Received: 3 April 2024 | Revised: 15 July 2024 | Accepted: 11 August 2024

DOI: 10.1111/liv.16078

Liver INTERNATIONAL WILEY

ORIGINAL ARTICLE

An ultrasound multiparametric method to quantify liver fat using magnetic resonance as standard reference

Laura De Rosa[1,2] | Antonio Salvati[3] | Nicola Martini[4] | Dante Chiappino[4] | Simone Cappelli[5] | Marcello Mancini[6] | Libertario Demi[2] | Lorenzo Ghiadoni[7] | Ferruccio Bonino[6] | Maurizia R. Brunetto[3,5,6] | Francesco Faita[1]

[1]Institute of Clinical Physiology, National Research Council, Pisa, Italy

[2]Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

[3]Hepatology Unit, Pisa University Hospital, Pisa, Italy

[4]Fondazione Toscana Gabriele Monasterio, Massa, Italy

[5]Department of Clinical and Experimental Medicine, Pisa University, Pisa, Italy

[6]Institute of Biostructure and Bioimaging, National Research Council, Naples, Italy

[7]Emergency Medicine Unit, Department of Clinical and Experimental Medicine, University Hospital of Pisa, Pisa, Italy

Correspondence
Maurizia Rossana Brunetto, Department of Clinical and Experimental Medicine, Via Savi 10, 56100, Pisa, Italy.
Email: maurizia.brunetto@unipi.it

Abstract

Background & Aims: There is an unmet need for a reliable and reproducible non-invasive measure of fatty liver content (FLC) for monitoring steatotic liver disease in clinical practice. Sonographic FLC assessment is qualitative and operator-dependent, and the dynamic quantification range of algorithms based on a single ultrasound (US) parameter is unsatisfactory.

This study aims to develop and validate a new multiparametric algorithm based on B-mode images to quantify FLC using Magnetic Resonance (MR) values as standard reference.

Methods: Patients with elevated liver enzymes and/or bright liver at US ($N=195$) underwent FLC evaluation by MR and by US. Five US-derived quantitative features [attenuation rate(AR), hepatic renal-ratio(HR), diaphragm visualization(DV), hepatic-portal-vein-ratio(HPV), portal-vein-wall(PVW)] were combined by mixed linear/exponential regression in a multiparametric model (Steatoscore2.0). One hundred and thirty-four subjects were used for training and 61 for independent validations; score-computation underwent an inter-operator reproducibility analysis.

Results: The model is based on a mixed linear/exponential combination of 3 US parameters (AR, HR, DV), modelled by 2 equations according to AR values. The computation of FLC by Steatoscore2.0 (mean ± std, 7.91% ± 8.69) and MR (mean ± std, 8.10% ± 10.31) is highly correlated with a low root mean square error in both training/validation cohorts, respectively ($R=0.92/0.86$ and RMSE = 5.15/4.62, $p<.001$). Steatoscore2.0 identified patients with MR-FLC≥5%/≥10% with sensitivity = 93.2%/89.4%, specificity = 86.1%/95.8%, AUROC = 0.958/0.975, respectively and correlated with MR ($R=0.92$) significantly ($p<.001$) better than CAP ($R=0.73$).

Abbreviations: AE, absolute error; AR, attenuation rate; AUROC, area under the curve of the receiver operating characteristic; BAA, Bland–Altman analysis; CAP, controlled attenuation parameter; CT, computed tomography; DV, diaphragm visualization; FLC, fatty liver content; GUI, graphical user interface; HPV, hepatic portal vein-ratio; HR, hepatic renal ratio; IC, interval of confidence; ICC, intraclass correlation coefficient; MASLD, metabolic dysfunction associated steatotic liver disease; MR, magnetic resonance; MRS, magnetic resonance spectroscopy; NPV, negative predictive value; PDFF, proton density fat fraction; PPV, positive predictive value; PVW, portal vein wall; QUS, quantitative ultrasound; RF, radiofrequency; RMSE, root mean square error; ROC, receiver operating characteristic; ROI, region of interest; SD, standard deviation; SLD, steatotic liver disease; TGC, time-gain compensation; US, ultrasound.

Conclusions: Multiparametric Steatoscore2.0 measures FLC providing values highly comparable with MR. It is reliable, inexpensive, easy to use with any US equipment and qualifies to be tested in larger, prospective studies as new tool for the non-invasive screening and monitoring of FLC.

KEYWORDS
fatty liver, MASLD, multiparametric model, quantitative ultrasound, Steatotic liver disease

1 | INTRODUCTION

In recent years, metabolic dysfunction associated steatotic liver disease (MASLD) has become the most common cause of chronic liver disease in Western countries and it is predicted to become also the most frequent indication for liver transplantation by 2030.[1] The global prevalence of MASLD differs among geographical areas: United States–30%, Middle East–32%, South America–30%, Asia–27%, Europe–24% and Africa–13%, and is influenced by the different lifestyles and genetic factors.[2] MASLD is closely associated with obesity, diabetes, insulin resistance, dyslipidemia and hypertension and is recognized as the hepatic manifestation of the metabolic syndrome. In addition, steatotic liver disease (SLD) is a comorbidity in many patients with chronic viral or autoimmune liver disease and major pathology in alcohol and many drug-induced liver diseases.[1] Thus, the early measurement of liver fat is of paramount clinical relevance for patients' management and decision-making; however, its evaluation by histology is limited because liver biopsy[3] is an invasive procedure that is unsuitable for routine monitoring and burdened by sampling error in cases of inhomogeneous distribution of intrahepatic fat.[4] Therefore, in recent years, much attention has been paid to imaging techniques to non-invasively quantify liver steatosis, including ultrasound (US), computerized tomography (CT)[5] and magnetic resonance (MR). Proton magnetic resonance spectroscopy (^{1}H-MRS) and proton density fat fraction (MRI-PDFF) are considered gold standard techniques,[6] but they are time consuming, expensive and not suitable for high throughput screening. CT is an ionizing technique, and, in addition, it demonstrated limited accuracy in steatosis detection.[7] US is the most used technique: it is widely available, low cost and well tolerated by patients; however, the diagnosis of fatty liver content (FLC) by US is still considered operator-dependent with poor sensitivity and unsuitable for quantification.[8] In fact, the degree of fat infiltration of the liver can be qualitatively estimated by grading some US features that include liver brightness, attenuation of the US signal across the liver parenchyma, contrast ratio between liver and kidney parenchyma, visualization of intrahepatic vessels and diaphragm.[9] Since the assessment of the features by sonographers is qualitative and operator-dependent, quantitative US methods were proposed for the quantification of hepatic steatosis using a single-US parameter associated with fat accumulation.[10] Controlled Attenuation Parameter (CAP, Fibroscan®, Echosens, Paris, France), continuous CAP (cCAP)[11] and quantitative ultrasound (QUS)[12] can be used to measure FLC, but both require the analysis of radiofrequency (RF) signals. Several different CAP cut-offs have been proposed for grading the steatosis severity classes suggesting the high variability of the technology. Furthermore, CAP has been shown to efficiently recognize low (<10%) and high (>33%) grades of steatosis with AUROC values higher than 0.90, but it is less efficient for intermediate degrees of steatosis.[13]

To overcome the limitations of the mono-parametric approach, few quantification systems have been developed combining the analysis of multiple US parameters[14,15] extracted from both B-mode and raw RF signals. Recently, even the World Federation for Ultrasound in Medicine and Biology guidelines promoted the development of multiparametric US for quantifying liver fat content, by including attenuation, speed-of-sound and shear wave dispersion, that are still extracted from raw RF data.[16] Accordingly, a multiparametric system based on ultrasound imaging that could be used both online and offline without the need of accessing the RF data would be extremely helpful. We recently proposed a quantitative multiparametric score (Steatoscore) representing the percentage of FLC obtained by a combination of five US parameters computed from B-mode US images using MR-derived measurements of FLC as gold standard.[17] Steatoscore was trained on 61 patients enrolled from two different clinical centres achieving a correlation of 0.84 and a root mean square error (RMSE) of 6.41. In order to give more robustness to the prediction model of FLC and to overcome some limitations of the previous model, in the current study, we present the validation of the updated version of the Steatoscore (Steatoscore2.0) which consists in a new model trained on a significantly larger cohort of subjects with MR fat measurement acquired in a single clinical centre.

HIGHLIGHTS

- There is an unmet need for non-invasive liver fat content measurement.
- Steatoscore2.0 exhibited excellent agreement with MR liver fat measurement.
- US multiparametric approach improves diagnostic accuracy in detecting steatotic liver disease.
- Steatoscore2.0 is inexpensive, easy to use and can be applied on any US equipment.

2 | MATERIALS AND METHODS

2.1 | Patients

The study enrolled 195 patients; 167 of them admitted at the Hepatology Unit, Pisa University Hospital, Italy, for liver disease clinical evaluation because of elevated liver enzymes (transaminases and/or gamma-glutamyl-transferase) and/or bright liver at US. Inclusion criteria were as follows: age >18 years old and not contraindications at underwent MRI examinations. Patients with clinical, instrumental or biochemical signs of cirrhosis were excluded from the study. Subjects were consecutively enrolled, and US examinations were performed following a specific acquisition protocol to allow image analysis for parameter calculation (previously reported in[17]). In all the patients US and MRI examinations were performed within 1 week a part. Data from 134 patients (males:females 72:62, age: 52 ± 13 years old) were used for the model generation (training cohort) and data of 33 subjects (males: females 18:15, age: 56 ± 13 years old) were used to validate the model (validation cohort). A second external cohort of subjects ($N = 28$ subjects, males:females 14:14, age: 49 ± 13 years old) has been used as external and independent cohort to further validate the model. Data on this second cohort have been acquired at IRCCS SDN Foundation of Naples and it was composed of consecutive individuals undergoing a hepatology visit, none of them with already diagnosed metabolic syndrome or elevated serum iron and/or ferritin levels.

2.2 | Ethics statement

The protocol was approved by the Ethics Committee of the University Hospital of Pisa (19/02/2016) and by the independent ethics committee of the IRCCS SDN Foundation (18/04/2013), for the population of Pisa and Naples, respectively. All the subjects signed a written informed consent.

2.3 | ^{1}H-MRS and MRI-PDFF liver fat content evaluation

MR imaging was performed with two MR scanners (Pisa: Philips Ingenia 3.0 T, Philips Healthcare, Best, The Netherlands; Naples: Philips Achieva 1.5 T 3.0 T, Philips Healthcare, Best, The Netherlands). We used two commonly adopted MR-based quantification techniques, ^{1}H-MRS and MRI-PDFF, for FLC evaluation. Details of the two MR techniques used for FLC quantification are described in the following.

2.3.1 | ^{1}H-MRS

Single-voxel magnetic resonance spectroscopy (^{1}H-MRS) was acquired with a point-resolved spectroscopy sequence (PRESS) with the same parameters described in.[17] Spectral analysis was done offline using the jMRUI software package.[18] Time-domain spectral fitting of the water peak at 4.7 ppm and of the lipid resonance (at 0.9, 1.3 and 2.1 ppm) was performed using the AMARES method.[19] ^{1}H-MRS values were assessed for each subject, normalizing the fitted signal amplitude of the fat to the sum of water and fat amplitudes.[20]

2.3.2 | MRI-PDFF

Proton density fat fraction (MRI-PDFF) measurements were acquired using a multi-echo gradient-echo MRI sequence (FOV 440 × 400 mm, flip angle 3°, number of echoes 12, TR 15 ms, first TE 1.1 ms, echo spacing 1.1 ms). PDFF maps were reconstructed offline using complex-based fitting of the source multi-echo images with a multicomponent water-fat model.[20] PDFF values were recorded for each region of interest (ROI)/segment, and a final right-lobe MRI-PDFF value for each participant was obtained by averaging the values of the four corresponding ROIs.

2.4 | CAP evaluations

CAP measurement by Fibroscan® 502 (Echosens, Paris, France) was obtained in a subgroup of 121 patients (all these patients have been enrolled in Pisa); M/XL probe was used when BMI </≥30 kg/m^2 or if skin-to-capsule distance is </>25 mm according to the EASL-ALEH Clinical Practice Guidelines.[21]

2.5 | US–acquisition

All US image acquisitions were performed with standard diagnostic US systems (Pisa: LogiQ E9, GE Healthcare, Bucking-Hamshire, UK and Naples: Philips iU22, Philips Healthcare, Bothell, WA, USA) both equipped with a 1.8–5 MHz convex probe. All acquisitions were performed by a single experienced operator (one for each centre) blinded to MRI fat quantifications and with patients in the supine position in a temperature-controlled room (22°C–24°C). The images were acquired according to a defined protocol, that prescribe the acquisition of an intercostal or subcostal longitudinal scan view with subject in supine/left lateral position, an oblique subcostal scan view and an acquisition of US images from a subcostal longitudinal view that is modified to correctly visualize a portion of the portal vein in the centre of the liver. All the images were acquired in breath-hold with the subject at maximal inspiration and maintaining the Time Gain Compensation (TGC) at a fixed value. All the US scans were then exported and saved on an external memory in order to be processed on a personal computer.

2.6 | US–images processing

Five parameters were obtained by processing the US images. The theoretical basis of these parameters has been widely described in[17] and they are summarized in the following.

Briefly:

- Attenuation Rate (AR) was calculated from the average of several grey level profiles (from 10 to 50 for each frame of the clip) along the US beam fitted by a decreasing exponential curve.[22] These profiles were extracted considering a homogenous portion of the hepatic parenchyma avoiding as much as possible vessels or other structures.
- Hepatic-renal ratio (HR) represents the intensity ratio between the mean grey levels of a ROI placed within the liver parenchyma and of a second ROI within the renal cortex.
- Diaphragm visualization (DV) represents the mean peak intensity of five grey-level profiles of the diaphragm line normalized for both the overall gain of the B-mode scan and the depth at which the diaphragm is located.
- Portal vein wall (PVW) represents the ratio of the mean grey levels of the hepatic parenchyma and of a ROI placed within the vessel wall of the portal vein.
- Hepatic portal vein ratio (HPV) represents the ratio of the mean grey levels of the hepatic parenchyma and of a ROI placed inside the portal vein.

Note that for all subjects in the study cohort, evaluation of single parameters and of the multiparametric score was successfully obtained. The US image processing algorithms for the evaluation of the US parameters were implemented on a graphical user interface (GUI) developed in MATLAB R2022b (The MathWorks Inc., Natick, MA, USA).

2.7 | US parameters variability

The US parameters were computed by two different operators on the whole dataset. The variability analysis was performed to study the repeatability of the computation of each parameter and of the Steatoscore2.0.

2.8 | Multivariable quantitative US models

Two classes of mathematical models (multiple linear and multiple exponential regression models) have been trained to build the predictive multiparametric score (Steatoscore2.0) for the assessment of FLC combining five[5] US parameters and using MR imaging measurements as the gold standard.

All models were trained both considering the full training database, as well as splitting the training database according to the sign of the AR parameter. This choice was based on the physical meaning of the AR parameter, i.e. attenuation or amplification of the ultrasound beam when AR>0 or AR≤0, respectively. In fact, a liver parenchyma that attenuates the US beam (positive AR) is generally characterized by a higher fat content than a tissue that does not show attenuation (AR negative), as the fatty tissue exhibits higher attenuation coefficients than other tissues.[23] The best model to predict the FLC was obtained following a 3 steps procedure:

1. Selection of the US parameters that linearly and/or exponentially correlate with a coefficient higher than 0.6 against the fat per cent value derived by magnetic resonance imaging;
2. Generation of the regression models:
   (i) Training on the main dataset of a fully linear, a fully exponential and mixed (linear and exponential) multiparametric models. In this latter case, US features were included in the model as linear or exponential component according to the minimum value of root mean square error (RMSE) of the linear/exponential mono-parametric regression;
   (ii) Iteration of the procedure described in steps 2a using the split dataset;
   (iii) Selecting the best combination of two models derived at step 2b on the two split datasets. The final model was obtained combining models with lowest RMSE;
3. Comparing all trained models and selecting the model (Steatoscore2.0) that showed the best RMSE on the full dataset. Training was performed using analytic forms of models' equations and the non-linear fit function Matlab command *nlinfit*. The models' coefficients were obtained iteratively with the minimization technique of the least square equation; RMSE, the coefficient of correlation (R) and the coefficient of determination (R^2) between models' prediction and MR-derived fat content percentage have been evaluated for each model.

Once the model was parametrized, we assessed the prediction performance on the training/test set (training dataset, $N=134$ subjects). Finally, we performed the validation of that model evaluating Steatoscore2.0 on the independent dataset (validation dataset, $N=33$ subjects) and a further validation on an external and independent dataset ($N=28$ subjects).

2.9 | Statistical analysis

Descriptive statistics were computed for all variables and were expressed as percentages/counts for categorical variables and mean ± standard deviation (SD) for continuous variables. Correlations between variables were examined using Pearson's correlation coefficients. The Bland–Altman Analysis (BAA) was performed to evaluate the agreement between prediction fat content by the model and the gold standard. In particular, limits of agreement (LoA) and bias were assessed. Absolute errors (AEs) were assessed between the MR-derived fat per cent and the estimation of fat content value by the models; the Student's t-test was evaluated between distributions of AE values to statically compare different models. Receiver operator characteristic (ROC) curves were used to assess specificity and sensitivity of the model in discriminating patients with MRI fat values higher than the clinically accepted cut-off of 5% to discriminate health subjects from those with steatosis and of 10% to classify

absent/mild from moderate/severe level of steatosis and the correspondent areas under the receiver operating characteristic curve (AUROC) were evaluated. Moreover, diagnostic accuracy, positive predictive value (PPV) and negative predictive value (NPV) were assessed by 2×2 contingency tables. The Intraclass Correlation Coefficient (ICC) and BAA were performed to evaluate the inter-operator variability. Correlation analysis and BAA were also performed to validate the model on an independent (validation dataset) and external (external validation dataset) datasets. Correlation coefficient was also measured to assess correlation of CAP and MRI-PDFF values, and a statistical comparison between correlations of CAP and the proposed multiparametric model versus MRI-PDFF was evaluated by performing William's test for comparing dependent correlations.[24] Correlations were considered significant with $R \geq 0.300$ and $p < .05$, while all other statistical tests were considered significant at $p < .05$. All statistical analyses were performed with SPSS Version 26 (IBM, Armonk, NY, USA) or with Matlab R2022b (The MathWorks Inc., Natick, MA, USA).

3 | RESULTS

The demographic and clinical characteristics of the three cohorts (training, validation and external validation datasets) are reported in Table 1.

3.1 | US–fat content estimation model

The quantitative values of the five parameters obtained in the training cohort are reported in Table 2. Table 3 reported the results of mono-parametric models (linear and exponential) for each of five US parameters (AR, HR, DV, HPW, PVW) on the training dataset. Correlation plots between the US parameters and the MR fat content are shown in Figure S1 of Supplementary Materials with both linear and exponential fitting lines reported for each parameter. More in detail, AR, HR and DV (exponential only) showed a coefficient of correlation (R) greater than 0.6 [AR: 0.731 (linear), HR: 0.761 (linear) and DV: −0.703 (exponential), respectively]. Hepatic portal vein-ratio (HPV) and portal vein wall (PVW) showed an R lower than 0.6 (both linear and exponential). Accordingly, HPV and PVW were excluded from the analysis for the generation of the fat content prediction model, while AR, HR and DV were included.

In Table 4 (first three rows) the results of multiparametric models on the training dataset ($R = 0.83/0.81/0.85$ and RMSE = 5.66/6.00/5.34 for fully-linear/fully-exponential/mixed models, respectively) are reported. Then, we performed the split of the dataset according to AR value (≤0 and >0) in order to differentiate subjects showing attenuation from amplification. The results of mono-parametric and multiparametric models (both linear and exponential) on the two datasets derived from the split of the training dataset according to AR values are reported in Table 5. The best multiparametric model trained on the split dataset was chosen to maximize the performances in terms of RMSE minimization, by selecting among models trained on the two split datasets and by combining the two models with the minimum RMSE. Finally, the overall RMSE value was calculated by combining the RMSEs of the two models. The RMSE of the final models was equal to 5.15 and the correlation coefficient with the fat per cent values obtained by MR was equal to 0.86. The aforementioned best-performing model on the split dataset showed lower RMSE than the three multiparametric models trained on the no-split dataset (Table 4, first column), thus it was chosen as the final model (Steatoscore2.0).

TABLE 2 US quantitative parameters (mean ± standard deviation) assessed on the training database.

	Mean ± std
AR	0.007 ± 0.019
HR	1.71 ± 0.79
DV	0.15 ± 0.075
HPV	0.15 ± 0.075
PVW	3.22 ± 1.12

TABLE 1 Characteristics of the training, validation and external validation cohorts.

Characteristic	Training ($N = 134$)	Validation ($N = 33$)	External validation ($N = 28$)	Significance level ($p = .05$)
Demographic				
Age (years)	52 ± 13 (18–75)	56 ± 13 (26–75)	50 ± 13 (21–71)	n.s.
Sex (male:female)	72:62	18:15	14:14	n.s.
Physical				
Body mass Index (kg/m^2)	26.7 ± 5.10 (15.8–41.7)	27.1 ± 5.0 (15.8–40.2)	25.3 ± 4.0 (15.3–31.5)	n.s.
Imaging				
MR–Fat (%)	8.13 ± 10.25 (0.45–50.97)	7.99 ± 10.68 (0.70–39.62)	4.84 ± 7.44 (0.27–32.24)	n.s.
CAP (dB/m)	260 ± 75 (100–400) $N = 97$	249 ± 71 (141–400) $N = 24$	221 ± 60 (122–332) $N = 12$	n.s

TABLE 3 Performances of the linear and exponential mono-parametric models (one for each US parameter) trained on the training database.

	AR	HR	DV	HPV	PVW
Linear					
RMSE	6.47	6.68	8.49	9.11	8.56
R	0.754**	0.761**	−0.566	−0.058	−0.347
p	<0.001	<0.001	<0.001	n.s.	<0.001
Exponential					
RMSE	7. 02	7.69	6.88	9.03	7.83
R	0.731**	0.665**	−0.703**	−0.068	−0.505
p	<0.001	<0.001	<0.001	n.s.	<0.001

**The cells with the lower RMSE value for both linear/exponential models.

TABLE 5 RMSE values of the mono/multi-parametric models (linear and exponential) trained on the split dataset.

	AR	HR	DV	Multiparametric (AR, HR, DV)
	RMSE			
AR ≤0				
Linear	2.21	2.22	2.19	2.14
Exponential	2.17*	2.18*	2.15*	2.13*
AR >0				
Linear	7.94*	7.90*	10.21	6.59
Exponential	8.36	8.62	8.84*	7.37
Mixed	n.a.	n.a.	n.a.	6.43*

*The cells with the lower RMSE value.

TABLE 4 Comparison between the multiparametric models on the training database (fully linear, fully exponential and the best combination) and on the split database (best model).

	RMSE	R	*p*	*t*-test between AEs Model vs Steatoscore2.0
Training Dataset				
Linear	5.66	0.83	<.001	<0.001
Exponential	6.00	0.81	<.001	<0.001
Mixed	5.34	0.85	<.001	<0.001
Steatoscore2.0				
Exponential (AR≤0) +mixed (AR >0)	5.15	0.86	<.001	1

Accordingly, the Steatoscore2.0 was based on a mixed linear and exponential combination of the US parameters, modelled by two different equations depending on the value of the AR attenuation parameter and its analytical form is expressed by:

$$\text{Steatoscore2.0} = \begin{cases} 3.21 \times exp(0.13 \times HR + 4.46 \times AR - 2.27 \times DV) & AR \leq 0 \\ 13.3 \times exp(-13 \times DV) + 3.53 \times HR + 219 \times AR - 4.6 & AR > 0. \end{cases}$$

The Steatoscore2.0 model has been proved to be statistically superior to both the multiparametric fully-linear model (training dataset–linear in Table 4, $p < .001$) and a similar result was achieved comparing the purely exponential model (training dataset–exponential in Table 4, $p < .001$). Finally, statistically significant better performances ($p < .001$) were obtained by comparing the Steatoscore2.0 with a mixed linear and exponential model (training dataset–mixed in Table 4).

3.2 | Comparison of FLC measured by MR, Steatoscore2.0 and CAP

The correlation plots of MR-based fat content versus the Steatoscore2.0 values with the regression line as obtained on the training, validation and external validation datasets are shown in Figure 1A,C,E, respectively. The regression slope in the training cohort was 0.75 and the regression intercept was 2.06 (95% CI: 0.67–0.821). The Pearson correlation coefficient was $R = 0.86$ and RMSE = 5.15. Bland–Altman plot of the comparison between Steatoscore2.0 and MR in the training dataset is shown in Figure 1B. The bias was non-significant (0.0, $p = 1.0$) and the lower and upper 95% limits of agreements (LoA) were − 10% and + 10%, respectively. In the validation cohort, the coefficient of correlation between MR-based and Steatoscore2.0 fat measurements is 0.92 with a RMSE of 4.62. Bland–Altman plot of the comparison between Steatoscore2.0 and MR in the validation dataset is shown in Figure 1D. The bias was non-significant (−0.95, $p = .24$) and the lower and upper 95% LoA were − 10.0% and + 8.0%, respectively. Results on the external validation cohort showed a coefficient of correlation between MR-based and Steatoscore2.0 fat measurements of 0.83 and an RMSE of 4.17; the Bland–Altman plot showed a non-significant bias (−0.49, $p = .49$) and a 95% LoA of −7.6% and 6.6% (Figure 1F).

On the whole study population, the ROC curve discriminating subjects with MRI values ≥5% or ≥ 10% (Figure 2A,B) showed AUROC of 0.958 and 0.975, 95% interval of confidence (IC) of [0.922, 0.993] and [0.950, 0.999], sensitivity of 93.22% and 89.36% and specificity of 86.11% and 95.83%, respectively. The diagnostic accuracy of Steatoscore2.0 in identifying patients with FLC ≥5% (Table S1 of Supplementary materials) and ≥ 10% (Table S2 of Supplementary materials) at MRI was 88.62% and 94.01%, with PPV of 95.88% and 95.83% and NPV of 78.57% and 89.36%, respectively.

In the subgroup in which CAP was available, the correlation analysis between CAP and MR values (Figure 3A) showed a $R = 0.73$ which was statistically lower than the correlation between Steatoscore2.0

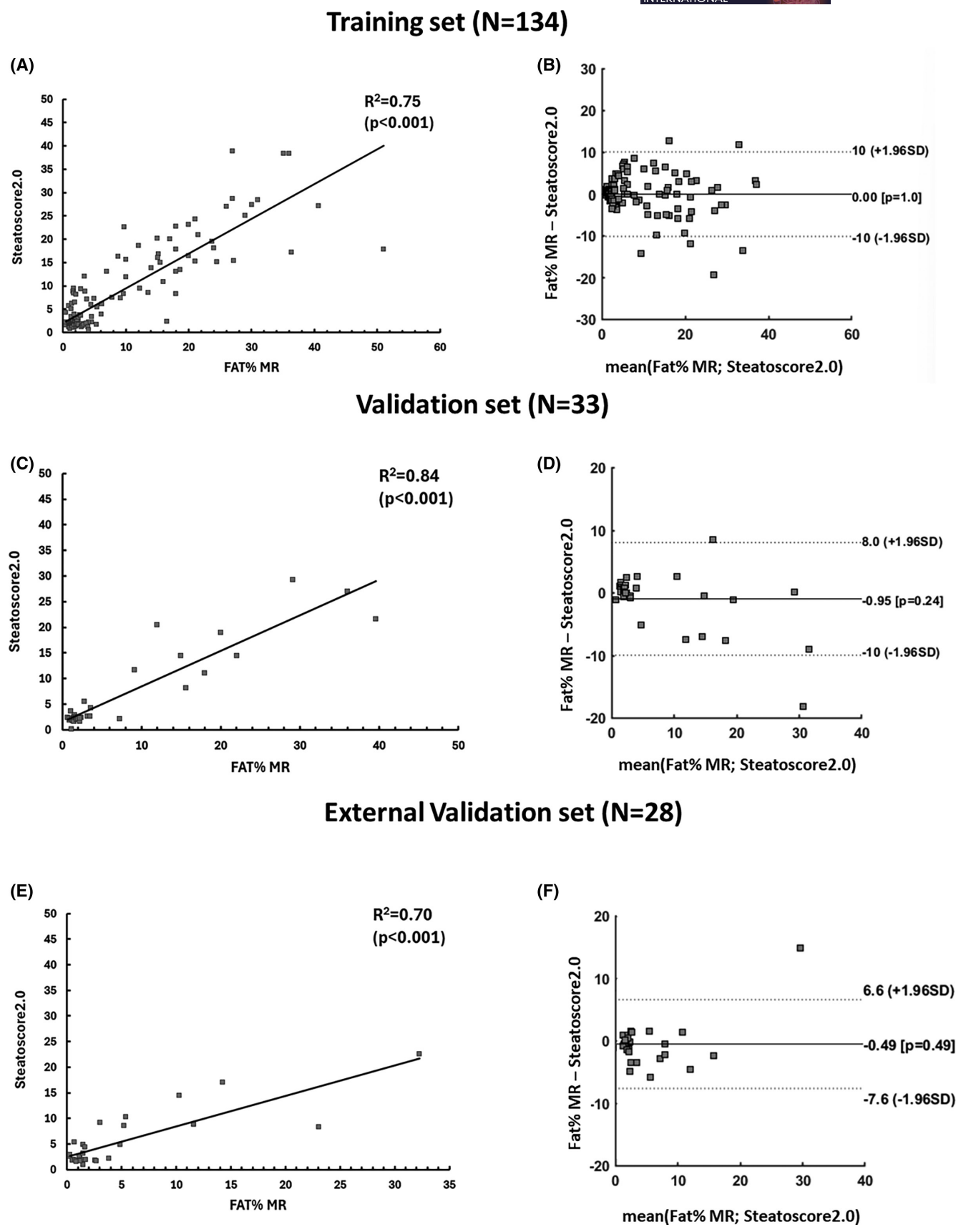

FIGURE 1 Comparisons between the fat percentage values calculated with gold standard (MR) and with Steatoscore2.0 on training ($N = 134$), validation ($N = 33$) and the external validation ($N = 28$) datasets. (A) Linear regression between Steatoscore2.0 and MR fat% on training set. (B) Bland–Altman plot (Steatoscore2.0 vs. MR fat%) on training set. (C) Linear regression between Steatoscore2.0 and MR fat% on validation set. (D) Bland–Altman plot (Steatoscore2.0 vs MR fat%) on validation set. (E) Linear regression between Steatoscore2.0 and MR fat% on the external validation set. (F) Bland–Altman plot (Steatoscore2.0 vs MR fat%) on the external validation set.

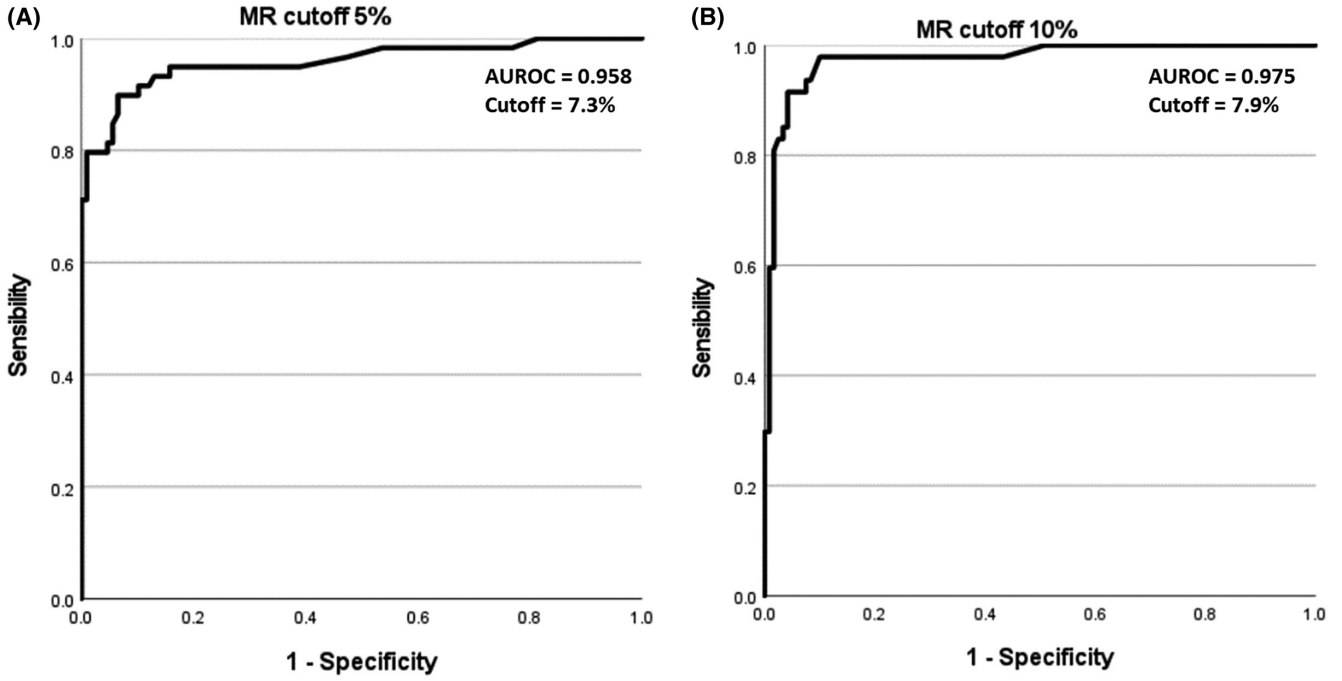

FIGURE 2 Receiver operating characteristic (ROC) curves of Steatoscore2.0 for the diagnosis of liver fat content obtained using MRI fat% >5% (A) and 10% (B) as the cut-off values.

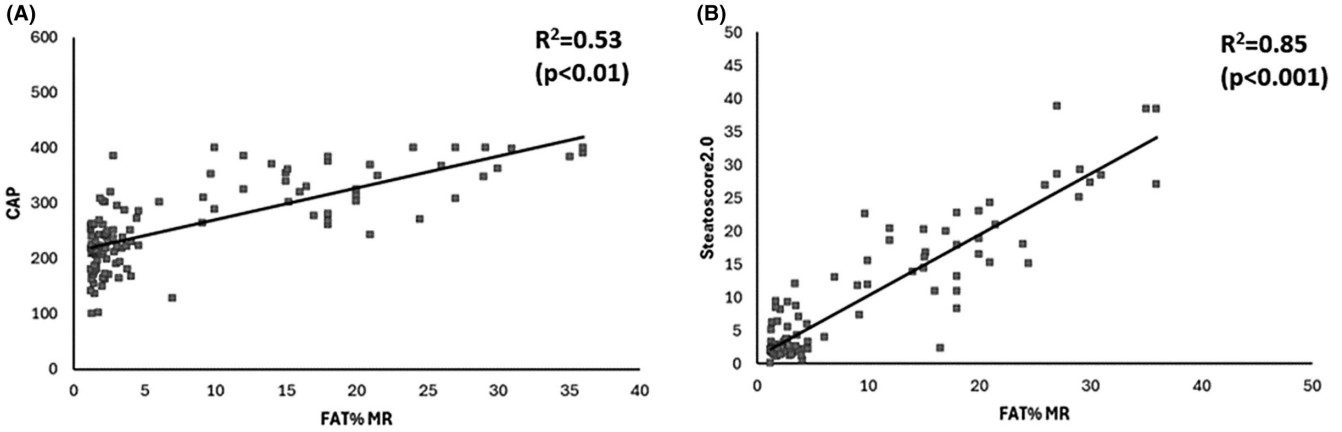

FIGURE 3 Correlation analysis between CAP and **MR** fat % and between Steatoscore2.0 and MR fat% on a subgroup of subjects ($N = 121$ patients). (A) CAP versus MR fat% values. (B) Steatoscore2.0 versus MR fat% values.

and MRI values in the same group of 121 patients ($R = 0.92$, $p < .001$) (Figure 3B).

3.3 | Reproducibility of quantitative US parameters

Two different operators (Op1 and Op2) evaluated US parameters (AR, HR, DV, HPV and PVW) and Steatoscore2.0 on the cohort ($N = 167$) of patients enrolled in Pisa. AR, HR, DV and Steatoscore2.0 computed by the first operator were strongly correlated with the measurement made by the second operator (AR: $R = 0.94$, HR: $R = 0.90$, DV: $R = 0.85$, Steatoscore2.0: $R = 0.96$; $p < .001$ for all). HPV and PVW values obtained by the two operators were moderately correlated (HPV: $R = 0.54$, PVW: $R = 0.62$ and $p < .001$ for both). The inter-observer ICC values obtained for AR, HR, DV and Steatoscore2.0 were 0.97, 0.95, 0.92 and 0.98, respectively. The analysis performed on HPV and PVW parameters resulted in lower ICC values (0.60, 0.72, respectively). The BAA performed on Steatoscore2.0 measurement shows an excellent agreement between the two operators, with a negligible bias (of −0.27, not significantly different from 0 and with $p = .14$) and satisfactory LoA (−4.8% and +4.3% for lower and upper limit, respectively) (Figure 4).

4 | DISCUSSION

The new multiparametric quantitative US imaging-based method (Steatoscore2.0) to measure fatty liver content (FLC) showed a good

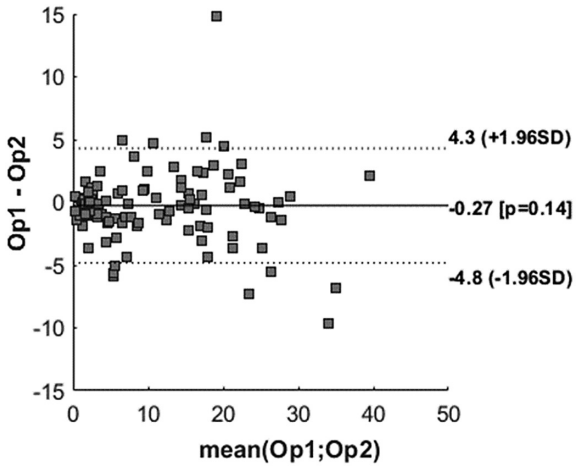

FIGURE 4 Study of inter-observer variability. Bland–Altman plot to evaluate the Steatoscore2.0 inter-observer reproducibility.

correlation ($R = 0.92$, RMSE = 5.15) with the reference standard MRI as demonstrated also by the absence of bias in the Bland–Altman analysis (BAA) (bias = 0.0, $p = 1.0$). The new model was implemented in a large cohort of 195 patients (134, 33 and 28 in training, validation and external validation datasets, respectively): a larger sample size than most of the studies proposing US-based methods to estimate liver fat content using MR values as reference standard.[25-31] In addition, the high diagnostic performance showed by Steatoscore2.0 in the training cohort where most of the patients had MRI FLC <10% qualifies the method as very promising tool to stratify patients in clinical practice.

The Steatoscore2.0 is powered by an optimized image processing software aimed to improve the reproducibility of the measurements as confirmed by the Intraclass Correlation Coefficient (ICC) of 0.98 and to promote the implementation of the method on commercial US devices. In this context, using a dataset focused on individuals with low-fat percentages has meant to improve the accuracy of FLC measurement in patients with low levels of steatosis, that is crucial but challenging for screening purposes.[32] Indeed, the proposed system is able to discriminate patients with MRI fat values ≥5% with a sensitivity, specificity, PPV, NPV and accuracy of 93.22%, 86.11%, 95.88%, 78.57% and 88.62%, respectively, and an AUROC of 0.958. Nevertheless, Steatoscore2.0 showed a high diagnostic performance (sensitivity, specificity, PPV, NPV and accuracy of 89.36%, 95.83%, 95.83%, 89.36%, 94.01%, respectively) in the identification of patients with higher fat content (≥10%) and an AUROC of 0.975. We also reported (Figure 2A,B) the optimal cut-offs obtained by maximizing the Youden indexes, which may be useful for a preliminary dichotomization of patients with or without >5%/>10% liver fat%. However, we recommend to use the quantitative value of the Steatoscore2.0 (instead of cut-off values) in the clinical practice, due to the great advantage of having a continuous measurement of the percentage of liver fat expressed in the same measurement unit of the recognized MRI gold standard. Therefore, Steatoscore2.0 appears as a suitable and reliable screening tool, but also it could be used for the follow-up and clinical management of the single patient with SLD/MASLD.

The methodology used to implement the model, namely the use of different statistical model in addition to multivariate linear regression, allowed to maximize the performances: specifically, US parameters were combined using linear, exponential and a mix of these models, which led to the identification of the three most relevant US parameters (AR, HR and DV) that are combined in the final prediction multiparametric model. Very recently, several non-imaging multiparametric models have also been proposed as predictor indexes of MASLD,[33-35] that may be mainly used to predict the presence/absence of MASLD but not to directly quantify liver fat content for which US imaging-based parameters are more suitable, as also recommended in.[36] To this regard, to date, there are few multiparametric systems capable to quantify the degree of hepatic steatosis from the processing of B-mode ultrasound images only, as most of the proposed systems are based on a single parameter. However, multiparametric US-based systems to monitor the progress of MASLD are very recently gaining increasing attention.[16] With the validation of the Steatoscore2.0, we demonstrated that multi-parametricity strengthens the accuracy of the estimate compared to single-parameter models. Moret et al.[37] studied the correlation between the "B-mode ratio" and MRI in subjects with chronic liver disease finding a lower correlation with respect to our HR (0.61[37] vs. 0.72 HR, $p < .001$ for both). In their studies, Jeon et al.[25] and Ferraioli et al.,[26] reported a good agreement between ATI and MRI-PDFF measurements of FLC, showing correlation values ranging between $R = 0.66$[25] and $R = 0.81$.[26] As regards multiparametric systems, Han et al.[38] proposed a system reporting a correlation with MRI-PDFF of 0.76, while Labyed et al.[39] proposed a two-parameter model for predicting MRI-PDFF, achieving a correlation of 0.87 and an RMSE of 4.5, comparable with our results ($R = 0.92$, RMSE = 4.62). Furthermore, in our work some single-parameter models were tested, from the simplest linear regression model to the exponential one. However, the significant improvement in performance was achieved only by switching to the multiparametric system. Indeed, the best mono-parametric model (using HR as a predictor) showed an R and RMSE of 0.76 and 6.68, respectively, which were significantly (William's test $p < .001$ and absolute errors t-test $p < .001$) lower than the prediction performance achieved by multiparametric Steatoscore2.0 ($R = 0.92$ and RMSE = 5.15). Furthermore, comparing CAP measurements with MRI values on a subgroup of subjects, we obtained an R of 0.73, significantly lower (William's test $p < .001$) than the correlation between Steatoscore2.0 and MRI values ($R = 0.92$).

Interestingly, the correlation between Steatoscore2.0 and MRI was comparable in the training, validation and external cohorts (0.92, 0.86 and 0.83) with similar RMSE (5.15, 4.62 and 4.17).

As regards the system reproducibility, inter-observer ICC values obtained for AR, HR, DV and Steatoscore2.0 demonstrate an excellent reproducibility (0.97, 0.95, 0.92 and 0.98, respectively). These

results have also been confirmed by the absence of a significant bias (−0.27, $p = .14$) and the range of LoA [−4.8; 4.3] in the BAA.

In our study, none of the subjects presented an estimated glomerular filtration rate <60 mL/min/1.73mq (median value of 98.1 and IQR 25%–75% of 88.3–104.9). Although we did not exclude patients with chronic kidney disease (CKD) a priori, our cohort was relatively young and did not present a high burden of risk factor for CKD. In the presence of advanced stages of kidney disease, the reduction in the cortical thickness and the "contracted" appearance of the kidney on ultrasound could limit the accuracy of the HR parameter, although this can be hampered by the multiparametric nature of the score. Future validation in this clinical setting would be needed.

Furthermore, the exclusion of patients with cirrhosis and the acquisition of ultrasound images by a single operator (one for each centre) are two main limitations of our study. The former was due to the need for a specific validation in cirrhotic patients, whose US images usually show a higher attenuation of the US beam, thus making them less suitable for the estimation of liver fat with a model using such a parameter. As well as for patients with CKD even at an early stage, which may impact on the value of HR parameter and, therefore, on the final score. Nevertheless, being the proposed model based on a multiparametric approach, it might prove in such patients (i.e. cirrhotic and advanced CKD patients) much more useful than mono-parametric models which are only based on the measurement of attenuation. The latter resulted from the decision to limit the inter-operator variability during the model generation phase. Future studies should therefore investigate the feasibility of Steatoscore2.0 in cirrhotic patients and the reproducibility of the results with data acquired from multiple sonographers. Finally, as the number of patients with high fat content (>24%) at MRI examination was limited (17 patients, 10% of the whole population) future studies should include these patients also. The challenging adoption of Steatoscore2.0 on patient with very high BMI (higher than 35 kg/m^2) can be considered another limitation of the study, as it is well known that distal attenuation is also dependent on the physical conformation of patients, and for subjects with high BMI, a part of the attenuation may be due to physical conformation and not necessarily to steatosis. However, the multiparametricity of the proposed score can reduce the impact of noisy distal attenuation.

In addition, difficulties in performing US examination due to meteorism or poor acoustic window, represent a limitation in assessing Steatoscore2.0, but it is worth emphasizing, as already demonstrated in the previous paper,[40] that right intercostal scanning can be used for comparison between renal cortex and liver parenchyma. This approach allows for clip acquisition even during the most challenging US examinations. An insurmountable limitation remains in the feasibility of the multiparametric score in all cases where comparison between cortex and liver parenchyma is not possible, such as renal agenesis, renal ectopia, nephrectomy, renal cortex thinning secondary to chronic disease and polycystic kidney disease. In these latter cases, Steatoscore2.0 calculation cannot be performed.

In conclusion, the proposed multiparametric approach based on three US-imaging features associated with intrahepatic fat accumulation is feasible and reliable to quantify fatty liver in clinical practice showing a good accuracy as compared to the standard MR fat percentage assessment. This method is quite inexpensive, easy to use and can be applied on any US equipment and qualifies as a new tool for fatty liver disease screening and monitoring response to therapies and lifestyle changes. These results prompt future studies aimed to evaluate the performance of this software tool in larger and different cohorts of patients including those with more advanced liver diseases and cirrhosis.

AUTHOR CONTRIBUTIONS

Laura De Rosa: Preparing first draft of the manuscript, analysis and interpretation of data, critical revision of the manuscript, approved final submission. Antonio Salvati: Analysis and interpretation of data, critical revision of the manuscript, approved final submission. Nicola Martini: Analysis and interpretation of data, critical revision of the manuscript, approved final submission. Marcello Mancini, Libertario Demi, Lorenzo Ghiadoni and Dante Chiappino: Critical revision of the manuscript, obtained funding, approved final submission. Simone Cappelli: Analysis and interpretation of data, critical revision of the manuscript, approved final submission. Ferruccio Bonino and Maurizia R. Brunetto: Study concept and design, analysis and interpretation of data, drafting the manuscript, critical revision of the manuscript, study supervision, approved final submission. Francesco Faita: Study concept and design, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript, obtained funding, study supervision, approved final submission.

ACKNOWLEDGEMENTS

None.

FUNDING INFORMATION

This study was supported by unrestricted grants from the Institute of Clinical Physiology and from the University of Pisa.

CONFLICT OF INTEREST STATEMENT

There is no conflict of interest for all authors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The study protocol was approved by the Ethics Committee of the University Hospital of Pisa (19/02/2016). All the subjects signed a written informed consent. Protocol n. 1179/2016.

ORCID

Laura De Rosa https://orcid.org/0000-0002-6030-8483
Maurizia R. Brunetto https://orcid.org/0000-0001-8364-9152
Francesco Faita https://orcid.org/0000-0002-6201-1843

REFERENCES

1. Rinella ME, Lazarus JV, Ratziu V, et al. A multi-society Delphi consensus statement on new fatty liver disease nomenclature. *J Hepatol*. 2023;79:E93-E94.
2. Carr RM, Oranu A, Khungar V. Nonalcoholic fatty liver disease: pathophysiology and management. *Gastroenterol Clin North Am [Internet]*. 2016;45(4):639-652. doi:10.1016/j.gtc.2016.07.003
3. Bravo AA, Sheth SG, Chopra S. Liver biopsy. *N Engl J Med*. 2001;344(7):495-500.
4. Smith EH. Complications of percutaneous abdominal fine-needle biopsy. *Review Radiology*. 1991;178(1):253-258.
5. Park YS, Park SH, Lee SS, et al. Biopsy-proven nonsteatotic liver in adults: estimation of reference range for difference in attenuation between the liver and the spleen at nonenhanced CT. *Radiology*. 2011;258(3):760-766.
6. Noureddin M, Lam J, Peterson MR, et al. Longitudinal comparison between MRI, MRS and histology-determined steatosis in NAFLD patients at two-time points in a randomized trial. *Hepatology (Baltimore, Md)*. 2013;58(6):1930-1940. http://www.ncbi.nlm.nih.gov/pubmed/23696515
7. Park SH, Kim PN, Kim KW, et al. Macrovesicular hepatic steatosis in living liver donors: use of CT for quantitative and qualitative assessment. *Radiology*. 2006;239(1):105-112.
8. Hashimoto E, Taniai M, Tokushige K. Characteristics and diagnosis of NAFLD/NASH. *J Gastroenterol Hepatol*. 2013;28(S4):64-70.
9. 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.
10. Bozic D, Podrug K, Mikolasevic I, Grgurevic I. Ultrasound methods for the assessment of liver steatosis: a critical appraisal. *Diagnostics*. 2022;12(10):2287.
11. Song SJ, Nogami A, Liang LY, et al. Performance of continuous controlled attenuation parameter and liver stiffness measurement by the novel SmartExam in metabolic dysfunction-associated steatotic liver disease. *Liver Int*. 2024;44(5):1167-1175.
12. 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.
13. Pu K, Wang Y, Bai S, et al. Diagnostic accuracy of controlled attenuation parameter (CAP) as a non-invasive test for steatosis in suspected non-alcoholic fatty liver disease: a systematic review and meta-analysis. *BMC Gastroenterol*. 2019;19(1):1-11.
14. Gao J, Lee R, Trujillo M. Reliability of performing multiparametric ultrasound in adult livers. *J Ultrasound Med* 2022 3;41(3):699–711.
15. Popa A, Bende F, Șirli R, et al. Quantification of liver fibrosis, steatosis, and viscosity using multiparametric ultrasound in patients with non-alcoholic liver disease: a "real-life" cohort study. *Diagnostics*. 2021;11(5):783.
16. Ferraioli G, Barr RG, Berzigotti A, et al. WFUMB guidelines/guidance on liver multiparametric ultrasound. Part 2: guidance on liver fat quantification. *Ultrasound Med Biol*. 2024; 50(8):1088-1098.
17. Di Lascio N, Avigo C, Salvati A, et al. Steato-score: non-invasive quantitative assessment of liver fat by ultrasound imaging. *Ultrasound Med Biol [Internet]*. 2018;44(8):1585-1596. doi:10.1016/j.ultrasmedbio.2018.03.011
18. Naressi A, Couturier C, Devos JM, et al. Java-based graphical user interface for the MRUI quantitation package. *MAGMA*. 2001;12(2–3):141-152.
19. Vanhamme L, Van Den Boogaart A, Van Huffel S. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. *J Magn Reson*. 1997;129(1):35-43.
20. Reeder SB, Cruite I, Hamilton G, Sirlin CB. Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. *J Magn Reson Imaging*. 2011;34(4):729-749.
21. Marchesini G, Day CP, Dufour JF, et al. EASL-EASD-EASO clinical practice guidelines for the Management of non-alcoholic fatty liver disease. *Obes Facts*. 2016;9:65-90.
22. Xia MF, Yan HM, He WY, et al. Standardized ultrasound hepatic/renal ratio and hepatic attenuation rate to quantify liver fat content: an improvement method. *Obesity*. 2012;20(2):444-452.
23. Thomas S. Diagnostic Ultrasound Imaging: Inside Out.
24. Williams EJ. The comparison of regression variables. *J R Stat Soc B Methodol*. 1959;21(2):396-399.
25. Jeon SK, Lee JM, Joo I, et al. Prospective evaluation of hepatic steatosis using ultrasound attenuation imaging in patients with chronic liver disease with magnetic resonance imaging proton density fat fraction as the reference standard. *Ultrasound Med Biol*. 2019;45(6):1407-1416.
26. Ferraioli G, Maiocchi L, Raciti MV, et al. Detection of liver steatosis with a novel ultrasound-based technique: a pilot study using MRI-derived proton density fat fraction as the gold standard. *Clin Transl Gastroenterol*. 2019;10(10):1-8.
27. Ferraioli G, Maiocchi L, Savietto G, et al. Performance of the attenuation imaging Technology in the Detection of liver steatosis. *J Ultrasound Med*. 2021;40(7):1325-1332.
28. Tada T, Kumada T, Toyoda H, et al. Attenuation imaging based on ultrasound technology for assessment of hepatic steatosis: a comparison with magnetic resonance imaging-determined proton density fat fraction. *Hepatol Res*. 2020;50(12):1319-1327.
29. Kwon EY, Kim YR, Kang DM, Yoon KH, Lee YH. Usefulness of US attenuation imaging for the detection and severity grading of hepatic steatosis in routine abdominal ultrasonography. *Clin Imaging*. 2021;76:53-59.
30. 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.
31. Imajo K, Toyoda H, Yasuda S, et al. Utility of ultrasound-guided attenuation parameter for grading steatosis with reference to MRI-PDFF in a large cohort. *Clin Gastroenterol Hepatol*. 2022;20(11):2533-2541.e7.
32. 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. *J Hepatol*. 2009;51(3):433-445.
33. Tang LJ, Sun DQ, Song SJ, et al. Serum PRO-C3 is useful for risk prediction and fibrosis assessment in MAFLD with chronic kidney disease in an Asian cohort. *Liver Int*. 2024;44(5):1129-1141.
34. Fan R, Yu N, Li G, et al. Machine-learning model comprising five clinical indices and liver stiffness measurement can accurately identify MASLD-related liver fibrosis. *Liver Int*. 2024;44(3):749-759.
35. Li G, Lin H, Sripongpun P, et al. Diagnostic and prognostic performance of the SAFE score in non-alcoholic fatty liver disease. *Liver Int*. 2024;44(1):15-26.
36. Ferraioli G, Berzigotti A, Barr RG, et al. Quantification of liver fat content with ultrasound: a WFUMB position paper. *Ultrasound in Medicine and Biology*. Vol 47. Elsevier Inc.; 2021:2803-2820.
37. Moret A, Boursier J, Houssel Debry P, et al. Evaluation of the hepatorenal B-mode ratio and the controlled attenuation parameter for the detection and grading of steatosis. *Ultraschall Med*. 2020;43:479-487.
38. 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.

39. Labyed Y, Milkowski A. Novel method for ultrasound-derived fat fraction using an integrated phantom. *J Ultrasound Med*. 2020;39(12):2427-2438.
40. De Rosa L, Salvati A, Bonino F, Brunetto MR, Faita F. Effects of different scan projections on the quantitative ultrasound-based evaluation of hepatic steatosis. *Healthcare (Switzerland)*. 2022;10(2):374-385.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: De Rosa L, Salvati A, Martini N, et al. An ultrasound multiparametric method to quantify liver fat using magnetic resonance as standard reference. *Liver Int*. 2024;00:1-12. doi:10.1111/liv.16078