Ultrasound in Med. & Biol., Vol. 46, No. 1, pp. 26–33, 2020

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2019.09.004

● *Original Contribution*

A NEW MULTIMODEL MACHINE LEARNING FRAMEWORK TO IMPROVE HEPATIC FIBROSIS GRADING USING ULTRASOUND ELASTOGRAPHY SYSTEMS FROM DIFFERENT VENDORS

ISABELLE DUROT,*,† ALIREZA AKHBARDEH,* HERSH SAGREIYA,*
ANDREAS M. LOENING,* and DANIEL L. RUBIN*,‡,§

* Department of Radiology, School of Medicine, Stanford University, Stanford, California, USA; † Institute of Radiology, Cantonal Hospital Aarau, Aarau, Switzerland; ‡ Department of Biomedical Data Science, Stanford University, Stanford, California, USA; and § Department of Medicine (Biomedical Informatics Research), Stanford University, Stanford, California, USA

(*Received* 7 *March* 2019; *revised* 14 *July* 2019; *in final from* 8 *September* 2019)

Abstract—The purpose of the work described here was to determine if the diagnostic performance of point and 2-D shear wave elastography (pSWE; 2-DSWE) using shear wave velocity (SWV) with a new machine learning (ML) technique applied to systems from different vendors is comparable to that of magnetic resonance elastography (MRE) in distinguishing non-significant (<F2) from significant (≥F2) fibrosis. We included two patient groups with liver disease: (i) 144 patients undergoing pSWE (Siemens) and MRE; and (ii) 60 patients undergoing 2-DSWE (Philips) and MRE. Four ML algorithms using 10 SWV measurements as inputs were trained with MRE. Results were validated using twofold cross-validation. The performance of median SWV in binary grading of fibrosis was moderate for pSWE (area under the curve [AUC]: 0.76) and 2-DSWE (0.84); the ML algorithm support vector machine (SVM) performed particularly well (pSWE: 0.96, 2-DSWE: 0.99). The results suggest that the multivendor ML-based algorithm SVM can binarily grade liver fibrosis using ultrasound elastography with excellent diagnostic performance, comparable to that of MRE. (E-mail: dlrubin@stanford.edu) © **2019 World Federation for Ultrasound in Medicine & Biology. All rights reserved.**

Key Words: **Machine learning, Ultrasound, Liver fibrosis, Shear wave elastography.**

INTRODUCTION

Chronic liver disease, caused by hepatic injury of various etiologies, is a crucial global health problem with rising incidence. Precise disease staging is paramount for patient management, treatment recommendations and accurate prognosis (Ferraioli et al. 2015). Liver biopsy has classically been the gold standard for fibrosis staging; however, non-invasive imaging methods, such as transient elastography (Fibroscan), point shear wave elastography (pSWE), 2-D shear wave elastography (2-DSWE) and magnetic resonance elastography (MRE), have been reported to be at least as accurate with fewer complications (Afdhal et al. 2015; Lurie et al. 2015; Zhang et al. 2019). Two-dimensional SWE and pSWE provide liver stiffness information using acoustic radiation force impulses (Friedrich-Rust et al. 2012), and MRE uses an external passive driver to generate hepatic shear waves that are imaged by MRE pulse sequences (Trout et al. 2016). MRE has been reported to be highly reproducible and accurate for liver stiffness measurement (Cui et al. 2016), as has ultrasound elastography (D'Onofrio et al. 2010; Rizzo et al. 2011; Bota et al. 2012), although with somewhat lower accuracy: (area under the curve [AUC]: pSWE 0.81; 2-DSWE 0.88 [Sigrist et al. 2017]; MRE >0.9 [Shi et al. 2014]). Ultrasound elastography is cheaper than MRE and widely used in clinics; nonetheless, it lacks an ideal sensitivity and specificity in grading liver fibrosis (Sigrist et al. 2017), which can negatively influence patient care. Furthermore, ultrasound elastography cutoff values for grading liver fibrosis based on velocity or stiffness values vary among manufacturers (Sigrist et al. 2017; Ferraioli et al. 2019); thus, results are not interchangeable from one system to another. In addition, studies with the necessary population size to define or improve these cutoff values are becoming harder to conduct because of the lack of gold standard biopsies being performed. There is a critical need for robust cutoff values

Address correspondence to: Daniel L. Rubin, Department of Radiology, School of Medicine, Stanford University, 1265 Welch Road, Room X-335, MC 5464, Stanford, CA 94305-5621. E-mail: dlrubin@stanford.edu

for standardized hepatic fibrosis grading that can be applied to all systems and diseases (Dietrich et al. 2017). This important unaddressed concern has been raised in the literature (Sigrist et al. 2017).

In recent years, machine learning (ML) approaches in diagnostic radiology have emerged and gained prominence (Erickson et al. 2017). Prior studies incorporating machine or deep learning algorithms sought to improve liver fibrosis grading with ultrasound elastography (Stoean et al. 2011; Fujimoto et al. 2013; Chen et al. 2017; Gatos et al. 2017). Nonetheless, there is no published study that has assessed an ML technology for characterizing liver fibrosis using ultrasound elastography velocity measurements obtained with pSWE and 2-DSWE to train and validate a scoring system that is comparable to MRE for grading liver fibrosis, which can also be applied to systems from different vendors.

Therefore, the purpose of this study was to determine if the diagnostic performance of pSWE and 2-DSWE for grading liver fibrosis using shear wave velocity (SWV) with a new ML technique is comparable to that of MRE in distinguishing non-significant (<F2) from significant (≥F2) fibrosis and can be applied to ultrasound systems from different vendors.

METHODS

This HIPPA-compliant retrospective study was approved by the institutional review board of our institution, and the requirement for written consent was waived for all participating patients. Exclusion criteria were non-diagnostic MRE and unreliable ultrasound elastography with an interquartile ratio (IQR) divided by the median (IQR/median >0.3). Figure 1 summarizes the study design.

Patient population

Group 1. From April 2014 to February 2017, 169 ultrasound elastography exams (pSWE) were performed (86 men—mean age: 53.8 y, range: 23−75 y; 80 women—mean age: 56.9 y, range: 22−80 y) in patients who also underwent an MRE examination within 12 mo (this time frame was chosen based on discussions with hepatologists from our institution as well as evidence in the literature (Pan et al. 2018). Twenty-five of 169 patients (14.8%) were excluded because of unreliable exams. All enrolled patients (144/144, 100%) had known chronic liver disease or elevated liver enzymes (Table 1).

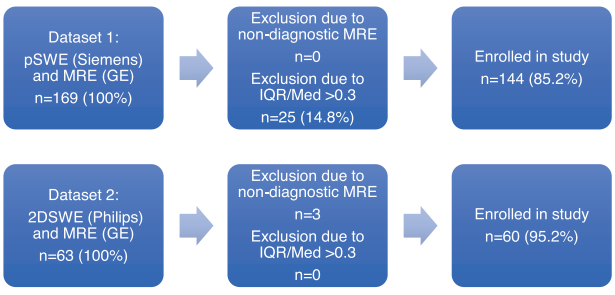


Fig. 1. Flow diagram of the enrollment process in this retrospective study. IQR = interquartile ratio; MRE = magnetic resonance elastography; pSWE = point shear wave elastography; 2-DSWE = 2-D shear wave elastography.

Group 2. From February 2016 to October 2017, 63 ultrasound elastography exams (2-DSWE) were performed (39 men: mean age: 53.9 y, range: 23−79 y; 24 women: mean age: 55.4 y, range: 22−73 y) in patients who underwent an MRE examination (median interval: 0 d, mean interval: 1 d). Three of 63 patients (4.8%) were excluded because of a non-diagnostic MRE exam. Chronic liver disease was known to be present in 58 of 60 enrolled patients (96.7%) (Table 1).

Ultrasound elastography image acquisition

Point SWE was performed in patients in group 1 in the Virtual Touch Tissue Quantification (VTTQ) mode on a clinical ultrasound scanner (Acuson S2000, Siemens Medical Solutions, Mountain View, CA, USA) coupled to a curved array transducer (6 C1 HD, Siemens Medical Solutions). Philips 2-DSWE (group 2) was performed using the prototype ElastQ software on an Epiq7 system coupled to a curved array transducer (C5-1, Philips Healthcare, Amsterdam, Netherlands).

Patients were asked to fast for at least 4 h before ultrasound imaging. SWV measurements of the liver were performed in group 1 by one of three sonographers with dedicated training in pSWE and in group 2 by one sonographer with dedicated training in 2-DSWE. Patients were

Table 1. Distribution of diseases between the two data sets

Diagnosis	pSWE + MRE	2-DSWE + MRE
Hepatitis B	50	4
Hepatitis C	41	10
Non-alcoholic fatty liver disease or steatohepatitis	19	23
Abnormal liver function studies	13	4
Alcohol abuse and alcoholic cirrhosis	7	7
Primary biliary cholangitis	6	1
Hemochromatosis cirrhosis	3	3
Cryptogenic cirrhosis	2	2
Autoimmune hepatitis		1
Drug-induced hepatitis		2
Budd−Chiari syndrome	1	
Morbus Wilson cirrhosis	1	
Cardiac cirrhosis	1	
Portal/mesenteric vein thrombosis		1
No known chronic liver disease		2

2-DSWE = 2-D shear wave elastography; MRE = magnetic resonance elastography; pSWE = point shear wave elastography.

placed in the supine position, and the right arm was elevated above the shoulder to widen the intercostal space. The regions of interest (ROIs; Siemens: 10 × 6 mm, Philips: 0.785 cm^2) were placed in liver segment 8 (Fig. 2). Ten consecutive SWV measurements (in m/s) were obtained from approximately the same location within 2 cm of Glisson's capsule and perpendicular to the liver capsule, without including large vessels or dilated bile ducts. Patients were asked to maintain breath-holding at a neutral position during measurements.

The results of all 10 measurements were automatically displayed by the systems at the end of the exam and either saved into the clinical picture archiving and communication system (PACS; Centricity; GE; group 1) or on an external hard drive disk (group 2).

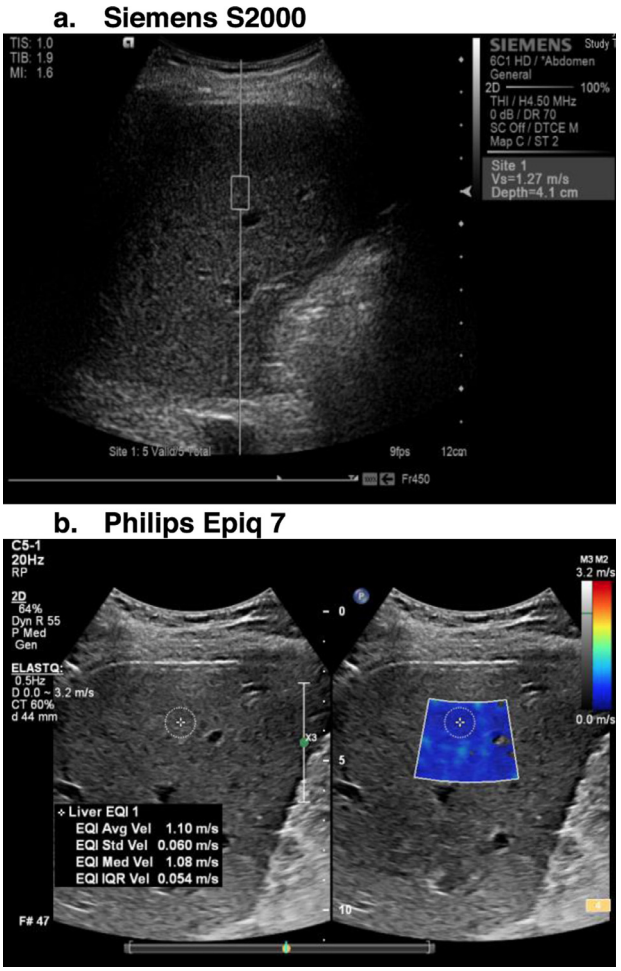


Fig. 2. Ultrasound elastography images of the liver in segment 8 (transverse plane) obtained in the two different groups. (a) Point shear wave elastography (pSWE) on a Siemens scanner in group 1 (51-y-old female patient with abnormal liver function studies). (b) Two-dimensional shear wave elastography (2-DSWE) on a Philips scanner in group 2 (42-y-old male patient with chronic hepatitis B).

Magnetic resonance elastography imaging acquisition

Patients were instructed to fast for 4 h before the MRE examination. All magnetic resonance (MR) elastography examinations were performed on a 3-T MR magnet (GE750, GE Healthcare, Waukesha, WI, USA) using a 32-channel torso phased-array receive coil, with a passive driver placed on the patient's right upper abdomen to allow the transmission of 60-Hz vibrations into the liver and a 2-D phase-sensitive echo-planar MR elastography sequence (MR-Touch, GE Healthcare). The sequence was acquired in a single expiratory breath hold (~20 s) with the passive driver activated. A direct inversion algorithm automatically created shear wave images and stiffness maps from the acquired data. Radiologists drew an ROI encompassing areas of the right hepatic lobe assessed to have reliable signal, measuring liver stiffness (complex shear modulus) in kilopascals.

Shear wave velocity-based grading and statistical analysis

When the 10 SWV measurements had an IQR divided by the median (IQR/median) >0.3, they were considered unreliable and excluded from the study (Ferraioli et al. 2015): group 1, 25 of 169 (14.8%); and group 2, 0 of 63 (0%). Non-diagnostic MREs were excluded from the study (group 1, $n=0$; group 2, $n=3$). Liver fibrosis was binarily classified as clinically non-significant ($<$F2) or significant ($\geq$F2) based on stiffness values for MRE with a published cutoff of 3.5 kPa (Venkatesh and Ehman 2014); and for ultrasound elastography based on median SWV using a cutoff value for Siemens of 1.34 m/s (Friedrich-Rust et al. 2012). At the time the present study was performed, Philips has not yet provided a published reference table for the just recently released ElastQ software to grade fibrosis.

For Siemens data, the accuracy of median SWV using the published cutoff value of 1.34 m/s with respect to MRE-based fibrosis grading was calculated. Essentially, median SWV from US elastography (USE) using a cutoff value of 1.34 m/s for Siemens divides the data set into clinically significant and clinically non-significant fibrosis, whereas for MRE, using a cutoff of 3.5 kPa also divides the data set into clinically significant and clinically non-significant fibrosis, and the accuracy of USE was compared with that of MRE for this determination. However, as Philips does not yet have a published cutoff value for clinically significant fibrosis, for both groups (Siemens and Philips) the performance of median SWV velocity measurements with respect to MRE-based binary fibrosis grades (true labels) was performed using a receiver operating characteristic (ROC) curve analysis. Hence, for both groups (in a technique that thus does not rely on a published cutoff value), median SWV measurements and MRE-based binary fibrosis grades were input to the MATLAB *perfcurve* function to generate

a receiver operating characteristic curve and calculate the area under the curve (AUC).

Machine learning-based grading and statistical analysis

Figure 3 summarizes the ML approach to binary hepatic fibrosis grading using the two groups: pSWE + MRE and 2-DSWE + MRE. Four supervised ML algorithms common in the literature (Erickson et al. 2017) were applied: generalized linear regression model (Dobson 1990), naïve Bayes (Hastie et al. 2009), quadratic discriminant analysis (Guo et al. 2007) and a non-linear support vector machine (SVM) (Schölkopf and Smola 2002).

Logistic regression, which falls under the category of generalized linear models, is a commonly used statistical technique that can be used to predict a categorical outcome value, most commonly binary, given a set of predictor values. If the positive event is coded as "1" and the negative event is coded as "0," then binary logistic regression provides the log odds of the outcome being positive given the predictor values. It uses the logit or sigmoid function, $f(t) = 1/(1 + e^t)$, which represents the log odds of observing the positive event. In this case, $t = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k$, where the values β represent the model parameters to be optimized and the vector X represents the input data (Dobson 1990; Erickson et al. 2017).

A *naïve Bayes classifier* uses the idea of prior or previous probabilities, derived from previous outcomes, and applies the Bayes theorem, which determines the probability of an event occurring taking advantage of these known prior probabilities. The classifier takes each outcome, such as "1" versus "0" in the binary case, and selects the one with the highest probability (Hastie et al. 2009; Erickson et al. 2017).

Linear discriminant analysis, in practical terms, seeks to discriminate between two groups. It does this by minimizing the distance between data points from the same class, while maximizing the distance between data points from different classes. *Quadratic discriminant analysis* is a variation of the aforementioned technique, in which a "pseudo-quadratic" transformation is applied to the data. While linear discriminant analysis naturally allows a linear decision boundary between classes, quadratic discriminant analysis allows quadratic equations to represent that decision boundary. Hence, while quadratic discriminant analysis allows for greater flexibility in the decision boundary, it requires more parameters to be calculated (Guo et al. 2007; Erickson et al. 2017).

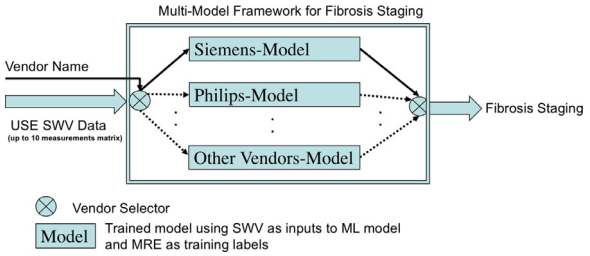


Fig. 3. Proposed multimodel framework for machine learning (ML)-based fibrosis staging. This approach will provide a fibrosis staging between 0 and 100 regardless of vendor. In this work we only tested ultrasound elastography shear wave velocity (USE SWV) measurements obtained using Siemens and Philips scanners, with magnetic resonance elastography (MRE) as ground truth. However, in the future this model could be extended to other vendors after additional training and validation on those data sets.

A SVM seeks to discriminate between classes by mapping each data point into a higher-dimensional space and creating an optimal separating hyperplane that maximizes the distance between each data point and that hyperplane, maximizing the differentiation between each class. This mapping into higher-dimensional space is accomplished by a kernel, and the particular kernel used in this study was the Gaussian radial basis kernel, which performs well with high-dimensional data (Schölkopf and Smola 2002; Erickson et al. 2017). We also used auto scaling with a box constraint of 1.

The 10 measurements of shear wave velocity served as inputs to these ML algorithms, and their accuracy for binary hepatic fibrosis grading was assessed. Twofold cross-validation was performed, that is, half of the data for training and half for testing and vice versa (Hastie et al. 2009). During each run, the group 1 training data set was used to train model 1 with MRE (Siemens-Model; Fig. 3), and the MATLAB *predict* function applied this model to the validation data and output a score representing the likelihood that the label came from each class, either clinically non-significant or significant fibrosis. The MATLAB *perfcurve* function then used these scores and true class labels (from all data) to generate ROC curves to calculate AUC, sensitivity, specificity, positive and negative predictive values and accuracy. Next, the group 2 data set (Philips) was similarly used to train the Philips model with MRE.

To determine if the improvement in AUC between the ML algorithm and median SWV was statistically significant, we performed the DeLong test.

All statistical analyses were performed in MATLAB R2015 b (MathWorks, Natick, MA).

RESULTS

Using the current clinically established standard of care (SOC) cutoff value for binary fibrosis grading for Siemens pSWE (group 1), median SWV measurements performed only fair compared with MRE with 60.4% accuracy (Table 2). Note that SOC versus MRE analysis was not performed in group 2 because of the lack of a

Table 2. Performance of median shear wave velocity of 10 measurements in predicting clinically non-significant versus significant fibrosis using a SOC cutoff value of 1.34 m/s for the pSWE data set (Friedrich-Rust et al. 2012) compared with the reference standard MRE, as well as ML-based staging (MRE equivalent) for group 1

pSWE	SOC versus MRE	SOC versus ML
Sensitivity	81.6	82.5
Specificity	49.5	47.1
Negative predictive value	83.9	87.5
Positive predictive value	45.5	37.5
Accuracy	60.4	56.9

ML = machine learning; MRE = magnetic resonance elastography; pSWE = point shear wave elastography; SOC = standard of care cutoff value.

published cutoff table for Philips 2-DSWE. Next, using the median of 10 consecutive SWV measurements in an analysis employing an ROC curve for both groups, the performance of binary fibrosis grading was moderate for the pSWE (AUC 0.76) and 2-DSWE (AUC 0.84) data sets (Tables 3 and 4, Fig. 4).

Next, performance was assessed using the four ML algorithms, with shear wave velocity measurements as inputs and binary fibrosis grading as determined by MRE as the gold standard (Tables 3 and 4, Fig. 4): the SVM had the highest level of performance of the ML algorithms in binary fibrosis grading, with an AUC of 0.96 for the pSWE data set and 0.99 for the 2-DSWE data set. For the 2-DSWE data sets, quadratic discriminant analysis yielded an AUC of 0.88. The other ML-based algorithms either reached the same or slightly higher AUC values than median SWV in both data sets.

Most notably, the difference in AUC between median shear wave velocity and SVM was statistically significant for both Siemens and Philips, although the *p* value was better for Siemens as it had a larger sample size (Table 5).

In the analysis of score separation between non-significant and significant hepatic fibrosis, median SWV exhibited worse score separation between the two classes (Fig. 5). With the ML-based algorithms, especially support vector machines, there was improved binary score separation for both data sets (Fig. 5).

DISCUSSION

In our study, the ML algorithm SVM outperformed median SWV in distinguishing between non-significant and significant hepatic fibrosis, with a diagnostic performance similar to that of MRE-based fibrosis grading. The ML-based algorithm SVM had excellent diagnostic performance in data sets acquired from Siemens and Philips systems, despite the fact that these two vendors used different elastography techniques.

Table 3. Performance of each machine learning algorithm as well as median shear wave velocity in predicting clinically non-significant versus significant fibrosis in the group 1 data set (pSWE)

Classifier	Sensitivity	Specificity	NPV	PPV	Accuracy	AUC	*p* Value
Median SWV	71.4	71.6	82.9	56.5	71.5	0.760	3.36*E*-07
GLRM	77.1	70.5	85.9	56.9	72.7	0.808	1.87*E*-09
Bayesian	71.4	76.8	83.9	61.4	75.0	0.776	5.88*E*-08
QDA	77.1	70.5	85.9	56.9	72.7	0.821	4.16*E*-10
SVM	81.3	94.7	90.9	88.6	90.2	0.962	1.93*E*-19

Sensitivity and specificity represent different points on the receiver operating characteristic (ROC) curve. *p* Values were calculated using a Wilcoxon rank-sum test.

GLRM = generalized linear regression model; NPV = negative predictive value, PPV = positive predictive value, AUC = area under the curve; QDA = quadratic discriminant analysis; SVM = support vector machine; SWV = shear wave velocity.

Table 4. Performance of each machine learning algorithm as well as median shear wave velocity (without cut-off value) in predicting clinically non-significant versus significant fibrosis in the group 2 data set (2-DSWE)

Classifier	Sensitivity	Specificity	NPV	PPV	Accuracy	AUC	*p* value
Median SWV	73.7	100.0	89.1	100.0	91.7	0.841	2.54 E-05
GLRM	84.2	75.6	91.2	61.5	78.3	0.858	1.16 E-05
Bayesian	78.9	80.5	89.2	65.2	80.0	0.886	1.60 E-06
QDA	78.9	80.5	89.2	65.2	80.0	0.881	2.55 E-06
SVM	89.5	100.0	95.4	100.0	96.7	0.987	1.61 E-09

Sensitivity and specificity represent different points on the receiver operating characteristic (ROC) curve. *p* Values were calculated using a Wilcoxon rank-sum test.

AUC = area-under-the-curve; GLRM = generalized linear regression model; NPV = negative predictive value; PPV = positive predictive value; QDA = quadratic discriminant analysis; SVM = support vector machine; SWV = shear wave velocity.

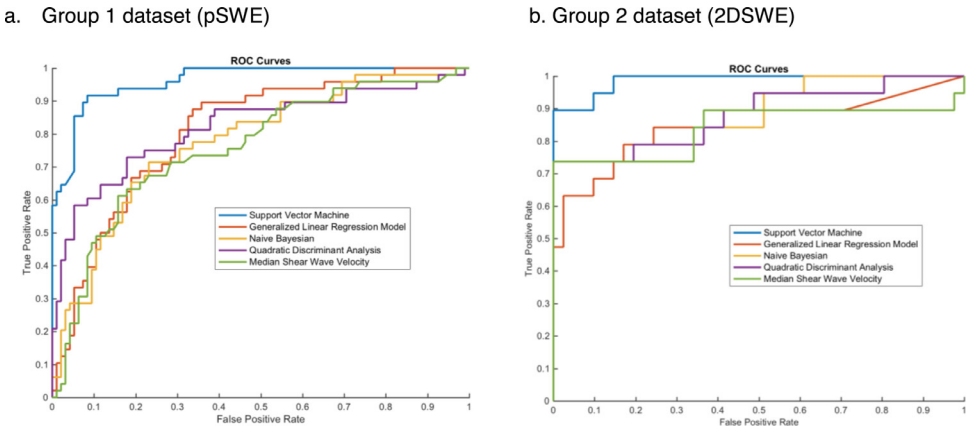


Fig. 4. Receiver operating characteristic (ROC) curves compare the performance of each machine learning (ML) algorithm and the baseline technique using median shear wave velocity to predict clinically non-significant versus significant liver fibrosis, as determined by magnetic resonance elastography (MRE) as gold standard. Support vector machines (*blue*) had the highest performance of all ML algorithms in both groups. pSWE = point shear wave elastography; 2-DSWE = 2-D shear wave elastography.

Table 5. Differences in areas under the curve between the ML algorithm, SVM and median SWV

	P Value	Significantly different
Siemens		
Median SWV versus SVM	4.95*E*-05	Yes
Median SWV versus QDA	0.19098	No
Median SWV versus Bayesian	0.46593	No
Median SWV versus GLRM	0.19098	No
Philips		
Median SWV versus SVM	0.036085	Yes
Median SWV versus QDA	0.32787	No
Median SWV versus Bayesian	0.22957	No
Median SWV versus GLRM	0.71877	No

Difference in areas under the curve between the ML algorithm, SVM and median SWV was statistically significant for both groups.

AUC = area-under-the-curve; GLRM = generalized linear regression model; QDA = quadratic discriminant analysis; SVM = support vector machine; SWV = shear wave velocity.

Significantly different = p-value <0.05.

We used ML in elastography by analyzing 10 shear wave velocity measurements obtained with systems from two different vendors as inputs and then training the algorithm with MRE; our study is the first to assess ML for characterization of liver fibrosis using pSWE and 2-DSWE data from different vendors.

A known issue with ultrasound elastography exams is the variability of the 10 measurements that might be owing to tissue properties (the higher the liver damage, the higher is the variability), operator performance and/or device precision. Currently, median shear wave velocity is used to grade liver fibrosis (Dietrich et al. 2017). One advantage of ML is that it is able to capture information on the data beyond just the median. Future studies need to be conducted to use ML to analyze data from spatial samples (elasticity maps from 2-DSWE) versus temporal samples (10 consecutive measurements, as in our present study). Analyzing spatial data from a single elasticity map would minimize operator dependency, by decreasing the number of maps to be acquired, and reduce scanning time; this would also better account for the heterogeneity of the liver tissue, especially in fibrotic/cirrhotic patients.

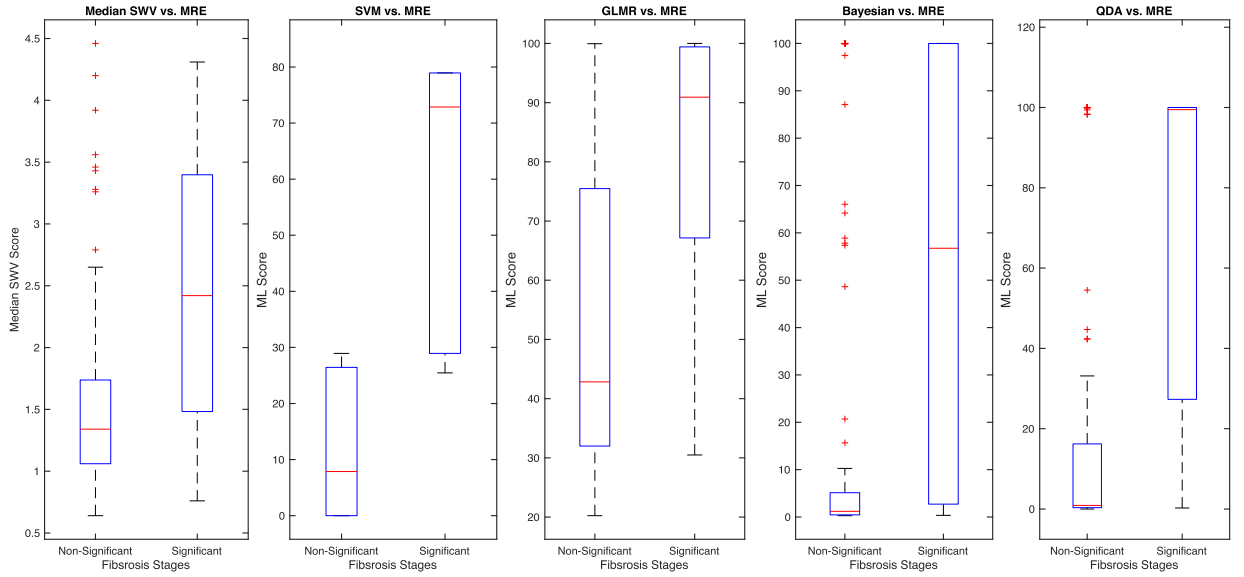
In recent years, ML has been further developed and used increasingly for imaging data analysis, including liver elastography. A prior study performed automatic fibrosis staging in hepatitis C patients using multivariate linear regression that characterized texture features derived from color maps from real-time elastography (Fujimoto et al. 2013). There is also published literature on ML approaches using elastography in other organs, such as the breast in cancer diagnosis (Zhang et al. 2016).

Our study has several limitations. First, the sample size was small and differed between the groups; nonetheless, we confirmed that the difference in AUC between the ML algorithm SVM and median SWV was statistically significant; future studies with more patients are warranted. Second, we trained, tested and validated the ML-based algorithm on systems from only two vendors; systems from other vendors need to be addressed in future studies. Third, our "study gold standard" was MRE; ideally our results will be confirmed in a study with a pathology-trained ML algorithm, although this would be challenging given the small number of patients who undergo liver biopsy at most institutions.

CONCLUSIONS

The new machine learning-based algorithm for grading liver fibrosis into clinically non-significant and significant categories with two different ultrasound elastography techniques from two vendors was found to have excellent diagnostic performance, comparable to that of MR

a. **Group 1 dataset (pSWE)**



b. **Group 2 dataset (2DSWE)**

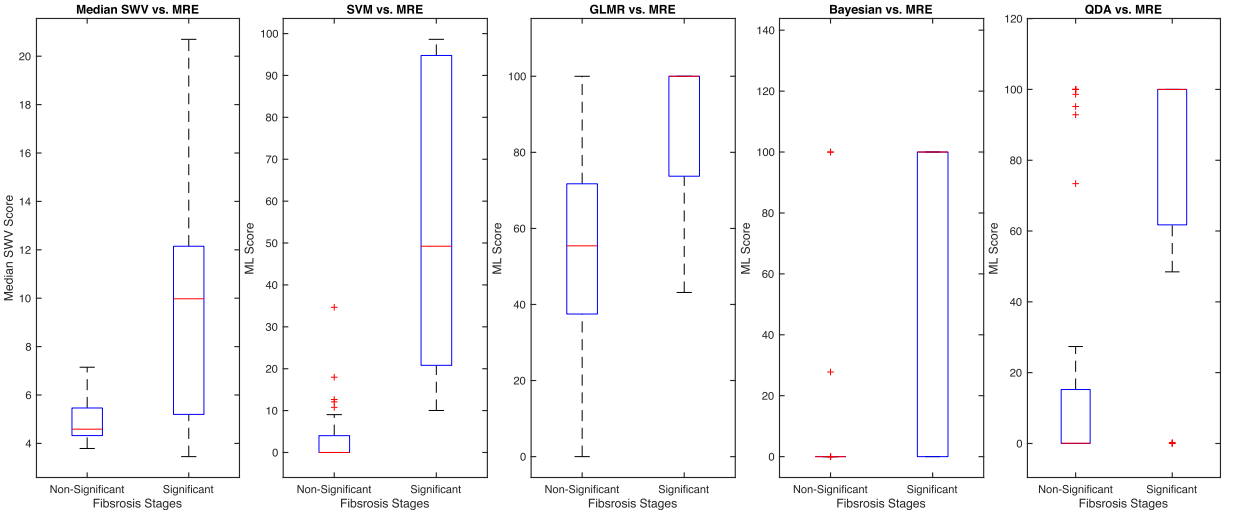


Fig. 5. Scores for non-significant and significant fibrosis separation using median shear wave velocity (SWV) as well as the new machine learning (ML) algorithms in data set 1, (a) pSWE, and data set 2, (b) 2-DSWE. The different scores reflect the likelihood that the label came from each class (non-significant or significant fibrosis). Boxplots reveal excellent score separation in both data sets when a support vector machine (SVM) is used to perform classification, compared with worse score separation with median SWV. Note that ML scores differ between systems from different vendors as well as for the different ML algorithms. MRE = magnetic resonance elastography; GLRM = generalized linear regression model; QDA = quadratic discriminant analysis. The *ends of the box* are the upper and lower quartiles; the *vertical line* inside the box represents the median; and the *whiskers* extend to the highest and lowest values.

elastography. The ML algorithm—support vector machines—outperformed median shear wave velocity. With additional validation in larger studies, this ML-based algorithm, along with a scoring system, might ultimately be included in routine ultrasound screening protocols for the liver for improved liver fibrosis grading, especially in the large patient population with chronic liver disease, without extending the acquisition time. The algorithm, along with a scoring system, could be integrated into the software of clinically established ultrasound elastography

systems from different vendors after being trained and validated for each of these vendors. The scoring system would have the same cutoff for differentiating non-significant from significant fibrosis in systems from all vendors and would provide comparable fibrosis staging, thus abrogating the need for establishing and implementing a different reference table for each vendor.

Acknowledgments—The authors acknowledge the profound impact on this work of the late Dr. Jürgen K. Willmann, who guided this project from start to finish. I.D. was personally supported by the Swiss Society of Radiology, not the project itself. H.S. was awarded an RSNA Research Fellow grant by the Radiological Society of North America and has received funding from the Stanford Cancer Imaging Training (SCIT) Program.

Conflict of interest disclosure—The authors declare no competing interests.

REFERENCES

Afdhal NH, Bacon BR, Patel K, Lawitz EJ, Gordon SC, Nelson DR, Challies TL, Nasser I, Garg J, Wei LJ, McHutchison JG. Accuracy of fibroscan, compared with histology, in analysis of liver fibrosis in patients with hepatitis B or C: A United States multicenter study. Clin Gastroenterol Hepatol 2015;13:772–779.e1−3.

Bota S, Sporea I, Sirli R, Popescu A, Danila M, Costachescu D. Intra- and interoperator reproducibility of acoustic radiation force impulse (ARFI) elastography—Preliminary results. Ultrasound Med Biol 2012;38:1103–1108.

Chen Y, Luo Y, Huang W, Hu D, Zheng RQ, Cong SZ, Meng FK, Yang H, Lin HJ, Sun Y, Wang XY, Wu T, Ren J, Pei SF, Zheng Y, He Y, Hu Y, Yang N, Yan H. Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. Comput Biol Med 2017;89:18–23.

Cui J, Heba E, Hernandez C, Haufe W, Hooker J, Andre MP, Valasek MA, Aryafar H, Sirlin CB, Loomba R. Magnetic resonance elastography is superior to acoustic radiation force impulse for the diagnosis of fibrosis in patients with biopsy-proven nonalcoholic fatty liver disease: A prospective study. Hepatology 2016;63:453–461.

Dietrich CF, Bamber J, Berzigotti A, Bota S, Cantisani V, Castera L, Cosgrove D, Ferraioli G, Friedrich-Rust M, Gilja OH, Goertz RS, Karlas T, de Knegt R, de Ledinghen V, Piscaglia F, Procopet B, Saftoiu A, Sidhu PS, Sporea I, Thiele M. EFSUMB Guidelines and recommendations on the clinical use of liver ultrasound elastography, Update 2017 (Long Version). Ultraschall Med 2017;38:e16–e47.

Dobson AJ. An introduction to generalized linear models. New York: Chapman & Hall; 1990.

D'Onofrio M, Gallotti A, Mucelli RP. 'Tissue quantification with acoustic radiation force impulse imaging: Measurement repeatability and normal values in the healthy liver'. AJR Am J Roentgenol 2010;195:132–136.

Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics 2017;37:505–515.

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 Med Biol 2015;41:1161–1179.

Ferraioli G, De Silvestri A, Lissandrin R, Maiocchi L, Tinelli C, Filice C, Barr RG. Evaluation of inter-system variability in liver stiffness measurements. Ultraschall Med 2019;40:64–75.

Friedrich-Rust M, Nierhoff J, Lupsor M, Sporea I, Fierbinteanu-Braticevici C, Strobel D, Takahashi H, Yoneda M, Suda T, Zeuzem S, Herrmann E. Performance of acoustic radiation force impulse imaging for the staging of liver fibrosis: A pooled meta-analysis. J Viral Hepat 2012;19:e212–e219.

Fujimoto K, Kato M, Kudo M, Yada N, Shiina T, Ueshima K, Yamada Y, Ishida T, Azuma M, Yamasaki M, Yamamoto K, Hayashi N, Takehara T. 'Novel image analysis method using ultrasound elastography for noninvasive evaluation of hepatic fibrosis in patients with chronic hepatitis C. Oncology 2013;84 (Suppl. 1):3–12.

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 Med Biol 2017;43:1797–1810.

Guo Y, Hastie T, Tibshirani R. Regularized linear discriminant analysis and its application in microarrays. Biostatistics 2007;8:86–100.

Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: Data mining, inference, and prediction. New York: Springer; 2009.

Lurie Y, Webb M, Cytter-Kuint R, Shteingart S, Lederkremer GZ. Non-invasive diagnosis of liver fibrosis and cirrhosis. World J Gastroenterol 2015;21:11567–11583.

Pan JJ, Bao F, Du E, Skillin C, Frenette CT, Waalen J, Alaparthi L, Goodman ZD, Pockros PJ. Morphometry confirms fibrosis regression from sustained virologic response to direct-acting antivirals for hepatitis C. Hepatol Commun 2018;2:1320–1330.

Rizzo L, Calvaruso V, Cacopardo B, Alessi N, Attanasio M, Petta S, Fatuzzo F, Montineri A, Mazzola A, L'Abbate L, Nunnari G, Bronte F, Di Marco V, Craxi A, Camma C. Comparison of transient elastography and acoustic radiation force impulse for non-invasive staging of liver fibrosis in patients with chronic hepatitis C. Am J Gastroenterol 2011;106:2112–2120.

Schölkopf B, Smola AJ. Learning with kernels: Support vector machines, regularization, optimization, and beyond. Cambridge, MA: MIT Press; 2002.

Shi Y, Guo Q, Xia F, Dzyubak B, Glaser KJ, Li Q, Li J, Ehman RL. MR elastography for the assessment of hepatic fibrosis in patients with chronic hepatitis B infection: Does histologic necroinflammation influence the measurement of hepatic stiffness?. Radiology 2014;273:88–98.

Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound elastography: Review of techniques and clinical applications. Theranostics 2017;7:1303–1329.

Stoean R, Stoean C, Lupsor M, Stefanescu H, Badea R. Evolutionary-driven support vector machines for determining the degree of liver fibrosis in chronic hepatitis C. Artif Intell Med 2011;51:53–65.

Trout AT, Serai S, Mahley AD, Wang H, Zhang Y, Zhang B, Dillman JR. Liver stiffness measurements with MR elastography: Agreement and repeatability across imaging systems, field strengths, and pulse sequences. Radiology 2016;281:793–804.

Venkatesh SK, Ehman RL. Magnetic resonance elastography of liver. Magn Reson Imaging Clin North Am 2014;22:433–446.

Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, Zheng H. Deep learning based classification of breast tumors with shear-wave elastography. Ultrasonics 2016;72:150–157.

Zhang W, Zhu Y, Zhang C, Ran H. Diagnostic accuracy of 2-dimensional shear wave elastography for the staging of liver fibrosis: A meta-analysis. J Ultrasound Med 2019;38:733–740.