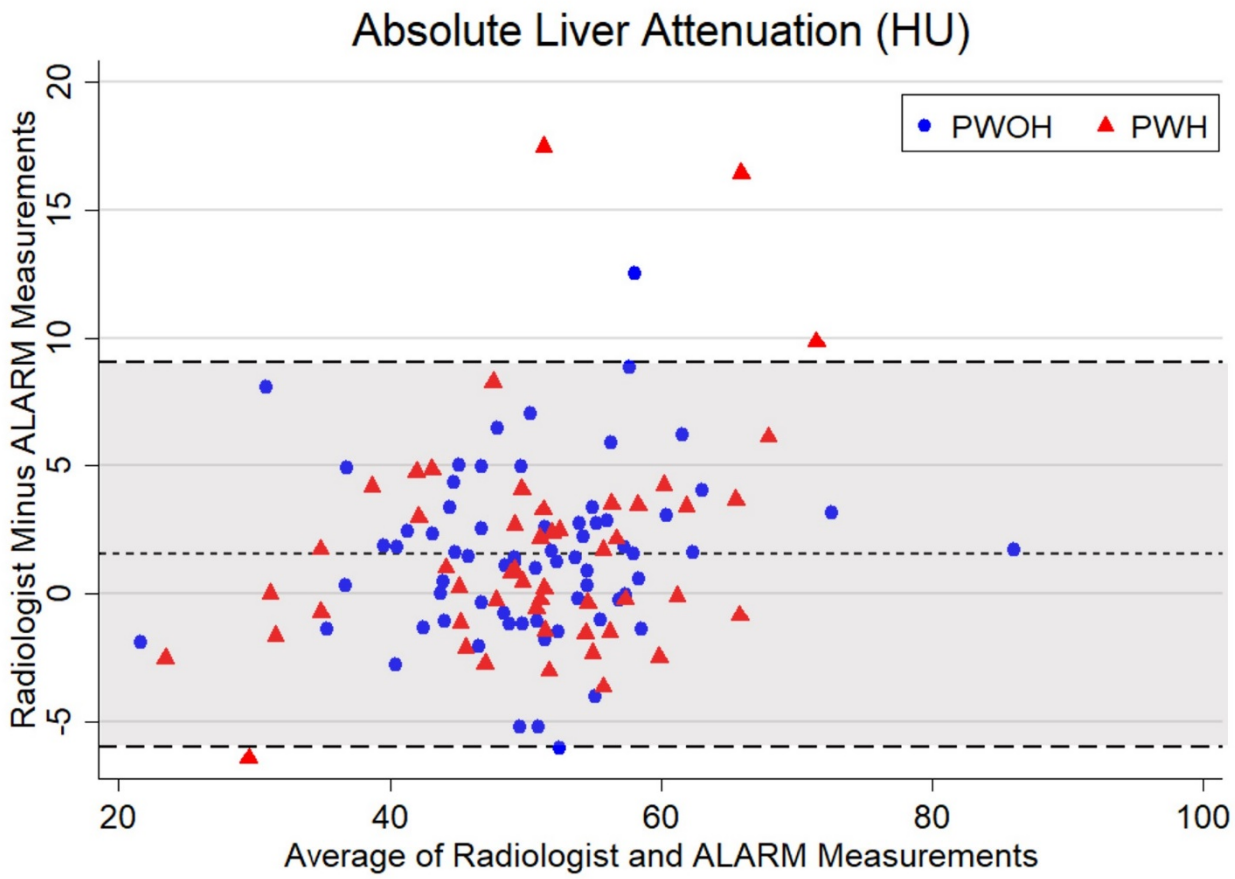


Figure 2. **Bland-Altman plot between mean radiologist measurement of liver attenuation and mean ALARM measurement of liver attenuation.** The gray area indicates the 95% confidence interval.



Performance of fully automated liver attenuation assessment for measurement of hepatic steatosis within noncontrast computed tomography scans among people with and without HIV

Jessie Torgersen, Scott Akers, Yuankai Huo, James G. Terry, J. Jeffrey Carr, Alexander T. Ruutiainen, Melissa Skanderson, Woody Levin, Joseph K. Lim, Tamar H. Taddei, Kaku So-Armah, Debika Bhattacharya, Christopher Rentsch, Li Shen, Rotonya Carr, Russell T. Shinohara, Michele McClain, Matthew Freiberg, Amy C. Justice, Vincent Lo Re III

Table of Contents:

Table S1…….. 2

Table S2…….. 3

Table S3…….. 4

Table S4…….. 5

Table S1. Performance characteristics of the Automatic Liver Attenuation Region-of-Interest-based Measurement (ALARM) for identification of moderate-to-severe hepatic steatosis, defined as liver attenuation <40 Hounsfield Units (HU), among people without human immunodeficiency virus (HIV) infection compared to manual radiologist review.

	Radiologist Assessment		
ALARM Assessment	≥40 HU	<40 HU	
Liver attenuation ≥40 HU	61	1	62
Liver attenuation <40 HU	2	5	7
Total	63	6	69
Sensitivity	83.3% (95% CI: 35.9-99.6%)		
Specificity	96.8% (95% CI: 89.0-99.6%)		
Positive Predictive Value	71.4% (95% CI: 29.0-96.3%)		
Negative Predictive Value	98.4% (95% CI: 91.3-99.9%)		
Abbreviations: CI, confidence interval; HU, Hounsfield units			

Table S2. Performance characteristics of the Automatic Liver Attenuation Region-of-Interest-based Measurement (ALARM) for identification of moderate-to-severe hepatic steatosis, defined as liver attenuation <40 Hounsfield Units (HU), among people with human immunodeficiency virus (HIV) infection compared to manual radiologist review.

	Radiologist Assessment		
ALARM Assessment	≥40 HU	<40 HU	
Liver attenuation ≥40 HU	43	0	43
Liver attenuation <40 HU	2	6	8
Total	45	6	51
Sensitivity	100% (95% CI: 91.8-100%)		
Specificity	95.5% (95% CI: 84.8-99.4%)		
Positive Predictive Value	75.0% (95% CI: 34.9-96.8%)		
Negative Predictive Value	100% (95% CI: 91.8-100%)		
Abbreviations: CI, confidence interval; HU, Hounsfield units			

Table S3. Performance characteristics of the Automatic Liver Attenuation Region-of-Interest-based Measurement (ALARM) for identification of moderate-to-severe hepatic steatosis, defined as liver attenuation <48 Hounsfield units (HU).

	Radiologist Assessment		
ALARM Assessment	≥48 HU	<48 HU	
Liver attenuation ≥48 HU	73	9	82
Liver attenuation <48 HU	3	35	38
Total	76	44	120
Sensitivity	79.5% (95% CI: 64.7-90.2%)		
Specificity	96.1% (95% CI: 88.9-99.2%)		
Positive Predictive Value	92.1% (95% CI: 78.6-98.3%)		
Negative Predictive Value	89.0% (95% CI: 80.2-94.9%)		
AUROC	0.91 (95% CI: 0.85-0.96)		
Abbreviations: AUROC, area under the receiving operator characteristic; CI, confidence interval; HU, Hounsfield units			

Table S4. Performance characteristics of the Automatic Liver Attenuation Region-of-Interest-based Measurement (ALARM) for identification of mild hepatic steatosis, defined as liver attenuation <51 Hounsfield units (HU).

	Radiologist Assessment		
ALARM Assessment	≥51 HU	<51 HU	
Liver attenuation ≥51 HU	52	8	60
Liver attenuation <51 HU	11	49	60
Total	63	57	120
Sensitivity	86.0% (95% CI: 74.2-93.7%)		
Specificity	82.5% (95% CI: 70.9-90.9%)		
Positive Predictive Value	81.7% (95% CI: 69.6-90.5%)		
Negative Predictive Value	86.7% (95% CI: 75.4-94.1%)		
AUROC	0.84 (95% CI: 0.78-0.91)		
Abbreviations: AUROC, area under the receiving operator characteristic; CI, confidence interval; HU, Hounsfield units			