Published in final edited form as:

Ultrasound Med Biol. 2020 October ; 46(10): 2667–2676. doi:10.1016/j.ultrasmedbio.2020.05.016.

Image Processing Pipeline for Liver Fibrosis Classification Using Ultrasound Shear Wave Elastography

Laura J. Brattain, PhD^{1,2}, Arinc Ozturk, MD², Brian A. Telfer, PhD¹, Manish Dhyani, MD³, Joseph R. Grajo, MD⁴, Anthony E. Samir, MD, MPH²

¹MIT Lincoln Laboratory, Lexington, MA, USA

²Center for Ultrasound Research & Translation, Department of Radiology Massachusetts General Hospital, Boston, MA, USA

³Department of Radiology, Lahey Hospital & Medical Center, Burlington, Massachusetts, USA

⁴Abdominal Imaging Division, Department of Radiology, University of Florida College of Medicine, Gainesville, FL, USA

Abstract

The purpose of this study was to develop an automated method for classifying liver fibrosis stage

F2 based on ultrasound shear-wave elastography (SWE), and to assess the system performance in comparison with a reference manual approach. The reference approach consists of manually selecting a region of interest from each of 8 or more SWE images, computing the mean tissue stiffness within each of the regions of interest, and computing a resulting stiffness value as the median of the means. The 527-subject database consisted of 5,526 SWE images and pathologist-scored biopsies, with data collected from a single system at a single site. The automated method integrates three modules that assess SWE image quality, select a region of interest from each SWE measurement, and perform machine learning-based, multi-image SWE classification for fibrosis stage F2. Several classification methods were developed and tested using five-fold cross validation with training, validation and test sets partitioned by subject. Performance metrics were: area under receiver operating characteristic (AUROC), specificity at 95% sensitivity, and the number of SWE images required. The final automated method yielded an AUROC of 0.93 (95% CI: 0.90–0.94) versus 0.69 (95% CI: 0.65–0.72) for the reference method, 71% specificity with 95% sensitivity versus 5%, and four images per decision versus 8 or more. In conclusion, the automated method reported in this study significantly improved the accuracy for F2 classification of SWE measurements as well as reducing the number of measurements needed, which has the potential to reduce clinical work flow.

Corresponding author contact info: Laura J. Brattain, MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02420, brattainl@ll.mit.edu, Phone: 781-999-5595.

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.

Keywords

Liver fibrosis; machine learning; multi-image classification; shear wave elastography; single-image classification

INTRODUCTION

Chronic liver disease (CLD) comprises a wide variety of diffuse hepatocellular disease processes that consume substantial healthcare and financial resources both in the United States and worldwide (Udompap et al. 2015). CLD may progress to irreversible end stage liver fibrosis, a condition termed cirrhosis. The burden of cirrhosis continues to increase in the United States, with an estimated prevalence of approximately 0.27% of all American adults (Scaglione et al. 2015).

Liver biopsy histopathology is the current CLD diagnosis and staging reference standard. Liver biopsy allows for analysis of hepatic inflammation, fibrosis, steatosis, and necrosis, but is limited by sampling error, intra- and inter-observer variability, morbidity, rare mortality, and cost (Sumida et al. 2014). Ultrasound shear wave elastography (SWE) shows promise as an inexpensive and noninvasive method for detecting advanced liver fibrosis (Dhyani et al. 2015). This technology uses ultrasound to generate and sense tissue shear waves, from which Young's modulus, i.e., tissue stiffness, is computed (Ozturk et al. 2018a). Several studies have shown areas under the receiver operating characteristic (AUROCs) ranging from 0.77 to 0.94 for detection of liver fibrosis METAVIR stage F2 (Feng et al. 2016).

Currently, measurement of tissue stiffness and fibrosis staging require an operator to manually select tissue, using ultrasound imaging guidance, for SWE excitation and then to measure tissue stiffness averaged across a manually selected region of interest within the excited tissue, and also by averaging the tissue stiffness from multiple (8 – 10 typical) independently acquired images. The manual nature of this current process increases inter-operator variability and the potential for errors, such as inadvertent inclusion of blood vessels, bile ducts, or the liver capsule in the region of interest, producing high shear wave velocities that are not representative of the underlying liver parenchyma (Ferraioli et al. 2015; Ferraioli et al. 2018). The study hypothesis is that automating image quality assurance, region of interest selection, and fibrosis stage classification will address these limitations.

This paper focuses on fibrosis stage classification F2, which is associated with elevated liver disease specific mortality, especially in patients infected with the hepatitis C virus (HCV) (Xu et al. 2015) and in non-alcoholic steatohepatitis (Angulo et al. 2015). These are among the most common human diseases and are associated with severe morbidity and large public health expenditure (Blachier et al. 2013; Thrift et al. 2016).

The purpose of this study is to develop an automated method for classifying liver fibrosis stage F2 based on SWE, and to assess the system performance (sensitivity, specificity, and the number of SWE measurements used) in comparison with the current manual approach.

This paper expands on initial work (Brattain et al. 2018b) with a larger study population, refined methods, improved results, and more in-depth analysis.

MATERIALS AND METHODS

Data Collection

This retrospective study was approved by the MGH Institutional Review Board (IRB) and was compliant with the Health Insurance Portability and Accountability Act. Informed consent was waived by IRB. Data from subjects (age 16–83) who underwent SWE examination immediately prior to their liver biopsy during the period January 2014 – September 2017 were included. Patients who underwent biopsy for liver allograft evaluation were excluded, as inflammation due to graft rejection may confound SWE assessment. Subjects known to have the rare diseases of granulomatous hepatitis and hepatic amyloidosis were also excluded as these diseases are known to confound SWE measurement (Trifanov et al. 2015).

All SWE imaging was performed by a sonographer with more than 5 years of experience using a SuperSonic Imagine Aixplorer (Aix-en-Provence, France) ultrasound imaging system. SWE images were acquired by protocol in the right anterior oblique position with a convex broadband probe (SC6–1, SuperSonic Imagine) during a subjectively small breath hold. All patients were fasting for at least six hours prior to sonography. A minimum of 8 SWE measurements were obtained per case from the upper right lobe of the liver, as this region is known to reflect the SWE measurements with highest accuracy (Samir et al. 2015). Biopsies were assessed with standard histopathology scoring systems by 6 subspecialist pathologists with more than 5 years of experience who were blinded to the SWE values and clinical information. METAVIR (Bedossa and Poynard 1996) was used as the standard biopsy scoring system for the machine learning analysis. Each biopsy was assigned a single METAVIR score ranging from F0 (no fibrosis) to F4 (cirrhosis).

Patient demographics

The database contained 527 subjects (230 men and 297 women, mean age 49.3 years), total of 5526 images. 139 subjects had fibrosis stage F2. The detailed distribution is provided in Table 1 Column 2. Preliminary analysis with 328 patients was previously reported (Brattain et al. 2018b). The current study expands on this by having a larger patient population, analyzed with refined methods.

Algorithm Design

To compute a previously evaluated manual baseline (or reference) score (Samir et al. 2015), the mean stiffness values measured manually by sonographers were first read from the stored Digital Imaging and Communications in Medicine (DICOM) images. The median of the mean values extracted from the 8 or more SWE images for each patient was then computed as the score. SWE images were also extracted from stored DICOM images. To address variability in SWE manual acquisition and measurement, the following image processing steps were implemented: (1) automated image quality check, (2) automated

Region of Interest (ROI) selection, and (3) fibrosis classification (Figure 1). Binary classification is evaluated to differentiate fibrosis stages of F2 vs. F0-F1.

The size of the SWE image box varies. A typical size in this dataset is 72 pixel $\times$ 72 pixel (~40 mm $\times$ 40 mm). Automated image quality assessment was performed by computing the ratio of the sum of the number of valid pixels (pixels for which a Young's modulus was measured by the device) to the total number of pixels for each SWE image, yielding a metric termed the Percentage of Color Fill-In (PCFI). Only images with PCFI above a threshold were included for further analysis. To select a threshold for the subsequent image processing, PCFI was computed for candidate thresholds from 50% to 80% at 5% increments. For each candidate threshold, the baseline score from the qualified images was calculated, and an AUROC value was computed.

Automated region of interest selection was performed by scanning a 32 $\times$ 32 pixel ROI over the SWE image at 1-pixel spacing to produce numerous candidate ROIs. Of all candidate ROIs, the ROI with the smallest standard deviation (SD) was selected for further analysis. This is based on previous studies which demonstrated that homogeneity is a useful consideration for ROI selection (Dhyani et al. 2015; Samir et al. 2015).

Two types of classifier inputs were investigated: 1) features to characterize the stiffness distribution (mean, mode, median, skew, kurtosis, entropy, and 54 histogram features), and 2) all 32 $\times$ 32 SWE pixel values in an ROI. Several classifiers were assessed: support vector machine (Cortes and Vapnik 1995), random forest (Breiman 2001), Convolutional Neural Network (CNN) (Krizhevsky et al. 2012), and Fully Connected Neural Network (FCNN) (Krizhevsky et al. 2012). See supplementary information for details. For multi-image classification, the mean output was computed from varying numbers of images chosen randomly from the qualified images per case.

Statistical Analysis

Classification performance was evaluated by computing the AUROC. Cross-validation was used such that SWE images were randomly assigned by subject to a training, validation, and test set, with 72%, 8% and 20% of the subjects allocated to each set, respectively. Dividing the data by subject ensures that images from one subject are only used in one of the sets above. This avoids leakage in model training and testing. The classifier was trained based on the training and validation set and tested on the independent test set. A set of true positive rates (TPRs) and false positive rates (FPRs) was generated by varying a threshold on the classifier output on the test set. This procedure was repeated for five different randomly selected training, validation and test sets, which is referred to as five-fold cross-validation. (Five-fold cross validation was sufficient to establish statistically significant results, so more folds were not tested.) The means of the TPRs from the five folds were used to generate the overall ROC and AUROC. The 95% confidence interval (CI) of the AUROC across the five folds was generated with a bootstrapping technique (Efron and Tibshirani 1986) to randomly draw samples with replacement from the mean of all TPRs and their corresponding FPRs, after which AUROCs were computed. This process was repeated for 1000 trials, which was sufficient for the CI to converge. MATLAB (2017b, Math Works, Natick, MA) was used for the analysis. Statistical confidence in AUROC values was also quantified by computing p-

value. In this case, the AUROC uncertainties were assumed to be Gaussian, a valid assumption for large N for ranksum tests (Mann-Whitney U tests) such as AUROC.

RESULTS

Automated Image Quality Check

Figure 2A shows an example of a good quality SWE image, where the SWE box is filled with color pixels, each representing a SWE measurement. (The color coding is used so that the SWE values can be distinguished from the gray-value B-mode ultrasound image upon which the SWE image is superimposed.) Figure 2B is an example of a poor quality SWE image, where the SWE box contains very few color pixels. Good quality images were selected for subsequent processing by applying an empiric threshold to the PCFI metric. A PCFI threshold of 55% significantly improved baseline method AUROC compared with baseline method AUROC without PCFI. While the AUROCs for higher PCFI thresholds were slightly better, a substantially higher number of images were also classified as low quality and removed. Subsequent image processing steps, namely ROI selection and classification, were evaluated using qualified cases (total 515 of 527, 137 cases of F2) and images (total 4075 of 5526) at PCFI of 55%. Table 1 Column 3 shows the distribution details for this subgroup.

Automated ROI Selection

Figure 3 shows examples of algorithm-selected ROIs in comparison to manually selected ROIs. In Figure 3A, the manually selected ROI is at a desired location, and the two ROIs match well. In Figure 3B, the manually selected ROI is erroneously positioned immediately adjacent to a blood vessel, as indicated by the red arrow in the ultrasound image at the bottom of Figure 3B. The estimated tissue Young's modulus of 29.9 kPa in this region is far higher than that typically encountered with liver tissue. As a comparison, the algorithm selected ROI (inset image) is chosen at a more representative location, away from the blood vessel. This selection was subsequently validated as better placement by an expert operator.

Classification results

The manually computed baseline ROC is plotted in Figure 4A, with an AUROC of 0.69 (95% CI: 0.65–0.72). Applying the baseline method to images that passed the quality check improved the AUROC to 0.74 (95% CI: 0.70–0.77). The full pipeline of image quality check, ROI selection, and classification was tested next. The support vector machine, random forest and FCNN operating on the statistical features further improved the AUROC to 0.75 (95% CI: 0.71–0.78), 0.76 (95% CI: 0.72–0.79) and 0.80 (95% CI: 0.77–0.82), respectively, even though this classification is based only on a single image rather than the 10 images typically used in clinical practice. The CNN operating directly on SWE pixels from the ROI also yielded an AUROC of 0.80. The best results were obtained from the FCNN that operated on a combination of statistical features and SWE pixels from the ROI, with an AUROC of 0.87 (95% CI: 0.84–0.88). The FCNN thus significantly improved the AUROC over that of the baseline approach, 0.87 vs 0.69 ($p\text{-value} = 4 \times 10^{-7}$). In addition, at a sensitivity of 95%, a marked improvement in specificity was observed, increasing from 5%

with conventional technique to 58%. Sensitivities and specificities for the other classifiers can be seen from the ROCs in Figure 4.

FCNN results were further improved by combining classification outputs for multiple images per case. Figure 4B plots the ROCs based on varying the number of images from two to eight. As shown in the figure, when making a decision with four images, there was an additional significant performance gain with an AUROC of 0.93 (95% CI: 0.90–0.94) vs 0.87 (95% CI: 0.84–0.88), and p-value of 0.027. Moreover, this results in a further specificity increase to 71% at 95% sensitivity. Further increasing the number of images offered minimal gains, indicating four images per decision is a good design choice for this classification task.

Neural Network Interpretability

A common weakness of neural network classifiers is that the results cannot be explained, with the internal workings of the network being a “black box.” However, in this work, interpretation of the internal workings of the network (the network weights) may be insightful. The AUROC improvement produced by training directly on SWE pixels vs. summary statistics suggests spatial information not contained within conventional summary statistics is being exploited to improve classification accuracy. To test this hypothesis, FCNN hidden layer weights corresponding to the image pixels were inspected. The 12-weight images shown in Figure 5 display varying textures and orientations, at a spatial resolution that matches the SWE images. For a typical SWE image of 72 pixel $\times$ 72 pixel (~40 mm $\times$ 40 mm), the physical extent of these weight images is approximately 18 mm $\times$ 18 mm. These weight images suggest the best-performing classifier does exploit spatial information in the SWE images, not simply summary statistics. This interpretation, although preliminary, also suggests that future work should strive to explain the cause for these spatial variations in stiffness for different fibrosis stages.

DISCUSSION

Several noninvasive diagnostic methods have been proposed. Blood-based biomarkers and calculation methods including platelets, FIB-4, APRI and Hepascore have been studied in many clinical trials (Vilar-Gomez and Chalasani 2017). As one of the most commonly used blood-based biomarkers, FIB-4 provided 0.70 (95% CI, 0.68–0.72) of AUROC value to differentiate F2 fibrosis in a study with 1716 patients (Yen et al. 2018). Although blood-based biomarkers are easy to use and are low cost options, these methods may be confounded by several patient and disease related factors including patient age, presence of diabetes, acute inflammation, sepsis and cholestasis (Poynard et al. 2007; Vilar-Gomez and Chalasani 2018).

Imaging biomarkers are also used to diagnose patients with F2 fibrosis. Transient elastography (TE) is a commonly used imaging based technique which quantifies liver stiffness and steatosis (Ozturk et al. 2018b). In a recent meta-analysis, TE has been shown to have AUROC value of 0.71 (95%CI, 0.66–0.74) for differentiating F2 fibrosis (Xia et al. 2018). Although TE is an accurate technique for fibrosis staging, it does not provide real-time ultrasound imaging, which is considered useful for diagnosing liver fibrosis

complications, especially lesions in cirrhosis patients. In comparison, SWE is typically acquired in conjunction with ultrasound B-mode imaging, both at real-time. The current biggest limitation for a wider adoption of SWE is the operator dependence on SWE acquisition and ROI selection. This is one of the challenges this proposed work aims to address.

The goal of this work was to apply image processing combined with machine learning-based classification to improve noninvasive SWE-based liver fibrosis staging sensitivity and specificity, while simultaneously reducing the number of images required for clinical decision-making.

Automated image quality determination, and automated ROI selection were both shown to improve AUROC. A series of machine learning experiments established that additional information in the SWE imagery can be exploited to significantly improve AUROC compared with conventional technique. Combining additional statistical features through machine learning improved the AUROC, even when the number of images used was based on single image per decision, versus 8 and more images used in the baseline approach. Combining statistical features with the SWE ROI pixel values through a FCNN further improved the AUROC. An inspection of the neural network weights (Figure 5) suggests that the FCNN is exploiting spatial variations in the SWE ROI. Finally, using four images per decision resulted in an AUROC of 0.93, versus the baseline AUROC of 0.69 from 8 or more images per decision. Considering a high sensitivity operating point of 95% for detecting F2 cases, our approach improves specificity from 5% to 71%.

Many recent investigations of machine learning applied to ultrasound have been conducted, although less commonly for SWE (Brattain et al. 2018a). Recent work on automated liver fibrosis scoring has included machine learning based on both handcrafted features as well raw SWE images (Gatos et al. 2017; Wang et al. 2019). One study used texture features and a support vector machine classifier, operating on 85 SWE images from 85 cases (Gatos et al. 2017). Another study used a CNN for assessing fibrosis levels in patients with hepatitis B, operating on 1990 images from 398 cases (Wang et al. 2019). The CNN was trained on the entire 256×256 pixel SWE image box. This second study also evaluated multiple images per decision. A key difference between our work and (Wang et al. 2019) is that we included automated image quality checking and ROI selection before applying classifiers to the ultrasound imagery. These steps automate key operator tasks known to introduce variability and degrade fibrosis staging.

This study has limitations. First, multiple pathologists reviewed the liver biopsy samples. It is well known that high interobserver variability limits histopathologic liver fibrosis staging despite its wide acceptance as the reference standard (Nalbantoglu and Brunt 2014). Secondly, the current study was conducted with a single type of SWE system and at a single site. In the future, we plan on expanding our work to include data acquired by other machines and additional sites, thus improving the generalizability of our results. In addition, the current patient population covers a number of chronic liver disease etiologies, including Non-alcoholic Fatty Liver Disease (NAFLD), viral hepatitis, alcoholic hepatitis, and autoimmune liver disease. These diseases have different liver fibrosis patterns, which may

limit our ability to consider them as a group. In the future, we will stratify the analysis by disease type, and expand classification results to include F3 as well as F2. Lastly, patient factors, such as high skin-to-capsule distance and abdominal obesity are known to degrade shear wave propagation and SWE data quality (Staugaard et al. 2016; Yoneda et al. 2015). Additionally, severe steatosis and/or inflammation may lead to overestimate of liver stiffness (Raizner et al. 2017). In future work, we aim to incorporate clinical and non-SWE imaging feature sets to further mitigate patient-specific confounders and operator bias.

In conclusion, the automated method developed in this study significantly improved the accuracy of SWE liver fibrosis F2 classification while reducing the number of SWE measurements needed. By dramatically increasing specificity from 5% for the reference manual approach to 71% with 95% sensitivity, accuracy has increased to a level which could allow for noninvasive screening to identify a smaller set of candidates for invasive biopsies. The significant reduction in number of SWE images required, four instead of 8 – 10, has the potential to significantly reduce the total examination time, which can currently require 20 to 30 minutes. Thus, this noninvasive approach, when matured, will have the potential to improve clinical work flow, reduce the need for invasive liver biopsies, and enable quantitative, longitudinal monitoring of patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the National Institute of Biomedical Imaging and Bioengineering and National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under Award Nos. K23 EB020710 and R01 DK119860. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Institutes of Health.

REFERENCES

- Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwithaya P, Mills PR, Keach JC, Lafferty HD, Stahler A, Haflidadottir S, Bendtsen F. Liver Fibrosis, but no Other Histologic Features, Associates with Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2015;149:389–397. [PubMed: 25935633]
- Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;24:289–293. [PubMed: 8690394]
- Blachier M, Leleu H, Peck-Radosavljevic M, Valla DC, Roudot-Thoraval F. The burden of liver disease in Europe: A review of available epidemiological data. *Journal of Hepatology* 2013;58:593–608. [PubMed: 23419824]
- Brattain LB, Telfer BA, Dhyani M, Grajo JR, Samir AE. Machine Learning for Medical Ultrasound: Status, Methods, and Future Opportunities. *Abdominal Radiology* 2018a;43:786–799. [PubMed: 29492605]
- Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE. Objective Liver Fibrosis Estimation from Shear Wave Elastography. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018b;:1–5.
- Breiman L Random forests. *Machine Learning* 2001;45:5–32.
- Cortes C, Vapnik V. Support-Vector Networks. *Machine Learning* 1995;20:273–297.

- Dhyani M, Anvari A, Samir AE. Ultrasound elastography: liver. *Abdominal Imaging* 2015;40:698–708. [PubMed: 25690689]
- Efron B, Tibshirani R. Bootstrap methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* 1986;1:54–77.
- Feng JC, Li J, Wu XW, Peng XY. Diagnostic accuracy of supersonic shear imaging for staging of liver fibrosis: A meta-analysis. *Journal of Ultrasound in Medicine* 2016;35:329–339. [PubMed: 26795041]
- Ferraioli G, Filice C, Castera L, Choi BI, Sporea I, Wilson SR, Cosgrove D, Dietrich CF, Amy D, Bamber JC, Barr R, Chou YH, Ding H, Farrokh A, Friedrich-Rust M, Hall TJ, Nakashima K, Nightingale KR, Palmeri ML, Schafer F, Shiina T, Suzuki S, Kudo M. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 3: Liver. *Ultrasound in Medicine and Biology* 2015;41:1161–1179. [PubMed: 25800942]
- Ferraioli G, Wong VWS, Castera L, Berzigotti A, Sporea I, Dietrich CF, Choi BI, Wilson SR, Kudo M, Barr RG. Liver Ultrasound Elastography: An Update to the World Federation for Ultrasound in Medicine and Biology Guidelines and Recommendations. *Ultrasound in Medicine and Biology* 2018;44:2419–2440. [PubMed: 30209008]
- Gatos I, Tsantis S, Spiliopoulos S, Karnabatidis D, Theotokas I, Zoumpoulis P, Loupas T, Hazle JD, Kagadis GC. A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography. *Ultrasound in Medicine and Biology* 2017;43:1797–1810. [PubMed: 28634041]
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information and Processing Systems (NIPS)* 2012;60:84–90.
- Nalbantoglu I, Brunt EM. Role of liver biopsy in nonalcoholic fatty liver disease. *World Journal of Gastroenterology* 2014;20:9026–9037. [PubMed: 25083076]
- Ozturk A, Grajo JR, Dhyani M, Anthony BW, Samir AE. Principles of ultrasound elastography. *Abdominal Radiology* 2018a;43:773–785. [PubMed: 29487968]
- Ozturk A, Grajo JR, Gee MS, Benjamin A, Zubajlo R, Thomenius K, Anthony B, Samir A, Dhyani M. Quantitative Hepatic Fat Quantification in Non-alcoholic Fatty Liver Disease Using Ultrasound-based Techniques: A Review of Literature and Their Diagnostic Performance. *Ultrasound in Medicine and Biology* 2018b;44:2461–2475. [PubMed: 30232020]
- Poynard T, Morra R, Halfon P, Castera L, Ratzliff V, Imbert-Bismut F, Naveau S, Thabut D, Lebre C, Zoulim F, Bourliere M, Cacoub P, Messous D, Munteanu M, de Ledinghen V. Meta-analyses of FibroTest diagnostic value in chronic liver disease. *BMC Gastroenterology* 2007;7:1–11. [PubMed: 17257428]
- Raizner A, Shillingford N, Mitchell PD, Harney S, Raza R, Serino J, Jonas MM, Lee CK. Hepatic inflammation may influence liver stiffness measurements by transient elastography in children and young adults. *Journal of Pediatric Gastroenterology and Nutrition* 2017;64:512–517. [PubMed: 27540711]
- Samir AE, Dhyani M, Vij A, Bhan AK, Halpern EF, Méndez-Navarro J, Corey KE, Chung RT. Shear-wave elastography for the estimation of liver fibrosis in chronic liver disease: determining accuracy and ideal site for measurement. *Radiology* 2015;274:888–96. [PubMed: 25393946]
- Scaglione S, Kliethermes S, Cao G, Shoham D, Durazo R, Luke A, Volk ML. The Epidemiology of Cirrhosis in the United States A Population-based Study. *Journal of Clinical Gastroenterology* 2015;49:690–696. [PubMed: 25291348]
- Staugaard B, Christensen PB, Mössner B, Hansen JF, Madsen BS, Sørensen J, Krag A, Thiele M. Feasibility of transient elastography versus real-time two-dimensional shear wave elastography in difficult-to-scan patients. *Scandinavian Journal of Gastroenterology* 2016;51:1354–1359. [PubMed: 27310486]
- Sumida Y, Nakajima A, Itoh Y. Limitations of liver biopsy and non-invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World Journal of Gastroenterology* 2014;20:475–485. [PubMed: 24574716]
- Thrift AP, El-Serag HB, Kanwal F. Global epidemiology and burden of HCV infection and HCV-related disease. *Nature Reviews Gastroenterology & Hepatology* 2016;14:122.

- Trifanov DS, Dhyani M, Bledsoe JR, Misdraji J, Bhan AK, Chung RT, Samir AE. Amyloidosis of the liver on shear wave elastography: case report and review of literature. *Abdominal Imaging* 2015;40:3078–3083. [PubMed: 26254907]
- Udompap P, Kim D, Kim WR. Current and Future Burden of Chronic Nonmalignant Liver Disease. *Clinical Gastroenterology and Hepatology* 2015;13:2031–2041. [PubMed: 26291665]
- Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *Journal of Hepatology* 2017;68:305–315. [PubMed: 29154965]
- Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. *Journal of Hepatology European Association for the Study of the Liver*, 2018;68:305–315.
- Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, Wu C, Liu C, Huang L, Jiang T, Meng F, Lu Y, Ai H, Xie X, Yin L, Liang P, Tian J. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B : a prospective multicentre study. *Gut* 2019;68:729–741. [PubMed: 29730602]
- Xia B, Wang F, Friedrich-rust M, Zhou F, Zhu J, Yang H, Ruan W, Zeng Z. Feasibility and Efficacy of Transient Elastography using the XL probe to diagnose liver fibrosis and cirrhosis. *Medicine* 2018;97:1–8.
- Xu F, Leidner AJ, Tong X, Holmberg SD. Estimating the number of patients infected with chronic HCV in the United States who meet highest or high-priority treatment criteria. *American Journal of Public Health* 2015;105:1285–1289. [PubMed: 25973816]
- Yen Y, Kuo F, Kee K, Chang K, Tsai M, Hu T, Lu S, Wang J, Hung C, Chen H. APRI and FIB-4 in the evaluation of liver fibrosis in chronic hepatitis C patients stratified by AST level. *PLoS ONE* 2018;13:1–16.
- Yoneda M, Thomas E, Sclair SN, Grant TT, Schiff ER. Supersonic Shear Imaging and Transient Elastography With the XL Probe Accurately Detect Fibrosis in Overweight or Obese Patients With Chronic Liver Disease. *Clinical Gastroenterology and Hepatology* 2015;13:1502–1509. [PubMed: 25804329]

Figure 1.
SWE image processing pipeline structure.

Figure 2.
Example SWE images and tissue stiffness measurements. (A) Good quality SWE image with full elastogram filling. (B) Poor quality SWE with poor elastogram filling.

Figure 3.

Comparison of manually selected and algorithm selected ROI. (A) Algorithm selected ROI is in agreement with the manually selected ROI. (B) Manually selected ROI is near a vessel, which negatively affected the measurement. Algorithm selected ROI avoids the vessel.

Figure 4.

Classification results. (A) Baseline (orange) vs image quality measure (green) vs classifiers based on single image. FCNN is the best performing classifier (blue). (B) FCNN with multiple images. Decision from 1 and 10 images per patient. ROIs selected by algorithm. Specificity improvements for 95% sensitivity are shown by red dashed line with arrows.

Figure 5.

Visualization of the first hidden layer of a 12-neuron fully connected network. White corresponds to the most positive weights; black corresponds to the most negative. The 32 rows and 32 columns in each image correspond to the 32 rows and 32 columns in the region of interest.

Table 1.

Patient Demographics

	Original	After quality check
Total # of cases/images	527/5526	515/4075
Mean age	49.34 (16–83)	49.35 (16–83)
Gender (Male/Female)	230/297	225/290
Mean age by gender (Male/Female)	47.80/50.55	46.67/50.65
Fibrosis stage (METAVIR System)	N (%)	
FO - Absent	202 (38.3%)	198 (38.4%)
F1 - Enlarged fibrotic portal tract	186 (35.3%)	181 (35.1%)
F2 - Few portal-portal septa but intact architecture	71 (13.5%)	69 (13.4%)
F3 - Many septa with architectural distortion but no obvious cirrhosis	49 (9.3%)	48 (9.3%)
F4 - Cirrhosis	19 (3.6%)	19 (3.7%)