Radiology

ORIGINAL RESEARCH • GASTROINTESTINAL IMAGING

Two-dimensional Convolutional Neural Network Using Quantitative US for Noninvasive Assessment of Hepatic Steatosis in NAFLD

Sun Kyung Jeon, MD, PhD • Jeong Min Lee, MD, PhD • Ijin Joo, MD, PhD • Jeong Hee Yoon, MD, PhD • Gunwoo Lee, PhD

From the Department of Radiology, Seoul National University Hospital and Seoul National University College of Medicine, 101 Daehangno, Jongno-gu, Seoul 03080, Korea (S.K.J., J.M.L., I.J., J.H.Y.); Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea (J.M.L.); and Ultrasound R&D 2 Group, Health & Medical Equipment Business, Samsung Electronics Co, Ltd, Seoul, Korea (G.L.). Received June 21, 2022; revision requested August 29; revision received September 30; accepted November 4. **Address correspondence to** J.M.L. (email: *jmsh@snu.ac.kr*).

Supported by a research grant from Samsung Medison (06-2020-2040) and a grant from the Seoul National University Hospital research fund (03-2018-0230).

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

See also the editorial by Sidhu and Fang in this issue.

Radiology 2023; 000:1–9 • https://doi.org/10.1148/radiol.221510 • Content codes: GI US AI

Background: Quantitative US (QUS) using radiofrequency data analysis has been recently introduced for noninvasive evaluation of hepatic steatosis. Deep learning algorithms may improve the diagnostic performance of QUS for hepatic steatosis.

Purpose: To evaluate a two-dimensional (2D) convolutional neural network (CNN) algorithm using QUS parametric maps and B-mode images for diagnosis of hepatic steatosis, with the MRI-derived proton density fat fraction (PDFF) as the reference standard, in patients with nonalcoholic fatty liver disease (NAFLD).

Materials and Methods: Consecutive adult participants with suspected NAFLD were prospectively enrolled at a single academic medical center from July 2020 to June 2021. Using radiofrequency data analysis, two QUS parameters (tissue attenuation imaging [TAI] and tissue scatter-distribution imaging [TSI]) were measured. On B-mode images, hepatic steatosis was graded using visual scoring (none, mild, moderate, or severe). Using B-mode images and two QUS parametric maps (TAI and TSI) as input data, the algorithm estimated the US fat fraction (USFF) as a percentage. The correlation between the USFF and MRI PDFF was evaluated using the Pearson correlation coefficient. The diagnostic performance of the USFF for hepatic steatosis (MRI PDFF ≥5%) was evaluated using receiver operating characteristic curve analysis and compared with that of TAI, TSI, and visual scoring.

Results: Overall, 173 participants (mean age, 51 years ± 14 [SD]; 96 men) were included, with 126 (73%) having hepatic steatosis (MRI PDFF ≥5%). USFF correlated strongly with MRI PDFF (Pearson r = 0.86, 95% CI: 0.82, 0.90; $P < .001$). For diagnosing hepatic steatosis (MRI PDFF ≥5%), the USFF yielded an area under the receiver operating characteristic curve of 0.97 (95% CI: 0.93, 0.99), higher than those of TAI, TSI, and visual scoring ($P = .015$, .006, and < .001, respectively), with a sensitivity of 90% (95% CI: 84, 95 [114 of 126]) and a specificity of 91% (95% CI: 80, 98 [43 of 47]) at a cutoff value of 5.7%.

Conclusion: A deep learning algorithm using quantitative US parametric maps and B-mode images accurately estimated the hepatic fat fraction and diagnosed hepatic steatosis in participants with nonalcoholic fatty liver disease.

ClinicalTrials.gov registration nos. NCT04462562, NCT04180631

Supplemental material is available for this article.

The prevalence of nonalcoholic fatty liver disease (NAFLD), the leading cause of chronic liver disease worldwide, has increased considerably over the past 2 decades (1). NAFLD ranges from isolated hepatic steatosis, the key histologic feature of NAFLD, to nonalcoholic steatohepatitis, advanced fibrosis, cirrhosis, and even hepatocellular carcinoma (2–4). NAFLD is also associated with metabolic syndrome and cardiovascular disease, such as coronary artery disease or atherosclerosis (5,6). Because timely management of hepatic steatosis can arrest or reverse the disease before irreversible changes occur, early detection and accurate staging of hepatic steatosis is clinically important in patients with NAFLD (3,7).

Although liver biopsy remains the reference standard for diagnosing NAFLD, its invasiveness and error rate necessitates development of a noninvasive diagnostic and monitoring method for hepatic steatosis (8). MRI-based techniques, including MRI proton density fat fraction (PDFF) measurement, can accurately and reproducibly quantify hepatic fat (9) but are not widely applied due to high cost and low accessibility. US is widely used for diagnosing hepatic steatosis, given its availability and cost-effectiveness. However, conventional B-mode US has drawbacks, such as its qualitative and subjective nature and modest accuracy, particularly in mild steatosis, where less than 30% of hepatocytes are affected (8,10).

To overcome these drawbacks of conventional B-mode US, quantitative US (QUS) techniques using raw radiofrequency data, which provide more comprehensive information about tissue composition, have been developed

Abbreviations

AUC = area under the receiver operating characteristic curve, CNN = convolutional neural network, NAFLD = nonalcoholic fatty liver disease, PDFF= proton density fat fraction, QUS = quantitative US, TAI = tissue attenuation imaging, TSI = tissue scatter-distribution imaging, 2D = two-dimensional, USFF = US fat fraction

Summary

A deep learning–based two-dimensional convolutional neural network algorithm using quantitative US parametric maps acquired from radiofrequency data can accurately quantify hepatic fat fraction and diagnose hepatic steatosis.

Key Results

- In 173 prospectively enrolled participants with suspected nonalcoholic fatty liver disease, a deep learning algorithm using quantitative US (QUS) parametric maps and B-mode images accurately estimated the hepatic fat fraction, which showed excellent correlation with the MRI proton density fat fraction (r = 0.86, $P < .001$).
- For diagnosing hepatic steatosis, the deep learning–estimated US fat fraction had an area under the receiver operating characteristic curve of 0.97, higher than those of the QUS parameters (tissue attenuation imaging and tissue scatter-distribution imaging) and visual score ($P = .015$, .006, and < .001, respectively).

(11–14). Various radiofrequency data–driven QUS parameters, such as the attenuation coefficient and backscatter coefficient, correlate significantly with hepatic steatosis and have good diagnostic performance (89.0%–96.4%) (11,13,15,16). Moreover, a multivariable logistic regression model using various QUS parameters has yielded high diagnostic accuracy for steatohepatitis and hepatic steatosis (12,17).

Deep learning–based US approaches have recently been introduced for hepatic steatosis assessment (18,19) and can improve the diagnostic performance of QUS. Additionally, deep learning–based two-dimensional (2D) convolutional neural networks (CNNs) can provide high accuracy and robustness by adopting 2D images as input data (20). We hypothesized that a 2D CNN algorithm can help improve the diagnostic performance of QUS using multiple QUS parametric maps and B-mode images as input data, thus allowing a multiparametric approach.

Therefore, we aimed to develop and prospectively validate a 2D CNN algorithm using QUS parametric maps and B-mode images for diagnosing hepatic steatosis, with the MRI-derived PDFF as the reference standard, in patients with NAFLD.

Materials and Methods

Study Design and Participants

This prospective study was conducted at Seoul National University Hospital, an academic tertiary care center in Korea (ClinicalTrials.gov registration no. NCT04462562). The institutional review board approved this prospective study (IRB no. 2002–020–1099) and all participants provided written informed consent. One major variable of the study was to evaluate the diagnostic performance of the deep learning algorithm for hepatic fat quantification using the MRI PDFF as the reference standard. This study was supported by a research grant from Samsung Medison. One author (G.L.) is an employee of the industry; however, the nonemployee authors had control of the data and information submitted for publication.

From July 2020 to June 2021, participants who were 18 years or older, who were referred to the radiology department for liver US evaluation for known or clinically suspected NAFLD or were scheduled to undergo hepatectomy for liver donation, and who had signed informed consent were prospectively enrolled. The exclusion criteria were as follows: *(a)* clinical, laboratory, or histologic evidence of liver disease other than NAFLD; *(b)* excessive alcohol consumption (≥14 and ≥7 drinks per week for men and women, respectively); *(c)* hepatotoxic or steatogenic medication use; *(d)* previous liver surgery; *(e)* contraindication for MRI; and *(f)* missing QUS or MRI data (Fig 1). Anthropometric data, including age, sex, body mass index, and biochemical tests that included aspartate aminotransferase and alanine aminotransferase levels, were recorded by nonauthor study coordinators. All participants underwent US examination and chemical shift–encoded MRI PDFF assessment on the same day whenever possible, or within a 14-day period.

US Data Acquisition

US examination and US radiofrequency data acquisition.— All participants underwent liver US examination with an RS85 system (Samsung Medison) with a convex probe (CA1-7) performed by one of two abdominal radiologists (S.K.J. and J.M.L., each with more than 8 years of abdominal US experience) who were blinded to MRI PDFF results. All participants were requested to fast at least 4 hours before the US examination. Participants were positioned supine with the right arm abducted during examination. Using a right intercostal plane near the hepatic hilum, the representative hepatic parenchyma of the right liver was visualized, avoiding major vasculature or focal liver lesions. A radiologist made five data acquisitions at the same location in the right liver lobe during a breath hold with a fixed set of time-gain compensation and focus position, and the radiofrequency data were automatically recorded. During US examination, visual hepatic steatosis scores were determined by the operator using the Hamaguchi scoring system, whereby 0 indicates none, 1 indicates mild, 2 indicates moderate, and 3 indicates severe steatosis (21) (Fig S1).

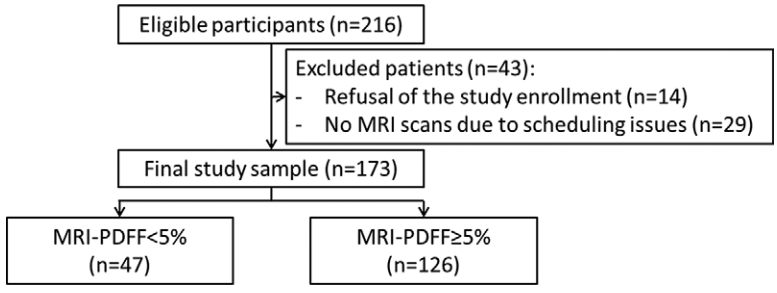

Figure 1: Flowchart shows participant inclusion. PDFF = proton density fat fraction.

QUS parametric map generation using US radiofrequency data.—Two QUS parameters, tissue attenuation imaging (TAI) and tissue scatter-distribution imaging (TSI), were computed from the US radiofrequency data, and TAI and TSI maps were created. The TAI map was a pixel-by-pixel map of the local center frequency associated with US attenuation properties (22,23). The TSI map consisted of local Nakagami parameters, which reflect the arrangement and concentration of scatterers within the tissue (24,25). The theories underlying these two parameters are summarized in Appendix S1.

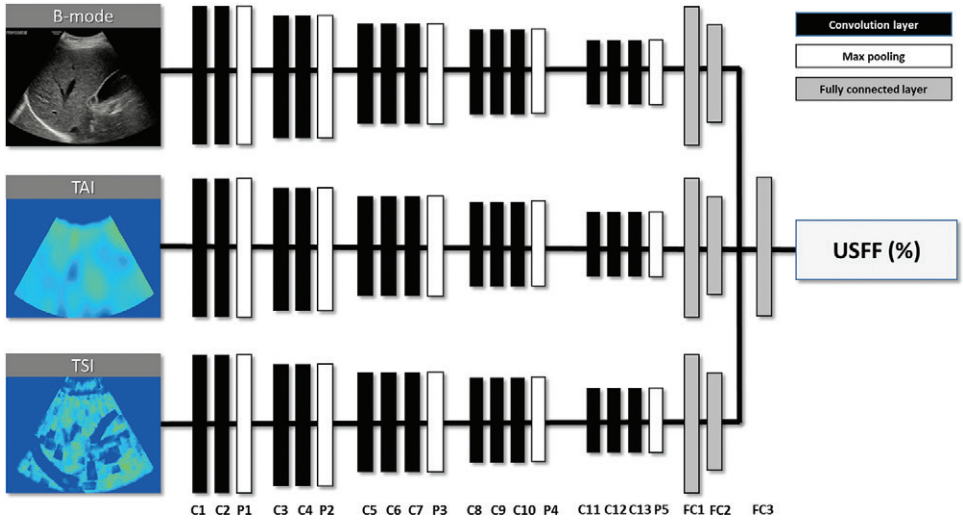

Figure 2: Schematic shows development of the two-dimensional convolutional neural network algorithm for estimating the hepatic fat fraction. B-mode images and tissue attenuation imaging (TAI) and tissue scatter-distribution imaging (TSI) maps generated from analysis of radiofrequency data are used as input data. Three input data sets (B-mode image, TAI map, and TSI map) generate one output, which is the deep learning–estimated US fat fraction (USFF). C = convolutional layer, FC = fully connected layer, P = pooling layer.

Measurement of QUS parameters.—Measurement of the two QUS parameters (TAI and TSI) was performed during US examination with the RS85 machine (Samsung Medison). For measurement, a 2 × 3–cm fan-shaped region of interest was placed on the right liver lobe at least 2 cm below the liver capsule, avoiding areas with large vessels, focal lesions, and reverberation artifacts or shadowing according to vendor recommendations. Areas with errors in the calculation of parameters, such as vascular structures, were automatically excluded and present as vacant on the TAI and TSI maps. Additionally, the reliability of measurements are presented as an R^2 value, such that the operator attempted to obtain values with an R^2 greater than or equal to 0.6. For each QUS parameter, the five measurements were averaged as the representative value for each participant according to vendor recommendations (Fig S2). The distributions of TAI and TSI are plotted in Figure S3.

Two-dimensional CNN Algorithm for Hepatic Fat Quantification

To train the deep learning–based 2D CNN algorithm, we used raw US radiofrequency data that were collected in a previous prospective study (ClinicalTrials.gov registration no. NCT04180631), which evaluated the performance of QUS parameters (calculated values) using radiofrequency data analysis for assessing hepatic steatosis in NAFLD (n = 120) (13). B-mode images and QUS parametric maps (TAI and TSI; ie, 2D data) derived from the radiofrequency data of all participants (120 of 120) in the previous study were used as the training and internal validation sets of the 2D CNN algorithm for hepatic fat quantification. In the present study, the performance of the 2D CNN algorithm was prospectively tested. Data of the 120 participants were only used for training and internal validation of the 2D CNN algorithm, not for the prospective test of its performance, and there was no overlap between training, internal validation, and test set participants (Table S1).

The CNN was configured to extract feature information from each input using a visual geometry group network, then concatenate these and derive a regression output through a fully connected layer (Fig 2). Details for development of the 2D CNN algorithm are presented in Appendix S2 and Table S2), and the code is available for research use at *https://github.com/SamsungLabs/NAFLD*. Three inputs from one US radiofrequency data set (B-mode image, TAI map, and TSI map) generated one output, which was the deep learning–estimated US fat fraction (USFF) presented as a percentage. As each participant had five US images, the 2D CNN algorithm generated five outputs per participant, which were averaged to yield per-participant estimates. Additionally, a 2D CNN algorithm using only QUS parametric maps (TAI and TSI maps), without B-mode images, as input data was developed separately and an estimated hepatic fat fraction was generated. The results are summarized in Appendix S3 and Figure S4. Deep CNN processing was performed using Python version 3.6.9 (Python Software Foundation; *https://www.python.org/psf*) and TensorFlow version 1.10.0 *(https://www.tensorflow.org/install/source)*.

Reference Standard: Chemical Shift–encoded MRI PDFF

Chemical shift–encoded MRI PDFF estimation was used as the reference standard because it is a well-validated, accurate, and safe method for assessing hepatic steatosis and is thus widely used as a reference standard for quantifying hepatic fat content (4,26).

All participants underwent chemical shift–encoded MRI PDFF assessment with a 3-T MRI scanner (MAGNETOM Skyra; Siemens Healthineers). Details of the MRI protocol and parameters are summarized in Appendix S4. One abdominal radiologist (J.S.K.), who was blinded to laboratory test and USFF results, manually placed a 1-cm-diameter circular region of interest in each of the nine Couinaud liver segments of the PDFF map (27). Averaged values from liver segment five through eight

were calculated and used as the reference hepatic fat content value to match the USFF and QUS data that were obtained from the right hepatic lobe (28). As previously proposed, the degree of hepatic steatosis was graded as mild (S1), moderate (S2), and severe (S3) at MRI PDFF thresholds of 5%, 15%, and 25%, respectively (4).

Statistical Analysis

The sample size was estimated based on one of the major variables of the study, which evaluated the diagnostic accuracy of three parameters. Details of sample size estimation are presented in Appendix S5.

The 2D CNN algorithm was evaluated at the participant level. The mean USFF values according to hepatic steatosis grade were compared using one-way analysis of variance with the Tukey post hoc test. The Pearson correlation coefficient was calculated to evaluate the USFF and MRI PDFF correlation. Bland-Altman analysis with 95% limits of agreement was additionally performed. Linearity was evaluated using a sequential test of polynomial fits (three-degree and two-degree polynomial linear regression) to plot the USFF versus MRI PDFF (29). When the linearity test over the entire MRI PDFF range failed, the linear range was identified by finding an MRI PDFF range in which the coefficients of nonlinear terms of the three-degree and two-degree polynomial fits were not statistically significantly different from zero. The linear regression slope, intercept, and R^2 were evaluated using linear regression analysis, and quadratic regression analysis was also performed. Receiver operating characteristic curve analysis was used to evaluate the performance of the USFF, QUS parameters (TAI and TSI), and visual score for evaluating hepatic steatosis based on MRI PDFF values greater than or equal to 5%, greater than or equal to 15%, and greater than or equal to 25%. The DeLong test was used to compare areas under the receiver operating characteristic curve (AUCs) of the USFF and QUS parameters or visual score. For comparison of AUCs, a Bonferroni-adjusted $P < .017$ (0.05/3) was considered to indicate a statistically significant difference. Optimal cutoff values for USFF, TAI, and TSI were determined using the maximal Youden index (30), with performance parameters including sensitivity, specificity, positive predictive value, and negative predictive value for evaluating hepatic steatosis based on MRI PDFF values greater than or equal to 5%, greater than or equal to 15%, and greater than or equal to 25%. USFF cutoff values for sensitivity and specificity exceeding 90% were also derived. Ninety percent winsorization was used to adjust outliers for USFF, TAI, and TSI scatterplots (Appendix S6) (31). All statistical analyses were performed using MedCalc version 18.11.6 (MedCalc Software) and SAS version 9.4 (SAS Institute). $P < .05$ was considered indicative of a statistically significant difference, except for the aforementioned AUC comparison.

Table 1: Participant Characteristics

Characteristic	Value (n = 173)
Age (y)*	51 ± 14 (19–74)
Sex	
M	96 (56)
F	77 (45)
Participants with clinically suspected or known NAFLD	148 (86)
Body mass index (kg/m^2)*	26.5 ± 3.5 (19.2–39.8)
Skin-to-liver capsule distance (mm)*	20.7 ± 4.3
Aspartate aminotransferase level (IU/L)*	28.7 ± 14.1 (9–99)
Alanine aminotransferase level (IU/L)*	39.2 ± 31.4 (7–211)
Hepatic steatosis visual score	
S0	55 (32)
S1	29 (17)
S2	66 (38)
S3	23 (13)
MRI PDFF (%)*	11.2 ± 7.8 (1.5–46.4)
<5%	47 (27)
≥5 to <15%	79 (46)
≥15% to <25%	37 (21)
≥25%	10 (6)

Note.—Except where indicated, data are numbers of participants, with percentages in parentheses. For hepatic steatosis visual scoring, S0 indicates none, S1 indicates mild, S2 indicates moderate, and S3 indicates severe steatosis. NAFLD = nonalcoholic fatty liver disease, PDFF = proton density fat fraction.

* Data are means ± SDs, with ranges in parentheses.

Results

Participant Characteristics

In total, 216 participants were screened and 43 were excluded due to refusal of study participation (n = 14) or lack of MRI scans due to scheduling issues (n = 29). Finally, 173 participants (mean age, 51 years ± 14 [SD]; 96 men) were included in the study (Fig 1). The mean MRI PDFF was 11.2% ± 7.8 (range, 1.5%–46.4%). A total of 126 participants had hepatic steatosis

Table 2: Distribution of Deep Learning–based USFF according to Hepatic Steatosis Grade

Parameter	MRI PDFF				P Value	Post Hoc P Value		
	<5% (S0)	5%–15% (S1)	15%–25% (S2)	≥25% (S3)		S0 vs S1	S1 vs S2	S2 vs S3
USFF (%)	3.5 ± 1.8	9.8 ± 3.7	16.7 ± 2.8	20.4 ± 3.2	<.001	<.001	<.001	.006

Note.—Except where indicated, data are means ± SDs. MRI PDFF thresholds indicate the hepatic steatosis grade such that S0 indicates none, S1 indicates mild, S2 indicates moderate, and S3 indicates severe steatosis. Post hoc P values were calculated using one-way analysis of variance with the Tukey post hoc test. USFF = US fat fraction, PDFF = proton density fat fraction.

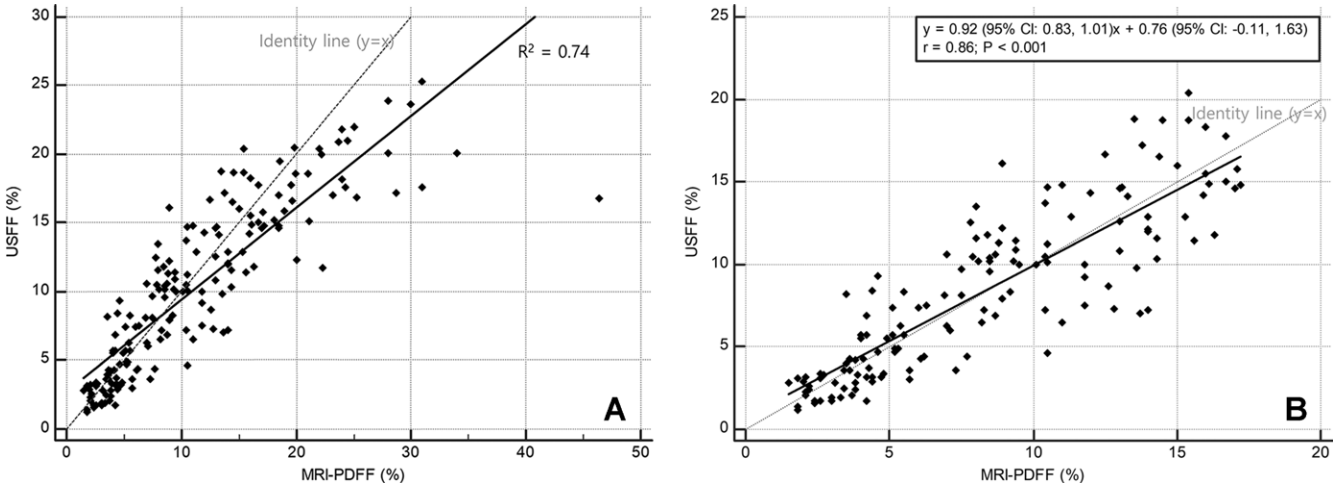

Figure 3: Scatterplots of the deep learning–estimated US fat fraction (USFF) versus the MRI proton density fat fraction (PDFF) show **(A)** the entire MRI PDFF range (MRI PDFF <47%) and **(B)** the linear range (MRI PDFF ≤18%), as well as the identity line and the linear regression line.

(MRI PDFF ≥5%). The median interval between US and MRI was 0 days (range, 0–14 days [IQR, 0–0 days]). Participant characteristics are summarized in Table 1.

Correlation between USFF and MRI PDFF

The distribution of USFF according to hepatic steatosis grade is summarized in Table 2. The USFF showed a correlation with the MRI PDFF (r = 0.86, 95% CI: 0.82, 0.90; $P < .001$). The USFF and MRI PDFF showed linearity for an MRI PDFF value less than or equal to 18%. Within the linear range (MRI PDFF ≤18%), linear regression analysis of the USFF against the MRI PDFF revealed a slope of 0.92 (95% CI: 0.83, 1.01), an intercept of 0.76 (95% CI: −0.11, 1.63), and an R^2 of 0.74 (Pearson, r = 0.86). The slope and intercept were not significantly different from 0 and 1, respectively ($P = .08$ and .09, respectively). For the entire MRI PDFF range (MRI PDFF <47%), regression analysis of the USFF against the MRI PDFF demonstrated an R^2 of 0.74 in the linear model and a polynomial R^2 of 0.83 in the quadratic regression model (Fig 3). The mean bias of the USFF over the entire MRI PDFF range was −0.97% ($P = .002$), and 95% limits of agreement ranged from −8.8% to 6.9% (Fig 4).

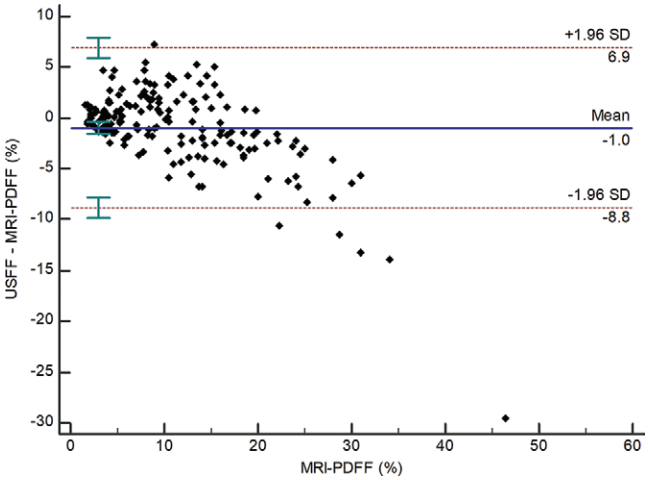

Figure 4: Bland-Altman plot shows the difference between the deep learning–estimated US fat fraction (USFF) and the MRI-derived proton density fat fraction (PDFF).

Diagnostic Performance of USFF for Evaluating Hepatic Steatosis

The USFF AUCs for evaluating hepatic steatosis based on MRI PDFF values greater than or equal to 5%, greater than or equal to 15%, and greater than or equal to 25% were 0.97 (95% CI: 0.93, 0.99; $P < .001$), 0.96 (95% CI: 0.92, 0.99; $P < .001$), and 0.95 (95% CI: 0.90, 0.97; $P < .001$), respectively. Cutoff values and the corresponding sensitivity, specificity, positive predictive value, and negative predictive value are shown in Table 3.

For diagnosing hepatic steatosis (MRI PDFF ≥5%), the USFF showed a sensitivity of 90% (114 of 126; 95% CI: 84, 95) and specificity of 91% (43 of 47; 95% CI: 80, 98) at the optimal cutoff value of 5.7%. Using the cutoff value of 5.7%, the positive predictive value and negative predictive value were 97% (114 of 118; 95% CI: 92, 99) and 78% (43 of 55; 95% CI: 68, 86), respectively. For diagnosing moderate to severe hepatic steatosis (MRI PDFF ≥15%), the USFF had a sensitivity of 89% (42 of 47; 95% CI: 77, 97), specificity of 91% (115 of 126; 95% CI: 85, 96), positive predictive value of 79% (42 of 53; 95% CI: 68, 87), and negative predictive value of 96% (115 of 120; 95% CI: 91, 98) at the optimal cutoff of 14.1%.

Comparison of Diagnostic Performance between USFF, QUS Parameters, and Visual Score for Hepatic Steatosis

Performance of the USFF, QUS parameters (TAI and TSI), and visual score for evaluating hepatic steatosis based on MRI PDFF values greater than or equal to 5%, greater than or equal to 15%, and greater than or equal to 25% is summarized in Table 4.

For evaluating hepatic steatosis based on MRI PDFF values greater than or equal to 5% and greater than or equal to 15%, the USFF had greater AUCs than TAI, TSI, and visual scoring, respectively (USFF vs TAI, TSI, and visual scoring: 0.97 [95% CI: 0.93, 0.99] vs 0.92 [95% CI: 0.87, 0.95], 0.91 [95% CI: 0.85, 0.94], and 0.84 [95% CI: 0.77, 0.89], $P = .015$, .006,

Table 3: Diagnostic Performance of Deep Learning–based USFF for Hepatic Steatosis

Hepatic Fat Content	AUC	Cutoff (%)	Sensitivity*	Specificity*	PPV*	NPV*
MRI PDFF ≥5%	0.97 [0.93, 0.99]					
Optimal threshold†		>5.7	90 (114/126) [84, 95]	91 (43/47) [80, 98]	97 (114/118) [92, 99]	78 (43/55) [68, 86]
Threshold for 90% sensitivity		>5.7	90 (114/126) [84, 95]	91 (43/47) [80, 98]	97 (114/118) [92, 99]	78 (43/55) [68, 86]
Threshold for 90% specificity		>5.7	90 (114/126) [84, 95]	91 (43/47) [80, 98]	97 (114/118) [92, 99]	78 (43/55) [68, 86]
MRI PDFF ≥15%	0.96 [0.92, 0.99]					
Optimal threshold†		>14.1	89 (42/47) [77, 97]	91 (115/126) [85, 96]	79 (42/53) [68, 87]	96 (115/120) [91, 98]
Threshold for 90% sensitivity		>12.6	91 (43/47) [80, 97]	87 (110/126) [80, 93]	73 (43/59) [63, 81]	96 (110/114) [92, 99]
Threshold for 90% specificity		>14.1	89 (42/47) [77, 97]	91 (115/126) [85, 96]	79 (42/53) [68, 87]	96 (115/120) [91, 98]
MRI PDFF ≥25%	0.95 [0.90, 0.97]					
Optimal threshold†		>16.7	100 (10/10) [70, 100]	87 (142/163) [81, 92]	32 (10/31) [24, 42]	100 (142/142) [97, 100]
Threshold for 90% sensitivity		>16.7	100 (10/10) [70, 100]	87 (142/163) [81, 92]	32 (10/31) [24, 42]	100 (142/142) [97, 100]
Threshold for 90% specificity		>20.0	60 (6/10) [26, 88]	96 (156/163) [92, 99]	46 (6/13) [26, 68]	98 (156/160) [95, 99]

Note.—Data in brackets are 95% CIs. AUC = area under the receiver operating characteristic curve, NPV = negative predictive value, PDFF = proton density fat fraction, PPV = positive predictive value, USFF = US fat fraction.

* Except as otherwise indicated, data are percentages, with numerators and denominators in parentheses.

† Optimal threshold–indicated cutoff values that maximize the Youden index.

and < .001, respectively, for MRI PDFF ≥5%; and 0.96 [95% CI: 0.92, 0.99] vs 0.91 [95% CI: 0.86, 0.95], 0.84 [95% CI: 0.78, 0.89], and 0.83 [95% CI: 0.77, 0.89], P = .009, < .001, and .002, respectively, for MRI PDFF ≥15%). For evaluating hepatic steatosis based on MRI PDFF values greater than or equal to 5%, the USFF AUC was not different than those of TAI, TSI, and visual scoring (P = .18, .02, and .72, respectively) (Fig 5).

Discussion

In recent years, there have been several efforts for noninvasive evaluation of hepatic steatosis in nonalcoholic fatty liver disease (NAFLD). Among various techniques, quantitative US (QUS) has an advantage in that it can provide more information about liver tissue composition by using raw radiofrequency data. A deep learning approach has potential to improve the performance of QUS but has not been well investigated. Thus, in the present study, we prospectively validated the diagnostic performance of a two-dimensional convolutional neural network algorithm using B-mode images and QUS parametric maps generated from US radiofrequency data for assessment of hepatic steatosis in patients with NAFLD, with the MRI proton density fat fraction (PDFF) as the reference standard. Our study results demonstrated that the deep learning–estimated US fat fraction (USFF) showed a strong correlation with the MRI PDFF (Pearson, r = 0.86; P < .001) and demonstrated excellent diagnostic performance for various stages of hepatic steatosis, from an MRI PDFF greater than or equal to 5% to an MRI PDFF greater than or equal to 25% (areas under the receiver operating characteristic curve, 0.95–0.97). Moreover, the diagnostic performance of the algorithm for hepatic steatosis (MRI PDFF ≥5%) was excellent and significantly higher than that of the QUS parameters (tissue attenuation imaging and tissue scatter-distribution imaging) and that of the subjective visual score (0.97 vs 0.92, 0.91, and 0.84; P = .015, .006, and < .001, respectively). Our study has demonstrated that the application of a deep learning algorithm can improve the performance of QUS parameters for diagnosing hepatic steatosis. Given the excellent performance of our algorithm, the USFF holds promise as a noninvasive and accurate screening tool for hepatic fat quantification in patients with NAFLD.

In our study, the USFF yielded high performance for diagnosis and staging of hepatic steatosis, comparable to or better than that reported in previous studies that used B-mode images (18,32) or radiofrequency signals (19) as input data in deep learning algorithms. The excellent diagnostic performance of the USFF could be attributed to the use of the QUS parametric map as the input data for the 2D CNN algorithm. QUS parametric maps derived from radiofrequency data analysis can provide more information about liver tissue composition, which can be lost during B-mode image generation (17). Furthermore, as radiofrequency data are not affected by dynamic range and filtering settings (12), using the QUS parametric map as the input data of the 2D CNN algorithm may be valuable

Table 4: Comparison of Diagnostic Performance Between USFF, QUS Parameters, and Visual Score for Hepatic Steatosis

Hepatic Fat Content	AUC	Cutoff	Sensitivity*	Specificity*	PPV*	NPV*
USFF (%)						
MRI PDFF ≥5%	0.97 [0.93, 0.99]	>5.7	90 (114/126) [84, 95]	91 (43/47) [80, 98]	97 (114/118) [92, 99]	78 (43/55) [68, 86]
MRI PDFF ≥15%	0.96 [0.92, 0.99]	>14.1	89 (42/47) [77, 97]	91 (115/126) [85, 96]	79 (42/53) [68, 87]	96 (115/120) [91, 98]
MRI PDFF ≥25%	0.95 [0.90, 0.97]	>16.7	100 (10/10) [70, 100]	87 (142/163) [81, 92]	32 (10/31) [24, 42]	100 (142/142) [97, 100]
TAI (dB/cm/MHz)						
MRI PDFF ≥5%	0.92 [0.87, 0.95]	>0.72	83 (104/126) [75, 89]	91 (43/47) [80, 98]	96 (104/108) [91, 99]	66 (43/65) [57, 74]
MRI PDFF ≥15%	0.91 [0.86, 0.95]	>0.83	79 (37/47) [64, 89]	91 (115/126) [85, 96]	77 (37/48) [65, 86]	92 (115/125) [87, 95]
MRI PDFF ≥25%	0.90 [0.84, 0.94]	>0.86	100 (10/10) [69, 100]	80 (130/163) [73, 86]	23 (10/43) [18, 29]	100 (130/130) [97, 100]
TSI						
MRI PDFF ≥5%	0.91 [0.85, 0.94]	>95.6	83 (105/126) [76, 89]	79 (37/47) [64, 89]	91 (105/115) [84, 95]	64 (37/58) [54, 73]
MRI PDFF ≥15%	0.84 [0.78, 0.89]	>98.4	94 (44/47) [90, 99]	64 (81/126) [55, 73]	49 (44/89) [43, 56]	96 (81/84) [90, 99]
MRI PDFF ≥25%	0.81 [0.75, 0.87]	>98.9	100 (10/10) [69, 100]	54 (88/163) [46, 62]	12 (10/85) [10, 14]	100 (88/88) [96, 100]
Visual score						
MRI PDFF ≥5%	0.84 [0.77, 0.89]	≥1 (mild)	87 (109/126) [79, 92]	81 (38/47) [67, 91]	92 (109/118) [87, 96]	69 (38/55) [59, 78]
MRI PDFF ≥15%	0.83 [0.77, 0.89]	≥2 (moderate)	100 (47/47) [93, 100]	67 (84/126) [58, 75]	53 (47/89) [47, 59]	100 (84/84) [80, 100]
MRI PDFF ≥25%	0.91 [0.85, 0.95]	≥3 (severe)	90 (9/10) [56, 100]	91 (149/163) [86, 95]	39 (9/23) [27, 53]	99 (149/150) [96, 100]

Note.—Data in brackets are 95% CIs. Cutoff values for the deep learning–based USFF and the QUS parameters (TAI and TSI) were determined using the maximal Youden index. AUC = area under the receiver operating characteristic curve, NPV = negative predictive value, PDFF = proton density fat fraction, PPV = positive predictive value, QUS = quantitative US, TAI = tissue attenuation imaging, TSI = tissue scatter-distribution imaging, USFF = US fat fraction.

* Except as otherwise indicated, data are percentages, with numerators and denominators in parentheses.

for developing robust algorithms for assessing hepatic steatosis. Although B-mode images can be affected by dynamic range and filtering settings, the magnitude of effect of those technical confounders could be minimized by data acquisition using a fixed set of time-gain compensation and focus position. As each QUS parameter and B-mode image depicts different tissue characteristics (ie, TAI, TSI, and B-mode images indicate attenuation, scattering distribution of signals, and brightness, respectively), the combined interpretation of those QUS parameters and B-mode images may provide more comprehensive tissue composition information. Of note, the use of a 2D CNN in our study enabled fine-tuning of large input data, achieving higher performance for diagnosis of hepatic steatosis.

Our study demonstrated significantly higher performance of the USFF than the QUS parameters or subjective visual score in evaluating hepatic steatosis based on MRI PDFF values greater than or equal to 5% and greater than or equal to 15%. Applying the 2D CNN algorithm and QUS parametric maps may overcome the modest performance of conventional US, which is commonly used as a first-line tool for assessing hepatic steatosis. Of note, the higher sensitivity of the USFF than that of the visual score may be more helpful in screening patients with clinically suspected NAFLD. Additionally, as the results can be presented as continuous quantitative values, the USFF can be helpful for longitudinal follow-up or treatment monitoring, although the reproducibility of the results should be further investigated.

We demonstrated that the USFF correlated strongly and linearly with MRI PDFF values less than or equal to 18%. However, for MRI PDFF values greater than 18%, a possible saturation effect was observed, which was consistent with the findings of Han et al (12). Indeed, previous studies also reported that QUS parameters were saturated in severe hepatic steatosis (12,13,33). Saturation of QUS parameters can be explained by the inherent limitations of US signals, which become weak and insensitive in severe hepatic steatosis, as it is already severely attenuated in the nearfield. Considering the saturation effects in severe hepatic steatosis, the USFF should be cautiously interpreted in the range of MRI PDFF values greater than 18%, and the application of the parameter itself is limited, although it may be directly applicable for MRI PDFF values less than or equal

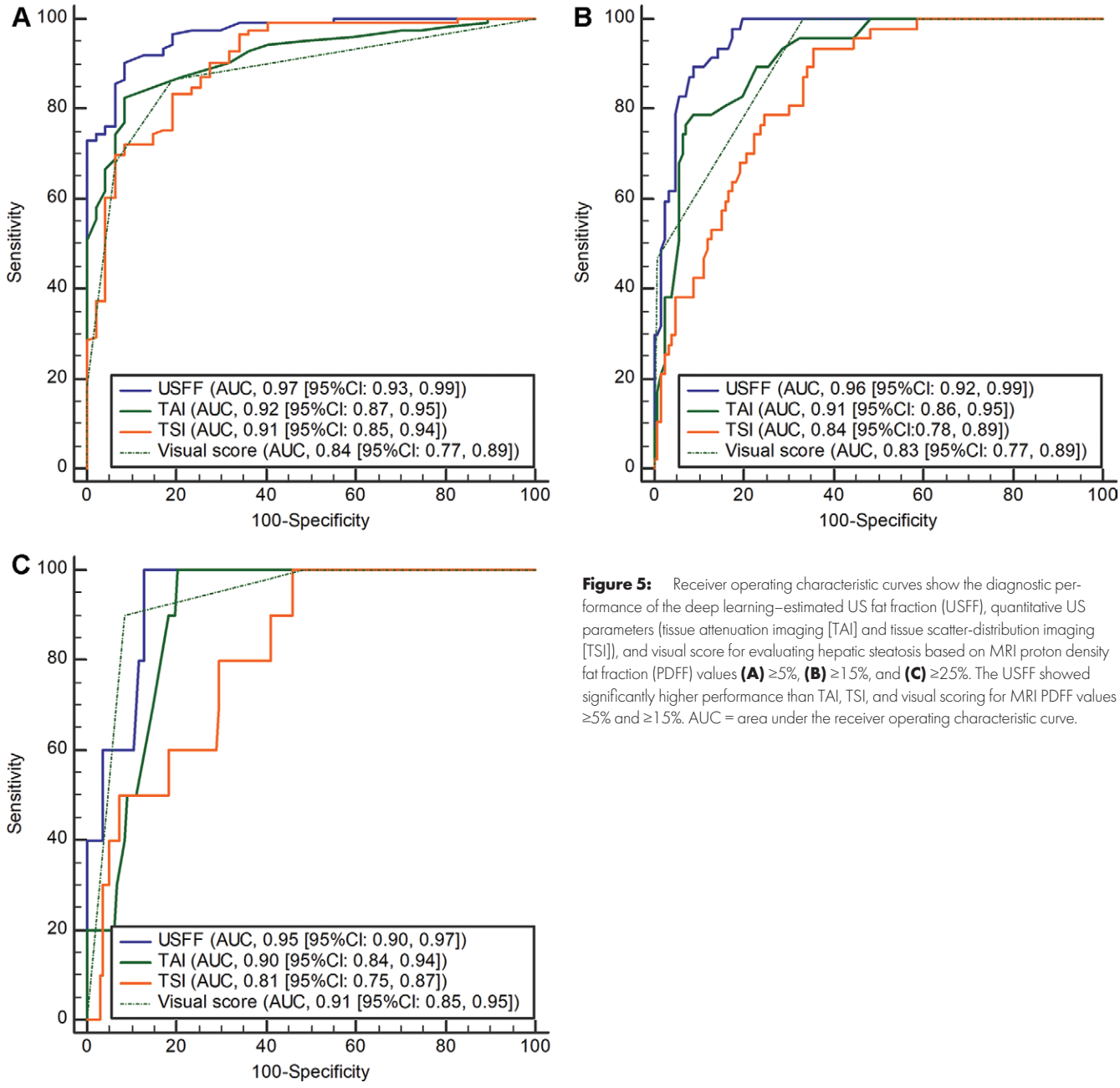

Figure 5: Receiver operating characteristic curves show the diagnostic performance of the deep learning–estimated US fat fraction (USFF), quantitative US parameters (tissue attenuation imaging [TAI] and tissue scatter-distribution imaging [TSI]), and visual score for evaluating hepatic steatosis based on MRI proton density fat fraction (PDFF) values **(A)** ≥5%, **(B)** ≥15%, and **(C)** ≥25%. The USFF showed significantly higher performance than TAI, TSI, and visual scoring for MRI PDFF values ≥5% and ≥15%. AUC = area under the receiver operating characteristic curve.

to 18%. In addition, efforts to minimize the saturation effect by applying a multiparametric approach, including speed of sound, dispersion, or texture analysis, would be valuable, and further studies are warranted.

Our study had several limitations. First, because all US data were obtained from a single scanner, further studies for assessing cross-platform generalizability are needed. Second, currently, radiofrequency data acquisition is not readily available in all clinical US systems; however, with most manufacturers beginning to provide radiofrequency acquisition capabilities, this may be widely available in the near future. Third, our study evaluated the performance of a deep learning algorithm that used both QUS parametric maps and B-mode images as input data, and the performance of the algorithm using B-mode images alone was not evaluated. Fourth, potential confounders, such as inflammation or fibrosis, could not be assessed. Fifth, the effect of body mass index and skin-to-liver capsule distance on the performance of the algorithm was not investigated, and further analysis with various ranges of body mass indexes and skin-to-liver capsule distances are needed. Finally, the reported sensitivity and specificity of the USFF in detecting various degrees of hepatic steatosis are likely to be overestimated as they were determined from data using a cutoff that was optimized for the same data. Further external validation of our study is needed in terms of the performance and reproducibility of the deep learning algorithm.

In conclusion, a deep learning algorithm that used quantitative US parametric maps and B-mode images could accurately

estimate the hepatic fat fraction and diagnose hepatic steatosis in nonalcoholic fatty liver disease. Further studies using histologic results as a reference standard are warranted to assess potential confounders, such as inflammation or fibrosis.

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

Data sharing: Data generated or analyzed during the study are available from the corresponding author by request.

Disclosures of conflicts of interest: **S.K.J.** No relevant relationships. **J.M.L.** Editorial board, *Radiology*; grants from Bayer Healthcare, Canon Healthcare, Philips Healthcare, GE Healthcare, Central Medical Servis, Guerbet, Samsung Medison, Bracco, Siemens Healthcare, and Dongkuk Pharma; consulting fees from Samsung Medison; lecture payment from Bayer Healthcare, Siemens Healthineers, Samsung Medison, Guerbet, Philips, GE Healthcare, and Clarify. **I.J.** No relevant relationships. **J.H.Y.** Honoraria from Bayer, Philips Healthcare, and Accuzen. **G.L.** No relevant relationships.

References

1. 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 2016;64(1):73–84.
2. Huang DQ, El-Serag HB, Loomba R. Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2021;18(4):223–238.
3. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. Nat Med 2018;24(7):908–922.
4. Starekova J, Hernando D, Pickhardt PJ, Reeder SB. Quantification of Liver Fat Content with CT and MRI: State of the Art. Radiology 2021;301(2):250–262.
5. Puchner SB, Lu MT, Mayrhofer T, et al. High-risk coronary plaque at coronary CT angiography is associated with nonalcoholic fatty liver disease, independent of coronary plaque and stenosis burden: results from the ROMICAT II trial. Radiology 2015;274(3):693–701.
6. Wu S, Wu F, Ding Y, Hou J, Bi J, Zhang Z. Association of non-alcoholic fatty liver disease with major adverse cardiovascular events: A systematic review and meta-analysis. Sci Rep 2016;6(1):33386.
7. Valenti L, Pelusi S. Redefining fatty liver disease classification in 2020. Liver Int 2020;40(5):1016–1017.
8. Machado MV, Cortez-Pinto H. Non-invasive diagnosis of non-alcoholic fatty liver disease. A critical appraisal. J Hepatol 2013;58(5):1007–1019.
9. Tang A, Desai A, Hamilton G, et al. Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. Radiology 2015;274(2):416–425.
10. Dasarathy S, Dasarathy J, Khiyami A, Joseph R, Lopez R, McCullough AJ. Validity of real time ultrasound in the diagnosis of hepatic steatosis: a prospective study. J Hepatol 2009;51(6):1061–1067.
11. Paige JS, Bernstein GS, Heba E, et al. A Pilot Comparative Study of Quantitative Ultrasound, Conventional Ultrasound, and MRI for Predicting Histology-Determined Steatosis Grade in Adult Nonalcoholic Fatty Liver Disease. AJR Am J Roentgenol 2017;208(5):W168–W177.
12. Han A, Zhang YN, Boehringer AS, et al. Assessment of Hepatic Steatosis in Nonalcoholic Fatty Liver Disease by Using Quantitative US. Radiology 2020;295(1):106–113.
13. Jeon SK, Lee JM, Joo I, Park SJ. Quantitative Ultrasound Radiofrequency Data Analysis for the Assessment of Hepatic Steatosis in Nonalcoholic Fatty Liver Disease Using Magnetic Resonance Imaging Proton Density Fat Fraction as the Reference Standard. Korean J Radiol 2021;22(7):1077–1086.
14. Oelze ML, Mamou J. Review of Quantitative Ultrasound: Envelope Statistics and Backscatter Coefficient Imaging and Contributions to Diagnostic Ultrasound. IEEE Trans Ultrason Ferroelectr Freq Control 2016;63(2):336–351.
15. Wan YL, Tai DI, Ma HY, Chiang BH, Chen CK, Tsui PH. Effects of fatty infiltration in human livers on the backscattered statistics of ultrasound imaging. Proc Inst Mech Eng H 2015;229(6):419–428.
16. Lin SC, Heba E, Wolfson T, et al. Noninvasive Diagnosis of Nonalcoholic Fatty Liver Disease and Quantification of Liver Fat Using a New Quantitative Ultrasound Technique. Clin Gastroenterol Hepatol 2015;13(7):1337–1345.e6.
17. Sugimoto K, Lee DH, Lee JY, et al. Multiparametric US for Identifying Patients with High-Risk NASH: A Derivation and Validation Study. Radiology 2021;301(3):625–634.
18. Byra M, Styczynski G, Szmigielski C, et al. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. Int J CARS 2018;13(12):1895–1903.
19. Han A, Byra M, Heba E, et al. Noninvasive Diagnosis of Nonalcoholic Fatty Liver Disease and Quantification of Liver Fat with Radiofrequency Ultrasound Data Using One-dimensional Convolutional Neural Networks. Radiology 2020;295(2):342–350.
20. Wu Y, Yang F, Liu Y, Zha X, Yuan S. A Comparison of 1-D and 2-D Deep Convolutional Neural Networks in ECG Classification. arXiv preprint arXiv:1810.07088. https://arxiv.org/abs/1810.07088. Posted October 16, 2018. Accessed January 5, 2022.
21. Hamaguchi M, Kojima T, Itoh Y, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. Am J Gastroenterol 2007;102(12):2708–2715.
22. Kim H, Varghese T. Attenuation estimation using spectral cross-correlation. IEEE Trans Ultrason Ferroelectr Freq Control 2007;54(3):510–519.
23. Liao YY, Yang KC, Lee MJ, Huang KC, Chen JD, Yeh CK. Multifeature analysis of an ultrasound quantitative diagnostic index for classifying nonalcoholic fatty liver disease. Sci Rep 2016;6(1):35083.
24. Mohana Shankar P. A general statistical model for ultrasonic backscattering from tissues. IEEE Trans Ultrason Ferroelectr Freq Control 2000;47(3):727–736.
25. Ho MC, Lee YH, Jeng YM, Chen CN, Chang KJ, Tsui PH. Relationship between ultrasound backscattered statistics and the concentration of fatty droplets in livers: an animal study. PLoS One 2013;8(5):e63543.
26. Ferraioli G, Kumar V, Ozturk A, Nam K, de Korte CL, Barr RG. US Attenuation for Liver Fat Quantification: An AIUM-RSNA QIBA Pulse-Echo Quantitative Ultrasound Initiative. Radiology 2022;302(3):495–506.
27. Tang A, Tan J, Sun M, et al. Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. Radiology 2013;267(2):422–431.
28. Jung J, Han A, Madamba E, et al. Direct Comparison of Quantitative US versus Controlled Attenuation Parameter for Liver Fat Assessment Using MRI Proton Density Fat Fraction as the Reference Standard in Patients Suspected of Having NAFLD. Radiology 2022;304(1):75–82.
29. Raunig DL, McShane LM, Pennello G, et al; QIBA Technical Performance Working Group. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 2015;24(1):27–67.
30. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3(1):32–35.
31. Dixon WJ. Simplified Estimation from Censored Normal Samples. Ann Math Stat 1960;31(2):385–391, 387.
32. Reddy DS, Bharath R, Rajalakshmi P. A Novel Computer-Aided Diagnosis Framework Using Deep Learning for Classification of Fatty Liver Disease in Ultrasound Imaging. In: 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom). IEEE, 2018; 1–5. https://ieeexplore.ieee.org/document/8531118.
33. Jeon SK, Joo I, Kim SY, et al. Quantitative ultrasound radiofrequency data analysis for the assessment of hepatic steatosis using the controlled attenuation parameter as a reference standard. Ultrasonography 2021;40(1):136–146.