Quantitative Approaches to Elastography Data

Hersh Sagreiya

PI: Daniel L. Rubin
School of Medicine, Stanford University

Background

- Liver cirrhosis a dire worldwide public health concern
  - > 1 million annual deaths.
- Common causes:
  - Hepatitis B/C
  - Nonalcoholic fatty liver disease
  - Alcoholic liver disease
  - Autoimmune liver disease.
- Varying degrees of fibrosis (liver stiffness) can progress to cirrhosis.
- If hepatic fibrosis is detected early and its cause identified, it can be stabilized/reversed using anti-viral, anti-fibrotic, or anti-inflammatory drugs.
- Quantitative assessment of hepatic fibrosis could be used to guide treatment choice/assess treatment response.

Methods of Evaluation

- Liver biopsy was the traditional method for assessing hepatic fibrosis, using the METAVIR histopathologic grading system.
- However, liver biopsy is invasive, with risks including bleeding and infection.
- Biopsy can also be prone to under-sampling or inter-observer differences in interpretation using METAVIR among pathologists.
- While magnetic resonance elastography (MRE) is also considered a gold standard examination for assessing liver fibrosis, it is expensive and not an option for many at risk for cirrhosis worldwide.

Ultrasound Elastography

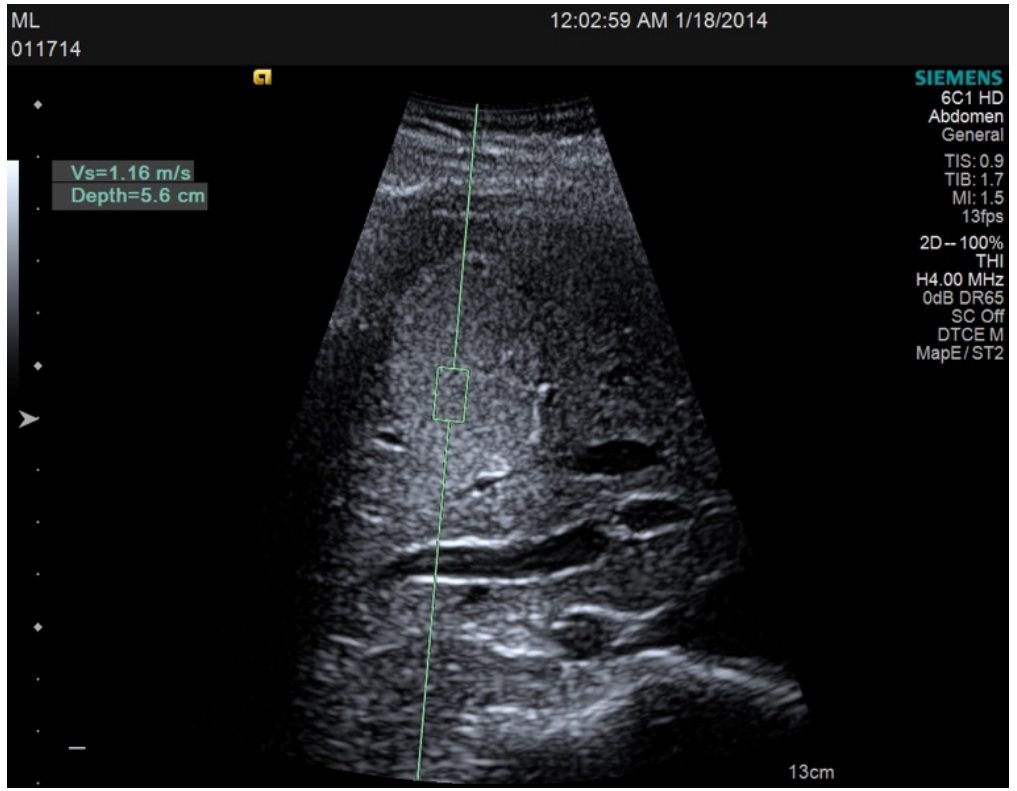
- Ultrasound elastography non-invasively measures tissue stiffness.
- There can be differences in the elasticity of soft tissues due to pathological or physiological processes.
- Fibrosis in the setting of chronic liver diseases reduces the elasticity of the liver relative to normal tissues due to collagen deposition and microstructural changes.
- Recent studies have shown the ability of ultrasound elastography to differentiate malignant from benign focal liver lesions with 97% sensitivity and 66% specificity .
- Elastography can help in differentiating diseased from normal tissue and assist in diagnosis by providing additional information to conventional ultrasound.

Ultrasound Elastography

- There are some limitations of ultrasound elastography:
  - Normal physiologic processes and disease states (passive hepatic congestion in cardiac insufficiency, cholestasis, and hepatic steatosis) can confound elastography measurements.
  - Due to current limitations in distinguishing between individual fibrosis stages, the World Federation for Ultrasound in Medicine and Biology guidelines recommend ultrasound elastography to be used to distinguish significant or advanced fibrosis from non-significant fibrosis.
  - The Society of Radiologists in Ultrasound recommend using elastography to discriminate no or minimal fibrosis from severe fibrosis or cirrhosis.

Current Dataset

- 3,637 patients who underwent point shear wave elastography with a Siemens Acuson S2000 ultrasound scanner.
- Involves getting shear wave velocity measurements from 10 regions of interest (ROIs), represented by boxes.
- Boxes are placed on the original grayscale image, and shear wave velocity is calculated 10 times.



Grading the Degree of Fibrosis

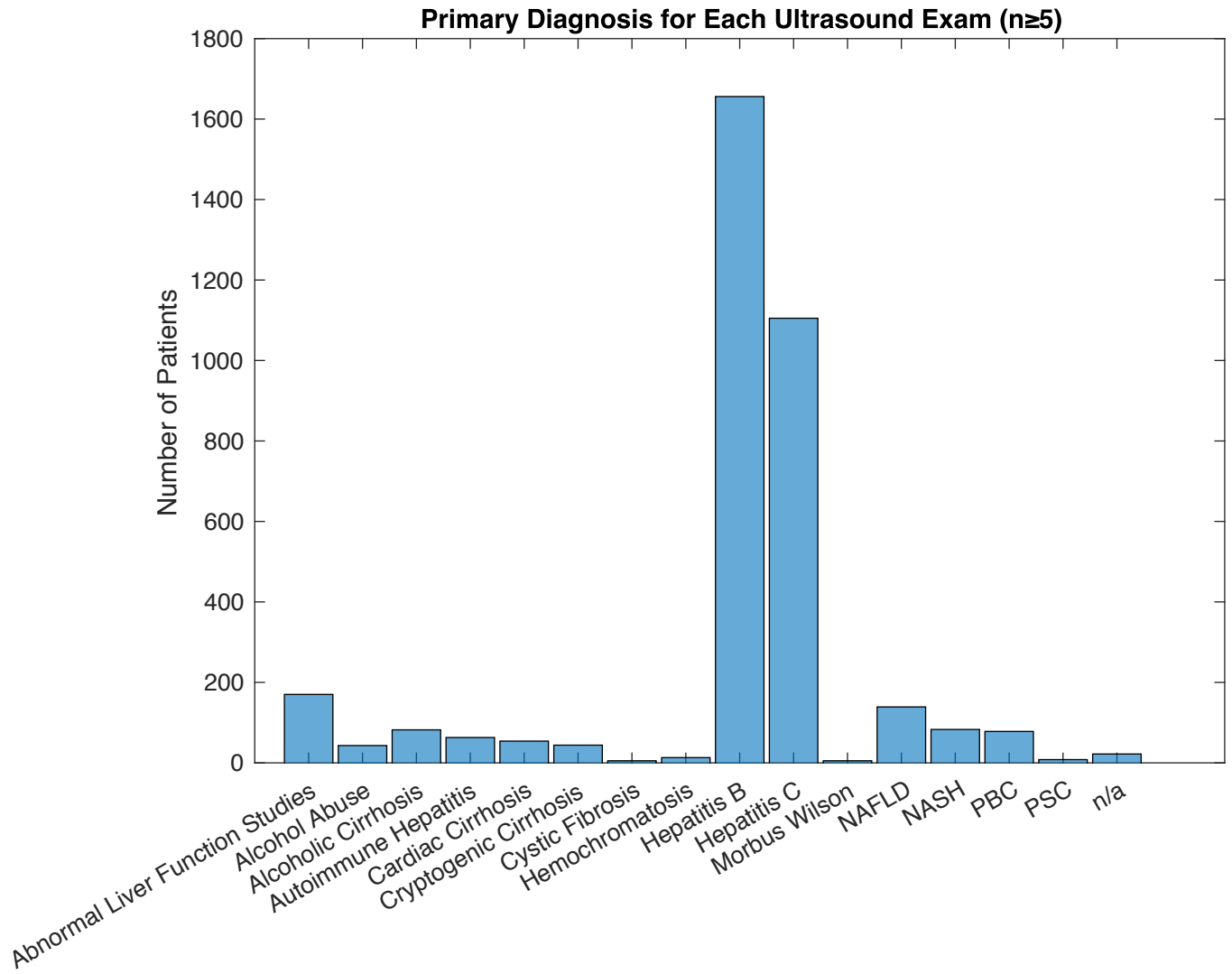
- The median shear wave velocity is used to grade the degree of fibrosis into four bins: F0/F1, F2, F3, F4.
- F0/F1 vs. F2-F4 represents the cut-off for “clinically significant fibrosis.”

Reference values for liver fibrosis:
F0/F1: For velocities ≤ 1.34 m/s
F2: For velocities > 1.34 m/s
F3: For velocities > 1.55 m/s
F4: For velocities > 1.8 m/s
(Friedrich-Rust, 2012)

Patient Population

Variable	Distribution*
Sex	1923 (M), 1714 (F)
Age	55.2 ± 14.2 (μ ± σ)
BMI	25.9 ± 5.5 (μ ± σ)
Cirrhotic	751 (Y), 2827 (N), 59 (n/a)
Ascites	135 (Y), 3443 (N), 59 (n/a)
Steatosis	828 (Y), 2750 (N), 59 (n/a)

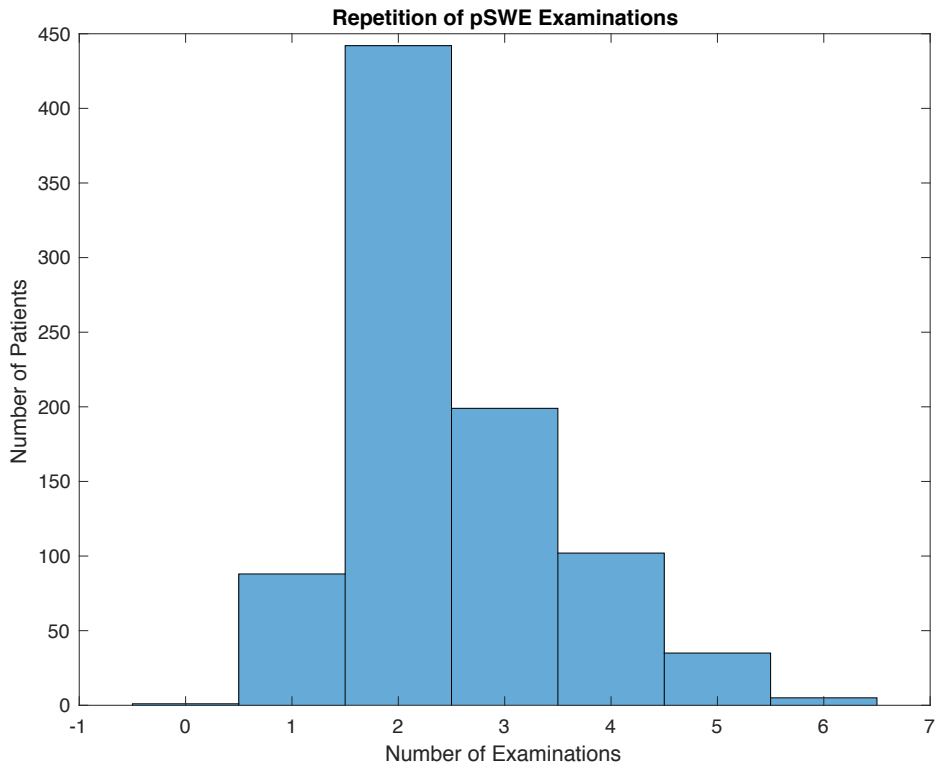
Clinical Indication



Primary Indication	n
'Abnormal Liver Function Studies'	170
'Alcohol Abuse'	43
'Alcoholic Cirrhosis'	82
'Autoimmune Hepatitis'	63
'Cardiac Cirrhosis'	54
'Cryptogenic Cirrhosis'	44
'Cystic Fibrosis'	5
'Hemochromatosis'	13
'Hepatitis B'	1656
'Hepatitis C'	1105
'Morbus Wilson'	5
'NAFLD'	139
'NASH'	83
'PBC'	78
'PSC'	8
'n/a'	22

Longitudinal Information

- ~800 patients have longitudinal studies, with 10 measurements of pSWE for each study.



Additional Data Collected from EMR

- Used Stanford Research Informatics to get additional information.
- This includes key lab values (with dates obtained):
  - AST
    ALT
  - Alkaline Phosphatase
  - WBC
- Also, extracted treatment with antiviral drugs (also with dates):
  - These are the treatment for hepatitis B & C.

Project 1:
Elastography Pathology Prediction

- **Study purpose:** To determine if shear wave velocity measurements from point shear wave elastography can predict biopsy-determined fibrosis grade using machine learning techniques.

Study Design

- Pathology is uncommonly performed because of its invasiveness.
- Used Stanford Research Informatics to find all the liver pathology reports of patients who underwent shear wave elastography in this dataset.
- A total of 73 examinations for 61 patients had liver biopsy within six months of the ultrasound elastography examination.
- Clinically significant fibrosis, as determined by pathology, was the ground truth for this study, essentially a binary outcome.
  - This bins together clinically nonsignificant fibrosis (F0/F1) and clinically significant fibrosis (F2-F4), similar to what is done clinically.

Machine Learning

- Using the ten measurements of shear wave velocity as inputs, four machine learning algorithms—AdaBoost, support vector machine (using the Gaussian radial basis function), a multilayer perceptron (10 layers), and k-nearest neighbors (k=3)–were evaluated in their ability to distinguish clinically significant and non-significant fibrosis, as determined by biopsy.
- Validated using two-fold cross-validation.

Statistical Analysis

- The Matlab *predict* function was applied to validation data to obtain scores representing the probability that the predicted class was correct.
- The Matlab *perfcurve* function used the scores and pathology outcomes to calculate sensitivity, specificity, positive and negative predictive value, accuracy, and ROC area-under-the-curve.
- Scores between the two classes were compared with a Wilcoxon rank-sum test.

Results

Algorithm	Sensitivity	Specificity	NPV	PPV	Accuracy	AUC	p-value
AdaBoost	82.4	91.3	70	95.5	85.1	0.93	7.0×10^{-10}
Support Vector Machine	84.3	91.3	72.4	95.6	86.5	0.93	2.2×10^{-10}
Multi-Layer Perceptron	62.7	90.9	51.3	94.1	71.2	0.81	1.5×10^{-6}
K Nearest Neighbors	54.9	95.7	48.9	96.5	67.6	0.78	4.3×10^{-5}

Performance of the different machine learning algorithms, using ten measurements of shear wave velocity from hepatic point shear wave elastography to classify between clinically-significant and non-significant fibrosis.

Conclusions

- Shear wave velocity measurements from hepatic point shear wave elastography used as inputs to machine learning correlate well with biopsy-determined fibrosis grade. AdaBoost and SVM performed best.
- Machine learning can successfully be used to predict biopsy-determined fibrosis grade from point shear wave elastography measurements.

Project 2: Fat Quantification

- **Study purpose:** To determine if there is a relationship between shear wave velocity measurements using ultrasound point shear wave elastography and fat quantification derived from MRI using machine learning (ML).

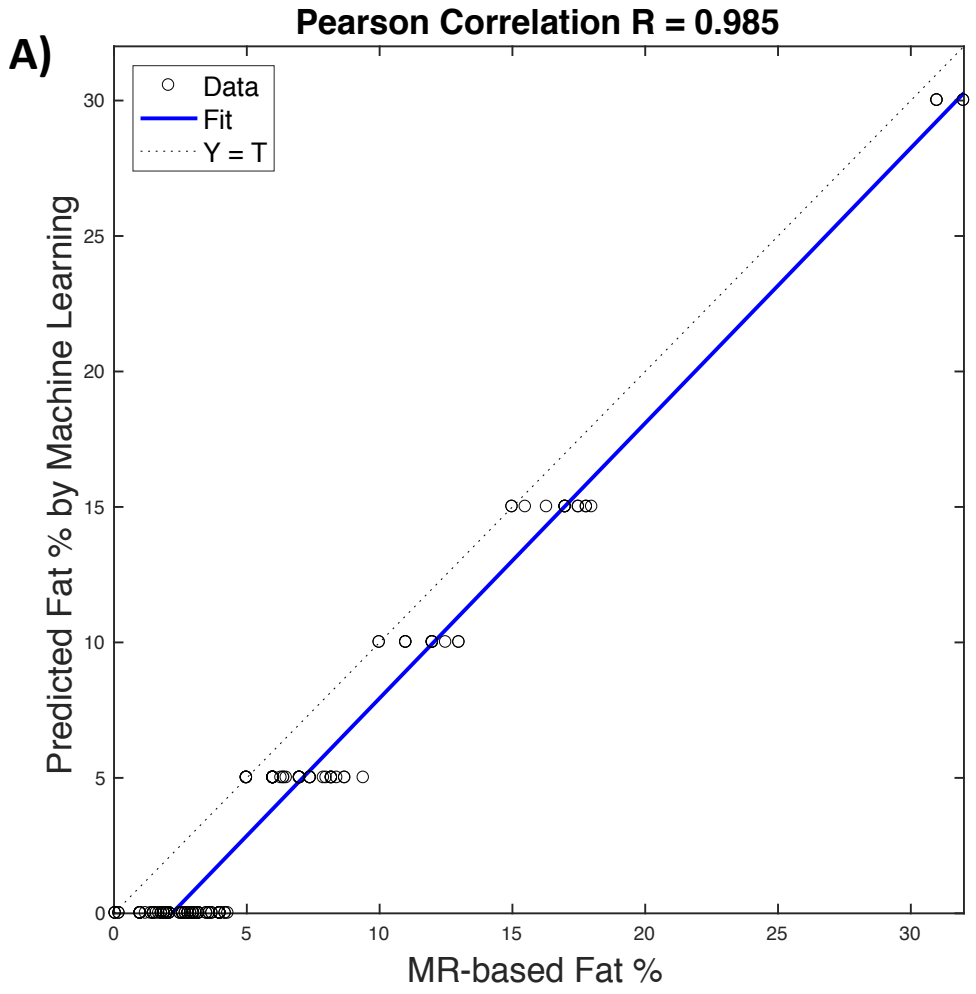
Fat Quantification

- 186 examinations from 113 patients had results from both point shear wave elastography from a Siemens ultrasound scanner and fat quantification from MRI.
- Fat quantification values were quantized into intervals of 5%, and a multi-model support vector machine (SVM) algorithm with the Gaussian radial basis function kernel was run with ten measurements of shear wave velocity as inputs.
- For each fat quantification interval, a dedicated SVM model was trained, and machine-learning based fat prediction was determined by fusing the results from all models.
- Intervals of 5% were chosen since the standard steatosis threshold clinically is 5%.

Fat Quantification

- Results were validated using leave-one-out cross-validation.
- Next, for each quantization interval, the p-value for the fat prediction, as determined by the SVM, was calculated using a Wilcoxon rank-sum test.
- Finally, the correlation between predicted and actual fat quantification was done via Pearson correlation.

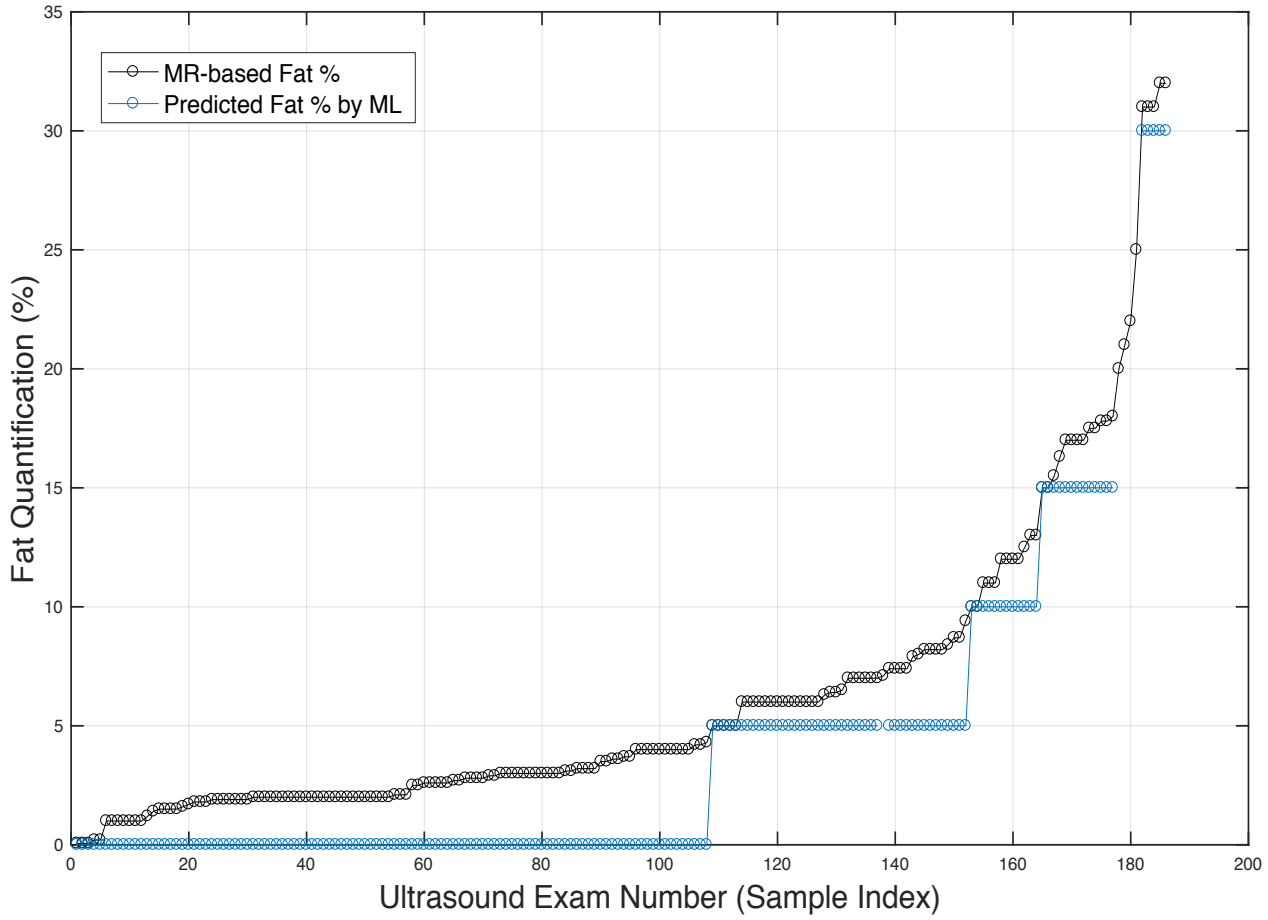
Results



Quantification Interval	p-value	Number of samples
0-5	2.05E-38	108
5-10	2.10E-10	44
10-15	2.34E-07	12
15-20	1.43E-10	13
20-25	Insufficient Data	3
25-30	Insufficient Data	1
30+	1.52E-05	5

A) Pearson correlation between MR-determined fat quantification and predicted fat quantification using machine learning. **B)** Strength of predictions at each fat quantification interval as determined by the Wilcoxon rank-sum test.

Results



C) MR-determined fat percent (black) and predicted fat percent by machine learning (blue).

Results Summary

- There was high correlation between predicted fat quantification interval from ML and MRI-based fat quantification interval ($r = 0.98$).
- For this dataset, the dynamic range for MRI-based fat percent was between 0% and 35%; however, five SVM models had enough data samples to be trained.
- There was good score separation in intervals under 20% ($p < 10^{-9}$), whereas intervals greater than 20% had insufficient samples in our dataset.
- Finally, for each ultrasound examination (sample index), a predicted fat quantification interval could be determined.

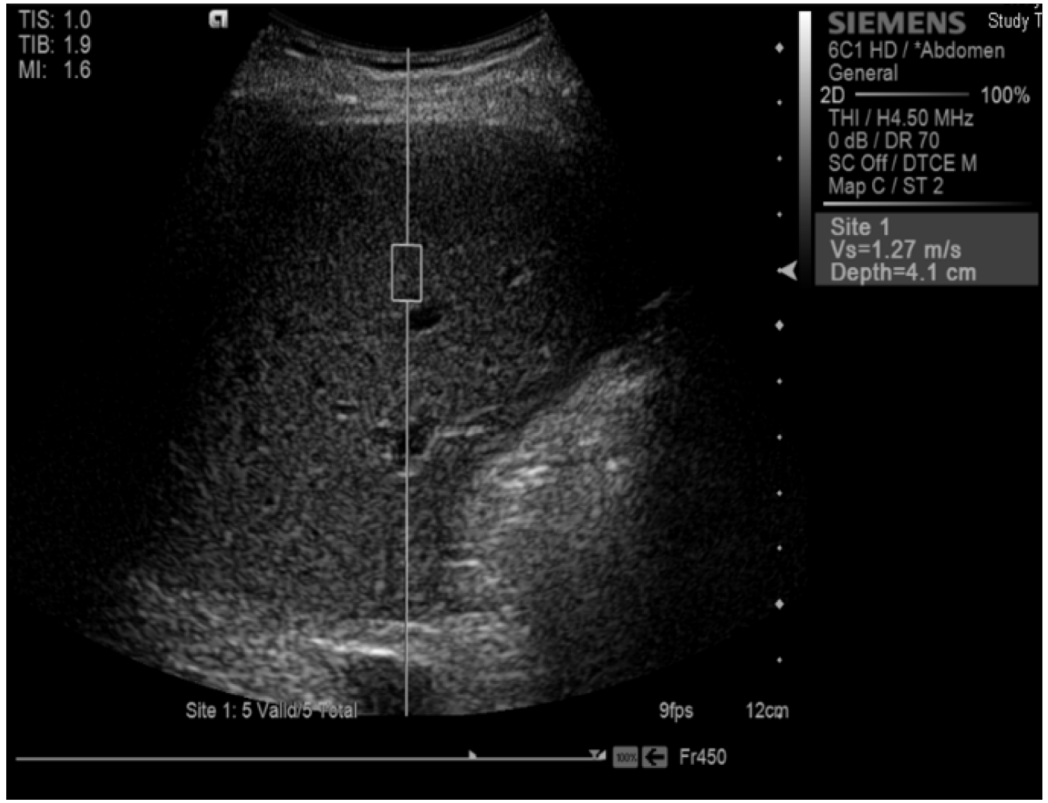
Conclusion

- ML with data from ultrasound point shear elastography correlates well with MRI-determined fat quantification levels.
- Additional investigation and training with a bigger dataset are necessary to further validate the robustness of this approach. In addition, a future study could add more data to the model (hepatic echogenicity, RF data).

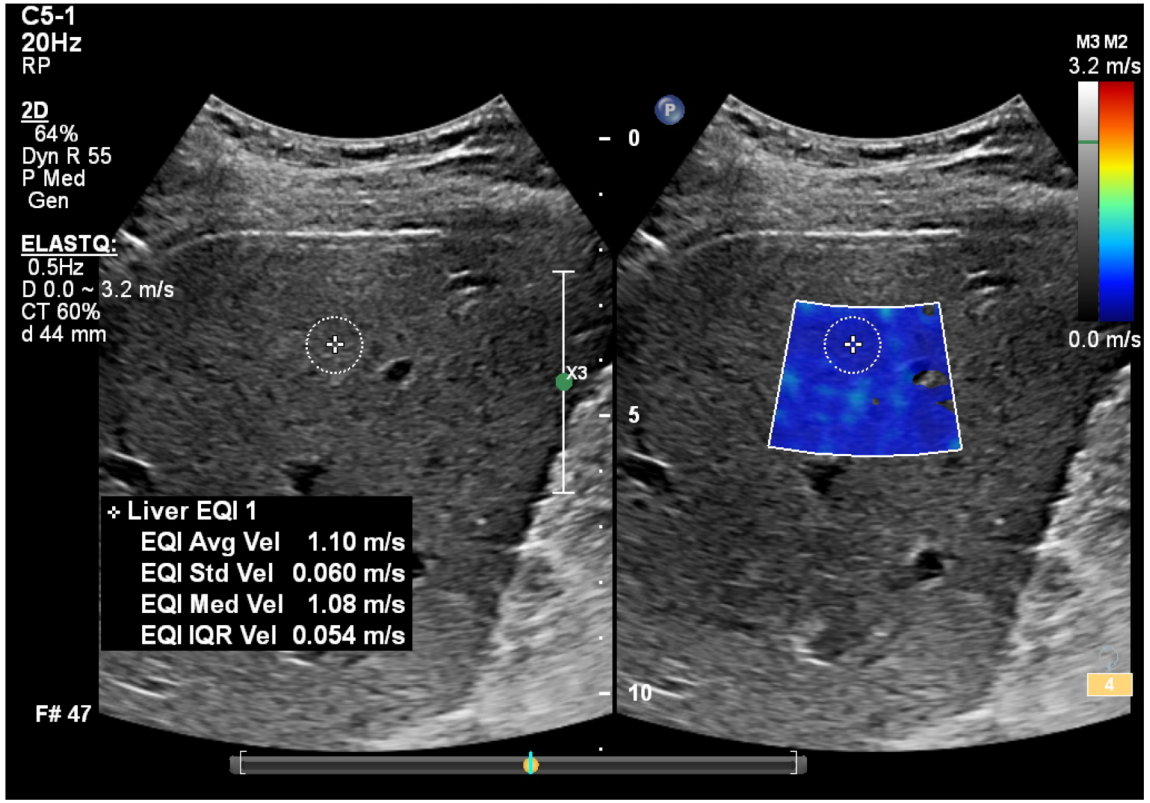
Project 3a: Standardizing Data from Different Vendors

- Different vendors use different technologies to grade hepatic fibrosis. Major clinical problem, because validated cut-off values for different fibrosis grades are only established for some technologies.
- Is it possible to standardize hepatic fibrosis grading via technologies from different vendors using MR elastography as the gold standard?
- Study Population:
  - 123 patients undergoing point shear wave elastography (pSWE) using a Siemens S2000 scanner and MRE.
  - 60 patients undergoing 2D shear wave elastography (Philips Epiq7) and MRE.
- Comparisons: Median shear wave velocity (current standard) and 4 ML algorithms: support vector machine, logistic regression, naïve Bayes, quadratic discriminant analysis.
- Validation: 2-fold CV.

Different Technologies

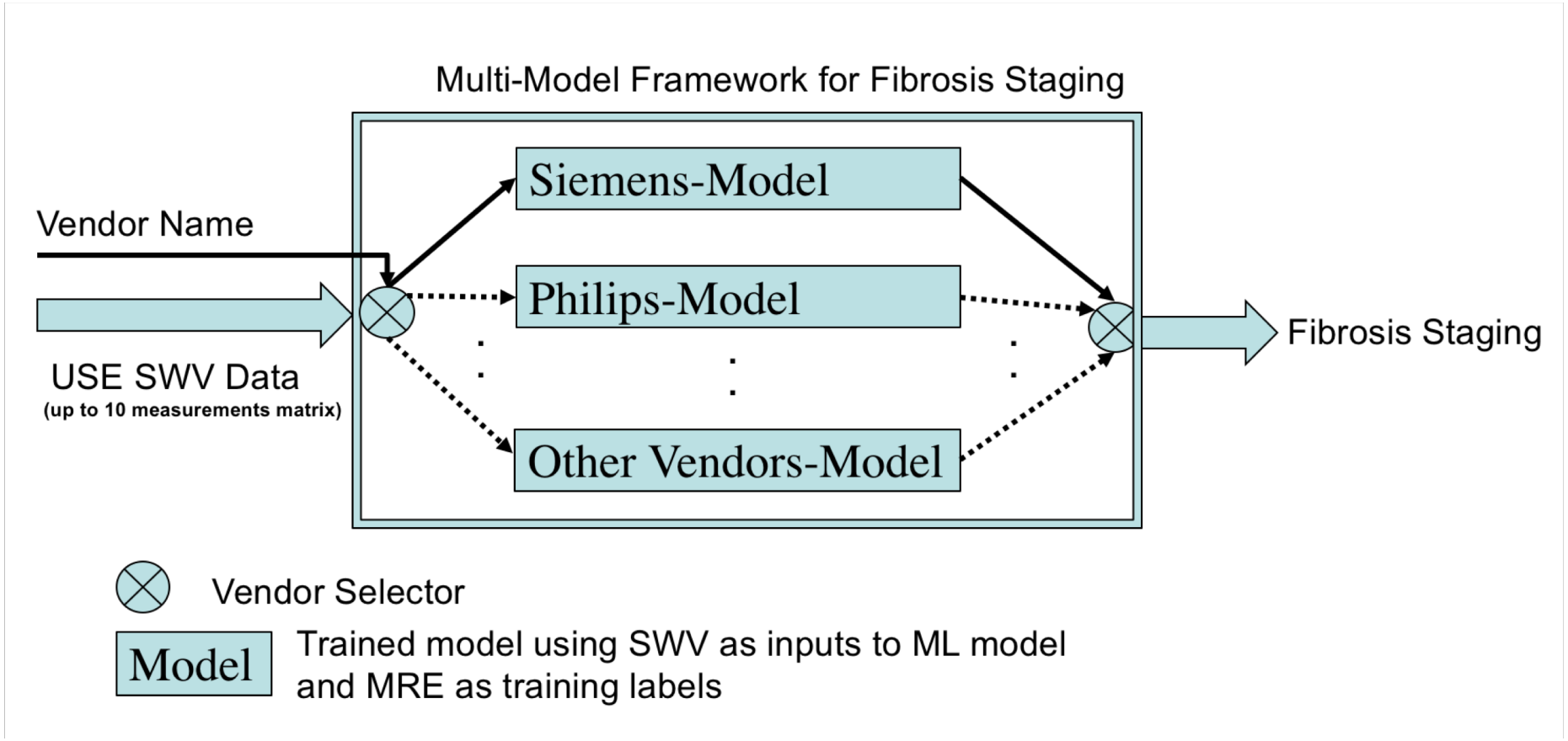


Point shear wave elastography (Siemens)

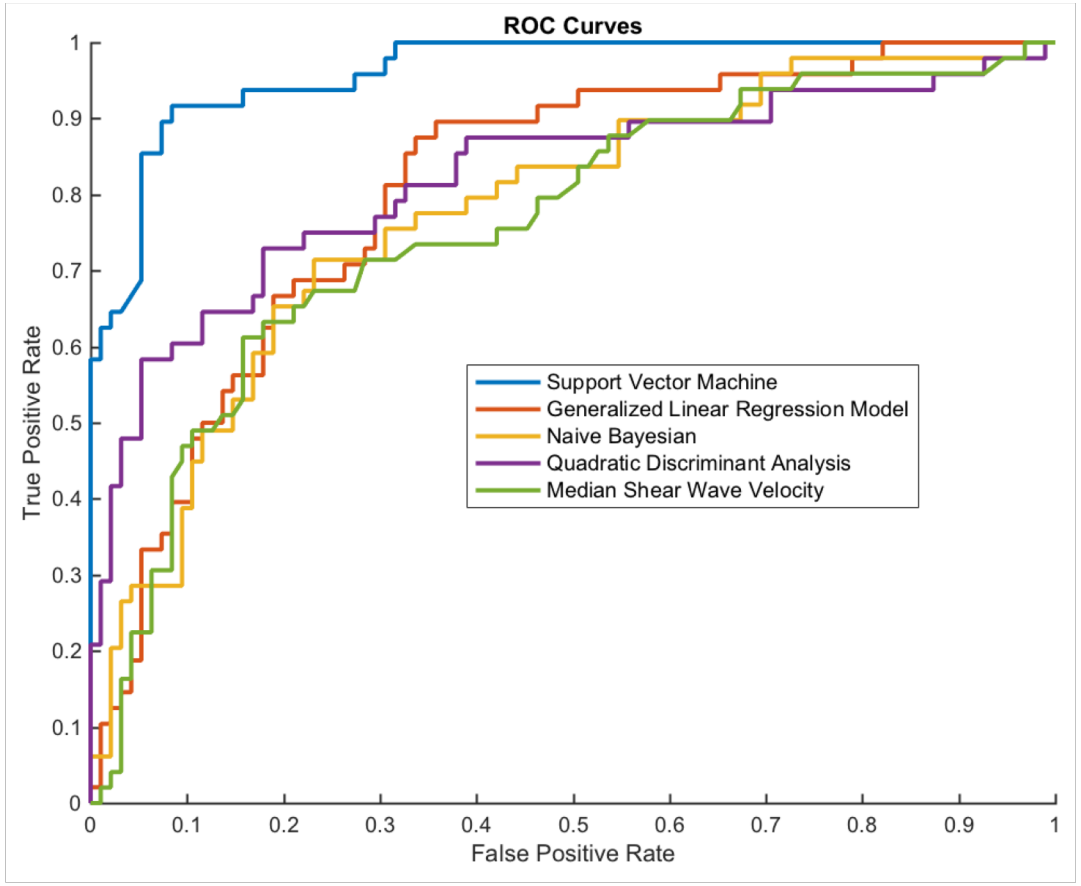


2D shear wave elastography (Philips)

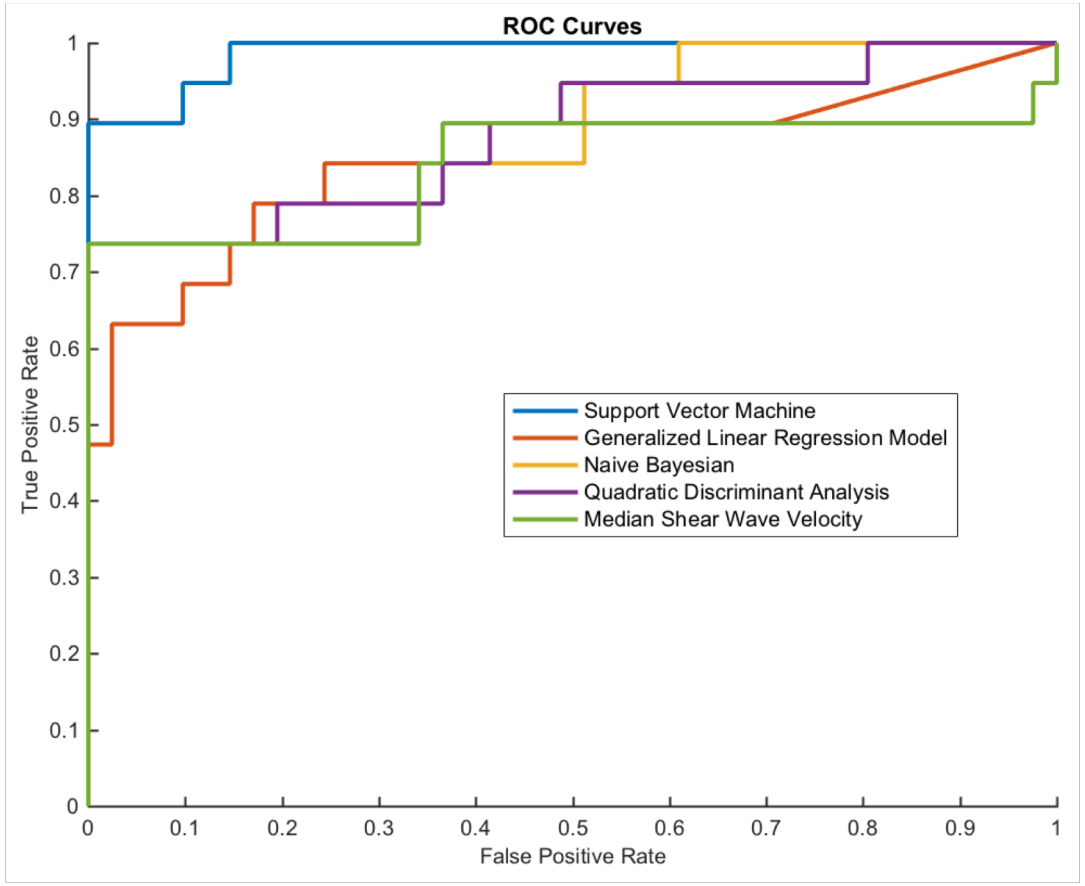
Overall Framework



Results



Point shear wave elastography (Siemens)
Median Shear Wave AUC = 0.76
SVM AUC = 0.96



2D shear wave elastography (Philips)
Median Shear Wave AUC = 0.84
SVM AUC = 0.99

Project 3b: Standardizing grading against another validated technology

- Transient elastography (TE) is the oldest and most highly validated ultrasound elastography technique.
- Widely performed in Europe.
- TE: Dynamic stress is generated by a mechanical device. No grayscale image is created.
- Comparatively newer technique: point shear wave elastography (pSWE).
- pSWE: An acoustic radiation force impulse is used to generate transverse shear waves at a single focal zone. The speed of these shear waves is related to liver stiffness.
- Advantages:
  - Uses grayscale images to select uniform hepatic tissue (avoids vessels/bile ducts).
  - Less sensitive to ascites and obesity.
- TE is an acceptable reference standard:
  - Accepted as a reliable substitute for liver biopsy, which itself demonstrates intra- and inter-observer variability and is invasive.
- **Purpose:** To use machine learning to improve the measurements from pSWE using TE as the gold standard.

Methods

- 308 patients with chronic hepatitis C in Italy.
- Imaged using pSWE (acoustic radiation force impulse quantification) with a Philips ElastPQ ultrasound scanner and TE with Fibroscan on same day.
- pSWE: 10 measurements of shear wave velocity were obtained.
- TE: Median value of 10 acquisitions was obtained and used as the reference standard with the cutoff of 7 kPa for clinically significant fibrosis (stage F2).
- Similar ML algorithms and validation as in 3a.
- The significance of the difference in AUC between each technique was evaluated using the DeLong method.

Demographics

Category	Value
Gender	182 male, 126 female
Age (mean, SD)	55.7 ± 14.7 years
BMI (mean, SD)	23.6 ± 3.8 kg/m^2
AST (median, IQR)	36 (23-58) U/L
ALT (median, IQR)	40 (22-70) U/L
GGT (median, IQR)	40 (24-81) U/L
ALP (median, IQR)	72 (61-89) U/L

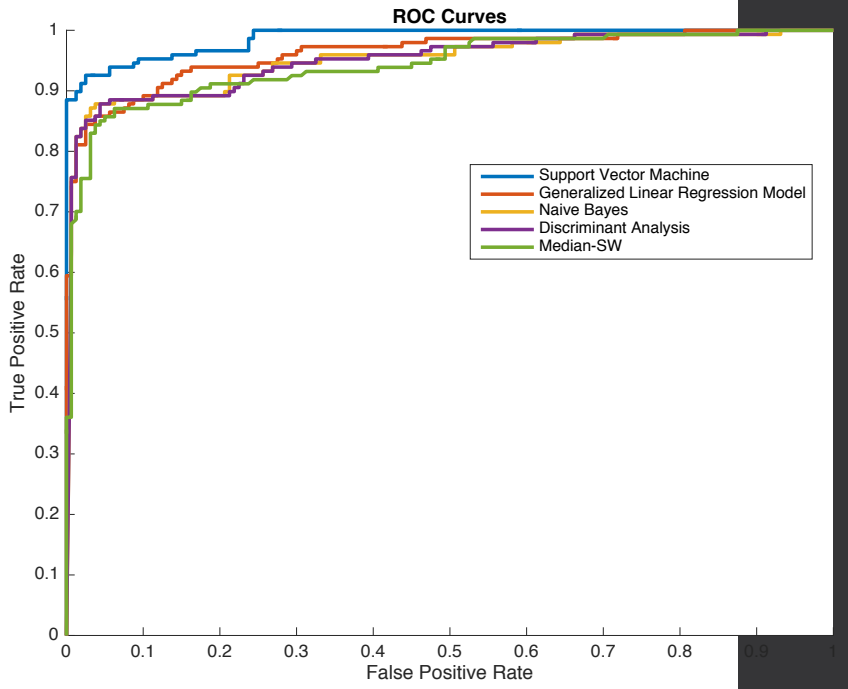
Demographics

Fibrosis Grade	Number
F0-F1	160
F2	45
F3	16
F4	87

Fibrosis Grade Determined by Transient Elastography

Results

Algorithm Type	SENS	SPEC	NPV	PPV	ACC	ROC AUC	P-VAL
Linear Regression	80.4	98.8	84.5	98.4	89.9	0.962	1.49 E-44
Support Vector Machine	80.4	100	84.7	100	90.6	0.987	6.92 E-50
Naïve Bayes	80.4	98.8	84.5	98.4	89.9	0.951	9.39 E-44
Quadratic Discriminant Analysis	80.4	98.8	84.5	98.4	89.9	0.951	9.30 E-44
Median Shear Wave Velocity	80.3	96.9	84.2	95.9	88.9	0.943	4.77 E-41



Performance of four algorithms using ten measurements of shear wave velocity from point shear wave elastography as inputs and a **cutoff of 7 kPa for significant fibrosis** via transient elastography serving as the reference standard. Performance was compared with that of simply taking the median of the ten velocity measurements from point shear wave elastography as the input.

Model Comparison

Combination	P-value	Significantly Different
SVM vs. Median SWV	*6.86 E-4*	*Yes*
SVM vs. QDA	*4.29 E-4*	*Yes*
SVM vs. Logistic Regression	*6.24 E-3*	*Yes*
SVM vs. Naïve Bayes	*2.14 E-3*	*Yes*
Logistic Regression vs. Median SWV	*9.69 E-3*	*Yes*
Logistic Regression vs. Naïve Bayes	0.0808	No
Logistic Regression vs. QDA	0.304	No
QDA vs. Median SWV	0.323	No
Naïve Bayes vs. Median SWV	0.252	No
Naïve Bayes vs. QDA	0.867	No

Comparison of AUC values for each combination of techniques using the DeLong method.

Discussion

- Shear wave velocity measurements using point shear wave elastography, a newer technology with key advantages, are consistent with the determination of fibrosis using the established TE method.
- Machine learning adds value by improving the sensitivity and specificity for fibrosis staging.
- Validation in a larger dataset is warranted.

Project 4: Deep Learning (in progress)

- Despite the ubiquity of clinical ultrasound, not all machines have shear wave elastography available, and doing the multiple measurements required for shear wave elasatography can be time-consuming.
- Traditional B-mode ultrasound is inexpensive, portable, real-time, and nearly ubiquitous worldwide.
- While radiologists can sometimes diagnose hepatic fibrosis by subjectively observing “coarsening of the hepatic echotexture,” this is not quantitative.

Deep Learning: Background

- While common at Stanford, most centers do not routinely perform ultrasound elastography, as it requires specialized ultrasound scanners and trained ultrasound technologists and radiologists.
- Since hepatic fibrosis is a heterogenous disease, ultrasound elastography may be subject to sampling error.
- It would be ideal if it were possible to use the original B-mode images, obviating the need to select small regions of interest and overcoming the limitations of ultrasound elastography.
- With the recent revolution in using deep learning to classify medical images, can try using deep learning techniques on the original grayscale images from ultrasound elastography to accurately stage liver fibrosis.

Deep Learning: Background

- The model will initially be trained and validated using clinically significant fibrosis as determined by ultrasound elastography.
- It will further be validated on a patient cohort that has a gold standard exam—MRE.
- Prior work used a much smaller dataset (279 vs 3,637) and did not have the clinical gold standard MRE for validation.
- This work could be highly impactful clinically, as it will enable rapid assessment of the liver using conventional B-mode ultrasound to automatically grade and longitudinally assess hepatic fibrosis, with accuracy comparable to MRE:
  - This could drastically reducing screening costs and providing a strategy for deployment worldwide.

Preliminary Results:

- Used Stanford Research IT to get the original DICOM images for 3600+ ultrasound elastography exams performed using a Siemens Acuson S2000 scanner.
- Ultrasound images were de-identified using the Stanford CTP server.
- **Initial Problem:** For each exam, the complete abdomen was imaged, including many images (1000+) that were not relevant. Unfortunately, analysis of the DICOM metadata did not help.
- **Solution:** Used the *tesserocr* (v2.2.2) and *pydicom* (v.1.0.2) Python packages to develop code to correctly identify images of the liver using *optical character recognition*. This is able to identify both the relevant image of the liver, as well as the corresponding shear wave velocity measurement.

Experimental Design:

- Use a machine with the NVIDIA Titan Xp graphics processing unit, the CUDA parallel computing platform (v9.0), the CuDNN deep neural network library (v7.0.4), PyTorch (v0.40).
- Train as a binary classifier to determine clinically-significant fibrosis. 70% of the data will be set aside for training, and ten-fold cross-validation will be performed
- Select appropriate model architecture (e.g. DenseNet, ResNet-50) and fine-tune parameters affecting model fit, such as regularization and data augmentation.
- Clinical variables (gender, BMI, medical diagnosis, and race) will be added to the model. Model success will be determined on the held-out test set with 30% of the data.
- We will also explore other model architectures and machine learning with pre-defined features.

Deep Learning Collaboration

- Have started a collaboration with China Medical University, which has 2000+ ultrasound images, 1000+ CT images, clinical variables, outcomes, and digital pathology slides. Currently, working on processing the IRB on our end.
- **Additional Goals:**
  - Determine how well our model generalizes to the external dataset (and vice-versa).
  - Add clinical data to the model to see if it improves performance.
  - See how well CT images can predict fibrosis.
  - Investigate whether imaging data can be correlated with digital pathology slides (rad-path correlation).
  - Evaluate whether ultrasound imaging combined with other data can predict clinical endpoints (development of HCC, survival).
  - Write a grant with these questions in mind.

Acknowledgements

- Dr. Daniel Rubin
- Dr. Isabelle Durot
- Dr. Alireza Akhbardeh
- Dr. Jared Dunnmon
- David Cohn
- Stanford SCIT Fellowship
- RSNA Research Fellow Grant
- NVIDIA GPU Grant

References

- Mokdad AA, Lopez AD, Shahraz S, et al. Liver cirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. BMC Med. 2014;12:145.
- Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. Hepatology 1996; 24:289.
- Kose S, Ersan G, Tatar B, Adar P, Sengel BE. Evaluation of Percutaneous Liver Biopsy Complications in Patients with Chronic Viral Hepatitis. Eurasian J Med. 2015;47(3):161-4.
- Meng, D, Zhang, L, Cao, G, et al. “Liver Fibrosis Classification Based on Transfer Learning and FCNet for Ultrasound Images.” IEEE Access. 2017;5:5804-5810.
- Barr RG. Elastography in clinical practice. Radiologic clinics of North America. 2014; 52: 1145-62.
- Udompap, Prowpanga, Donghee Kim, and W. Ray Kim. "Current and future burden of chronic nonmalignant liver disease." *Clinical Gastroenterology and Hepatology* 13.12 (2015): 2031-2041.
- Pellicoro, Antonella, et al. "Liver fibrosis and repair: immune regulation of wound healing in a solid organ." *Nature Reviews Immunology* 14.3 (2014): 181.
- Czaja, Albert J. "Hepatic inflammation and progressive liver fibrosis in chronic liver disease." *World journal of gastroenterology: WJG* 20.10 (2014): 2515.
- Goodman, Zachary D. "Grading and staging systems for inflammation and fibrosis in chronic liver diseases." *Journal of hepatology* 47.4 (2007): 598-607.
- Shiina T, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 1: basic principles and terminology. Ultrasound in medicine & biology. 2015; 41: 1126-47.
- Kamaya, Aya, et al. "New technologies in clinical ultrasound." *Seminars in roentgenology*. Vol. 48. No. 3. Elsevier, 2013.
- Goya, Cemil, et al. "The role of acoustic radiation force impulse elastography in the differentiation of infectious and neoplastic liver lesions." *Ultrasonic imaging* 37.4 (2015): 312-322.
- Tang, An, et al. "Ultrasound elastography and MR elastography for assessing liver fibrosis: part 1, principles and techniques." *American journal of roentgenology* 205.1 (2015): 22-32.
- Ferraioli G, et al. WFUMB Guidelines and Recommendations for Clinical Use of Ultrasound Elastography: Part 3: Liver. Ultrasound in medicine & biology. 2015; 41: 1161-79.
- Barr, Richard G., et al. "Elastography assessment of liver fibrosis: society of radiologists in ultrasound consensus conference statement." *Radiology* 276.3 (2015): 845-861.
- Lakhani, Paras, and Baskaran Sundaram. "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks." *Radiology* 284.2 (2017): 574-582.
- Cheng, Jie-Zhi, et al. "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans." *Scientific reports* 6 (2016): 24454.
- Liao, Shu, et al. "Representation learning: a unified deep learning framework for automatic prostate MR segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Berlin, Heidelberg, 2013.