# Automated Machine Learning in the Sonographic Diagnosis of Non-alcoholic Fatty Liver Disease

*Sriharsha Gummadi, MD [a,b], Nirmal Patel [a], Haresh Naringrekar, MD [a], Laurence Needleman, MD [a], Andrej Lyshchik, MD PhD [a], Patrick O'Kane, MD [a], Jesse Civan, MD [c], John R Eisenbrey, PhD [a,*]*

[a] *Department of Radiology, Thomas Jefferson University, PA, USA;* [b] *Department of Surgery, Lankenau Medical Center, PA, USA;* [c] *Department of Medicine, Division of Gastroenterology & Hepatology, Thomas Jefferson University, PA, USA*

***Objective:*** This study evaluated the performance of automated machine-learning to diagnose non-alcoholic fatty liver disease (NAFLD) by ultrasound and compared these findings to radiologist performance.

***Methods:*** 96 patients with histologic (33) or proton density fat fraction MRI (63) diagnosis of NAFLD and 100 patients without evidence of NAFLD were retrospectively identified. The "Fatty Liver" label included 96 patients with 405 images and the "Not Fatty Liver" label included 100 patients with 500 images. These 905 images made up a "Comprehensive Image" group. A "Radiology Selected Image" group was then created by selecting only images considered diagnostic by a blinded radiologist, resulting in 649 images. Cloud AutoML Visionbeta (Google LLC, Mountain View, CA) was used for machine learning. The models were evaluated against three blinded radiologists.

***Results:*** The "Comprehensive Image" group model demonstrated a sensitivity of 88.6% (73.3–96.8%) and a specificity of 95.3% (84.2–99.4%). Radiologist performance on this image group included a sensitivity of 81.0% (74.3–87.6%) and specificity of 86.0% (72.6–99.5%). The model's overall accuracy was 92.3% (84.0–97.1%), compared with mean individual performance (83.8%, 78.4–89.1%). The "Radiology Selected Image" group model demonstrated a sensitivity of 88.6% (73.3 – 96.8%) and specificity of 87.9% (71.8–96.6%). Mean radiologist sensitivity was 92.4% (86.9–97.9%) and specificity was 91.9% (83.4–100%). The model's overall accuracy was 88.2% (78.1–94.8%) which was comparable to the individual radiologist performance (92.2%, 90.1–94.2%) and consensus performance (95.6%, 87.6–99.1%).

***Conclusions:*** An automated machine-learning algorithm may accurately detect NAFLD on ultrasound.

***Key words:*** Non-alcoholic fatty liver disease; Machine learning; Ultrasound

Non-alcoholic fatty liver disease (NAFLD) is defined as the presence of hepatic steatosis not attributable to other causes of liver dysfunction, including alcoholic liver disease, viral hepatitis, and hemochromatosis. NAFLD may be present without inflammation (non-alcoholic fatty liver, or NAFL) or with inflammation (non-alcoholic steatohepatitis, or NASH). Variable levels of fibrosis up to frank cirrhosis may be present. Importantly, NAFLD is now the leading cause of chronic liver disease in the U.S, representing up to 75.1% of cases with an overall prevalence as high as 11.0% [1, 2].

The diagnosis of NASH is complicated due to a lack of consensus on the role of laboratory findings, imaging, and histopathology from biopsy. The magnitude of liver function test (LFT) abnormalities does not correlate with degree of inflammation or fibrosis [3]. Additionally, the absence of LFT elevations does not

exclude the presence of NAFLD. Imaging modalities for diagnosis include computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound. CT has a sensitivity of 50% and specificity of 83% in the detection of NAFLD, but radiation concerns may limit routine use [4]. MRI performs better with a sensitivity of 88% and specificity of 63% [4]. The use of in- phase and out-phase sequences on MRI may be used to improve test statistics with a sensitivity and specificity of 95% and 98% respectively by quantifying degree of hepatic steatosis [5, 6]. However, lack of access, high cost, and contraindications related metallic implants are prevalent. While liver biopsy is considered the gold-standard, the test is invasive and typically reserved for patients with an unclear diagnosis.

Ultrasound may be used to screen for significant steatosis in patients with NAFLD, with a sensitivity up to 85% and specificity of up to 94% respectively [7]. However, this sensitivity decreases to as low as 49% in the presence of morbid obesity [8]. Fatty liver may be diagnosed sonographically by hepatic hyperechogenicity relative to the renal cortex and spleen, ultrasound wave attenuation with a loss of diaphragm definition, poor delineation of the intrahepatic architecture, and a loss of echogenic fat within portal triads [7]. Compared with CT and MRI, ultrasound is less expensive, readily available, faster, and portable. There are no concerns with claustrophobia, radiation, or metallic contraindications. A reliable ultrasound approach to screen for NAFLD would be beneficial for early diagnosis. We hypothesize that a trained deep-learning model can diagnose NAFLD by ultrasound accurately.

## Materials and Methods

Data (ultrasound images) for this institutional review board (IRB) approved study for model training, validation, and testing purposes were extracted from our institutional Picture Archiving and Communication (PACS) system in a retrospective fashion via consecutive sampling. Informed consent was waived by our local ethics board as this study only utilized retrospective data. A binary classification system was created for labelling purposes: "Fatty Liver" and "Not Fatty Liver." Images were processed using Cloud AutoML Visionbeta (Google LLC, Mountain View, CA) for model training and evaluation. This is a cloud-based freeware program available for Beta testing. Per the product guidelines, at least 100 – 500 images per label is recommended.

Ultrasound images were originally created on a variety of scanners including a GE Logiq 9 and E9 (GE Healthcare), an IU22 and Epiq 7 (Philips), and an Aplio 500 and i800 (Cannon). The "Fatty Liver" label consisted of ultrasound pictures from patients identified either by tissue histology via liver biopsy or after fat quantification by MRI. Fat quantification was performed using proton density fat fraction (PDFF) on one of two Optima MR450W 1.5T scanner using Ideal IQ (GE Healthcare). Patients with histological confirmation of NAFLD were selected from patients identified via PACS undergoing ultrasound-guided liver biopsy for a presumed diagnosis of NAFLD or NASH. As many images of each patient as considered representative and unique were extracted. Patients with an MRI fat quantification diagnosis of NAFLD were selected from patients identified via PACS undergoing MRI with fat quantification greater than 6.4% with a diagnostic ultrasound available within the prior 6 months [9]. A goal of 5 pictures per patient was predicted to be extracted with a total label goal of between 100 – 500 images. Inclusion and exclusion criteria for the "Fatty Liver" label is further detailed in Table 1.

**Table 1**   Inclusion and exclusion criteria for patients within the "Fatty Liver" label

| Inclusion Criteria | Exclusion Criteria |
| --- | --- |
| Histologic diagnosis of NAFLD (NASH or hepatic steatosis) after liver biopsy or MRI with fat quantification > 6.4% AND | |
| Diagnostic ultrasound within 6 months | TIPS |
| Retrievable through the PACS system | Hemochromatosis or hemosiderosis with an iron content > 2.0 mg Fe / g Liver |
| Age > 18 | Alcoholic hepatitis or uncontrolled viral hepatitis |
| | Hepatocellular carcinoma or hepatic adenoma |
| | Systemic hepatobiliary pathology (PBC, PSC, autoimmune hepatitis) |

The "Not Fatty Liver" label consisted of ultrasound images from patients identified retrospectively through PACS system after undergoing a right upper quadrant ultrasound or abdominal ultrasound for a variety of indications with a normal liver function panel and without ultrasound findings of hepatic steatosis, chronic

hepatitis, or other acute pathologic findings. Importantly, patients with type II diabetes mellitus were excluded from the "Not Fatty Liver" label to limit the inclusion of sub-clinical or undiagnosed NAFLD. However, because clinically patients within this label were not considered to necessarily be at risk for NAFLD, most did not have confirmatory MRI fat quantification or tissue histology. Inclusion and exclusion criteria for the "Not Fatty Liver" label is further detailed in Table 2. A goal of 5 unique images per patient was predicted.

**Table 2** Inclusion and exclusion criteria for patients within the "Not Fatty Liver" label

| Inclusion Criteria | Exclusion Criteria |
| --- | --- |
| Right upper quadrant or abdominal ultrasound | Abnormal liver function panel |
| No sonographic findings of hepatic steatosis | Alcoholic hepatitis or uncontrolled viral hepatitis |
| No acute pathologic findings present on examination | Hepatocellular carcinoma or hepatic adenoma |
| Retrievable through the PACS system | Increased renal echogenicity or hydronephrosis |
| Age > 18 | Biliary dilatation or other biliary abnormality |
| | Cholecystectomy or other hepatobiliary operation within the last 6 months |
| | Systemic hepatobiliary pathology (PBC, PSC, autoimmune hepatitis) |
| | Clinical diagnosis of possible NAFLD |
| | Type II diabetes mellitus |

Images were subsequently cropped in a manual fashion by a non-radiologist to only include the ultrasound image to eliminate nondiagnostic image annotations. Samples of cropped images within the "Fatty Liver" label and within the "Not Fatty Liver" label are provided in Figure 1. This set of images was considered the "Comprehensive Image Group." An independent board-certified radiologist (H.N with 3 years of experience and who did not serve as a blinded reader) then reviewed all images and selected images considered technically adequate for diagnosis to produce the "Radiology Selected Image" group. Both image groups were independently uploaded to Cloud AutoML Visionbeta for model creation, optimization, and analysis.
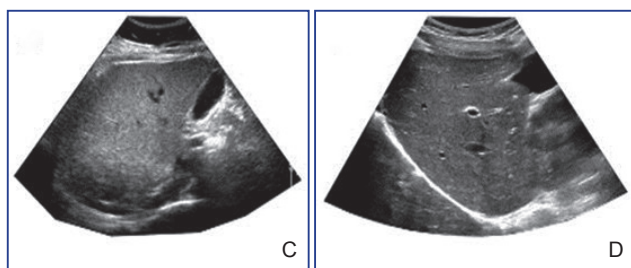


**Figure 1** Panel A demonstrates B-Mode hepatic ultrasound of a patient with known NAFLD. Panel B demonstrates B-mode hepatic ultrasound of a patient without clinical evidence of NAFLD. The gallbladder is partially within view in both windows.

Cloud AutoML Visionbeta allows for the creation of custom models trained on uploaded images using a convolutional neural network pre-trained through transfer learning. Of note, this product is within the beta launch stage. Image groups were sampled randomly by AutoML into a training set (roughly 80%), validation set (roughly 10%), and test set (roughly 10%); this was performed separately for both the "Comprehensive Image" group and "Radiology Selected Image" group. The training set is used for model training and the validation set is used for internal hyper- parameter optimization. The final test set (for which the model is naïve to) is used for evaluation of the model and for comparison against expert analysis.

The test set from the "Comprehensive Image" group was first used for blinded interpretation and assessment by three-independent board-certified radiologists with no additional clinical information available. After at least one month (to limit recall bias), these same radiologists were presented the test set from the "Radiologist Selected Image" group. The only interpretation that was asked of the radiologists during each assessment is whether they would make the diagnosis of hepatic steatosis. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy was subsequently calculated for the model as well as for the mean statistics of three blinded readers (L.N, A.L, and P.OK, each with at least 15 years of experience in ultrasound) for comparison against the gold standard label assignment; tissue histology or MRI fat quantification for the "Fatty Liver" label and a lack of any clinical data suggesting NAFLD required for the "Not Fatty Liver" label.

Demographic data (age, gender, and body mass index as calculated per National Institute of Health) was pooled per image group and compared among labels [10]. Independent sample t-test and chi-square analysis was performed on SPSS v25 (IBM, Armonk, New York). Test metrics were calculated on Excel (Microsoft, Redmond, WA). Clopper-Pearson confidence interval was calculated for non-mean test metrics. A p-value of 0.05 was designated as significant when utilized. Performance metrics are presented as means with 95% confidence intervals (CI).

## Results

Figure 2 diagrams the flow of enrolled patients and ultrasound images. Patients within the "Fatty Liver" label came from two sources from PACS: tissue histology ($N = 33$ patients, 102 pictures) and MRI fat quantification ($N = 63$ patients, 303 pictures), resulting in a total of 405 pictures. The "Not Fatty Liver" label consisted of a total of 100 patients with 500 ultrasound pictures. Thus, the "Comprehensive Image" group consists of a total of 905 pictures. From this, the "Radiology Selected Image" group consisted of a total of 649 images (292 within the "Fatty Liver" label and 357 from the "Not Fatty Liver" label).

Univariate analysis of patients within the "Comprehensive Image" group and "Radiology Selected" group as stratified per "Fatty Liver" label and "Not Fatty Liver" label is detailed in Table 3. Within both image groups, patients with the "Fatty Liver" label had a statistically significantly higher BMI compared with the "Not Fatty Liver" label ($P < 0.0001$ in both). The mean BMI of patients designated with the "Not Fatty Liver" label within both the "Comprehensive Image" group (28.3 ± 7.7) and "Radiology Selected Image" group (28.0 ± 7.6) fell within the National Institutes of Health BMI stratification as "overweight" (BMI 25 – 30)[10]. In contrast, the mean BMI of patients with the "Fatty Liver" label within both the "Comprehensive Image" group (34.5 ± 7.2) and "Radiology Selected Image" group (34.6 ± 6.7) fell within the National Institutes of Health BMI stratification as "obese" (BMI 30+)[10].
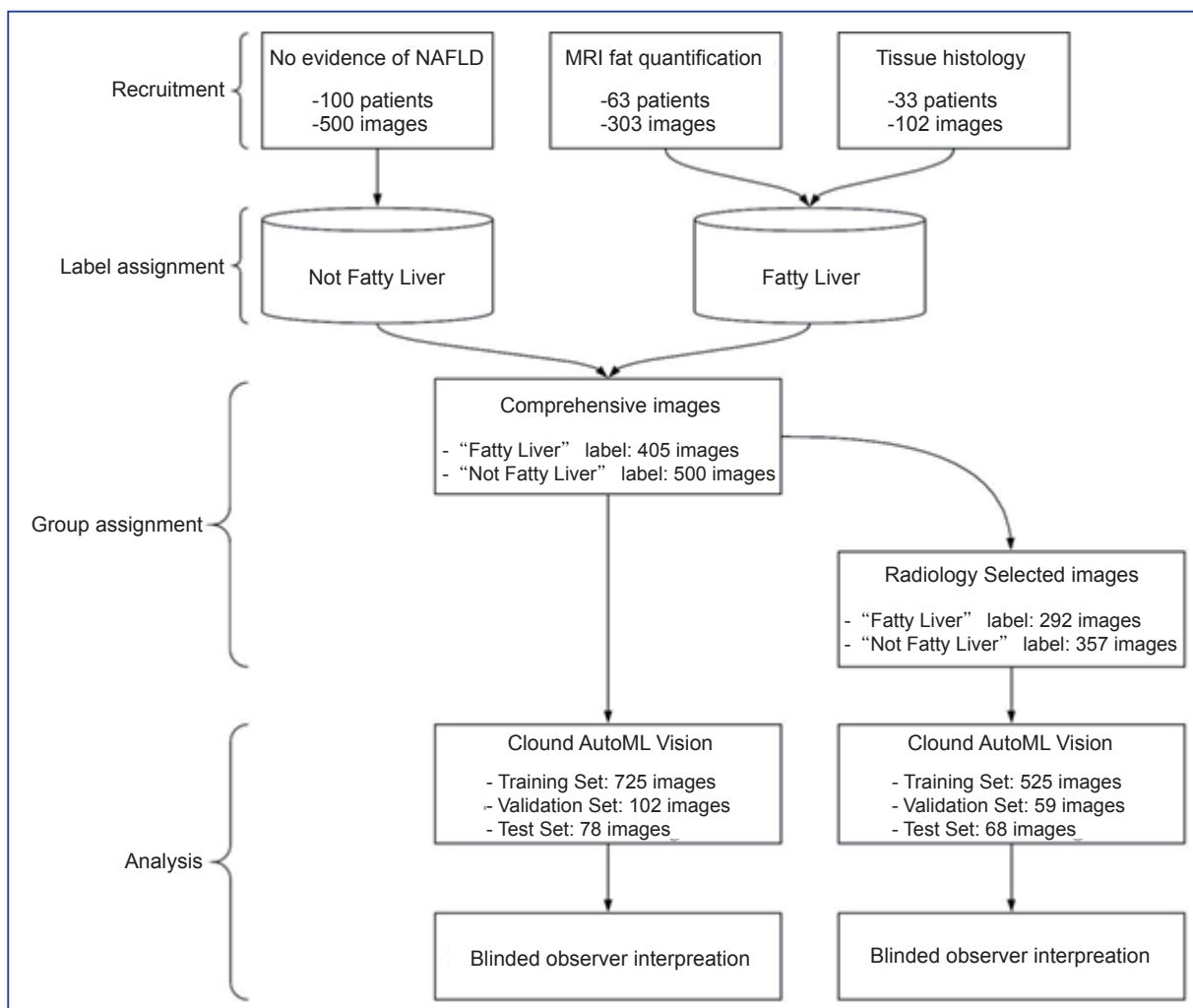


**Figure 2**   Workflow of data used for model creation and validation.

Additionally, the median hepatic steatosis grade by pathology for patients within the "Fatty Liver" label in both image groups was grade 2 moderate steatosis (33 – 66% of the examined histologic sample surface area visually determined to be involved by steatosis) [11]. The mean fat quantification percentage by PDFF MRI in patients with the "Fatty Liver" label within the "Comprehensive Image" group (17.3% ± 8.0) and "Radiology Selected Image" group (16.9% ± 7.8) both fell within the range of grade 1 (6.5 – 17.4%) mild hepatic steatosis but approached grade 2 (17.5 – 22.1%) moderate hepatic steatosis [9].

Randomized sampling of each image group was performed from within the Cloud AutoML Visionbeta API for model testing, validation, and testing. Within the "Comprehensive Image" group, 725 images were used for training, 102 images were used for validation, and 78 images were used for testing. Within the "Radiology Selected Image" group, 525 images were used for training, 59 images were used for validation, and 68 images were used for testing.

Furthermore, because these sets were populated randomly, the testing set for the "Comprehensive Image" group is different from the testing set in the "Radiology Selected Image" group; as are the corresponding training and validation sets. The model produced from the "Comprehensive Image" group performed with a sensitivity of 88.6% (95% CI = 73.3 – 96.8%) and a specificity of 95.3% (95% CI = 84.2 – 99.4%) for the detection of NAFLD. Similarly, PPV and NPV was 93.9% (95% CI = 79.9 – 98.4%) and 91.1% (95% CI = 80.3 – 96.3%), respectively. The overall accuracy

of the model was 92.3% (95% CI = 84.0 – 97.1%). Comparatively, the board-certified ultrasound trained radiologists performed with a mean sensitivity of 81.0% (95% CI = 74.3 – 87.6%), mean specificity of 86.0% (95% CI = 72.6 – 99.5%), mean PPV of 84.6% (95% CI = 72.4 – 96.7), and mean NPV of 85.1% (95% CI = 81.8 - 88.3%). A calculated Fleiss kappa score of inter-observability demonstrates moderate agreement at 0.636. Test metrics with a 95% CI are detailed in Table 4.

The model produced from the "Radiology Selected Image" group performed with a sensitivity of 88.6% (95% CI = 73.3 – 96.8%), specificity of 87.9% (95% CI = 71.8 – 96.6%), PPV of 88.6% (95% CI = 75.4 – 95.1%) and NPV of 87.9% (95% CI = 74.1 – 94.8%). The overall accuracy of the model is 88.2% (95% CI = 78.1 – 94.8%). The radiology group performed with a mean sensitivity of 92.4% (95% CI = 86.9 – 97.9%), mean specificity of 91.9 % (95% CI = 83.4 – 100%), mean PPV of 93.1% (95% CI = 86.1% – 100), and mean NPV of 92.4% (95% CI = 87.8 – 97.0%). A calculated Fleiss kappa score of inter-observability demonstrates improved agreement at 0.784. Test metrics with a 95% CI are detailed in Table 4.

As a secondary analysis, test statistics for consensus read by the radiology group were calculated for each image group. That is, the more common diagnosis made by the three radiologists for each test image was selected as the final interpreting diagnosis and compared against the true label. Within the "Comprehensive Image" group, a sensitivity of 88.6% (95% CI = 73.3 – 96.8%), specificity of 93.0% (95% CI = 80.9 – 98.5%), PPV of 91.2% (95% CI = 71.5 – 96.9%) and NPV of 90.9% (95% CI = 79.9 – 96.2%) was reported.

**Table 3** Univariate analysis of patients enrolled in each image group

| Item | "Comprehensive Image" group | | | "Radiology Selected Image" group | | |
| --- | --- | --- | --- | --- | --- | --- |
| Label | "Not Fatty Liver" | "Fatty Liver" | *P* value | "Not Fatty Liver" | "Fatty Liver" | *P* value |
| Mean Age | 44.5 (± 16.7) | 50.5 (± 13.3) | 0.006 | 44.1 (± 16.7) | 51.0 (± 13.3) | 0.002 |
| Gender | 29.0% male | 40.6% male | 0.087 | 29.2% male | 44.4% male | 0.035 |
| Mean BMI | 28.3 (± 7.7) | 34.5 (± 7.2) [a] | 0.001 | 28.0 (± 7.6) | 34.6 (± 6.7) [c] | 0.001 |
| Median Steatosis Grade by Histology | NA | 2 (*N* = 31 pts) [b] | NA | NA | 2 (*N* = 21 pts) [d] | NA |
| Mean % Fat by MRI | NA | 17.3% (± 8.0) (*N* = 63 pts) | NA | NA | 16.9% (± 7.8) (*N* = 60 pts) | NA |

[a] 6 patients omitted due to no recorded BMI within a year of evaluation; [b] 2 patients omitted due to no reported hepatic steatosis grade in pathology;
[c] 5 patients omitted due to no recorded BMI within a year of evaluation; [d] 2 patients omitted due to no reported hepatic steatosis grade in pathology

Within the "Radiology Selected Image" group, a sensitivity of 97.1% (95% CI = 85.1 – 99.9%), specificity of 93.9% (95% CI = 79.8 – 99.3%), PPV of

94.4% (95% CI = 81.6 – 98.5%), and NPV of 96.9% (95% CI = 81.8 – 99.5%). Test metrics with a 95% CI are detailed in Table 4.

## Discussion

The early and accessible diagnosis of NAFLD is important, especially given its increasing prevalence and potential for irreversible hepatic injury. As a modality, ultrasound is well poised to meet this clinical need. To this end, machine learning may have the potential to accurately and automatically assist in making an NAFLD diagnosis on B-mode ultrasound.

The model produced by the "Comprehensive Image" group performed with a sensitivity 88.6% and specificity 95.3% that was comparable to that of three board-certified ultrasound trained radiologists. Even after selection of only images considered diagnostic by an independent radiologist (the model produced by the "Radiology Selected Image" group), the model performance remained comparable. It is possible that by explicitly selecting for images considered diagnostic by a trained radiologist, the "Radiology Selected Image" group is now implicitly biased towards improving the ability of observers to detect NAFLD. However, even despite this, the model still performed well, with an overall accuracy of 88.2%.

As a secondary analysis, test statistics utilizing consensus ultrasound interpretation was calculated (Table 4). As expected, all four parameters (sensitivity, specificity, PPV, and NPV) improved with consensus compared with independent mean test statistics in both image groups (though remained comparable within a 95% confidence interval). This suggests that observer interpretation accuracy may be improved when pooled across the experience of multiple experienced radiologists. To that end, the potential for improvement in model performance still exists.

Machine-learning for sonographic decision support in the diagnosis of NAFLD is an active area of research, with accuracies between 80 – 98% reported [12]. However, limitations in methodologies exist including low label sizes and limited feature dimensionality. Alternative transfer learning neural network algorithms with manual hyperparameter optimization have been applied to the assessment of hepatic steatosis [13]. Cloud AutoML Visionbeta and other automated hyperparameter optimization protocols provide the advantage of rapidly and automatically optimizing the model per the user's application.

Machine learning presents a notable challenge in the medical radiologic space. The "black box" of deep learning systems makes independent expert validation difficult without the use of test sets for performance evaluation. Additionally, pathologies can be radiologically subtle and require a volumetric comprehension of a series of images [14]. Ultrasound offers further challenges including variable pathology visualization (given operator dependency), significant variation in technical parameterization (such as gain, depth, focal zone, resolution, and instrumentation), and an even more complex volumetric comprehension across all three axes in real-time. While AI-based identification methods are relatively new to ultrasound as an imaging modality, studies are increasingly utilizing these approaches to improve or automate diagnosis [15,16]. Creating large and standardized image databases is an important step in attempting to overcome these obstacles by attempting to encompass the wide variety of case examples.

Consequently, several limitations exist in this initial exploration. In the absence of feature extraction, is possible that irrelevant features were recognized as important in designating test image labels. To that end, images were manually cropped (by a non-radiologist observer) and images did span across a variety of ultrasound manufacturers with different technical parameters. Overfitting to irrelevant features was attempted to be overcome by aiming for the higher end of the recommended size of 100 – 500 images per label using sample size augmentation by including multiple different images per patient (divided among training, validation, and test sets). However, this does increase the risk of data leakage. Additionally, due to an increasing presence of NAFLD in the general population, it is possible that in the absence of chemical-shift MRI or liver biopsy, patients within the "Not Fatty Liver" group may have had subclinical or unrecognized NAFLD. For example, retrospective quality control found that 5 of the 100 patients within the "Not Fatty Liver" label later had a diagnosis of type II diabetes (perhaps not unexpected given the BMIs presented in table 3). The use of the described exclusion criteria was used to attempt to limit the chance of undiagnosed NAFLD in this population. Similarly, as patients were identified through retrospective chart review, the veracity of medical documentation is not always guaranteed. Finally, because NAFLD represents systemic hepatic disease, it is possible that within any single ultrasound image, variable degrees of hepatic steatosis may be present.

Despite these limitations, extensive future directions exist. Specifically, the image database and training set may be improved by standardizing the examination technique and technical specifications (such as the ultrasound machine and parameters). Additionally, identification of the relevant significant features used by the model for label discrimination may be used for clinical and research applications. Moving forward it may be possible to create a "Fatty Liver Ultrasound Index", in which the responsible clinician is provided a score to interpret.

## Conflict of Interest

The authors have no conflict of interest to declare.

## References

[1] Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016; 64: 73-84.

[2] Younossi ZM, Stepanova M, Afendy M, Fang Y, Younossi Y, Mir H, et al. Changes in the prevalence of the most common causes of chronic liver diseases in the united states from 1988 to 2008. *Clin Gastroenterol Hepatol* 2011; 9: 524-30.

[3] Sheth SG, Chopra S. Epidemiology, clinical features, and diagnosis of nonalcoholic fatty liver disease in adults. Waltham, MA: UpToDate Inc; 2018.

[4] Cho CS, Curran S, Schwartz LH, Kooby DA, Klimstra DS, Shia J, et al. Preoperative radiographic assessment of hepatic steatosis with histologic correlation. *J Am Coll Surg* 2008; 206: 480-8.

[5] Springer F, Machann J, Claussen CD, Schick F, Schwenzer NF. Liver fat content determined by magnetic resonance imaging and spectroscopy. *World J Gastroenterol* 2010; 16: 1560-6.

[6] Borra RJ, Salo S, Dean K, Lautamaki R, Nuutila P, Komu M, et al. Nonalcoholic fatty liver disease: Rapid evaluation of liver fat content with in-phase and out-of-phase MR imaging. *Radiology* 2009; 250: 130-6.

[7] Hernaez R, Lazo M, Bonekamp S, Kamel I, Brancati FL, Guallar E, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: A meta-analysis. *Hepatology* 2011; 54: 1082-90.

[8] Mottin CC, Moretto M, Padoin AV, Swarowsky AM, Toneto MG, Glock L, et al. The role of ultrasound in the diagnosis of hepatic steatosis in morbidly obese patients. *Obes Surg* 2004; 14: 635-7.

[9] Tang A, Tan J, Sun M, Hamilton G, Bydder M, Wolfson T, et al. Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. *Radiology* 2013; 267: 422-31.

[10] Janssen I, Katzmarzyk PT, Ross R. Body mass index, waist circumference, and health risk: Evidence in support of current national institutes of health guidelines. *Arch Intern Med* 2002; 162: 2074-9.

[11] Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the american association for the study of liver diseases. *Hepatology* 2018; 67: 328-57.

[12] Acharya UR, Fujita H, Bhat S, Raghavendra U, Gudigar A, Molinari F, et al. Decision support system for fatty liver disease using gist descriptors extracted from ultrasound images. *Information Fusion* 2016; 29: 32-39.

[13] Byra M, Styczynski G, Szmigielski C, Kalinowski P, Michalowski L, Paluszkiewicz R, et al. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images. *Int J Comput Assist Radiol Surg* 2018; 13: 1895-903.

[14] Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep learning: A primer for radiologists. *Radiographics* 2017; 37: 2113-31.

[15] Daniels K, Gummadi S, Zhu Z, Wang S, Patel J, Swendseid B, et al. Machine learning by ultrasonography for genetic risk stratification of thyroid nodules. *JAMA Otolaryngol Head Neck Surg* 2019, 10.1001/jamaoto.2019.30731-6.

[16] S Wang, JB Liu, ZY Zhu, Eisenbrey J. Artificial intelligence in ultrasound imaging: Current research and applications. *Advanced Ultrasound in Diagnosis and Therapy* 2019; 3: 53-61.