COMPUTER-AIDED RENAL ALLOGRAFT ASSESSMENT USING ULTRASOUND ELASTOGRAPHY AND MACHINE LEARNING

A THESIS
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA

BY

LUIZ HENRIQUE DE ARAUJO VASCONCELOS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

Dr. MATTHEW W. URBAN

MAY 2022

Acknowledgements

We would like to acknowledge Mayo Clinic's Department of Radiology and Mrs. Jennifer L. Poston for administrative support. We are grateful to the multiple study coordinators and sonographers that facilitated the patient studies that provided the data used in Chapters 2, 3, 5 and 6. We also would like to acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported in Chapter 5.

The thesis was supported in part by grant R01DK092255, from the National Institutes of Health, and Chapter 4 was partially supported by the Ministry of Science and Higher Education of Poland under agreement no. 0177/E-356/STYP/13/2018 for Dr. Piotr Kijanka's contributions.

The content is solely the responsibility of authors and does not necessarily represent the official views of the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institutes of Health or the Ministry of Science and Higher Education of Poland.

Dedication

I would like to thank the support of my mother, Marise, my father, Cláudio, and my sister, Erica, whom I sorrowfully left back in Brazil. Although physically apart, they are, and always will be, an integral part of my life. These troubled times kept us from meeting as much as we wanted, and needed, though, I am grateful every day for their understanding, strength, and wellbeing.

My wife, Rebeca, whom, not only took the challenge of leaving her family and friends behind, but also, made the dream her own. Our future can only be brighter, no matter where this life takes us.

I would like to thank my dear friends back in Brazil for always being around since high school. Our online gatherings and gaming always bring me closer to home, no matter the latency. I am also grateful for making such strong friendships in Minnesota, although the winter may be long and tiresome, there is always warmth of heart available.

Finally, I would like to thank all the giants upon whose shoulders I now stand. My mentors and friends Matthew Urban and Ivan Nenadic, the Bioinformatics and Computational Biology's chairs Yuk Sham and Chad Myers, thesis committee members John Osborn Jr and David Holmes III, and my dear friends Sara Aristizabal and Piotr Kijanka. I will always look up to you all.

Abstract

Rheological tissue parameters have been shown to correlate with specific histological characteristics related to different pathologies and specifically to kidney rejection. For decades, kidney function tests and biopsy have been used as the main assessment methods for allograft health. With this work we are creating novel approaches for reliable and non-invasive allograft assessment tools by using shear wave elastography measurements with different machine learning algorithms to model the mechanical properties and the pathological changes in the tissue. We also propose to interpret the findings leveraging game theory analysis of the model inputs and outputs to understand what parameters are contributing most for the model prediction. We also intend to better comprehend the progress of kidney rejection from microscopic to macroscopic scales using histology-based models of shear wave propagation. Finally, we propose to create a fast, reliable, and non-invasive allograft assessment method by analyzing the shear wave propagation with minimal signal processing, leveraging convolutional neural networks architectures to retrieve the features from two-dimensional Fourier transform analysis of shear wave data, without the use of complex mathematical and physical models.

Table of Contents

List of Tables vi
List of Figures vii
Chapter 1 – Introduction 1
1.1 Background 1
1.1.1 Chronic Kidney Disease 1
1.1.2 Renal Biopsy 1
1.1.3 Non-invasive Kidney Evaluation 2
1.1.4 Elasticity Imaging 2
1.1.5 Shear Wave Elastography 3
1.1.6 Machine Learning 5
1.1.7 Significance 6
1.2 Specific Aims 7
1.2.1 Specific Aim 1 7
1.2.2 Specific Aim 2 8
Chapter 2 – Kidney allograft assessment using shear wave elastography and clinical biomarkers 11
2.1 Introduction 11
2.2 Methods 13
2.3 Results 17
2.4 Discussion 20
2.5 Conclusion 22
2.6 Appendix 25
Chapter 3 – Interpretable machine learning models for kidney biopsy score prediction based on shear wave elastography 28
3.1 Introduction 28
3.2 Materials and Methods 32
3.3 Results 36
3.4 Discussion 39
3.5 Conclusion 42
3.6 Appendix 43
Chapter 4 – Viscoelastic parameter estimation using simulated shear wave motion and convolutional neural networks 47

4.1 Introduction 47
4.2 Methods 49
4.3 Results 55
4.4 Discussion 60
4.5 Conclusion 64
Chapter 5 – Kidney cortex shear wave motion simulations based on segmented biopsy slides 66
5.1 Introduction 66
5.2 Methods 68
5.3 Results 72
5.4 Discussion 77
5.5 Conclusion 79
Chapter 6 – Renal allograft interstitial inflammation detection based on convolutional neural networks and shear wave motion 2D Fourier transform 81
6.1 Introduction 81
6.2 Methods 83
6.3 Results 86
6.4 Discussion 87
6.5 Conclusion 88
Chapter 7 – Summary and Future Work 89
7.1 Introduction 89
7.2 Summary of Results 90
7.3 Limitations 94
7.4 Future Research 95
7.5 Conclusion 96
7.7 Academic Achievements 97
7.7.1 Peer-reviewed Papers 97
7.7.2 Conference Abstracts 98
7.7.3 Fellowships and Awards 98
Bibliography 99

List of Tables

Table 2-1 – SWE measurements and clinical observations 15
Table 2-2 – Study demographics 17
Table 2-3 – Banff score distribution 18
Table 2-4 – Logistic regression performance 20
Table 2-5 – Full dataset 25
Table 2-6 – L_MID dataset 26
Table 2-7 – Median dataset 27
Table 3-1 – Study cohort demographics 32
Table 3-2 – SWE measurements and clinical observations 33
Table 3-3 – Banff patient score distribution 34
Table 3-4 – Banff scan score distribution 34
Table 3-5 – Low p-value features (p-value < 0.10) 35
Table 3-6 – Top 3 SVM models for each diagnostic outcome 37
Table 4-1 – Model codes 53
Table 4-2 – Elasticity results 56
Table 4-3 – Viscosity results 56
Table 5-1 – Study cohort demographics 68
Table 5-2 – Diagnostic outcome and spatial composition 70
Table 5-3 – Parametric analysis 70
Table 5-4 – *in silico* and *in vivo* comparison 76
Table 5-5 – Average spatial composition 78
Table 6-1 – Study cohort demographics 84
Table 6-2 – Banff patient and scan score distribution 86
Table 6-3 – CNN performance metrics 86

List of Figures

Figure 1-1 – Diagrams of Maxwell (a) and Kelvin-Voigt (b) viscoelastic models. $\sigma(t)$ is the stress applied to the model, whereas $\varepsilon(t)$ is the strain sustained by the model. The parameters in the models are μ_1 as the shear elastic modulus, while μ_2 is the shear viscosity..... 4

Figure 2-1 – Combinations of plane and position used for the patient studies. Longitudinal (L) and transverse (T) planes, and upper pole (UP), middle region (MID) and lower pole (LP). *Image sourced and adapted from Biorender.....14

Figure 2-2 – Box plot of the variation of IQR-to-median ratios across all 13 elastic and viscoelastic measurements, the table on the right summarize the median values for every location. Longitudinal-middle (L_MID) have the lowest overall IQR-to-median ratio.....18

Figure 3-1 – Example of ROC for 5-fold cross-validation test and its SHAP evaluations of a SVM classifier trained with the entire dataset, all features available, no IQR-to-median ratio filtering and undersampling. Each dot represents one evaluation, the color encodes the feature value, and the horizontal position its impact to the model output. In this example, it is possible observe diastolic blood pressure has positive importance, as it increases so does the output of the model. *Only top 20 features are displayed. 36

Figure 3-2 – AUROC distribution for every diagnostic category aggregated by dataset, feature set, IQR-to-median ratio filtering and undersampling (Banff i only). The red line is the median value, the grey box displays the interquartile range (IQR) and the maximum and minimum whiskers set to 1.5 IQR. Outliers are displayed as a blue vertical line. The list of all results is available in the chapter Appendix. 37

Figure 3-3 – Scatter plot of feature mentions at 1st, 2nd, and 3rd importance rank. The marker color encodes which diagnostic outcome was analyzed, and the marker shape encodes the trend (e.g., upright triangle indicates positive trend). Positive trends occur when the output of the classifier increases as the feature value increase, whereas negative trend occurs when the classifier output increases when the feature value decreases. ...38

Figure 4-1 – General schema of a convolutional neural networks. Convolutional and max pooling layers can be stacked for deeper networks. The input image is convolved with a sliding window (yellow square), resulting in a set of feature maps that are processed by the max-pooling layers. The output of the final max-pooling is then processed by a set of

dense layers responsible to interpret the feature maps and provide the estimated value of elasticity or viscosity as output.....48

Figure 4-2 – Images of excitation profiles for F/N = (a) 1 and (b) 2 based on the normalized beam intensity. Lower f-numbers generate more concentrated pushes, whereas with higher f-numbers the push profile is spread axially (z), creating more planar wave fronts. 50

Figure 4-3 – (a) Example of segmentation performed during the preprocessing stage for F/N = 2, μ_1 = 2 kPa and μ_2 = 4 Pa·s. The z-axis origin is the center of excitation (z_f = 25 mm from Fig. 4-1); (b) and (c) examples of 20 and 5 dB SNR. 52

Figure 4-4 – Examples of validation loss (a) and MAE (b) for second round of evaluations, where the combinations were reduced to 3 convolutional layers; 3 and 4 dense layers; 32 nodes per layer (convolutional and dense); 0.001, 0.0005 and 0.0001 learning rates. The lower the values of validation loss and MAE indicated a better model. From these plots 3 convolutional layers, 4 dense layers with 32 nodes and a 0.0005 learning rate were chosen for elasticity regression of F/N = 1 (pink arrow). The curves are plotted using Tensorboard with a smoothing factor of 0.5 (solid color lines) over the calculated values (shaded color lines). 54

Figure 4-5 – Elasticity and viscosity error box plots aggregated by true values, the top ten plots are results for the CNNs trained, validated and tested on data from the same and different F/N datasets, whereas the bottom four were evaluated using Fourier analysis (FT) for comparison. The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing. 57

Figure 4-6 – Elasticity and viscosity error box plots aggregated by vertical position, the top ten plots are results for the CNNs trained, validated and tested on data from the same and different F/N datasets, whereas the bottom four were evaluated using Fourier analysis (FT) for comparison. The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing. 58

Figure 4-7 – Elasticity and viscosity error box plots aggregated by noise levels, the top ten plots are results for the CNNs trained, validated and tested on data from the same and different F/N datasets, whereas the bottom four were evaluated using Fourier analysis

(FT) for comparison. The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing. 59

Figure 4-8 – Elasticity and viscosity error box plots aggregated by noise levels, the top ten plots are results for the CNNs trained, validated and tested on data from the same and different F/N datasets, whereas the bottom four were evaluated using Fourier analysis (FT) for comparison. The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing. 60

Figure 4-9 – Elasticity and viscosity error box plots aggregated by the F/N of the test subject. The model pair was trained, validated and tested on both F/N datasets (E1+2 and V1+2). 60

Figure 4-10 – Example of μ_1 = 18 kPa and μ_2 = 2 Pa·s, at z = +0.90 cm and 20 dB SNR, with F/N = 1 (a) and F/N = 2 (b). It is possible to observe that the response using F/N = 1 has a wider wave front, which might simulate the viscosity effect which also widens the wavefront.....62

Figure 4-11 – Example of μ_1 = 18 kPa and μ_2 = 0 Pa·s , at z = +0.90 cm and 20 dB SNR, with F/N = 1 (a) and F/N = 2 (b). The fringing effect present in F/N = 1 case might explain the elasticity bias present in cross F/N evaluations.....62

Figure 5-1 – (a) Original 100x histology image, (b) segmented version of digitized histology, (c) 2.5 x 5 mm sample of 1.0 x 2.0 cm composite mask used for SGFD simulation. The black lines divide the different cores composing the image.....69

Figure 5-2 – Shear wave velocity maps calculated from baseline simulations of each subject. The rightmost high velocity artifacts are caused by the low amplitude of the shear wave, as it attenuates away from the push at origin. The striated pattern is caused by the limited number of cores available for digitization, and therefore its repetition over the mask composition. 73

Figure 5-3 – Compiled results for each patient, iteration, and biopsy outcome. Each biopsy outcome cluster is composed by the six mechanical parameters that compose the cortex morphology: elasticity and viscosity; for the glomeruli, tubule, and interstitial space. The vertical bars display the standard deviation across the three patients that composed each biopsy outcome. Shear viscosity and slope showed the most sensitivity to rheological

alterations, while alterations to tubules and interstitial space altered the measurements the most. 74

Figure 5-4 – Comparison of each diagnostic outcome depicted in Figure 5-3, against normal patients. The percent difference from each diagnostic outcome mean was calculated and plotted. Inflammation was found to affect the rheological measurements the most. 75

Figure 5-5 – Scatter plot of *in vivo* and *in silico* elastography estimations, linear regression is also displayed. The color of the dots decodes the patient respective diagnostic outcome. 77

Figure 6-1 – General convolutional neural network architecture implemented for 2D FT image classification. Convolutional and max-pooling pairs can be stacked for deeper feature extraction. The feature extracted are resized into a one-dimensional array that is then processed by the fully connected layers to produce the image classification. 83

Figure 6-2 – (a) Original ROI shear motion, (b) shear motion with DC component removed, (c) shear motion after Tukey windowing, (d) final four quadrant 2D FT image used for training. 85

Figure 6-3 – ROC plots for 6.2:1 (a), 2:1 (b) and 1:1 (c) testing. The orange line is the ROC curve, while the dashed blue line is the chance performance (AUROC = 0.50). 87

Chapter 1 – Introduction

1.1 Background

1.1.1 Chronic Kidney Disease

Proper kidney function is critical for maintaining body homeostasis. Chronic kidney disease (CKD), also called chronic kidney failure, describes a long-term gradual decrease in function of the kidneys. In its final stages, CKD can progress to end-stage renal disease (ESRD), which is optimally treated by kidney transplant, with hemodialysis as a temporary option [1]. Although transplant offers better outcomes and patient quality of life, it may not be a life-long solution due to rejection and other systemic conditions. Kidney allografts can also suffer from CKD and ESRD. At the point of failure, the patient ultimately has to return to hemodialysis and transplant queues.

The fact that patients can outlive their allografts is critical, as in 2018 over 90,000 patients were waitlisted to receive transplantation in the Unites States alone [2]. Therefore, the allograft health assessment is crucial to prolong allograft survival and patients' quality of life and reduce care costs. Given the importance of allograft continuous care, the options for surveillance are still limited. Frequent reassessments allow clinicians to administer immunosuppressant medications to maximize allograft lifespan without jeopardizing the patient's immune response [3].

1.1.2 Renal Biopsy

The gold standard procedure for evaluating kidney health is renal biopsy. Using the Banff diagnostic criteria, such as interstitial fibrosis and tubular atrophy (IFTA) [1], [3], nephropathologists can directly assess tissue alterations, even before kidney function is affected. Despite its allograft assessment capabilities, biopsies are invasive and might impose complications such as infection and bleeding, limiting the frequency in which the procedure can be performed [4]. At Mayo Clinic, protocol biopsies are performed at 4 months, and 1, 2, 4, 7, and 10 years post-transplant. Additionally, at Mayo Clinic, biopsies can cost up to 3.5 times that of an abdominal ultrasound exam, and, in the event of complication, the costs associated with the procedures can reach even higher amounts.

1.1.3 Non-invasive Kidney Evaluation

Functional surrogates such as serum creatinine and glomerular filtration rate (GFR) hold important insight to kidney failure. Creatinine is a metabolic subproduct that is freely filtered from circulation by the kidneys, and therefore, its elevated presence is directly related to renal function. GFR is inversely proportional to serum creatinine, as it describes the volume of fluid filtered by the kidneys and is the main index for functional assessment. Unfortunately, direct GFR evaluation is not easily performed, as it requires exogenous filtration monitoring through long periods of time. For that reason, endogenous techniques leveraging serum creatinine and/or cystatin C are common in clinical practice for estimated GFR (eGFR) calculation [5].

Although serum tests can be useful and affordable, morphological alterations of renal tissues can precede renal function alteration [6]. Noninvasive imaging techniques have been implemented for kidney structural and functional assessment [7]–[9]. Computed tomography (CT) has been used to investigate fluid collections and renal artery alterations [10], [11]. Magnetic resonance imaging (MRI) has also been shown to differentiate normal and abnormal kidney cortex through the evaluation of kidney anatomy and perfusion characteristics [12].

CT and MRI are important imaging modalities, nevertheless, tend to lack the affordability and availability of ultrasound. Cortical ultrasound echogenicity has been found to correlate with kidney sclerosis [13]. Spectral Doppler ultrasound can also be used for indirect morphological assessment. Changes in pulsatility and compliance of renal arteries can alter resistive index (RI) indicators. Although more correlated with systemic alterations, Doppler RI is an important biomarker for kidney prognosis [14].

1.1.4 Elasticity imaging

An alternative for allograft assessment is elasticity imaging. By leveraging shear wave motion as a contrast mechanism, it is possible to assess changes in renal allografts directly related to pathology. It has been shown that the shear modulus of biological tissues in the body can vary over six orders of magnitude [15]. Renal allografts have been studied using several elasticity imaging techniques over the last decade [16]–[19]. In contrast to most organ transplant procedures, native kidneys are not removed during surgery and, therefore, kidney allografts do not spatially replace native kidneys. Instead,

the renal allograft is placed in the lower abdomen, more superficially, when compared to native kidneys [20]. Shear wave applications benefit from renal allograft positioning, as wave propagation and detection are facilitated with less tissue attenuation.

Magnetic resonance elastography (MRE) uses an external mechanical actuator to create shear waves in soft tissues and MRI techniques to measure the 3D shear wave motion [21]. MRE has been used to evaluate renal arterial stenosis as well as interstitial fibrosis in renal cortex [22]. Even though MRE has shown potential to characterize different tissue structures, including allografts [23], [24], it is not widely available, and scans can take considerable time. Another technique that has been evaluating liver and renal allograft noninvasive evaluation is transient elastography (TE) [25], [26]. It also relies on external actuators to provide a single cycle of low-frequency (typically around 50 Hz) vibration and ultrasound methods to track the resulting motion [17]. While this method is effective for liver evaluation, it lacks imaging guidance and may not be suitable for evaluation of renal allografts and specific assessment of the renal cortex.

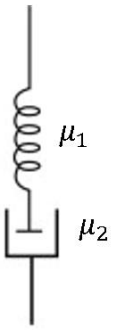
1.1.5 Shear wave elastography

Shear wave elastography (SWE) techniques have been very useful for the analysis of tissue rheological mechanical properties and pathologies. SWE uses acoustic radiation force (ARF) to generate shear waves and high frame rate ultrasonic tracking methods to measure shear wave motion. The motion is analyzed to retrieve elastic and viscoelastic property measurements [15]. Previous studies showed significant correlation of SWE properties, such as shear wave velocity (SWV) [18], [27], [28], and allograft fibrosis. In an elastic medium, the SWV is also proportional to the shear modulus or stiffness. Noninvasive imaging approaches could provide a means to frequently monitor patients with renal allografts if the parameters measured with the imaging method correlate well with biopsy and laboratory results.

In addition to examining the elastic material properties of the kidney, various mechanical testing techniques have been used to examine the viscoelastic properties of cortical and medullary tissues [29]. Viscoelastic materials exhibit time-dependent deformation when a stress is applied. Previous work performed by our laboratory has shown that renal viscoelasticity can gauge mechanical changes in *ex vivo* porcine models that underwent acute renal arterial stenosis [18]. Shear wave phase velocity variation with

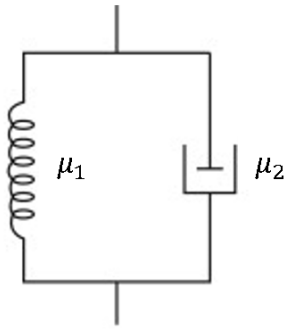
frequency (dispersion) can be used to quantitatively characterize the viscoelasticity of the tissue [30], [31] and assess the degree of fibrosis and steatosis of the liver [32].

(a) Maxwell model



$$\frac{1}{\mu_1}\frac{d\sigma(t)}{dt}+\frac{\sigma(t)}{\mu_2}=\frac{d\varepsilon(t)}{dt}$$

(b) Kelvin-Voigt model



$$\sigma(t)=\mu_1(t)\ +\ \mu_2\frac{d\varepsilon(t)}{dt}$$

Figure 1-1 – Diagrams of Maxwell (a) and Kelvin-Voigt (b) viscoelastic models. $\sigma(t)$ is the stress applied to the model, whereas $\varepsilon(t)$ is the strain induced by the stress. The parameters in the models are μ_1 as the shear elastic modulus, while μ_2 is the shear viscosity.

Several mathematical models have been used to characterize viscoelastic tissue, such as the Kelvin-Voigt (KV) and Maxwell models [33]. Both describe the viscoelastic material as a spring and dashpot system, where $\sigma(t)$ is the stress applied to the model, whereas $\varepsilon(t)$ is the strain sustained by the model, μ_1 is known as the shear elastic modulus (spring), while μ_2 is the shear viscosity (dashpot) (Fig. 1-1). The Maxwell model is composed of a spring and dashpot arranged in series while the KV model has the spring and dashpot arranged in parallel. Therefore, in the Maxwell model, both components share the same stress (and independent strain), whereas in the KV model they are subject to the same strain (with independent stress). For SWE, the KV model has been frequently used as the strain, or displacement, behaves approximately according to a KV model [33]. It is possible then to analyze the KV model for a vibrational system, and, by evaluation of the frequency response, one can derive the SWV as a function of the SW frequency, also known as the SW dispersion. The dispersion equation is given by

$$c_p(\omega)=\sqrt{\frac{2\left(\mu_1^2+\omega^2\mu_2^2\right)}{\rho\left(\mu_1+\sqrt{\mu_1^2+\omega^2\mu_2^2}\right)}}\,, \tag{1-1}$$

where the shear wave velocity (c_p) is described as a function of the angular frequency (ω), the media density ρ, and the viscoelastic parameters μ_1 and μ_2.

Eq. 1-1 allows for the calculation of the viscoelastic parameters μ_1 and μ_2 based on the dispersion characteristics of the observed shear wave propagation. The dispersion curve is most commonly obtained by the two-dimensional (2D) Fourier transform of the acquired spatio-temporal tissue motion. Although different elastic and viscoelastic parameters have been shown to correlate with specific histological characteristics and renal function variables, optimization of the current shear wave data processing methods are needed to make SWE more robust for clinical translation.

Finite difference simulations have been an important tool to better understand shear wave interactions. In this thesis, the staggered-grid finite-difference (SGFD) approach, introduced by Kijanka, *et al*. [34], can be used to implement a KV material using the Navier-Stokes equations to calculate the shear motion caused by ARF. SGFD allows researchers to have full flexibility of simulation parameters, including temporal and spatial resolutions, and custom acoustic radiation force profiles for material excitation. Another advantage is the pixel-based viscoelastic parameter assignment. The staggered-grid viscoelastic parameters can be defined independently for each node and, therefore, highly heterogeneous materials can be modeled and simulated.

1.1.6 Machine Learning

Machine learning (ML) is the computer science field dedicated to creating algorithms capable of learning and making predictions from a given dataset. Substantial progress has been made in recent decades on both algorithms and computational power to perform data training and predictions in feasible time [35]. ML can identify patterns in datasets that otherwise might be impossible to fit by standard statistical models, or by evaluating each feature on its own. ML framework may allow biomedical researchers to predict well-established gold standard biopsy scores using more novel sets of measurements and/or biomarkers.

Recent progress in neural networks (NN), a form of ML, also enabled the expansion of the field and its adoption in many research and commercial fields, as they are particularly powerful for extracting non-linear features from data [36]. Novel NN architectures, such as convolutional neural networks (CNN) [37], have also shown

unprecedented accuracy for classifying two-dimensional datasets such as images. The use of such capabilities might open new possibilities for ultrasound imaging diagnostics.

After proper implementation, NN techniques have the potential of performing data processing more efficiently, as they require much lower computational costs to predict outcomes from a trained model than current image processing techniques. Neural network architectures may allow the use of data in a lower pre-processing state, such as dispersion curves, Fourier representations from two-dimensional Fourier transform (2D FT) analysis, and shear wave velocity maps, to classify the allograft health in a more generalizable and time efficient fashion, enabling real-time computer-aided diagnostics (CAD) implementations.

In recent years, there has been an increasing interest in the employment of ML in SWE applications. The combination of NN and 2D SWE have been used for various implementations, including diagnosis of liver fibrosis [38], reconstruction of 2D shear modulus maps, and lesion segmentation [39]. Although, the results have shown the potential of such applications the employment of ML techniques to augment performance of SWE applications is still in its initial stages.

To possibly achieve classification capabilities comparable to biopsies, classic statistical and ML methods must face the high inter- and intra-operator variability of ultrasound imaging, which requires large amounts of data to be utilized [35]. Although the correlation between elastography measurements and kidney alterations have been documented in literature [16], elastography still relies on complex pre-processing techniques, such as directional and band-pass filtering, that can introduce variability and bias.

1.1.7 Significance

Continual renal evaluation is critical for allograft preservation. Even though renal biopsy is the gold standard for allograft assessment, the invasive nature of the procedure does not allow its frequent use by clinicians. Non-invasive functional tests can be leveraged to indirectly assess renal state but are not optimal, as functional alterations may not reflect the full health of the kidney. Rheological tests can be more sensitive to morphological changes, and therefore, allow for early diagnosis. Elastic and viscoelastic parameters have been shown to correlate with renal histological and functional alterations, but the current shear wave data processing techniques still lack the reliability for robust

clinical translation in nephrology applications. SWE implementations are still heavily dependent on the device characteristics and multiple pre-processing stages are necessary for viscoelastic quantification.

The use of SWE combined with machine learning modeling techniques might enable more efficient allograft reassessment, as trained classifiers can achieve higher classification performance than biomarkers evaluated individually. The use of efficient and generalizable implementations can open the path for clinical translation and real-time applications, even in inter-institutional applications [38]. ML algorithms are capable of reaching high levels of performance, but in multiple instances it lacks the interpretability of what data patterns are being modeled, this “black box” characteristic can hinder clinical translation, as it may not inspire practitioners’ confidence. The goal of this research is to introduce faster, cheaper, non-invasive, and interpretable techniques capable of assessing allograft health reliably and more frequently, enabling more efficient immunosuppressant administration and, therefore, better quality of life for transplant patients.

1.2 Specific Aims

1.2.1 Specific Aim 1

Develop supervised machine learning models that can predict renal transplant biopsy scores based on SWE measurements and other physiological biomarkers with >80% specificity and sensitivity.

Hypothesis: SWE measurements of kidney allograft can be used to predict renal transplant biopsy outcomes and, therefore, assess rejection. Machine learning methods might allow elastographic measurements, such as shear wave velocity and attenuation, along with other physiological biomarkers, to classify allograft health with clinically relevant sensitivity and specificity.

Rationale: Noninvasive imaging approaches could provide means to frequently monitor patients with renal allografts if the parameters measured with the imaging method correlate well with biopsy and laboratory results. It has been shown by other groups and in our preliminary results that SWE characterization of renal allografts correlates with declining function and histological features [29], [30], [32]. However, the features analyzed

do not hold the characterization power necessary for clinical translation individually, and so, a multi-parametric elastographic characterization of renal allografts is needed. The use of machine learning algorithms opens the possibility of combining the measurements obtained with SWE to produce a prediction model for biopsy's scores that would not be feasible with classic statistical models.

Chapter 2: We investigated the use of ultrasound shear wave elastography methods and clinical observations for kidney allograft health assessment. The SWE measurements and biomarkers were obtained prior to routine biopsies of kidney transplants at Mayo Clinic. General estimating equations and odds ratio statistical analyses were performed to evaluate the significance of each parameter individually, and logistic regression was used for combining the parameters into a single binary outcome probability for each of the biopsy scores. The p-values, odds ratios and logistic regression area under the receiver operating characteristic curve (AUROC) were reported.

Chapter 3: Support vector machines (SVMs) were evaluated and interpreted for classification of kidney allograft's biopsy scores based on the viscoelastic, physiological, and demographical observations described in Chapter 2. Different feature sets were used to test viscoelastic parameters and biomarkers in isolation along with the subset of features with p-value < 0.10, as evaluated in Chapter 2. The IQR-to-median ratio was also tested as a data quality filter by calculation data acquisition variability over each viscoelastic feature, position, and patient. Classification performance indicators such as AUROC, sensitivity and specificity were reported.

1.2.2 Specific Aim 2

Develop supervised machine learning models that can predict renal transplant biopsy scores based on analysis of data from shear wave propagation simulation and patient data with minimal pre-processing.

Hypothesis: SWE imaging of kidney allograft can be used to predict biopsy outcomes and, therefore, assess rejection, without the necessity of the use of models and calculation of rheological parameters, such as shear wave speed and attenuation. Machine learning can be leveraged to evaluate histopathological changes in the tissue that modify shear wave propagation characteristics.

Rationale: It has been shown that parameters from SWE characterization of renal allografts correlate with declining function and histological features [29], [30], [32]. However, the current implementations rely on mathematical models developed under mechanical assumptions that do not always hold true in clinical practice, such as homogeneity, isotropy and linearity. The use of machine learning techniques, such as convolutional neural networks (CNN), may allow for a direct assessment of the physiological and morphological disease progression underlying the rheological measurement changes. The ML performance can be evaluated at multiple stages of the current image processing pipeline to obtain maximum classification accuracy at optimum processing speed and generalizability.

Expected Outcomes: We aspire to create novel machine learning models capable of classifying allograft health from patient SWE data with minimal signal processing. We also intended to further develop the simulation models created during the preliminary study, and better understand how histology changes can alter the rheological characteristics of the tissue. The NN models developed will be trained and tested to maximize accuracy and generalizability, creating a reliable computer-aided diagnostic tool for allograft assessment.

Chapter 4: The capabilities of CNNs to extract features from wave propagation and estimate elasticity and viscosity values from wave motion images sourced from staggered-grid finite-difference (SGFD) simulations were investigated. A total of 1050 SGFD models were simulated using a wide range of elasticity and viscosity values for two different values of excitation F/N. The dataset was used to train, validate and test CNN architectures configured for data regression using mean squared error as loss function. The viscoelastic parameters were also estimated using the conventional 2D Fourier transform (2D FT) analysis approach with Kelvin-Voigt dispersion curve fitting and compared against the CNN results.

Chapter 5: The evaluation of heterogenous kidney cortex simulations based on human renal allograft biopsies was performed. Staggered-grid finite-difference (SGFD) were leveraged to simulate shear wave propagation of 12 patients from the cohort examined in Chapters 2 and 3 (3 normal, 3 with interstitial inflammation, 3 with interstitial fibrosis and 3 with tubular atrophy). Time-to-peak analysis and KV dispersion curve fitting was used to estimate common viscoelastic paraments to the simulated data as well as the SWE scans performed on the same patients. The Pearson correlation coefficient was used to calculate the correlation between both *in vivo* and *in silico* estimations.

Chapter 6: A proof-of-concept study was proposed to evaluate a novel 2D FT CNN technique for kidney interstitial inflammation detection. The patient dataset used in Chapters 2, 3 and 5 was leveraged in a preprocessed form. The shear wave motion was extracted from the raw in-phase/quadrature acquisition data and processed using 2D FT in space and time. The 2D FT images were used to train a CNN to detect interstitial inflammation, classification performance indicators such as AUROC, recall and precision were reported.

Chapter 2 – Kidney allograft assessment using shear wave elastography and clinical biomarkers

2.1 Introduction

Chronic kidney disease (CKD) encompasses a long-term decrease in function of the kidneys. CKD can progress to kidney failure or end-stage renal disease (ESRD), which is treated by hemodialysis or, preferably, kidney transplant. Patient and renal graft short-term survival rates have increased over the past two decades, but long-term survival of grafts is still an issue [40]. Patients can outlive the transplanted organs requiring hemodialysis and new transplants.

It is critical to assess kidney function in management of transplant patients to save and prolong life, improve quality of life, and reduce care costs. Various indicators such as serum biomarkers and noninvasive imaging methods are utilized to provide clinicians with feedback on the health of a patient's kidneys. However, renal biopsy is still the gold standard for diagnosis of kidney health, but it cannot be used frequently due to its invasive characteristics that can cause complications. An ideal indicator would be noninvasive and could be used frequently enough for clinical feedback to be obtained on disease progression and treatment regimens.

Noninvasive indicators such as serum creatinine (SCr), glomerular filtration rate (GFR) and classical medical imaging can be leveraged to provide insights into kidney disease state. Because creatinine is freely removed from the circulatory system by the kidneys, its levels are inversely proportional to GFR. GFR is considered the best overall index of kidney health and disease [1], [5], [41]–[43]. However, direct measurement of GFR is not easily performed in clinical practice, so equations have been developed that include the SCr or serum cystatin levels, age, race and sex (eGFR) [5], [44], [45]. Additionally, spectral Doppler ultrasound can be leveraged to evaluate the pulsatility and compliance of renal arteries with the calculation of the resistive index (RI). Although more correlated with systemic alterations, it holds an important biomarker for kidney prognosis [14].

Biopsies can help to guide treatment and establish a prognosis, especially in assessing allograft rejection by evaluating the degree of fibrosis or inflammation [46]–[51]. It has become more evident that examining fibrosis and inflammation in combination,

rather than alone, in 1 year protocol biopsies has more predictive power to determine which allografts may have functional decline [52]–[54]. At Mayo Clinic, protocol biopsies are performed at 4 months, and 1, 2, 4, 7, and 10 years post-transplant. These protocol biopsies can be valuable indicators of graft rejection [48]–[51], [55].

Renal biopsies are routinely performed, but they are associated with complications such as small and large hematomas, gross hematuria, arteriovenous fistula, and even death [56]–[58]. The complication rate has been reported to be in range of 0.1-6% [59], [60]. With a projection of an increase in the number of biopsies [61], the potential number of patients with complications also increases.

Elasticity imaging is a noninvasive imaging modality that uses the elasticity of tissue as the contrast mechanism. Shear wave elastography (SWE) uses acoustic radiation force (ARF) to generate shear waves and high frame rate ultrasonic tracking methods measure the propagation speed. Previous studies have reported significant correlation between shear wave velocity and the level of fibrosis found in biopsy examinations, allograft RI, the dysfunction of allografts physiology and negative correlation with eGFR [16], [19], [62]–[64].

In addition to examining the elastic material properties of the kidney, various mechanical testing techniques have been used to examine the viscoelastic and nonlinear mechanical properties of cortical and medullary tissues [29], [65]–[68].

Viscoelastic materials exhibit time-dependent deformation when a stress is applied. The modulus of nonlinear materials varies with the applied stress or strain. Shear waves can be used to investigate the viscoelastic properties of tissue. Shear wave phase velocity dispersion, or variation with frequency, can be used to quantitatively characterize the viscoelasticity of the medium [69]. Measurements in liver have shown that the shear elasticity and viscosity increase as the degree of fibrosis and steatosis increases [32], [70]–[74].

Our group has developed methods to measure viscoelasticity with fitting phase velocity dispersion, variation with frequency, with a model-based approach using a rheological model [69], [75] as well as methods that do not depend on a rheological model, but just the phase velocity and attenuation [76]. This model-free method has been very successful in separating patients with and without acute cellular rejection in liver transplants [32], [77], [78].

In this study, we evaluated how elastic and viscoelastic mechanical properties were associated with biopsy scores to evaluate renal transplants undergoing protocol biopsies.

2.2 Methods

The patient study was conducted from October 2017 to July 2019, following the protocol approved by the Mayo Clinic Institutional Review Board (IRB# 11-003249). The patients were recruited at the time of their routine follow-up visits and provided written informed consent. The ultrasound scan was performed before the subject's biopsy procedure using a General Electric Logiq E9 scanner (General Electric Healthcare, Wauwatosa, WI, USA) using a curvilinear array (C1-6, General Electric Healthcare, Wauwatosa, WI, USA). A research mode was enabled on the ultrasound scanner to store the raw in-phase/quadrature (IQ) data for later processing.

All ultrasound scans were performed at Mayo Clinic's Methodist Hospital by a team of sonographers trained in shear wave elastography ultrasound imaging. The sonographers were advised to avoid physical pressure at skin surface, because transducer force can alter SWE measurements, as kidney allografts are placed at a shallow depth [42].

The ultrasound SWE scans were performed at three different positions of the allograft (lower pole, middle region, upper pole) and in different planes (longitudinal and transverse) (Fig. 2-1). At every combination of position and plane, 10-13 shear wave excitations were performed and recorded. The kidney dimensions and Doppler resistive index (RI) measurements were also recorded.

For the processing of the SWE data, the B-mode image was reconstructed from the IQ data and a region-of-interest (ROI) within the kidney cortex was manually selected for shear wave motion reconstruction. With the cortical shear wave motion isolated, it was then possible to analyze a variety of elastic parameters such as group velocity, c_g, the elastic shear modulus, μ, and viscoelastic parameters such as shear wave velocities and attenuation at different frequencies and rheological model parameters estimated from fitting shear wave velocity dispersion.

The group velocity, c_g, and shear modulus, μ, were reconstructed using the methods developed in our laboratory [79], [80]. The values of the shear modulus were calculated based on $\mu = \rho c_g^2$ where $\rho = 1000$ kg/m^3 is assumed.

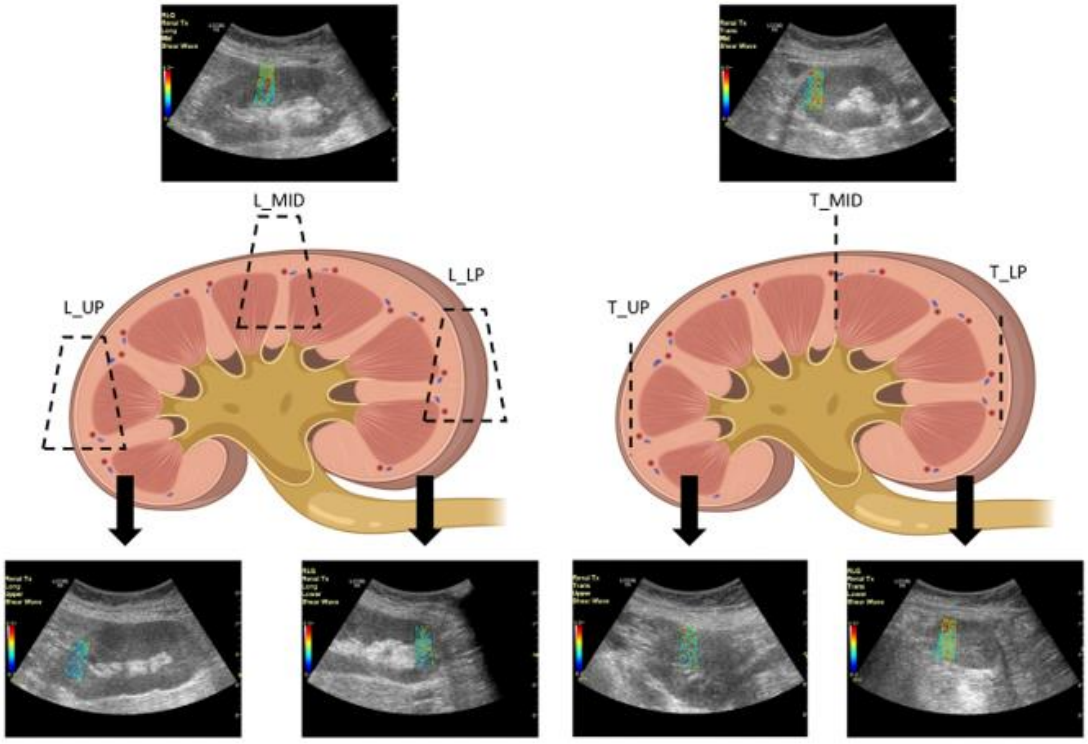


Figure 2-1 – Combinations of plane and position used for the patient studies. Longitudinal (L) and transverse (T) planes, and upper pole (UP), middle region (MID) and lower pole (LP). *Image sourced and adapted from Biorender.

The dispersion curves were obtained using the generalized Stockwell transformation combined with a slant frequency-wavenumber analysis (GST-SFK) [81] technique developed by our group. This algorithm produces lower errors over a wider frequency band in comparison to a two-dimensional Fourier transform (2D FT) approach [82].

To obtain the values for shear elasticity, μ_1, and shear viscosity, μ_1, the dispersion curves were then fit to the theoretical Kelvin-Voigt model. The Kelvin-Voigt model is a commonly used rheological model, as it has been shown to describe the shear wave velocity dispersion successfully, particularly within the frequency range analyzed [83]–[85].

$$c_p(\omega) = \sqrt{\frac{2\left(\mu_1^2+\omega^2\mu_2^2\right)}{\rho\left(\mu_1+\sqrt{\mu_1^2+\omega^2\mu_2^2}\right)}}, \tag{2-1}$$

where the shear wave speed (c_p) is described as a function of the angular frequency (ω), the media density ρ, and the viscoelastic parameters μ_1 and μ_2.

We also used the 2D FT, or k-space approach, to extract phase velocity and attenuation at various frequencies. From the k-space data, we estimated attenuation using $\alpha = \left(\pi/\sqrt{3}\right)FWHM_k$ where $FWHM_k$ is the full-width at half maximum in the k-direction (spatial frequency) [78]. The characterization of the phase velocity and attenuation does not require a rheological model and has been very successful in evaluating acute cellular rejection in liver transplants [78]. All measurements evaluated during this study are summarized by Table 2-1.

Table 2-1 – SWE measurements and clinical observations

Column	**Description**	**Column**	**Description**
Location	Scan position	Sex	Patient sex
c_g	Group velocity, m/s	Kidney Depth, mm	Cortex depth on ultrasound image
μ	Shear Modulus, m/s	BP_S, mmHg	Systolic blood pressure
μ_1	Kelvin-Voigt μ_1, Pa	BP_D, mmHg	Diastolic blood pressure
μ_2	Kelvin-Voigt μ_2, Pa·s	RI_{UP}	Restive index at upper pole
c_{2DFT}^{100}	2D FT phase velocity at 100 Hz, m/s	RI_{MID}	Restive index at middle
c_{2DFT}^{200}	2D FT phase velocity at 200 Hz, m/s	RI_{LP}	Restive index at lower pole
α_{2DFT}^{100}	2D FT phase attenuation at 100 Hz, Np/m	Time After Transplant (TAT), months	Months after transplant
α_{2DFT}^{200}	2D FT phase attenuation at 200 Hz, Np/m	BMI, kg/m^2	Body mass index
c_{SFK}^{100}	GST-SFK phase velocity at 100 Hz, m/s	SCr, mg/dL	Serum creatinine
c_{SFK}^{150}	GST-SFK phase velocity at 150 Hz, m/s	eGFR, mL/min/1. 73 m^2	Estimated glomerular filtration rate
c_{SFK}^{200}	GST-SFK phase velocity at 200 Hz, m/s	Banff i	Interstitial inflammation score
c_{SFK}^{250}	GST-SFK phase velocity at 250 Hz, m/s	Banff ci	Interstitial fibrosis score
c_{SFK}^{300}	GST-SFK phase velocity at 300 Hz, m/s	Banff ct	Tubular atrophy score
Age	Patient age, yrs	IFTA	Interstitial fibrosis and tubular atrophy (ci + ct)

The measurements were also evaluated for their reliability by calculating the ratio of the interquartile range (IQR) and the median of the parameter values within the 10-13 measurements done at each subject, scan position and plane. The IQR-to-median ratio is widely used in the elastography field as a reliability measure of quantitative measurements, conventionally IQR-to-median ratios below 0.3 are deemed reliable [86], [87].

To evaluate the predictive power of the techniques tested, the biopsy findings from nephropathologists were recorded. The Banff criteria was used to document the level of injury sustained by the allografts [46], [88]. Interstitial inflammation (Banff i), interstitial fibrosis (Banff ci) and tubular atrophy (Banff ct) scores range from 0 to 3, from not present to severe injury [88]–[90]. Additionally, Banff ci and Banff ct were combined into interstitial fibrosis and tubular atrophy (IFTA) by aggregating the scores. IFTA is defined by the floor of the sum of Banff ci and ct.

Three datasets were analyzed in this study. The first was the full dataset with all individual measurements and positions. One subset used only the L_MID location data as that had the lowest IQR-to-median ratio values. Another subset was created by calculating the median value of each parameter at each of the six locations, to decrease the level of noise and outliers.

We used a statistical analysis method called general estimating equations (GEE), a method of dealing with correlated data when, except for the correlation among responses, the data can be modeled as a generalized linear model [91]. The GEE parameters are robust to covariance, and therefore, appropriate in the analysis of the data from this research work, as the patient's elastography measurements can be correlated [92]. The GENMOD procedure from SAS version 9.4 (SAS Inc. Cary, NC), allow us to perform the analysis to account for this correlation while assessing the associations between an outcome and predictor variables.

Outcome variables in our research are all binary thus the GENMOD procedure from SAS allows us to use a logit link which provides a means to quantify the magnitude of the associations as odds ratios (OR) and 95% confidence intervals, as in the case of usual logistic regression. If the number of regression parameters is not matched by the number of response pairs for estimating correlation, the model estimation does not converge and therefore no OR and/or p-value are reported.

Additionally, to evaluate the predictive performance of the dataset using multiple measurements and observations, we implemented logistic regression algorithms in Python (Python Software Foundation, Wilmington, DE, USA) and the scikit-learn module [93]. This analysis is performed at an observation level rather than at a patient level. Thus, it treats each observation as an independent observation. The model can describe the relationship between multiple independent variables and a single binary outcome, and therefore, can be used for prediction purposes.

Logistic regression has been used extensively in the medical field to predict mortality and severity of different diagnostics based on observed patient characteristics [73]. The logistic regression hyper-parameters were adjusted using nested 5-fold cross-validation and the final model was tested using 5-fold cross-validation, the results from each fold were compiled to create a single mean receiver operating characteristic (ROC) curve and the area under the ROC (AUROC) was also calculated. For each dataset analyzed (full, L_MID, and median), the logistic regression was performed using all patient measurements, elastography measurements only, biomarker observations only, and, for purposes of dimensionality reduction, a subset containing only the variables with p-value < 0.1 for the specific dataset was also evaluated.

2.3 Results

This study is composed by 206 patients totaling 223 studies, because some patients were assessed at multiple follow-up visits, but each visit is treated as an independent study. There was a total of 122 male and 84 female participants. The age at time of study was 49.72 ± 14.02 years-old (mean $\pm$ standard deviation). The average time between transplant and the scan was 16.12 ± 14.74 months, and the BMI at time of scan was 27.85 ± 5.45. The study demographics are provided in Table 2-2. The patients' Banff score distributions are summarized in Table 2-3. To evaluate the most reliable position and plane combination, the median IQR-to-median ratio for all measurements was taken and the results are displayed in Figure 2-2.

Table 2-2 – Study demographics

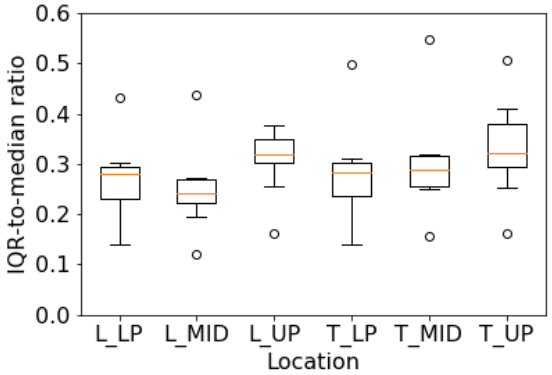
	Patients (n=206)		Studies (n=223)	
	Number	%	Number	%
Sex				
Male	122	59.22	128	57.40
Female	84	40.78	95	42.60
	Mean (Std. Dev)	Range	Mean (Std. Dev)	Range
Age, yrs	49.83 (13.98)*	22-76	49.72 (14.02)	22-76
BMI, kg/m^2	27.59 (5.28)*	13.1-42.3	27.85 (5.45)	13.1-42.3
TAT (months)	14.53 (14.92)*	4-60	16.12 (14.74)	4-60
*At first scan				

The GEE and OR results are summarized in appendix Tables 2-5, 2-6 and 2-7, for the original dataset, L_MID position only, and median datasets, respectively. All datasets showed correlations of some of the elastic and viscoelastic properties in

addition to clinical markers. For 0.05 < p ≤ 0.10, we interpret these parameters as having a diagnostic trend (yellow), whereas p-values ≤ 0.05 were interpreted as statistically significant (orange). In some cases, the GEE model did not converge, and therefore no value of the OR and/or p-value is displayed (red).

Table 2-3 – Banff score distribution

Score	Banff i (n=223)		Banff ci (n=223)		Banff ct (n=223)		IFTA (n=223)	
	Number	%	Number	%	Number	%	Number	%
0	192	86.10	127	56.95	72	32.29	127	56.95
1	17	7.62	84	37.67	139	62.33	84	37.67
2	10	4.48	11	4.93	11	4.93	11	4.93
3	4	1.79	1	0.45	1	0.45	1	0.45
>0	31	13.90	96	43.05	151	67.71	96	43.05



LOCATION	MEDIAN
L_LP	0.28
L_MID	0.24
L_UP	0.32
T_LP	0.28
T_MID	0.29
T_UP	0.32

Figure 2-2 – Box plot of the variation of IQR-to-median ratios across all 13 elastic and viscoelastic measurements, the table on the right summarize the median values for every location. Longitudinal-middle (L_MID) have the lowest overall IQR-to-median ratio.

The full dataset showed several statistically significant correlations between the measurements analyzed and the four biopsy scores evaluated. For Banff i, c_{2DFT}^{200}, α_{2DFT}^{200} and BP_D showed statistical significance, while c_{SFK}^{200} and sex showed statistical trend. Banff ci had statistical significance found with c_{SFK}^{200}, Age, TAT and BMI; and statistical trends with c_g, c_{SFK}^{150}, c_{SFK}^{250}, c_{SFK}^{300} and RI_{UP}. Clinical biomarkers showed good performance for Banff ct, with Age, BP_S, TAT and SCr p-values below statistical significance, along with c_g and μ. Only α_{2DFT}^{200} showed statistical trend with Banff ct diagnosis. IFTA showed good correlations with both viscoelastic and clinical parameters. c_{SFK}^{150} showed statistical trend, while c_{SFK}^{200} and c_{SFK}^{250} showed statistical significance. Sex, RI_{UP}, TAT and BMI showed statistical significance with IFTA diagnosis.

The L_MID dataset showed lower predictive power when compared with the full dataset. Sex, eGFR and c_g were statistically significant and BP_D showed statistical trend

for Banff i. Banff ci diagnosis had statistically significant p-values for Age and time after transplant only. Banff ct had the best predictive results, with α_{2DFT}^{100}, α_{2DFT}^{200}, Age, BP_S, TAT and SCr had statistical significance, and RI_{MID} showing statistical trend. IFTA also did not showed any significant correlation with viscoelastic measurements in this specific dataset. Age and TAT showed statistical significance, whereas BP_S had statistical trend.

The median dataset had similar results to the L_MID dataset. The parameters μ_1, sex and eGFR showed statistical significance for inflammation diagnosis. The parameters c_g and BP_S showed statistical trend for Banff ci diagnosis, whereas α_{2DFT}^{200}, Age and TAT showed statistical significance. For Banff ct, similarly to L_MID, α_{2DFT}^{100}, α_{2DFT}^{200}, Age, BP_S, TAT and SCr had statistical significance. Finally, α_{2DFT}^{100}, Age and TAT showed statistical significance, with BP_S showing statistical trend.

The OR analysis showed TAT to have positive correlation with the prevalence of ci, ct and IFTA, from 1.026 OR (L_MID/IFTA) to 1.679 OR (Full/ct). Age also had positive correlation with OR ranging from 1.057 for inflammation up to 1.218 for tubular atrophy. Serum creatinine also showed strong positive correlation with the prevalence of ct (ranging from 1.492 to 1.522 OR), whereas eGFR showed negative correlation with Banff i (0.845-0.8741 OR). Overall, the elastography measurements showed nearly unitary OR, apart from 2D FT attenuation at 200 Hz (α_{2DFT}^{200}) which showed negative correlation with inflammation for the full dataset at 0.965 OR.

The logistic regression results are shown in Table 2-4, the AUROCs varied from chance (0.50) to 0.72. The standard deviation ranged from 0.03 to 0.10. Overall, the L_MID dataset had the best results in allograft classification, with 0.63, 0.66 and 0.72 for Banff i, ci and ct, respectively, when applying the low p-value ($p < 0.10$) measurements. L_MID also performed best for IFTA classification with 0.66 AUROC when fit to biomarkers only, although it performed similarly when fitted to all measurement and low p-value only, at 0.63 and 0.64, respectively. The full, L_MID only and median datasets had an overall mean of 0.58, 0.61 and 0.60, respectively.

Banff ct showed the best classification potential as it was the only score to reach AUROC above 0.70 and a mean of 0.62, whereas Banff i, had the weakest results, with a single test above the 0.60 mark and a mean of 0.55. Banff ci and IFTA had very similar performance throughout the tests, both with a mean of AUROC = 0.61.

Elastography measurements alone performed the weakest overall without reaching 0.60 for any of the tests performed and an overall mean of 0.54. The low p-value

set performed best with two tests reaching 0.70, for Banff ct classification using the L_MID only dataset and median dataset. Both low p-value and biomarkers measurement sets had a mean of 0.63, while classifying using all measurements followed closely at AUROC = 0.62.

Table 2-4 – Logistic regression performance

Dataset	Measurements	Mean AUROC ± st. dev.			
		Banff i	Banff ci	Banff ct	IFTA
Full	All	0.58 ± 0.09	0.62 ± 0.09	0.61 ± 0.10	0.61 ± 0.06
	p-value < 0.10	0.59 ± 0.06	0.60 ± 0.03	0.67 ± 0.07	0.62 ± 0.07
	Elastography	0.51 ± 0.03	0.54 ± 0.03	0.51 ± 0.03	0.52 ± 0.05
	Biomarkers	0.51 ± 0.08	0.62 ± 0.03	0.63 ± 0.08	0.58 ± 0.06
L_MID only	All	0.51 ± 0.08	0.64 ± 0.05	0.64 ± 0.08	0.63 ± 0.08
	p-value < 0.10	0.63 ± 0.09	0.66 ± 0.09	0.72 ± 0.06	0.64 ± 0.04
	Elastography	0.56 ± 0.07	0.58 ± 0.04	0.50 ± 0.05	0.56 ± 0.04
	Biomarkers	0.52 ± 0.07	0.64 ± 0.09	0.66 ± 0.08	0.66 ± 0.03
Median	All	0.52 ± 0.08	0.62 ± 0.07	0.64 ± 0.07	0.63 ± 0.05
	p-value < 0.10	0.56 ± 0.07	0.66 ± 0.06	0.70 ± 0.05	0.62 ± 0.05
	Elastography	0.55 ± 0.06	0.55 ± 0.02	0.52 ± 0.04	0.54 ± 0.04
	Biomarkers	0.57 ± 0.10	0.64 ± 0.08	0.64 ± 0.10	0.65 ± 0.07

2.4 Discussion

The general estimating equations and odds ratio analysis revealed meaningful insight on the datasets produced. The positive correlation and significance between temporal variables, Age and TAT, showed an expected trend of degradation of the allograft with time, although it was interesting to notice that this behavior was not evident for inflammation. While fibrosis and tubular atrophy increase with the age of the patient and the TAT, inflammation does not seem to show the same correlation and might even be more prevalent at the earlier post-transplant stages, as the <1 odds ratio indicate. Additionally, systolic blood pressure showed significance for the prevalence of interstitial fibrosis and tubular atrophy (ci, ct and IFTA), with a positive trend that indicates these injuries might be affected by hypertension. Inflammation, alternatively, showed positive correlation with diastolic blood pressure instead. Blood pressure prevalence is known to increase with age and cause injury to kidney function, and therefore, the findings are in agreement with literature [94].

Serum creatinine and eGFR showed significance for tubular atrophy and inflammation, respectively. Creatinine should be freely filtrated form circulation, and the positive correlation with tubular atrophy agrees with the decline in kidney function. The

eGFR behaves oppositely, with a negative correlation with inflammation. Differently from interstitial fibrosis and tubular atrophy, inflammation is not permanent and may not be histologically present at routine evaluations, hindering the course of treatment. The overall time and function biomarkers' trends agreed with literature, as fibrosis is usually the result of a chronic inflammatory state [52], [54].

All datasets also showed significance between the AMUSE attenuation measurements and tubular atrophy and, in the Median dataset, with fibrosis as well. Attenuation also had statistical significance with respect to inflammation when the full dataset was evaluated. The group and phase velocities showed different significance for different datasets. Group velocity and shear modulus showed significance for ci and ct, on the full dataset; Banff i, on the L_MID dataset and i and ci, on the Median dataset. Phase velocity measurements calculated from the 2D FT had significance only for inflammation diagnostic using the full dataset, whereas GST-SFK showed significance in all frequencies for fibrosis using the full dataset. Overall, attenuation measurements showed negative correlation with injury, while velocity measurements have positive correlation with injury, in agreement with previous findings reported by our group [69], [73], [75], [78].

The logistic regression results agreed with the IQR-to-median ratios calculated, L_MID had superior performance when compared to evaluating all positions. The L_MID position tends to have the simplest geometry, anisotropy profile and, in most cases, is the shallower region of the allograft. These characteristics lead to better shear wave propagation, and therefore, tissue characterization. Other positions impose complex curved geometries and variable anisotropy that might confound the SWE measurements.

As expected, the clinical biomarkers performed reasonably well when isolated from the SWE measurements, these parameters are widely used clinically and can be used for biopsy pre-screening. The elastography measurements did not perform well when isolated with a mean AUROC ranging from 0.51-0.58. The logistic regression approach implies that all observations are independent and show little to no multicollinearity, which does not hold true for SWE measurements as phase velocity at multiple frequencies, for example, hold high correlation. Despite these findings, the low p-value variables' set performed the best, when including few SWE measurements along with significant biomarkers. When combining the L_MID dataset with the low p-value approach, the logistic regression obtained its best results with a mean of AUROC = 0.66 and a high of AUROC = 0.72 for Banff ct classification. Therefore, it was shown that SWE measurements hold classification significance when combined with well stablished biomarkers. The use of all observations

had an average performance of AUROC = 0.62, and it might be affected by the “curse of dimensionality” as linear models tend to suffer greatly from higher dimensional problems. The statistical significance approach for feature extraction has proven to be effective but it does not account for possible interactions between the different observations.

Another limitation imposed by the characteristics of patient study is the lack of severe observations of the analyzed scores (scores ≥ 2). Over 90% of the biopsy outcomes were non-present to mild manifestations, therefore, the rheological and functional alterations to the allograft can be minimal. Subjects with more advanced rejection or acute rejection, should be targeted for recruitment in present and future studies.

Additionally, even though this study was performed within a single institution, there is still some level of operator variability that may confound the methods proposed. The pressure applied by the sonographer can alter SWE measurements as it compresses the tissue [95], mainly in higher BMI patients where adipose tissue might prevent optimal ultrasound imaging and ARF excitation. There is also a certain level of variability within the nephropathologist team, and in future multi-institutional studies, the implementation of a voting diagnostic system between multiple operators will be imperative to ensure label consistency.

These results show the potential of the methods proposed for the classification and pre-screening of kidney allograft health. The use of machine learning techniques that overcome the limitations of the logistic regression might enable superior performance for the classification task.

2.5 Conclusion

In this study, we evaluated the use of ultrasound shear wave elastography methods and clinical observations for kidney allograft health assessment. The SWE measurements and biomarkers were obtained during kidney transplanted patients’ routine biopsies at Mayo Clinic. The patients were scanned at bedside and subsequently had the graft biopsy performed. The SWE measurements were acquired from three different kidney locations (upper pole, middle region, and lower pole) and two different transducer orientations (longitudinal and transverse). Group velocity, shear modulus, Kelvin-Voigt dispersion curve fitting (μ_1 and μ_2), AMUSE (2DFT phase velocity and attenuation) and GST-SFT (phase velocity) were implemented to obtain the quantification of the rheological

properties of the kidney. Standard of care biomarkers, such as, serum creatinine and eGFR, were also used for allograft assessment. The biopsies were studied for the presence of inflammation, interstitial fibrosis and tubular atrophy. The measurements and observations were then assessed to estimate the biopsy outcomes provided by the institution's nephropathologists. In addition to the original full dataset, other two subsets were created to evaluate the quality of the acquisition process, in terms of acquisition location (longitudinal middle only) and intra-operator variability (acquisitions median).

GEE and OR statistical analyses were performed to evaluate the significance of each parameter individually, and logistic regression was used for combining the parameters into a single binary outcome probability for each of the biopsy scores.

Attenuation measurements showed strong significance (p-value < 0.05) for tubular atrophy with negative correlation (OR < 1), while velocity associated measurements showed mild significance (p-value < 0.10) for inflammation and fibrosis with positive correlation (OR > 1). Time after transplant (TAT) and Age showed strong statistical significance for interstitial fibrosis and tubular atrophy (independently and combined as IFTA) with a positive correlation, but showed no significance for inflammation, in agreement with the chronic nature of inflammation when compared to the more progressive and acute characteristics of fibrosis and tubular atrophy. The eGFR also showed significance with inflammation in some datasets, with negative correlation, and serum creatinine (SCr) showed strong significance for tubular atrophy for all datasets tested, with positive correlation, both in agreement with the declining kidney function cause by the injuries evaluated.

The logistic regression was tested on all three dataset and fitted to four different subsets of parameters for dimensionality reduction assessment: All measurements and observations, low p-value only (p-value < 0.1), elastography measurements only and clinical observations only. Overall, the longitudinal middle (L_MID) dataset performed the best during the logistic regression tests, with a peak tubular atrophy classification performance of AUROC = 0.72 when fit to low p-value parameters only. Inflammation and interstitial fibrosis also benefited the most form the L_MID/low p-value setting, with AUROCs of 0.63 and 0.66, respectively.

Logistic regression was not capable of producing good models from settings with larger number of dimensions such as elastography only. The non-multicollinearity assumptions and "curse of dimensionality" characteristics of this method might have confounded the performance models. Other classification approaches such as support

vector classifiers and dimensionality reduction techniques such as principal component analysis can be implemented to address the limitations of the method proposed.

The patient study evaluated in this work is still ongoing and larger study pools will be available in the future. At present, the dataset is predominantly composed by non-existent to mild injury, therefore, the groups differentiation may be minimal. The methods proposed can benefit considerably from larger cohorts of mild to severe chronic rejection and acute rejection.

The initial results are encouraging and in agreement with previous findings. The kidney's geometric complexity, heterogeneity, anisotropy makes SWE implementation challenging, larger and more diverse cohorts, as well as more developments to the acquisition and processing workflow will be necessary for successful clinical translation.

2.6 Appendix

Table 2-5 – Full dataset									
		Banff i		**Banff ci**		**Banff ct**		**IFTA**	
	Unit of Increment	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
c_g	0.5	1.0003 (0.9999, 1.0006)	0.1481	0.9998 (0.9997, 1.0000)	0.0848	0.9980 (0.9967, 0.9993)	0.0031	0.9999 (0.9997, 1.0000)	0.1274
μ	1000	1.0001 (0.9999, 1.0003)	0.2506	0.9999 (0.9999, 1.0000)	0.2182	0.9992 (0.9986, 0.9997)	0.0044	1.0000 (0.9999, 1.0000)	0.2846
μ_1	500	0.9999 (.9949, 1.0050)	0.9823	1.0001 (0.9999, 1.0004)	0.2629	1.0002 (0.9998, 1.0005)	0.3834	1.0001 (0.9999, 1.0003)	0.2814
μ_2	1	1.0 (0.9835, 1.0167)	0.9958	0.9999 (0.9993, 1.0004)	0.6195	1.0014 (0.9994, 1.0035)	0.1637	0.9999 (0.9993, 1.0004)	0.6274
c_{2DFT}^{100}	0.5	0.9967 (0.9664, 1.0279)	0.8328	1.0006 (0.9997, 1.0015)	0.1939	1.0022 (0.9994, 1.0049)	0.1205	1.0006 (0.9997, 1.0015)	0.1719
c_{2DFT}^{200}	0.5		0.0027	1.0001 (0.9994, 1.0008)	0.7138	0.9997 (0.9983, 1.0012)	0.7251	1.0001 (0.9994, 1.0008)	0.7768
α_{2DFT}^{100}	100	1.0022 (0.9701, 1.0353)	0.8962	1.0003 (0.9987, 1.0019)	0.7371	0.9942 (.9878, 1.0007)	0.0817	1.0002 (0.9986, 1.0017)	0.8364
α_{2DFT}^{200}	100	0.9654 (0.9419, 0.9895)	0.0051	1.0012 (0.9997, 1.0027)	0.1235	1.0032 (0.9964, 1.0100)	0.3538	1.0012 (0.9997, 1.0026)	0.1160
c_{SFK}^{100}	0.5	1.0338 (0.9884, 1.0814)	0.1472	1.0007 (0.9994, 1.0020)	0.2771	1.0000 (0.9978, 1.0022)	0.9966	1.0007 (0.9994, 1.0020)	0.2759
c_{SFK}^{150}	0.5	1.0266 (0.9947, 1.0596)	0.1034	1.0009 (0.9999, 1.0019)	0.0870	0.9999 (0.9983, 1.0014)	0.8594	1.0009 (0.9999, 1.0018)	0.0868
c_{SFK}^{200}	0.5	1.0170 (0.9973, 1.0371)	0.0910	1.0009 (1.0, 1.0018)	0.0424	1.0012 (0.9995, 1.0030)	0.1661	1.0009 (1.0001, 1.0018)	0.0367
c_{SFK}^{250}	0.5	1.0120 (0.9940, 1.0303)	0.1942	1.0008 (1.0, 1.0016)	0.0540	1.0015 (0.9995, 1.0035)	0.1445	1.0008 (1.0000, 1.0015)	0.0451
c_{SFK}^{300}	0.5	1.0140 (0.9957, 1.0327)	0.1351	1.0006 (1.0, 1.0013)	0.0675	1.0007 (0.9996, 1.0019)	0.2047	1.0006 (1.0000, 1.0013)	0.0588
Age	5	1.0579 (0.9216, 1.2142)	0.4240		0.0037	1.2177 (1.0899, 1.3604)	0.0005		0.0030
Sex	1	0.4846 (0.2195, 1.0699)	0.0730	0.6875 (0.4020, 1.1759)	0.1712	1.1967 (0.6756, 2.1198)	0.5382	0.7037 (0.4109, 1.2049)	0.2003
Kidney Depth	10	0.9999 (0.9995, 1.0003)	0.7343	1.0001 (0.9999, 1.0003)	0.2422	0.9992 (0.9972, 1.0011)	0.4071	1.0001 (0.9999, 1.0003)	0.4280
BP_S	5	1.0602 (0.9546, 1.1774)	0.2746			1.1286 (1.0232, 1.2448)	0.0156		
BP_D	5	1.2748 (1.0298, 1.5781)	0.0258	1.0157 (0.8872, 1.1628)	0.8214	1.0877 (0.9471, 1.2493)	0.2338	1.0212 (0.8918, 1.1695)	0.7611
RI_{UP}	0.1	1.1543 (0.7218, 1.8459)	0.5493		0.0978	1.3109 (0.9157, 1.8768)	0.1392		0.0041
RI_{MID}	0.1	0.8855 (0.5748, 1.3641)	0.5813		0.5304	1.3680 (0.9406, 1.9895)	0.1011		0.3573
RI_{LP}	0.1	1.1697 (0.7422, 1.8434)	0.4995	1.0502 (0.7423, 1.4858)	0.7819	1.0846 (0.7491, 1.5703)	0.6673	1.0295 (0.7269, 1.4581)	0.8699
TAT	12	0.8292 (0.5705, 1.2053)	0.3264	1.3580 (1.0673, 1.7279)	0.0128	1.6793 (1.1395, 2.4747)	0.0088	1.3491 (1.0606, 1.7160)	0.0147
BMI	5	1.1966 (0.7902, 1.8119)	0.3965		<0.0001	1.0960 (0.8399, 1.4303)	0.4995		<0.0001
SCr	0.5	1.2848 (0.8893, 1.8561)	0.1818		0.3249	1.5119 (1.0441, 2.1893)	0.0286		0.3317
eGFR	10	0.8741 (0.7402, 1.0323)	0.1129	0.9229 (0.8105, 1.0508)	0.2255	0.9102 (0.7913, 1.0470)	0.1878	0.9253 (0.8129, 1.0533)	0.2401

	p-value < 0.1 (Trend)
	p-value < 0.05 (Significant)
	Model did not converge

Table 2-6 – L_MID dataset									
		Banff i		**Banff ci**		**Banff ct**		**IFTA**	
	Unit of Increment	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
c_g	0.5	1.0 (1.0000, 1.0000)	0.0487	1.0000 (1.0000, 1.0000)	0.9950	1.0018 (0.9795, 1.0245)	0.9516	1.0 (1.0000, 1.0000)	0.9957
μ	1000	1.0 (1.0000, 1.0000)	0.1264	1.0000 (1.0000, 1.0000)	0.9969	0.9997 (0.9904, 1.0091)	0.9817	1.0 (1.0000, 1.0000)	0.7085
μ_1	500	1.0 (0.9997,1.0002)	0.9051	1.0 (1.0000, 1.0000)	0.9877	0.9976 (0.9939, 1.0014)	0.3508	1.0 (1.0000, 1.0000)	0.9841
μ_2	1	0.9999 (0.9992,1.0006)	0.7853	1.0 (1.0000, 1.0000)	0.9909	1.0067 (0.9959, 1.0176)	0.4515	1.0 (1.0000, 1.0000)	0.9862
c_{2DFT}^{100}	0.5	0.9995 (0.9983, 1.0006)	0.3728	1.0 (1.0000, 1.0000)	0.9997	1.0176 (0.9951, 1.0405)	0.1867	1.0 (1.0000, 1.0000)	0.9988
c_{2DFT}^{200}	0.5	0.9997 (0.9992, 1.0002)	0.2135	1.0 (1.0000, 1.0000)	0.9957	1.0050 (0.9991, 1.0109)	0.5999	1.0 (1.0000, 1.0000)	0.9933
α_{2DFT}^{100}	100	0.9995 (0.9983, 1.0007)	0.4245	1.0 (1.0000, 1.0000)	0.9997	0.9907 (0.9888, 0.9925)	<0.0001	1.0 (1.0000, 1.0000)	0.9985
α_{2DFT}^{200}	100			1.0 (1.0000, 1.0000)	0.9896	0.9860 (0.9832, 0.9888)	<0.0001	1.0 (1.0000, 1.0000)	0.9892
c_{SFK}^{100}	0.5	1.0005 (0.9988, 1.0022)	0.8333	1.0 (1.0000, 1.0000)	0.9915	1.0147 (0.9887, 1.0414)	0.4706	1.0 (1.0000, 1.0000)	0.9904
c_{SFK}^{150}	0.5	1.0 (0.9984, 1.0016)	0.9929	1.0 (1.0000, 1.0000)	0.9883	0.9949 (0.9872, 1.0028)	0.7647	1.0 (1.0000, 1.0000)	0.9845
c_{SFK}^{200}	0.5	1.0003 (0.9991, 1.0014)	0.8949	1.0 (1.0000, 1.0000)	0.9115	0.9876 (0.9683, 1.0073)	0.3954	1.0 (1.0000, 1.0000)	0.9999
c_{SFK}^{250}	0.5	1.0 (0.9991, 1.0010)	0.9805	1.0 (1.0000, 1.0000)	0.9964	0.9873 (0.9695, 1.0054)	0.1671	1.0 (1.0000, 1.0000)	0.9944
c_{SFK}^{300}	0.5	1.0001 (0.9992, 1.0009)	0.9652	1.0 (1.0000, 1.0000)	0.9992	0.9853 (0.9675, 1.0035	0.1842	1.0 (1.0000, 1.0000)	0.9978
Age	5	1.0626 (0.9174, 1.2309)	0.4005	1.1915 (1.0741, 1.3217)	0.0008	1.1780 (1.0643, 1.3039)	0.0016	1.1938 (1.0762, 1.3243)	0.0007
Sex	1	0.4354 (0.1960, 0.9671)	0.0344	0.6780 (0.3944, 1.1657)	0.1518	1.0917 (0.6104, 1.9527)	0.7450	0.6954 (0.4040, 1.1969)	0.1819
Kidney Depth	10	0.9998 (0.9994, 1.0001)	0.1898	0.9999 (0.9997, 1.0001)	0.9805	0.9040 (0.7989, 1.0229)	0.1650	0.9999 (0.9997, 1.0001)	0.9804
BP_S	5	1.0416 (0.9360, 1.1592)	0.4751	1.0680 (0.9858, 1.1571)	0.1100	1.1256 (1.0214, 1.2404)	0.0076	1.0744 (0.9906, 1.1654)	0.0855
BP_D	5	1.1845 (0.9491, 1.4783)	0.0955	1.0086 (0.8791, 1.1571)	0.9032	1.0765 (0.9365, 1.2375)	0.2938	1.0146 (0.8841, 1.1643)	0.8375
RI_{UP}	0.1	1.1661 (0.7282, 1.8674)	0.5285	1.3066 (0.9199, 1.8559)	0.1238	1.3022 (0.9151, 1.8530)	0.1182	1.3135 (0.9248, 1.8655)	0.1164
RI_{MID}	0.1	0.8955 (0.5745, 1.3958)	0.6576	1.1209 (0.8006, 1.5694)	0.5099	1.3579 (0.9368, 1.9683)	0.0784	1.1370 (0.8108, 1.5945)	0.4600
RI_{LP}	0.1	1.2346 (0.7590, 2.0083)	0.4166	1.0771 (0.7592, 1.5282)	0.6771	1.0863 (0.7519, 1.5693)	0.6376	1.0551 (0.7430, 1.4983)	0.7645
TAT	12	0.8193 (0.5575, 1.2041)	0.3080	1.3734 (1.0661, 1.7693)	0.0084	1.6623 (1.1217, 2.4637)	0.0012	1.0262 (1.0048, 1.0480)	0.0100
BMI	5	1.1125 (0.7365, 1.6803)	0.5542	1.0522 (0.8198, 1.3505)	0.6866	1.0886 (0.8332, 1.4222)	0.4973	1.0094 (0.9601, 1.0613)	0.7106
SCr	0.5	1.3757 (0.9589, 1.9737)	0.1159	1.1777 (0.8676, 1.5987)	0.2942	1.5216 (1.0524, 2.2001)	0.0178	1.3769 (0.7478, 2.5352)	0.3015
eGFR	10	0.8452 (0.7212, 0.9905)	0.0377	0.9131 (0.7997, 1.0425)	0.1656	0.9104 (0.7868, 1.0534)	0.1626	0.9913 (0.9783, 1.0045)	0.1838

	Trend
	p less than 0.05 (Significant)
	Model did not converge

Table 2-7 – Median dataset									
		Banff i		**Banff ci**		**Banff ct**		**IFTA**	
	Unit of Increment	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
c_g	0.5	1.0007 (0.9990, 1.0023)	0.4191	1.0000 (1.0000, 1.0000)	0.0767	1.0000 (1.0000, 1.0000)	0.6452	1.0000 (1.0000, 1.0000)	0.1015
μ	1000	1.0002 (0.9995, 1.0010)	0.5518	1.0000 (1.0000, 1.0000)	0.1143	1.0000 (1.0000, 1.0000)	0.6974	1.0000 (1.0000, 1.0000)	0.1458
μ_1	500	0.9997 (0.9994, 0.9999)	0.0101	1.0000 (1.0000, 1.0000)	0.8528	1.0000 (1.0000, 1.0000)	0.5420	1.0000 (1.0000, 1.0000)	0.8107
μ_2	1	0.9995 (0.9990, 1.0001)	0.1161	1.0000 (1.0000, 1.0000)	0.9050	1.0000 (1.0000, 1.0000)	0.4312	1.0000 (1.0000, 1.0000)	0.9612
c_{2DFT}^{100}	0.5	0.9993 (0.9982, 1.0005)	0.2491	1.0000 (1.0000, 1.0000)	0.2392	1.0000 (1.0000, 1.0000)	0.5349	1.0000 (1.0000, 1.0000)	0.2311
c_{2DFT}^{200}	0.5	0.9998 (0.9990, 1.0006)	0.6005	1.0000 (1.0000, 1.0000)	0.6592	1.0000 (1.0000, 1.0000)	0.3227	1.0000 (1.0000, 1.0000)	0.6029
α_{2DFT}^{100}	100	1.0002 (0.9975, 1.0030)	0.8792	1.0000 (1.0000, 1.0000)	0.2316	1.0000 (1.0000, 1.0000)	0.0309	1.0000 (1.0000, 1.0000)	0.2205
α_{2DFT}^{200}	100			1.0000 (1.0000, 1.0000)	0.0305	1.0000 (1.0000, 1.0000)	0.0131	1.0000 (1.0000, 1.0000)	0.0269
c_{SFK}^{100}	0.5	0.9992 (0.9978, 1.0006)	0.2602	1.0000 (1.0000, 1.0000)	0.1648	1.0000 (1.0000, 1.0000)	0.9331	1.0000 (1.0000, 1.0000)	0.1657
c_{SFK}^{150}	0.5	1.0002 (0.9990, 1.0014)	0.7439	1.0000 (1.0000, 1.0000)	0.6361	1.0000 (1.0000, 1.0000)	0.4037	1.0000 (1.0000, 1.0000)	0.6540
c_{SFK}^{200}	0.5	1.0002 (0.9991, 1.0013)	0.7539	1.0000 (1.0000, 1.0000)	0.5550	1.0000 (1.0000, 1.0000)	0.3566	1.0000 (1.0000, 1.0000)	0.5825
c_{SFK}^{250}	0.5	1.0000 (0.9990, 1.0010)	0.9705	1.0000 (1.0000, 1.0000)	0.8554	1.0000 (1.0000, 1.0000)	0.2963	1.0000 (1.0000, 1.0000)	0.8489
c_{SFK}^{300}	0.5	1.0000 (0.9991, 1.0009)	0.9515	1.0000 (1.0000, 1.0000)	0.8652	1.0000 (1.0000, 1.0000)	0.1840	1.0000 (1.0000, 1.0000)	0.9186
Age	5	1.0621 (0.9248, 1.2198)	0.3936	1.1913 (1.0747, 1.3205)	0.0009	1.1808 (1.0668, 1.3070)	0.0013	1.1935 (1.0767, 1.3230)	0.0008
Sex	1	0.4148 (0.1904, 0.9036)	0.0268	0.6742 (0.3944, 1.1524)	0.1495	1.1461 (0.6491, 2.0236)	0.6382	0.6917 (0.4042, 1.1837)	0.1787
Kidney Depth	10	0.9999 (0.9977, 1.0022)	0.9582	1.0000 (1.0000, 1.0000)	0.4361	1.0000 (1.0000, 1.0000)	0.5160	1.0000 (1.0000, 1.0000)	0.4474
BP_S	5	1.0494 (0.9457, 1.1645)	0.3636	1.0715 (0.9895, 1.1602)	0.0892	1.1256 (1.0213, 1.2406)	0.0171	1.0778 (0.9943, 1.1684)	0.0688
BP_D	5	1.1921 (0.9591, 1.4817)	0.1134	1.0070 (0.8796, 1.1528)	0.9193	1.0904 (0.9492, 1.2526)	0.2211	1.0127 (0.8844, 1.1596)	0.8551
RI_{UP}	0.1	1.2084 (0.7633, 1.9128)	0.4193	1.3010 (0.9170, 1.8457)	0.1404	1.2904 (0.9027, 1.8447)	0.1620	1.3076 (0.9217, 1.8549)	0.1328
RI_{MID}	0.1	0.9374 (0.6149, 1.4291)	0.7640	1.1221 (0.8021, 1.5699)	0.5011	1.3420 (0.9216, 1.9543)	0.1250	1.1379 (0.8120, 1.5946)	0.4529
RI_{LP}	0.1	1.2292 (0.7763, 1.9463)	0.3788	1.0581 (0.7474, 1.4980)	0.7502	1.0643 (0.7343, 1.5426)	0.7422	1.0363 (0.7313, 1.4686)	0.8411
TAT	12	0.8169 (0.5734, 1.1639)	0.2628	1.3564 (1.0658, 1.7263)	0.0132	1.6775 (1.1389, 2.4708)	0.0088	1.3473 (1.0590, 1.7140)	0.0152
BMI	5	1.1975 (0.8041, 1.7834)	0.3752	1.0818 (0.8460, 1.3831)	0.5309	1.0764 (0.8270, 1.4010)	0.5841	1.0779 (0.8421, 1.3796)	0.5514
SCr	0.5	1.1998 (0.8285, 1.7374)	0.3349	1.1350 (0.8383, 1.5367)	0.4126	1.4924 (1.0345, 2.1531)	0.0322	1.1316 (0.8361, 1.5314)	0.4233
eGFR	10	0.8557 (0.7267, 1.0077)	0.0617	0.9203 (0.8085, 1.0476)	0.2089	0.9127 (0.7941, 1.0490)	0.1982	0.9235 (0.8115, 1.0509)	0.2275

	p-value < 0.1 (Trend)
	p-value < 0.05 (Significant)
	Model did not converge

Chapter 3 – Interpretable machine learning models for kidney biopsy score prediction based on shear wave elastography

3.1 Introduction

Chronic kidney disease (CKD), also called chronic kidney failure, describes a long-term gradual decrease in renal function. In its final stages, CKD can progress to kidney failure, which is treated by hemodialysis or kidney transplant. Patient survival rates with transplant have increased over the past decades, but patients still often outlive the allografts due to rejection, what might require hemodialysis and new transplants [40]. Renal biopsy is the gold standard for diagnosis of kidney health, but is an invasive procedure that can cause complications, and so, cannot be used frequently. Serum creatinine (SCr), glomerular filtration rate (GFR), and classical medical imaging are generally used for noninvasive assessment of kidney transplant rejection [96].

Creatinine is a waste product of muscle metabolism, and, therefore, is filtered freely by the kidneys with minimal reabsorption. The presence of serum creatinine in abnormal levels is directly related to the filtration capacity, and, consequently, an important biomarker for kidney insufficiency. In spite of GFR direct measurement complexity, it is still considered one of the main surrogates of kidney function. Indirect methods using SCr, serum cystatin C, age and sex were developed to provide estimated GFR (eGFR) with standard blood tests [97].

The renal blood flow can also be evaluated using spectral Doppler ultrasound at the arcuate or interlobar arteries for resistive index (RI) measurement. The RI assesses the pulsatility and compliance of the arteries evaluated, and so, holds more correlation with current systemic vascular alterations, nevertheless, it is an important parameter for kidney health prognosis [14].

Renal biopsy is an important tool for kidney rejection diagnosis and staging, as the samples are evaluated for the presence of multiple histological alterations, such as fibrosis and inflammation. When combined, fibrosis and inflammation have shown predictive performance for establishing allograft accurate prognosis at 1 year protocol biopsies [46], [98], [99]. Nephropathologists evaluate the tissue sample to score the injury levels

following the standardized Banff classification criteria [46], [100]. These scores are used to establish rejection prognosis and treatment.

Even though routine biopsies are the gold standard for allograft prognosis, they are performed sparsely during the allograft's life. Mayo Clinic's post-transplant protocol dictates that routine biopsies should be performed at 4 months, 1, 2, 4, 7 and 10 years after transplant. Renal biopsies are invasive procedures that are associated with complications such as arteriovenous fistulas, bleeding and, in rare cases, death [101]–[103]. The reported complication rate can vary from 0.1 to 6%, with a tendency of events to increase due to the increase of transplants performed and, consequently, biopsy employment [103]–[105].

Elasticity imaging methods have been studied over the last decade to determine their pathology discriminative capabilities as an alternative to invasive procedures like biopsy. Different physiological and structural changes induced by pathological processes can be identified as rheological alterations found during elastographic evaluations. Shear wave elastography is an elastography method that uses acoustic radiation force (ARF) to generate shear waves within the tissue under study. The propagation of the shear wave is then measured using high frame rate ultrasound imaging and the motion is processed to retrieve characteristics such as the overall propagation velocity (group velocity) as well as the velocity at specific wave frequencies (phase velocity). Processes such as fibrosis, and different biomarkers, such as RI and eGFR, have been found to correlate with shear wave velocity alterations [16], [19], [62]–[64].

Human tissue exhibits viscoelastic mechanical properties; therefore, it is important to evaluate these parameters in addition to the elastic properties. The time-dependent deformation of viscoelastic materials give rise to shear wave velocity dispersion, which is defined by the variation of shear wave velocity with respect to the wave frequency. The phase velocity is commonly obtained with a method that uses a two-dimensional Fourier transform (2D FT) [82]. By analyzing the phase velocity at multiple frequencies, it is possible to quantitively evaluate the viscoelastic properties of a tissue [29], [31], [65]. The quantification of viscoelastic parameters can be derived from rheological models, such as Kelvin-Voigt, or through the calculation of propagation characteristics such as phase velocity and phase attenuation [78], [85].

In previous work including that shown in Chapter 2, it was possible to correlate different elastography measurements and biomarkers to the incidence of interstitial inflammation (Banff i), interstitial fibrosis (Banff ci) and tubular atrophy (Banff ct) and

interstitial fibrosis and tubular atrophy (IFTA) during biopsy assessment [46], [100]. Each Banff category (i, ci, ct) ranges from 0-4, with 0 indicating no presence and the other values indicating increasing amounts of inflammation, interstitial fibrosis, or tubular atrophy. IFTA can be calculated by the floor average of ci and ct, e.g. if ci = 0 and ct = 1 then IFTA = 0 [98], [100]. In this study, we distinguished between scores of 0 and > 0.

The increasing complexity of modern dataset structures have given rise to new demands for techniques capable to model such applications. In recent years, the computer science field of machine learning (ML) has developed to efficiently perform tasks such as clustering, prediction, and regression by modeling such complex datasets that conventional linear and statistical techniques could not perform. Most ML approaches embed nonlinearities, and therefore, can represent the intrinsic nonlinearities of real datasets.

One of the most effective and versatile algorithms is the support vector machine (SVM). The SVM has been successfully applied to solve various complex tasks, such as image classification, text categorization and protein classification without the data volume constraints present in neural network implementations [106]–[109]. The objective of the SVM is to minimize classification error while optimizing the separating hyperplanes between the different data classes. The optimal hyperplane is defined by the fine adjustment of its margins, the tighter the margins the better the fitting to the training data, while looser margins will yield better generalization. The penalty for incorrect classifications is defined by the hinge loss function, increasing linearly with the distance from the data point to the separating hyperplane. The margins' slackness is then defined by the penalty parameter C applied to the hinge loss function. As C increases, the loss function slope increases and, consequently, the misclassification loss increases, enforcing better class separation with smaller margins. As with most ML algorithms, SVM results depend on proper balance between accuracy and generalization. Higher values of C can lead to high training accuracy, but poor test performance in the real world, which is referred to as overfitting [108].

In complex tasks, the linear hyperplane may not be able to successively separate the designated classes, in such cases, hyperspace transformations can be employed to achieve geometrical separation. The "kernel trick" embeds nonlinearities that transform the data into a higher dimension hyperspace where a linear hyperplane can be fit. SVM kernel development is a research field on its own, and different kernels can model different dataset characteristics [110]. In this work, the radial basis function kernel (RBF) was used

due to its flexibility to model complex datasets. The main RBF parameter, γ, defines the influence radius of a single observation to the calculation of the separating hyperplane.

In recent years, there has been an increasing concern not only on the quality of ML models, but also in their interpretability. The “black box” characteristics of most ML implementations are not suitable for critical applications such as health care, and directly impact the confidence on such models. Some ML algorithms, such as Decision Trees, have direct access to feature importance, others have feature importance surrogates, such as convolutional neural network’s feature maps, that can give insights on what the algorithm is converging for. Such solutions may not be available for SVM when a kernel transformation is applied to the feature space. In order to introduce interpretability to the RBF SVM, Shapley additive explanations (SHAP) [111] were implemented.

SHAP leverages the game theory solution concept of Shapley values to iteratively compute the feature importance based solely on the input and output of the trained model, therefore, this approach is mostly agnostic to the algorithm implemented. The Shapley value was introduced in 1951 by Lloyd Shapley and consists of a solution to fairly distribute the total gains to the players that collaborate towards a particular task, according to each player’s efforts [112]. According to the Shapley value, the player i worth contribution to a given collaborative game result is:

$$\varphi_i(v) = \sum_{S \in N/\{i\}} \frac{|S|!(n-|S|-1)!}{n!} (v(S \cup \{i\}) - v(S)), \quad (3\text{-}1)$$

where v is the characteristic function, which describes the total expected reward for a given coalition of players S. The variable n is the total number of players, N is the coalition of all players, and the sum extends over all coalitions S that do not contain player i. Each individual player is evaluated and added to the coalition at a time, therefore, the average of the contributions over all possible permutations is interpreted as the fair compensation for each particular player [112]. It is then possible to translate the game theory to the interpretability of ML models, where the features are the players, and the model output is the game reward. By iteratively evaluating how the model output reacts to feature variations it is possible to assign Shapley values to each feature, not only for the whole model, but for each data point individually.

In the previous chapter, the renal transplant dataset, composed by 206 patients, was analyzed using general estimating equations (GEE) to obtain the statistical significance (p-values) of each of the features, elastography based and biomarkers. Although, multiple high statistically significant elastography features were found, logistic

regression was not able to produce relevant odds ratios, when evaluating features separately ($0.965 \leq OR \leq 1.017$), or areas under the receiver operating characteristic (ROC) curve above 0.63 for inflammation, 0.66 for fibrosis and IFTA, and 0.72 for tubular atrophy. In this study we intend to further develop and interpret the clinical application of SWE for kidney allograft evaluation with the use of support vector machine and SHAP.

3.2 Materials and Methods

The data set used for this study was composed from 206 patients with kidney transplants that had SWE scans performed with a General Electric Logiq E9 ultrasound system (General Electric, Wauwatosa, WI, USA), right before their routine biopsy procedure at Mayo Clinic. It is important to note that some patients were scanned more than once during the course of the protocol, totaling 223 studies. The data acquisition was conducted from October 2017 to July 2019, under the protocol approved by the Mayo Clinic Institutional Review Board (IRB# 11-003249). The study cohort demographics are displayed in Table 3-1.

Table 3-1 – Study cohort demographics

	Patients (n=206)		Studies (n=223)	
	Number	%	Number	%
Sex				
Male	122	59.2	128	57.4
Female	84	40.8	95	42.6
	Mean (Std. Dev)	Range	Mean (Std. Dev)	Range
Age, yrs	49.8 (14.0)*	22-76	49.7 (14.0)	22-76
BMI, kg/m^2	27.6 (5.3)*	13.1-42.3	27.9 (5.5)	13.1-42.3
TAT (months)	14.5 (14.9)*	4-60	16.1 (14.7)	4-60
*At first scan				

Biopsy protocols for kidney rejection assessment rely on the diffusiveness of injury throughout the allograft tissue to correlate local samples to the whole kidney state [32]. Although the same premise can be applied to shear wave elastography, non-invasive techniques allow more comprehensive evaluation of the allograft. Therefore, each patient was scanned with 10-13 acquisitions at three kidney locations (upper pole, middle region, and lower pole) on both longitudinal and transversal views, generating an average of 60-80 measurements per patient.

The SVM was evaluated to assess its capabilities to predict the Banff i, ci, ct and IFTA scores from the rheological, physiological, and demographical features. The parameters obtained from the SWE measurements included group velocity, shear modulus, phase

velocity and attenuation at frequencies of 100 and 200 Hz. The Stockwell transform combined with a slant frequency-wavenumber analysis (GST-SFK) [81] was also used to retrieve the phase velocity dispersion curves at a wider frequency range (100-300 Hz). Additionally, the models also used clinical observation features such as blood pressure (systolic and diastolic), resistive index, serum creatinine and eGFR. The complete list of features and labels used are displayed in Table 3-2.

Table 3-2 – SWE measurements and clinical observations

Indicator	**Description**	**Indicator**	**Description**
Features:		Kidney Depth	Cortex depth on ultrasound image, mm
Location	Scan position		
c_g	Group velocity, m/s	BP_S	Systolic blood pressure, mmHg
μ	Shear modulus, Pa	BP_D	Diastolic blood pressure, mmHg
μ_1	Kelvin-Voigt shear elasticity μ_1, Pa	RI_{UP}	Restive index at upper pole
μ_2	Kelvin-Voigt shear viscosity μ_2, Pa·s	RI_{MID}	Restive index at middle
c_{2DFT}^{100}	2D FT phase velocity at 100 Hz, m/s	RI_{LP}	Restive index at lower pole
c_{2DFT}^{200}	2D FT phase velocity at 200 Hz, m/s	TAT	Time-after-transplant, months
α_{2DFT}^{100}	2D FT phase attenuation at 100 Hz, Np/m	BMI	Body mass index, kg/m^2
α_{2DFT}^{200}	2D FT phase attenuation at 200 Hz, Np/m	SCr	Serum creatinine, mg/dL
c_{SFK}^{100}	GST-SFK phase velocity at 100 Hz, m/s	eGFR	Estimated glomerular filtration rate, mL/min/1.73 m^2
c_{SFK}^{150}	GST-SFK phase velocity at 150 Hz, m/s	Labels:	
c_{SFK}^{200}	GST-SFK phase velocity at 200 Hz, m/s	Banff i	Interstitial inflammation score
c_{SFK}^{250}	GST-SFK phase velocity at 250 Hz, m/s	Banff ci	Interstitial fibrosis score
c_{SFK}^{300}	GST-SFK phase velocity at 300 Hz, m/s	Banff ct	Tubular atrophy score
Age	Patient age, years	IFTA	Interstitial fibrosis and tubular atrophy (ci + ct)

The models were implemented in Python (Python Software Foundation, Wilmington, DE, USA) and scikit-learn libraries [113]. All the features were normalized prior to modeling. The hyperparameters and other components, such as penalty C, γ and kernels, were evaluated using nested 5-fold cross validation and the final model was tested using 5-fold cross validation. All k-fold cross-validations were implemented with group shuffle split algorithm, not allowing different scans from a same study to be present at train and test sets at the same time.

After hyperparameter grid search, 'scale' showed overall better prediction results when compared to 'auto', and fixed γ settings. The scikit-learn algorithm defines the 'scale' γ by the equation $\gamma = 1/(n_{feat} * X.var)$, therefore, the influence radius of a single observation is inversely proportional to the number of features (n_{feat}) and the training data variance ($X.var$). The algorithm showed better performance with C ranging from 0.1 to

10, and optimal tolerances within 0.001 and 0.01. The class weight parameter was set to 'balanced' to help accommodate for small skewness present on datasets. The positive Banff diagnostic outcomes (greater than zero) were combined to create binary labels. The Banff diagnostic outcome distribution per patient and per scan is shown in Table 3-3 and 3-4.

Table 3-3 – Banff patient score distribution

Score	Banff i (n=223)		Banff ci (n=223)		Banff ct (n=223)		IFTA (n=223)	
	Patients	%	Patients	%	Patients	%	Patients	%
0	192	86.10	127	56.95	72	32.29	127	56.95
1	17	7.62	84	37.67	139	62.33	84	37.67
2	10	4.48	11	4.93	11	4.93	11	4.93
3	4	1.79	1	0.45	1	0.45	1	0.45
> 0	31	13.90	96	43.05	151	67.71	96	43.05

Table 3-4 – Banff scan score distribution

Score	Scans	%	Scans	%	Scans	%	Scans	%
0	14024	86.26	9352	57.52	5478	33.69	9352	57.52
1	1160	7.13	6033	37.11	9907	60.94	6033	37.11
2	767	4.72	796	4.90	796	4.90	796	4.90
3	307	1.89	77	0.47	77	0.47	77	0.47
> 0	2234	13.74	6906	42.48	10780	66.31	6906	42.48

For each Banff binary diagnostic outcome (i, ci, ct, IFTA), several models were trained on a variety of dataset and feature setups. Primarily, three datasets were used: one with all the data available for training (Full), one limited to the longitudinal plane, middle position (L_MID) and one with the median values for all viscoelastic measurements throughout the 10-13 acquisitions and any given position (Median). Four feature sets were used to train the models: all features available, elastography only, biomarkers only and low p-value features ($p < 0.10$, as given in Appendix Table 3-5). To evaluate the SWE assessment variability, the IQR-to-median ratio for each patient's position was calculated throughout every elastography measurement [114]. This information was evaluated as a data quality parameter, by filtering positions within the kidney (longitudinal/transverse, upper/lower pole, middle region) with IQR-to-median ratio > 0.4. In cases where not all viscoelastic features were being used, IQR-to-median ratio filtering was tested on all features (extensive), regardless of use, and only on the features used for training (specific, e.g., low p-value). In the case of Banff i, specifically, due to its skewness, IQR-to-median ratio based undersampling was also applied to balance the majority class. The healthy patients with lower average IQR-to-median ratio were selected to compose the balanced Banff i dataset (n = 62).

The receiver operating characteristic (ROC) curves from the five folds were compiled to produce a single average ROC curve, additionally, the area under ROC (AUROC), sensitivity, specificity, and accuracy (at Youden's index optimal cutoff) were recorded. All ROC averages were within a standard deviation threshold of 0.10 to ensure model stability. The ROC standard deviation is important to ensure each of the k-folds are fair representations of the dataset and, therefore, certify the mean ROC as a trustworthy representation of the model's real performance. For each fold, the Shapley values of each test observation were recorded. In addition to the overall importance, it is possible to also identify the feature effect trend, that is, the behavior of the of the output given the increment or reduction of a given feature. The SHAP typical output is exemplified in Figure 3-1 [111].

Table 3-5 – Low p-value features (p-value < 0.10)

Diagnostic Outcome	Full	L_MID	Median
Banff i	c_{2DFT}^{200} α_{2DFT}^{200} c_{SFK}^{200} BP_D	c_g BP_D eGFR	μ_1 eGFR
Banff ci	c_g c_{SFK}^{150} c_{SFK}^{200} c_{SFK}^{250} c_{SFK}^{250} Age RI_{UP} TAT BMI	Age TAT	c_g α_{2DFT}^{200} Age BP_S TAT
Banff ct	c_g μ α_{2DFT}^{100} Age BP_S TAT SCr	α_{2DFT}^{100} α_{2DFT}^{200} Age BP_S RI_{MID} TAT SCr	α_{2DFT}^{100} α_{2DFT}^{200} Age BP_S TAT SCr
IFTA	c_{SFK}^{150} c_{SFK}^{200} c_{SFK}^{250} Age RI_{UP} TAT BMI	Age BP_S TAT	α_{2DFT}^{200} Age BP_S TAT

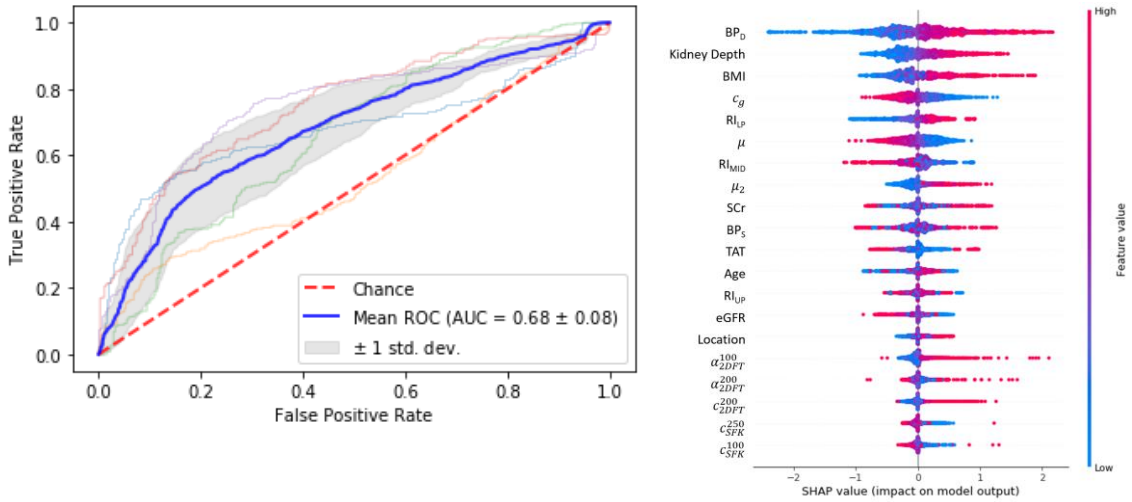


Figure 3-1 – Example of ROC for 5-fold cross-validation test and its SHAP evaluations of a SVM classifier trained with the entire dataset, all features available, no IQR-to-median ratio filtering and undersampling. Each dot represents one evaluation, the color encodes the feature value, and the horizontal position its impact to the model output. In this example, it is possible observe diastolic blood pressure has positive importance, as it increases so does the output of the model. *Only top 20 features are displayed.

3.3 Results

The AUROC, sensitivity, specificity, and accuracy for each model was recorded. For the sake of conciseness, the analysis will be focused on the AUROC. The values of sensitivity, specificity, and accuracy are available in the Appendix of this chapter. Figure 3-2 shows the AUROC performance distribution for each dataset variation proposed, dataset (Full, L_MID and Median), feature set (All, Elastography, Biomarkers, low p-value), IQR-to-median ratio filtering (no, specific, and general).

Banff ci, ct and, consequently IFTA, had similar performance distributions, in contrast with inflammation (Banff i). Discrimination of interstitial fibrosis and tubular atrophy show better performance with the L_MID dataset and leveraging all features. The elastography feature set showed poor performance for all three diagnostic outcomes, with AUROCs close to the 0.50 mark. The use of all features consistently showed best median performance with AUROCs in the 0.62-0.70 range. Although the p-value feature set was the only to reach the 0.70 mark, IQR-to-median ratio filtering showed better median performance for Banff ci and ct, but worst median performance for IFTA.

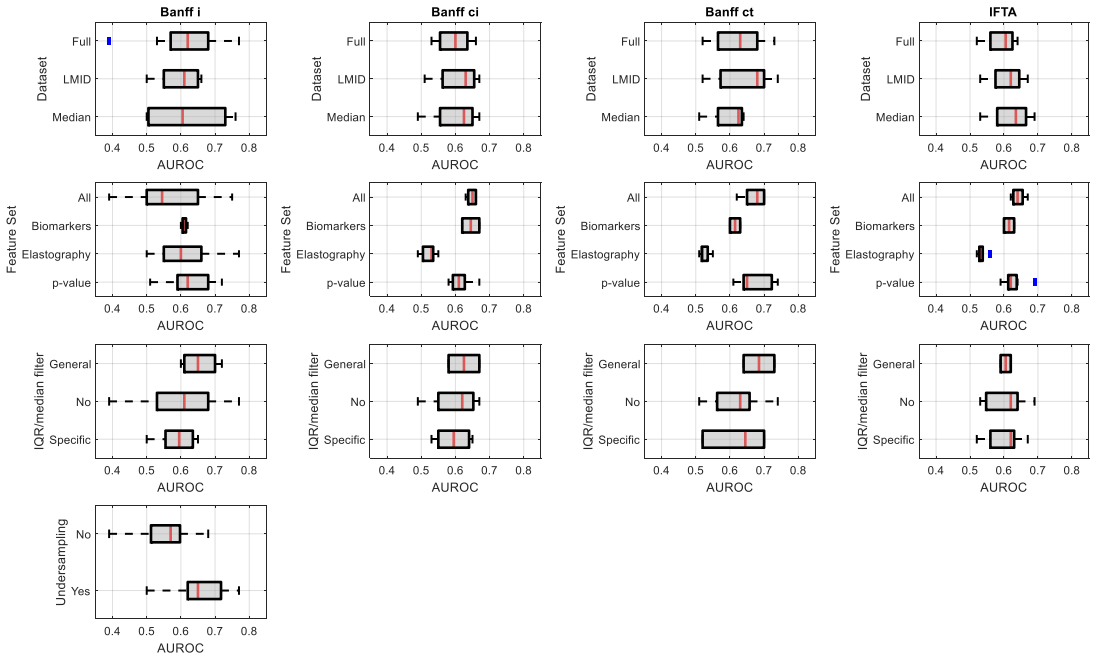


Figure 3-2 – AUROC distribution for every diagnostic category aggregated by dataset, feature set, IQR-to-median ratio filtering and undersampling (Banff i only). The red line is the median value, the grey box displays the interquartile range (IQR) and the maximum and minimum whiskers set to 1.5 IQR. Outliers are displayed as a blue vertical line. The list of all results is available in the chapter Appendix.

Table 3-6 – Top 3 SVM models for each diagnostic outcome

Diagnostic Outcome	Rank	Dataset	Feature set	IQR-to-median ratio Filter	Undersampling	AUROC
Banff i	1st	Median	Elastography	No	Yes	0.76
	2nd	Median	All	No	Yes	0.75
	3rd	Full	p-value	No	Yes	0.72
Banff ci	1st	L_MID	p-value	General	N/A	0.67
	2nd	Median	Biomarkers	No	N/A	0.67
	3rd	Full	All	No	N/A	0.66
Banff ct	1st	L_MID	p-value	No	N/A	0.74
	2nd	Full	p-value	General	N/A	0.73
	3rd	L_MID	All	No	N/A	0.70
IFTA	1st	Median	p-value	No	N/A	0.69
	2nd	L_MID	All	Specific	N/A	0.67
	3rd	L_MID	All	No	N/A	0.65

Banff i, in contrast, showed the best median performance when trained with the full dataset, low p-value features, but with IQR-to-median ratio applied to all features, applying the IQR-to-median ratio only to the features at hand decreased variability but did not improve the overall performance. Banff i greatly benefited from undersampling with an increase in median of 14%.

Table 3-6 shows the three best performing models for each diagnostic outcome. Banff i and Banff ct had models that were able to reach AUROC > 0.70, with 0.76 and 0.74, respectively. It is important to note that, in some cases, the best overall performance does not align with the median trends shown in Figure 3-2. Banff ci showed the worst overall performance, with a maximum of AUROC = 0.67. IFTA showed slightly better performance with a top AUROC = 0.69. A table with all models tested and their performances is supplied in the chapter Appendix.

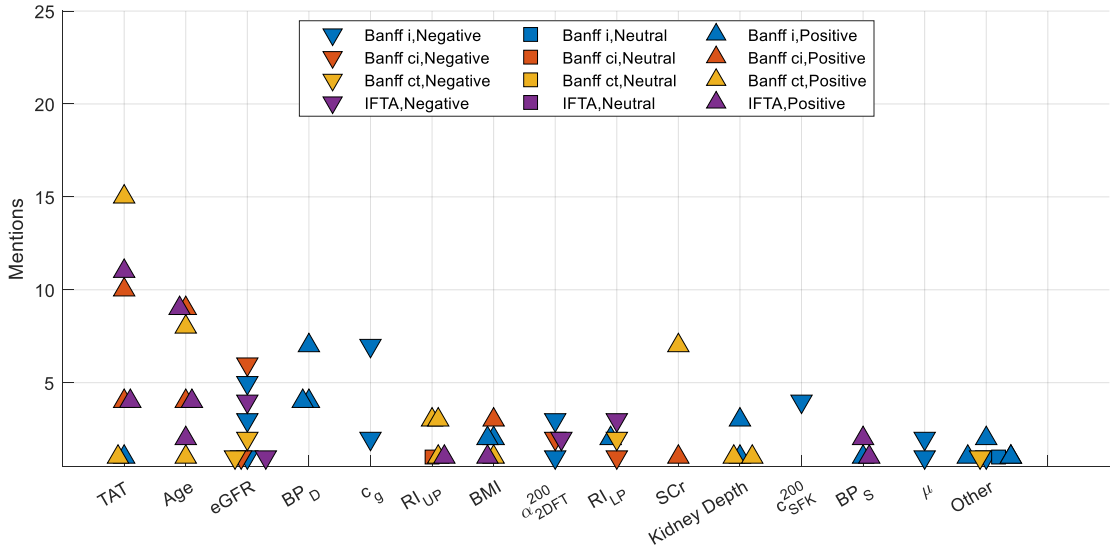


Figure 3-3 – Scatter plot of feature mentions at 1st, 2nd, and 3rd importance rank. The marker color encodes which diagnostic outcome was analyzed, and the marker shape encodes the trend (e.g., upright triangle indicates positive trend). Positive trends occur when the output of the classifier increases as the feature value increase, whereas negative trend occurs when the classifier output increases when the feature value decreases.

Figure 3-3 summarizes the top 3 importance mentions for each model tested with AUROC > 0.60 as calculated by the SHAP algorithm. Any models below AUROC = 0.60 were deemed too close to chance (AUROC = 0.50), and the SHAP analysis should not hold significance. The SHAP analysis showed the strong importance of the time dependent biomarkers Age and TAT for Banff ci, ct and consequently IFTA. These features consistently showed positive trends, and therefore, the probability of positive prediction increases as Age and TAT increases. It is important to note that are only 14 longitudinal studies are present in the data set, and therefore TAT and Age are not directly correlated. The eGFR showed strong importance for all diagnostic outcomes analyzed, with a negative trend, as the eGFR decreases the probability of positive prediction increases. Differently from other diagnostic outcomes, Banff i, showed strong diastolic

blood pressure importance with this biomarker as top importance for 16 models, always with a positive behavior.

From the elastography features, group velocity (c_g) and 2D FT phase attenuation at 200 Hz (α_{2DFT}^{200}) showed higher importance with 18 and 17 mentions, respectively. Overall, Banff i was the diagnostic prediction most impacted by viscoelastic alterations, although Banff ci, ct and IFTA had better results when viscoelastic features were combined with the biomarkers.

3.4 Discussion

In this study, we evaluated and interpreted the implementation of support vector machines for classification of kidney allograft's biopsy scores based on viscoelastic, physiological, and demographical observations. Group velocity (c_g), shear modulus (μ), Kelvin-Voigt dispersion curve fitting (μ_1 and μ_2), AMUSE (2D FT phase velocity and attenuation, c_{2DFT} and α_{2DFT}) and GST-SFK (phase velocity, c_{SFK}) were used to assess renal cortex rheological properties. Kidney depth, blood pressure (systolic and diastolic), resistive index (at upper pole, middle region, and lower pole), time-after-transplant (TAT), Age, BMI, serum creatinine and eGFR were used to assess the performance of common bedside biomarkers. The SVMs were trained and tested on different subsets of patient data, the original full dataset, one containing the L_MID position only, and median dataset, calculated over all scans performed at each specific patient and position. Different feature sets were used to test viscoelastic parameters and biomarkers in isolation along with the subset of features with p-value < 0.10, as evaluated in Chapter 2. The IQR-to-median ratio was also tested as a data quality filter by calculation data acquisition variability over each viscoelastic feature, position, and patient.

Several models reached AUROCs > 0.70, a level of performance compatible with previous work performed by our group and others [17], [21], [22], [33]–[36]. The results were encouraging but still lack the performance necessary for clinical translation. Nevertheless, it was possible to evaluate different characteristic of the allograft injuries as related to the most common biomarkers and experimental viscoelastic measurements. The SHAP importance analysis agreed with the generalized estimating equations (GEE) and odds ratio analysis performed previously (Chapter 2), showing the reliability of SVM for model classification and interpretability.

Overall, inflammation was shown to be the most detectable injury by elastography measurements, demonstrating classification AUROC > 0.70 even when biomarkers were not present. Inflammation may tend to have a transient occurrence, with tubular atrophy and interstitial fibrosis being a consequence of recurring inflammation [120]. It is possible to note how temporal biomarkers, TAT and Age, do not give insight into the inflammation prediction. Diastolic blood pressure also showed high importance for inflammation detection, with a positive trend, suggesting that hypertensive patients might be more associated with having transplant inflammation. Group velocity also had reasonably high importance, with 9 mentions in the top 3 ranking. Its trend was negative, which might be due to the increase of non-scarred interstitial volume with infiltrate of plasma cells and edema between tubules at the area affected [100]. Limiting the features to the most significant, as analyzed by GEE, prove to be the most reliable technique, although the best performance was reached using all the viscoelastic measurements and no biomarkers. Best reliability as reached with the use of a general IQR-to-median ratio filter, by filtering subject positions where any viscoelastic measurement had IQR-to-median ratio is above 0.40.

Regardless, the use of the median dataset proves to be a viable solution for reducing data variability and improving quality. Even when combined with no IQR-to-median ratio filtering it was able to produce the best Banff i models. Banff i was also the least frequent injury present in the patient study, with less then 14% prevalence. Therefore, the models significantly benefited from majority class undersampling, the use of the IQR-to-median ratio to retain the most reliable negative scores also prove to be a viable technique for better models. It is important to note that, in order to clinically translate such models, IQR-to-median ratio calculations should be embedded to the acquisition process, to ensure data quality similar to how it has been done with liver SWE [114]. Further investigation on inflammation prediction is necessary, which would involve studies with more patients with moderate and severe allograft inflammation are imperative for the advancement of the proposed techniques.

Banff ci, ct and IFTA, had very similar behavior in terms of performance distribution for dataset, feature set and IQR-to-median ratio filter. The L_MID dataset showed the best reliability for all three of the diagnostic outcomes, possibly due to a more diffuse characteristic of the interstitial fibrosis and tubular atrophy. As discussed previously, L_MID is the best quality acquisition position with lower average IQR-to-median ratio, and therefore it might provide reasonable evaluations without relying on other orientations or

kidney locations. For Banff ci and ct, the L_MID dataset also showed the best overall performance. Discrimination of interstitial fibrosis and tubular atrophy relied heavily on biomarkers for accurate classification. Time associated biomarkers such as TAT and Age showed high importance with a positive trend, this agrees with the progressive characteristics of the injury. In this study, the number of longitudinal patient acquisitions was limited, with only 14 patients (6.8%) being scanned at multiple timepoints, therefore, the are no strong correlation between TAT and Age. The Age positive trend importance then suggests a higher prevalence of ci, ct and IFTA in older patients. These findings also agree with the GEE and odds ratio evaluations performed previously.

Besides inflammation, tubular atrophy was the only diagnostic outcome to reach over AUROC > 0.70, at a maximum of 0.74. Although TAT and Age showed higher importance, as expected, the models benefitted from the addition of viscoelastic measurements, especially 2D FT shear wave attenuation at 200 Hz, α_{2DFT}^{200}, which showed a negative trend, the decrease in attenuation is consistent with the interstitial inflammation process, where interstitial contents expand and compress, decreasing the overall attenuation of the tissue [23][38][39]. The eGFR also demonstrated high importance in multiple models for the different outcomes, its negative trend relates directly to the allograft function, as the injury progresses the filtration capacity of the kidney tends to decline [45]. Interstitial fibrosis and IFTA had the worst performance at a maximum AUROC of 0.67 and 0.69, respectively. All Banff ci patients evaluated presented tubular atrophy as well, which did not occur with Banff i and Banff ct. The lack of isolated fibrosis cases might confound the classification of both Banff ci and IFTA. Further patient studies are necessary to isolate fibrosis and ensure proper training.

All patient studies were performed in a single institution what can impose levels of bias to the dataset. Future studies should aim at multi-institutional assessment to evaluate a broader demographic representation and different allograft care methodologies. The patients present in this study also showed minimal to no injury prevalence, which might confound the performance of the algorithms tested. Inflammation and isolated interstitial fibrosis had limited representation on the dataset, with 14% and no representation, respectively. Conversely, the single institutional characteristic of this study helps to control intra-operator variability of the scans performed as both sonographers and nephropathologists underwent similar training standards. Similar to other machine learning algorithms, RBF SVM assumes feature non-collinearity. Although SWE measurement can have different levels of multicollinearity that might confound the method

proposed, the low p-value feature selection did not alter the performance of the datasets significantly, as shown in Figure 3-2.

Overall, the work proposed showed encouraging results for early inflammation detection, even on limited datasets. With more Banff i cases, the classification performance may be improved further. Additionally, it is possible to relate the acquisition position with the biopsy location annotation, allowing better understanding of the localization of inflammation within the allograft. More frequent assessment of allograft state might enable better control of the inflammation process and, consequently, reduce the prevalence of progressive injuries such as fibrosis and tubular atrophy.

3.5 Conclusion

Overall, prediction of inflammation showed the best results, reaching AUROC = 0.77. Banff i, was also the diagnostic more correlated to rheological alterations, reaching best performance without the use of biomarker features. Undersampling was necessary in other to achieve superior results what indicates that more positive cases are necessary to create a more balanced, but still diverse, dataset. Interstitial fibrosis and tubular atrophy classification, differently, relied more on temporal biomarkers such as Age and time-after-transplant (both with positive trends), this agrees with the progressive characteristics of these injuries, being related to the recurrence of inflammation, in a scarring process. Regardless of biomarkers higher importance, Banff ci, ct and IFTA benefited from the inclusion of viscoelastic parameters such as 2D FT shear wave attenuation.

The results reported agree with previous GEE and odds ratio analysis, which reinforces the reliability of machine learning methods. The interpretation of SVM models enabled important insights into the rheological alterations causes by the injuries sustained during allograft rejection. Further investigation is necessary to better understand how the micro-structure changes caused by repeated inflammation, progressing fibrosis and tubular atrophy can dictate the macro viscoelastic changes observed. The kidney is a complex, anisotropic, and heterogenous organ, what poses a variety of data acquisition, processing and analysis challenges, machine learning might be an important tool to advance the field towards clinical translation.

3.6 Appendix

Diagnostic	Dataset	QA	Feature set	Under sampled	AUC	Sens	Spec	Cutoff	Acc	1st	2nd	3rd
Banff i	Median	No	Elastography	Yes	0.76	0.58	0.84	0.36	0.68	cp	mu	ap_100
Banff i	Median	No	All	Yes	0.75	0.67	0.76	0.48	0.68	cg	mu	ap_100
Banff i	Full	No	p-value	Yes	0.72	0.63	0.76	0.46	0.68	BPD	cp_SFK_200	ap_200
Banff i	Full	General	p-value	Yes	0.72	0.74	0.68	0.51	0.62	BPD	cp_SFK_200	ap_200
Banff i	Median	No	p-value	Yes	0.71	0.77	0.59	0.59	0.64	cg	BPD	eGFR
Banff i	Full	No	All	Yes	0.68	0.49	0.81	0.25	0.61	BPD	Kidney Depth	BMI
Banff i	Full	General	p-value	No	0.68	0.61	0.68	0.94	0.89	BPD	cp_SFK_200	cp_200
Banff i	L_MID	No	All	Yes	0.65	0.60	0.69	0.11	0.60	BMI	RI_LP	eGFR
Banff i	L_MID	Specific	Elastography	Yes	0.65	0.71	0.55	0.52	0.58	ap_200	cp_SFK_100	cp_SFK_150
Banff i	L_MID	Specific	p-value	Yes	0.65	0.72	0.55	0.54	0.59	cg	eGFR	BPD
Banff i	Full	Specific	All	Yes	0.64	0.41	0.85	0.17	0.59	Kidney Depth	BPD	BMI
Banff i	Full	Specific	Elastography	Yes	0.63	0.55	0.67	0.46	0.59	cg	ap_100	mu
Banff i	L_MID	Specific	p-value	No	0.63	0.85	0.42	0.94	0.87	cg	eGFR	BPD
Banff i	L_MID	No	p-value	Yes	0.62	0.93	0.30	0.73	0.58	cg	eGRF	BPD
Banff i	Full	No	Biomarkers	Yes	0.62	0.34	0.92	0.45	0.55	BMI	RI_LP	TAT
Banff i	Full	Specific	p-value	Yes	0.62	0.67	0.57	0.51	0.57	BPD	cp_SFK_200	ap_200
Banff i	L_MID	General	p-value	No	0.62	0.74	0.60	0.95	0.88	BPD	cg	eGFR
Banff i	Full	No	Biomarkers	No	0.61	0.37	0.84	0.80	0.87	eGFR	RI_M	BPD
Banff i	Median	No	Biomarkers	Yes	0.61	0.75	0.58	0.49	0.67	Kidney Depth	BPD	BPS
Banff i	L_MID	General	p-value	Yes	0.60	0.55	0.72	0.45	0.39	BPD	cg	eGFR
Banff i	Median	No	Biomarkers	No	0.60	0.85	0.45	0.85	0.87	Kidney Depth	BPD	eGFR
Banff i	L_MID	No	p-value	No	0.59	0.55	0.67	0.87	0.87	cg	eGFR	BPD
Banff i	Full	No	p-value	No	0.57	0.41	0.77	0.82	0.86	BPD	ap_200	cp_200
Banff i	Full	Specific	All	No	0.57	0.58	0.59	0.94	0.86	BPD	RI_LP	RI_UP
Banff i	Full	Specific	Elastography	No	0.57	0.50	0.63	0.86	0.86	ap_100	cp_SFK_100	cp_200

Diagnostic	Dataset	QA	Feature set	Under sampled	AUC	Sens	Spec	Cutoff	Acc	1st	2nd	3rd
Banff i	Full	Specific	p-value	No	0.56	0.83	0.31	0.93	0.86	BPD	cp_SFK_200	ap_200
Banff i	L_MID	Specific	Elastography	No	0.55	0.53	0.59	0.87	0.86	cg	ap_200	cp_SFK_100
Banff i	Median	No	All + IQR	No	0.53	0.90	0.17	0.51	0.46	Kidney Depth	eGFR	SCr
Banff i	L_MID	Specific	All	No	0.52	0.49	0.60	0.93	0.83	BPD	eGFR	Kidney Depth
Banff i	Median	No	p-value + IQR	No	0.52	0.39	0.72	0.84	0.87	BPD	cg	eGFR
Banff i	Median	No	p-value	No	0.51	0.39	0.74	0.85	0.87	BPD	cg	eGFR
Banff i	Median	No	Elastography + IQR	No	0.51	0.77	0.31	0.87	0.86	ap_200	ap_100	cp_SFK_250
Banff i	L_MID	No	All	No	0.50	0.12	0.94	0.52	0.87	BPD	eGFR	Kidney Depth
Banff i	L_MID	Specific	All	Yes	0.50	0.68	0.50	0.61	0.39	SCr	RI_LP	mu1
Banff i	Median	No	All	No	0.50	0.27	0.77	0.83	0.87	BPD	RI_M	eGFR
Banff i	Median	No	Elastography	No	0.50	0.78	0.26	0.86	0.72	cp_SFK_250	ap_200	cp_SFK_300
Banff i	Full	No	All	No	0.39	0.03	0.98	0.01	0.34	BPD	RI_UP	TAT
Banff ci	L_MID	General	p-value	N/A	0.67	0.68	0.64	0.51	0.55	TAT	Age	BMI
Banff ci	Median	No	Biomarkers	N/A	0.67	0.63	0.70	0.33	0.68	TAT	Age	eGFR
Banff ci	Full	No	All	N/A	0.66	0.51	0.73	0.49	0.62	TAT	Age	eGFR
Banff ci	L_MID	No	All	N/A	0.66	0.67	0.59	0.63	0.62	TAT	Age	eGFR
Banff ci	Full	Specific	All	N/A	0.65	0.60	0.61	0.63	0.54	TAT	Age	RI_LP
Banff ci	L_MID	Specific	All	N/A	0.64	0.56	0.63	0.59	0.58	TAT	Age	eGFR
Banff ci	L_MID	No	p-value	N/A	0.63	0.65	0.59	0.6	0.62	Age	TAT	BMI
Banff ci	Median	No	All	N/A	0.63	0.48	0.73	0.5	0.59	TAT	Age	eGFR
Banff ci	Full	No	Biomarkers	N/A	0.62	0.51	0.71	0.56	0.62	TAT	Age	eGFR
Banff ci	Median	No	p-value	N/A	0.62	0.61	0.59	0.56	0.59	Age	TAT	ap_200
Banff ci	Full	No	p-value	N/A	0.61	0.46	0.72	0.48	0.58	Age	TAT	RI_UP
Banff ci	Median	No	p-value + IQR	N/A	0.61	0.63	0.58	0.57	0.59	Age	TAT	ap_200
Banff ci	L_MID	Specific	p-value	N/A	0.60	0.57	0.65	0.56	0.55	TAT	Age	BMI
Banff ci	Median	No	All + IQR	N/A	0.60	0.48	0.72	0.44	0.57	TAT	eGFR	SCr
Banff ci	Full	Specific	p-value	N/A	0.59	0.67	0.48	0.67	0.57	Age	TAT	RI_UP

Diagnostic	Dataset	QA	Feature set	Under sampled	AUC	Sens	Spec	Cutoff	Acc	1st	2nd	3rd
Banff ci	Full	General	p-value	N/A	0.58	0.32	0.83	0.01	0.48	TAT	BMI	Age
Banff ci	L_MID	Specific	Elastography	N/A	0.55	0.48	0.61	0.56	0.55	cp_SFK_100	cg	cp_100
Banff ci	Full	Specific	Elastography	N/A	0.53	0.4	0.66	0.52	0.54	cp_100	cp_200	cg
Banff ci	Median	No	Elastography + IQR	N/A	0.52	0.51	0.53	0.57	0.58	cp_100	mu2	cp_SFK_200
Banff ci	Median	No	Elastography	N/A	0.49	0.88	0.13	0.35	0.67	cg	mu	cp_100
Banff ct	L_MID	No	p-value	N/A	0.74	0.67	0.72	0.30	0.69	TAT	Age	SCr
Banff ct	Full	General	p-value	N/A	0.73	0.68	0.69	0.04	0.68	TAT	Age	SCr
Banff ct	L_MID	No	All	N/A	0.70	0.64	0.71	0.12	0.68	TAT	kidney Depth	RI_UP
Banff ct	Full	Specific	All	N/A	0.70	0.73	0.63	0.28	0.69	TAT	eGFR	RI_UP
Banff ct	L_MID	Specific	p-value	N/A	0.70	0.78	0.53	0.44	0.67	TAT	Age	SCr
Banff ct	Median	No	p-value + IQR	N/A	0.69	0.66	0.68	0.32	0.66	TAT	Age	SCr
Banff ct	L_MID	Specific	All	N/A	0.68	0.71	0.61	0.25	0.67	RI_UP	RI_M	TAT
Banff ct	Full	No	All	N/A	0.66	0.49	0.78	0.14	0.69	TAT	RI_LP	RI_UP
Banff ct	Full	No	p-value	N/A	0.65	0.45	0.78	0.16	0.68	TAT	Age	SCr
Banff ct	L_MID	General	p-value	N/A	0.64	0.52	0.79	0.04	0.65	TAT	Age	BMI
Banff ct	Median	No	p-value	N/A	0.64	0.58	0.65	0.54	0.62	TAT	Age	SCr
Banff ct	Median	No	All + IQR	N/A	0.63	0.49	0.77	0.20	0.68	TAT	RI_LP	Age
Banff ct	Median	No	Biomarkers	N/A	0.63	0.70	0.51	0.59	0.59	TAT	RI_UP	kidney Depth
Banff ct	Median	No	All	N/A	0.62	0.45	0.75	0.12	0.66	TAT	RI_UP	eGFR
Banff ct	Full	Specific	p-value	N/A	0.61	0.47	0.71	0.02	0.63	TAT	Age	SCr
Banff ct	Full	No	Biomarkers	N/A	0.60	0.31	0.88	0.26	0.68	TAT	RI_UP	eGFR
Banff ct	Full	Specific	Elastography	N/A	0.52	0.65	0.40	0.40	0.64	cp_SFK_200	cp_SFK_150	cg
Banff ct	L_MID	Specific	Elastography	N/A	0.52	0.68	0.39	0.42	0.63	cp_SFK_300	cp_SFK_200	cp_SFK_100
Banff ct	Median	No	Elastography	N/A	0.51	0.66	0.39	0.58	0.57	cg	mu	ap_200
Banff ct	Median	No	Elastography + IQR	N/A	0.50	0.74	0.27	0.34	0.67	ap_200	IQR	ap_100
IFTA	Median	No	p-value	N/A	0.69	0.60	0.70	0.28	0.69	Age	TAT	ap_200
IFTA	L_MID	Specific	All	N/A	0.67	0.46	0.78	0.41	0.61	TAT	eGFR	Age

Diagnostic	Dataset	QA	Feature set	Under sampled	AUC	Sens	Spec	Cutoff	Acc	1st	2nd	3rd
IFTA	L_MID	No	All	N/A	0.65	0.76	0.47	0.74	0.60	TAT	Age	RI_LP
IFTA	Full	No	p-value	N/A	0.64	0.41	0.81	0.42	0.61	Age	TAT	BMI
IFTA	Median	No	All	N/A	0.64	0.56	0.65	0.57	0.59	TAT	Age	RI_LP
IFTA	Full	Specific	All	N/A	0.63	0.66	0.50	0.71	0.59	TAT	Age	RI_LP
IFTA	L_MID	Specific	p-value	N/A	0.63	0.57	0.68	0.56	0.57	Age	TAT	BPS
IFTA	Median	No	p-value + IQR	N/A	0.63	0.69	0.54	0.59	0.59	Age	TAT	ap_200
IFTA	Median	No	Biomarkers	N/A	0.63	0.70	0.51	0.59	0.59	TAT	Age	eGFR
IFTA	Full	No	All	N/A	0.62	0.40	0.78	0.35	0.58	TAT	Age	eGFR
IFTA	L_MID	No	p-value	N/A	0.62	0.60	0.61	0.55	0.60	TAT	Age	BPS
IFTA	L_MID	General	p-value	N/A	0.62	0.46	0.80	0.49	0.55	TAT	BPS	Age
IFTA	Median	No	All + IQR	N/A	0.62	0.43	0.77	0.42	0.57	TAT	Age	eGFR
IFTA	Full	Specific	p-value	N/A	0.61	0.61	0.58	0.59	0.6	TAT	Age	RI_UP
IFTA	Full	No	Biomarkers	N/A	0.60	0.65	0.57	0.58	0.57	TAT	Age	eGFR
IFTA	Full	General	p-value	N/A	0.59	0.72	0.43	0.71	0.50	TAT	Age	BMI
IFTA	L_MID	Specific	Elastography	N/A	0.56	0.42	0.74	0.37	0.53	cp_SFK_100	ap_200	mu
IFTA	Median	No	Elastography	N/A	0.53	0.28	0.78	0.53	0.57	cp_100	mu2	cg
IFTA	Full	Specific	Elastography	N/A	0.52	0.27	0.78	0.45	0.54	cp_100	cp_SFK_150	cp_SFK_100
IFTA	Median	No	Elastography + IQR	N/A	0.52	0.24	0.86	0.58	0.57	cp_SFK_100	ap_200	mu2

Chapter 4 – Viscoelastic parameter estimation using simulated shear wave motion and convolutional neural networks

4.1 Introduction

Ultrasound imaging is one of the main imaging modalities used in healthcare. It is widely used for multiple applications due to its low cost, relative portability and safety [123]. One of the applications that has developed popularity over the last several years is ultrasound shear wave elastography (SWE). SWE typically uses focused ultrasound beams to produce acoustic radiation force (ARF) that pushes within the tissue [124], [125]. When the ARF ceases, the locally perturbed tissue returns to its equilibrium position, but also generates waves traveling perpendicular to the push direction. These shear waves propagate through the tissue, and the motion is measured using high frame rate ultrasound techniques [125],[126].The wave motion can be analyzed to extract information about the wave velocity and attenuation. These parameters can also be used to characterize rheological properties of the tissue such as elasticity and viscosity.

The estimation of rheological parameters is dependent on mathematical models, such as the Kelvin-Voigt and Maxwell models [31], [128]. To date, methods have been developed to directly measure the wave velocity and attenuation of shear waves as they vary with frequency [32], [75], [83], [121], [122], [128]–[131]. However, the shear wave attenuation can be difficult to directly measure [78]. With the wave velocity dispersion information alone, it is possible to estimate viscoelastic properties for different rheological models [31], [75], [83], [128], [132], [133]. Most of these methods involve Fourier analysis to extract dispersion curves that are used for fitting to the rheological models. Each of the approaches that have been studied have different levels of bias and variability. Additionally, different aspects of the SWE acquisition such as the ARF distribution and beam shape and measurement noise can affect the wave motion and confound these methods [134]–[136]. Taken together, there is still a need for robust methods for viscoelastic parameter evaluation from measured shear wave motion.

Machine learning (ML) is the computer science field dedicated to creating algorithms capable of modeling complex datasets, and performs tasks such as prediction, regression or clustering. These methods commonly have high computational costs. In recent years, there has been substantial progress on both algorithms and computational

power to perform data training and predictions in feasible time with available computational resources. ML can identify patterns within datasets that otherwise would be very difficult to fit by standard statistical models or by evaluating each feature on its own. The use of ML might enable SWE to reach higher applicability, by naturally modeling the intrinsic complexity of viscoelastic tissues [38], [39], [137], [138].

Artificial neural networks (ANN), or simply neural networks (NN), were inspired by the central nervous system structure and its physiology. They are composed by nodes, or neurons, interconnected into multiple layers, from input to output. Each node has adjusted weights and activation functions that simulate the synaptic reinforcement plasticity. The adjustment of the weights is implemented by solving the convex problem that minimizes the loss of the networks, which evaluates the prediction assigned by the network against the ground truth output [36], in a process called supervised learning.

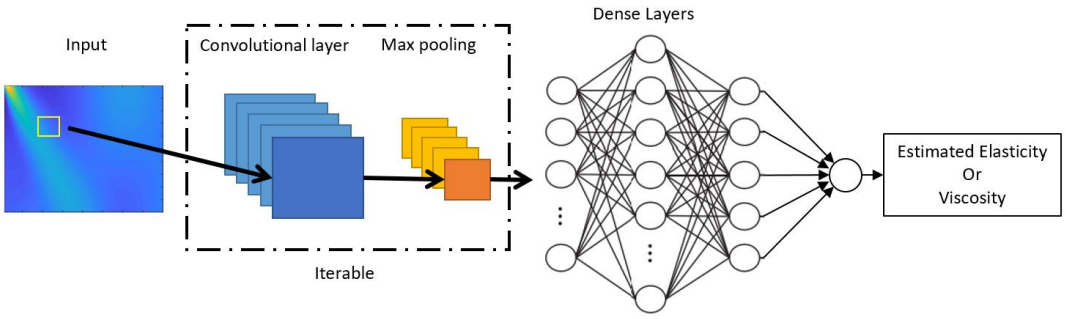


Figure 4-1 – General schema of a convolutional neural networks. Convolutional and max pooling layers can be stacked for deeper networks. The input image is convolved with a sliding window (yellow square), resulting in a set of feature maps that are processed by the max-pooling layers. The output of the final max-pooling is then processed by a set of dense layers responsible to interpret the feature maps and provide the estimated value of elasticity or viscosity as output.

The convex problem is solved using numerical approaches, such as gradient descent, which are iterative, and therefore, can take time and large amounts of data. The process of weight adjustment is called optimization and it is an area of research of its own. Some of the most common optimizers used for neural networks are Stochastic Gradient Descent [139] and Adam [140].

Convolutional neural networks (CNN), as other NN architectures, were also inspired by neurophysiology. The architecture replicates the layer-based structure from the human visual system, where the convolutional layers work as the receptive fields present in the brain, while the max-pooling layers act as the lateral geniculate nucleus by reducing the level of information and extracting the relevant features for the task in hand

[37]. The input is convolved with the feature weights to produce a feature map which is processed by a non-linear activation function (rectified linear units, traditionally). The pooling layers then reduce the spatial resolution of the feature maps to achieve spatial invariance to input distortions and translations. Max-pooling layers aggregate and propagate the maximum value in a receptive field to the next layers. After stacking multiple convolutional and pooling layers, a fully connected layer is implemented to interpret the feature representations extracted by previous layers, much like the human visual cortex.

The fully connected layer is a traditional NN, with different output layer activation functions and loss functions, depending on the task [37]. The CNN basic schema is described in Fig. 4-1. CNNs have been extensively used for image recognition tasks with high performance and generalizability [37], [137].

In this proof-of-concept study, we propose to analyze, in a controlled setting, the capabilities of ML to model wave propagation and retrieve elasticity and viscosity values from wave motion images simulated using the staggered-grid finite-difference (SGFD) approach to implement a Kelvin-Voigt (KV) material model [34], [141], [142]. Due to the two-dimensional (2D) image characteristics of the task, the CNN architecture was deemed the best for the objectives of this proof-of-concept. We also compared the ML performance to the two-dimensional Fourier transform (2D FT) analysis approach for evaluating dispersion curves and fitting to the KV theoretical dispersion curve [31], [128], [132].

The modeling of viscoelastic wave propagation is useful for understanding wave phenomena in complex media. Finite-difference modeling approaches have great importance for geophysical exploration, reservoir engineering, and military applications because they provide complete wavefield responses at reasonable computational cost [141], [143]. The SGFD method was used to simulate wave propagation in viscoelastic homogeneous media. SGFD adopts a finite difference operator with fixed-order accuracy to calculate spatial derivatives for a homogeneous or heterogeneous medium. The method's accuracy, efficiency and stability are improved by the use of staggered-grid for spatial discretization [141], [142].

4.2 Methods

For this study, the simulations were created using SGFD simulations implemented in MATLAB (Mathworks, Natick, MA, USA) and computed on GTX 1080 and RTX 2080Ti GPU cards (Nvidia, Santa Clara, CA, USA). Each simulation took approximately 1-8 hours

of computing time depending on the GPU card used. The SGFD simulations implement a KV material in the Navier-Stokes equations to calculation the motion caused by acoustic radiation force from ultrasound push beams. The shear elasticity and viscosity ranges for the KV model were chosen so they would cover a wide range of values for normal and pathological soft tissues [33]. The shear elasticity, μ_1, ranged from 1-25 kPa with increments of 1 kPa, and shear viscosity, μ_2, ranged from 0-10 Pa·s with increments of 0.5 Pa·s. The 25 values of shear elasticity and 21 values of shear viscosity yield 525 unique combinations of viscoelastic parameters. The medium is assumed to be homogeneous, isotropic, linear, and incompressible [34].

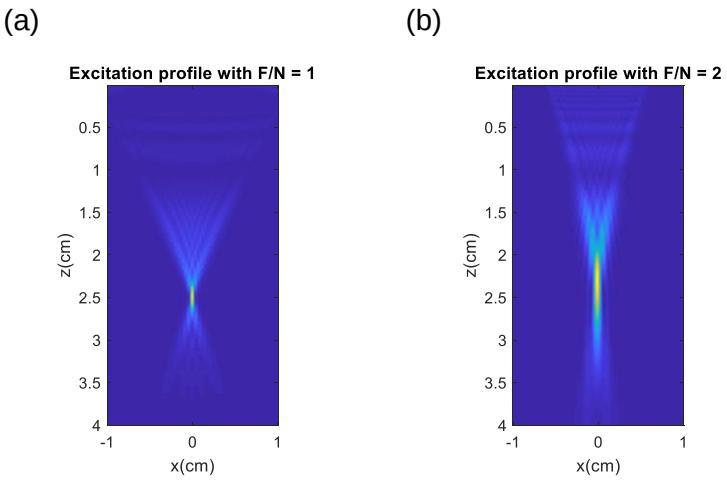


Figure 4-2 – Images of excitation profiles for F/N = (a) 1 and (b) 2 based on the normalized beam intensity. Lower f-numbers generate more concentrated pushes, whereas with higher f-numbers the push profile is spread axially (z), creating more planar wave fronts.

We used two different ARF excitation focus profiles with f-numbers (F/N = focal length/width of aperture) 1 and 2 (Fig. 4-2) with a focal distance of z_f = 2.5 cm, adding up to 1050 simulations. These values of F/N are used frequently in practice and provided cases of a very focused beam (F/N = 1) and a less focused beam (F/N = 2). The ARF was simulated using FOCUS using the model of an L7-4 linear array transducer (Philips Healthcare, Andover, MA) [144], [145]. The element pitch was set to 0.0308 cm, the element kerf was 0.0025 cm, the element height was 0.7 cm, the elevation focus was 2.5 cm, and the frequency was 4 MHz. The medium was modeled using a density of 1000 kg/m^3, speed of sound of 1540 m/s, and ultrasound attenuation of 0.5 dB/cm/MHz. A focal

depth of 2.5 cm was used and for F/N = 1 and 2, such that 82 and 40 elements were used, respectively. The pressure was simulated, and the intensity was calculated as it is proportional to the ARF in absorbing soft tissues. For the models, the force was applied for 200 μs. For each combination of μ_1, μ_2, and F/N, a dataset was produced.

An important factor of finite-difference methods is the model's boundary conditions. The boundary conditions have to absorb and attenuate the waves without creating artifacts such as reflections. It is also important that boundary conditions do not overly increase computational costs. Perfectly matched layers (PML) are considered the optimal absorbing boundary condition for finite-difference methods, and therefore, were implemented in this study [146].

Another important aspect of SGFD simulations is the Courant-Friederichs-Lewy criterion (CFL), as it describes the conditional stability that allows convergent behavior of the solution. The space-time discretization is dictated by the smallest shear elasticity (μ_1) and the highest shear viscosity (μ_2) to be simulated, and so, the CFL criterion determines the time increment and thus the number of time steps to achieve a certain simulation length for a given shear wave velocity.

$$C = \frac{\Delta t \sqrt{V_p^2 + V_s^2}}{\Delta x} \ll 1 \ , \tag{4-1}$$

where the dimensionless number C is called the Courant number, V_p and V_s are the compressional and shear wave velocities, Δt is the time step, and Δx the spatial step in the x dimension. In this study's simulations, the Δx and Δz (spatial step in the z dimension) were set to 0.1 mm, in a field of 44 mm width and 32 mm height. To optimize the simulation run time, two different values of C were used in the simulations depending on the media properties: 0.2 if $\mu_1 > 5$ kPa and 0.05 if $\mu_1 \leq 5$ kPa and $\mu_2 \geq 4$ Pa·s. The simulation Δt was then calculated based on Eq. (4-1), for a total period of 25 ms. To meet the conditions described above, the temporal resolution was very high (5-7 MHz). After completion, the simulated data was resampled in time to 16 kHz, to ensure that all models would have the same time discretization. This sampling frequency could be reached with interpolated data that would be upsampled for precise time-of-flight estimates for evaluation of shear elasticity estimations [133].

The three-dimensional simulations (z, x, t) were then cropped and sampled in the z-direction to create the 2D x-axis vs. time wave motion dataset for training. The excitation generates two identical wave fronts propagating perpendicular to the push, therefore, the simulation region-of-interest (ROI) used to create the dataset was composed solely by the

right hand-side wave front (Fig. 4-3a). The ROI has a width of 2 cm with its origin at 0.2 cm laterally from the push center, to eliminate the high amplitude motion of the push from the dataset images. The time range adopted was 20 ms. The data that was used for the CNN input was the motion from particular depths in the z-direction. To augment the dataset the right hand-side lateral propagation was used from five different vertical positions relative to the push focus (Fig. 4-3a), at -0.90, -0.45, 0.00 (center), 0.45 and 0.90 cm (Fig. 4-3). Another technique implemented to augment the number of samples, was adding different levels of Gaussian noise to the lateral propagation images to generate motion with signal-to-noise ratios (SNRs) consistent to literature, at 20, 10 and 5 dB (Figs. 3b and 3c) [34], [147], [148]. The segmentation and augmentation of the dataset was also implemented in MATLAB. With the five depth positions and three noise levels and 1050 unique combinations of the viscoelastic parameters, the total number of images used for the study was 15,750.

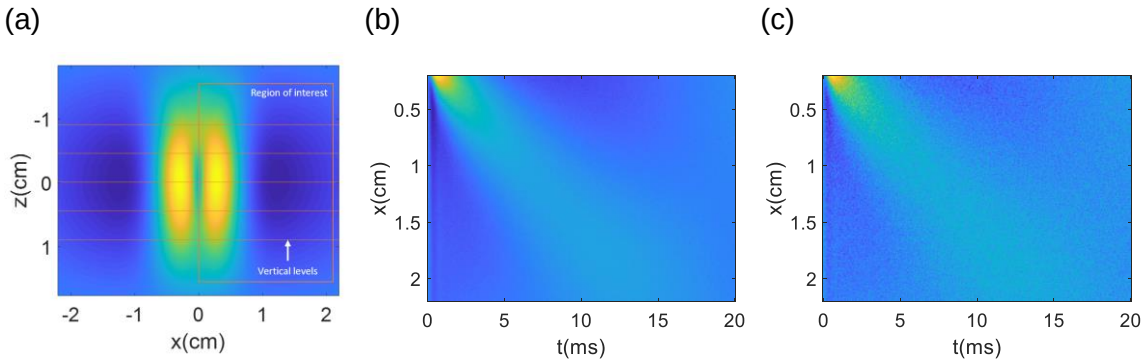


Figure 4-3 – (a) Example of segmentation performed during the preprocessing stage for F/N = 2, μ_1 = 2 kPa and μ_2 = 4 Pa·s. The z-axis origin is the center of excitation (z_f = 25 mm from Fig. 4-1); (b) and (c) examples of 20 and 5 dB SNR.

The CNN architecture was implemented in Python (Python Software Foundation, Wilmington, DE, USA), using Keras [149] with Tensorflow (Google Brain, Mountain View, CA, USA) backend, and trained on a Z820 workstation (Hewlett-Packard, Palo Alto, CA, USA) equipped with a GTX 1080 GPU (Nvidia, Santa Clara, CA, USA). The data in each image was normalized by subtracting the minimum amplitude from the whole ROI and dividing all motion amplitude values by the maximum amplitude after subtraction, creating images where all the values had an amplitude between 0 and 1. The labels were also normalized with a normalizing factor of the maximum label value (μ_1 = 25 kPa for elasticity and μ_2 = 10 Pa·s for viscosity) plus a 20% margin, to leave a margin for model overestimation in the top of the range. The same margin was not used at the bottom of

the range (μ_1 = 1 kPa for elasticity and μ_2 = 0 Pa·s for viscosity), because negative values are not physically possible.

The training, validation and test sets were defined using a group shuffle split of 60/20/20, which means that images derived from the same simulation (μ_1, μ_2, F/N combination) could not be present in more than one set simultaneously. This ensures that simulation subjects presented to the CNN in the validation and test datasets were never seen during training in any of their augmented versions. Due to the regression characteristics of the task, sigmoid and mean squared error (MSE) were implemented as output layer activation function and loss function, respectively.

The elasticity and viscosity determine different characteristics of the wave propagation. The elasticity is related to the time-of-flight (e.g., slope of the wave motion distribution in the x-t plane, Fig. 4-3b), whereas the viscosity affects the widening of the wave at a particular location.

Therefore, for each dataset, a pair of CNN models were implemented, one for elasticity regression, and one for viscosity regression. To facilitate the compilation of results, a coding system was applied to name the models and results in terms of what the training/validation and test f-numbers were used (Table 4-1). The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing.

Table 4-1 – Model codes

Pair code	Training F/N (number of images)	Validation F/N (number of images)	Test F/N (number of images)
E11, V11	1 (4725)	1 (1575)	1 (1575)
E12, V12	1 (4725)	1 (1575)	2 (7875)
E22, V22	2 (4725)	2 (1575)	2 (1575)
E21, V21	2 (4725)	2 (1575)	1 (7875)
E1+2, V1+2	1 & 2 (9450)	1 & 2 (3150)	1 & 2 (3150)

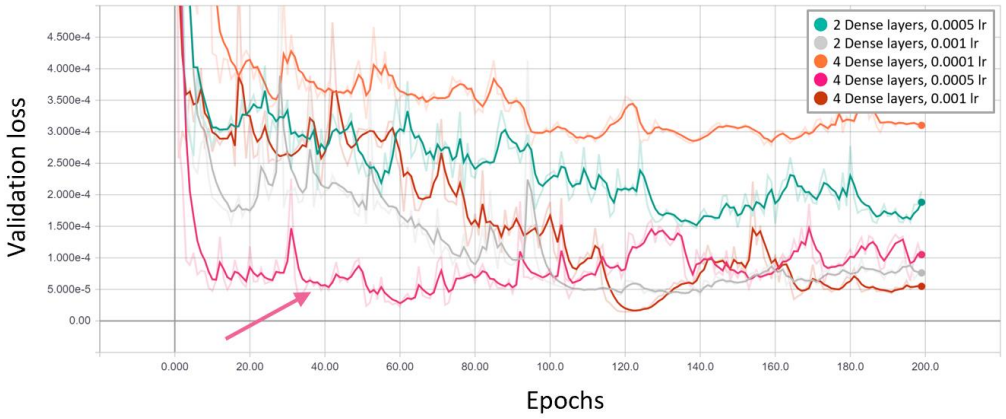
The CNN models trained, validated and tested on data from both f-numbers were denoted by the X1+2 designation. The CNN models were optimized for multiple combinations of 1, 2 and 3 convolutional layers; 2, 3 and 4 dense layers; 16 and 32 nodes per layer (convolutional and dense); 0.001, 0.0005 and 0.0001 learning rates; totaling 54 models. Tensorboard (Google Brain, Mountain View, CA, USA) was used to monitor and evaluate the models during the hyperparameter grid search. By evaluating the validation loss and mean absolute error (MAE) during 200 epochs, it is possible to determine when

the learning reached an asymptote or demonstrated overfitting. MAE is defined by the equation:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{x}_i - x_i| \ , \tag{4-2}$$

where N is the number of test subjects, $\hat{x}_i$ is the estimated value of the subject and x_i is the true value of the parameter.

(a)



(b)

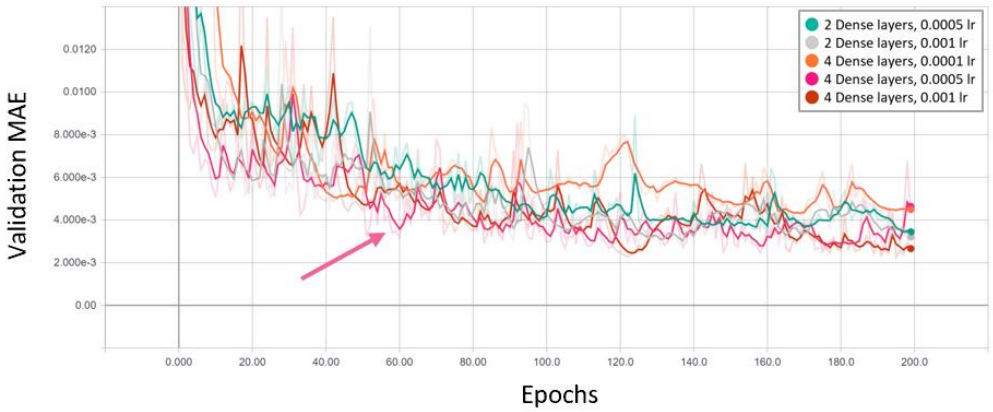


Figure 4-4 – Examples of validation loss (a) and MAE (b) for second round of evaluations, where the combinations were reduced to 3 convolutional layers; 3 and 4 dense layers; 32 nodes per layer (convolutional and dense); 0.001, 0.0005 and 0.0001 learning rates. The lower the values of validation loss and MAE indicated a better model. From these plots 3 convolutional layers, 4 dense layers with 32 nodes and a 0.0005 learning rate were chosen for elasticity regression of F/N = 1 (pink arrow). The curves are plotted using Tensorboard with a smoothing factor of 0.5 (solid color lines) over the calculated values (shaded color lines).

Each 200-epoch training session took approximately 1 hour of GPU computing time on the GTX 1080 GPU card. For each grid search, the top 20% CNN models with the

lowest validation MAE were selected and re-trained for a subsequent round of evaluation until one hyperparameter set remained (Fig. 4-4). This approach was used to mitigate hyperparameter overfitting, because each iteration of cross evaluation is performed on different random group splits. The final model was then trained, also for 200 epochs, using the optimal hyperparameters. Checkpoint recall was used to store the best model generated and prevent overfitting. The models were then evaluated against the test set.

The datasets were also evaluated using 2D FT analysis and the estimated dispersion curves were fit to the theoretical Kelvin-Voigt model to obtain the viscoelastic parameters. This analytical approach was implemented in MATLAB. The 2D fast Fourier transform was performed with a zero padding factor of 4096 for optimum frequency and wavenumber resolution of the frequency domain representation (k-space). The dispersion curves were than estimated from the k-space using the relationship $c_p(\omega_p) = \omega_p / k_{max}$, where c_p is the phase velocity at ω_p frequency from 100-600 Hz, and k_{max} is the wavenumber of maximum k-space energy at ω_p. Additionally, $c_p$values below 0 m/s and above 10 m/s were discarded from further analysis. The dispersion curves were then fit to the theoretical Kelvin-Voigt model (Eq. 4-3), with the density ρ = 1000 kg/m^3, to obtain μ_1 and μ_2 values. The estimated values were then compared to the CNN results.

$$c_p(\omega_p) = \frac{\sqrt{2(\mu_1^2+\omega_p^2\mu_2^2)}}{\rho\left(\mu_1+\sqrt{\mu_1^2+\omega_p^2\mu_2^2}\right)} \quad . \tag{4-3}$$

4.3 Results

The testing results were evaluated using the MAE between the true value of the viscoelastic property and the CNN prediction output. A linear regression was performed to compare the results from the 2D FT analysis as well as the CNN prediction to the identity line, and the slope, intercept and R^2 of the regression line were also calculated and reported. The optimal hyperparameters were also noted and reported in Tables 4-2 and 4-3 for the elasticity and viscosity CNN models, respectively. All elasticity models trained and tested on the same f-numbers showed very good fitting with MAE below 0.079 kPa with slopes and intercept of approximately one and zero, respectively. The viscosity models trained and tested on the same f-numbers also showed very good agreement with MAE below 0.091 Pa·s with slopes and intercept of approximately one and zero, respectively. Alternatively, CNN models tested on f-numbers different from the

training/validation F/N had diminished performance with MAE of at least 1.542 Pa·s (V12) and as high as 2.834 kPa for E12.

Table 4-2 – Elasticity results								
CNN								
Result Code	Hyperparameters				Results			
	Convolutional Layers	Dense Layers	Layer Size	Learning Rate	MAE, kPa	Slope	Intercept	R^2
E11	3	4	32	0.0005	0.051	1.000	-0.001	1.000
E12					2.834	0.666	4.951	0.836
E22	2	2	32	0.0001	0.079	1.001	-0.033	1.000
E21					1.663	0.848	0.547	0.958
E1+2	4	4	32	0.0005	0.065	0.998	0.021	1.000
2D FT								
E1					2.380	0.936	0.404	0.762
E2					2.211	0.901	1.252	0.775

Table 4-3 – Viscosity results								
CNN								
Result Code	Hyperparameters				Results			
	Convolutional Layers	Dense Layers	Layer Size	Learning Rate	MAE, Pa·s	Slope	Intercept	R^2
V11	4	3	32	0.0001	0.065	0.997	0.007	0.999
V12					1.644	0.678	-0.030	0.856
V22	3	2	32	0.0005	0.052	0.998	0.001	0.999
V21					1.611	1.046	1.296	0.885
V1+2	3	2	32	0.0001	0.069	1.003	-0.015	0.999
2D FT								
V1					3.173	1.580	-0.022	0.693
V2					2.925	1.593	-0.240	0.695

The errors for each CNN model were also aggregated and plotted versus the true value (Fig. 4-5). The data were reported with boxplots where the red line represents the median, and the edges of the box represent the interquartile range (IQR, 25th and 75th quantiles). The whiskers represent 1.5*IQR above or below the 75th or 25th quantile, respectively. Values denoted by markers outside of whiskers are denoted as outliers. For cases E11, E22, and E1+2, the level of bias was minimal over the range of true shear elasticity. For E12, μ_1 was overestimated when the true μ_1 < 15 kPa, and underestimated for larger values of μ_1. For E21, a consistent underestimation of μ_1 was observed. In the cases of V11, V22, and V1+2, the bias was relatively low, where the MAE was less than 0.069 Pa·s. For V12, μ_2 was consistently underestimated and for V21 μ_2 was consistently overestimated.

When compared to the 2D FT analysis, the CNN results were found to be significantly more reliable. The Fourier analysis showed reduced performance at higher

viscosity and lower elasticity combinations, driving elasticity underestimation and viscosity overestimation. In these types of materials, the ARF excitation was not able to generate energy at higher frequencies (300-600 Hz), and therefore it limited the quality of Kelvin-Voigt fitting. Although the linear regression technique was able to retrieve acceptable values for slope and intercept for the elasticity estimations, the errors still drove MAE values above 2.2 kPa, and R2 below 0.78. The viscosity estimations were more inferior to the performance achieved with the CNNs, with slopes above 1.5, MAE above 2.9 Pa·s and R^2 below 0.7.

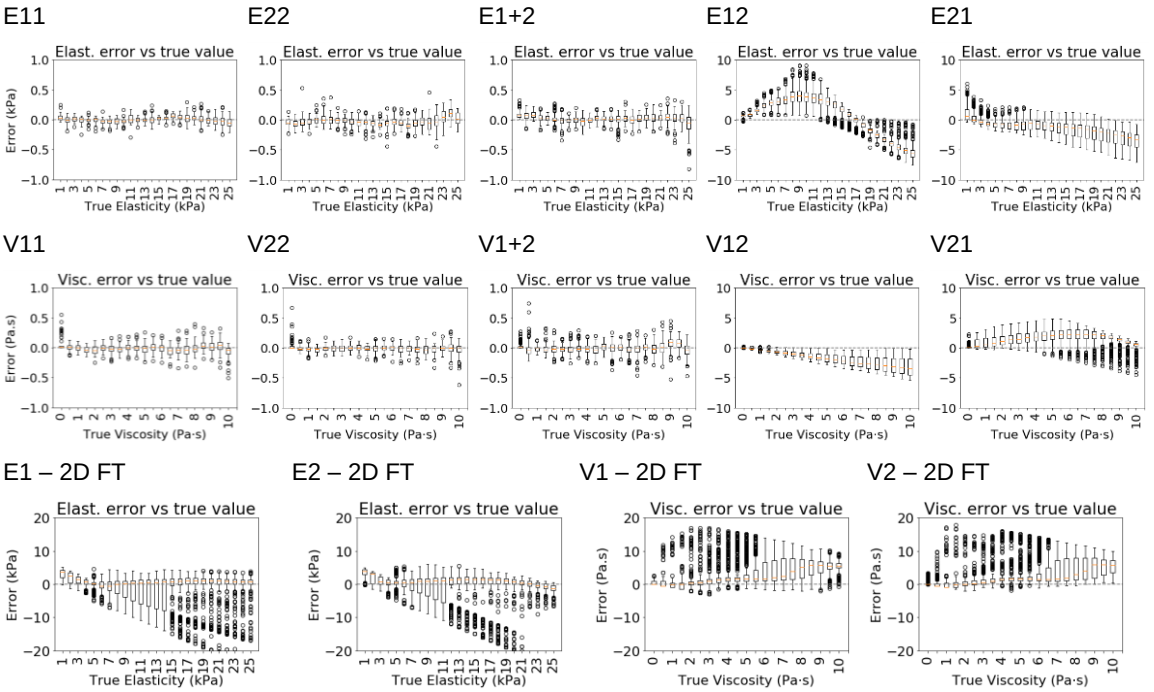


Figure 4-5 – Elasticity and viscosity error box plots aggregated by true values, the top ten plots are results for the CNNs trained, validated and tested on data from the same and different F/N datasets, whereas the bottom four were evaluated using Fourier analysis (FT) for comparison. The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing.

The errors were also plotted versus vertical position in Fig. 4-6. Similar trends were observed for these additional parameters as compared to those discussed above. The prediction of μ_1 was largely insensitive to the vertical position of the wave motion used, though the results for model E1+2 exhibited more outliers. For E12, the bias is positive (average of 0.14 kPa), and 1.5*IQR of over 0.63 kPa showing a very large variability, but insensitive versus vertical position. For E21, there was a larger negative bias for data from the pre-focal locations (-0.27 kPa at -0.90 cm and -0.1 kPa at the focus), while larger

numbers of outliers for post-focal datasets. For V11, V12, and V1+2, the bias was small, but the number of outliers slightly increased for the V1+2 case. The errors in μ_2 for V12 were largely insensitive to the position, though for V21 the error increased both in bias and variability as the data farther from the focus was used, with 0.1 ± 0.18 kPa at the focus, 0.15 ± 0.22 Pa·s at -0.90 cm and 0.28 ± 21 Pa·s at 0.90 cm. When compared to the 2D FT analysis, the CNN is more reliable with respect to the depth position, when trained with the proper f-number. The same sensitivity to off-focus position was also observed using the 2D FT analysis for F/N = 1 (Fig. 4-6 – E1-2D FT and V1-2D FT).

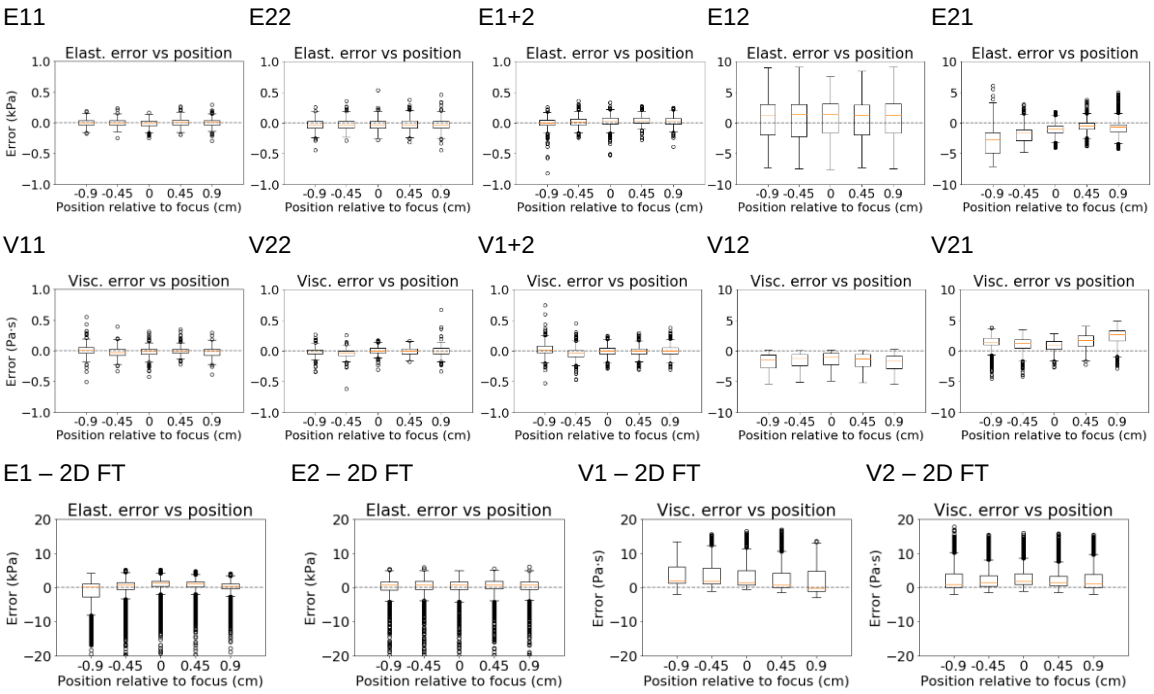


Figure 4-6 – Elasticity and viscosity error box plots aggregated by vertical position, the top ten plots are results for the CNNs trained, validated and tested on data from the same and different F/N datasets, whereas the bottom four were evaluated using Fourier analysis (FT) for comparison. The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing.

The errors for the models as a function of noise level are shown in Fig. 4-7. The model results for the elasticity prediction were not largely affected by the level of noise. The cases of E1+2 and E21 did demonstrate a larger number of outliers. The models trained/validated and tested with different f-numbers showed bias invariance to the noise levels, nevertheless, E11 showed an increase of 21% and 82% in 1.5*IQR from 20 dB to 10 dB and 10 dB to 5 dB, respectively, while V11 showed an increase of 60% and 38% in 1.5*IQR from 20 dB to 10 dB and 10 dB to 5 dB, respectively. The results for model V1+2

showed more outliers than V11 and V22 models. The 2D FT analysis showed a larger number of outliers, when compared with the CNN. While the number of outliers were relatively constant, the 2D FT results were similar across different levels of noise.

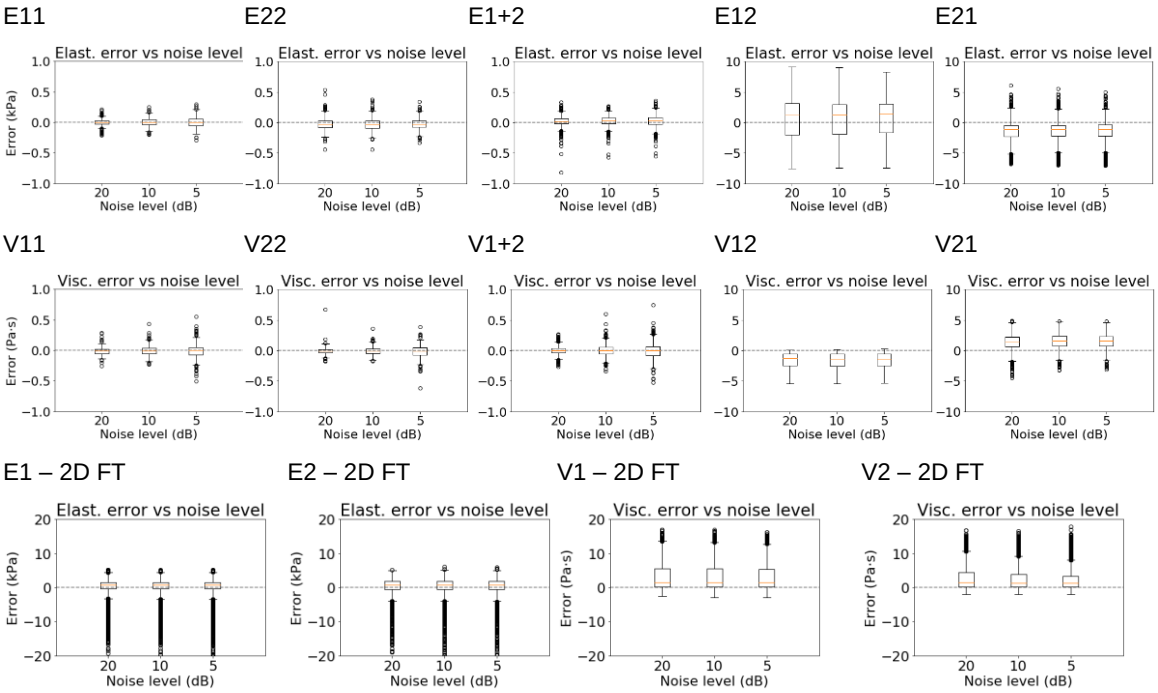


Figure 4-7 – Elasticity and viscosity error box plots aggregated by noise levels, the top ten plots are results for the CNNs trained, validated and tested on data from the same and different F/N datasets, whereas the bottom four were evaluated using Fourier analysis (FT) for comparison. The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing.

For models trained, validated and tested with both f-numbers, the errors were also compiled by f-number (Fig. 4-8). No relevant difference was found between the errors for both elasticity and viscosity. Because all the data is aggregated, there are a high number of outliers present, but the error has a similar overall range for all models and F/N values (Tables 4-2 and 4-3). The outliers cover a similar range for both F/N values for V1+2. Models trained and tested on the same f-numbers generally showed very low absolute errors with no relevant bias and standard deviation of less than 0.5 kPa or 0.5 Pa·s throughout the true value range.

Models E21 and V12 showed a pattern of higher bias and IQR with the increase of true value, whereas E12 and V21 showed more complex patterns of error across the true value range. In some cases, even with the overestimation margin applied to the label normalization, the higher value estimations were saturated and drove the bias negatively,

e.g., E11 and V11. It is also notable that model V1+2 presented a higher level of outliers and positive bias at 0 Pa·s.

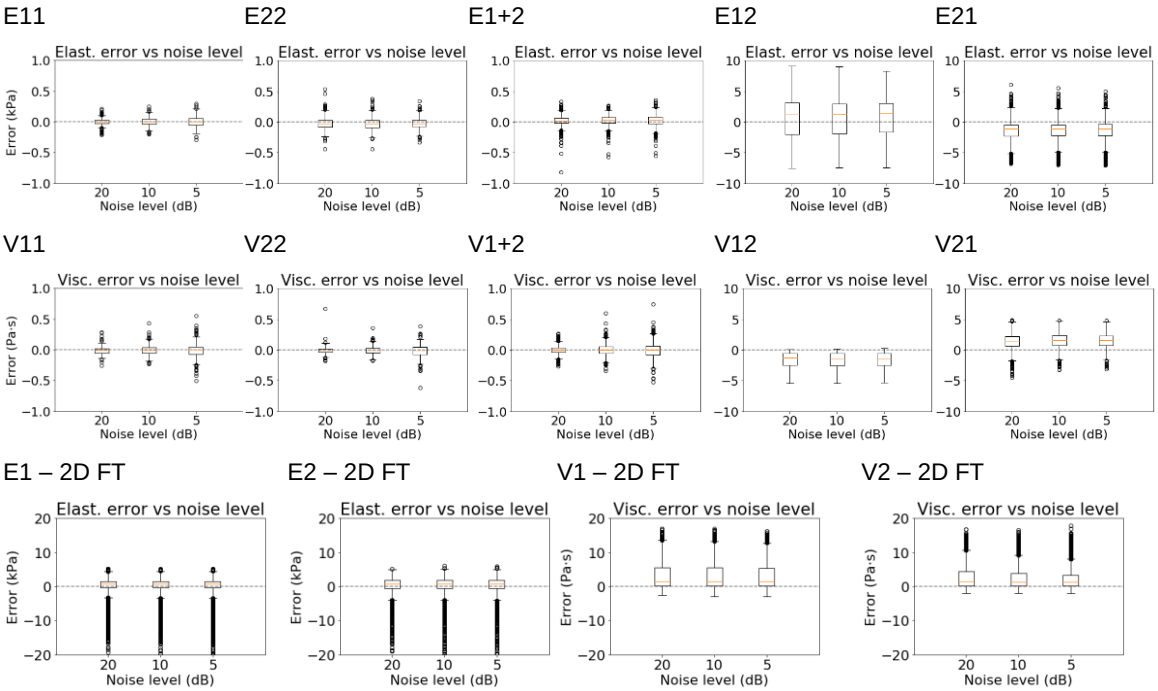


Figure 4-8 – Elasticity and viscosity error box plots aggregated by noise levels, the top ten plots are results for the CNNs trained, validated and tested on data from the same and different F/N datasets, whereas the bottom four were evaluated using Fourier analysis (FT) for comparison. The letter designates what label was used, elasticity (E) or viscosity (V); the first number specifies the F/N used for training and validation, whereas the second number designates the F/N used for testing.

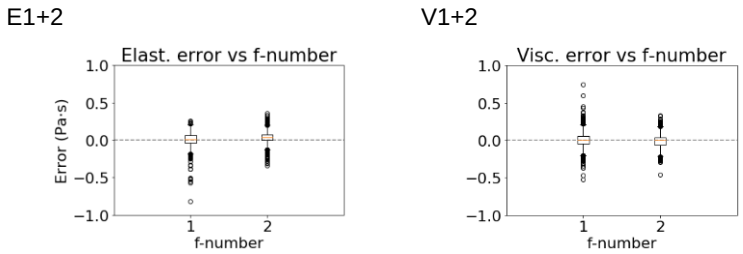


Figure 4-9 – Elasticity and viscosity error box plots aggregated by the F/N of the test subject. The model pair was trained, validated and tested on both F/N datasets (E1+2 and V1+2).

4.4 Discussion

The results shown are encouraging and exhibit the potential of the use of convolutional neural networks in ultrasound SWE for viscoelastic characterization. The

results demonstrated that the CNN models were robust to noise, vertical position and partially to F/N. When tested on datasets with the same excitation F/N for training and testing, the elasticity and viscosity estimations were very good and showed robustness to noise and vertical position.

The CNN models were marginally reliable when presented with test data from a different F/N than used for the training, with a trend of overestimation for E12 and V21, and underestimation for E21 and V12. Vertical position showed some level of variability in models trained with data using F/N = 2 and tested on data with F/N = 1 (X21), possibly due to the planar characteristics of higher F/Ns, when compared to a lower F/N, which produced more curved wave fronts.

Higher f-numbers have less focused beams that might cause wider lateral shear wave profiles that could simulate the widening effect of viscosity in the wave propagation. When trained on data using F/N = 1 and tested on data when F/N = 2 (X12), the results showed a consistent underestimation profile increasing linearly with the true viscosity (V12, Fig. 4-5), while the opposite effect occurred with V21, showing consistent viscosity overestimation. The V21 overestimation is exaggerated with the distance from the focus, as F/N = 1 has a more curved intensity profile to produce the wave front while F/N = 2 has a more planar wave front (Fig. 4-2). This behavior is well exemplified in Figs. 4-6 and 4-9. This characteristic is important because it suggests the possibility of reliable estimation correction for cases of F/N for which the model was not trained.

When different F/N values were used, the elasticity estimations showed the opposite behavior from the viscosity, overestimation for E12 and underestimation for E21. It is possible that the more curved intensity profile in the z-direction used for the ARF push beam, and the presence of more pre-focal and post-focal intensity for F/N = 1 also leads to errors in elasticity estimation, as it might change the perceived slope of the wave motion (Fig. 4-10).

Similarly to viscosity, the elasticity showed correlation between absolute bias and the distance to push focus when trained with F/N = 2 and tested on F/N = 1 (X21). Models trained with F/N = 2 are tuned to the planar wave front characteristics, therefore, when presented with data from a lower F/N for testing, show higher levels of error off-center (Fig. 4-6, V21 and E21).

Noise did not affect the estimations significantly in any of the scenarios tested, which suggests the robustness of the technique to SNR changes.

When trained with both f-numbers the CNNs were able to overcome the shortcomings discussed above and performed similarly as the pairs trained, validated, and tested with a single F/N. The results indicate that the CNNs can discern between wave front profiles with high accuracy, as long as they are trained properly.

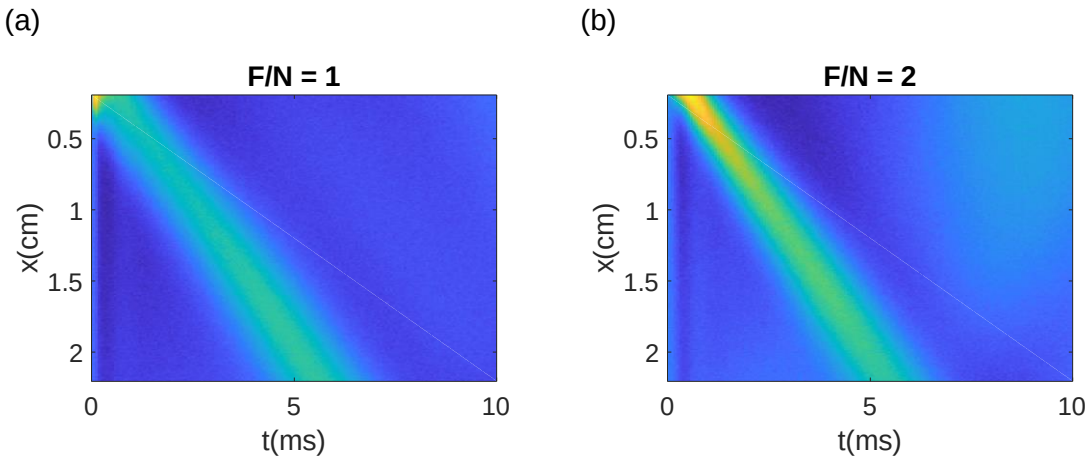


Figure 4-10 – Example of μ_1 = 18 kPa and μ_2 = 2 Pa·s, at z = +0.90 cm and 20 dB SNR, with F/N = 1 (a) and F/N = 2 (b). It is possible to observe that the response using F/N = 1 has a wider wave front, which might simulate the viscosity effect which also widens the wavefront.

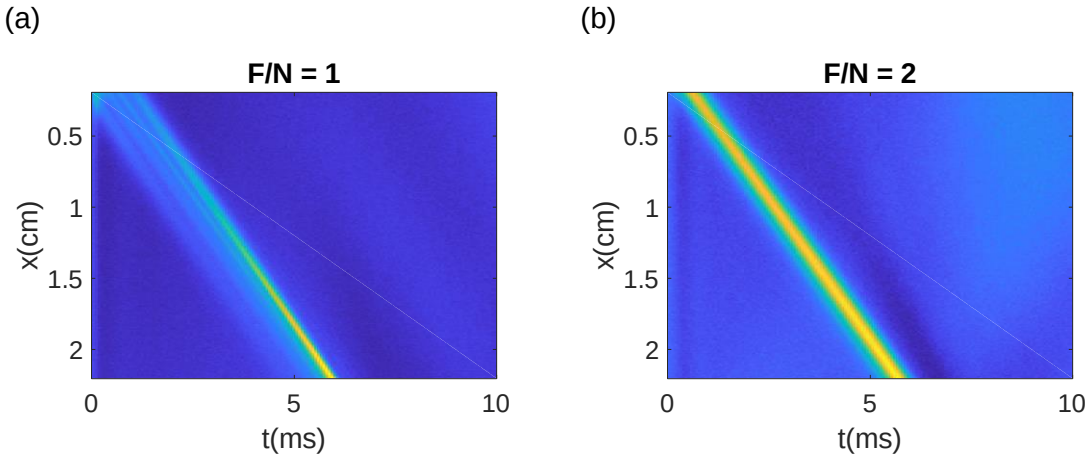


Figure 4-11 – Example of μ_1 = 18 kPa and μ_2 = 0 Pa·s , at z = +0.90 cm and 20 dB SNR, with F/N = 1 (a) and F/N = 2 (b). The fringing effect present in F/N = 1 case might explain the elasticity bias present in cross F/N evaluations.

Overall, this proof-of-concept study showed that the CNN models were able to successfully retrieve the viscoelastic properties from the wave motion images. The approach proposed required minimal preprocessing, without the employment of filters or transforms. The conventional 2D FT analysis approach showed inferior performance on

the same datasets, with worse MAE and R^2 levels found compared to those found with CNN regression, even in cases where the CNN was trained and tested with different f-numbers. The 2D FT analysis was especially affected by materials where the effects of the viscosity was more prominent compared to the elasticity (e.g., high μ_2 and low μ_1), whereas the CNN did not show the same behavior and was reliable over the full viscoelastic parameter range. As noted in the Results, in the cases where the viscosity effects were substantial, energy at higher frequencies was diminished and curve fitting was more error prone. These findings with the CNN models are particularly encouraging, because changes in viscoelastic properties are related to pathological changes to the tissue, e.g. fibrotic tissue is correlated with higher μ_1 [32], [132]. Therefore, patients and practitioners can benefit from ML classification of tissue state based on SWE, which is an affordable and non-invasive measurement.

Although encouraging, the methods discussed still need further investigation. The method needs to be extended to allow for heterogeneous cases and other noise characteristics. The practical use of ML in ultrasound SWE has specific challenges such as the noise due to ultrasound detection and physiological motion, the standardization of images, and simulation model limitations that will have to be addressed in future work.

In this study, we used simulated motion from the SGFD based on a Kelvin-Voigt rheological model. Although it has been shown in literature, in multiple works, that the Kelvin-Voigt model does describe shear wave velocity dispersion over the typical elastography analysis ranges of frequency, such as 100-600 Hz [85], [150]–[152], the use of the KV model on SGFD simulations might impair the use of simulations for *ex vivo* and *in vivo* transfer learning as not all tissues may be well-described by a KV rheological model. We also chose to use the 2D x-t wave motion images to directly estimate the values of μ_1 and μ_2. Other ML approaches or NN architectures could be utilized to estimate the dispersion curves for fitting to different rheological models.

We did not incorporate the ultrasound detection in these simulated data, which may add additional bias, but this proof-of-concept approach shows the promise of using shear wave motion and the overall shape of the wave fronts to estimate the viscoelastic properties. We did explore adding Gaussian noise to the wave motion data, and the models were largely robust to different noise levels. Higher levels of noise could be explored in future work. Physiological motion will also introduce noise as well that is not accounted for in this work. Each of the images used for these models used a fixed range of lateral space and measurement time. For *in vivo* implementation, this would have to be

standardized for a given application or transducer. We also limited the range of depths over which we made the evaluations. Typically, viscoelastic measurements are made within or near the depth-of-focus of the ARF push beams, but an evaluation of parameter estimation versus depth will be investigated in our ongoing work.

Ultrasound SWE implementations have a wide variety of characteristics, such as F/N, transducer profile, motion detection approaches, all of which with specific use cases, strengths and weaknesses, and therefore, need to be evaluated separately. Ideally, for *in vivo* implementation, an acquisition approach with the aforementioned parameters would need to be thoroughly modeled and likely used in the training and validation and later testing for a robust technique. The entire pipeline including ARF push beam modeling, wave motion modeling with different material models, ultrasound detection, and motion estimation and filtering would need to be incorporated to all for transfer learning with experimental data from phantoms or *ex vivo* and *in vivo* soft tissues. This work represents a first step to demonstrate the promise of this type of approach as well as identify some of the next steps to be taken in this technology development.

4.5 Conclusion

In this study, we proposed to evaluate the capabilities of CNNs to model wave propagation and retrieve elasticity and viscosity values from wave motion images sourced from staggered-grid finite-difference (SGFD) simulations. A total of 1050 SGFD models were simulated using a wide range of elasticity and viscosity values for two different values of excitation F/N. The dataset was then augmented using five different vertical positions to generate the spatiotemporal images and using three different levels of noise. The dataset was used to train, validate and test CNN architectures configured for data regression using mean squared error as loss function. The viscoelastic parameters were also estimated using the conventional 2D Fourier analysis approach with Kelvin-Voigt dispersion curve fitting and compared against the CNN results.

The overall results showed the CNNs' potential to be an alternative to complex mathematical analyses such as Fourier analysis and dispersion curve estimation used currently for shear wave viscoelastic parameter estimation. The CNN approach does not impose the biases and variabilities associated with these methods and requires minimal amount of preprocessing, such as filters and transforms. The methods proposed might enable simpler and more reliable non-invasive evaluation of tissue injuries that alter

rheological parameters such as liver and kidney fibrosis. The use of transfer learning is also a possibility to enable *ex vivo* and *in vivo* viscoelastic estimations [153], and this will be pursued in ongoing work.

More evaluations regarding other f-numbers are necessary, although higher values will yield even more planar wave fronts and should provide easier translation to other CNN models. The initial robustness to noise was also encouraging as it is a strong characteristic of ultrasound applications, but other noise settings need to be evaluated. This work serves as a proof-of-concept to demonstrate the use of CNN models for estimation of viscoelastic properties directly from shear wave motion. We will continue to develop this technology to test on experimental data from phantoms and soft tissues.

Chapter 5 – Kidney cortex shear wave motion simulations based on segmented biopsy slides

5.1 Introduction

Chronic kidney disease (CKD) is defined by the progressive decrease in renal filtration. In later stages of failure, when the kidneys are no longer capable of sustaining homeostasis, hemodialysis or kidney transplant are necessary to increase patient survival and quality of life. Although patient survival rates have increased with the improvement of transplant protocols, improving allograft longevity is still challenging. Transplanted patients with allograft failure are required to return to dialysis and return to transplant waiting lists [40]. To monitor allograft health, and adjust the course of treatment accordingly, several biomarkers are commonly used to evaluate kidney function, such as serum creatinine (SCr) and glomerular filtration rate (GFR) [96]. Doppler imaging techniques can also be leveraged to calculate the resistive index (RI) in multiple locations of the allograft [14]. Although important surrogates of kidney health, biopsy is still the gold standard for allograft assessment, due to its capabilities of detecting microscopic tissue alterations before kidney function is affected significantly.

The most common injuries used for allograft prognosis from biopsy samples, as described by the Banff criteria [46], [100], are interstitial inflammation (i), interstitial fibrosis (ci) and tubular atrophy (ct) [98], [99]. The Banff scores can range from 0, indicating no presence of disease, to 4, with increasing severity. In spite of their usefulness in the clinical practice, protocol biopsies cannot be performed frequently, because the invasive nature of the procedure can cause complications such as bleeding, fistulas, infections and even death [101]–[103].

Over the last decades the use of elasticity imaging techniques has been evaluated to discriminate from normal and abnormal tissue. Different pathological processes can physiologically and morphologically modify the structure of organs, and therefore, possibly modify its rheological characteristics. Shear wave elastography (SWE) is a technique that has emerged with the potential for non-invasive and affordable tissue evaluation. The SWE method relies on exciting shear waves with a focused ultrasound beam, which produces acoustic radiation force. The wave propagation is then measured with high frame rate ultrasound imaging. The wave motion is processed to assess a variety of shear

wave characteristics, such as shear wave group velocity, phase velocity and attenuation [16], [19], [62]–[64].

In addition to the more established elastic properties, the viscoelasticity of human tissue may also change with disease. The viscosity of a given material dictates time-dependent deformation properties, which can be estimated using the shear wave velocity dispersion, i.e., the change in shear wave speed in respect to frequency. The velocity dispersion curves can be obtained by applying a two-dimensional Fourier transform (2D FT), in space and time, to the shear wave motion data. It is possible then to evaluate the dispersion curves with the use of model-free parameters such as specific phase velocities and attenuation [78], [82], [85], or perform fitting to rheological models, such as Kelvin-Voigt, to estimate the shear elasticity and viscosity of the material [29], [31], [65].

It has been shown that various elastography measurements correlate to the prevalence of interstitial inflammation, fibrosis and tubular atrophy, in different levels [64], [117]–[119]. However, there is still a gap in knowledge of how the morphological alterations in the kidney microstructure can alter the macroscopic shear wave propagation for different injuries.

Viscoelastic material simulation through finite difference (FD) numerical techniques have been vastly adopted in geophysical applications due to their capabilities of modelling wave propagation in complex discretized media at feasible computational cost [141], [143]. Staggered-grid finite-difference (SGFD) scheme have been shown to successfully to implement the Kelvin-Voigt (KV) material model for human tissue [34], [154]. SGFD calculates the spatial derivatives for a given media with fixed-order accuracy. The use of the staggered-grid method for spatial discretization enhances the stability, efficiency and accuracy of FD models [141], [142].

In this study, heterogeneous models based on kidney allograft biopsy histology were simulated using SGFD to investigate the interaction of the morphological alterations commonly caused by rejection with the rheological alterations as perceived by SWE. The parametric *in silico* experiment was developed based on digitized biopsy cores of 12 kidney transplanted patients (3 normal, 3 with interstitial inflammation, 3 with interstitial fibrosis and 3 with tubular atrophy). The main components of kidney morphology (glomeruli, tubule, interstitial, and fluid) were segmented and assigned baseline mechanical properties. The baseline properties were then altered independently to evaluate the impact to the overall shear wave propagation for all the different subjects.

5.2 Methods

The subjects from this study were obtained from a cohort of 200+ patients already evaluated by our group in previous clinical studies. The patients' biopsies were collected at routine allograft surveillance visits at Mayo Clinic from February 2016 to June 2019. Our SWE study followed a protocol approved by the Mayo Clinic Institutional Review Board (11-003249). In addition to the biopsy Banff scores annotation and core digitization, each patient was also scanned using the SWE capabilities of a Logiq E9 ultrasound system (General Electric Healthcare, Wauwatosa, WI, USA). The in-phase and quadrature (IQ) data were stored for post-processing and viscoelastic parameter estimation. The 12 subjects were selected based on having low IQR-to-median ratio values available in the dataset, to minimize intra-operator variability. The subjects' demographics are displayed in Table 5-1.

Table 5-1 – Study cohort demographics

	Patients (n=12)	
	Number	%
Sex		
Male	5	42
Female	7	58
	Mean (Std. Dev)	Range
Age, yrs	43.7 (13.2)*	24-60
BMI, kg/m^2	25.0 (4.0)*	20.7-35.7
TAT (months)	18.3 (20.3)*	4-60

The biopsy cores were digitized at 100x magnification using trichrome staining due to its higher color contrast between the interstitial and tubule components. The segmentation was performed in MATLAB (Mathworks, Natick, MA, USA) leveraging the RGB component differences between interstitial space, tubules, and fluid. The RGB approach was not capable of discerning the glomeruli from the interstitial and tubular space, and therefore, a user interface was implemented to manually segment the glomeruli due to its complex morphology. An example of histological sample and its respective segmented image is displayed in Figures 5-1a and 5-1b. At 100x magnification was not possible to digitize the entire core into a single image, therefore, the multiple segmented images from a single core were concatenated. The whole core segmented images were then randomly selected, flipped horizontally and vertically, and concatenated to create 12 composite 1.0 x 2.0 cm 2D models (Fig. 5-1c), one for each subject. The diagnostic outcome scores and relative morphological component area, for each subject analyzed are listed in Table 5-2.

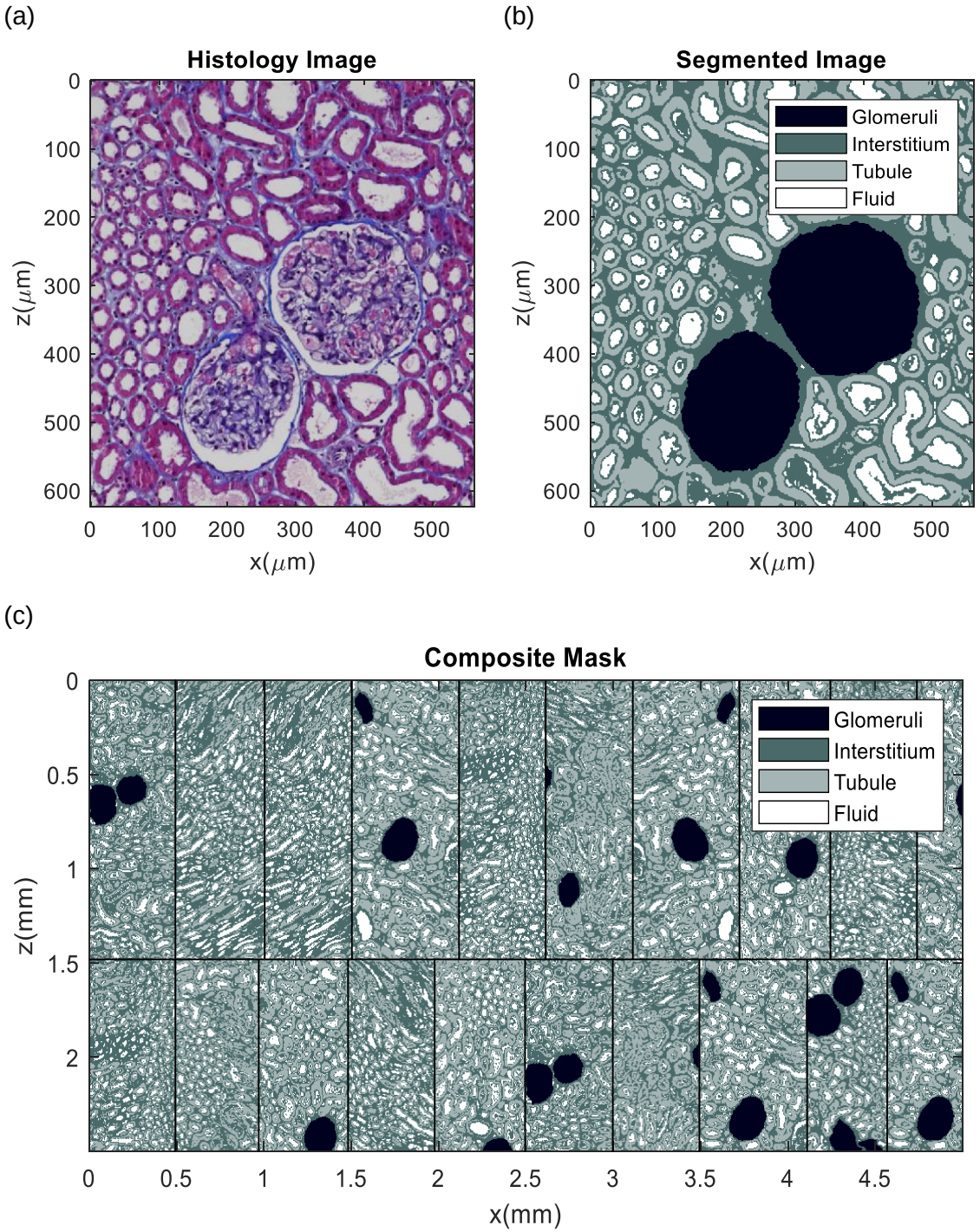


Figure 5-1 – (a) Original 100x histology image, (b) segmented version of digitized histology, (c) 2.5 x 5 mm sample of 1.0 x 2.0 cm composite mask used for SGFD simulation. The black lines divide the different cores composing the image.

Table 5-2 – Diagnostic outcome and spatial composition							
Diagnostic	**Banff i**	**Banff ci**	**Banff ct**	**G%**	**T%**	**I%**	**Fl%**
Normal 1	0	0	0	5.6%	46.1%	38.3%	10.0%
Normal 2	0	0	0	4.3%	38.0%	42.2%	15.5%
Normal 3	0	0	0	6.3%	33.1%	47.8%	12.7%
Banff i 1	2	0	0	7.9%	42.4%	41.5%	8.2%
Banff i 2	1	0	0	5.7%	40.0%	46.0%	8.3%
Banff i 3	2	0	0	5.5%	20.7%	69.9%	4.0%
Banff ci 1*	0	1	1	6.1%	40.1%	42.9%	10.9%
Banff ci 2*	0	2	2	3.2%	55.4%	33.5%	8.0%
Banff ci 3*	0	1	1	3.7%	25.5%	57.9%	12.9%
Banff ct 1	0	0	1	5.8%	47.7%	37.1%	9.3%
Banff ct 2	0	0	1	3.1%	41.7%	44.0%	11.1%
Banff ct 3	0	0	1	4.1%	48.7%	41.6%	5.5%
G: Glomeruli, T: Tubules, I: Interstitial space, Fl: Fluid *All patients from the IRB cohort that presented interstitial fibrosis also presented with tubular atrophy.							

The KV materials were implemented using the Navier-Stokes equations to calculate shear wave propagation excited by an acoustic radiation force (ARF) push [155]. The push was generated using Field II [156], with f-number (F/N = focal length/width of aperture) of 2. The transducer model was set to L7-4 linear array transducer (Philips Healthcare, Andover, MA), with an element pitch of 0.0308 cm, kerf of 0.0025 cm, height of 0.7 cm, and the frequency was set to 4 MHz. A 200 μs ARF excitation was applied using 40 elements to simulate the push at a focal depth of 0.5 cm.

The media was assumed to be isotropic, linear, and incompressible [34]. All models were set with a density of 1000 kg/m^3, speed of sound of 1540 m/s, and ultrasound attenuation of 0.5 dB/cm/MHz. The shear elasticity, μ_1, and shear viscosity, μ_2, of each kidney cortex component was determined using baseline values found in the literature [151], [157]–[161]. From baseline values, two other values were assigned, above and below baseline, to perform a parametric study with a total of 156 models. The baseline values and parametric analysis are listed in Table 5-3.

Table 5-3 – Parametric analysis				
Cortex component	**Parameter**	**Lower value**	**Baseline**	**Upper value**
Glomeruli	μ_1, kPa	2.00	4.00	6.00
	μ_2, Pa·s	0.10	0.25	1.00
Tubules	μ_1, kPa	1.00	1.50	2.00
	μ_2, Pa·s	0.10	0.25	1.00
Interstitium	μ_1, kPa	0.50	1.00	2.00
	μ_2, Pa·s	0.50	1.00	2.00
Fluid	μ_1, kPa	N/A	0.00	N/A
	μ_2, Pa·s	N/A	0.10	N/A

Boundary conditions are crucial to the effectiveness simulating wave propagation in viscoelastic models, the interaction of pressure and shear waves with the borders of the model can generate artifacts that would not be present in real applications. The use of perfectly matched layers (PML) was implemented to the models for optimal absorption of pressure and shear waves at the boundaries of the material [146]. The spatial resolution was set to 0.05 mm for both axes, and the simulation time was set to 15 ms. The simulation sampling frequency was defined by the Courant-Friederichs-Lewy criterion (CFL), allowing for model's conditional stability and convergent behavior. The CFL determines the temporal step Δt based on the compressional and shear wave velocities, V_p and V_s, and the spatial step, Δx, in the x dimension (Eq. 5-1). In this study, a CFL of 0.2 was used for proper model stability.

$$\Delta t = \frac{CFL* \Delta x}{\sqrt{V_p^2+V_s^2}}, \tag{5-1}$$

The SGFD models were implemented in MATLAB (Mathworks, Natick, MA, USA) and processed using computing resources from the Minnesota Supercomputing Institute at the University of Minnesota. The simulations leveraged CUDA (Nvidia Corporation, Santa Clara, CA, USA) for GPU cluster utilization. The simulations computing time was approximately 1-2 hours when queued for Nvidia V100 (Nvidia Corporation, Santa Clara, CA, USA) nodes, and approximately 3-4 hours if the Nvidia K40 (Nvidia Corporation, Santa Clara, CA, USA) clusters nodes were used. The shear motion data was resampled then saved with a sampling frequency of 8 kHz for further processing.

To evaluate the masks' rheological heterogeneity, the shear wave velocity (SWV) map of the baseline models for each patient was calculated. The map was calculated using a window of 0.30 x 0.30 mm with a patch size of 0.15 mm [162]. The dispersion curves, from the simulations and patient scans, were retrieved using generalized Stockwell transformation combined with a slant frequency-wavenumber analysis (GST-SFK) [81] from 100-400 Hz, due to its wider and more reliable frequency band when compared to the standard 2D Fourier transform approach [34], [82]. The dispersion curves were then fit to the Kelvin-Voigt model (Eq. 5-2) to estimate the shear elasticity, μ_1, and shear viscosity, μ_2

$$c_p(\omega) = \sqrt{\frac{2\left(\mu_1^2+\omega^2\mu_2^2\right)}{\rho\left(\mu_1+\sqrt{\mu_1^2+\omega^2\mu_2^2}\right)}}, \tag{5-2}$$

where the shear wave phase velocity (c_p) is described as a function of the angular frequency (ω), the media density ρ. Additionally, the group velocity was calculated using time-to-peak algorithm and a linear fit (Eq. 5-3) was applied to the simulation dispersion curves to estimate the dispersion slope (dc/df) and phase velocity at 200 Hz ($c_p(200)$) [32]

$$c_p(f) = c_0 + \frac{dc}{df} f \ , \tag{5-3}$$

where c_0 is the shear wave velocity intercept at 0 Hz and f is the phase frequency.

5.3 Results

A total of 156 models were simulated for the 12 subjects' masks. The models' rheological heterogeneity can be observed in Fig. 5-2. The shear wave speed velocities were able to successfully illustrate the spatial variation of mechanical properties induced by the presence of the various different components of kidney morphology and its different mechanical properties, particularly as the shear wave propagates through glomerular structures.

In some cases, due to the limitation of available biopsy cores for digitization, the shear wave velocity map showed some level of striation, as the cores had to be repeated more often to generate the model mask, creating the pattern observed for patients 1 and 3 in the Banff ct category. The x-axis origin was positioned at push focus, therefore, the maps only cover about 15.5 mm. It is important to note that all shear waves showed significant attenuation after approximately 12 mm of lateral propagation, the higher velocity values estimated after shear wave dissipation (yellow to red hue) are, therefore, artifacts.

The group velocity, μ_1, μ_2, dispersion slope and intercept at 200 Hz were calculated and reported. For simplification the three different subjects composing each diagnostic outcome under evaluation (normal, Banff i, ci and ct) were compiled into mean and standard deviation for each of the simulated variations from baseline listed in Table 5-3. The slope calculation showed the highest level of standard deviation, with a mean of 0.14 m/s/kHz, with a maximum standard deviation of 0.63 m/s/kHz for the upper interstitial viscosity variation (μ_2 = 2 Pa·s from 1 Pa·s baseline). All other measurements showed significantly lower standard deviations, with a maximum of 0.16 Pa·s for the shear

viscosity for the upper interstitial viscosity variation (μ_2 = 2 Pa·s). The results are displayed in Figure 5-3.

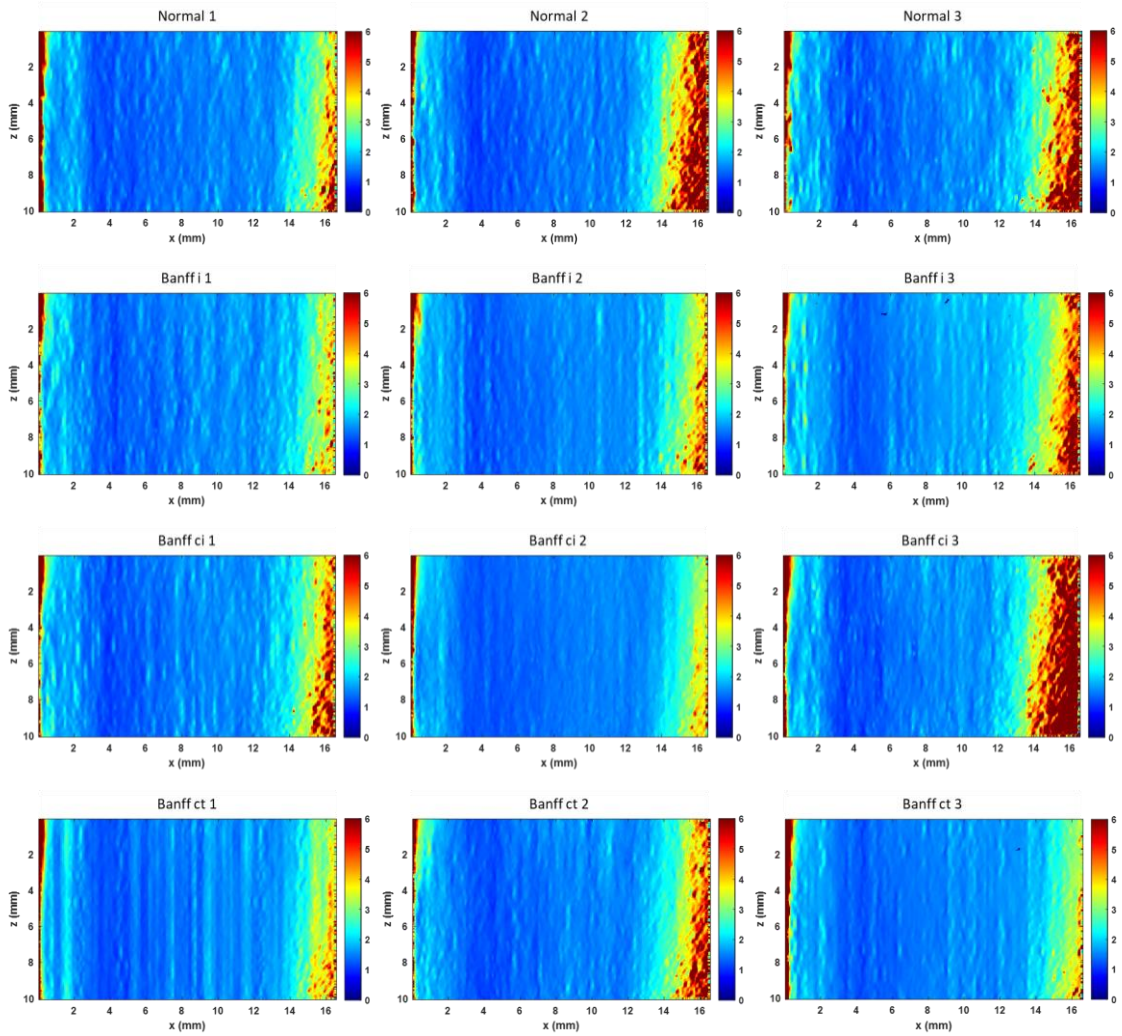


Figure 5-2 – Shear wave velocity maps calculated from baseline simulations of each subject. The rightmost high velocity artifacts are caused by the low amplitude of the shear wave, as it attenuates away from the push at origin. The striated pattern is caused by the limited number of cores available for digitization, and therefore its repetition over the mask composition.

For diagnosis comparison, the average measurements' difference from normal within each diagnosis were calculated and divided by the normal subjects' average to obtain a percentage change from normal (Eq. 5-4). The results regarding inflammation showed the highest levels of viscoelastic change, with a mean change of over 25% in μ_2 and 32% in slope. The results from inflammation and tubular atrophy subjects showed slight changes from normal in μ_1, group velocity and intercept, ranging from 1.8% (Banff ct intercept) to 4.9% (Banff ct shear elasticity). Banff ct showed more significant change

for interstitial alterations specifically, with -0.3 to -0.5% μ_2 change, and -6.2 to -11.7% slope change. The cases of interstitial fibrosis were the least affected injury in terms of elasticity, with 0.7%, 1.0% and 1.2% change for group velocity, intercept, and μ_1, respectively. Viscosity evaluations showed larger change trends for interstitial fibrosis, within 6.2% to 8.9% change in μ_2 and slope, respectively. The percent change plots are displayed in Figure 5-3.

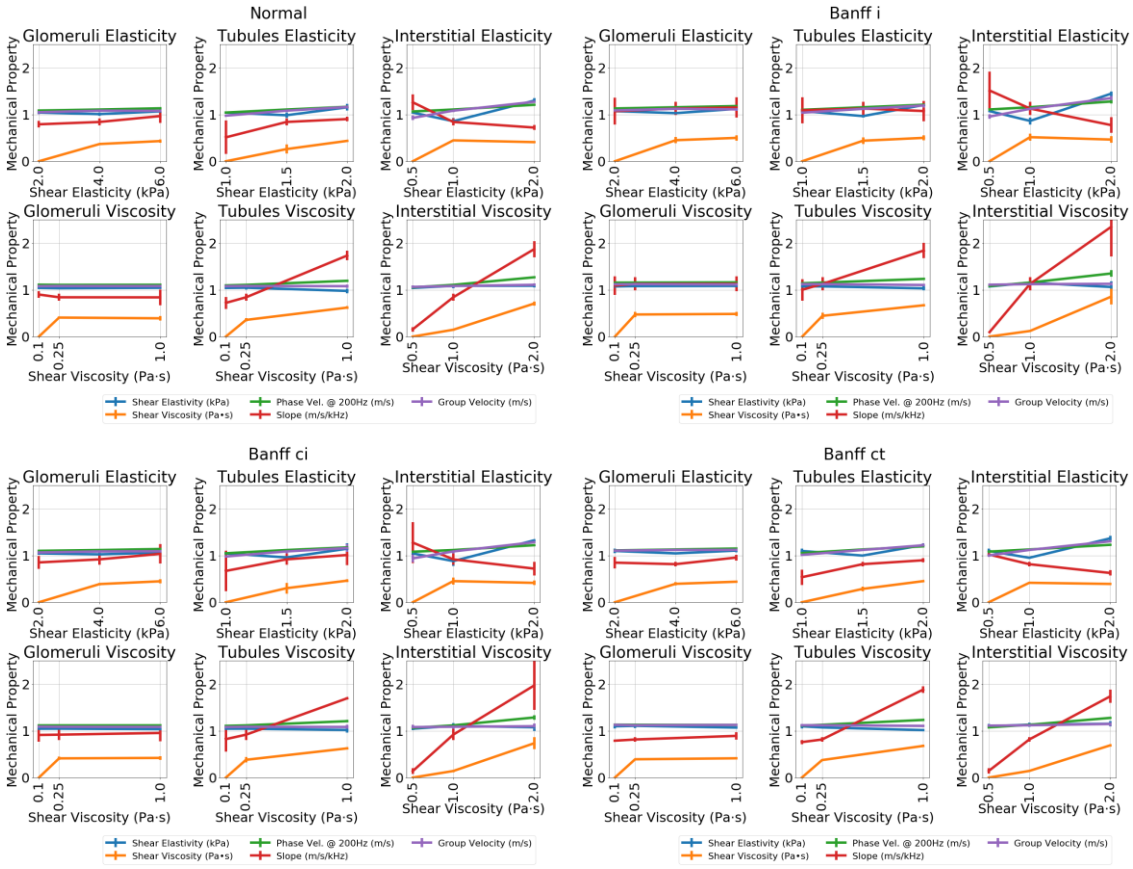


Figure 5-3 – Compiled results for each patient, iteration, and biopsy outcome. Each biopsy outcome cluster is composed by the six mechanical parameters that compose the cortex morphology: elasticity and viscosity; for the glomeruli, tubule, and interstitial space. The vertical bars display the standard deviation across the three patients that composed each biopsy outcome. Shear viscosity and slope showed the most sensitivity to rheological alterations, while alterations to tubules and interstitial space altered the measurements the most.

$$Injury\ \%\ Change = \frac{Injury_{avg} - Normal_{avg}}{Normal_{avg}} \tag{5-4}$$

To compare the simulation results with the measurements obtained *in vivo*, SWE evaluations were performed using a Logiq E9 ultrasound system acquisitions performed on the same patients prior to biopsy procedure. The Pearson correlation coefficient

between *in silico* and *in vivo* was calculated for the median values of each patient and the aggregated biopsy outcome. The results are displayed in Figure 5-5 and Table 5-4. The p-values for the *in silico* measurements as predictors of *in vivo* estimations was also calculated, all p-values were <0.001, aside from μ_1 with p-value of 0.09. For all parameters evaluated the in vivo estimations showed significantly higher values, particularly for μ_2 and dispersion slope at 6.72 and 6.85 average multiplication factor, respectively. Group velocity, intercept, and μ_1 had an average multiplication factor of 1.14 to 2.32. Due to its intrinsic challenges (e.g., operator variability, kidney position variability), the *in vivo* measurements also showed significant higher levels of variability when compared to simulation.

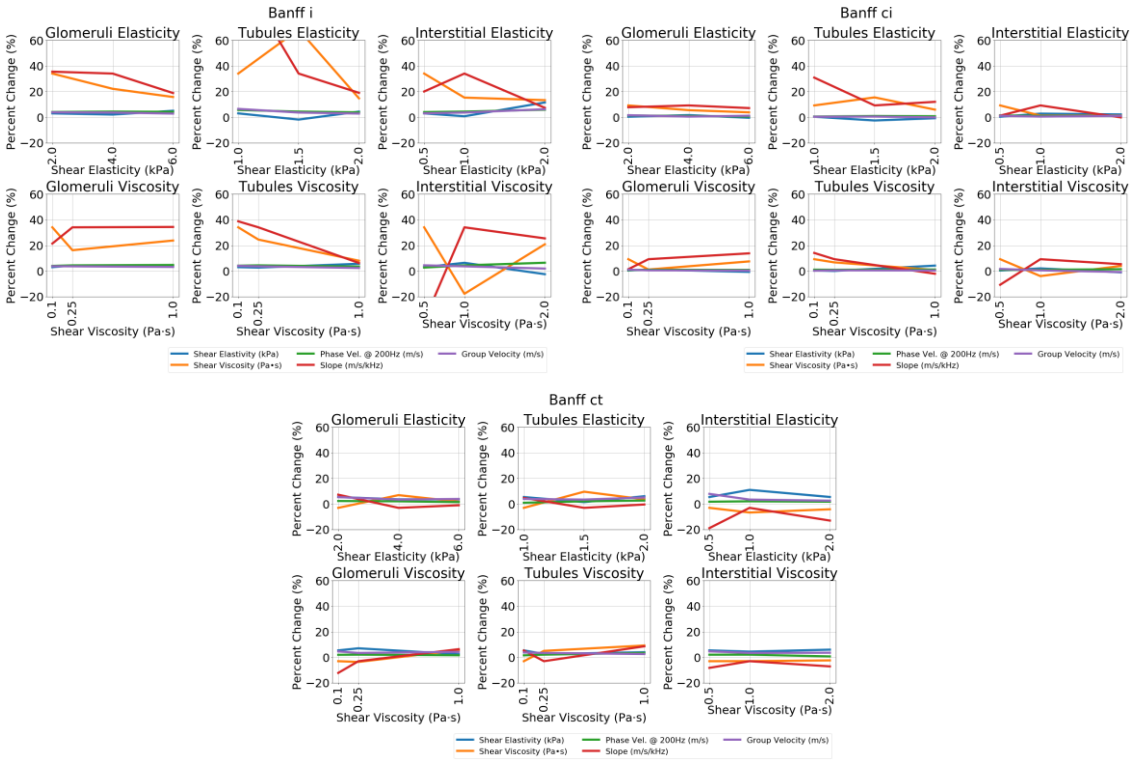


Figure 5-4 – Comparison of each diagnostic outcome depicted in Figure 5-3, against normal patients. The percent difference from each diagnostic outcome mean was calculated and plotted. Inflammation was found to affect the rheological measurements the most.

Although group velocity did not show correlation within the patient-wise median analysis, mean diagnostic group velocity showed high correlation between the *in silico* and *in vivo* cases, with Pearson correlation coefficient of 0.95 and p-value < 0.001. Therefore, group velocity variability impaired the patient-wise collinearity, but did not

confound the diagnostic outcome trend, where patients with inflammation and tubular atrophy tend to present higher group velocity estimations.

Table 5-4 – *in silico* and *in vivo* comparison

Patient outcome analysis										
Diagnostic/ Patient	**Group Velocity (m/s)**		**Phase Velocity at 200 Hz (m/s)**		**Slope (m/s/kHz)**		**Shear Elasticity (kPa)**		**Shear Viscosity (Pa·s)**	
	In silico	*In vivo*	*In silico*	*In vivo*	*In silico*	*In vivo*	*In silico*	*In vivo*	*In silico*	*In vivo*
Normal 1	1.13	1.92	1.12	2.23	0.84	6.16	1.07	1.16	0.40	2.48
Normal 2	1.05	2.62	1.09	2.08	0.75	5.22	1.02	1.93	0.38	2.89
Normal 3	1.05	2.41	1.13	2.22	1.00	6.39	1.03	1.06	0.44	2.83
Banff i 1	1.14	2.84	1.14	2.03	0.89	7.54	1.09	0.85	0.44	2.80
Banff i 2	1.11	2.62	1.13	2.10	1.00	5.91	1.06	0.88	0.45	3.00
Banff i 3	1.09	2.52	1.19	2.58	1.37	7.46	1.09	1.73	0.56	3.25
Banff ci 1	1.09	2.08	1.12	2.38	0.92	8.15	1.05	1.20	0.43	2.92
Banff ci 2	1.14	2.72	1.12	2.05	0.73	3.84	1.09	1.24	0.37	2.68
Banff ci 3	1.02	2.69	1.12	2.30	1.18	5.22	0.98	1.37	0.46	3.31
Banff ct 1	1.13	2.82	1.12	2.30	0.80	5.09	1.08	1.16	0.39	3.10
Banff ct 2	1.09	2.86	1.12	2.32	0.88	6.50	1.06	1.14	0.41	3.32
Banff ct 3	1.14	2.54	1.15	2.04	0.91	8.43	1.13	0.82	0.42	2.60
Pearson coefficient	0.01		0.43		0.35		-0.43		0.46	
Diagnostic outcome analysis										
Normal	1.08 ± 0.04	2.33 ± 0.30	1.11 ± 0.02	2.17 ± 0.07	0.87 ± 0.10	5.92 ± 0.51	1.04 ± 0.02	1.38 ± 0.39	0.41 ± 0.02	2.73 ± 0.18
Banff i	1.11 ± 0.02	2.66 ± 0.13	1.16 ± 0.03	2.24 ± 0.25	1.09 ± 0.21	6.97 ± 0.75	1.08 ± 0.01	1.15 ± 0.41	0.48 ± 0.05	3.02 ± 0.18
Banff ci	1.08 ± 0.05	2.50 ± 0.29	1.12 ± 0.00	2.24 ± 0.14	0.94 ± 0.19	5.74 ± 1.80	1.04 ± 0.04	1.27 ± 0.07	0.42 ± 0.04	2.97 ± 0.26
Banff ct	1.12 ± 0.02	2.74 ± 0.14	1.13 ± 0.02	2.22 ± 0.13	0.86 ± 0.05	6.67 ± 1.37	1.09 ± 0.03	1.04 ± 0.15	0.40 ± 0.01	3.01 ± 0.30
Pearson coefficient	0.95		0.58		0.52		-0.93		0.45	

The phase velocity at 200 Hz, dispersion slope and μ_2 showed similar correlation results with diagnostic-wise Pearson correlation coefficient of approximately 0.50. Patients with inflammation showed higher levels of shear wave dispersion than normal subjects, whereas ci and ct alterations where not as prominent, with both the measurements diverging within the standard deviation magnitude.

Although μ_1 showed the lowest scale factor, of approximately 1.14, μ_1 was the only measurement to show negative correlation, both patient and outcome-wise. The μ_1 estimation showed a high outcome-wise Pearson correlation, above 0.90, but with a negative trend.

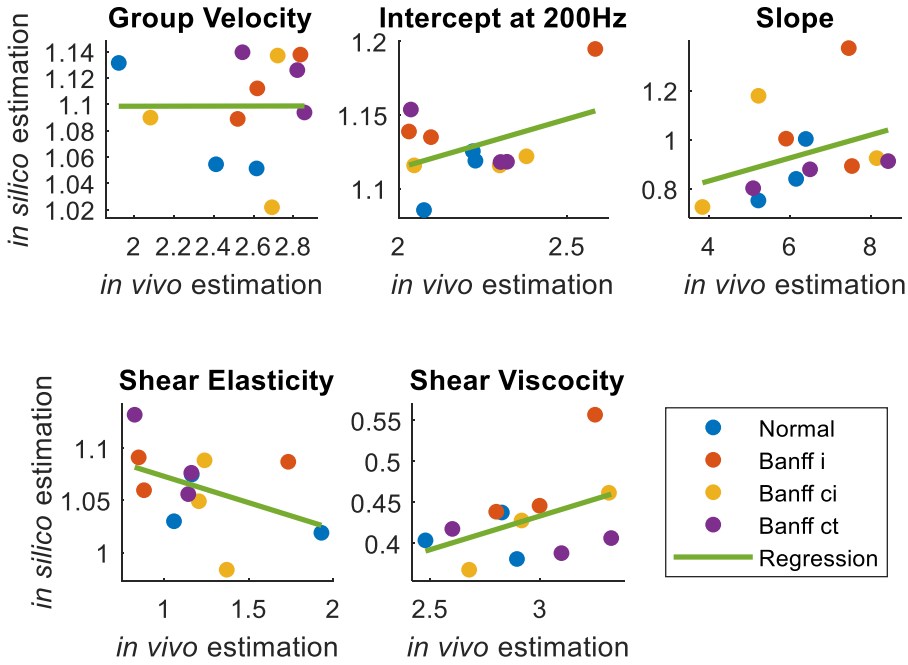


Figure 5-5 – Scatter plot of *in vivo* and *in silico* elastography estimations, linear regression is also displayed. The color of the dots decodes the patient respective diagnostic outcome.

5.4 Discussion

The results shown are encouraging and successfully demonstrated the capabilities of biopsy-based SGFD simulations. The shear wave speed maps were able to capture the micro-structure heterogeneity of the kidney cortex created by the biopsy masks. The spatial reconstruction resolution of the Logiq E9, as configured, was 0.3 mm, and therefore not capable of detecting micro-structures underlying the macroscopic rheological changes. For instance, the average glomeruli diameter in humans is 0.2 mm [163], and consequently not detectable by the Logiq E9 equipment. The use of simulations enable analysis with higher spatial resolution (0.05 mm), capable of assessing both micro- and macro-structural alterations in a controlled setting.

The average morphological composition for each diagnosis was calculated and displayed in Table 5-5. Inflammation differed the most from the other cases with higher interstitial and glomerular area (52.4% and 6.4%, respectively), and lower tubular and fluid relative area (34.1% and 6.8%, respectively). Tubular atrophy showed the highest tubular area, at 46%, and lower fluid composition, in agreement with the thickening of tubular basement membranes in addition to the tubular flow constriction [164]. Normal and interstitial fibrosis showed similar component distributions with an average deviation of ±1.6%.

Table 5-5 – Average spatial composition

	Avg_G%	Avg_T%	Avg_I%	Avg_Fl%
Normal	5.4%	39.1%	42.7%	12.8%
Banff i	6.4%	34.4%	52.4%	6.8%
Banff ci	4.3%	40.3%	44.8%	10.6%
Banff ct	4.4%	46.0%	40.9%	8.7%

G: Glomeruli, T: Tubules, I: Interstitial space, Fl: Fluid

Inflammation was found to alter the kidney composition the most, and therefore, change the cortex mechanical properties the most. This finding agreed with the results from the *in vivo* study described in Chapter 2, showing significantly higher values for all measurements, when comparing normal to Banff i patients. The increase in interstitial space in detriment of fluid and tubular composition is compatible with higher tissue stiffness, and viscosity, and therefore compatible with the elastography findings. Interstitial fibrosis and tubular atrophy did not appear to have such differentiation potential. The lack of patients with exclusively interstitial fibrosis in the subject pool for this study may have confounded Banff ci characterization, since ci positive patients also presented Banff ct.

The simulations showed, for the most part, consistent correlation with the *in vivo* studies performed. Although helpful for the understanding of the morphological alterations and its consequences, the simulation setting proposed in this study inherit several limitations that impair its capacity of numerically matching the *in vivo* assessments. We did not perform ultrasound detection of the moving tissue to avoid any potential introduced bias to the evaluations. The clinical setting also imposes different levels of variability, such as operator variability, not present in simulations. One of the main challenges, and the aim for future investigations, is the effect of perfusion to the elastography assessment. Several groups investigated these phenomena [150], [165]–[167], nevertheless rheological parameters for each specific cortex component *in vivo* and under perfusion were not

available at the time of development. Additionally, the models proposed did not account for the kidney cortex anisotropy, which can also induce bias. However, we only used SWE measurements in patients from the longitudinal plane of the kidney in the middle of the organ to reduce effects of anisotropy that may occur closer to the kidney's poles.

The only parameter to show negative correlation was μ_1. The *in vivo* data showed approximately 5 times more dispersion than *in silico*, even though with diagnostic-wise positive correlation above 0.45. The higher levels of dispersion tend to highly affect the shear elasticity estimation through Kelvin-Voigt model fitting. The increase of dispersion tends to decrease the dispersion curve zero intercept and, therefore, lower the shear elasticity estimation proportionally to the increase of slope/viscosity. The implementation of such simulations allows further investigation of these findings and artifacts and will be aims of future work.

Overall, the results shown agreed with those presented in Chapters 2 and 3. The inflammation cases were found to affect rheological characteristics the most, while group velocity and specific phase velocities showed the most sensitivity for inflammatory alterations. Inflammation diagnosis is critical for continuous allograft surveillance, as it is the main cause of more progressive injuries, such as fibrosis and tubular atrophy [98]. The use of SWE techniques might be a viable solution for more frequent allograft assessment.

5.5 Conclusion

In this study, we proposed to evaluate the simulation of heterogenous kidney cortex models based on human renal allograft biopsies. Staggered-grid finite-difference (SGFD) were leveraged to simulate shear wave propagation in 12 patient models. The results showed the method was capable of successfully simulate shear wave propagation in heterogeneous media based on segmented biopsy images. Although the elastography measurements, *in silico* and *in vivo*, diverged quantitatively, the results showed good correlation with the results obtained from the biopsy volunteers. Inflammation showed the most differentiation, with higher elasticity and viscosity levels when compared to the control group. Group velocity showed a Pearson diagnostic-wise correlation of over 0.9, although the same level was not found patient-wise. Intercept at 200 Hz, dispersion slope and shear viscosity showed reasonable correlation, above 0.40, whereas shear wave elasticity showed negative correlation between *in silico* and *in vivo* measurements. These

implementation gaps and limitations, such as anisotropy, should be addressed as the computation capabilities for more complex simulations advances.

Despite the disparities between *in silico* and *in vivo* limitations, the results were encouraging and present SGFD heterogeneous simulations as a viable testing framework for better understating on how the pathological changes to organs' micro-structure can alter the macro-structure rheological characteristics. Compared to interstitial fibrosis and tubular atrophy, interstitial inflammation was shown to alter the kidney cortex's morphology the most and therefore its mechanical properties the most both *in vivo* and *in silico*, this finding is particularly important as inflammation is a recurring injury that gives rise to fibrosis and tubular atrophy, and consequently, one of the main targets for more frequent allograft assessment.

Chapter 6 – Renal allograft interstitial inflammation detection based on convolutional neural networks and shear wave motion 2D Fourier transform

6.1 Introduction

Kidney function is essential to homeostasis and survival. Chronic kidney disease (CKD) is characterized by declining renal filtration and ultimately leads to end-stage renal disease (ESRD). Depending on the level of renal failure, the kidneys are no longer able to sustain homeostasis, and consequently, hemodialysis and transplantation are the standard course of treatment. Renal transplant is the preferred procedure as it increases patient survival and quality of life. Although transplantation outcomes have been improved since the introduction of the procedure, allografts can also be affected by rejection and other systemic disorders, such as hypertension. Renal allograft failure can reverse the benefits of transplantation eventually requiting hemodialysis and patient's return to transplant waiting lists [40].

Proper surveillance is critical to ensure allograft longevity and patient's quality of life. Biomarkers, such as serum creatinine (SCr), glomerular filtration rate (GFR) and resistive index (RI) have been used to assess kidney function and detect filtration anomalies [14], [96]. Although useful, these biomarkers are surrogates of kidney function, but not necessarily of kidney tissue state. The gold standard for kidney assessment is renal biopsy, as nephropathologists can directly evaluate tissue alterations and score tissue state using the Banff criteria, even before the renal function is affected [46], [100], [168]. Nevertheless, biopsies are an invasive procedure that incur higher clinical costs and can cause complications, such as hemorrhage, infection and ectasia [101], [105], [169].

Interstitial inflammation is the most common initial alteration presented during kidney rejection, although remediable, it can initiate scarring processes that can lead to interstitial fibrosis and tubular atrophy [98], [170]. It is, therefore, imperative to minimize inflammation processes with proper surveillance and treatment. Because common biomarkers cannot detect tissue alterations before function decline and biopsies cannot be performed frequently due to invasiveness, other non-invasive renal assessment modalities should be developed.

Over the past two decades, elasticity imaging techniques have been proven to be a viable non-invasive tool for tissue evaluation [21], [123]. Different pathological processes can alter the tissue morphological structure, that can translate to rheological alterations detectable by elasticity imaging techniques. Shear wave elastography (SWE) has emerged as a promising non-invasive and affordable tissue evaluation approach. The methodology leverages acoustic radiation force (ARF) pushes to generate shear waves that propagate through the investigated tissue. The propagation is then measured using high frame rate ultrasound image and reconstructed to estimate rheological parameters [171].

A variety of SWE parameters and have been leveraged for kidney assessment, from model-free estimations, such as group velocity, phase velocity and attenuation, for model-based estimations such as Kelvin-Voigt (KV) fitting [16]. Most of the techniques for rheological estimation rely on some form of time-to-frequency transformation, the most common methodology is the shear wave velocity dispersion calculation based on the two-dimensional Fourier transform (2D FT) of the shear motion in space and time. The Fourier frequency resolution is defined by the signal discretization characteristics (signal sampling frequency and length), while its bandwidth is defined by the Nyquist frequency (half the sampling frequency) alone. Even though, different acquisitions can present different discretization characteristics the 2D FT approach is able to output standardized frequency spectra. The frequency bandwidth can be truncated at the smallest Nyquist frequency available, and the frequency resolution can be adjusted by truncating or zero padding the signal to the required length. Analogous signal standardization in the original time-space dimensions require the use of interpolation techniques, that may introduce bias and variability to the signal.

The shear wave velocity dispersion, i.e., the velocity dependency on frequency, can then be evaluated for tissue characterization [78], [82], [132]. Although such measurements have been shown to correlate with the prevalence of interstitial inflammation and other alterations [64], [117]–[119], the quantification of rheological properties require complex post-processing techniques, such as directional filtering, noise filtering and curve fitting, that can introduce different levels of bias and variability, as they rely on premises, such as isotropy and linearity, that do not hold true, especially in the renal perspective [172], [173].

During the last decade, convolutional neural networks (CNNs) have been developed to model increasingly complex non-linear datasets, especially for imaging

tasks, such as computer vision and inclusion detection. The CNN implements a layer-based architecture where weighted kernel convolutions are performed to mimic a receptive field, which is then reduced by a pooling layer to extract the most relevant features. The output of multiple convolutional and pooling layer pairs is then fed through fully connected neural network to perform the final image interpretation [37]. The training objective is defined by the reduction of a given loss function defined by the task (e.g. binary cross-entropy for binary classification), the weights are then adjusted with the optimization task in mind, usually by leveraging stochastic gradient descent techniques [139], [140]. In this proof-of-concept study, we propose to evaluate the capabilities of CNNs to detect interstitial inflammation from shear wave 2D FT images with binary biopsy score supervision (Fig. 6-1).

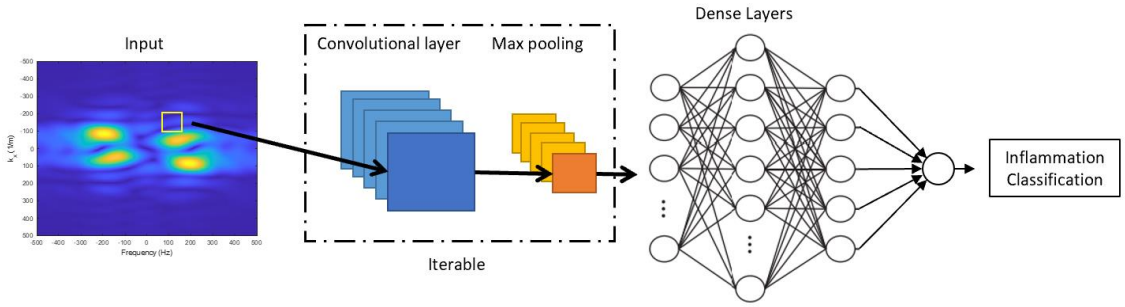


Figure 6-1 – General convolutional neural network architecture implemented for 2D FT image classification. Convolutional and max-pooling pairs can be stacked for deeper feature extraction. The feature extracted are resized into a one-dimensional array that is then processed by the fully connected layers to produce the image classification.

6.2 Methods

The dataset was taken from a cohort of 223 SWE studies performed with General Electric (GE) Logiq E9 ultrasound system (General Electric, Wauwatosa, WI, USA) prior to routine allograft biopsy, under the protocol approved by Mayo Clinic's Institutional Review Board (IRB# 11-003249). A small fraction of patients (n = 13) had studies performed at multiple routine visits, all 206 participants provided written informed consent. The inflammation biopsy scoring was performed and annotated by the Mayo Clinic's nephropathologists for supervised training. Each subject was scanned at lower pole, middle portion, and upper pole, in longitudinal and transverse planes. The data quality of each subject was evaluated using ratio of the interquartile range (IQR) and median (IQR-to-median ratio) for common elastography measurements, such as group velocity, phase

velocity and shear wave attenuation from 100 to 300 Hz, and KV shear elasticity and viscosity.

Table 6-1 – Study cohort demographics				
	Patients (n=206)		Studies (n=223)	
	Number	%	Number	%
Sex				
Male	122	59.2	128	57.4
Female	84	40.8	95	42.6
	Mean (Std. Dev)	Range	Mean (Std. Dev)	Range
Age, yrs	49.8 (14.0)*	22-76	49.7 (14.0)	22-76
BMI, kg/m^2	27.6 (5.3)*	13.1-42.3	27.9 (5.5)	13.1-42.3
TAT (months)	14.5 (14.9)*	4-60	16.1 (14.7)	4-60
*At first scan				

The dataset showed high skewness, with only 31 subjects with detected inflammation, therefore, majority under sampling was used to balance the dataset. All 31 positive subjects ware used for training, whereas the control subjects were ranked according to mean IQR-to-median ratio over the elastography measurements.

To produce the 2D FT images for training, a MATLAB (Mathworks, Natick, MA, USA) script was implemented to pre-process the shear wave motion images. To eliminate the signal's DC component, the mean value was subtracted from the shear motion field. Because boundary discontinuity can introduce additional spectral leakage to the 2D FT evaluation, a 2D Tukey window ($\alpha = 0.2$) was implemented prior to transform [174]. The 2D FT was applied to the spatiotemporal (x-t) image to obtain the 2D FT wavenumber (k_x)-frequency (f) image (Fig. 6-2). In all these cases, the magnitude of the 2D FT was used. Because the GE Logic E9 system applies two simultaneous ARF pushes, on each side of the region-of-interest (ROI), the motion data contains waves traveling in the positive x-direction and negative x-direction. As a result, the four quadrants were used to generate the training images, and therefore, eliminating the necessity of directional filtering. Although the sampling frequency was constant throughout the acquisitions, the ROIs were defined per acquisition by the operators, therefore, zero padding was employed to generate consistent 2D FT discretization in both axes. The distributions in the first and third quadrants represent the waves traveling in the positive x-direction and the distributions in the second and fourth quadrants represent the waves traveling in the negative x-direction. The distributions in these quadrants have complex conjugate symmetry.

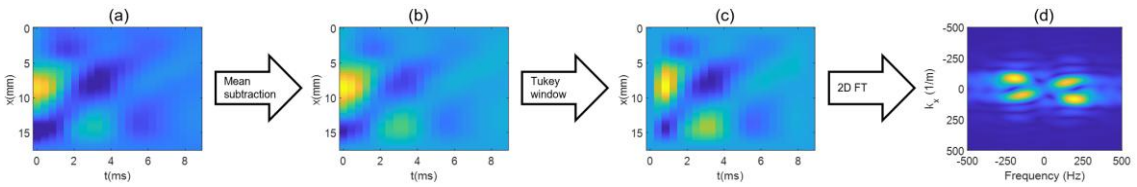


Figure 6-2 – (a) Original ROI shear motion, (b) shear motion with DC component removed, (c) shear motion after Tukey windowing, (d) final four quadrant 2D FT image used for training.

The CNN was implemented in Python (Python Software Foundation, Wilmington, DE, USA), using Keras [149] with a Tensorflow (Google Brain, Mountain View, CA, USA) backend. The models were trained on a Z280 workstation (Hewlett-Packard, Palo Alto, CA, USA) equipped with a GTX 1080 GPU (Nvidia, Santa Clara, CA, USA). All subjects positive for inflammation were loaded while, the control subjects with the lowest IQR-to-median ratio values were loaded according to the defined under sampling. The Banff inflammation scores from the biopsy evaluation equal or above 1 were defined as 1, for binary classification. All images loaded were then normalized and flipped vertically for dataset augmentation, consequently doubling the number of data samples. Group shuffle split was used to define the training, validation, and test sets (60/20/20), reducing the chance that overfitting occurs due to data leak, as different images from a single subject cannot be presents in multiple split sets.

The number and size of convolutional, dense and dropout [175] layers were varied to evaluate the models' performance using the Tensorboard API (Google Brain, Mountain View, CA, USA). Additionally, batch normalization was implemented for all layers. The training was set to a maximum of 1000 epochs and early stopping was implemented [175] with a minimum validation loss δ of 0.0001 over 20 epochs. The models' loss, accuracy and area under the Precision-Recall curve were recorded for each epoch. The Adam optimizer [140] was used to adjust the weights according to the binary cross-entropy loss. The best hyperparameter sets were selected for final testing and reported. The final test was performed on the segregated test set. The receiver operating characteristic curve and its area (AUROC) were recorded. To better evaluate the inflammation detection capabilities in different skewness settings the precision (Eq. 6-1), recall (Eq. 6-2) and F1-score (Eq. 6-3) at Youden's index optimal cutoff, were also recorded [176].

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}, \tag{6-1}$$

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}, \tag{6-2}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (6\text{-}3)$$

6.3 Results

The results showed the potential for interstitial inflammation detection using convolutional neural networks on the 2D FT images. No under sampling (6.2:1), 1:1 and 2:1 control/abnormal ratios were used to train different models. The original Banff score distribution in terms of patient and scans are displayed in Table 6-2.

Table 6-2 – Banff patient and scan score distribution

Score	Banff i (n=223)	
	Patients	%
0	192	86.10
1	17	7.62
2	10	4.48
3	4	1.79
> 0	31	13.90
Score	Scans	%
0	14024	86.26
1	1160	7.13
2	767	4.72
3	307	1.89
> 0	2234	13.74

The best architecture for 2:1 and 1:1 datasets was found to be 3 convolutional layers, with 16 nodes each, 2 dense layers with 16 nodes each, and dropout of 0.3 at each layer interface. The original 6.2:1 dataset was tuned to 3 convolutional layers, with 16 nodes each, 4 dense layers with 16 nodes each, and dropout of 0.1 at each layer interface.

The 1:1 and 2:1 ratio datasets were able to achieve F1-scores of 0.75 and 0.69, respectively. Although the 1:1 dataset managed to reach a high level of inflammation detection at 0.79 recall rate, it performed similarly to 2:1 in terms for specificity and precision, at 0.72 and 0.71 respectively. The high skewness of the dataset heavily confounded the original 6.2:1 dataset training with an AUROC of 0.57 and F1-score below 0.3. The performance metrics for each CNN model and the ROC curves are displayed in Table 6-1 and Figure 6-3, respectively.

Table 6-3 – CNN performance metrics

Skewness	AUROC	Cutoff	Specificity	Sensitivity/ Recall	Precision	F1-score
6.2:1	0.57	0.09	0.35	0.77	0.15	0.26
2:1	0.75	0.49	0.73	0.68	0.70	0.69
1:1	0.81	0.47	0.72	0.79	0.71	0.75

(a) 6.2:1 (b) 2:1 (c) 1:1

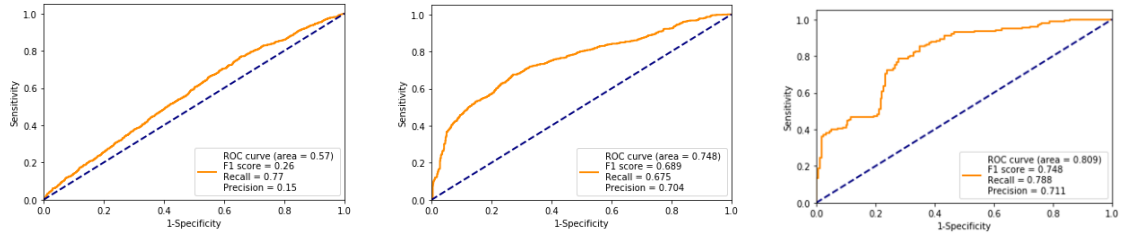


Figure 6-3 – ROC plots for 6.2:1 (a), 2:1 (b) and 1:1 (c) testing. The orange line is the ROC curve, while the dashed blue line is the chance performance (AUROC = 0.50).

6.4 Discussion

This proof-of-concept study showed that a CNN using 2D FT magnitude images was capable of detecting interstitial inflammation at similar levels presented in literature [63], [117], [119] and in our group previous studies, without resorting to complex image processing techniques such as directional filtering. Given the limited availability of kidney allograft datasets and the morphological complexity of kidney cortex, when compared to more homogeneous and isotropic organs, such as liver [78], [132], the results were encouraging and further investigation will be necessary to improve the methodology.

In this study, a more balanced dataset was achieved using a data quality driven majority under sampling, this is not ideal as it can diminish the generalizability of the trained model. The number of moderate to severe inflammation (Banff i > 1) was 6.27% of the total cohort, while mild (Banff i = 1) cases accounted for 7.62% of the patient cohort. The ongoing expansion of the patient study will allow the group to collect more moderate to severe cases, enhancing binary classification, as well as allowing for categorical classification. The models proposed would also benefit from a more diverse cohort of patients, with different demographic backgrounds and standard of care. Interstitial inflammation treatment can vary significantly depending on patients' secondary diagnostics and institutional protocols [177]. The expansion of this study to other institutions may be beneficial, as standards of care and patients' common underlying conditions (such as hypertension and obesity) may affect the observed severity distribution [178], [179].

Inter-institutional training imposes specific data standardization challenges. In such perspective, the 2D FT approach also holds potential for inter-institutional training. High levels of data processing can enhance data acquisition biases, as such techniques

often are equipment and methodology specific. The use of minimal processing and 2D FT images allows a more prompt adaptation of different acquisition profiles and facilitate transfer learning applications [153], i.e., different sampling frequencies and ROIs can be easily adjusted during Fourier transform. The use of all four 2D FT quadrants combined with CNNs' translation and rotation invariance may also be beneficial to accommodate different ARF profiles.

Even though scans at 3 different locations within kidney cortex were used for training, the biopsy was performed at a single location (usually the closest to surface). Both biopsy and this study leverage the premise of diffuse inflammation within kidney cortex, which may not hold true. In future work, this premise will be investigated with the use of the biopsy location annotation and employment of weak supervision learning techniques. In addition, classification uncertainty techniques [180] might also be implemented to produce easily interpretable uncertainty quantification to clinicians, and better support decision making. It is also important to note that other injuries such as tubular atrophy and interstitial fibrosis were also present, what might confound inflammation discrimination.

6.5 Conclusion

Ultrasound SWE has shown its potential for low-cost and non-invasive tissue assessment. The use of machine learning techniques, such as CNNs, for imaging applications have also showed increasingly performance when compared to well established processing techniques. The proof-of-concept study introduced a novel 2D FT CNN technique for kidney interstitial inflammation detection. The method was applied to a dataset of 31 subjects with interstitial inflammation and 31 to 192 control subjects (1:1, 2:1 and 6.2:1 control/inflammation ratios). The 1:1 balanced dataset was able to reach an AUROC of over 0.8 and a F1-score of 0.75, showing the potential of this technique for allograft inflammation detection. Further investigations will be performed to advance the technique and mitigate its limitations.

Chapter 7 – Summary and Future Work

7.1 Introduction

This thesis objectives were to explore machine learning (ML) capabilities for kidney allograft biopsy classification based on *in vivo* and *in silico* shear wave elastography (SWE) evaluations. The use of SWE for non-invasive and affordable tissue assessment has been researched for decades, but well established clinical implementations are still limited. The kidney morphological characteristics add to the challenges SWE wider clinical adoption of renal assessment. Given that the kidney is highly heterogeneous, anisotropic and non-linear, machine learning applications might be able to better model the renal complexity without the common homogeneous, isotropic and linear assumptions of classical mathematical models, or by intrinsically modelling the biases and noise introduced by such assumptions.

For *in vivo* evaluation, data from a patient study performed at Mayo Clinic from October 2017 to July 2019 was leveraged. Each transplanted patient was scanned with a SWE capable ultrasound equipment prior to biopsy procedure. Six different cortical positions were evaluated; upper pole, lower pole and middle region at longitudinal and transverse orientations. The biopsies were evaluated by Mayo Clinic's nephropathologists, and the degree of interstitial inflammation, fibrosis and tubular atrophy was annotated to the dataset. Other clinical biomarkers were also recorded, such as blood pressure, serum creatinine and resistive index.

In Specific Aim 1, we investigated the ML capabilities as well as classical statistical modalities, such as general estimation equations, to model and classify renal allograft biopsy scores based on stablished SWE measurements and stablished protocol biomarkers. Specific Aim 2 investigated the correlation between the morphological alterations caused by the pathological processes common to allograft rejection, and how they affect the rheological characteristics of cortical tissue as perceived by SWE measurements. Specific Aim 2 also investigated the capabilities of convolutional neural networks for modelling the viscoelastic measurements (regression task), and biopsy outcomes (classification task), based on shear wave motion, and 2D FT images, respectively.

7.2 Summary of Results

7.2.1 Specific Aim 1

Develop supervised machine learning models that can predict renal transplant biopsy scores based on SWE measurements and other physiological biomarkers with >80% specificity and sensitivity.

In Chapter 2, the *in vivo* dataset was used to investigate the significance of a variety of common SWE measurements. Group velocity, shear modulus, Kelvin-Voigt dispersion curve fitting (μ_1 and μ_2), 2D FT phase velocity, shear wave attenuation using AMUSE, and GST-SFK (phase velocity) were used quantify the rheological properties of the kidney. Clinical biomarkers, such as, serum creatinine and eGFR, were also used for allograft assessment. The presence of inflammation, interstitial fibrosis and tubular atrophy were evaluated and annotated by nephropathologists.

Attenuation measurements showed strong significance (p-value < 0.05) for tubular atrophy with negative correlation (0.97 ≤ OR < 1.00), while velocity associated measurements showed mild significance (p-value < 0.10) for inflammation and fibrosis with positive correlation (1.00 < OR ≤ 1.02).

Time after transplant (TAT) and Age showed strong statistical significance for interstitial fibrosis and tubular atrophy (independently and combined as IFTA) with a positive correlation, but showed no significance for inflammation, in agreement with the chronic nature of inflammation when compared to the more progressive and acute characteristics of fibrosis and tubular atrophy.

The eGFR also showed significance with inflammation in some datasets, with negative correlation, and serum creatinine (SCr) showed strong significance for tubular atrophy for all datasets tested, with positive correlation, both in agreement with the declining kidney function cause by the injuries evaluated.

Overall, the longitudinal middle position performed the best with the logistic regression tests, with the best classification performance of AUROC = 0.72 for tubular atrophy when fit to low p-value parameters only. Inflammation and interstitial fibrosis also benefited the most form the L_MID/low p-value setting, with AUROCs of 0.63 and 0.66, respectively.

In Chapter 3, the patient dataset was evaluated for biopsy binary classification using the support vector machine (SVM). For model interpretability Shapley explanations (SHAP) was implemented. Several models reached AUROCs > 0.70, a level of performance compatible with previous work performed by our group and others [17], [21], [22], [33]–[36]. The SHAP importance analysis agreed with the generalized estimating equations (GEE) and odds ratio analysis performed in Chapter 2, showing the reliability of SVM for model classification and interpretability.

Inflammation was the least frequent injury present in the patient study, with less than 14% prevalence. Therefore, the models significantly benefited from majority class undersampling, the use of the IQR-to-median ratio to retain the most reliable majority samples also prove to be a viable technique for mitigating skewness. Inflammation was the most detectable injury by elastography measurements, demonstrating classification AUROC > 0.70 even when biomarkers were not present. As noted in Chapter 2, temporal biomarkers TAT and Age do not give insight into the inflammation prediction. Diastolic blood pressure also showed high importance for inflammation detection, with a positive trend, suggesting that hypertensive patients might be more associated with having transplant inflammation. Further investigation on inflammation prediction is necessary, which would involve studies with more patients with moderate and severe allograft inflammation are imperative for the advancement of the proposed techniques.

Besides inflammation, tubular atrophy was the only diagnostic outcome to reach AUROC > 0.70, at a maximum of AUROC = 0.74. Although TAT and Age showed higher importance, as expected, the models benefitted from the addition of viscoelastic measurements, especially 2D FT shear wave attenuation at 200 Hz, α_{2DFT}^{200}, which showed a negative trend, the decrease in attenuation is consistent with the interstitial inflammation process, where interstitial contents expand and compress, decreasing the overall attenuation of the tissue [23][38][39]. The eGFR also demonstrated high importance in multiple models for the different outcomes, its negative trend relates directly to the allograft function, as the injury progresses the filtration capacity of the kidney tends to decline [45].

Unfortunately, the 80% sensitivity and specificity targets were not able to be reached. Logistic regression reached a maximum of AUROC = 0.72 for fibrosis classification, with sensitivity and specificity below 0.65 at optimal Youden's index. Best classification performance was found with the SVM implementation, with sensitivity and specificity for AUROCs > 0.7 varied from 0.58-0.77 and 0.53-0.84, at optimal Youden's index. As indicated in Chapter 2, the cutoff for classification can be adjusted to favor

sensitivity or specificity. Lower cutoff will increase sensitivity in detriment of specificity. This solution is not ideal as false positive rate will increase, but can be leveraged in primary care setting, as indication for further investigation.

7.2.2 Specific Aim 2

Develop supervised machine learning models that can predict renal transplant biopsy scores based on analysis of data from shear wave propagation simulation and patient data with minimal pre-processing.

In Chapter 4, shear wave motion simulations were created using staggered-grid finite-difference (SGFD) method. The shear elasticity and viscosity ranges for the KV model were chosen so they would cover a wide range of values for normal and pathological soft tissues. The shear elasticity, μ_1, ranged from 1-25 kPa with increments of 1 kPa, and shear viscosity, μ_2, ranged from 0-10 Pa·s with increments of 0.5 Pa·s. Two different ARF excitation focus profiles were used, with f-numbers (F/N = focal length/width of aperture) 1 and 2 with a focal distance of z_f = 2.5 cm, totaling 1050 simulations.

The results shown in Chapter 4 were encouraging and exhibit the potential of the use of convolutional neural networks in ultrasound SWE for viscoelastic characterization. The results demonstrated that the CNN models were robust to noise, vertical position and partially to F/N. When tested on datasets with the same excitation F/N for training and testing, the elasticity and viscosity estimations were very good and showed robustness to noise and vertical position of the measurement.

The testing results were evaluated using the MAE between the true value of the viscoelastic property and the CNN prediction output. A linear regression was performed to compare the results from the 2D FT analysis as well as the CNN prediction to the identity line, and the slope, intercept and R^2 of the regression line were also calculated and reported. All elasticity models trained and tested on the same F/N showed very good fitting with MAE below 0.079 kPa with slopes and intercept of approximately one and zero, respectively. The viscosity models trained and tested on the same f-numbers also showed very good agreement with MAE below 0.091 Pa·s with slopes and intercept of

approximately one and zero, respectively. Alternatively, CNN models tested on f-numbers different from the training/validation F/N had diminished performance.

When trained with both F/Ns the CNNs were able perform similarly as the pairs trained, validated, and tested with a single F/N. The results indicate that the CNNs can discern between wave front profiles with high accuracy, as long as they are trained properly. Overall, the proof-of-concept study showed that the CNN models were able to successfully retrieve the viscoelastic properties from the wave motion images. The approach proposed required minimal preprocessing, without the employment of filters or transforms.

In Chapter 5, heterogenous kidney cortex models based on human renal allograft biopsies were evaluated. Staggered-grid finite-difference (SGFD) were leveraged to simulate shear wave propagation in 12 patient models based on segmented biopsy images. Although the elastography measurements, *in silico* and *in vivo*, diverged quantitatively, the results showed good correlation with the results obtained from the transplant patients. Inflammation showed the most differentiation, with higher elasticity and viscosity levels when compared to the control group. Group velocity showed a Pearson correlation coefficient of over 0.90 when considering correlations within a diagnostic outcome. Intercept at 200 Hz, dispersion slope and shear viscosity showed reasonable correlation, above 0.40, whereas shear wave elasticity showed negative correlation.

Despite the disparities between *in silico* and *in vivo* implementations, the results were encouraging and present SGFD heterogeneous simulations as a viable testing framework for better understating on how the pathological changes to organs' micro-structure can alter the macro-structure rheological characteristics. Compared to interstitial fibrosis and tubular atrophy, interstitial inflammation was shown to alter the kidney cortex's morphology and rheology the most, this finding is particularly important as inflammation is a recurring injury that gives rise to fibrosis and tubular atrophy, and consequently, an important target for more frequent allograft assessment.

In Chapter 6, the 2D FT CNN was evaluated for inflammation classification. The architecture proposed was able to reach AUROCs above the performances reported in Chapter 2 and 3 for balanced inflammation datasets (AUROC = 0.81). The proof-of-concept study proved it is possible to obtain better classification performance from less pre-processed datasets, although further investigations are necessary to further advance the technique, including the expansion of the current dataset.

The objectives proposed by Specific Aim 2 were concluded, not only with the introduction of novel methodologies but also with valuable insights on the usefulness of the combination of SWE and ML, and the association between pathological and rheological alterations.

7.3 Limitations

Chapters 2, 3, 5 and 6 used data from the *in vivo* patient study. The patient study includes several limitations in terms of operator variability. To mitigate inter-operator variability all sonographers have performed the same training for SWE capable ultrasound applications. The level of compression applied by the sonographer during scan can modify kidney stiffness. All sonographers were instructed to apply minimal compression as possible and be cognizant of allograft displacement at B-mode imaging. Although beyond the scope of this research, have been investigated by our group [77].

The complex kidney morphology also imposes specific limitations, the common assumptions of isotropy and homogeneity do not hold, especially for renal tissue. To account for anisotropy the scans were performed in both transverse and longitudinal planes, and, during ML training with elastography parameters, the location of scan was always provided so the SVM could possibly model the biases introduced by fiber orientation. The restriction to only longitudinal middle position also proved a viable mitigation strategy, as it had the best IQR-to-median ratio results.

All biopsy labeling was performed by Mayo Clinic's pathology department. All nephropathologists use the well-established Banff criteria for injury scoring, therefore inter-operator variability within the label annotation process was deemed minimal. For future studies voting processes can be implemented to further reduce variability, although such strategies were beyond the scope of this thesis and not included in Mayo's Clinic protocol transplant evaluations.

The patient dataset is also limited by the demographics of patient cohort, all patients enrolled in the research protocol received treatment a single institution, therefore, the lack of diversity in standard of care might have limited the variability of patient outcomes. The number of patients with moderate-to-severe (Banff > 1) injury were below 6% for all injuries evaluated, which can hinder classification performance. Interstitial fibrosis and tubular atrophy had 43.1% and 67.7% mild-to-severe (Banff > 0) rate, respectively, whereas interstitial inflammation had only 13.9% mild-to-severe prevalence.

Highly skewed datasets can hinder ML training, therefore, balanced datasets were created using the IQR-to-median ratio, for all inflammation evaluations.

The elastography quantification techniques implemented in Specific Aim 1 can also introduce different levels of variability to the dataset, especially in noisy environments such *in vivo* applications. The use of ML to reduce the data processing overhead was investigated in Chapter 4 and 6, by quantitatively and qualitatively evaluating *in silico*, and *in vivo* datasets with minimal pre-processing.

Chapters 4 and 5 leveraged SGFD simulations. Due to computational resource restrictions, ultrasound detection was not implemented to the simulated data, which may add bias to elastography measurements. Although Gaussian noise was added to mitigate noise discrepancies between *in silico* and *in vivo* data, it does not replicate all ultrasound noise characteristics and can further introduce bias.

The homogeneous assumption limitations were investigated in Chapter 5. The SGFD method was capable of reproducing isotropic cortical heterogeneity. The results showed correlation between heterogeneous *in silico* and *in vivo* elastography estimations, however it was not possible to match the results numerically. The *ex vivo* evaluations used for glomeruli, tubule and interstitial mechanical property assignment did not take perfusion into consideration. Perfusion affects cortical rheology, as demonstrated by our group in previous research [166], and should be taken under consideration. The perfusion analysis was beyond the scope of the proof-of-concept study introduced, nonetheless, the framework implemented can be used for perfusion investigation, due to the flexibility of assigning renal cortex viscoelastic parameters individually.

7.4 Future Research

The patient study that produced the data used in Chapters 2, 3, 5 and 6 is still active, therefore, more data have become available. Dr. Urban's group will further advance the investigation proposed in this thesis with the newly acquired data, possibly mitigating inflammation skewness characteristics. The study has expanded to other Mayo Clinic sites, in Florida and Arizona, to augment data diversity.

The simulation framework introduced in Chapter 5 will also be leveraged for further renal cortex parametric analysis. Future research will be devoted to adjusting the rheological parameters to account for perfusion alterations and better understand not only

the morphological changes to the micro-structure, but also the individual component (i.e., glomeruli) rheological alterations.

Chapter 6 introduced a novel 2D FT data representation for CNN training. With the advance of the patient study more data will be available for training and other architectures may be evaluated. The current implementation assumes the injuries present at the biopsy location are present in other regions of the kidney. The group also envision the implementation of weak learning techniques [181], by assigning the scan locations near to biopsy site as strong labels, and other locations as weak labels for better performance and generalization.

The content in Chapter 4 was published in Computers in Medicine and Biology, and the content in Chapters 2, 3, 5 and 6 are, at time of thesis submission, under co-authors' review for future submission as peer-reviewed articles.

7.5 Conclusion

This thesis work consisted of investigating the renal allograft assessment capabilities of shear wave elastography combined with machine learning algorithms. The group also investigated the explainability of the proposed support vector machine algorithms and showed that machine learning agrees with classic statistical approaches such as generalized estimating equations while delivering better performance.

Convolutional neural networks architectures were leveraged to estimate viscoelastic parameters from homogeneous shear wave motion simulation, as well as predict allograft interstitial inflammation based on two-dimensional Fourier images. Additionally, annotated biopsy slides were used to construct heterogeneous staggered-grid finite-difference and investigate the interaction between the morphological alterations caused by the rejection processes, and the rheological changes perceived by shear wave elastography.

Interstitial inflammation was shown to be the best target for shear wave elastography evaluation, with higher levels cortical morphological and rheological alterations. Recurring inflammation is known to originate common scarring injuries such as interstitial fibrosis and tubular atrophy, and therefore, inflammation assessment and treatment is imperative to extent allograft longevity. Shear wave elastography and machine learning have the potential to be important tools for allograft surveillance and better patient care.

We envision that the novel techniques introduced in this thesis will allow clinicians to better understand the ML capabilities and how they relate to SWE and tissue pathology. The level of confidence introduced by novel technologies is equally important as performance indicators. We hope to strengthen the interpretability, reliability and performance of ML and SWE alike, so, in the future, SWE can be widely implemented as a reliable, cheap, fast and non-invasive methodology allograft diagnostic.

7.7 Academic Achievements

7.7.1 Peer-reviewed Papers

Vasconcelos, L.; Kijanka, P.; Grande, J.; Urban, M. W.; *Viscoelastic parameter estimation using simulated shear wave motion and convolutional neural networks*; Computers in Biology and Medicine 2021 Jun; 133:104382. doi: 10.1016/j.compbiomed.2021.104382. Epub 2021 Apr 11.

Vasconcelos, L.; Kijanka, P.; Mandrekar, J.; Grande, J.; Amador, C.; Aristizabal, S.; Oliveira, P.; Rule, A.; Atwell, T.; Urban, M. W.; *Kidney allograft assessment using shear wave elastography and clinical biomarkers.* In preparation.

Vasconcelos, L.; Kijanka, P.; Grande, J.; Amador, C.; Aristizabal, S.; Oliveira, P.; Rule, A.; Atwell, T.; Urban, M. W.; *Interpretable machine learning models for kidney biopsy score prediction based on shear wave elastography.* In preparation.

Vasconcelos, L.; Kijanka, P.; Grande, J.; Amador, C.; Aristizabal, S.; Oliveira, P.; Rule, A.; Atwell, T.; Urban, M. W.; *Kidney cortex shear wave motion simulations based on segmented biopsy slides.* In preparation

Vasconcelos, L.; Kijanka, P.; Amador, C.; Aristizabal, S.; Oliveira, P.; Urban, M. W.; *Renal allograft interstitial inflammation detection based on convolutional neural networks and shear wave motion 2D Fourier transform.* In preparation.

Kijanka, P.; **Vasconcelos, L.;** Urban, M. W.; *Evaluation of robustness of S-transform based phase velocity estimation in viscoelastic phantoms and renal transplants.* Submitted to Acta Biomaterilia.

7.7.2 Conference Abstracts

Vasconcelos, Luiz; Kijanka, Piotr; Grande, Joseph; Urban, Mattew; *Shear wave motion simulations of renal cortex based on histological morphology*. IEEE 2021 International Ultrasonics Symposium, Xian, China, Poster presentation.

Vasconcelos, Luiz; Kijanka, Piotr; Urban, Mattew; *Viscoelastic Parameter Estimation Using Shear Wave Elastography and Convolutional Neural Networks*, IEEE 2021 International Symposium on Biomedical Imaging, Nice, France, Poster presentation.

Vasconcelos, Luiz; Amador, Carolina; Aristizabal, Sara; Urban, Matthew; Nenadic, Ivan; *Evaluating Kidney Biopsy Prediction Using Shear Wave Elastography and Machine Learning Techniques and Shapley Explanation Techniques*. IEEE 2020 International Ultrasonics Symposium, Las Vegas, USA, Oral presentation.

Vasconcelos, Luiz; Amador, Carolina; Aristizabal, Sara; Urban, Matthew; Nenadic, Ivan; Kidney biopsy classification using shear wave elastography and machine learning techniques. IEEE 2019 International Ultrasonics Symposium, Glasgow, Scotland, Oral presentation.

Vasconcelos, Luiz; Amador, Carolina; Aristizabal, Sara; Urban, Matthew; Nenadic, Ivan; *Kidney biopsy score prediction based on shear wave elastography measurements and machine learning*. In: IEEE 2018 International Ultrasonics Symposium, Kobe, Japan, Poster presentation.

7.7.3 Fellowships and Awards

Bioinformatics and Computational Biology Fellowship
University of Minnesota Rochester
2018-2019 Academic year

Bibliography

[1] R. Saran, Y. Li, B. Robinson, K. C. Abbott, L. Y. C. Agodoa, J. Ayanian, J. Bragg-Gresham, R. Balkrishnan, J. L. T. Chen, E. Cope, P. W. Eggers, D. Gillen, D. Gipson, S. M. Hailpern, Y. N. Hall, K. He, W. Herman, M. Heung, R. A. Hirth, D. Hutton, S. J. Jacobsen, K. Kalantar-Zadeh, C. P. Kovesdy, Y. Lu, M. Z. Molnar, H. Morgenstern, B. Nallamothu, D. V. Nguyen, A. M. O'Hare, B. Plattner, R. Pisoni, F. K. Port, P. Rao, C. M. Rhee, A. Sakhuja, D. E. Schaubel, D. T. Selewski, V. Shahinian, J. J. Sim, P. Song, E. Streja, M. Kurella Tamura, F. Tentori, S. White, K. Woodside, and R. A. Hirth, "US renal data system 2015 annual data report: epidemiology of kidney disease in the united states," *Am. J. Kidney Dis.*, vol. 67, no. 3 Suppl 1, p. Svii, S1-305, Mar. 2016, doi: 10.1053/j.ajkd.2015.12.014.

[2] J. H. Wang and A. Hart, "Global perspective on kidney transplantation: united states," *Kidney360*, vol. 2, no. 11, pp. 1836–1839, Nov. 2021, doi: 10.34067/KID.0002472021.

[3] K. Solez, R. A. Axelsen, H. Benediktsson, J. F. Burdick, A. H. Cohen, R. B. Colvin, B. P. Croker, D. Droz, M. S. Dunnill, and P. F. Halloran, "International standardization of criteria for the histologic diagnosis of renal allograft rejection: the Banff working classification of kidney transplant pathology," *Kidney Int.*, vol. 44, no. 2, pp. 411–422, Aug. 1993, doi: 10.1038/ki.1993.259.

[4] D. S. Marwah and S. M. Korbet, "Timing of complications in percutaneous renal biopsy: What is the optimal period of observation?," *American Journal of Kidney Diseases*, vol. 28, no. 1, pp. 47–52, Jul. 1996, doi: 10.1016/S0272-6386(96)90129-8.

[5] A. S. Levey, L. A. Stevens, C. H. Schmid, Y. L. Zhang, A. F. Castro, H. I. Feldman, J. W. Kusek, P. Eggers, F. Van Lente, T. Greene, J. Coresh, and CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration), "A new equation to estimate glomerular filtration rate," *Ann Intern Med*, vol. 150, no. 9, pp. 604–612, May 2009, doi: 10.7326/0003-4819-150-9-200905050-00006.

[6] "Chapter 1: Definition and classification of CKD," *Kidney Int Suppl*, vol. 3, no. 1, pp. 19–62, Jan. 2013, doi: 10.1038/kisup.2012.64.

[7] K. M. Gooding, C. Lienczewski, M. Papale, N. Koivuviita, M. Maziarz, A.-M. D. Andersson, K. Sharma, P. Pontrelli, A. G. Hernandez, J. Bailey, K. Tobin, V.

Saunavaara, A. Zetterqvist, D. Shelley, I. Teh, C. Ball, S. Puppala, M. Ibberson, A. Karihaloo, K. Metsärinne, R. E. Banks, P. S. Gilmour, M. Mansfield, M. Gilchrist, D. de Zeeuw, H. J. L. Heerspink, P. Nuutila, M. Kretzler, M. W. Smith, L. Gesualdo, D. Andress, N. Grenier, A. C. Shore, M. F. Gomez, S. Sourbron, and for the Bea.-D. Consortium, "Prognostic imaging biomarkers for diabetic kidney disease (iBEAt): study protocol," *BMC Nephrology*, vol. 21, 2020, doi: 10.1186/s12882-020-01901-x.

[8] N. Grenier, P. Merville, and C. Combe, "Radiologic imaging of the renal parenchyma structure and function," *Nat Rev Nephrol*, vol. 12, no. 6, pp. 348–359, Jun. 2016, doi: 10.1038/nrneph.2016.44.

[9] R. Köhnke, D. Kentrup, K. Schütte-Nütgen, M. Schäfers, U. Schnöckel, V. Hoerr, and S. Reuter, "Update on imaging-based diagnosis of acute renal allograft rejection," *Am J Nucl Med Mol Imaging*, vol. 9, no. 2, pp. 110–126, 2019.

[10] C. Sebastià, S. Quiroga, R. Boyé, C. Cantarell, M. Fernandez-Planas, and A. Alvarez, "Helical CT in renal transplantation: normal findings and early and late complications," *Radiographics*, vol. 21, no. 5, pp. 1103–1117, Oct. 2001, doi: 10.1148/radiographics.21.5.g01se131103.

[11] R. Rosenbaum, P. E. Hoffsten, R. J. Stanley, and S. Klahr, "Use of computerized tomography to diagnose complications of percutaneous renal biopsy," *Kidney Int*, vol. 14, no. 1, pp. 87–92, Jul. 1978, doi: 10.1038/ki.1978.93.

[12] J.-P. Laissy, J.-M. Idée, P. Fernandez, M. Floquet, F. Vrtovsnik, and E. Schouman-Claeys, "Magnetic resonance imaging in acute and chronic kidney diseases: present status," *Nephron Clin Pract*, vol. 103, no. 2, pp. c50-57, 2006, doi: 10.1159/000090609.

[13] H. Hricak, C. Cruz, R. Romanski, M. H. Uniewski, N. W. Levin, B. L. Madrazo, M. A. Sandler, and W. R. Eyler, "Renal parenchymal disease: sonographic-histologic correlation," *Radiology*, vol. 144, no. 1, pp. 141–147, Jul. 1982, doi: 10.1148/radiology.144.1.7089245.

[14] C. McArthur, C. C. Geddes, and G. M. Baxter, "Early measurement of pulsatility and resistive indexes: correlation with long-term renal transplant function," *Radiology*, vol. 259, no. 1, pp. 278–285, Apr. 2011, doi: 10.1148/radiol.10101329.

[15] A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. B. Fowlkes, and S. Y. Emelianov, "Shear wave elasticity imaging: a new ultrasonic technology of medical

diagnostics," *Ultrasound Med Biol*, vol. 24, no. 9, pp. 1419–1435, Nov. 1998, doi: 10.1016/s0301-5629(98)00110-0.

[16] M. W. Urban, A. D. Rule, T. D. Atwell, and S. Chen, "Novel uses of ultrasound to assess kidney mechanical properties," *Kidney360*, vol. 2, no. 9, pp. 1531–1539, Sep. 2021, doi: 10.34067/KID.0002942021.

[17] V. Lukenda, I. Mikolasevic, S. Racki, I. Jelic, D. Stimac, and L. Orlic, "Transient elastography: a new noninvasive diagnostic tool for assessment of chronic allograft nephropathy," *Int Urol Nephrol*, vol. 46, no. 7, pp. 1435–1440, Jul. 2014, doi: 10.1007/s11255-014-0697-y.

[18] T. Syversveen, K. Brabrand, K. Midtvedt, E. H. Strøm, A. Hartmann, J. A. Jakobsen, and A. E. Berstad, "Assessment of renal allograft fibrosis by acoustic radiation force impulse quantification--a pilot study," *Transpl. Int.*, vol. 24, no. 1, pp. 100–105, Jan. 2011, doi: 10.1111/j.1432-2277.2010.01165.x.

[19] N. P. Ghonge, M. Mohan, V. Kashyap, and S. Jasuja, "Renal allograft dysfunction: evaluation with shear-wave sonoelastography," *Radiology*, vol. 288, no. 1, pp. 146–152, Jul. 2018, doi: 10.1148/radiol.2018170577.

[20] A. Bueno Jimenez, L. Larreina, J. Serradilla, F. de Borja Nava, R. Lobato, S. Rivas, P. Lopez-Pereira, L. García, L. Espinosa, and M. J. Martinez-Urrutia, "Upside-down kidney placement: An alternative in pediatric renal transplantation," *J Pediatr Surg*, vol. 56, no. 8, pp. 1417–1420, Aug. 2021, doi: 10.1016/j.jpedsurg.2020.09.019.

[21] A. Manduca, P. J. Bayly, R. L. Ehman, A. Kolipaka, T. J. Royston, I. Sack, R. Sinkus, and B. E. Van Beers, "MR elastography: Principles, guidelines, and terminology," *Magn Reson Med*, vol. 85, no. 5, pp. 2377–2390, May 2021, doi: 10.1002/mrm.28627.

[22] L. Warner, M. Yin, K. J. Glaser, J. A. Woollard, C. A. Carrascal, M. J. Korsmo, J. A. Crane, R. L. Ehman, and L. O. Lerman, "Noninvasive In vivo assessment of renal tissue elasticity during graded renal ischemia using MR elastography," *Invest Radiol*, vol. 46, no. 8, pp. 509–514, Aug. 2011, doi: 10.1097/RLI.0b013e3182183a95.

[23] S. Singh, S. K. Venkatesh, A. Keaveny, S. Adam, F. H. Miller, P. Asbach, E. M. Godfrey, A. C. Silva, Z. Wang, M. H. Murad, S. K. Asrani, D. J. Lomas, and R. L. Ehman, "Diagnostic accuracy of magnetic resonance elastography in liver transplant recipients: A pooled analysis," *Ann Hepatol*, vol. 15, no. 3, pp. 363–376, Jun. 2016, doi: 10.5604/16652681.1198808.

[24] S. D. Serai and M. Yin, “MR elastography of the abdomen: experimental protocols,” *Methods Mol Biol*, vol. 2216, pp. 519–546, 2021, doi: 10.1007/978-1-0716-0978-1_32.

[25] A. C. Winters, R. Mittal, and T. D. Schiano, “A review of the use of transient elastography in the assessment of fibrosis and steatosis in the post-liver transplant patient,” *Clin Transplant*, vol. 33, no. 10, p. e13700, Oct. 2019, doi: 10.1111/ctr.13700.

[26] C. Sommerer, M. Scharf, C. Seitz, G. Millonig, H. K. Seitz, M. Zeier, and S. Mueller, “Assessment of renal allograft fibrosis by transient elastography,” *Transpl Int*, vol. 26, no. 5, pp. 545–551, May 2013, doi: 10.1111/tri.12073.

[27] K. F. Stock, B. S. Klein, M. T. Vo Cong, O. Sarkar, M. Römisch, C. Regenbogen, M. Büttner, T. Schuster, E. Matevossian, K. Amann, D. A. Clevert, U. Heemann, and C. Küchle, “ARFI-based tissue elasticity quantification in comparison to histology for the diagnosis of renal transplant fibrosis,” *Clin. Hemorheol. Microcirc.*, vol. 46, no. 2–3, pp. 139–148, 2010, doi: 10.3233/CH-2010-1340.

[28] W.-Y. He, Y.-J. Jin, W.-P. Wang, C.-L. Li, Z.-B. Ji, and C. Yang, “Tissue elasticity quantification by acoustic radiation force impulse for the assessment of renal allograft function,” *Ultrasound Med Biol*, vol. 40, no. 2, pp. 322–329, Feb. 2014, doi: 10.1016/j.ultrasmedbio.2013.10.003.

[29] S. Nasseri, L. E. Bilston, and N. Phan-Thien, “Viscoelastic properties of pig kidney in shear, experimental results and modelling,” *Rheol. Acta*, vol. 41, no. 1, pp. 180–192, Jan. 2002, doi: 10.1007/s003970200017.

[30] C. Amador, M. W. Urban, S. Chen, and J. F. Greenleaf, “Shearwave Dispersion Ultrasound Vibrometry (SDUV) on swine kidney,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 58, no. 12, pp. 2608–2619, Dec. 2011, doi: 10.1109/TUFFC.2011.2124.

[31] S. Chen, M. Fatemi, and J. F. Greenleaf, “Quantifying elasticity and viscosity from measurement of shear wave speed dispersion,” *J. Acoust. Soc. Am.*, vol. 115, no. 6, pp. 2781–2785, Jun. 2004, doi: 10.1121/1.1739480.

[32] K. R. Nightingale, N. C. Rouze, S. J. Rosenzweig, M. H. Wang, M. F. Abdelmalek, C. D. Guy, and M. L. Palmeri, “Derivation and analysis of viscoelastic properties in human liver: impact of frequency on fibrosis and steatosis staging,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 62, no. 1, pp. 165–175, Jan. 2015, doi: 10.1109/TUFFC.2014.006653.

[33] I. Z. Nenadic, M. W. Urban, J. F. Greenleaf, J.-L. Gennisson, M. Bernal, and M. Tanter, *Ultrasound elastography for biomedical applications and medicine*. John Wiley & Sons, 2019.

[34] P. Kijanka and M. W. Urban, “Dispersion curve calculation in viscoelastic tissue-mimicking materials using non-parametric, parametric, and high-resolution methods,” *Ultrasonics*, vol. 109, p. 106257, Jan. 2021, doi: 10.1016/j.ultras.2020.106257.

[35] L. J. Brattain, B. A. Telfer, M. Dhyani, J. R. Grajo, and A. E. Samir, “Machine learning for medical ultrasound: status, methods, and future opportunities,” *Abdom Radiol (NY)*, vol. 43, no. 4, pp. 786–799, 2018, doi: 10.1007/s00261-018-1517-0.

[36] R. Pagariya and M. Bartere, “Review paper on artificial neural networks,” *International Journal of Advanced Research in Computer Science*, vol. 4, no. 6, 2013, doi: 10.26483/ijarcs.v4i6.1699.

[37] W. Rawat and Z. Wang, “Deep convolutional neural networks for image classification: a comprehensive review,” *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, Jun. 2017, doi: 10.1162/neco_a_00990.

[38] K. Wang, X. Lu, H. Zhou, Y. Gao, J. Zheng, M. Tong, C. Wu, C. Liu, L. Huang, T. Jiang, F. Meng, Y. Lu, H. Ai, X.-Y. Xie, L.-P. Yin, P. Liang, J. Tian, and R. Zheng, “Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study,” *Gut*, vol. 68, no. 4, pp. 729–741, 2019, doi: 10.1136/gutjnl-2018-316204.

[39] S. Ahmed, U. Kamal, and Md. K. Hasan, “DSWE-Net: A deep learning approach for shear wave elastography and lesion segmentation using single push acoustic radiation force,” *Ultrasonics*, vol. 110, p. 106283, Feb. 2021, doi: 10.1016/j.ultras.2020.106283.

[40] S. A. Lodhi and H.-U. Meier-Kriesche, “Kidney allograft survival: the long and short of it,” *Nephrology Dialysis Transplantation*, vol. 26, no. 1, pp. 15–17, Jan. 2011, doi: 10.1093/ndt/gfq730.

[41] S. M. Bagshaw and R. T. N. Gibney, “Conventional markers of kidney function,” *Crit Care Med*, vol. 36, no. 4 Suppl, pp. S152-158, Apr. 2008, doi: 10.1097/CCM.0b013e318168c613.

[42] C. A. Jones, G. M. McQuillan, J. W. Kusek, M. S. Eberhardt, W. H. Herman, J. Coresh, M. Salive, C. P. Jones, and L. Y. Agodoa, “Serum creatinine levels in the

US population: third National Health and Nutrition Examination Survey," *Am J Kidney Dis*, vol. 32, no. 6, pp. 992–999, Dec. 1998, doi: 10.1016/s0272-6386(98)70074-5.

[43] P. J. Swedko, H. D. Clark, K. Paramsothy, and A. Akbari, "Serum creatinine is an inadequate screening test for renal failure in elderly patients," *Arch Intern Med*, vol. 163, no. 3, pp. 356–360, Feb. 2003, doi: 10.1001/archinte.163.3.356.

[44] A. S. Levey, J. P. Bosch, J. B. Lewis, T. Greene, N. Rogers, and D. Roth, "A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group," *Ann Intern Med*, vol. 130, no. 6, pp. 461–470, Mar. 1999, doi: 10.7326/0003-4819-130-6-199903160-00002.

[45] L. A. Stevens, J. Coresh, T. Greene, and A. S. Levey, "Assessing kidney function — measured and estimated glomerular filtration rate," *New England Journal of Medicine*, vol. 354, no. 23, pp. 2473–2483, Jun. 2006, doi: 10.1056/NEJMra054415.

[46] H. J. Jeong, "Diagnosis of renal transplant rejection: Banff classification and beyond," *Kidney Res Clin Pract*, vol. 39, no. 1, pp. 17–31, Mar. 2020, doi: 10.23876/j.krcp.20.003.

[47] S. Sund, A. V. Reisaeter, P. Fauchald, O. Bentdal, K. S. Hall, and T. Hovig, "Living donor kidney transplants: a biopsy study 1 year after transplantation, compared with baseline changes and correlation to kidney function at 1 and 3 years," *Nephrol Dial Transplant*, vol. 14, no. 10, pp. 2445–2454, Oct. 1999, doi: 10.1093/ndt/14.10.2445.

[48] F. G. Cosio, J. P. Grande, H. Wadei, T. S. Larson, M. D. Griffin, and M. D. Stegall, "Predicting subsequent decline in kidney allograft function from early surveillance biopsies," *Am J Transplant*, vol. 5, no. 10, Art. no. 10, Oct. 2005, doi: 10.1111/j.1600-6143.2005.01050.x.

[49] D. Rush, "Protocol transplant biopsies: an underutilized tool in kidney transplantation," *Clin J Am Soc Nephrol*, vol. 1, no. 1, Art. no. 1, Jan. 2006, doi: 10.2215/CJN.00390705.

[50] M. Mengel, J. R. Chapman, F. G. Cosio, M. W. Cavaillé-Coll, H. Haller, P. F. Halloran, A. D. Kirk, M. J. Mihatsch, B. J. Nankivell, L. C. Racusen, I. S. Roberts, D. N. Rush, A. Schwarz, D. Serón, M. D. Stegall, and R. B. Colvin, "Protocol biopsies in renal transplantation: insights into patient management and

pathogenesis," *Am J Transplant*, vol. 7, no. 3, Art. no. 3, Mar. 2007, doi: 10.1111/j.1600-6143.2006.01677.x.

[51] O. Thaunat, C. Legendre, E. Morelon, H. Kreis, and M.-F. Mamzer-Bruneel, "To biopsy or not to biopsy? Should we screen the histology of stable renal grafts?," *Transplantation*, vol. 84, no. 6, pp. 671–676, Sep. 2007, doi: 10.1097/01.tp.0000282870.71282.ed.

[52] W. D. Park, M. D. Griffin, L. D. Cornell, F. G. Cosio, and M. D. Stegall, "Fibrosis with inflammation at one year predicts transplant functional decline," *J Am Soc Nephrol*, vol. 21, no. 11, Art. no. 11, Nov. 2010, doi: 10.1681/ASN.2010010049.

[53] F. G. Cosio, M. El Ters, L. D. Cornell, C. A. Schinstock, and M. D. Stegall, "Changing Kidney Allograft Histology Early Posttransplant: Prognostic Implications of 1-Year Protocol Biopsies," *Am J Transplant*, vol. 16, no. 1, Art. no. 1, Jan. 2016, doi: 10.1111/ajt.13423.

[54] M. Gago, L. D. Cornell, W. K. Kremers, M. D. Stegall, and F. G. Cosio, "Kidney allograft inflammation and fibrosis, causes and consequences," *Am J Transplant*, vol. 12, no. 5, pp. 1199–1207, May 2012, doi: 10.1111/j.1600-6143.2011.03911.x.

[55] P. N. Furness, C. M. Philpott, M. T. Chorbadjian, M. L. Nicholson, J.-L. Bosmans, B. L. Corthouts, J. J. P. M. Bogers, A. Schwarz, W. Gwinner, H. Haller, M. Mengel, D. Seron, F. Moreso, and C. Cañas, "Protocol biopsy of the stable renal transplant: a multicenter study of methods and complication rates," *Transplantation*, vol. 76, no. 6, Art. no. 6, Sep. 2003, doi: 10.1097/01.tp.0000082542.99416.11.

[56] C. Manno, G. F. M. Strippoli, L. Arnesano, C. Bonifati, N. Campobasso, L. Gesualdo, and F. P. Schena, "Predictors of bleeding complications in percutaneous ultrasound-guided renal biopsy," *Kidney International*, vol. 66, no. 4, Art. no. 4, Oct. 2004, doi: 10.1111/j.1523-1755.2004.00922.x.

[57] S. M. Soares, F. C. Fervenza, D. J. Lager, M. A. Gertz, F. G. Cosio, and N. Leung, "Bleeding complications after transcutaneous kidney biopsy in patients with systemic amyloidosis: single-center experience in 101 patients," *Am J Kidney Dis*, vol. 52, no. 6, Art. no. 6, Dec. 2008, doi: 10.1053/j.ajkd.2008.05.022.

[58] I. D. Maya, P. Maddela, J. Barker, and M. Allon, "Percutaneous renal biopsy: comparison of blind and real-time ultrasound-guided technique," *Semin Dial*, vol. 20, no. 4, Art. no. 4, Aug. 2007, doi: 10.1111/j.1525-139X.2007.00295.x.

[59] I. D. Maya and M. Allon, “Percutaneous renal biopsy: outpatient observation without hospitalization is safe,” *Semin Dial*, vol. 22, no. 4, Art. no. 4, Aug. 2009, doi: 10.1111/j.1525-139X.2009.00609.x.

[60] T. D. Atwell, J. C. Spanbauer, B. P. McMenomy, A. H. Stockland, G. K. Hesley, C. D. Schleck, W. S. Harmsen, and T. J. Welch, “The Timing and Presentation of Major Hemorrhage After 18,947 Image-Guided Percutaneous Biopsies,” *AJR Am J Roentgenol*, vol. 205, no. 1, Art. no. 1, Jul. 2015, doi: 10.2214/AJR.14.13002.

[61] S. W. Kwan, M. Bhargavan, R. K. Kerlan, and J. H. Sunshine, “Effect of advanced imaging technology on how biopsies are done and who does them,” *Radiology*, vol. 256, no. 3, Art. no. 3, Sep. 2010, doi: 10.1148/radiol.10092130.

[62] B. J. Kim, C. K. Kim, and J. J. Park, “Non-invasive evaluation of stable renal allograft function using point shear-wave elastography,” *Br J Radiol*, vol. 91, no. 1081, p. 20170372, Jan. 2018, doi: 10.1259/bjr.20170372.

[63] S. D. Bolboacă, F. I. Elec, A. D. Elec, A. M. Muntean, M. A. Socaciu, G. Iacob, R. Zaro, A.-I. Andrieș, R. M. Bădulescu, R. M. Ignat, M. Iancu, and R. I. Badea, “Shear-wave elastography variability analysis and relation with kidney allograft dysfunction: a single-center study,” *Diagnostics (Basel)*, vol. 10, no. 1, p. E41, Jan. 2020, doi: 10.3390/diagnostics10010041.

[64] J. Hwang, H. W. Kim, P. H. Kim, C. H. Suh, and H. M. Yoon, “Technical performance of acoustic radiation force impulse imaging for measuring renal parenchymal stiffness: a systematic review and meta-analysis,” *J Ultrasound Med*, Feb. 2021, doi: 10.1002/jum.15654.

[65] M. Farshad, M. Barbezat, P. Flüeler, F. Schmidlin, P. Graber, and P. Niederer, “Material characterization of the pig kidney in relation with the biomechanical analysis of renal trauma,” *J Biomech*, vol. 32, no. 4, Art. no. 4, Apr. 1999, doi: 10.1016/s0021-9290(98)00180-8.

[66] K. Miller, “Constitutive modelling of abdominal organs,” *Journal of Biomechanics*, vol. 33, no. 3, Art. no. 3, Mar. 2000, doi: 10.1016/S0021-9290(99)00196-7.

[67] T. Bschleipfer, D. Kallieris, E. W. Hauck, W. Weidner, and R. A. Pust, “Blunt renal trauma: biomechanics and origination of renal lesions,” *Eur Urol*, vol. 42, no. 6, Art. no. 6, Dec. 2002, doi: 10.1016/s0302-2838(02)00437-2.

[68] J. G. Snedeker, M. Barbezat, P. Niederer, F. R. Schmidlin, and M. Farshad, “Strain energy density as a rupture criterion for the kidney: impact tests on porcine organs, finite element simulation, and a baseline comparison between human and porcine

tissues," *Journal of Biomechanics*, vol. 38, no. 5, Art. no. 5, May 2005, doi: 10.1016/j.jbiomech.2004.05.030.

[69] S. Chen, M. Fatemi, and J. F. Greenleaf, "Quantifying elasticity and viscosity from measurement of shear wave speed dispersion," *J. Acoust. Soc. Am.*, vol. 115, no. 6, Art. no. 6, Jun. 2004, doi: 10.1121/1.1739480.

[70] L. Huwart, F. Peeters, R. Sinkus, L. Annet, N. Salameh, L. C. ter Beek, Y. Horsmans, and B. E. Van Beers, "Liver fibrosis: non-invasive assessment with MR elastography," *NMR Biomed*, vol. 19, no. 2, Art. no. 2, Apr. 2006, doi: 10.1002/nbm.1030.

[71] N. Salameh, F. Peeters, R. Sinkus, J. Abarca-Quinones, L. Annet, L. C. Ter Beek, I. Leclercq, and B. E. Van Beers, "Hepatic viscoelastic parameters measured with MR elastography: correlations with quantitative analysis of liver fibrosis in the rat," *J Magn Reson Imaging*, vol. 26, no. 4, Art. no. 4, Oct. 2007, doi: 10.1002/jmri.21099.

[72] L. Huwart, C. Sempoux, N. Salameh, J. Jamart, L. Annet, R. Sinkus, F. Peeters, L. C. ter Beek, Y. Horsmans, and B. E. Van Beers, "Liver fibrosis: noninvasive assessment with MR elastography versus aspartate aminotransferase-to-platelet ratio index," *Radiology*, vol. 245, no. 2, Art. no. 2, Nov. 2007, doi: 10.1148/radiol.2452061673.

[73] S. Chen, W. Sanchez, M. R. Callstrom, B. Gorman, J. T. Lewis, S. O. Sanderson, J. F. Greenleaf, H. Xie, Y. Shi, M. Pashley, V. Shamdasani, M. Lachman, and S. Metz, "Assessment of liver viscoelasticity by using shear waves induced by ultrasound radiation force," *Radiology*, vol. 266, no. 3, Art. no. 3, Mar. 2013, doi: 10.1148/radiol.12120837.

[74] T. Deffieux, J.-L. Gennisson, L. Bousquet, M. Corouge, S. Cosconea, D. Amroun, S. Tripon, B. Terris, V. Mallet, P. Sogni, M. Tanter, and S. Pol, "Investigating liver stiffness and viscosity for fibrosis, steatosis and activity staging using shear wave elastography," *J Hepatol*, vol. 62, no. 2, Art. no. 2, Feb. 2015, doi: 10.1016/j.jhep.2014.09.020.

[75] S. Chen, M. W. Urban, C. Pislaru, R. Kinnick, Y. Zheng, A. Yao, and J. F. Greenleaf, "Shearwave dispersion ultrasound vibrometry (SDUV) for measuring tissue elasticity and viscosity," *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 56, no. 1, Art. no. 1, Jan. 2009, doi: 10.1109/TUFFC.2009.1005.

[76] C. Amador, M. W. Urban, S. Chen, and J. F. Greenleaf, "Loss tangent and complex modulus estimated by acoustic radiation force creep and shear wave dispersion,"

Phys Med Biol, vol. 57, no. 5, Art. no. 5, Mar. 2012, doi: 10.1088/0031-9155/57/5/1263.

[77] C. T. Barry, Z. Hah, A. Partin, R. A. Mooney, K.-H. Chuang, A. Augustine, A. Almudevar, W. Cao, D. J. Rubens, and K. J. Parker, “Mouse liver dispersion for the diagnosis of early-stage Fatty liver disease: a 70-sample study,” *Ultrasound Med Biol*, vol. 40, no. 4, Art. no. 4, Apr. 2014, doi: 10.1016/j.ultrasmedbio.2013.10.016.

[78] I. Z. Nenadic, B. Qiang, M. W. Urban, H. Zhao, W. Sanchez, J. F. Greenleaf, and S. Chen, “Attenuation Measuring Ultrasound Shearwave Elastography and in vivo application in post-transplant liver patients,” *Phys Med Biol*, vol. 62, no. 2, Art. no. 2, Jan. 2017, doi: 10.1088/1361-6560/aa4f6f.

[79] P. Song, H. Zhao, A. Manduca, M. W. Urban, J. F. Greenleaf, and S. Chen, “Comb-push ultrasound shear elastography (CUSE): a novel method for two-dimensional shear elasticity imaging of soft tissues,” *IEEE Trans Med Imaging*, vol. 31, no. 9, Art. no. 9, Sep. 2012, doi: 10.1109/TMI.2012.2205586.

[80] P. Song, M. W. Urban, A. Manduca, H. Zhao, J. F. Greenleaf, and S. Chen, “Comb-push ultrasound shear elastography (CUSE) with various ultrasound push beams,” *IEEE Trans Med Imaging*, vol. 32, no. 8, Art. no. 8, Aug. 2013, doi: 10.1109/TMI.2013.2257831.

[81] P. Kijanka and M. W. Urban, “Phase Velocity Estimation With Expanded Bandwidth in Viscoelastic Phantoms and Tissues,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1352–1362, May 2021, doi: 10.1109/TMI.2021.3054950.

[82] M. Bernal, I. Nenadic, M. W. Urban, and J. F. Greenleaf, “Material property estimation for tubes and arteries using ultrasound radiation force and analysis of propagating modes,” *J Acoust Soc Am*, vol. 129, no. 3, pp. 1344–1354, Mar. 2011, doi: 10.1121/1.3533735.

[83] T. Deffieux, G. Montaldo, M. Tanter, and M. Fink, “Shear wave spectroscopy for in vivo quantification of human soft tissues visco-elasticity,” *IEEE Trans Med Imaging*, vol. 28, no. 3, pp. 313–322, Mar. 2009, doi: 10.1109/TMI.2008.925077.

[84] J.-L. Gennisson, T. Deffieux, E. Macé, G. Montaldo, M. Fink, and M. Tanter, “Viscoelastic and anisotropic mechanical properties of in vivo muscle tissue assessed by supersonic shear imaging,” *Ultrasound Med Biol*, vol. 36, no. 5, pp. 789–801, May 2010, doi: 10.1016/j.ultrasmedbio.2010.02.013.

[85] S. Chen, W. Sanchez, M. R. Callstrom, B. Gorman, J. T. Lewis, S. O. Sanderson, J. F. Greenleaf, H. Xie, Y. Shi, M. Pashley, V. Shamdasani, M. Lachman, and S.

Metz, “Assessment of liver viscoelasticity by using shear waves induced by ultrasound radiation force,” *Radiology*, vol. 266, no. 3, pp. 964–970, Mar. 2013, doi: 10.1148/radiol.12120837.

[86] T. Chapman, T. Dubinsky, and R. G. Barr, “Ultrasound Elastography of the Liver: What the Clinician Needs to Know,” *Journal of Ultrasound in Medicine*, vol. 36, no. 7, pp. 1293–1304, 2017, doi: 10.7863/ultra.16.08001.

[87] C. Fang, O. S. Jaffer, G. T. Yusuf, E. Konstantatou, D. J. Quinlan, K. Agarwal, A. Quaglia, and P. S. Sidhu, “Reducing the number of measurements in liver point shear-wave elastography: factors that influence the number and reliability of measurements in assessment of liver fibrosis in clinical practice,” *Radiology*, vol. 287, no. 3, pp. 844–852, Jun. 2018, doi: 10.1148/radiol.2018172104.

[88] K. Solez, R. B. Colvin, L. C. Racusen, M. Haas, B. Sis, M. Mengel, P. F. Halloran, W. Baldwin, G. Banfi, A. B. Collins, F. Cosio, D. S. R. David, C. Drachenberg, G. Einecke, A. B. Fogo, I. W. Gibson, D. Glotz, S. S. Iskandar, E. Kraus, E. Lerut, R. B. Mannon, M. Mihatsch, B. J. Nankivell, V. Nickeleit, J. C. Papadimitriou, P. Randhawa, H. Regele, K. Renaudin, I. Roberts, D. Seron, R. N. Smith, and M. Valente, “Banff 07 classification of renal allograft pathology: updates and future directions,” *Am J Transplant*, vol. 8, no. 4, pp. 753–760, Apr. 2008, doi: 10.1111/j.1600-6143.2008.02159.x.

[89] L. C. Racusen, K. Solez, R. B. Colvin, S. M. Bonsib, M. C. Castro, T. Cavallo, B. P. Croker, A. J. Demetris, C. B. Drachenberg, A. B. Fogo, P. Furness, L. W. Gaber, I. W. Gibson, D. Glotz, J. C. Goldberg, J. Grande, P. F. Halloran, H. E. Hansen, B. Hartley, P. J. Hayry, C. M. Hill, E. O. Hoffman, L. G. Hunsicker, A. S. Lindblad, and Y. Yamaguchi, “The Banff 97 working classification of renal allograft pathology,” *Kidney Int*, vol. 55, no. 2, Art. no. 2, Feb. 1999, doi: 10.1046/j.1523-1755.1999.00299.x.

[90] K. Solez, R. B. Colvin, L. C. Racusen, B. Sis, P. F. Halloran, P. E. Birk, P. M. Campbell, M. Cascalho, A. B. Collins, A. J. Demetris, C. B. Drachenberg, I. W. Gibson, P. C. Grimm, M. Haas, E. Lerut, H. Liapis, R. B. Mannon, P. B. Marcus, M. Mengel, M. J. Mihatsch, B. J. Nankivell, V. Nickeleit, J. C. Papadimitriou, J. L. Platt, P. Randhawa, I. Roberts, L. Salinas-Madriga, D. R. Salomon, D. Seron, M. Sheaff, and J. J. Weening, “Banff ’05 meeting report: differential diagnosis of chronic allograft injury and elimination of chronic allograft nephropathy ('can’),” *Am J*

Transplant, vol. 7, no. 3, pp. 518–526, Mar. 2007, doi: 10.1111/j.1600-6143.2006.01688.x.

[91] K.-Y. LIANG and S. L. ZEGER, “Longitudinal data analysis using generalized linear models,” *Biometrika*, vol. 73, no. 1, Art. no. 1, Apr. 1986, doi: 10.1093/biomet/73.1.13.

[92] J. W. Hardin, “Generalized estimating equations (GEE),” in *Encyclopedia of Statistics in Behavioral Science*, John Wiley & Sons, Ltd, 2005. doi: 10.1002/0470013192.bsa250.

[93] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: machine learning in python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[94] P. A. James, S. Oparil, B. L. Carter, W. C. Cushman, C. Dennison-Himmelfarb, J. Handler, D. T. Lackland, M. L. LeFevre, T. D. MacKenzie, O. Ogedegbe, S. C. Smith, L. P. Svetkey, S. J. Taler, R. R. Townsend, J. T. Wright, A. S. Narva, and E. Ortiz, “2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8),” *JAMA*, vol. 311, no. 5, pp. 507–520, Feb. 2014, doi: 10.1001/jama.2013.284427.

[95] M. Urban, C. Amador Carrascal, and S. Aristizabal, “Compressing and shearing the transplanted kidney,” *Journal of the Acoustical Society of America*, vol. 140, pp. 3137–3137, Oct. 2016, doi: 10.1121/1.4969830.

[96] W. R. Zhang and C. R. Parikh, “Biomarkers of acute and chronic kidney disease,” *Annu Rev Physiol*, vol. 81, pp. 309–333, Feb. 2019, doi: 10.1146/annurev-physiol-020518-114605.

[97] A. S. Levey, L. A. Inker, and J. Coresh, “GFR estimation: from physiology to public health,” *Am J Kidney Dis*, vol. 63, no. 5, pp. 820–834, May 2014, doi: 10.1053/j.ajkd.2013.12.006.

[98] P. C. Wilson, M. Kashgarian, and G. Moeckel, “Interstitial inflammation and interstitial fibrosis and tubular atrophy predict renal survival in lupus nephritis,” *Clinical Kidney Journal*, vol. 11, no. 2, pp. 207–218, Apr. 2018, doi: 10.1093/ckj/sfx093.

[99] M. D. Patel, S. W. Young, J. Scott Kriegshauser, and N. Dahiya, “Ultrasound-guided renal transplant biopsy: practical and pragmatic considerations,” *Abdom Radiol (NY)*, vol. 43, no. 10, pp. 2597–2603, Oct. 2018, doi: 10.1007/s00261-018-1484-5.

[100] M. Haas, A. Loupy, C. Lefaucheur, C. Roufosse, D. Glotz, D. Seron, B. J. Nankivell, P. F. Halloran, R. B. Colvin, E. Akalin, N. Alachkar, S. Bagnasco, Y. Bouatou, J. U. Becker, L. D. Cornell, J. P. Duong van Huyen, I. W. Gibson, E. S. Kraus, R. B. Mannon, M. Naesens, V. Nickeleit, P. Nickerson, D. L. Segev, H. K. Singh, M. Stegall, P. Randhawa, L. Racusen, K. Solez, and M. Mengel, “The Banff 2017 Kidney Meeting Report: Revised diagnostic criteria for chronic active T cell-mediated rejection, antibody-mediated rejection, and prospects for integrative endpoints for next-generation clinical trials,” *Am J Transplant*, vol. 18, no. 2, pp. 293–307, Feb. 2018, doi: 10.1111/ajt.14625.

[101] M. Bojić, H. Regele, H. Herkner, G. Berlakovich, J. Kläger, C. Bauer, C. Seitz, and Ž. Kikić, “Tubular ectasia in renal allograft biopsy: associations with occult obstructive urological complications,” *Transplantation*, vol. 104, no. 1, pp. 145–153, Jan. 2020, doi: 10.1097/TP.0000000000002699.

[102] J. Naranjo, F. Villanego, J. M. Cazorla, C. León, M. J. Ledo, A. García, T. García, M. Ceballos, and A. Mazuecos, “Acute renal failure in kidney transplantation due to subcapsular hematoma after a renal allograft biopsy: report of two cases and literature review,” *Transplant Proc*, vol. 52, no. 2, pp. 530–533, Mar. 2020, doi: 10.1016/j.transproceed.2019.09.018.

[103] S.-F. Tsai, C.-H. Chen, K.-H. Shu, C.-H. Cheng, T.-M. Yu, Y.-W. Chuang, S.-T. Huang, J.-L. Tsai, and M.-J. Wu, “Current safety of renal allograft biopsy with indication in adult recipients: an observational study,” *Medicine (Baltimore)*, vol. 95, no. 6, p. e2816, Feb. 2016, doi: 10.1097/MD.0000000000002816.

[104] T. D. Atwell, J. C. Spanbauer, B. P. McMenomy, A. H. Stockland, G. K. Hesley, C. D. Schleck, W. S. Harmsen, and T. J. Welch, “The Timing and Presentation of Major Hemorrhage After 18,947 Image-Guided Percutaneous Biopsies,” *AJR Am J Roentgenol*, vol. 205, no. 1, pp. 190–195, Jul. 2015, doi: 10.2214/AJR.14.13002.

[105] L. Trajceska, G. Severova-Andreevska, P. Dzekova-Vidimliski, I. Nikolov, G. Selim, G. Spasovski, I. Rambabova-Busletik, V. Ristovska, L. Grcevska, and A. Sikole, “Complications and risks of percutaneous renal biopsy,” *Open Access Maced J Med Sci*, vol. 7, no. 6, pp. 992–995, Mar. 2019, doi: 10.3889/oamjms.2019.226.

[106] C. Leslie, E. Eskin, and W. S. Noble, “The spectrum kernel: a string kernel for svm protein classification,” in *Biocomputing 2002*, Kauai, Hawaii, USA, Dec. 2001, pp. 564–575. doi: 10.1142/9789812799623_0053.

[107] M. A. Chandra and S. S. Bedi, “Survey on SVM and their application in imageclassification,” *Int. j. inf. tecnol.*, vol. 13, no. 5, pp. 1–11, Oct. 2021, doi: 10.1007/s41870-017-0080-1.

[108] H. P. Bhavsar and M. H. Panchal, “A review on support vector machine for data classification,” 2012.

[109] Y. Shao and R. S. Lunetta, “Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 70, pp. 78–87, Jun. 2012, doi: 10.1016/j.isprsjprs.2012.04.001.

[110] A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Schölkopf, and G. Rätsch, “Support Vector Machines and Kernels for Computational Biology,” *PLOS Computational Biology*, vol. 4, no. 10, p. e1000173, Oct. 2008, doi: 10.1371/journal.pcbi.1000173.

[111] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *Advances in Neural Information Processing Systems*, 2017, vol. 30. Accessed: Feb. 07, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[112] L. S. Shapley, “Notes on the n-person game — ii: the value of an n-person game,” RAND Corporation, Aug. 1951. Accessed: Feb. 07, 2022. [Online]. Available: https://www.rand.org/pubs/research_memoranda/RM0670.html

[113] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, “Scikit-learn: machine learning in python,” *MACHINE LEARNING IN PYTHON*, p. 6.

[114] G. Ferraioli, C. Filice, L. Castera, B. I. Choi, I. Sporea, S. R. Wilson, D. Cosgrove, C. F. Dietrich, D. Amy, J. C. Bamber, R. Barr, Y.-H. Chou, H. Ding, A. Farrokh, M. Friedrich-Rust, T. J. Hall, K. Nakashima, K. R. Nightingale, M. L. Palmeri, F. Schafer, T. Shiina, S. Suzuki, and M. Kudo, “WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 3: liver,” *Ultrasound Med Biol*, vol. 41, no. 5, pp. 1161–1179, May 2015, doi: 10.1016/j.ultrasmedbio.2015.03.007.

[116] S. Chen, M. W. Urban, C. Pislaru, R. Kinnick, Y. Zheng, A. Yao, and J. F. Greenleaf, “Shearwave dispersion ultrasound vibrometry (SDUV) for measuring tissue elasticity and viscosity,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 56, no. 1, pp. 55–62, Jan. 2009, doi: 10.1109/TUFFC.2009.1005.

[117] S. S. Leong, J. H. D. Wong, M. N. Md Shah, A. Vijayananthan, M. Jalalonmuhali, and K. H. Ng, “Shear wave elastography in the evaluation of renal parenchymal stiffness in patients with chronic kidney disease,” *Br J Radiol*, vol. 91, no. 1089, p. 20180235, Sep. 2018, doi: 10.1259/bjr.20180235.

[118] S. S. Leong, J. H. D. Wong, M. N. Md Shah, A. Vijayananthan, M. Jalalonmuhali, T. K. Chow, N. H. M. Sharif, and K. H. Ng, “Shear wave elastography accurately detects chronic changes in renal histopathology,” *Nephrology (Carlton)*, vol. 26, no. 1, pp. 38–45, Jan. 2021, doi: 10.1111/nep.13805.

[119] F. Bob, I. Grosu, I. Sporea, S. Bota, A. Popescu, A. Sima, R. Şirli, L. Petrica, R. Timar, and A. Schiller, “Ultrasound-based shear wave elastography in the assessment of patients with diabetic kidney disease,” *Ultrasound Med Biol*, vol. 43, no. 10, pp. 2159–2166, Oct. 2017, doi: 10.1016/j.ultrasmedbio.2017.04.019.

[120] T. D. Hewitson, “Renal tubulointerstitial fibrosis: common but never simple,” *Am J Physiol Renal Physiol*, vol. 296, no. 6, pp. F1239-1244, Jun. 2009, doi: 10.1152/ajprenal.90521.2008.

[121] P. Kijanka and M. W. Urban, “Two-point frequency shift method for shear wave attenuation measurement,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 67, no. 3, pp. 483–496, 2020, doi: 10.1109/TUFFC.2019.2945620.

[122] S. Bernard, S. Kazemirad, and G. Cloutier, “A frequency-shift method to measure shear-wave attenuation in soft tissues,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 3, pp. 514–524, Mar. 2017, doi: 10.1109/TUFFC.2016.2634329.

[123] J.-L. Gennisson, T. Deffieux, M. Fink, and M. Tanter, “Ultrasound elastography: Principles and techniques,” *Diagnostic and Interventional Imaging*, vol. 94, no. 5, pp. 487–495, May 2013, doi: 10.1016/j.diii.2013.01.022.

[124] M. W. Urban, “Production of acoustic radiation force using ultrasound: methods and applications,” *Expert Rev Med Devices*, vol. 15, no. 11, pp. 819–834, 2018, doi: 10.1080/17434440.2018.1538782.

[125] T. Shiina, K. R. Nightingale, M. L. Palmeri, T. J. Hall, J. C. Bamber, R. G. Barr, L. Castera, B. I. Choi, Y.-H. Chou, D. Cosgrove, C. F. Dietrich, H. Ding, D. Amy, A.

Farrokh, G. Ferraioli, C. Filice, M. Friedrich-Rust, K. Nakashima, F. Schafer, I. Sporea, S. Suzuki, S. Wilson, and M. Kudo, “WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 1: basic principles and terminology,” *Ultrasound Med Biol*, vol. 41, no. 5, pp. 1126–1147, May 2015, doi: 10.1016/j.ultrasmedbio.2015.03.009.

[126] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, “Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 56, no. 3, pp. 489–506, Mar. 2009, doi: 10.1109/TUFFC.2009.1067.

[127] P. Song, M. Macdonald, R. Behler, J. Lanning, M. Wang, M. Urban, A. Manduca, H. Zhao, M. Callstrom, A. Alizad, J. Greenleaf, and S. Chen, “Two-dimensional shear-wave elastography on conventional ultrasound scanners with time-aligned sequential tracking (TAST) and comb-push ultrasound shear elastography (CUSE),” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 62, no. 2, pp. 290–302, Feb. 2015, doi: 10.1109/TUFFC.2014.006628.

[128] S. Catheline, J.-L. Gennisson, G. Delon, M. Fink, R. Sinkus, S. Abouelkaram, and J. Culioli, “Measurement of viscoelastic properties of homogeneous soft solid using transient elastography: An inverse problem approach,” *The Journal of the Acoustical Society of America*, vol. 116, no. 6, pp. 3734–3741, Dec. 2004, doi: 10.1121/1.1815075.

[130] E. Budelli, J. Brum, M. Bernal, T. Deffieux, M. Tanter, P. Lema, C. Negreira, and J.-L. Gennisson, “A diffraction correction for storage and loss moduli imaging using radiation force based elastography,” *Phys Med Biol*, vol. 62, no. 1, pp. 91–106, 07 2017, doi: 10.1088/1361-6560/62/1/91.

[131] P. Kijanka and M. W. Urban, “Local phase velocity based imaging: a new technique used for ultrasound shear wave elastography,” *IEEE Trans Med Imaging*, vol. 38, no. 4, pp. 894–908, 2019, doi: 10.1109/TMI.2018.2874545.

[132] X. Chen, Y. Shen, Y. Zheng, H. Lin, Y. Guo, Y. Zhu, X. Zhang, T. Wang, and S. Chen, “Quantification of liver viscoelasticity with acoustic radiation force: a study of hepatic fibrosis in a rat model,” *Ultrasound Med Biol*, vol. 39, no. 11, pp. 2091–2102, Nov. 2013, doi: 10.1016/j.ultrasmedbio.2013.05.020.

[133] M. L. Palmeri, M. H. Wang, J. J. Dahl, K. D. Frinkley, and K. R. Nightingale, “Quantifying hepatic shear modulus in vivo using acoustic radiation force,”

Ultrasound Med Biol, vol. 34, no. 4, pp. 546–558, Apr. 2008, doi: 10.1016/j.ultrasmedbio.2007.10.009.

[134] M. W. Urban, S. Chen, and J. F. Greenleaf, “Error in estimates of tissue material properties from shear wave dispersion ultrasound vibrometry,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 56, no. 4, pp. 748–758, Apr. 2009, doi: 10.1109/TUFFC.2009.1097.

[135] T. Deffieux, J. Gennisson, B. Larrat, M. Fink, and M. Tanter, “The variance of quantitative estimates in shear wave imaging: Theory and experiments,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 11, pp. 2390–2410, Nov. 2012, doi: 10.1109/TUFFC.2012.2472.

[136] Y. Deng, N. C. Rouze, M. L. Palmeri, and K. R. Nightingale, “On system-dependent sources of uncertainty and bias in ultrasonic quantitative shear-wave imaging,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 63, no. 3, pp. 381–393, Mar. 2016, doi: 10.1109/TUFFC.2016.2524260.

[137] A. Treacher, D. Beauchamp, B. Quadri, D. Fetzer, A. Vij, T. Yokoo, and A. Montillo, “Deep learning convolutional neural networks for the estimation of liver fibrosis severity from ultrasound texture,” *Proc SPIE Int Soc Opt Eng*, vol. 10950, Feb. 2019, doi: 10.1117/12.2512592.

[138] Q. Zhang, Y. Xiao, W. Dai, J. Suo, C. Wang, J. Shi, and H. Zheng, “Deep learning based classification of breast tumors with shear-wave elastography,” *Ultrasonics*, vol. 72, pp. 150–157, 2016, doi: 10.1016/j.ultras.2016.08.004.

[139] H. E. Robbins, “A stochastic approximation method,” 2007. doi: 10.1214/aoms/1177729586.

[140] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” *arXiv:1412.6980 [cs]*, Jan. 2017, Accessed: Jan. 28, 2020. [Online]. Available: http://arxiv.org/abs/1412.6980

[141] B. Hustedt, S. Operto, and J. Virieux, “Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling,” *Geophys J Int*, vol. 157, no. 3, pp. 1269–1296, Jun. 2004, doi: 10.1111/j.1365-246X.2004.02289.x.

[142] Y. Q. Zeng and Q. H. Liu, “A staggered-grid finite-difference method with perfectly matched layers for poroelastic wave equations,” *The Journal of the Acoustical Society of America*, vol. 109, no. 6, pp. 2571–2580, Jun. 2001, doi: 10.1121/1.1369783.

[143] J. Virieux, “P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method,” *GEOPHYSICS*, vol. 51, no. 4, pp. 889–901, Apr. 1986, doi: 10.1190/1.1442147.

[144] D. Chen and R. J. McGough, “A 2D fast near-field method for calculating near-field pressures generated by apodized rectangular pistons,” *J Acoust Soc Am*, vol. 124, no. 3, pp. 1526–1537, Sep. 2008, doi: 10.1121/1.2950081.

[145] X. Zeng and R. J. McGough, “Evaluation of the angular spectrum approach for simulations of near-field pressures,” *J Acoust Soc Am*, vol. 123, no. 1, pp. 68–76, Jan. 2008, doi: 10.1121/1.2812579.

[146] F. H. Drossaert and A. Giannopoulos, “A nonsplit complex frequency-shifted PML based on recursive integration for FDTD modeling of elastic waves,” *GEOPHYSICS*, vol. 72, no. 2, pp. T9–T17, Jan. 2007, doi: 10.1190/1.2424888.

[147] P. Kijanka, B. Qiang, P. Song, C. Amador Carrascal, S. Chen, and M. W. Urban, “Robust phase velocity dispersion estimation of viscoelastic materials used for medical applications based on the multiple signal classification method,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 65, no. 3, pp. 423–439, 2018, doi: 10.1109/TUFFC.2018.2792324.

[148] C. Amador Carrascal, S. Chen, A. Manduca, J. F. Greenleaf, and M. W. Urban, “Improved shear wave group velocity estimation method based on spatiotemporal peak and thresholding motion search,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 64, no. 4, pp. 660–668, Apr. 2017, doi: 10.1109/TUFFC.2017.2652143.

[149] F. Chollet, *Keras*. GitHub, 2015. [Online]. Available: https://github.com/fchollet/keras

[150] C. Amador, M. Urban, R. Kinnick, S. Chen, and J. F. Greenleaf, “In vivo swine kidney viscoelasticity during acute gradual decrease in renal blood flow: pilot study,” *Rev Ing Biomed*, vol. 7, no. 13, pp. 68–78, Jan. 2013.

[151] C. Amador, R. R. Kinnick, M. W. Urban, M. Fatemi, and J. F. Greenleaf, “Viscoelastic tissue mimicking phantom validation study with shear wave elasticity imaging and viscoelastic spectroscopy,” in *2015 IEEE International Ultrasonics Symposium (IUS)*, Oct. 2015, pp. 1–4. doi: 10.1109/ULTSYM.2015.0287.

[152] S. Aristizabal, C. Amador, Bo Qiang, I. Z. Nenadic, J. F. Greenleaf, and M. W. Urban, “Viscoelastic characterization of transverse isotropic tissue mimicking phantoms and muscle,” in *2014 IEEE International Ultrasonics Symposium*, Sep. 2014, pp. 228–231. doi: 10.1109/ULTSYM.2014.0058.

[153] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *J Big Data*, vol. 3, no. 1, p. 9, May 2016, doi: 10.1186/s40537-016-0043-6.

[154] L. Vasconcelos, P. Kijanka, and M. W. Urban, “Viscoelastic parameter estimation using simulated shear wave motion and convolutional neural networks,” *Comput Biol Med*, vol. 133, p. 104382, Jun. 2021, doi: 10.1016/j.compbiomed.2021.104382.

[155] J. Bercoff, M. Tanter, M. Muller, and M. Fink, “The role of viscosity in the impulse diffraction field of elastic waves induced by the acoustic radiation force,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 51, no. 11, pp. 1523–1536, Nov. 2004, doi: 10.1109/TUFFC.2004.1367494.

[156] D. Bæk, J. A. Jensen, and M. Willatzen, “Modeling transducer impulse responses for predicting calibrated pressure pulses with the ultrasound simulation program Field II,” *J Acoust Soc Am*, vol. 127, no. 5, pp. 2825–2835, May 2010, doi: 10.1121/1.3365317.

[157] H. M. Wyss, J. M. Henderson, F. J. Byfield, L. A. Bruggeman, Y. Ding, C. Huang, J. H. Suh, T. Franke, E. Mele, M. R. Pollak, J. H. Miner, P. A. Janmey, D. A. Weitz, and R. T. Miller, “Biophysical properties of normal and diseased renal glomeruli,” *Am. J. Physiol., Cell Physiol.*, vol. 300, no. 3, pp. C397-405, Mar. 2011, doi: 10.1152/ajpcell.00438.2010.

[158] A. E. Embry, H. Mohammadi, X. Niu, L. Liu, B. Moe, W. A. Miller-Little, C. Y. Lu, L. A. Bruggeman, C. A. McCulloch, P. A. Janmey, and R. T. Miller, “Biochemical and cellular determinants of renal glomerular elasticity,” *PLOS ONE*, vol. 11, no. 12, p. e0167924, Dec. 2016, doi: 10.1371/journal.pone.0167924.

[159] L. W. Welling and J. J. Grantham, “Physical properties of isolated perfused renal tubules and tubular basement membranes,” *J. Clin. Invest.*, vol. 51, no. 5, pp. 1063–1075, May 1972, doi: 10.1172/JCI106898.

[160] A. L. Wentland, E. A. Sadowski, A. Djamali, T. M. Grist, B. N. Becker, and S. B. Fain, “Quantitative MR measures of intrarenal perfusion in the assessment of transplanted kidneys: initial experience,” *Acad Radiol*, vol. 16, no. 9, pp. 1077–1085, Sep. 2009, doi: 10.1016/j.acra.2009.03.020.

[161] X. Liu, N. Li, T. Xu, F. Sun, R. Li, Q. Gao, L. Chen, and C. Wen, “Effect of renal perfusion and structural heterogeneity on shear wave elastography of the kidney: an in vivo and ex vivo study,” *BMC Nephrology*, vol. 18, no. 1, p. 265, Aug. 2017, doi: 10.1186/s12882-017-0679-2.

[162] P. Song, M. W. Urban, A. Manduca, H. Zhao, J. F. Greenleaf, and S. Chen, “Comb-push ultrasound shear elastography (CUSE) with various ultrasound push beams,” *IEEE Trans Med Imaging*, vol. 32, no. 8, pp. 1435–1447, Aug. 2013, doi: 10.1109/TMI.2013.2257831.

[163] T. Samuel, W. E. Hoy, R. Douglas-Denton, M. D. Hughson, and J. F. Bertram, “Applicability of the glomerular size distribution coefficient in assessing human glomerular volume: the Weibel and Gomez method revisited,” *J Anat*, vol. 210, no. 5, pp. 578–582, May 2007, doi: 10.1111/j.1469-7580.2007.00715.x.

[164] M. A. Lusco, A. B. Fogo, B. Najafian, and C. E. Alpers, “Ajkd atlas of renal pathology: tubular atrophy,” *American Journal of Kidney Diseases*, vol. 67, no. 6, pp. e33–e34, Jun. 2016, doi: 10.1053/j.ajkd.2016.04.007.

[165] C. Helfenstein, J.-L. Gennisson, M. Tanter, and P. Beillas, “Effects of pressure on the shear modulus, mass and thickness of the perfused porcine kidney,” *J Biomech*, vol. 48, no. 1, pp. 30–37, Jan. 2015, doi: 10.1016/j.jbiomech.2014.11.011.

[166] C. Amador, M. W. Urban, R. Kinnick, S. Chen, and J. F. Greenleaf, “Viscoelastic measurements on perfused and non-perfused swine renal cortex in vivo,” in *2010 IEEE International Ultrasonics Symposium*, Oct. 2010, pp. 1530–1533. doi: 10.1109/ULTSYM.2010.5935672.

[167] J.-L. Gennisson, N. Grenier, C. Combe, and M. Tanter, “Supersonic shear wave elastography of in vivo pig kidney: influence of blood pressure, urinary pressure and tissue anisotropy,” *Ultrasound Med Biol*, vol. 38, no. 9, pp. 1559–1567, Sep. 2012, doi: 10.1016/j.ultrasmedbio.2012.04.013.

[168] M. Mengel, J. R. Chapman, F. G. Cosio, M. W. Cavaillé-Coll, H. Haller, P. F. Halloran, A. D. Kirk, M. J. Mihatsch, B. J. Nankivell, L. C. Racusen, I. S. Roberts, D. N. Rush, A. Schwarz, D. Serón, M. D. Stegall, and R. B. Colvin, “Protocol biopsies in renal transplantation: insights into patient management and pathogenesis,” *Am J Transplant*, vol. 7, no. 3, pp. 512–517, Mar. 2007, doi: 10.1111/j.1600-6143.2006.01677.x.

[169] C. Manno, G. F. M. Strippoli, L. Arnesano, C. Bonifati, N. Campobasso, L. Gesualdo, and F. P. Schena, “Predictors of bleeding complications in percutaneous ultrasound-guided renal biopsy,” *Kidney International*, vol. 66, no. 4, pp. 1570–1577, Oct. 2004, doi: 10.1111/j.1523-1755.2004.00922.x.

[170] W. D. Park, M. D. Griffin, L. D. Cornell, F. G. Cosio, and M. D. Stegall, “Fibrosis with inflammation at one year predicts transplant functional decline,” *J Am Soc Nephrol*, vol. 21, no. 11, pp. 1987–1997, Nov. 2010, doi: 10.1681/ASN.2010010049.

[171] J. R. Doherty, G. E. Trahey, K. R. Nightingale, and M. L. Palmeri, “Acoustic radiation force elasticity imaging in diagnostic ultrasound,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 60, no. 4, pp. 685–701, Apr. 2013, doi: 10.1109/TUFFC.2013.2617.

[172] M. Bruce, O. Kolokythas, G. Ferraioli, C. Filice, and M. O’Donnell, “Limitations and artifacts in shear-wave elastography of the liver,” *Biomed Eng Lett*, vol. 7, no. 2, pp. 81–89, May 2017, doi: 10.1007/s13534-017-0028-1.

[173] P. Bouchet, J.-L. Gennisson, A. Podda, M. Alilet, M. Carrié, and S. Aubry, “Artifacts and technical restrictions in 2d shear wave elastography,” *Ultraschall Med*, vol. 41, no. 3, pp. 267–277, Jun. 2020, doi: 10.1055/a-0805-1099.

[174] F. J. Harris, “On the use of windows for harmonic analysis with the discrete Fourier transform,” *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978, doi: 10.1109/PROC.1978.10837.

[175] Y. Assiri, “Stochastic Optimization of Plain Convolutional Neural Networks with Simple methods,” *arXiv:2001.08856 [cs]*, Jan. 2020, Accessed: Apr. 05, 2022. [Online]. Available: http://arxiv.org/abs/2001.08856

[176] B. J. Erickson and F. Kitamura, “Magician’s corner: 9. performance metrics for machine learning models,” *Radiol Artif Intell*, vol. 3, no. 3, p. e200126, May 2021, doi: 10.1148/ryai.2021200126.

[177] C. Kodner and A. Kudrimoti, “Diagnosis and management of acute interstitial nephritis,” *AFP*, vol. 67, no. 12, pp. 2527–2534, Jun. 2003.

[178] J. E. Hall, M. W. Brands, and J. R. Henegar, “Mechanisms of hypertension and kidney disease in obesity,” *Annals of the New York Academy of Sciences*, vol. 892, no. 1, pp. 91–107, 1999, doi: 10.1111/j.1749-6632.1999.tb07788.x.

[179] J. E. Hall, J. M. do Carmo, A. A. da Silva, Z. Wang, and M. E. Hall, “Obesity, kidney dysfunction and hypertension: mechanistic links,” *Nat Rev Nephrol*, vol. 15, no. 6, pp. 367–385, Jun. 2019, doi: 10.1038/s41581-019-0145-4.

[180] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, “A review of uncertainty quantification in deep learning: Techniques, applications

and challenges," *Information Fusion*, vol. 76, pp. 243–297, Dec. 2021, doi: 10.1016/j.inffus.2021.05.008.

[181] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, Jan. 2018, doi: 10.1093/nsr/nwx106.