

Reference free quantitative ultrasound classification of fatty liver

Trong Nguyen, *Student Member, IEEE*, and Michael Oelze, *Member, IEEE*

Abstract—The backscatter coefficient (BSC) describes the scattering properties of a medium and can be used to characterize tissue, such as fatty liver. To calculate the BSC, a calibration spectrum is needed, which can be acquired using a reference phantom method before or after the clinical scanning procedure is complete. This requires yet another scanning procedure that can disrupt the busy clinical flow. Therefore, we have explored the use of a convolutional neural network (CNN) as a means of eliminating an external reference step while still capable of classifying liver disease in a rabbit model of fatty liver. Sixty New Zealand white rabbits were separated into five cohorts with each cohort maintained on a special high fat diet to induce different degrees of fatty liver. One week before scanning, rabbits were placed on normal chow. Rabbit livers were scanned *in vivo* using an L9-4/38 linear array connected to a SonixOne scanner. Raw RF data were collected from the scans and used for quantitative ultrasound (QUS) analysis. Immediately after scanning, the rabbit liver was extracted and the percent lipids in the liver were estimated using the Folch assay. The rabbit livers were classified into two classes: high fat (lipid percentage $\geq 5\%$) and low fat (lipid percentage $< 5\%$). The 5% threshold was equal to the median of lipid percentages of all the rabbits. Livers were classified using traditional QUS approaches and compared with a CNN approach where no reference was utilized, and the raw RF signals were used as the inputs. The attenuation and the BSC were estimated from the RF using a reference phantom technique. The attenuation slope, attenuation midband-fit, the effective scatterer diameter, and the effective acoustic concentration were extracted from the attenuation and the BSC and used as features to train a kernel support vector machine (SVM) for the task of liver lipid classification. A CNN was designed to simultaneously extract the features from the raw RF and perform liver lipid classification without using a reference phantom. Six-fold cross validation was performed to quantify the accuracy of the SVM classifier using the QUS parameters and the CNN classifier using the RF. The average training and testing accuracies across six folds using the QUS approach was 68.94% and 59.12%, respectively. The average training and test accuracy using the CNN approach were 81.03% and 73.81% for training and testing, respectively. The results demonstrate that the CNN can be used to classify fatty liver without the need for an external reference scan, i.e., reference-free QUS.

Index Terms—Machine learning, quantitative ultrasound, steatosis

I. 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 [1]. With up to a third of the United States population affected by NALFD (70-90% of

obese or Type 2 diabetic patients have NAFLD), NAFLD represents a significant medical concern. There remains an unmet clinical need to develop imaging techniques for the non-invasive evaluation of liver steatosis.

Different methods using ultrasound have been explored to detect steatosis. Based on histological grading, Lin et al. [2] 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 for grade 1, 2 and 3, respectively). In our previous study of 15 rabbits on high fat diets, QUS classified liver with accuracies up to 84% [3].

The first goal of the current study is to evaluate QUS as a noninvasive method to quantitatively assess liver 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. Independent correlations between QUS parameters and lipid percentages were computed. Because the system settings when acquiring rabbits liver scans were kept unchanged, we hypothesized that a 1D convolutional neural network (CNN) could compensate for the system-dependent and tissue-dependent effects, and perform classification in a reference-free manner. 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).

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 requires a reference scan to classifying liver state. This could eliminate a time consuming task of a busy clinical work flow.

II. METHODS

A. 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 x 5 factorial design with rabbits on a fatty diet for 0, 1, 2, 3 or 6 weeks The

T. Nguyen and M. L. Oelze are with the Beckman Institute for Advanced Science and Technology, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, IL, 61801, USA. e-mail: oelze@illinois.edu.

rabbits were returned to normal diet for about a week before ultrasonic scanning.

B. 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 [4]. Following scanning, the rabbits were euthanized via CO₂ while still under anesthesia.

C. 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 [5]. The Folch was used to quantify the lipid levels in the liver, respectively.

D. Quantitative Ultrasound Procedures

The attenuation and BSC curves were extracted from raw RF data by segmenting the liver regions from the B-mode images of the livers, i.e., regions of interest were chosen for each frame. Care was taken to omit regions in the liver that were shadowed by the ribs or other structures or that contained large blood vessels. This 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 [4]. The reference phantom composition and properties are described in detail in [6]. 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 get a mean attenuation curve for each rabbit liver [4], [7]. A slope and a mid-band fit at 4.5 MHz were estimated from the fitted line to the average attenuation curve. ESDs and EACs were derived from the BSC curves using a spherical Gaussian scattering form factor [8]. Correlations between each QUS parameter and the lipid percentages were calculated.

A linear regression of the four QUS parameters (ESD, EAC, attenuation slope, attenuation midband fit) was performed individually versus lipid. Specifically, linear regression of lipid percentage y_i^1 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:

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

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

where i is the index of the rabbits' ID.

E. Convolutional Neural Network Architecture

The model-based QUS approach requires the use of the reference phantom to derive the BSCs and the attenuation from the power spectrum of the RF data. On the other hand, the CNN approach did not require the use of a reference phantom or a model. However, in this study the ultrasonic scanner settings used in all of the rabbit scans were the same, the differences manifested in the RF data were hypothesized to come from the fatty diet. Using a CNN, the feature extraction and classification can be accomplished simultaneously through the concatenation of the convolutional layers and the fully connected layers. To prevent overfitting of the CNN classifier, only the problem of classification was considered: e.g., classification of high lipid vs. low lipid. The CNN was used to classify from the liver images only, without the help of reference signals to remove system effects.

The gated RF lines inside the same liver segmentation when extracting the QUS parameters were used as inputs to the 1D 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 used, 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. Six-fold cross validation was used to prevent overfitting. 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. To keep the number of rabbits in each class balanced, four rabbits were of class 1 and four rabbits were of class 0 in the test set. 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. The accuracy was evaluated on a frame by frame basis.

The first layer of the network is the 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. We utilized the Adam optimizer [9], which is a gradient descent method, due to its ability to escape local minima of the cost function. The cost function used for multi-class classification was the cross-entropy between the true class vector \mathbf{p} and the predicted class vector \mathbf{q} defined as follows:

$$H(\mathbf{p}, \mathbf{q}) = - \sum_i p \log(q). \quad (3)$$

The output vectors \mathbf{p} and \mathbf{q} were normalized via a softmax function so that each vector summed to unity and could be

interpreted as the probability vector of different class outputs. The output ground truth vector \mathbf{p} was constructed using one-hot encoding.

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. 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 [10], 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 [11] was also used to ensure the weights at each layer had zero mean and unit variance.

There were four hidden convolutional layers, four pooling layers, two fully connected layers and a four softmax output layer. 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 defined probability (0.5 was used in this study). Dropout helps redistribute the weights (importances) 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 important to classification, effectively reducing the dimension of the final classifier (or reducing overfitting).

The loss function was constructed for classifying into two groups. The loss function used for the binary classification was the cross-entropy between the ground truth and the predicted output:

$$L = - \sum_{i=1}^N y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i), \quad (4)$$

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. When the output of the sigmoid function is greater than 0.5, the input is classified as high lipid, otherwise low lipid.

TABLE I: 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

To compare the performance of the CNN with traditional QUS parameters, a support vector machine (SVM) was used with BSC- and attenuation-derived parameters. Specifically, the ESD, EAC, attenuation mid-band fit and attenuation slope were combined in the SVM. Like the CNN, testing and training accuracies were reported using six fold validation.

III. RESULTS

A. QUS Parameters

We sought 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 an additional rabbit from the study, which had a liver too small to segment.

Table I lists the averaged ESD, EAC, attenuation slope, attenuation midband-fit and their p-values for differentiating between the two lipid level classes. The attenuation curves had more differentiating power than the BSC curves. 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. 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.

Figure 1 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%.

B. CNN for lipid classification

The QUS parameters correlated well with the lipid changes. Table II shows the classification results of 1D CNN for classifying two classes: low and high lipid with a threshold of 5%. Table III 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 SVM was used. The results show that the CNN outperforms the QUS approach for classification of lipid changes.

IV. DISCUSSION

The objective of the study was to investigate the relationship between QUS parameters and liver steatosis through non-invasive ultrasonic interrogation and the application of a CNN

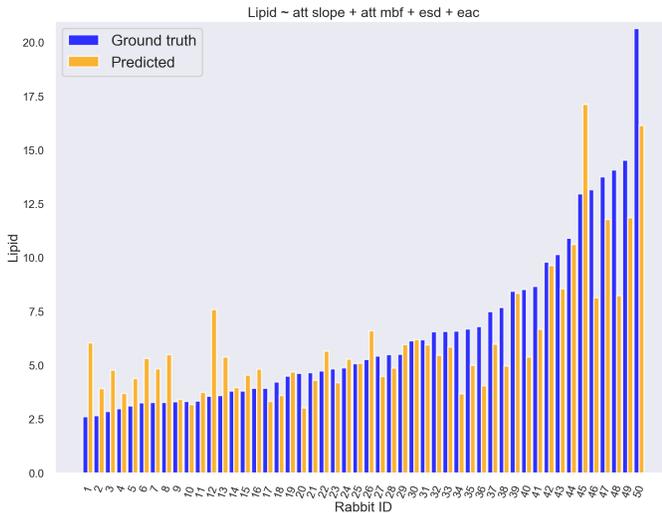


Fig. 1: Linear regression of att slope, att intercept, ESD, EAC to lipid percentages. Blue bars are the ground truth, orange bars are the regressed values.

TABLE II: Training and testing accuracy of the 1D convolutional 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 %

to the problem of liver classification allowing both a model-free analysis and reference-free scanning configuration. The rabbits were divided into five groups maintained on a fatty diet over different durations. According to histological analysis, rabbits had lipodosis and some with mild fibrosis, none of the rabbits developed cirrhosis or heavy fibrosis.

A median of the lipid percentages of 5% was chosen as a threshold for classification between low 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. Attenuation increased for increased steatosis. However, attenuation slope and midband fit were more sensitive to lipid liver percentages. ESD and EAC were not sensitive to lipid.

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. To test the feasibility of 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. More recent developed architectures like ResNet [12] or DenseNet Spectrum [13]

TABLE III: Training and testing accuracies of a SVM classifier with four QUS parameters. The same rabbits in each fold were used to compare with the CNN approach.

	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 %

might be tested in future work; however, those newer methods require more data. The CNN loses some of its interpretability. The CNN approach outperformed the QUS approach when classifying steatosis (74% versus 59%). Misclassification might be caused by unaccounted-for transmission losses caused by the layer between the transducer and the liver.

ACKNOWLEDGMENT

The authors acknowledge the support of a grant from the National Institutes of Health (R21EB020766).

REFERENCES

- [1] A. Wieckowska and A. E. Feldstein, "Diagnosis of nonalcoholic fatty liver disease: invasive versus noninvasive," in *Seminars in Liver Disease*, vol. 28, pp. 386–395, Thieme Medical Publishers, 2008.
- [2] T. Lin, J. Ophir, and G. Potter, "Correlation of ultrasonic attenuation with pathologic fat and fibrosis in liver disease," *Ultrasound in Medicine and Biology*, vol. 14, no. 8, pp. 729–734, 1988.
- [3] T. N. Nguyen, A. S. Podkova, A. Y. Tam, E. C. Arnold, R. J. Miller, T. H. Park, M. N. Do, and M. L. Oelze, "Characterizing fatty liver in vivo in rabbits, using quantitative ultrasound," *Ultrasound in medicine & biology*, 2019.
- [4] L. X. Yao, J. A. Zagzebski, and E. L. Madsen, "Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors," *Ultrasonic Imaging*, vol. 12, no. 1, pp. 58–70, 1990.
- [5] J. Folch, M. Lees, G. Sloane Stanley, *et al.*, "A simple method for the isolation and purification of total lipids from animal tissues," *Journal of Biological Chemistry*, vol. 226, no. 1, pp. 497–509, 1957.
- [6] K. Nam, I. M. Rosado-Mendez, L. A. Wirtzfeld, V. Kumar, E. L. Madsen, G. Ghoshal, A. D. Pawlicki, M. L. Oelze, R. J. Lavarello, T. A. Bigelow, *et al.*, "Cross-imaging system comparison of backscatter coefficient estimates from a tissue-mimicking material," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1319–1324, 2012.
- [7] K. J. Parker, "Ultrasonic attenuation and absorption in liver tissue," *Ultrasound in Medicine and Biology*, vol. 9, no. 4, pp. 363–369, 1983.
- [8] M. L. Oelze, J. F. Zachary, and O. W. D., "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*, vol. 112, no. 3 Pt 1, pp. 1202–11, 2002.
- [9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [11] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [13] J. Andén and S. Mallat, "Deep scattering spectrum," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4114–4128, 2014.