



Published in final edited form as:

Ultrasound Med Biol. 2021 March ; 47(3): 556–568. doi:10.1016/j.ultrasmedbio.2020.10.025.

Use of a convolutional neural network and quantitative ultrasound for diagnosis of fatty liver

Trong N. Nguyen^{a,b}, Anthony S. Podkova^{a,b}, Trevor H. Park^c, Rita J. Miller^{a,b}, Minh N. Do^a, Michael L. Oelze^{a,b,*}

^aBeckman Institute for Advanced Science and Technology, 405 N Mathews, Urbana, IL, 61801

^bDepartment of Electrical and Computer Engineering, 306 N. Wright St, Urbana, IL, 61801

^cDepartment of Statistics, 725 S Wright St, Champaign, IL, 61820

Abstract

Quantitative ultrasound (QUS) was used to classify rabbits that were induced to have liver disease by placing them on a fatty diet for a defined duration and/or periodically injecting them with CCl₄. The ground truth of the liver state was based on lipid liver percents estimated via the Folch assay and hydroxyproline concentration to quantify fibrosis. Rabbits were scanned ultrasonically *in vivo* using a SonixOne scanner and an L9–4/38 linear array. Liver fat percentage was classified based on the ultrasonic backscattered radio-frequency (RF) signals from the livers using either QUS or a 1D convolutional neural network (CNN). Use of QUS parameters with linear regression and canonical correlation analysis (CCA) demonstrated that the QUS parameters could differentiate between livers with lipid levels above or below 5%. However, the QUS parameters were not sensitive to fibrosis. The CNN was implemented by analyzing raw RF ultrasound signals without using separate reference data. The CNN output the classification of liver as either above or below a threshold of 5% fat level in the liver. The CNN outperformed the classification utilizing the QUS parameters combine with a support vector machine (SVM) in differentiating between low and high lipid liver levels, i.e., accuracies of 74% versus 59% on the testing data. Therefore, while the CNN did not provide a physical interpretation of the tissue properties, e.g., attenuation of the medium or scatterer properties, the CNN had much higher accuracy in predicting fatty liver state and did not require an external reference scan.

Keywords

Quantitative Ultrasound; Liver Disease; Machine Learning

*Corresponding Author: Michael L. Oelze, 405 N Mathews, Urbana, IL, 61801; oelze@illinois.edu; Phone, 217-333-9226.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Management of liver disease, including fatty and fibrotic liver, is an important clinical problem. Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease in the United States (Wieckowska and Feldstein, 2008). With up to a third of the United States population affected by NAFLD (70–90% of obese or Type 2 diabetic patients have NAFLD), NAFLD represents a significant medical concern. Inflammation associated with NAFLD can also lead to liver fibrosis and other liver diseases. Liver fibrosis is the cause of chronic damage to the liver including infections, toxins, and autoimmune disorder, and it can lead to cirrhosis and hepatocellular carcinoma.

NAFLD is often used to describe a range of liver conditions: fatty liver, steatohepatitis, advanced fibrosis, cirrhosis (Angulo, 2002). Fibrosis is the result of hepatic injury, which causes the gradual replacement of hepatocytes by extracellular matrix proteins such as collagens (Bataller and Brenner, 2005). According to the Center for Disease Control (CDC), cirrhosis, the end stage of fibrosis, affects 900,000 patients in the US, resulting in 30,000 deaths per year (CDC, 2019). Evaluation of hepatic fibrosis degree is essential for prognosis, surveillance and treatment decisions in patients with chronic liver disease. Early detection of fibrosis is clinically important because fibrosis is potentially reversible if caught at an early stage. Liver biopsy, currently the gold standard for assessing liver fibrosis and steatosis, suffers from sampling errors, inter-observer variability, and invasiveness, and has potential complications including morbidity and mortality.

There remains an unmet clinical need to develop imaging techniques for the non-invasive evaluation of liver steatosis and fibrosis. Ultrasound is an attractive imaging modality because it is safe, real time, portable and inexpensive. Different methods using ultrasound have been explored to detect fibrosis and steatosis. Conventional B-mode image features were shown to be inaccurate predictors of early and significant liver fibrosis (Choong et al., 2012). Because liver stiffness increases with the degree of fibrosis, transient elastography (TE) and acoustic radiation force impulse imaging have been successfully demonstrated to quantify liver stiffness and to correlate estimates with the degree of fibrosis (Sandrin et al., 2003; Ziol et al., 2005). Other ultrasonic methods based on echo amplitude distribution, such as the acoustic structure quantification (Tsui et al., 2016), can be used to quantify liver fibrosis via the Nakagami parameter. The Nakagami parameter, which quantifies the type of scattering conditions based on the envelope distribution, decreased with an increase in the histological fibrosis stage (which is inversely proportional to the severity of liver fibrosis). Attenuation was also used as an indicator of fibrosis. Suzuki et al. (1992) looked at the fatty and fibrotic rabbit liver model independently and found that attenuation and hydroxyproline scores had a positive correlation (0.87). Based on histological grading, Lin et al. (1988) observed an increase of attenuation coefficient for increased fat infiltration *in vitro* in humans and a less pronounced increase of attenuation for higher grades of fibrosis (0.63 ± 0.16 , 0.83 ± 0.26 and 0.87 ± 0.12 dB cm⁻¹ MHz⁻¹ for grade 1, 2 and 3, respectively). More recent studies have evaluated quantitative ultrasound (QUS) techniques in humans for classifying liver disease with accuracies of 55.0% and 68.3% based on attenuation and backscatter coefficient (BSC), respectively, when compared to MRI-estimated proton density fat fraction (PDFF) (Paige et al., 2017).

The first goal of the current study is to evaluate QUS as a noninvasive method to quantitatively assess liver fibrosis and steatosis in an *in vivo* rabbit model. The second goal of the study is to evaluate using the raw ultrasound backscattered signals to classify liver state without taking a reference spectrum. In a previous study, we demonstrated that QUS techniques could detect the presence of steatosis in a rabbit model of fatty liver alone with a classification accuracy of 84.11 % (Nguyen et al., 2019). In this study, we explored the effects of fibrosis and fat in the liver on QUS analysis. Independent correlations between QUS parameters and lipid percentages and fibrosis scores were computed. Correlations between the QUS parameters and linear combinations of lipid percentages and fibrosis scores were also calculated via canonical correlation analysis (CCA) to determine the relationship between QUS parameters and the codependency of steatosis and fibrosis.

The system settings were kept unchanged when acquiring rabbits liver scans. Therefore, we hypothesized that a 1D convolutional neural network (CNN) could separate the system-dependent signal from the tissue-dependent signal, and perform classification in a reference-free manner. A CNN learns a nonlinear mapping from the input to the output via stacking of multiple connected convolutional filter layers at different resolutions (Zeiler and Fergus, 2014). The learned features from the convolutional layers can then be concatenated into a vector and classified by fully connected layers. In the case of ultrasonic tissue characterization, the problem can be formulated as a supervised learning strategy using a CNN where the input is the backscattered RF data and the output is the pathological indicator (e.g. fatty/non-fatty) when the task is classification or the degree of fatty liver (in lipid percentage) or the fibrosis scores when the task is regression.

To our knowledge, no published work exists comparing CNNs for the problem of fatty liver classification using ultrasonic data to QUS-based classifiers. Treacher et al. tested 100 different CNN architectures using B-mode images as inputs and showed that there was no substantial predictive power to detect fibrosis (Treacher et al., 2019). The traditional spectral-based QUS approach does not utilize the phase information in the RF signal, because only the magnitudes of the power spectra from RF data were computed. We hypothesized that a CNN could extract classification power from the lost phase information from the time-domain RF and perform feature extraction and classification simultaneously. In this work we compare the reference-free CNN approach with more traditional QUS approaches that require a reference scan to classifying liver state. The CNN is reference free in the sense that no calibration data from a reference phantom are used with the CNN because the same system and settings were used to collect all data. The CNN is hypothesized to differentiate between the tissue signal and the signal properties from the system. However, if a different ultrasonic system, probe, setting or frequency range were used, the CNN would need to be retrained using the new configuration. If it is feasible to remove a reference scan from the process of liver classification, this could reduce the number of scan steps during the busy clinical work flow.

Materials and Methods

Animal Procedures

The protocol was approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Illinois at Urbana-Champaign. Sixty male New Zealand White rabbits were used in the study. The study was a 3×5 factorial design with rabbits on a fatty diet for 0, 1, 2, 3 or 6 weeks and injected with carbon tetrachloride (CCl_4) to induce fibrosis. Initially, rabbits were injected for 11 weeks at concentrations of 0, 0.2, 0.4 or 0.6 mL/kg (30 rabbits). However, three rabbits died during the study and an alternate injection protocol was required, via consultation with IACUC, for the remaining 30 rabbits, which consisted of weekly CCl_4 injections for 8 weeks at a concentration of 0, 0.035 or 0.07 mL/kg. The premature deaths were eventually ruled idiosyncratic. The rabbits were on fatty diet during the 8 week or 11 week period. During an injection, CCl_4 and olive oil were emulsified. The rabbits received weekly injections based on their weights. At the end of each 8 or 11 weeks, the rabbits were returned to normal diet and received no injection for about a week before ultrasonic scanning. While changing the injection protocol in the middle of the study is not ideal, labels for fibrosis were based on the hydroxyproline assay, such that the changes in injection protocols were still tied directly to the assay results and could be correlated to QUS parameters.

Chemical Assay Procedures

Immediately following euthanasia, the liver was removed *en mass*. A portion was fixed in neutral buffered formalin, embedded in paraffin, sectioned, and stained with hematoxylin and eosin for histopathological analysis by a board-certified pathologist. Another portion was flash frozen in liquid nitrogen and stored at -80°C for use in the Folch assay (Folch et al., 1957) and hydroxyproline assay using an hydroxyproline assay kit (Sigma Aldrich, St. Louis, MO). The Folch and the hydroxyproline assays were used to quantify the lipid levels and fibrosis in the liver, respectively. In the Folch assay, liver tissue was homogenized, and lipid content was extracted using chloroform and methanol (Ghoshal et al., 2012). Fat content was then expressed as the liver lipid percent. The hydroxyproline assay was used to estimate the collagen content as a measure of fibrosis in the liver. Liver tissue was homogenized and hydrolyzed overnight in 12 M HCl. The hydrolysate was evaporated and color was developed on the hydrolysate by reaction of oxidized hydroxyproline with 4-Deimethylamini benzaldehyde and the absorbance was read at 562 nm. The hydroxyproline level was expressed in mg/g of liver.

Ultrasonic Scanning Procedures

Before scanning with ultrasound, rabbits were anesthetized using isoflurane gas. The skin area above the liver was shaved and depilated prior to scanning to improve coupling of the ultrasound. Warm ultrasound gel was also placed on the skin surface to improve coupling. The liver was scanned *in vivo* with an L9-4/38 transducer using the SonixOne system (Analogic Corporation, Boston, MA, USA) providing an analysis bandwidth of 3 to 6 MHz. Fifty frames of post-beamformed RF data sampled at 40 MHz were acquired for each rabbit and saved for offline processing. A well-characterized reference phantom was scanned using the same system and system settings for calibration of the BSC and attenuation estimation

(Yao et al., 1990). Following scanning, the rabbits were euthanized via CO₂ while still under anesthesia.

Quantitative Ultrasound Procedures

In our supervised learning task, the input was the RF data and the output was either the classification or regression of liver lipid and fibrosis. The classification was useful when the summary statistics of the QUS parameters (i.e., slope and mid-band fit of attenuation curve) were meaningful and could absorb the errors incurred in the Folch and hydroxyproline assay results, where the regression alone might not be able to explain these experimental errors.

The desired outputs are the lipid percentage from the Folch assay and the hydroxyproline level from the hydroxyproline assay. Pathology descriptions were not incorporated into the classifier. The binary classification task of fatty liver was based on a 5 percent threshold, which is the median of all lipid percentages from the group of rabbits: high lipid ($y = 1$) if the lipid percentage is greater than or equal to 5 percent, low lipid ($y = 0$) if the lipid percentage is smaller than 5 percent. The median value was chosen to get an equal number of rabbits in each class to reduce the bias of the classifier. To obtain a complete picture of the relationship between QUS parameters and lipid and fibrosis scores, individual correlations were computed and the regression problem was also considered.

The attenuation and BSC curves were extracted from raw RF data by manually segmenting the liver regions from the B-mode images of the livers, i.e., regions of interest were chosen for each frame. Figure 1 shows a representative B-mode image of a liver along with the segmented area. The segmentation of the ROI can occur throughout the field of the image; however, segmentation occurred in regions that visibly appeared homogeneous in scattering, that were not shadowed by ribs or other structures and with care to omit large vessels. Segmentation was limited to depths between 0.5 and 3 cm depth. The liver regions were selected such that they were always further than 0.5 cm from the transducer surface. Each region of interest was divided into various data blocks of size of 15 by 15 wavelengths (5.1 mm by 5.1 mm) of the center frequency of the array probe, i.e., 4.5 MHz. Each data block had a 75% overlap with other data blocks. The BSC was calculated for each data block using the reference phantom method (Yao et al., 1990). The attenuation curve in each data block was estimated using the spectral log difference method and averaged over all blocks in an image frame and over all image frames for a rabbit to have a mean attenuation curve for each rabbit liver (Yao et al., 1990; Parker, 1983). A slope and a mid-band fit at 4.5 MHz were estimated from the fitted line to the average attenuation curve. Effective scatterer diameters (ESDs) and effective acoustic concentrations (EACs) were derived from the BSC curves using a spherical Gaussian scattering form factor (Oelze et al., 2002; Oelze and Mamou, 2016). Additionally, correlations between each QUS parameter and the lipid percentages and hydroxyproline levels were computed.

A linear regression of the four QUS parameters (ESD, EAC, attenuation slope, attenuation midband fit) was performed individually versus lipid and fibrosis scores. Specifically, linear regression of lipid percentage y_i^1 and the hydroxyproline level y_i^2 to QUS parameters vector

$\mathbf{x}_i = (x_i^1, x_i^2, x_i^3, x_i^4)$, where x_i^1 is the ESD, x_i^2 is the EAC, x_i^3 is the attenuation slope and x_i^4 is the attenuation midband-fit:

$$\mathbf{w}_1, b_1 = \operatorname{argmin}_{\mathbf{w}_1, b_1} \sum_i \|y_i^1 - \mathbf{w}_1 \mathbf{x}_i^T - b_1\|^2, \quad (1)$$

$$\mathbf{w}_2, b_2 = \operatorname{argmin}_{\mathbf{w}_2, b_2} \sum_i \|y_i^2 - \mathbf{w}_2 \mathbf{x}_i^T - b_2\|^2, \quad (2)$$

where i is the index of the rabbits' ID.

Canonical correlation analysis (CCA) (Hardoon et al., 2004) was also performed to estimate the correlation of the linear combination of the four QUS parameters with the linear combination of the lipid and fibrosis, which would provide a value describing the codependence of these two outputs. Let $\mathbf{y}_i = (y_i^1, y_i^2)$ be a vector of lipid and hydroxyproline values from the rabbits, the vector $\mathbf{S}_x \mathbf{w}_x^T$ is the weighted linear combination of the input QUS parameters where \mathbf{w}_x represents the weights, and the vector $\mathbf{S}_y \mathbf{w}_y^T$ is the weighted linear combination of the output parameters:

$$\mathbf{S}_x \mathbf{w}_x^T = (\mathbf{x}_1 \mathbf{w}_x^T, \dots, \mathbf{x}_n \mathbf{w}_x^T), \quad (3)$$

$$\mathbf{S}_y \mathbf{w}_y^T = (\mathbf{y}_1 \mathbf{w}_y^T, \dots, \mathbf{y}_n \mathbf{w}_y^T). \quad (4)$$

Then, the CCA finds the weights \mathbf{w}_x and \mathbf{w}_y such that the correlation between $\mathbf{S}_x \mathbf{w}_x^T$ and $\mathbf{S}_y \mathbf{w}_y^T$ is maximized:

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} \operatorname{corr}(\mathbf{S}_x \mathbf{w}_x^T, \mathbf{S}_y \mathbf{w}_y^T) = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\langle \mathbf{S}_x \mathbf{w}_x^T, \mathbf{S}_y \mathbf{w}_y^T \rangle}{\|\mathbf{S}_x \mathbf{w}_x^T\| \|\mathbf{S}_y \mathbf{w}_y^T\|}, \quad (5)$$

where \mathbf{w}_x and \mathbf{w}_y can be considered as the axes where the projections of \mathbf{x}_i and \mathbf{y}_i have maximum correlations. The term n denotes the total number of rabbits.

Convolutional Neural Network Architecture

The model-based QUS approach requires the use of the reference phantom to derive the BSCs and the attenuation from power spectra calculated from the RF data. In this study, we explored the use of the CNN approach that did not require the use of a reference phantom or a model. Furthermore, the ultrasonic scanner settings used in all of the rabbits scans were the same, the differences manifested in the RF data were hypothesized to come solely from the fatty diet and/or the CCl_4 injections. Using a CNN, the feature extraction and classification can be accomplished simultaneously through the concatenation of the convolutional layers and the fully connected layers.

Because the CNN involves stacking of multiple layers of convolutions with many parameters, large amounts of data are required to train the network. Unfortunately,

biological data is expensive and time-consuming to acquire; therefore, the number of animals in the study was small. Thus, to prevent overfitting of the CNN classifier, in conjunction with applying regularization techniques such as batch normalization, drop-out and early stopping, only the problem of binary classification was considered: i.e., classification of high lipid vs. low lipid. In each training fold division, there were approximately 100,000 training examples and 10,000 testing examples (each sample is one RF line) while the number of weight parameters in the CNN were 2178, so the network could generalize to new data. Note that the CNN was used to classify from the liver images only, without the help of reference signals to remove the system effects.

The same gated RF lines inside the liver segmentation when extracting the QUS parameters were used as inputs to the CNN. The length of each RF signal data segment was 5.1 mm, corresponding to 15 wavelengths axially. Only the RF data of the liver images were employed, the RF data from the reference phantom were not utilized. The RF data and the corresponding labels from the Folch assay were collected to form a data set to train and test the CNN. Out of 57 rabbits, five rabbits had very low SNR images where we could not segment an ROI of 15 by 15 wavelengths, so they were not included in the classification. Six-fold cross validation was used to assess the model performance. In each fold, 52 rabbits were randomly divided into 44 rabbits for training and 8 rabbits for testing, and the process repeated five additional times. In each testing fold, the number of rabbits in each class were kept balanced: four rabbits were of class 1 and four rabbits were of class 0. The accuracy was evaluated on a frame by frame basis. The predicted label of each frame was the majority of the predicted labels in all ROIs in that frame. The average accuracy of the classifier was computed across all frames.

The CNN consisted of stacking successive layers of convolutions, downsampling and dropout. Each hidden convolutional layer in the CNN applied different convolutional filters to the previous layer output,

$$f_i^{n+1}[m] = \sigma \left(\sum_{k=1}^N f_i^n[m-k] h_i^n[k] \right), \quad (6)$$

where $f_i^n[m]$ is the i -th features map at layer n , $h_i^n[k]$ is the i -th-filter to be learned at layer n . In the hidden layer n , different numbers of filters were designed to extract different features from the data. The activation function $\sigma(x)$ applies nonlinearity to output of the convolution, and effectively introduces nonlinearity in the decision boundaries of the feature space. In this work, we utilized the rectified linear unit (RELU) function as an activation function:

$$g(x) = \max(0, x). \quad (7)$$

The RELU activation function suppresses the convolution output that is smaller than zero.

The first layer of the network was the gated RF input from a data block selected from the image of the sample. The filter coefficients, $h_i^n[k]$, which are learned during the training phase via backpropagation, extract different features from the RF signal. Due to its ability to escape local minima of the cost function, Adam optimizer (Kingma and Ba, 2014) was

employed to perform gradient descent on the loss function. The loss function used for the binary classification was the cross-entropy between the ground truth and the predicted output:

$$L = - \sum_{i=1}^N y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i), \quad (8)$$

where y_i is the true output which takes the values of 1 for high lipid or 0 for low lipid, \hat{y}_i is the current output of the forward pass. The summation is over all training examples. The final output node used a sigmoid function to suppress the output into a scalar between 0 and 1 (\hat{y}_i in Equation 8). When the output of the sigmoid function is greater than 0.5, the input is classified as high lipid, otherwise low lipid.

In practice, instead of using all the training examples to calculate the gradients for updating the weights, a small batch size of 16 or 32 examples was used to update the network weights. This approach was empirically determined to accelerate training and did not affect the accuracy (Masters and Luschi (2018)). During testing, the unseen RF input was passed through the network to get an output prediction and accuracy was calculated based on the predicted output and the true output.

The network architecture used in this study resembles the VGG (named after Visual Geometry Group, University of Oxford) network architecture in (Simonyan and Zisserman, 2014), where downsampling at each layer was employed to reduce the dimensions of the feature space. Downsampling or max-pooling was applied by keeping the maximum values inside a sliding window across the hidden layer output map. The idea of concatenation of convolution, nonlinearity and max pooling was to select only the features in the input that strongly contributed to the output prediction. The initial weights of the convolutional filters were randomized before training. To prevent feature weights drifting and exploding, batch normalization (Ioffe and Szegedy, 2015) was also used to ensure the weights at each layer had zero mean and unit variance.

The CNN is composed of four hidden convolutional layers, four pooling layers, two fully connected layers and a four softmax output layer (see Table 1). After the fourth layer, all the output features were concatenated to get a feature vector of 54 features. Then, two fully connected layers were applied to those features to transform the features into a two-class classification. The two fully connected layers reorder the extracted features and thresholding is applied (via the RELU function) to arrive at the final prediction. To prevent overfitting, dropout was used at the fully connected layers, which randomly sets the node output to zero with a predefined probability (0.5 was used in this study). Dropout helps redistribute the weights (importance of features) to other parts of the networks, since only 50% of the weights are nonzero during training. Because node outputs are randomly dropped-out, only a few of them are essential to classification, effectively reducing the dimension of the final classifier (or reducing overfitting).

Results

Hydroxyproline Levels and Lipid Percentages

Table 2 provides the hydroxyproline levels for the different injection protocols that were used. The data indicate that the second protocol resulted in less fibrosis of the liver. Figure 2 shows the plots of hydroxyproline vs. lipid percentage for five different diet groups: 0, 1, 2, 3 and 6 weeks. The blue points are the rabbits without injections and the orange points are the rabbits with injections. The rabbits without injections rapidly developed fat in the liver over an increasing number of weeks, while the lipid in the rabbits receiving the CCl₄ injections increased more slowly over the weeks on diet. For example, for the 3 weeks diet group, all the rabbits with CCl₄ injections had lipid values less than 7% while the rabbits without injections had lipid values greater than 10%. These results suggest that the injections of CCl₄ resulted in increased fibrosis of the liver but may have slowed the accumulation of fat in the liver, which has also been observed in a mouse model combining a high fat diet with CCL4 injections (Kubota et al., 2013)

QUS Parameters

Treating fibrosis as an unobserved variable, we seek to classify the rabbits into two groups of steatosis using the four QUS parameters. There were 26 rabbits in the low lipid group and 30 rabbits in the high lipid group. We had to remove one rabbit from the study, which had very low SNR in the acquired RF, thus the segmented liver ROI was too small for the QUS estimation.

Figures 3 and 4 provide plots of the averaged BSC and attenuation curves for two classes with threshold of 5%. The attenuation curves had more differentiating power than the BSC curves. Figure 5 shows B-mode images of three rabbit livers for a range of lipid liver levels. The images show that as the lipid levels increase that attenuation in the B-mode appears to increase. Accompanying the B-mode figures are graphs of the BSC and attenuation estimates for the three rabbits. Differences in both the BSC and attenuation are observed as the liver lipid levels increase. Statistically significant differences (p-value < 0.05) were observed for the attenuation slope and attenuation midband-fit values between the high and low lipid livers. When incorporating rabbits with both fatty diet and CCl₄ injection, ESD and EAC were not strongly correlated to lipid changes. Using the BSCs, or its derived features ESD and EAC, did not result in the ability to differentiate between the low and high lipid level groups. Table 3 lists the averaged ESD, EAC, attenuation slope, attenuation midband-fit and their p-values for differentiating between the two lipid level classes.

Figure 6 plots the linear regression of the combined four QUS parameters fit to the lipid percentages. The estimated lipid percentage levels were typically higher than the actual lipid levels for rabbits with low lipid level percentage and lower for rabbits with a high lipid level percentage. The coefficient of determination r^2 was 0.69, suggesting the QUS parameters can linearly track the lipid changes. On average, the predicted lipid percentage and the ground truth lipid percentage differed by 2%. Table 4 shows the coefficients of linear regression and their corresponding p-values. Attenuation slope and mid-band fit were linearly correlated to lipid and fibrosis (p-values < 0.05) while ESD and EAC were not.

Similarly, Fig. 7 shows the regressed hydroxyproline output to the four QUS parameters. The r^2 coefficient was 0.20, which indicates that the QUS parameters could not explain the variances in the hydroxyproline output. The lack of discrimination using the spectral-based QUS approaches indicates that these parameters were not sensitive to the degree of liver fibrosis.

CCA analysis

Table 5 lists the CCA analysis results of QUS parameters and lipid/fibrosis scores: their weights w_x and w_y and the correlation between their linear combination. There is a strong correlation between the linear combinations of QUS parameters and the linear combination of lipid and fibrosis in the first canonical dimension (correlation of 0.71). The linear combination of QUS parameters in the first canonical dimension is

$$X_{cca} = w_x x^T = 1.73x_1 + 0.58x_2 + 0.0123x_3 + 0.0098x_4, \quad (9)$$

where x_1 is the attenuation slope and x_2 is the attenuation midband fit, x_3 is the ESD, and x_4 is the EAC. This variable X_{cca} is the horizontal axis in Fig. 8. The output linear combination is $0.2408 \times y_1$ (lipid percentage) $-0.1285 \times y_2$ (hydroxyproline score), which is the vertical axis in Fig. 8. This linear transformation of four QUS parameters and two output scores gives the maximum correlation of the two axes in Fig. 8. The second canonical dimension had a low correlation coefficient of 0.26 and is not plotted.

Figure 8 plots the first canonical correlation of the weighted QUS parameters and weighted lipid and hydroxyproline level. The opposite signs of the weights of the lipid percentage and hydroxyproline level in Table 4 suggest the competing effect between lipid and fibrosis, similar to Fig. 2.

CNN for lipid classification

The QUS parameters correlated well with the lipid changes, but not with the fibrosis changes. Thus, to be comparable to the QUS approach, the CNN was trained only on the task of fatty liver classification. It is understood that the fibrosis induced in the liver would affect the accuracy of the CNN classifier, and will be the subject of future work. Table 6 shows the classification results of 1D CNN for classifying two classes: low and high lipid with a threshold of 5%. Table 7 lists the accuracies when using the QUS approaches with four parameters: ESD, EAC, attenuation slope, attenuation midband fit. To classify the lipid classes using QUS approaches, a kernel support vector machine (SVM) was used. The results indicate that the CNN outperforms the QUS approach for classification of lipid changes. Figure 9 plots the ROC curves for the six folds training on the test data. The ROC curves reflects that the CNN did not generalize well to the some testing fold due to our limited dataset.

Discussion

The objective of the study was to investigate the relationship between QUS parameters and liver fibrosis and steatosis through noninvasive ultrasonic interrogation and the application of a CNN to the problem of liver classification allowing both a model-free analysis and

reference-free scanning configuration. The rabbits were divided into different groups based on different levels of CCl₄ injections and five groups maintained on a fatty diet over different durations. The hydroxyproline and lipid results reported in Fig. 2 indicate that the injection of CCl₄ inhibited the fat accumulation in the liver from the fatty diet but also produced fibrosis. According to histological analysis from the pathologist, none of the rabbits developed cirrhosis or heavy fibrosis. Some of the injections were small dose, which allowed the rabbits to recover from toxicological insult before observable fibrosis occurred. The fibrosis scored for the rabbits ranged from zero to mild to moderate.

Attenuation increased for both increased fibrosis and increased steatosis, indicating that the combined effects of both in general will make the attenuation increase. However, attenuation slope and midband fit were more sensitive to lipid liver percentages and much less so to fibrosis. ESD and EAC were not sensitive to lipid and fibrosis. Comparing the results of this study with our previous study on *in vivo* QUS analysis of fatty livers in rabbits showed similar trends but important differences (Nguyen et al., 2019). Specifically, in the previous study, the ESD decreased from 167 μm to 119 μm and the EAC increased from 21.1 dB to 31.8 dB between low and high lipid livers. These trends match the absolute numbers in the current study and are similar to results provided in Table 3, i.e., ESD decreased from 127 μm to 117 μm and EAC increased from 33.9 dB to 35.5 dB. Similar trends were observed with the attenuation estimates where in the previous study the attenuation slope increased from 0.6 dB/MHz/cm to 1.09 dB/MHz/cm and in the present study the attenuation slope increased from 0.69 to 0.97 dB/MHz/cm. An important difference between the current study and the previous study was the threshold set between the high and low lipid levels. In the previous study half of the rabbit livers were above 9% lipid liver content and half below. In this study, half of the rabbit livers were below 5% threshold and half above. These findings suggest that QUS techniques perform better at differentiating between a high and low lipid liver level if that threshold is higher. However, in order to maintain a balanced number of samples in the high and low lipid level groups and to reduce bias in the classifier we used the 5% threshold in the current study. Hence, QUS techniques could be used to identify and characterize steatosis noninvasively but would be less successful at detecting fibrosis, which has also been observed in other studies (Suzuki et al., 1992; Treacher et al., 2019). It is worth noting that in a recent study QUS parameters were able to identify changes in liver fibrosis in the same rabbit model but when using high ultrasonic frequencies, i.e., 20 MHz and 40 MHz probes in excised liver samples (Franceschini et al., 2019). On the other hand, because the QUS parameters at clinical frequencies are much less sensitive to fibrosis, they can be used to identify steatosis and be combined with other diagnostic approaches that are sensitive to fibrosis, such as TE or shear wave elastography.

In a specific case where the QUS parameters did not track well the lipid changes (rabbit IDs 12 and 45) might be due to the hidden variable of fibrosis. Rabbit 12 had a hydroxyproline value of 0.93 mg/g and moderate to marked periportal fibrosis with occasionally bridging fibrosis. Rabbit 45 had a hydroxyproline level of 1.84 mg/g of liver and mild but extensive fibrosis. If the number of weeks (0, 1, 2, 3, 6) and injection indicator information (with injection = 1, without injection = 0) were included as additional explanatory variables in the linear regression, the r^2 coefficient was increased to 0.76 compared to 0.69 without including injection.

Although the fibrosis has a limiting effect on the lipid development which implicitly impacts the predictive power of the QUS parameters, there was still a positive correlation of the QUS parameters to lipid changes. A median of the lipid percentages of 5% was chosen as a threshold for classification between low liver lipid and high liver lipid. This threshold was chosen because the value placed an equal number of rabbits in the high and low liver lipid classes. However, CCA analysis in Fig. 8 suggested a more appropriate threshold of 9% for determining two classes.

We also employed a 1D CNN to characterize the liver state using the raw RF signals acquired from the liver. It was empirically shown that the CNN can classify steatosis without using a model for scattering and without using the reference phantom when the system settings of all the scans were the same. The CNN was not used to detect fibrosis because there was no correlation using the QUS approach and consequentially we had no baseline for comparison.

To test the feasibility of the CNN in classifying the lipid without using the reference phantom, a general convolutional architecture commonly used in computer vision tasks was adapted to classifying the RF signal. An LSTM network (Zeyer et al., 2017) is hypothesized to be more capable of capturing temporal dependency such as speech signals or in our case RF data. However, we observed lower accuracy when using a LSTM network on our RF data. This might be due to the shorter gated window lengths that we used in our CNN for fair comparison with the QUS approach. More recent developed architectures like ResNet (He et al., 2016) or DenseNet (Huang et al., 2017) might be tested in future work; however, those newer methods require more data to train. Clinical translation of a reference-free CNN approach at this time would require a manufacturer to train their system, settings and probe on liver data from patients with known labels. In the future, transfer learning from one system to another could occur. However, additional study needs to be conducted to determine under what conditions results from one system and settings will transfer to another. Finally, the CNN loses some of its interpretability due to the general structure of stacking convolutions. To include the reference phantom as additional input to the CNN, a fusion mechanism can be used in the network where a separate convolution network could be used to extract features from the reference signal, and the features of the liver and the reference could be merged before feeding into the fully connected layers. This will be the subject of future work.

To test the robustness of the proposed CNN, we performed two ablation experiments: reduced the number of RF lines in one ROI and downsampling the RF line closer to the Nyquist frequency. For the first experiment where only 50% of the RF lines were randomly chosen for testing and this process was repeated 5 times, the average test accuracy was 71 % and the accuracy standard variation was 1% (across 5 times). For the second experiment where the RF data was downsampled to 10 MHz, the average testing accuracy was 65%. This suggests that the CNN is more robust to missing RF lines than missing higher frequency information when performing classification.

The CNN approach outperformed the traditional QUS approach when classifying steatosis (74% versus 59%). The misclassification might be caused by the fibrosis slowing the lipid

progression, or by unaccounted-for transmission losses caused by tissue layers, such as the skin, muscle and fat layers, between the transducer and the liver.

Conclusions

Rabbits were induced to have liver disease by placing them on a fatty diet for a defined duration and/or periodically injecting them with CCl₄. Rabbits were scanned ultrasonically *in vivo* and livers were classified based on the ultrasonic backscattered signals from the liver. Ground truth liver state was based on lipid liver percents using the Folch assay and hydroxyproline concentration to quantify fibrosis. Use of QUS parameters with linear regression and CCA demonstrated that the QUS parameters could differentiate between high and low lipid levels. However, the QUS parameters were not sensitive to fibrosis. Next, a CNN was implemented that took in raw RF signals and output classification of liver class in terms of a 5% lipid level threshold for the liver. The CNN outperformed classification using the QUS parameters combined with an SVM to classify liver state, 74% versus 59%. Therefore, while the CNN did not provide a physical interpretation of the tissue properties, e.g., attenuation of the medium or scatterer properties, the CNN had much higher accuracy in predicting fatty liver state.

Acknowledgements

This work was supported by a grant from National Institutes of Health (NIH) (R21 EB020766). We wish to acknowledge the assistance of Prof. Matthew Wallig and his pathology expertise in this work. We wish to acknowledge the assistance of Alexander Tam and Eben Arnold in data collection and Joe Rowles for his assistance with the Folch assay.

References

- Angulo P Nonalcoholic fatty liver disease. *New England Journal of Medicine*, 2002;346:1221–1231.
- Battaller R, Brenner DA. Liver fibrosis. *The Journal of Clinical Investigation*, 2005;115:209–218. [PubMed: 15690074]
- CDC. Chronic liver disease and cirrhosis. <https://www.cdc.gov/nchs/fastats/liver-disease.htm>, Accessed: March 3, 2019.
- Choong CC, Venkatesh SK, Siew EP. Accuracy of routine clinical ultrasound for staging of liver fibrosis. *Journal of Clinical Imaging Science*, 2012;2.
- Folch J, Lees M, Stanley GS. A simple method for the isolation and purification of total lipids from animal tissues. *Journal of Biological Chemistry*, 1957;226:497–509.
- Franceschini E, Escoffre JM, Novell A, Auboire L, Mendes V, Benane YM, Bouakaz A, Basset O. Quantitative ultrasound in ex vivo fibrotic rabbit livers. *Ultrasound in medicine & biology*, 2019;45:1777–1786. [PubMed: 31023499]
- Ghoshal G, Lavarello RJ, Kemmerer JP, Miller RJ, Oelze ML. Ex vivo study of quantitative ultrasound parameters in fatty rabbit livers. *Ultrasound in Medicine and Biology*, 2012;38:2238–2248. [PubMed: 23062376]
- Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 2004;16:2639–2664. [PubMed: 15516276]
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016 pp. 770–778.
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017 pp. 4700–4708.

- Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167, 2015.
- Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
- Kubota N, Kado S, Kano M, Masuoka N, Nagata Y, Kobayashi T, Miyazaki K, Ishikawa F. A high-fat diet and multiple administration of carbon tetrachloride induces liver injury and pathological features associated with nonalcoholic steatohepatitis in mice. *Clinical and Experimental Pharmacology and Physiology*, 2013.
- Lin T, Ophir J, Potter G. Correlation of ultrasonic attenuation with pathologic fat and fibrosis in liver disease. *Ultrasound in Medicine and Biology*, 1988;14:729–734. [PubMed: 3062866]
- Masters D, Luschi C. Revisiting small batch training for deep neural networks. arXiv preprint arXiv:1804.07612, 2018.
- Nguyen TN, Podkowa AS, Tam AY, Arnold EC, Miller RJ, Park TH, Do MN, Oelze ML. Characterizing fatty liver in vivo in rabbits, using quantitative ultrasound. *Ultrasound in medicine & biology*, 2019.
- Oelze ML, Mamou J. Review of quantitative ultrasound: Envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 2016;63:336–351.
- Oelze ML, Zachary JF, D OW. Characterization of tissue microstructure using ultrasonic backscatter: theory and technique for optimization using a Gaussian form factor. *The Journal of the Acoustical Society of America*, 2002;112:1202–11. [PubMed: 12243165]
- Paige JS, Bernstein GS, Heba E, Costa EAC, Fereirra M, Wolfson T, Gamst AC, Valasek MA, Lin GY, Han A, Erdman JWJ, O'Brien WDJ, Andre MP, Loomba R, Sirlin CB. A pilot comparative study of quantitative ultrasound, conventional ultrasound, and MRI for predicting histology-determined steatosis grade in adult nonalcoholic fatty liver disease. *American Journal of Roentgenology*, 2017;208:W168–W177. [PubMed: 28267360]
- Parker KJ. Ultrasonic attenuation and absorption in liver tissue. *Ultrasound in Medicine and Biology*, 1983;9:363–369. [PubMed: 6649154]
- Sandrin L, Fourquet B, Hasquenoph JM, Yon S, Fournier C, Mal F, Christidis C, Ziol M, Poulet B, Kazemi F, Beaugrand M, Palau R. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound in Medicine and Biology*, 2003;29:1705–1713. [PubMed: 14698338]
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- Suzuki K, Hayashi N, Sasaki Y, Kono M, Kasahara A, Fusamoto H, Imai Y, Kamada T. Dependence of ultrasonic attenuation of liver on pathologic fat and fibrosis: examination with experimental fatty liver and liver fibrosis models. *Ultrasound in Medicine and Biology*, 1992;18:657–666. [PubMed: 1440987]
- Treacher A, Beauchamp D, Quadri B, Fetzer D, Vij A, Yokoo T, Montillo A. Deep learning convolutional neural networks for the estimation of liver fibrosis severity from ultrasound texture In: *Medical Imaging 2019: Computer-Aided Diagnosis*. Vol. 10950 International Society for Optics and Photonics, 2019 p. 109503E.
- Tsui PH, Ho MC, Tai DI, Lin YH, Wang CY, Ma HY. Acoustic structure quantification by using ultrasound Nakagami imaging for assessing liver fibrosis. *Scientific Reports*, 2016;6:33075. [PubMed: 27605260]
- Wieckowska A, Feldstein AE. Diagnosis of nonalcoholic fatty liver disease: invasive versus noninvasive In: *Seminars in Liver Disease*. Vol. 28 Thieme Medical Publishers, 2008 pp. 386–395. [PubMed: 18956295]
- Yao LX, Zagzebski JA, Madsen EL. Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors. *Ultrasonic Imaging*, 1990;12:58–70. [PubMed: 2184569]
- Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision Springer*, 2014 pp. 818–833.

- Zeyer A, Doetsch P, Voigtlaender P, Schlüter R, Ney H. A comprehensive study of deep bidirectional lstm rnns for acoustic modeling in speech recognition. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) IEEE, 2017 pp. 2462–2466.
- Ziol M, Handra-Luca A, Kettaneh A, Christidis C, Mal F, Kazemi F, de Lédighen V, Marcellin P, Dhumeaux D, Trinchet JC., Beaugrand M. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. *Hepatology*, 2005;41:48–54. [PubMed: 15690481]

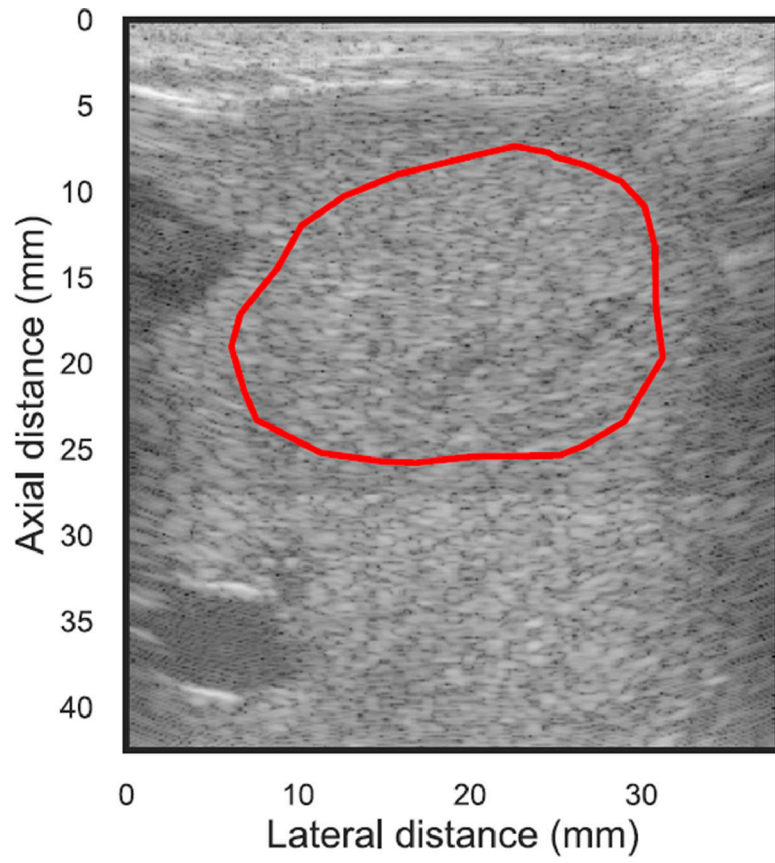


Figure 1:
B-mode image of rabbit liver with segmented region.

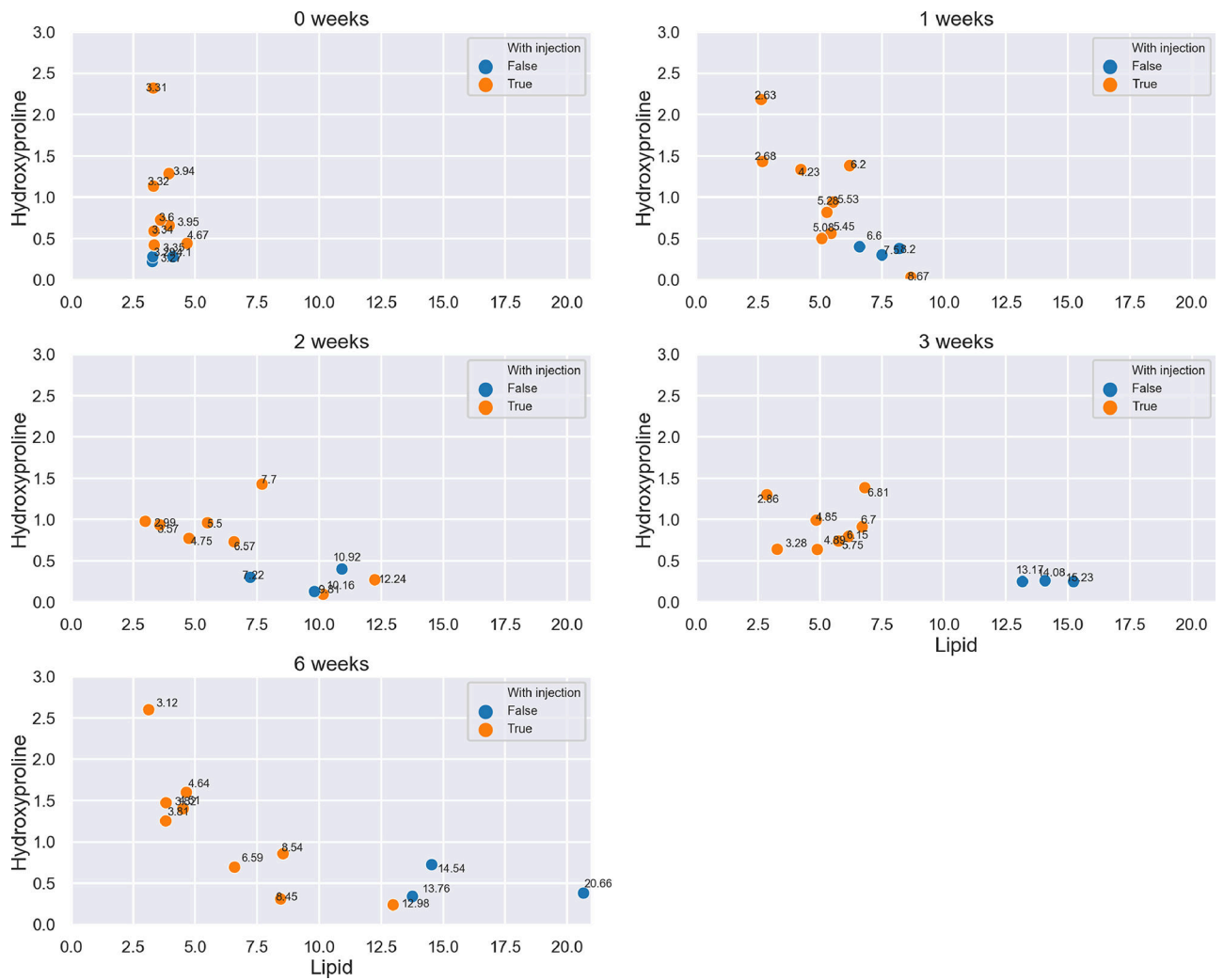


Figure 2: Hydroxyproline level (mg/g) vs. lipid percentage for 5 different diet groups. The blue circles are the rabbits without injection, the orange circles are the rabbits with injection.

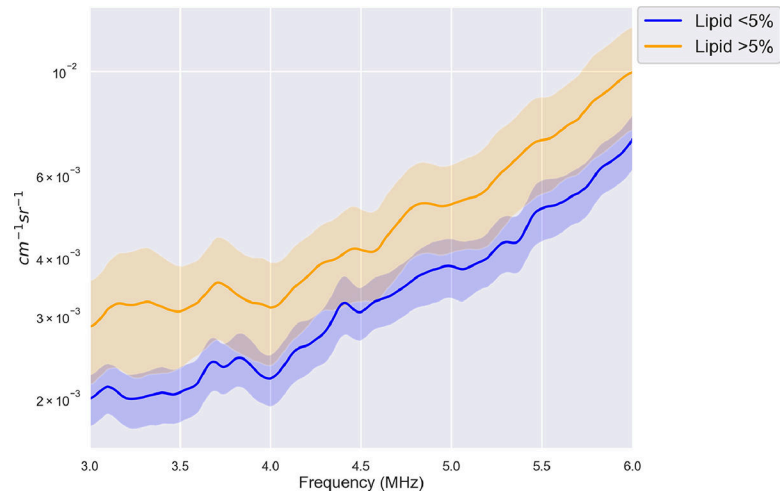


Figure 3:
BSC curves for two classes : low lipid (< 5%) and high lipid (> 5%).

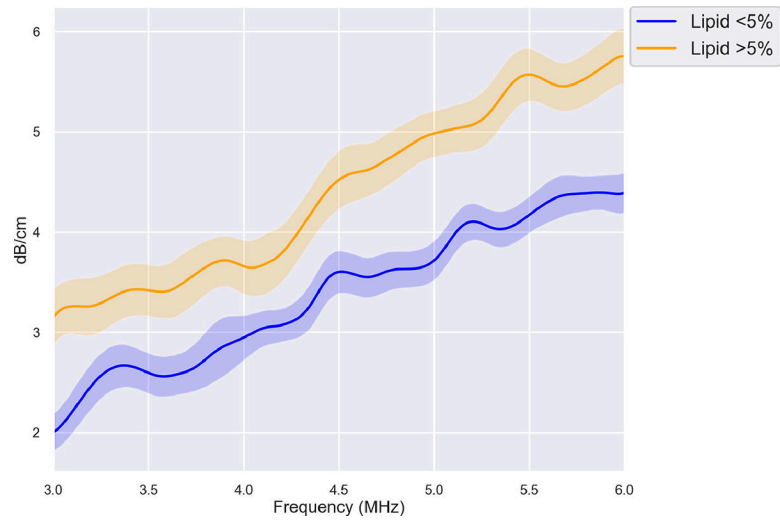


Figure 4:
Attenuation curves for two classes : low lipid (< 5%) and high lipid (> 5%).

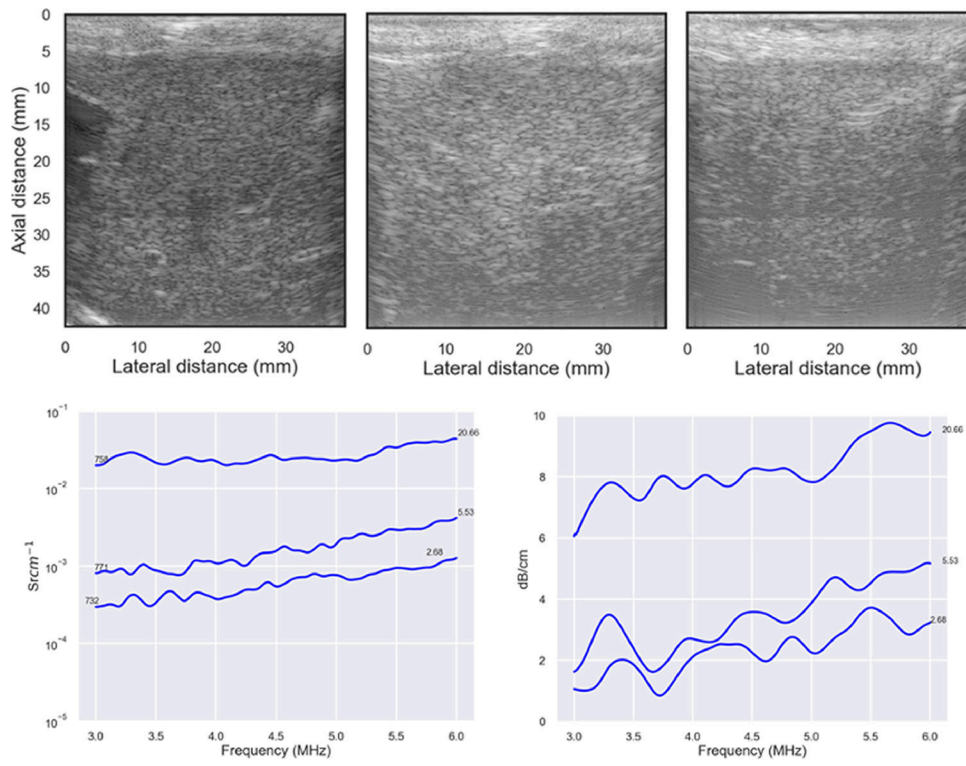


Figure 5: B-mode images of rabbit livers (top row) with lipid liver percentages from left to right of 2.68, 5.53 and 20.66. (Second row) Corresponding BSC and attenuation curves for the three rabbits shown in the B-mode images.

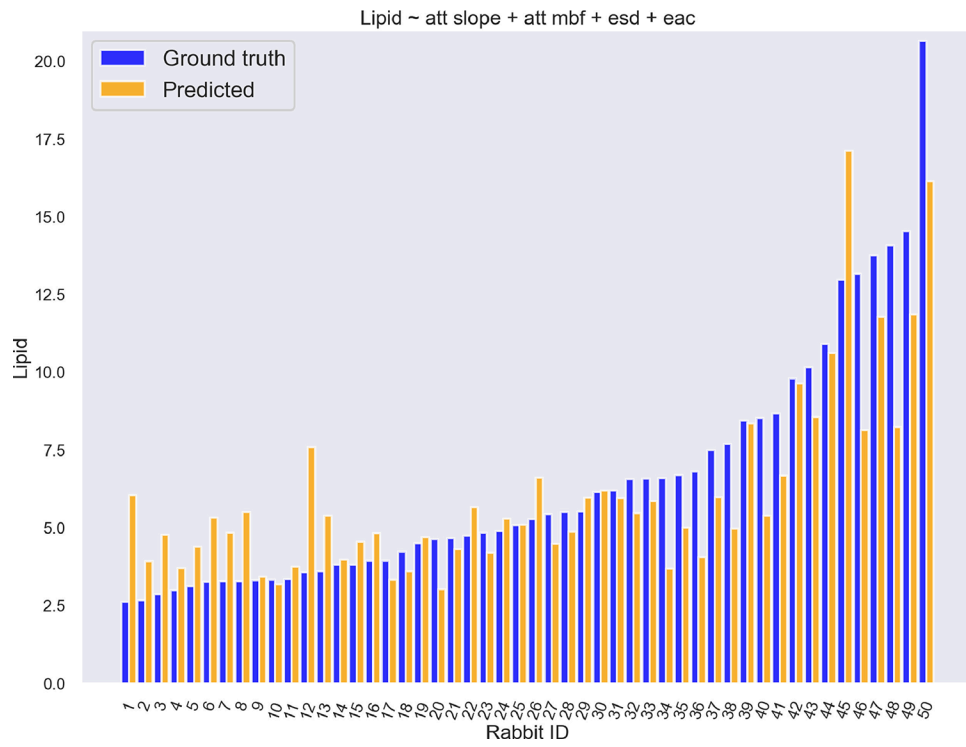


Figure 6: Linear regression of att slope, att intercept, ESD, EAC to lipid percentages. Blue bars are the ground truth, orange bars are the regressed values.

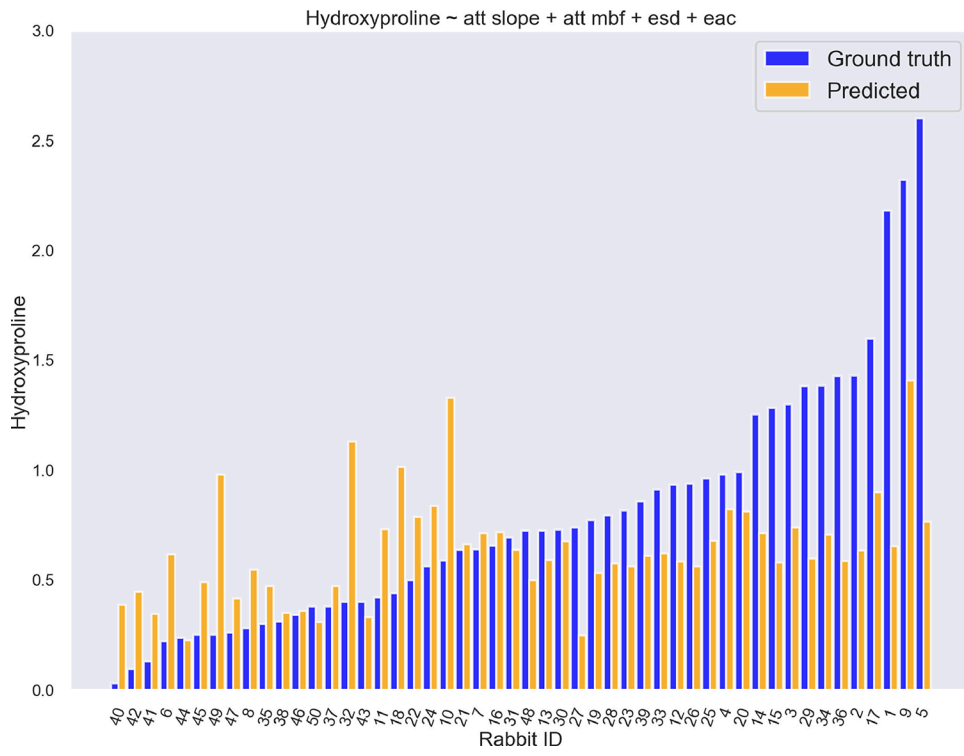


Figure 7: Linear regression of QUS parameters to hydroxyproline levels.

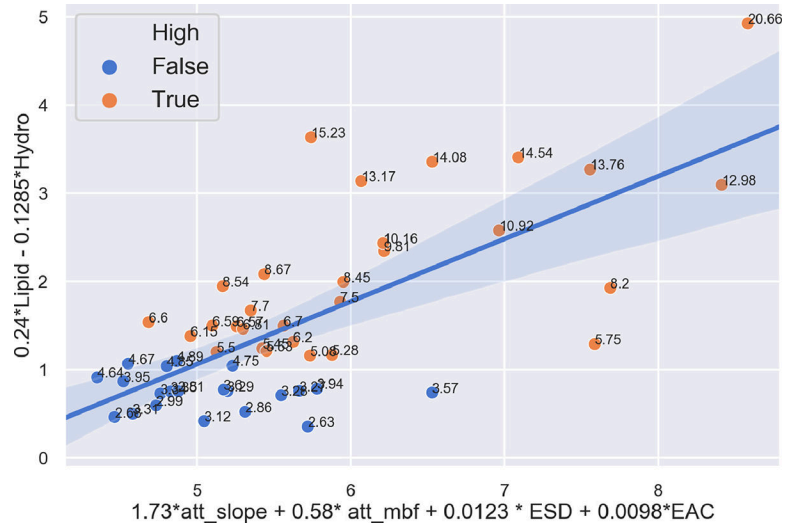


Figure 8: Plot of first canonical dimension correlation: the horizontal axis is the linear combination of inputs and vertical axis is linear combination of outputs. The blue circle denotes lipid percentages < 5% and the orange circle denotes lipid percentages ≥ 5%. The text next to the point is the actual lipid percentage.

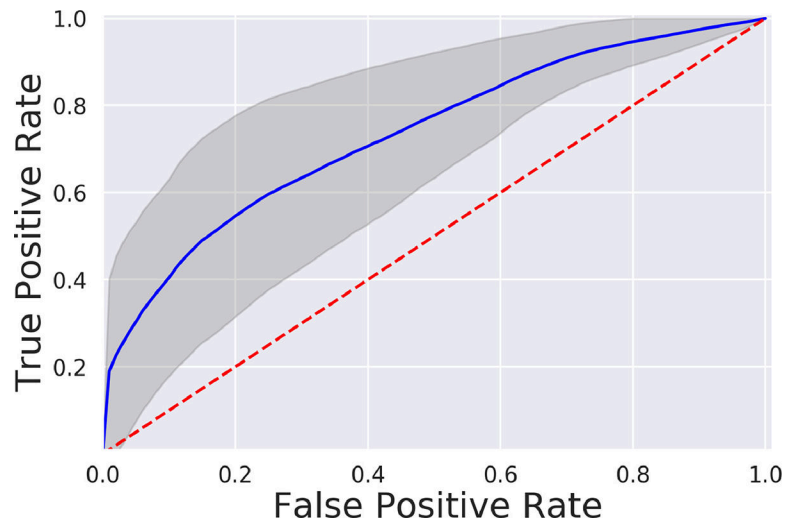


Figure 9: Average ROC curve of the six folds (grey band). The dotted red line indicates the no-discrimination line (random guess).

Table 1:

1D convolutional neural network architecture.

Layers	Output shape	Parameter numbers
1D Convolution	(347,7)	126
Max Pooling	(173,7)	0
1D Convolution	(165, 7)	448
Max Pooling	(82,7)	0
1D Convolution	(78,5)	180
Max Pooling	(39,5)	0
1D Convolution	(37,3)	48
Max Pooling	(18,3)	0
Flatten	54	0
Dense	16	1100
Dense	8	210
Dense	2	66
Activation	1	0

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Average Hydroxyproline values estimated for each injection group.

Injection (per week)	Number of weeks	Hydroxyproline level [mg/g]
0.0 mL/kg	11	0.28
0.2 mL/kg	11	1.24
0.3 mL/kg	11	0.97
0.6 mL/kg	11	1.67
0.0 mL/kg	8	0.34
0.035 mL/kg	8	0.90
0.07 mL/kg	8	1.13

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

QUS parameters for differentiating two classes with threshold of 5% lipid liver levels.

	ESD	EAC	Attenuation slope (dB/cm.MHz)	Attenuation midband-fit (dB/cm)
Low fat	127.16 ± 42.78	33.85 ± 16.14	0.69 ± 0.36	3.50 ± 0.88
High fat	117.16 ± 43.89	35.5 ± 16.75	0.97 ± 0.27	4.37 ± 1.44
p-value	0.38	0.71	0.03	0.003

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

Linear regression coefficients and their p-values.

		ESD	EAC	Attenuation slope	Attenuation midband-fit
Hydroxyproline	Coefficient	-0.0062	-0.0073	-0.8773	-0.0754
	p-value	0.20	0.54	0.01	0.25
Lipid percentage	Coefficient	0.031	0.023	4.71	1.72
	p-value	0.23	0.89	9.72e-06	3.72e-08

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5:

CCA input and output weights and their correlations.

	First canonical dimension	Second canonical dimension
Attenuation slope weight	1.7305	-3.9614
Attenuation midband fit weight	0.5819	0.5369
ESD weight	0.0123	-0.0339
EAC weight	0.0098	-0.0048
Lipid weight	0.2408	0.1531
Hydroxyproline weight	-0.1285	2.0118
Correlation	0.71	0.26

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6:

Training and testing accuracy of the 1D convolution neural network.

	Training accuracy	Test accuracy
Fold 1	82.16 %	78.56 %
Fold 2	80.18 %	77.47 %
Fold 3	81.62 %	68.46 %
Fold 4	81.24 %	65.39 %
Fold 5	80.5%	76.29 %
Fold 6	80.49 %	76.69 %
Average accuracy across folds	81.03 %	73.81%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7:

Training and testing accuracies of an SVM classifier using four QUS parameters: ESD, EAC, attenuation slope, attenuation midband fit. The same rabbits in each fold were used for the CNN and the QUS with SVM classifiers.

	Training accuracy	Test accuracy
Fold 1	66.14 %	67.04 %
Fold 2	70.79 %	38.17 %
Fold 3	69.64 %	62.17 %
Fold 4	66.96 %	68.53 %
Fold 5	70.62 %	61.13 %
Fold 6	69.49 %	57.68 %
Average accuracy across folds	68.94 %	59.12%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript