

Research Article

Full-Range Liver Fat Fraction Estimation in Magnitude MRI Using a Signal Shape Descriptor

Yuri A. Costa ¹, Carlos P. Filho,² Gabriela A. Flores,³
Evandro L. L. Rodrigues,¹ and Fernando F. Paiva ³

¹Department of Electrical and Computer Engineering, University of São Paulo, São Carlos, SP 13566-590, Brazil

²Institute of Math and Computer Sciences, University of São Paulo, São Carlos, SP 13566-590, Brazil

³São Carlos Institute of Physics, University of São Paulo, São Carlos, SP 13566-590, Brazil

Correspondence should be addressed to Fernando F. Paiva; paiva@ifsc.usp.br

Received 28 March 2019; Accepted 2 June 2019; Published 9 July 2019

Academic Editor: Massimiliano F. Peana

Copyright © 2019 Yuri A. Costa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Current methods for estimation of proton density fat fraction (PDFF) of the liver using magnitude magnetic resonance (MR) imaging face the challenge of correctly estimating it when fat is the dominant molecule; i.e., PDFF is more than 50%. Therefore, the accuracy of the methods is limited to half-range operation. We introduce a method based on neural networks for regression capable of estimating over the full range of fat fractions. We built a neural network based on the angles and distances between the data in the discrete MR signal (ADALIFE), using these as features associated with different PDFFs and as input for the network. Tests were performed using ADALIFE and Multi-interference, a state-of-the-art method to estimate PDFFs, with simulated signals at various signal-to-noise (SNR) values. Results were compared in order to verify repeatability and agreement using Bland-Altman and REC curves. Results for Multi-interference were similar to its *in vivo* literature, showing the relevance of a simulation. ADALIFE was able to correctly estimate fat fractions up to 100%, breaking the current paradigm for full-range estimation using only offline postprocessing. Within half range, our method outperformed Multi-interference in repeatability and agreement, with narrower limits of agreement and lower expected error at any SNR.

1. Introduction

Nonalcoholic fatty liver disease (NAFLD) is characterized by hepatic steatosis unrelated to alcohol abuse and may lead to chronic liver disease. It has been estimated by a metastudy that NAFLD has a prevalence of 25% among global population [1] and of 30% in the United States according to another study including 6000 subjects [2].

Among noninvasive quantitative approaches for estimation of hepatic fat content, magnetic resonance imaging (MRI) has shown great reproducibility [3, 4] and good agreement to clinical and research gold standards, biopsy, and magnetic resonance spectroscopy, respectively [5, 6]. Quantitative assessment of liver steatosis with MRI can be done by estimating the hepatic proton density fat fraction based on chemical shift images (from here on called proton density fat fraction, PDFF), which may carry magnitude or complex information. Complex imaging means obtaining

magnitude and phase images and whereas it allows full-range fat fraction estimation i.e., from 0 to 100% PDFF, phase errors must be accounted for as they can cause a bias of about 5% PDFF [7]. Additionally, complex acquisition and correction algorithms are not standard on clinical MRI scanners, being sold separately as a vendor-dependent software [8]. Magnitude-based images on the other hand are clinical routine and do not include phase errors. However, fat fraction estimation is prone to their signal-dependent rician noise [9] and is known to be limited to estimations up to 50% PDFF due to an effect called water-fat dominance ambiguity [10].

Current state-of-the-art method for fat fraction estimation using magnitude imaging is Multi-interference, which requires a set of multiple images and calculations can be performed offline. This method estimates tissue parameters within a set of registered voxels or regions of interest (ROIs). It calculates mainly water and fat proton densities, and a single apparent transverse magnetization decay $T2^*$ by fitting

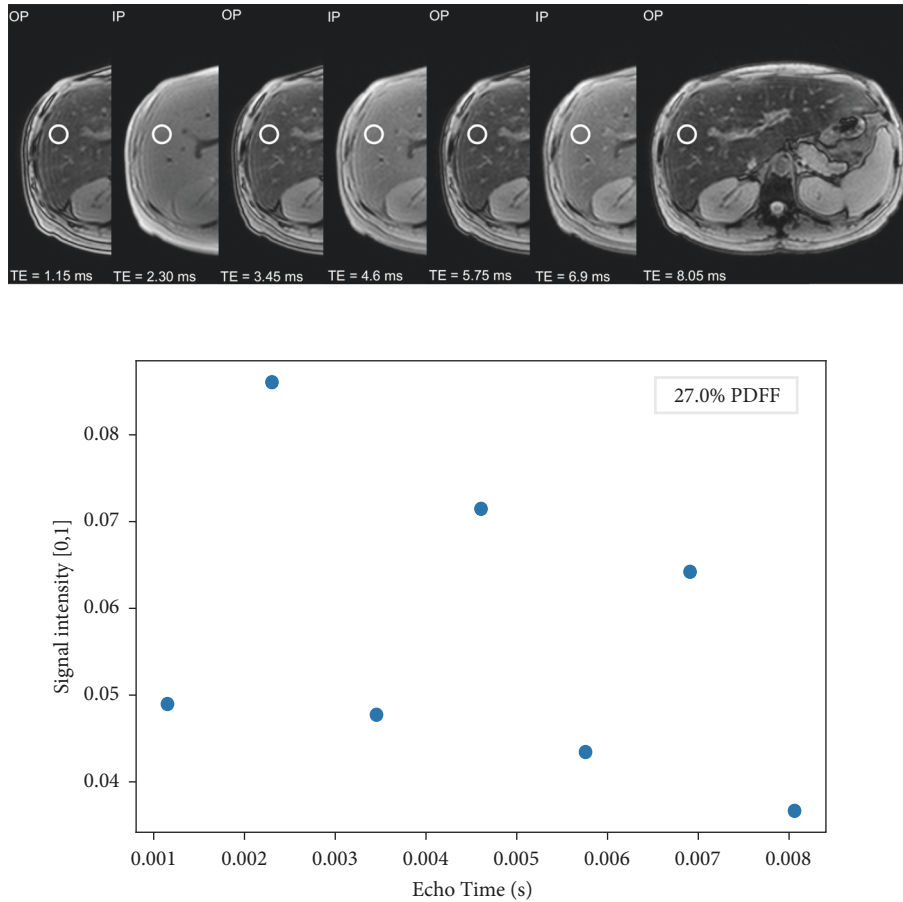


FIGURE 1: A set of GRE OP and IP images with circular white ROIs drawn on the liver and a plot of the corresponding average signal within those ROIs for each image at the corresponding echo time. Images were acquired at a magnetic field strength of 3 Tesla.

the data using nonlinear least squares approximation. With increase in computational power very recent works obtained success approaching this problem from advanced spectral models and machine learning perspectives [11, 12], bringing this challenge back to light.

In this paper, we propose a technique to estimate liver fat fraction in the full range of 0 to 100% based on magnitude MRI signal by extracting shape features (angle; distance) from simulated gradient echo signals to train an artificial neural network (ANN) for regression. We name this method Angle and DistAncE for LIver Fat Estimation (ADALIFE). Multi-interference and our method are evaluated to estimate simulated PDFF values from signals contaminated with different noise levels. Statistics are performed to verify their performance within the range of 0-50% and 0-100% PDFF.

2. Theory

An abdominal MRI exam for assessment of liver steatosis generates a stack of slice images of the patient's abdomen. In a gradient echo (GRE) multiecho sequence, for each slice position, multiple signals are repeatedly acquired within a time window of milliseconds. When combined, they form images spaced in time. A typical way to estimate liver fat fraction is by analyzing how these images change over that

time window. This section explores the concepts of these images and the properties of their signals and how literature estimates fat fraction based on them.

2.1. Magnitude Gradient Echo Signal. A typical magnitude GRE in-phase (IP) and opposed-phase (OP) image set aims to explore the difference in water and fat content among tissues. For a fixed slice position, the test consists of a series of images acquired at different echo times (TE) after an excitation pulse. As the ^1H magnetization of distinct molecules precesses at characteristic frequencies, echo times are chosen such as to cause water and fat magnetizations to be in-phase or opposed-phase for each acquisition. That means, in a very simplistic way, magnitude IP images carry a signal that is proportional to the absolute value of the summation of proton densities of water and fat, whereas magnitude OP images carry the magnitude of their subtraction. A set of liver GRE IP-OP images acquired on a 3 Tesla system and the corresponding plot of the average signal intensity (S) versus echo time TE of a region of interest in each image are shown in Figure 1.

The signal of a voxel for a magnitude GRE image at different echo times such as plotted in Figure 1 can be mathematically modeled according to the physical principles of magnetic resonance imaging as long as tissue properties

and imaging parameters are known. Equation (1) represents the signal of the magnitude gradient echo sequence for n different molecules, where S is the voxel signal, k is a machine-dependent constant, ρ_n is the proton density of the given molecule, α is the flip angle in radians, TR is the repetition time in seconds, $T1_n$ is the longitudinal relaxation time of the molecule in seconds, TE is the echo time in seconds, $T2_n^*$ is the apparent transverse relaxation time of the molecule in seconds, and f_n is the magnetization precession frequency of the molecule in hertz. Tissue properties are ρ_n , $T1_n$, $T2_n^*$, and f_n . Imaging parameters are TR , TE , and α .

Equation (2) is an approximation of (1) for low α and short TR , which minimize $T1$ effects, as previously proposed in literature [10, 13]. Therefore, (2) describes the signal intensity S in a voxel for the magnitude GRE sequence with small flip angle at a given TE as a constant k multiplied by the modulus of a sum of exponentially damped complex sine waves with different amplitudes ρ , frequencies f , and decay rates $1/T2^*$.

Equation (3) defines the precession frequency in Hertz of a given molecule as used in (1) and (2). γ is the hydrogen gyromagnetic ratio in MHz/T, B_0 is the main static magnetic field in Tesla, and δ_n indicates the chemical shift of the n molecule in parts per million, i.e., its precession deviation in relation to that of a reference (usually tetramethylsilane, $\text{Si}(\text{CH}_3)_4$). The convenient echo times for in-phase and opposed-phase can then be calculated according to (4), where f_{water} and f_{fat} are water and fat frequencies, respectively, resulting in half the period of the difference in precession between those molecules.

For IP-OP GRE imaging or spectroscopy as defined here, water and methylene (CH_2) chemical shifts are used ($\delta_{water} = 4.7$ and $\delta_{fat} = 1.3$ ppm, respectively) [14], the latter being the most abundant fat moiety in liver [15]. However, fat is not a molecule as simple as water, and it may be formed by different combinations of glycerol and fatty acids which must be taken into account for a precise estimation.

$$S = \left| k \sum_n \rho_n \frac{\sin(\alpha) [1 - \exp(-TR/T1_n)]}{1 - \cos(\alpha) \exp(-TR/T1_n)} \cdot \exp\left(-\frac{TE}{T2_n^*}\right) \exp(2\pi i f_n TE) \right| \quad (1)$$

$$S = \left| k \alpha \sum_n \rho_n \exp\left(-\frac{TE}{T2_n^*}\right) \exp(2\pi i f_n TE) \right| \quad (2)$$

$$f_n = \gamma B_0 \delta_n \quad (3)$$

$$TE = \frac{1}{2} \frac{1}{f_{water} - f_{fat}} \quad (4)$$

2.2. Fat Fraction Estimation in Literature. Proton density fat fraction or simply fat fraction PDFF is defined according to (5), where ρ_f and ρ_w are fat and water proton densities, respectively. In order to obtain those values, the first method was proposed by Dixon in 1984 [14] using a pair of opposed-phase and in-phase images at echo times calculated for water and methylene, which allowed for estimation of only this fat

moiety. Calculation is performed arithmetically according to (6). Later magnitude-based methods added a third echo to include the effects of a single $T2^*$ exponential decay, usually acquiring IP-OP-IP images for better SNR [16, 17].

$$PDFF = \frac{\rho_f}{\rho_w + \rho_f} \quad (5)$$

$$\rho_w \approx \frac{(S_{IP} + S_{OP})}{2} \quad (6)$$

$$\rho_f \approx \frac{(S_{IP} - S_{OP})}{2}$$

Studies of the *in vivo* liver fat spectrum [18] using high resolution spectroscopy identified not only CH_2 but six fat moieties in fixed proportions across subjects. Three of them are considered more prominent in humans, located at 0.9, 1.3, and 2.1 ppm, accounting for about 91% total liver fat [15]. The other peaks are located at 2.75, 4.2, and 5.3 ppm. These fat peaks and water (4.7 ppm) partially overlap, making it difficult to characterize them. Based on those studies, Yokoo et al. proposed the Multi-interference method [19], taking into consideration the three major peaks (which might be adjusted to compensate the missing peaks [15, 18, 20]) and a single $T2^*$ value. They showed that it was possible to use nonlinear least squares to calculate ρ_f and ρ_w by fitting the signal of six echoes to (7), a reduction of the previously presented GRE equation.

Equation (7) results in a signal S that is $k\alpha$ times the magnitude of the water proton density (ρ_w) and the weighted sum (ρ_f) of three complex sinusoids with different, fixed amplitudes c_n , and frequencies Δf_n relative to water, and the result multiplied by an exponential decay dependent on a single $T2^*$ value. The amplitudes c_n are equivalent to the fixed proportion between the fat peaks observed by Hamilton et al. adjusted to consider all peaks of fat [15]. Please note that, for the ultimate goal of PDFF calculation, k and α do not need to be estimated in (7), as they multiply ρ_w and ρ_f later to be divided by themselves in (5).

The presented methods for estimation using magnitude images at low flip angles and fixed echo times undergo water-fat dominance ambiguity. This effect is the inability to tell whether water or fat proton density is greater when phase information is discarded and modulus operation is applied in (1), as single data point cannot give information to resolve if its value is due to the subtraction of fat from water or vice versa. Water is then assumed to be dominant based on *in vivo* findings [20] and magnitude estimation is usually limited to 0-50% PDFF when using standard clinical protocols. However, a thorough analysis of the signal across the echo times may show other relevant clues on this issue.

$$S(TE, \alpha) = k\alpha \left| \rho_w + \rho_f \sum_{n=1}^3 c_n \exp(2\pi i \Delta f_n TE) \right| \cdot \exp\left(-\frac{TE}{T2^*}\right); \quad \text{with } \sum_{n=1}^3 c_n = 1 \quad (7)$$

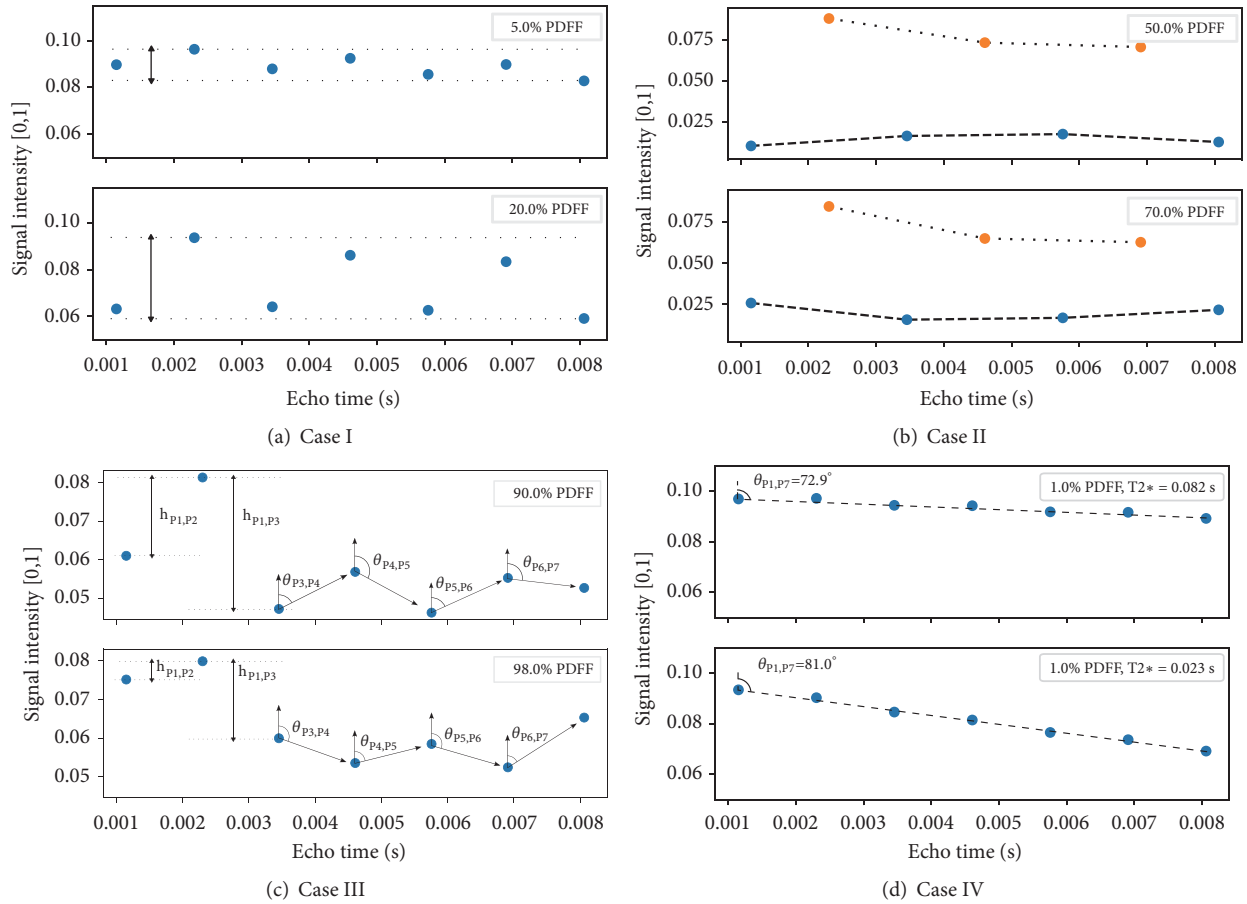


FIGURE 2: Shape particularities in geometric distribution of signal as fat fraction varies. Case I (a) describes how distance changes for IP-OP echoes as fat fraction increases. Case II (b) shows a change in the concavity of OP points as fat becomes dominant. Case III (c) shows a swap in the position of latter echoes as PDFF increases around 90 to 100%. Case IV makes the change in overall decay caused by parameter $T2^*$ explicit.

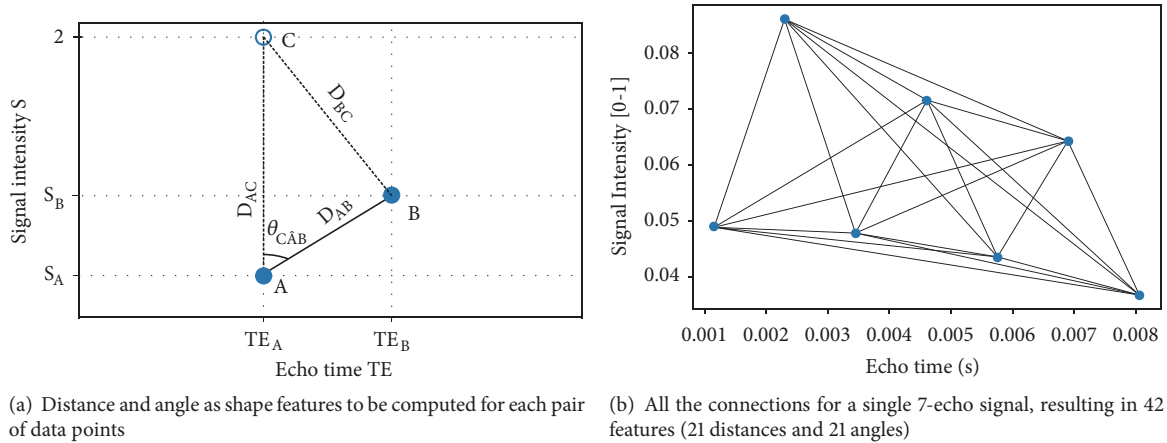
3. Materials and Methods

3.1. Signal Shape and Feature Extraction. By simulating the liver signal at several fat fractions we observed interesting particularities in the amplitude distribution of the data points across the time domain. For the next examples we used (2) for the simulations considering the frequencies and proportions of water molecule and six lipids as described in [10, 15, 21]. All molecules had the same $T2^*$ value of 0.082 seconds and no noise was added. Each signal was calculated for seven echo times considering water and methylene interaction for setting IP and OP times and a main static magnetic field of 3 Tesla.

Figure 2 explores four cases depicting particularities we found to highlight differences in signals at different fat fractions. Case I in Figure 2(a) is an example of two fat fractions far apart within the range of 0-50% in which we observe an evident increase in the y-axis distance between IP and OP points. This is expected within this range where water is dominant and IP represent $\rho_w + \rho_f$ and OP is $\rho_w - \rho_f$. Distances decrease back after 55% PDFF, as expected when fat becomes dominant. The reason for being about 55% is that at that point the three major fat moieties account for about 50% of total signal (and as most estimation methods only

consider these moieties, it is said to be 50% fat). Case II in Figure 2(b) shows a line connecting IP points (dotted line; orange points) and another for OP points (dashed line; blue points). The concavity of the dashed line aims downward if PDFF is less than 55% and then changes to upward if greater. Pointing downward becomes more evident at fat fraction close to 50% and pointing upward when close to 90%. The connection between IP points seems to change only in terms of steepness. Case III as depicted in Figure 2(c) demonstrates that for the latter echoes their relative position (to each other) is practically reversed. The angle θ between y-axis and a line between two consecutive data points shows this change. Moreover, in this fat-dominant example it can be observed that the y-axis distances are shortened in the initial echoes, as commented before. Lastly, Case IV was simulated using two different common $T2^*$ values for all molecules and the same fat fraction. It can be seen a difference in slope, represented by the change in angle for a line connecting the first and last echoes. Evidently, applying different relaxation values for each molecule would result in different signals, but the overall effect is a change in slope.

Considering the unit vector of a signal (dividing each amplitude value by the L^2 -Norm of the signal) in order to



(a) Distance and angle as shape features to be computed for each pair of data points (b) All the connections for a single 7-echo signal, resulting in 42 features (21 distances and 21 angles)

FIGURE 3: Features to be calculated for each pair of data points (a) and all the connections for a complete signal (b).

eliminate amplification effects, we can compute the Euclidean distance between any pair of points A and B at echo times TE_A and TE_B and with signal intensities SI_A and SI_B . We can also compute the angle between the y-axis and the straight line that connects both points, as shown in Figure 3(a). If we calculate these features for all possible connections between the seven echoes we end up with a network such as the one in Figure 3(b), which results in a descriptor containing 42 features that can be used in a machine learning approach. Intuitively, distances between in-phase and opposed-phased points will reflect the height, distances and angles within IP or OP between inner and outer points will reflect the concavity, and connections between extremity points will reflect decay, which should suffice to differentiate between fat fractions. The distance between points A and B can be calculated using (8) and the angle using law of cosines for a triangle with points A and B and an auxiliary point C located at $(TE_A, 2)$ for convenience.

$$D_{AB} = \sqrt{(S_B - S_A)^2 + (TE_B - TE_A)^2} \quad (8)$$

$$D_{BC} = \sqrt{(2 - S_B)^2 + (TE_B - TE_A)^2}$$

$$D_{AC} = 2 - S_A \quad (9)$$

$$\theta_{CAB} = \arccos\left(\frac{D_{AC}^2 + D_{AB}^2 - D_{BC}^2}{2D_{AC}D_{AB}}\right)$$

The set of equations presented in (9) allows the calculation of the sides D_{AC} and D_{BC} of the triangle and the angle θ_{CAB} . It is useful to simply use the index of the echo times instead of the actual values to make it independent of main magnetic field strength but keep the order. In order to scale the resulting distance and angle to values between zero and one, we must figure out maximum and minimum possible values for distance and angle. Distance is minimum when two points are consecutive and horizontally aligned and maximum when they are distant by the maximum number of echoes and greatest difference in signal (considering the signal to be between zero and one). Angle is minimum when two points are consecutive and the earlier and latter

points have minimum and maximum signals, respectively. The maximum value can be obtained by subtracting the minimum value from π , which represents consecutive points and swapped signal values.

3.2. *Obtaining a Complete GRE Signal.* Signals with seven echoes were generated based on (2) considering the effects of water molecule and six fat moieties, at echo times chosen for water and methylene to be in-phase and opposed-phase in a 3 Tesla magnetic field, according to (4). It has been suggested that the proportion between fat moieties tends to be similar regardless the fat fraction [15]. That means that fixed ρ_n values for each fat moiety can be used and then multiplied by a desired overall fat proton density ρ_f .

Thus, values 0.088, 0.7, 0.12, 0.006, 0.039, and 0.047 (total equals 1) were used for fat peaks at 0.9, 1.3, 2.1, 2.75, 4.2, and 5.3 ppm, respectively [15]. $T2^*$ values are known to vary according to their molecular surrounding, presence, or iron deposits in the liver and magnetic field inhomogeneities. However, typical values have been described in literature [10, 15, 21], allowing us to simulate the variability of these values using truncated normal distributions. Lower limit was 0.001 and upper limit was 1 second. The following average values in seconds were used: $T2^*_{water} = 0.027 \pm 0.007$, $T2^*_{fat0.9} = 0.083 \pm 0.016$, $T2^*_{fat1.3} = 0.061 \pm 0.011$, and $T2^*_{fat2.1} = 0.052 \pm 0.030$, where the \pm sign indicates standard deviation. Fat peaks located at 2.75, 4.2, and 5.3 ppm were assigned fixed $T2^*$ values of 0.051, 0.025, and 0.039 seconds, respectively, due to lack of information in literature for a proper distribution. The frequency of each molecule was determined by (3), terms k and α were set to 1, and ρ_f and ρ_w were set using random positive numbers between 0 and 1 that would satisfy a given fat fraction value, according to (5). Finally, Rician distributed noise [22] was added with signal-to-noise ratios (SNR) between 25 and 200; values we experienced in clinical images and similar to previously reported in literature for liver and water-fat separation imaging [23, 24]. Simulations were performed using Python 3.

3.3. *Building the Artificial Neural Network.* The model consisted on a fully connected regression neural network of the type multilayer perceptron (MLP), with 42 inputs, three hidden layers with sizes 32, 16, and 2, and one output. Activation function was Rectified Linear Unit function (ReLU) and optimizer method was ADAM [25] with default settings (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e - 7$, and decay = 0.0). The size of the training batches was 400 and the network was trained during 2000 epochs using separate train and validation datasets. The model was built and trained using Keras with TensorFlow backend in Python 3.

3.4. *Datasets for Training, Validation, and Testing.* Datasets for training, validation, and testing differ in size and in how data is distributed. The training dataset consisted of 50000 signals with 1000 different fat fraction values linearly distributed between 0.1 and 100%. In this way, each PDFF is presented in 50 random combinations of ρ_w , ρ_f , $T2^*_n$, and SNR values. A validation dataset is used during the learning process of the network for fine tuning. For this task, we used a dataset containing 20000 signals with single, random fat fraction values uniformly distributed between 0.1 and 100%, random $T2^*_n$ and SNR values. Test datasets are the ones used to finally evaluate the performance of the trained network and Multi-interference methods.

There were 16 test datasets, each consisting of 5000 instances of signals with fixed SNR values of 25, 50, 100, and 200 and PDFF ranging from 0 to 45 or 100%. For a given range of PDFF, the dataset corresponding to each SNR had random and unique fat fraction values uniformly distributed within its range. For each dataset, a repetition set with same fat fractions was generated from different parameters for repeatability analysis. When the datasets were meant to be used by the neural network, there was a conversion step to calculate the distances and angles for each signal as explained before.

3.5. *Statistical Analysis.* Multi-interference and ADALIFE were used to estimate the values of the test datasets. For each test dataset (SNR 200, 100, 50, and 25), Multi-interference estimated fat fractions ranging from 0.1 to 45% and ADALIFE estimated within ranges 0.1-45% and 0.1-100%. Multi-interference was adjusted to compensate all six fat peaks according to Hamilton et al. [15]. Repeatability tests of the methods were performed using two test datasets with the same fat fractions generated from different parameters, for example, changing $T2^*$ and noise values. This should lead to slightly different estimations for the same value in contrast to an ideal method that would estimate the same given PDFF independently of any other parameter of the signal. This approach was used as otherwise simulations would lead invariably to the same results.

In order to assess the repeatability, for each method and SNR we calculated the repeatability coefficient R_c and Bland-Altman (BA) analysis [26]. Agreement between a given method and simulated fat fractions (True PDFF) was estimated using Lin's concordance correlation coefficient ρ_c [27], BA analysis adjusted for comparing to a reference [28], and regression error characteristic (REC) curves with

TABLE 1: Repeatability results for Multi-interference and ADALIFE in half-range and full-range estimation.

Method	SNR	R_c	Outliers†
Multi-interference (half-range only)	200	0.0219	0
	100	0.0219	0
	50	0.0250	0
	25	0.0346	0
ADALIFE (half-range, full-range)	200	0.0067, 0.0088	4, 9
	100	0.0085, 0.0101	2, 2
	50	0.0134, 0.0133	5, 2
	25	0.0265, 0.0229	4, 5

Repeatability was assessed using two estimations of the same values. As R_c is 1.96 SD, the closest to zero the better. Both methods presented zero mean as estimated to four decimal places, which makes BA 95% agreement limits simply $\pm R_c$.

respective areas over the curves (AOC) [29]. A Kolmogorov-Smirnov two-sample test was used to assess the significance of difference between the REC curves for Multi-interference and ADALIFE under the null hypothesis that the error generated by the models follow the same distribution. Adjustments for outliers were performed for the BA and repeatability analysis when needed considering only the data within plus/minus five standard deviations (SD) of the mean of the differences of limits of agreement or repeatability. For all other analyses, the whole data was included. All the data analyses were performed using StatsModels and SciPy Python modules.

4. Results

In this section we present the repeatability and agreement results for Multi-interference and ADALIFE in half range (0.01-45% PDFF) and for ADALIFE in full range (0.01-100% PDFF). For the sake of simplicity and as all statistics except for the Concordance Correlation Coefficient are given in PDFF units, in this section fat fractions will be stated as decimals.

4.1. *Repeatability.* Table 1 presents the repeatability coefficient (R_c) calculated as 1.96 SD of the differences and the number of outliers excluded for each method and SNR. As both methods presented zero mean considering four decimal places, 95% limits can be directly obtained as $\pm R_c$ and 95% confidence intervals are in the order of 10^{-4} . For the calculation of repeatability, a few outliers had to be removed from the ADALIFE results in order to adjust the tails for a normal distribution and the Bland-Altman analysis make sense, averaging 0.08% of the datasets. Their incidence did not seem to be related to the SNR but to the size of the dataset (e.g., $n = 500$ would present no outliers), although no statistics were performed to confirm this relationship. We also verified that, as SNR decreases, the repeatability coefficient becomes worse. ADALIFE had a higher repeatability coefficient than Multi-interference at every signal-to-noise ratio (maximum difference = 0.0152; minimum difference = 0.0081), though Multi-interference showed a more stable repeatability to the decrease in SNR. The coefficient for ADALIFE in full range

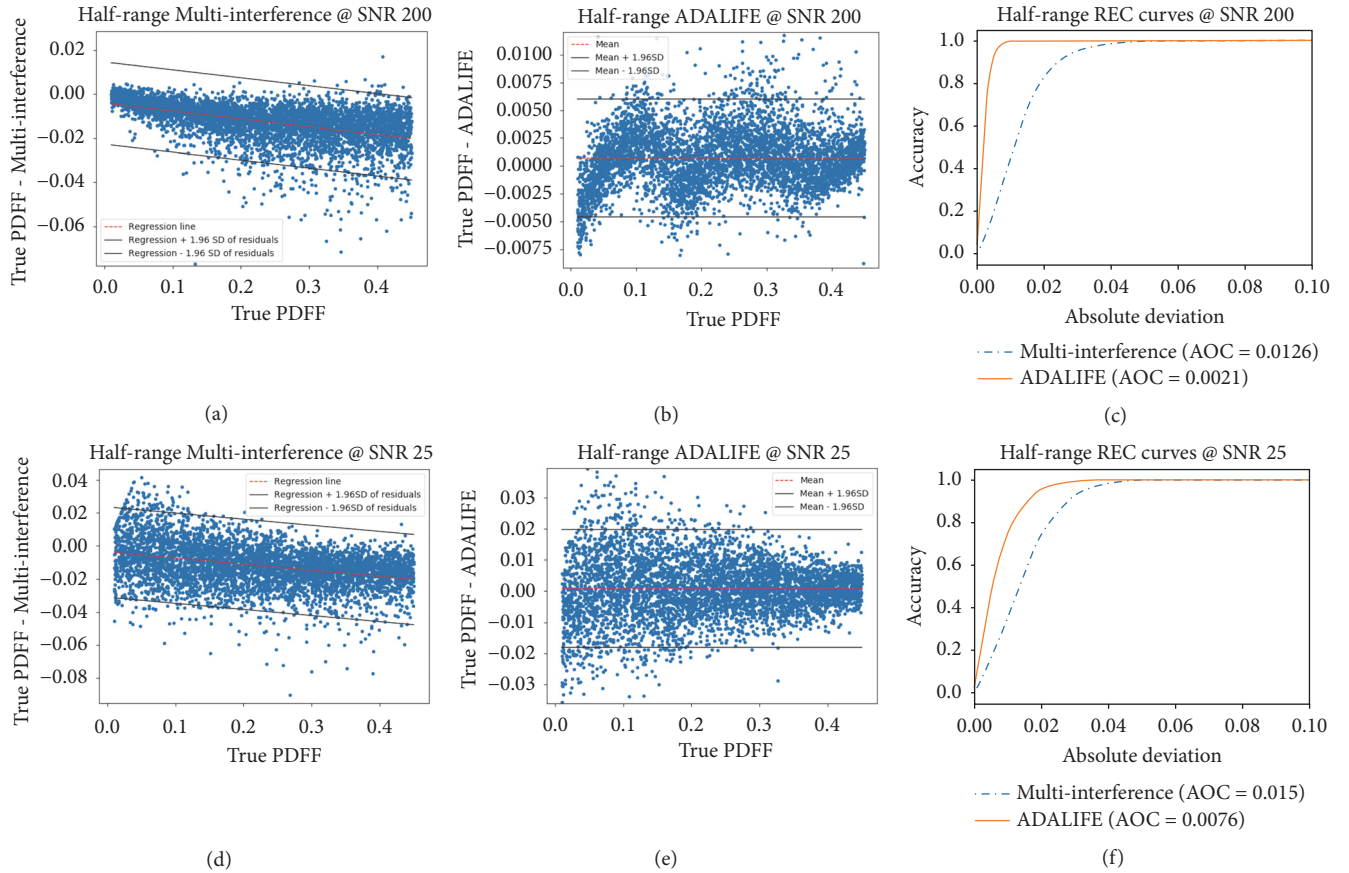


FIGURE 4: Bland-Altman plots (a), (b), (d), and (e) and REC curves (c) and (f), for Multi-interference and ADALIFE at SNR 200 and 25. Notice the difference in scales. Data points omitted on the plot due to scaling (out of 5000): (a) 1, (b) 8, (d) 3, and (e) 9.

was generally higher than in half range, but closely related (maximum difference = $0.0036PDFF$).

4.2. Agreement. Agreement between a selected model and true fat fraction values was estimated using the concordance correlation coefficient ρ_c , Bland-Altman, and REC curves plots. Figure 4 shows the BA plots in (a), (b), (d), and (e), and REC curves in (c) and (f) for Multi-interference and ADALIFE at SNR 200 and 25, for half-range estimation. Similar plots for full-range estimation are shown in Figure 5 except that REC curves compare ADALIFE performance over different SNR. Values for ρ_c , AOC, and BA analysis are presented in Table 2 for Multi-interference and Table 3 for ADALIFE. Outliers were removed from ADALIFE results for calculation of BA analysis, averaging 0.05% of the datasets.

Comparing half-range results, concordance correlation coefficient of ADALIFE was higher at every SNR except at 25. Area over the curve was lower for ADALIFE and the mean and 95% limits of agreement did not need to include dependency on True PDFF level. These limits at 0.0625 PDFF, a binary classification threshold in literature [20] considering SNR 200, are $[-0.1460, 0.0871]$ for Multi-interference and $[-0.0045, 0.0060]$ for ADALIFE. 95% confidence intervals for Multi-interference are in the order of 10^{-3} in slope and 95% limits and 10^{-4} in intercept. In ADALIFE they are in the order of 10^{-4} for mean and 95% limits of agreement. REC

curves presented in Figures 4(c) and 4(f) show a slow climb at start for Multi-interference which is due to bias in slope as verified in (a) and (d), in contrast to the steep ascent for ADALIFE due to the concentration of most differences being around zero. The Kolmogorov-Smirnov test for differences between the REC curves confirmed that they are significantly different with $p < 0.001$ at any SNR. Area over the curve estimates expected error for a given method, and ADALIFE obtained values six to two times smaller depending on the SNR. A noticeable feature is that expected error and limits of agreement increase more for ADALIFE than for Multi-interference as SNR decreases.

ADALIFE's full-range results are comparable to those in half range presenting higher ρ_c , lower AOC (except for SNR 200), lower mean, and similar limits of agreement. The model was able to estimate fat fractions between 0.001 and 1 in the worst case of SNR 25 with maximum expected error of 0.0062 and 95% limits of agreement of $[-0.0162, 0.0166]$, which are still lower values than the ones for Multi-interference operating in half range only. 95% confidence intervals are in the order of 10^{-4} for mean and limits of agreement. We identified a break in the continuity of the estimation around 0.56 PDFF which resulted in errors of up to ± 0.0175 . At the same time, noise seems to have a much stronger effect at the ends of the range of operation rather than in the middle. These features are noticeable in Figures 5(a) and 5(b). The

TABLE 2: Summary of agreement results for Multi-interference (half-range)*.

SNR	ρ_c	AOC	Intercept	Slope	Lower 95%	Upper 95%
200	0.9928	0.0125	-0.0041	-0.0360	-0.0227	0.0146
100	0.9927	0.0126	-0.0041	-0.0361	-0.0234	0.0152
50	0.9921	0.0132	-0.0033	-0.0389	-0.0244	0.0178
25	0.9901	0.0149	-0.0037	-0.0371	-0.0310	0.0236

Due to the nonuniform differences in relation to the values of True PDFF, in order to obtain the estimated error value at a given fat fraction PDFF one must calculate intercept + slope*PDFF. For 95% limits of agreement, the calculation is the same with lower or upper limits replacing intercept. No outliers had to be removed for these calculations. ρ_c is best when close to 1 and AOC when close to 0.

TABLE 3: Summary of agreement results for ADALIFE.

SNR	ρ_c	AOC*	Mean	Lower 95%	Upper 95%	Outliers†
Half-range estimation						
200	0.9994	0.0021	0.0007	-0.0045	0.0060	1
100	0.9968	0.0027	0.0006	-0.0059	0.0073	3
50	0.9955	0.0040	0.0007	-0.0091	0.0106	1
25	0.9829	0.0076	0.0008	-0.0180	0.0197	3
Full-range estimation						
200	0.9998	0.0024	0.0001	-0.0065	0.0068	6
100	0.9998	0.0027	0.0001	-0.0070	0.0074	3
50	0.9998	0.0037	0.0000	-0.0096	0.0097	3
25	0.9991	0.0062	0.0001	-0.0162	0.0166	2

*The Kolmogorov-Smirnov test for differences between the Multi-interference and ADALIFE REC curves (half-range) confirmed that they differ with $p < 0.001$ at any SNR.

†Outliers do not apply to ρ_c and AOC.

REC curves plotted in (c) show visually the difference in area according to different SNR values. Finally, Figures 6(a) and 6(b) depict PDFF maps obtained using Multi-interference and ADALIFE, in color scale ranging from 0 to 1, the latter corresponding to 100% PDFF. For the same ROI, Multi-interference estimated 24.6% and ADALIFE estimated 25.2%.

5. Discussion

This study proposed a new method for hepatic proton density fat fraction estimation using magnitude MR signal. We simulated signals with parameters based on *in vivo* studies and PDFFs ranging from 0.1% to 45% or 100%. Fat fraction values in half range were then estimated by a state-of-the-art model, Multi-interference, and the proposed model, and in full range by ADALIFE only. The maximum value of 45% PDFF instead of 50% prevents Multi-interference from estimation errors that appear after that level, according to our simulations. Our first concern was if simulations would adequately reproduce *in vivo* results. In fact, comparing our Multi-interference simulation results to the original *in vivo* validation publications [19, 30], we observe that our repeatability, intercept, and slope results differ in less than 2% PDFF, which is within the confidence intervals provided in those studies.

The simulations demonstrated ADALIFE to be more reproducible and to have better agreement to the true PDFF. Despite the nonlinear trajectory around the mean that can be observed for ADALIFE estimation in Figure 4(b), the errors followed a *quasi*-normal distribution, except for the very few

outliers mentioned before. These points were far enough to obscure the normality of the curve; however, they were few enough to be discarded using a ± 5 SD rule. The trajectory can be associated with the nonlinear solution found by the neural network which averages close to zero. It is also present at other SNR values although overwhelmed by noise-derived results. The same applies when operating in full range as in Figures 5(a) and 5(b). Multi-interference presented a skewed and leptokurtic normal distribution of errors due to the bias in slope. This time, however, removing possible outliers based on the same ± 5 SD rule as applied for ADALIFE did not cause any improvement, so we considered that it would be best to keep the whole data set. Therefore, we recurred to linear regression in the BA analysis around which the residuals are normally distributed. This resulted in limits of agreement that are dependent on the fat fraction for Multi-interference and fixed for ADALIFE. Overall, these limits were narrower for our method, and more importantly, narrower at 0.0625 PDFF, a suggested threshold for classification of fatty liver according to a comprehensive study involving 2349 participants [20].

Full-range estimation showed a discontinuity at 55% PDFF. As we mentioned before, at that fraction the major fat peaks become dominant, what makes of it also a problematic point for other traditional methods such as Dixon [14] and triple-echo [17], where they start to confound water and fat. In our case it is a characteristic of the solution curve found by the MLP and should not be a problem as long as the associated deviation is acceptable. Results also showed that as SNR decreases, limits of agreement and AOC worsen at a higher pace for ADALIFE than for Multi-interference,

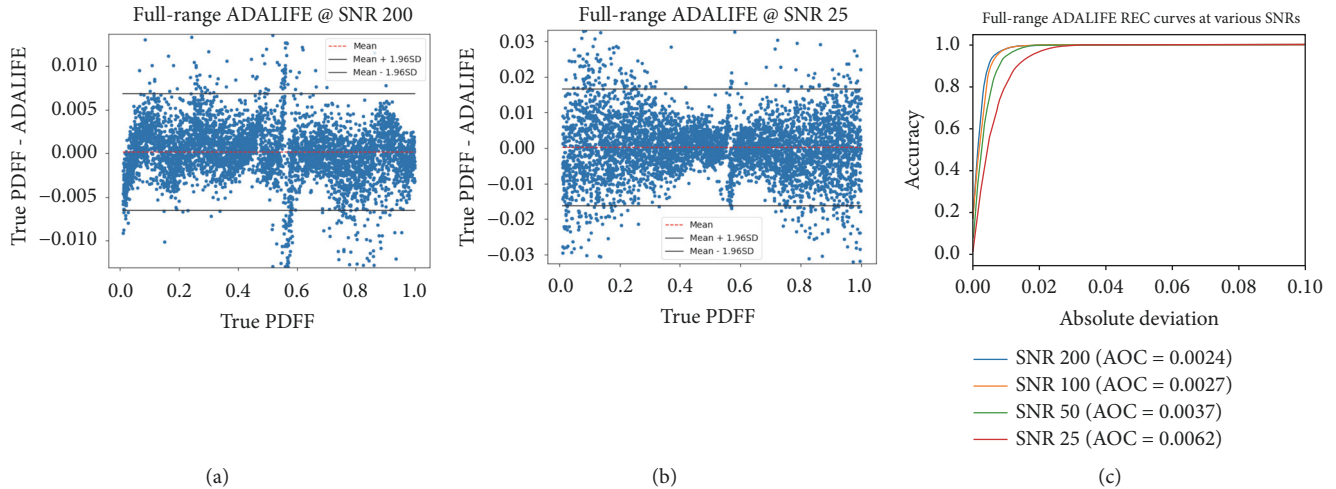


FIGURE 5: Bland-Altman plots (a) and (b) at SNR 200 and 25 and REC curves (c) at SNR 200, 100, 50, and 25 ADALIFE. Notice the difference in scales. Data points omitted on the plot due to scaling (out of 5000): (a) 43; (b) 12.

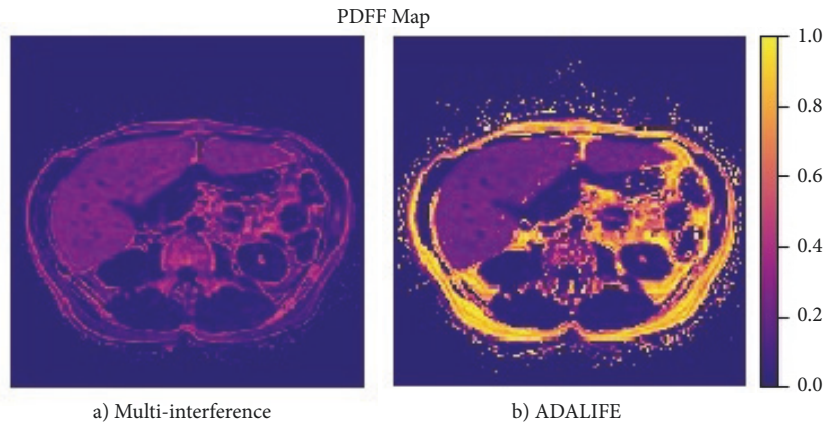


FIGURE 6: PDFF maps generated using (a) Multi-interference and (b) ADALIFE. For the same ROI, Multi-interference estimated 24.6% and ADALIFE estimated 25.2%. The scale on the image is decimal, corresponding to 0-100% PDFF.

indicating that the latter might be more robust to noise. However, ADALIFE performed better in absolute values.

The PDFF maps depicted in Figure 6 are not intended for a thorough image analysis, which shall be performed in a future work. Rather than that, they serve well as a good preview of the behavior of the methods when trained artificially and applied to *in vivo* images. Both methods present similar results within the liver area. Subcutaneous and intramuscular body fat also becomes easily distinguishable due to the full range of estimation of the proposed method, even though the dataset was trained using the fat spectrum of the liver, which may indicate they have common components. The ambiguity effect discussed earlier as a drawback of Multi-interference method can be also seen as this method reports these tissues with low PDFF in the color scale.

ADALIFE successfully estimated fat fraction values in full range with errors even smaller than the ones obtained in half range by Multi-interference, thus breaking the paradigm that magnitude signals do not provide enough information for water-fat separation at the low and high fat ends of

the PDFF spectrum. The key explored in ADALIFE is that the dephasing between fat peaks becomes significant in the resulting signal at longer echo times and by looking how it changes the signal geometrically. For instance, by analyzing only the first three echoes in all signals from Figure 2, one cannot easily distinguish low and high fat fractions, especially when there is added noise. While a nonlinear least squares approach tries to estimate the parameters of a reduced model of the signal in order to estimate PDFF, the proposed method models fat fraction as a direct function of changes in the signal. In fact, by consulting the matrix of weights generated by the training procedure, we noticed that the most important connections as depicted in Figure 3(b) are the ones between two consecutive points, corroborating with observations in Figure 2. We tried other combination of features such as only consecutive points, only distances, only angles, adding areas, and adding heights. The optimal combination was the one presented here. We also tried these features with other regression methods, starting with a simple decision tree (which worked loosely but was an indicator that the idea was

plausible) k-nn for regression, and different combinations of hidden layers and parameters for the MLP model. In literature we found a single previous work that proposed full-range PDFF estimation using two sets of magnitude MRI obtained with different flip angles [16], a technique with limited implementation as it requires a nontraditional acquisition. What we propose here relies on a standard protocol available at any clinical site and postprocessing that can be easily implemented offline with open-source software. The presented results will be validated using *in vivo* imaging in a future work.

The limitations of the liver fat model we used are in the sense that there is not much information about each fat moiety $T2^*$ values at different MR fields, especially the ones less prominent in human liver. These characterization studies are rather complex as they require large groups of volunteers. For the same reason we were not able to compare ADALIFE to models for full-range PDFF estimation based on complex signal rather than magnitude, as a realistic simulation of those would require even more biological parameters that have not been fully characterized in literature. A current limitation of ADALIFE is the need of seven echoes due to the late expression of phase accumulation of fat, although we believe this task can be performed with even less information. Application of ADALIFE is also limited in the sense that the descriptor that will train a regression method is obtained for a very specific set of acquisition parameters. On the other hand, the method was thought at first aiming clinical applications, which follow strict protocols. Considering that, changes in acquisition parameters would require retraining of the network. Future works should investigate a minimum number of echoes and improved robustness to noise. Also, ADALIFE estimates PDFF as a function of the geometrical characteristics of the signal, whereas the other methods first estimate tissue parameters that can be used for other calculations, including fat fraction. Nevertheless, we believe that the same principles behind ADALIFE should allow the regression of parameters other than PDFF, given adjustments on the training set. An example of application is the regression of $T2^*$ values, which is commonly used for calculation of liver iron content. As high resolution spectroscopy advances and maps the spectrum of other tissues in the human body, the proposed method can be extended for peripheral or visceral fat estimation, or musculoskeletal applications such as fatty degeneration.

Finally, ADALIFE is a signal shape descriptor capable of augmenting data and a relatively simple approach to the problem of fat estimation that can be used by regression models, here exemplified by a neural network model. It has an enormous potential of development considering current tendencies in machine learning that have been recently found applications in the field of medical imaging analysis.

6. Conclusions

We demonstrated how shape features of a MR signal of the liver can be used to train an artificial neural network and estimate fat fraction without additional acquisition techniques, a method we call ADALIFE to solve this open problem in liver

MR imaging. The simulation results compared repeatability and agreement between known PDFF values and estimations performed by Multi-interference and ADALIFE. Our method outperformed the state of the art in the simulations at every tested SNR. Future works aim at *in vivo* validation, improving robustness to noise and deep network modeling and expanding applications to other fatty tissues.

Data Availability

The numerical data used in simulations to support the findings of this study are available from the corresponding author upon request.

Disclosure

This paper is based on Yuri Ajala Da Costa's master thesis (Brazil, 2018), available at <http://www.teses.usp.br/teses/disponiveis/18/18152/tde-01102018-083519/en.php>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Brazilian National Council for Scientific and Technological Development (CNPq) under grant number 153070/2016-1 and by the University of Sao Paulo (USP) under grant number PUB 1582/2018.

References

- [1] Z. M. Younossi, A. B. Koenig, D. Abdelatif, Y. Fazel, L. Henry, and M. Wymer, "Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes," *Hepatology*, vol. 64, no. 1, pp. 73–84, 2016.
- [2] M. H. Le, P. Devaki, N. B. Ha et al., "Prevalence of non-alcoholic fatty liver disease and risk factors for advanced fibrosis and mortality in the United States," *PLoS ONE*, vol. 12, no. 3, Article ID e0173499, 2017.
- [3] D. Hernando, S. D. Sharma, M. Aliyari Ghasabeh et al., "Multi-site, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using a fat-water phantom," *Magnetic Resonance in Medicine*, vol. 77, no. 4, pp. 1516–1524, 2017.
- [4] T. Hayashi, K. Fukuzawa, H. Yamazaki et al., "Multicenter, multivendor phantom study to validate proton density fat fraction and $T2^*$ values calculated using vendor-provided 6-point DIXON methods," *Clinical Imaging*, vol. 51, pp. 38–42, 2018.
- [5] D. B. Parente, R. S. Rodrigues, F. F. Paiva et al., "Is MR spectroscopy really the best MR-based method for the evaluation of fatty liver in diabetic patients in clinical practice?" *PLoS ONE*, vol. 9, no. 11, Article ID e112574, 2014.
- [6] B. Kang, M. Kim, S. Song, D. w. Jun, and K. Jang, "Feasibility of modified Dixon MRI techniques for hepatic fat quantification

- in hepatic disorders: validation with MRS and histology,” *British Journal of Radiology*, no. 1089, Article ID 20170378, 2017.
- [7] D. Hernando, C. D. Hines, H. Yu, and S. Reeder, “Addressing phase errors in fat-water imaging using a mixed magnitude/complex fitting method,” *Magnetic Resonance in Medicine*, vol. 67, no. 3, pp. 638–644, 2012.
- [8] J. Kühn, D. Hernando, B. Mensel et al., “Quantitative chemical shift-encoded MRI is an accurate method to quantify hepatic steatosis,” *Journal of Magnetic Resonance Imaging*, vol. 39, no. 6, pp. 1494–1501, 2014.
- [9] D. Hernando, Z. Liang, and P. Kellman, “Chemical shift-based water/fat separation: a comparison of signal models,” *Magnetic Resonance in Medicine*, vol. 64, no. 3, pp. 811–822, 2010.
- [10] M. Bydder, T. Yokoo, G. Hamilton et al., “Relaxation effects in the quantification of fat using gradient echo imaging,” *Magnetic Resonance Imaging*, vol. 26, no. 3, pp. 347–359, 2008.
- [11] A. Triay Bagur, C. Hutton, B. Irving, M. L. Gyngell, M. D. Robson, and M. Brady, “Magnitude-intrinsic water–fat ambiguity can be resolved with multipeak fat modeling and a multipoint search method,” *Magnetic Resonance in Medicine*, vol. 82, no. 1, Article ID mrm.27728, pp. 460–475, 2019.
- [12] J. Cho and H. Park, “Robust water–fat separation for multi-echo gradient-recalled echo sequence using convolutional neural network,” *Magnetic Resonance in Medicine*, vol. 82, no. 1, pp. 476–484, 2019.
- [13] C. Liu, C. A. McKenzie, H. Yu, J. H. Brittain, and S. B. Reeder, “Fat quantification with IDEAL gradient echo imaging: correction of bias from T1 and noise,” *Magnetic Resonance in Medicine*, vol. 58, no. 2, pp. 354–364, 2007.
- [14] W. T. Dixon, “Simple proton spectroscopic imaging,” *Radiology*, vol. 153, no. 1, pp. 189–194, 1984.
- [15] G. Hamilton, T. Yokoo, M. Bydder et al., “In vivo characterization of the liver fat 1H MR spectrum,” *NMR in Biomedicine*, vol. 24, no. 7, pp. 784–790, 2011.
- [16] H. K. Hussain, T. L. Chenevert, F. J. Londy et al., “Hepatic fat fraction: MR imaging for quantitative measurement and display—early experience,” *Radiology*, vol. 237, no. 3, pp. 1048–1055, 2005.
- [17] B. Guiu, J.-M. Petit, R. Loffroy et al., “Quantification of liver fat content: comparison of triple-echo chemical shift gradient-echo imaging and in vivo proton MR spectroscopy,” *Radiology*, vol. 250, no. 1, pp. 95–102, 2009.
- [18] R. Longo, P. Pollesello, C. Ricci et al., “Proton MR spectroscopy in quantitative in vivo determination of fat content in human liver steatosis,” *Journal of Magnetic Resonance Imaging*, vol. 5, no. 3, pp. 281–285, 1995.
- [19] T. Yokoo, M. Bydder, G. Hamilton et al., “Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T,” *Radiology*, vol. 251, no. 1, pp. 67–76, 2009.
- [20] L. S. Szczepaniak, P. Nurenberg, D. Leonard et al., “Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population,” *American Journal of Physiology-Renal Physiology*, vol. 288, no. 2, pp. E462–E468, 2005.
- [21] B. Guiu, R. Loffroy, J. Petit et al., “Mapping of liver fat with triple-echo gradient echo imaging: validation against 3.0-T proton MR spectroscopy,” *European Radiology*, vol. 19, no. 7, pp. 1786–1793, 2009.
- [22] H. Gudbjartsson and S. Patz, “The Rician distribution of noisy MRI data,” *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 910–914, 1995.
- [23] T. Yokoo, Q. Yuan, J. Sénégas, A. J. Wiethoff, and I. Pedrosa, “Quantitative R2* MRI of the liver with rician noise models for evaluation of hepatic iron overload: simulation, phantom, and early clinical experience,” *Journal of Magnetic Resonance Imaging*, vol. 42, no. 6, pp. 1544–1559, 2015.
- [24] Z. Wen, S. B. Reeder, A. R. Pineda, and N. J. Pelc, “Noise considerations of three-point water-fat separation imaging methods,” *Medical Physics*, vol. 35, no. 8, pp. 3597–3606, 2008.
- [25] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 167–176, 2014.
- [26] J. Martin Bland and D. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement,” *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.
- [27] L. I-Kuei Lin, “A concordance correlation coefficient to evaluate reproducibility,” *Biometrics*, vol. 45, no. 1, pp. 255–268, 1989.
- [28] J. S. Krouwer, “Why Bland–Altman plots should use $(Y+X)/2$ when X is a reference method,” *Statistics in Medicine*, vol. 27, no. 5, pp. 778–780, 2008.
- [29] J. Bi and K. P. Bennett, “Regression error characteristic curves,” in *Proceedings of the 20th International Conference on Machine Learning (ICML '03)*, pp. 43–50, Menlo Park, Calif, USA, August 2003.
- [30] T. Yokoo, M. Shiehorteza, G. Hamilton et al., “Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T,” *Radiology*, vol. 258, no. 3, pp. 749–759, 2011.

