10.1148/radiol.2020191160

Appendix E1. Transmit Focal Range and Time Gain Compensation Settings

The transmit focal range was adjusted for each participant during the RF data acquisition. The transmit focal range values used for the RF data acquisition varied from 2 to 14 cm. The distribution of the transmit focal range used is shown in Table E1.

The time gain compensation was adjusted for each participant during the RF data acquisition. Figure E1 shows the actual analog gain as a function of depth applied to the radiofrequency signal.

Appendix E2. Network Architectures and Implementation Details

CNNs are commonly applied to images to automatically extract very subtle patterns, with each network layer learning increasingly abstract and higher-level representations of the input data. We designed 1D-CNNs applicable to 1D RF signals, where the one dimension represented the time axis. We implemented both 1D-CNN algorithms (ie, binary classifier and fat fraction estimator) in Tensorflow 1.7.0 (Google, Inc.; open source) and Python 2.7 (Python Software Foundation; open source).

Both algorithms were developed, tuned, and trained using the training group (n = 102), and evaluated using the separate test group (n = 102). The tuning, including network architecture optimization and hyperparameter selection, was performed within the training group through cross-validation, without using data from the separate test group. Specifically, multiple networks and hyper-parameters were compared through cross-validation within the training group, and the best performing network and hyper-parameters were selected. The selected model was then trained using data from the entire training group, yielding trained algorithms that can be readily tested in the test group.

The main architectures tested were 1D-CNNs in which the convolutional block was followed by two dense (fully connected) layers. The selected network for the classifier consisted of three convolutional layers, followed by two dense layers and a softmax layer (32). The classifier output was a NAFLD classification score p1 with a value between zero and one, as well as a normal classification score p0 = 1-p1. The selected network for the fat fraction estimator had a similar architecture except that the second dense layer served as the output without the softmax layer. The output value was the predicted fat fraction (%). Architecture details are presented in Figure E2. All activation layers were implemented with the tanh function defined by $\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$.

The network was trained from scratch with random initialization of weights. We used the Adagrad optimizer (33) with a learning rate of 0.005. The batch size was 256. The classifier was trained by minimizing the cross-entropy loss (18). The fat fraction estimator was trained by minimizing the mean square error between the algorithm output and the MRI-derived proton density fat fraction. The training stopped at the 50th epoch. Line level outputs for each

participant were ensembled to generate the participant level output for that participant according to the procedure described in Materials and Methods.

Appendix E3. Comparison between Radiofrequency and Envelope Signals as Deep Learning Input Data

The radiofrequency (RF) data were chosen as the input data for our 1D-CNN models based on the premise that the RF data contain more information than the envelope data. However, it was unknown whether the additional information contained in the RF data could make a difference with the 1D-CNN models. For completeness, we retrained and retested our 1D-CNN models (the classifier and the fat fraction estimator) using envelope data as the input.

Two groups of envelope data were tested, the raw envelope data and the processed envelope data, representing different degrees of processing involved. The raw envelope data were obtained by taking the absolute value of the Hilbert transform of the RF data [ie, raw envelope = abs (hilbert (RF))]. The processed envelope data were obtained by log-compressing the raw envelope data [ie, $20\log_{10}$(raw envelope)] and setting a dynamic range of 60 dB. The above operations were performed on the RF data at the original sampling frequency (40 MHz). The downsampling by a factor of 4 was performed after the above operations.

To be consistent with the RF methodology, both the RF data without TGC and the RF data with TGC were used to derive the envelope data, resulting in four types of envelope data: raw envelope data without TGC, processed envelope data without TGC, raw envelope data with TGC, and processed envelope data with TGC. The raw envelope data without TGC hypothetically contained less information than the RF data without TGC but contained more information than the processed envelope data without TGC. Among the four types of envelope data, the processed envelope data with TGC were closest to the underlying data of the B-mode images displayed on clinical scanners. Therefore, the processed envelope with TGC and the RF without TGC were the primary pair for purposes of comparing RF and envelope data, although the results of all types of input data were reported to provide further insight (Tables E2–E4).

For the classifier, all four types of envelope data yielded AUC values > 0.90 in the test group (Table E2), suggesting that all four types of envelope data could be effective for NAFLD diagnosis. The AUC estimate for the processed envelope with TGC [0.95 (95% CI, 0.90–0.99)] and the AUC estimate for the RF without TGC [0.98 (95% CI, 0.94–1.00)] did not differ (P = .14).

When a predetermined threshold of 0.5 in the composite NAFLD classification score was used for NAFLD diagnosis in the test group, the specificity and overall classification accuracy for the processed envelope data with TGC [specificity, 66% (95% CI, 47%–81%), 21/32; accuracy, 87% (95% CI: 79%–93%), 89/102] (Table E3) were lower than those for the RF data without TGC [specificity, 94% (95% CI, 79%–99%), 30/32; accuracy, 96% (95% CI, 90%–99%), 98/102], with P = .006 for specificity and P = .02 for accuracy. The sensitivity between the two cases were identical [97% (95% CI, 90%–100%), 68/70].

For the fat fraction estimator, the Pearson correlation coefficient values between the predicted fat fraction and the MRI-PDFF were 0.85, 0.78, 0.78, 0.80, 0.67, 0.65 for RF without TGC, raw envelope data without TGC, processed envelope data without TGC, RF with TGC, raw envelope data with TGC, and processed envelope data with TGC, respectively (P < .001 for

all cases) (Table E4). The Pearson correlation coefficient was lower for the processed envelope with TGC than for the RF without TGC ($P < .001$).

Overall, the additional information contained in the RF data relative to the envelope data were shown to positively affect the performance of the 1D-CNN fat fraction estimator, and to a lesser degree, the performances of the 1D-CNN classifier.

References

32. Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, England: Cambridge University Press, 2000.

33. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 2011;12:2121–2159. http://jmlr.org/papers/v12/duchi11a.html.

Table E1. The distribution of transmit focal range used for ultrasound radiofrequency data acquisition

Transmit Focal Range (cm)	Number of Participants		
	Training Group	Test Group	Total
2	1	0	1
4	1	0	1
6	26	34	60
8	60	54	114
10	14	12	26
14	0	2	2

Table E2. Area under the receiver operating characteristic curve values of the composite NAFLD classification scores obtained by the classifier in the test group, for various types of input ultrasound signals

	Input without TGC			Input with TGC		
	Input = RF	Input = Raw Envelope	Input = Processed Envelope	Input = RF	Input = Raw Envelope	Input = Processed Envelope
AUC (95% CI)	0.98 (0.94–1.00)	0.96 (0.90–1.00)	0.97 (0.93–1.00)	0.95 (0.91–0.99)	0.91 (0.85–0.97)	0.95 (0.90–0.99)

Footnote: AUC = area under the receiver operating characteristic curve, CI = confidence interval, NAFLD = nonalcoholic fatty liver disease, TGC = time gain compensation.

Table E3. Performance metrics for NAFLD diagnosis in the test group using the composite NAFLD classification scores generated by the binary classifier based on the predetermined threshold of 0.5, for various types of input ultrasound signals

	Input without TGC			Input with TGC		
	Input = RF	Input = Raw Envelope	Input = Processed Envelope	Input = RF	Input = Raw Envelope	Input = Processed Envelope
Sensitivity (95% CI) (%) [fraction]	97 (90–100) [68/70]	96 (88–99) [67/70]	96 (88–99) [67/70]	91 (82–97) [64/70]	91 (82–97) [64/70]	97 (90–100) [68/70]
Specificity (95% CI) (%) [fraction]	94 (79–99) [30/32]	84 (67–95) [27/32]	84 (67–95) [27/32]	88 (71–96) [28/32]	75 (57–89) [24/32]	66 (47–81) [21/32]
PPV (95% CI) (%) [fraction]	97 (90–99) [68/70]	93 (86–97) [67/72]	93 (86–97) [67/72]	94 (86–98) [64/68]	89 (81–94) [64/72]	86 (79–91) [68/79]
NPV (95% CI) (%) [fraction]	94 (79–98) [30/32]	90 (75–96) [27/30]	90 (75–96) [27/30]	82 (68–91) [28/34]	80 (64–90) [24/30]	91 (72–98) [21/23]
Accuracy (95% CI) (%) [fraction]	96 (90–99) [98/102]	92 (85–97) [94/102]	92 (85–97) [94/102]	90 (83–95) [92/102]	86 (78–92) [88/102]	87 (79–93) [89/102]

Footnote: CI = confidence interval, NAFLD = nonalcoholic fatty liver disease, NPV = negative predictive value, PPV = positive predictive value, RF = radiofrequency, TGC = time gain compensation.

Table E4. Pearson correlation coefficient values between the predicted fat fraction and the MRI-derived proton density fat fraction for various types of input ultrasound signals

	Input without TGC			Input with TGC		
	Input = RF	Input = Raw Envelope	Input = Processed Envelope	Input = RF	Input = Raw Envelope	Input = Processed Envelope
Pearson correlation coefficient	0.85 (*P* < .001)	0.78 (*P* < .001)	0.78 (*P* < .001)	0.80 (*P* < .001)	0.67 (*P* < .001)	0.65 (*P* < .001)

Footnote: RF = radiofrequency, TGC = time gain compensation.