STAGING LIVER FIBROSIS WITH STATISTICAL OBSERVERS

by

Jonathan Frieman Brand

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF OPTICAL SCIENCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2016

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Jonathan Frieman Brand, titled Staging Liver Fibrosis with Statistical Observers and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

____ Date: 04/19/16
Dr. Lars R. Furenlid

____ Date: 04/19/16
Dr. Maria I. Altbach

____ Date: 04/19/16
Dr. Matthew Kupinski

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 04/19/16
Dissertation Director: Dr. Lars R. Furenlid

STATEMENT BY AUTHOR

SIGNED:

Jonathan Frieman Brand

ACKNOWLEDGEMENTS

First, I must thank my advisor, Dr. Lars R. Furenlid, who guided me in my work during this project. I also want to thank my committee member Dr. Maria I. Altbach, who was instrumental in the coordination of my training in MRI data collection as well as connecting me with the team at the Department of Pathology. This project would also not be possible with the final member of my committee, Dr. Matthew Kupinski, who provided a majority of my education in mathematical observers and provided crucial feedback on my experiments. This work is not possible either without Dr. Diego Martin, who provided insight into radiologist's and patient's needs for non-invasive quantitative detection methods. Without their collective guidance and knowledge this dissertation would not have been possible.

I am in debt to the researchers supporting this project, allowing the formation of an inter-departmental network to provide training in the operation of a clinical MRI scanner. Dr. Jean-Philippe Galons and Mr. Scott Squire provided the necessary support for me to complete my experiments at the Department of Medical imaging. Also, thanks to Mr. Tulshi Bhattacharyya, who coordinated our work with the Department of Pathology

This project would not have been possible without our collaborators at the Department of Pathology. Led by Dr. Achuyt Bhattacharyya, who volunteered his time to provide pathology reads of our samples and train me in basic pathology practices. Others thanks at the Department of Pathology must go to Dr. Cornel Moga, Dr. Bruce Parks, Dr. Richard Sobonya, and Mr. Samuel Kilgorn for helping provide the phantoms that are a crucial focus of this dissertation.

A special thanks goes to Dr. Eric Clarkson for his knowledge and contributions during my study of model observers. Along with others at who taught many classes or provided clarity in various subjects, including, Dr. Art Gmitro, Dr. Theodore Trouard, Dr. Ali Bilgin, and Dr. Matthew Dubin.

A thank you to the students who I have studied with during my time at the University of Arizona, including, Tomoe Hagio, Abhishek Pandey, Zhitao Li, Dr. Ben Berman, Dr. Gregory Cohoon, and many others.

Thank you Archer and Foucault, who inspire procrastination everyday.

Finally, thank you, Alison Schultz, whose fantastic patience, love, and adventurous nature keeps me excited to see what mess we get into next.

DEDICATION

A Ph.D. candidate is a single person, I am so lucky to have the support of so many loving people in my life, who have supported me along the way.

I dedicate this dissertation to my parents who always let me find my own way, in all parts of my life, my aunt and uncle, Stephanie Brand and Mark Elder, who provided sanctuary in Sedona on many weekends.

Table of Contents

List of Figures

List of Tables

Abstract

Chronic liver disease is a worldwide health problem, and hepatic fibrosis (HF) is one of the hallmarks of the disease. Pathology diagnosis of HF is based on textural change in the liver as a lobular collagen network that develops within portal triads. The scale of collagen lobules is characteristically on order of 1mm, which close to the resolution limit of in vivo Gd-enhanced MRI.

In this work the methods to collect training and testing images for a Hotelling observer are covered. An observer based on local texture analysis is trained and tested using wet-tissue phantoms. The technique is used to optimize the MRI sequence based on task performance.

The final method developed is a two stage model observer to classify fibrotic and healthy tissue in both phantoms and *in vivo* MRI images. The first stage observer tests for the presence of local texture. Test statistics from the first observer are used to train the second stage observer to globally sample the local observer results. A decision of the disease class is made for an entire MRI image slice using test statistics collected from the second observer. The techniques are tested on wet-tissue phantoms and *in vivo* clinical patient data.

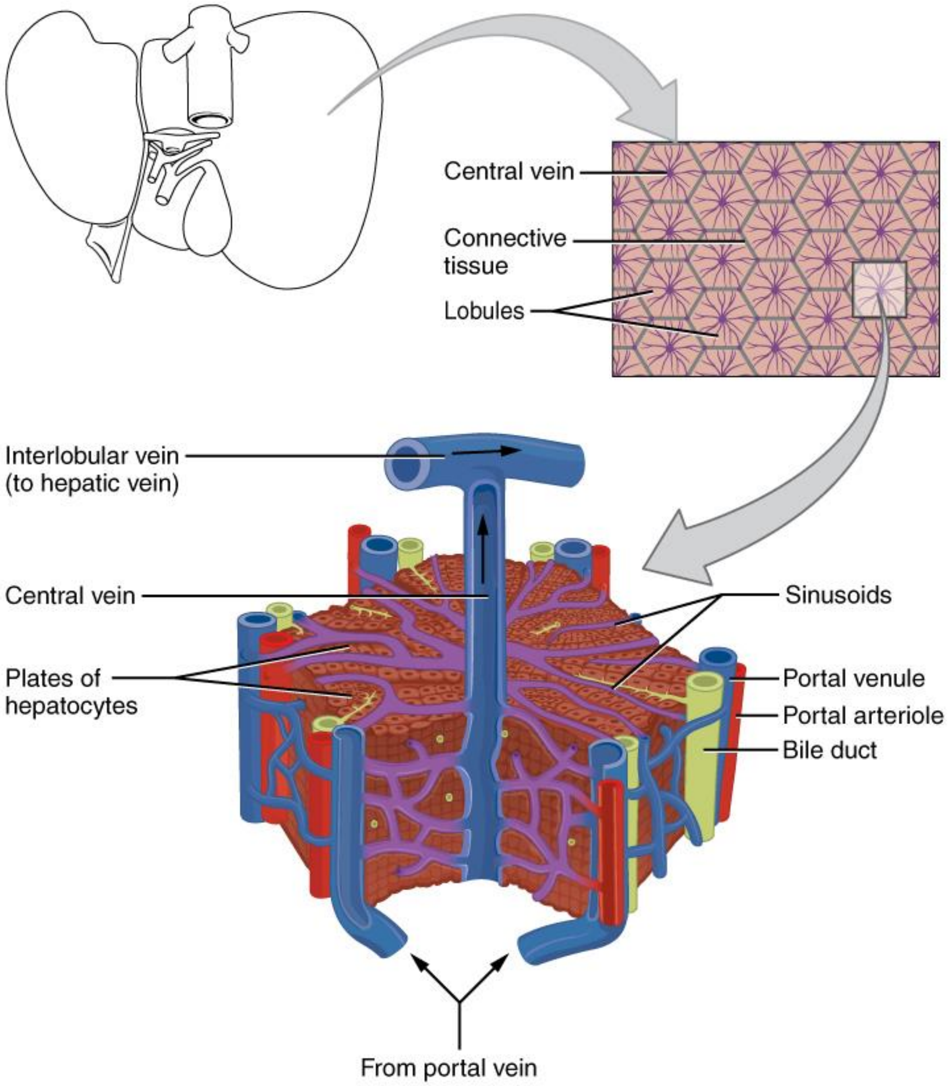
1. Introduction

The early detection of hepatic fibrosis (HF), as a result of chronic liver disease (CLD), is a significant need for patients and physicians. In this body of work, we combine magnetic resonance imaging (MRI) with statistical observers introduce a novel non-invasive method to stage liver disease. This introduction is an overview of the function of the liver, we target features that will be used for detection of HF, and current methods to stage HF. A basic introduction to MRI is provided as well, highlighting the methods for achieving contrast in images. The concepts of detection and classification with model observers is introduced along with basics of statistical decision theory. The properties of the linear and quadratic Hotelling Observer are described in this chapter. Finally, techniques for local texture analysis are discussed. The goal of this chapter is to present the tools that are necessary to understand the experiments described in the subsequent chapters.

1.1 The Liver and its Function

As the largest internal organ in the human body, the liver is responsible for multiple tasks including: regulation of the hepatobiliary system; working with the stomach, gallbladder, and pancreas in digestion, removing toxins from the blood and lymph systems; synthesizing albumin and clotting factors for use throughout the body and storage of a variety of necessary molecular components including vitamins and minerals.[1,2] Blood is supplied to the liver from both the hepatic artery and hepatic portal vein. After blood is transported through the liver, it exits through the hepatic vein to the inferior vena cava and back into the circulatory system.[2] Figure 1.1.1 shows the basic anatomy of the liver, including the basic operating unit, the hepatic lobule.

The boundary of the hepatic lobules is defined by a network of portal venules and arterioles. It is observed that these lobules range in size from approximately 0.8 mm to 1.5 mm in diameter. Each hepatic lobule contains groupings of the liver’s basic physiological cells, the hepatocytes, which filter incoming blood on its way to the hepatic vein. Hepatocytes perform the liver’s primary function of filtering nutrients and toxins from the blood for either storage or excretion from the body. If blood flow to the hepatocytes is restricted, this can lead to liver failure. One condition that can impede blood flow is hepatic fibrosis which is discussed in later sections.



From OpenStax[1]

Figure 1.1.1: A macroscopic to microscopic view of the liver showing the general liver shape and the arrangement of lobules. The hepatocytes are located within the lobules.[1]

1.2 Chronic Liver Disease and Hepatic Fibrosis

In the US in 2013 alone, 11.5 deaths out of 100,000 were related to CLD, making it the 12th in rank ordering of causes of death.[3] Many different diseases fall under the description of CLD. For example, CLD includes damage from drugs, such as alcohol, that strain the liver.[4,5] CLD also includes viral hepatitis, which comprises a variety of viral infections that attack the liver. Non-alcoholic steatohepatitis (NASH) is a condition the liver falls into when a patient is consuming a high-fat diet.[6,7] The hallmark symptom of these diseases is HF. If treatment is not given to a patient, the liver will likely progress to cirrhosis and the likelihood for hepatocellular carcinoma or liver failure greatly increases.

CLDs damage the liver in various ways, however, they all share the common result of the accumulation of an extracellular matrix (ECM) of collagen through the liver.[4] Figure 1.2.1 shows images of healthy and cirrhotic liver samples collected at the University of Arizona Banner-Health Center. All human studies were performed under informed consent with a protocol approved by the University of Arizona Institutional Review Board. Figure 1.2.1 (A) shows a healthy liver and has no ECM in the tissue. Figure 1.2.1 (B) is a photograph of a pathology slide taken from a biopsy of the liver in Figure 1.2.1 (A). In Figure 1.2.1 (C), the ECM has infiltrated the entire volume of the liver and a tumor has developed in the upper section. Figure 1.2.2 shows the same slide as Figure 1.2.1 D, but with the collagen network outlined for emphasis. These are the features pathologists look for when assessing fibrosis content. The pathology slides confirm that the healthy liver has no signs of an ECM whereas the cirrhotic liver has an ECM surrounding the lobules. In the presence of the ECM, liver function is inhibited, putting a patient at risk for disease.

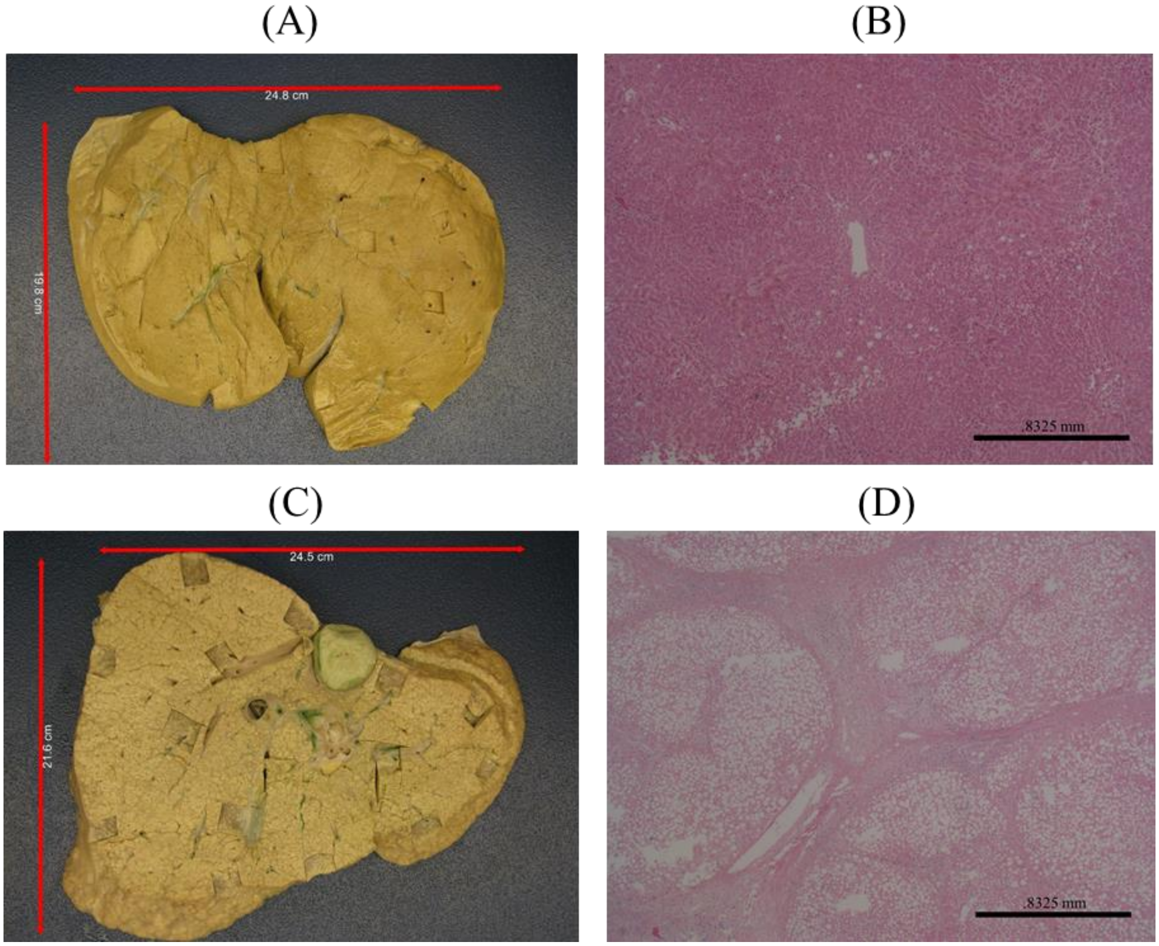


Figure 1.2.1: (A) Photograph of a healthy formalin fixed liver from autopsy and (B) an H&E pathology slide from biopsy. (C) Photograph of a cirrhotic liver recovered from autopsy and (D) an H&E pathology slide from biopsy.

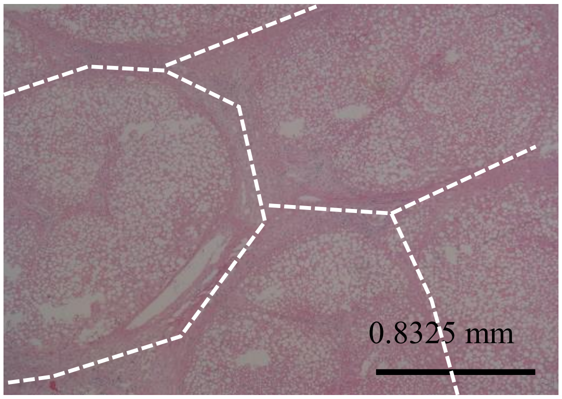


Figure 1.2.2: H&E pathology slide from biopsy from Figure 1.2.1 D with the collagen network highlighted

1.3 Needle Biopsy as the Gold Standard of Detection

The gold standard for diagnosis for HF is a needle biopsy.[6,8–15] Needle biopsy is an invasive surgical procedure where a $1mm^3$ tissue sample is taken from a patient's liver. The biopsy sample is fixed and sectioned for a pathologist to observe under a microscope. The basic stain used by pathologists is an H&E stain, which enhances the nuclei of the cells. A trichrome stain is used to mark collagen if the pathologist requires the additional contrast in order to make a diagnosis.

Biopsy offers a high-resolution image of the tissue, allowing a pathologist to observe the state of the tissue and report findings on fat content, carcinoma, and iron content, etc., along with HF. The disadvantages of biopsy are poor sampling of the subject as well as the risk for procedural complication.[6,10,11,14,16–18] The liver weighs approximately 1.3 kg in adults. The $1mm^3$ biopsy collected represents 1/50,000 of the entire liver. This large under-sampling has the chance of missing a region of the liver with early stage HF and sampling only healthy tissue, resulting in a false negative. In addition, pathologists only have tools to make a semi-quantitative diagnosis. Pathologists make calls based on pre-determined scales agreed upon by the practicing community.[11,12] Grading on these scales requires extensive training for pathologists and there is an inevitability of variation between pathologists.

When the pathologist reports the degree of HF to the physician there are different scales used to interpret the patient's status. One that has become popular with both pathologists and radiologists is the METAVIR score.[11,15,19–21] The METAVIR score conveys the fibrotic content of the liver and is independent of other features.[11] It is a five stage score, ranging from healthy with no fibrosis present as *F0,* to the end stage of the disease, cirrhosis, with a score of *F4.* The intermediate stages characterize points in the progression of the ECM as more damage is done to the liver.[12] The earliest stage where fibrosis is present is *F1*, where HF is characterized by the ECM developing around the portal veins and arteries.[11,12] These are the lattice points of the lobule structure and HF will progress if left untreated. *F2* is defined as the fibrosis from the portal veins branching into septa in the lobule lattice.[11,12] At this point, the start of an ECM lattice is observable. Once the septa branches begin to bridge the ECM between portal veins, this is described as *F3.* Once the ECM is fully connected, and the lobules are completely circumscribed by the ECM, the state of cirrhosis is reported and the METAVIR score is *F4.*[11,12] Figure 1.3.1 shows the progression of these stages.

Histopathology methods are capable of producing images with cellular resolution. As mentioned in section 1.2, the scale of fibrosis is 1-5mm. However, the low sampling rate of biopsy and variance in pathologist observers results in inconsistencies between diagnoses a risk of under reporting.

It is the goal of this dissertation to develop an alternative to needle biopsy performed in clinical practices. However, due to its ability to provide high-resolution images to a well-trained pathologist, and acceptance by the medical community, biopsy will serve as the gold standard for the experiments described in the forthcoming chapters.

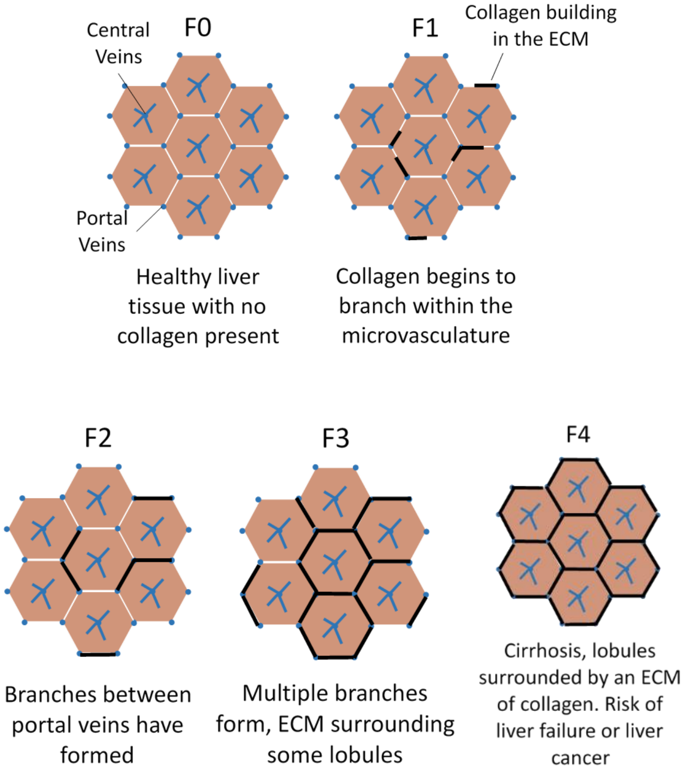


Figure 1.3.1 Representation of the METAVIR scoring system. Collagen is the black lined connecting portal veins.

1.4 Treatments for Chronic Liver Disease and Hepatic Fibrosis

Some diseases in the liver are treatable with changes in patient behavior, such as abstaining from alcohol in the case of alcoholic liver disease, or reducing weight in the presence of non-alcoholic fatty liver disease (NAFLD).[22] In other cases, it is necessary to identify and treat the cause of HF and not HF directly, such as with alcoholism.[22] Some drug treatments use anti-inflammatory drugs, however they do not directly treat fibrosis.[22]

Many different treatments have been attempted to reverse HF with various success rates. There are numerous pharmacological agents with both direct and indirect effects on fibrosis, but none have emerged as an ideal treatment.[23] Some drugs inhibit angiotensin II, such as Losartan.[23,24] Another group of drugs use interferon-α to treat damage to the liver, but there is limited documented success in human studies.[25–28] Another therapy is pentoxifylline, which has shown to affect down-regulation of collagen formation and other factors effecting HF.[23,29,30] A large summary of therapeutic options is available in *Pinzani et al.*[23]

CLD and HF are treatable diseases for patients if caught early. Biopsy, magnetic resonance elastography, and delayed phase contrast-enhanced MRI are all used for tracking the progression of CLD, however, an early non-invasive quantitative detection and screening method does not currently exist. The earlier the disease is caught the more treatment options are at the disposal of physicians.

1.5 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a powerful, still-developing, clinical imaging modality capable of non-invasively collecting tomographic images with strong soft tissue contrast.[14,31] MRI takes advantage of the presence of hydrogen nuclei in the body and recovers the response of the nuclear spin to changes in a locally generated magnetic field.

MRI is possible due to nuclear magnetic resonance (NMR), a phenomenon discovered in 1937 by Isidor Rabi, for which the Nobel prize was awarded in 1944. Rabi discovered the magnetic moment of the nucleus. NMR is a quantum physics phenomenon, but it is possible to understand MRI from a classical approach. The connection between the quantum and classical mechanics of NMR and MRI is possible by use of Ehrenfest's theorem applied to a collection of hydrogen nuclei. Ehrenfest's theorem states that the expectation values of quantum systems that correspond to classical variables obey classical laws of motion.[31–33]

It is often stated that the hydrogen nuclei behave like spinning tops, however this is only what the systems looks like when Ehrenfest's theorem is applied to a population of hydrogen atoms.[32]

The hydrogen atom consists of a proton in the nucleus and single electron; both particles are Fermions with spin ½. A spin ½ particle has only two possible eigenstates, spin up or spin down with respect to an externally defined z-axis. When the hydrogen atoms are in the body, it is the expectation value of the orientation of the spin angular momentum that is observed, and the result indeed processes like a spinning top in a gravitational field. However, what is truly occurring is the ratio of the states the hydrogen nuclei occupy are changing. The dynamics of spin angular momentum in an externally applied magnetic field are the basis of the NMR technique.

MRI is based on the NMR phenomenon of hydrogen atoms. The hydrogen atom consists of a proton in its nucleus and thus, has angular momentum, $\vec{\Psi}$, the arrow denotes a vector, and a magnetic moment, $\vec{\mu}$**,**

$$\vec{\mu} = \gamma\vec{\Psi}\boldsymbol{.}\ ^{33}$$

(1.5.1)

where, γ, a scalar, is the gyromagnetic ratio and has units of radians per second per tesla.

If a collection of nuclei, N_A, are in a volume, the microscopic magnetizations are summed to construct a bulk magnetization moment vector,

$$\vec{M}_0 = \sum_{n=1}^{N_A} \vec{\mu}_n.$$

(1.5.2)

When the bulk magnetic moment enters a magnetic field, $\overrightarrow{B_0}$, $\overrightarrow{M_0}$, will rotate with an angular velocity,

$$\omega_0 = \gamma B_0,\ ^{31,33}$$

(1.5.3)

known as the Larmor frequency. B_0 is the magnitude of the stationary magnetic field. Equation 1.5.3 has units of radians per second. When converted to cycles per second an equivalent expression for the Larmor frequency is,

$$\nu_0 = \frac{\gamma B_0}{2\pi}.$$

(1.5.4)

Knowledge of the Larmor frequency is useful for multiple reasons. First, and most important the radio frequency (RF) pulse that is used to excite the spins out of equilibrium, described later in this chapter, must have the same frequency as the Larmor frequency in order to observe the NMR phenomenon. The second advantage is that the Larmor frequency helps simplifying the mathematical formulations used to explain the NMR signal. The signals of the magnetization in the laboratory (stationary) frame of reference are dependent on the Larmor frequency, however if the frame of reference rotates at the Larmor frequency, the equations governing the signal will simplify.

So far, in our discussion, the spins are in equilibrium with M_0 along the direction of B_0 (typically defined along the z-axis). In order to observe $\overrightarrow{M_0}$, a temporary perturbation in the energy of the system is introduced to force the spins out of equilibrium. This is accomplished with an RF pulse, with frequency equal to the Larmor frequency, applied perpendicular to the external magnetic field. The RF pulse is capable of flipping the spins by an angle α (the flip angle, FA) from the z-axis into the transverse X-Y plane yielding a transverse component to the magnetization, $\overrightarrow{M_{xy}}$, defined as:

$$\overrightarrow{M_{xy}} = \overrightarrow{M_0} sin\alpha e^{i\phi}.^{33}$$

(1.5.5)

Once the RF pulse is turned off a receiver will detect the electro-magnetic field (emf) signal generated in the transverse plane.

Two relaxation mechanisms have an impact on the detected signal. The first is the longitudinal relaxation, T_1, which is an exponential recovery along z representing the move back toward equilibrium. The longitudinal magnetization is given by,

$$M_z(t) = M_0 \left(1 - (1 - \cos\alpha) e^{-t/T_1}\right)^{31}$$

(1.5.6)

and is dependent on both FA, and T_1. The relaxation of the transverse magnetization is characterized by T_2 and is a measure of the dephasing of all the spins in the transverse plane,

$$\overrightarrow{M_{xy}}(t) = \overrightarrow{M_0} \sin\alpha\, e^{i\phi} e^{-\frac{t}{T_2^*}}.^{33}$$

(1.5.7)

T_1 and T_2 are properties of the material in the magnetic field and offer two of the three main contrast mechanisms for MRI images. They have units of time, often reported in milliseconds, t in this section is time. The third type of contrast mechanism is proton density. By setting up different pulse sequences, images can have very different contrast between anatomical features.

During transverse relaxation, the collection of spins will process at different rates due to interactions with the fluctuating magnetic fields generated by neighboring spins (T_2) or local static fields (${T_2}^{\dagger}$). ${T_2}^{*}$is a combination of the two relaxation phenomena:

$$\frac{1}{{T_2}^{*}} = \frac{1}{T_2} + \frac{1}{{T_2}^{\dagger}} \tag{1.5.8}$$

Dephasing due to ${T_2}^{\dagger}$ relaxation can be reversed using a spin-echo experiment.[31,33] Dephasing due to T_2 cannot be rephased and the loss of signal is determined by the time between the excitation of spins into the transverse plane and the mid-point of signal detection (known as the time to-echo or *TE*) as well as the tissue properties.

MR images, as previously mentioned, have three different sources of contrast: T_1, T_2, and proton density. The MRI pulse sequence parameters determine what property dominates the signal of the observed image.

In order to recover a T_1-weighted image, a sample is excited rapidly, multiple times. The repetition time, *TR*, is the time between which the spins' successive excitations into the transverse plane. *TR* is set at a time close to the T_1 value of the sample in a T_1-weighted image. In order to minimize the effect of the T_2 relaxation, the TE is kept short. Recovering a T_2-weighted image requires TE to be set approximately to the value of T_2 in the sample, while keeping a long TR. This allows the effect on the signal to decay according to T_2, creating contrast in the T_2 realm.

The final contrast is proton density, which requires $M_{\mathbf{z}}$ to fully recover back to M_0 before each RF excitation. This is accomplished with a long TR and a short TE (to minimize $T_{\mathbf{2}}$-decay effects). The result is a signal that reflects the different proportions of hydrogen atoms within a voxel of tissue. TR, TE, and FA are three major variables controlled in an MRI pulse sequence. The precise arrangement of RF pulses and magnetic gradients are integral to MR image acquisition. An example of a gradient echo pulse sequence is shown in Figure 1.5.1.

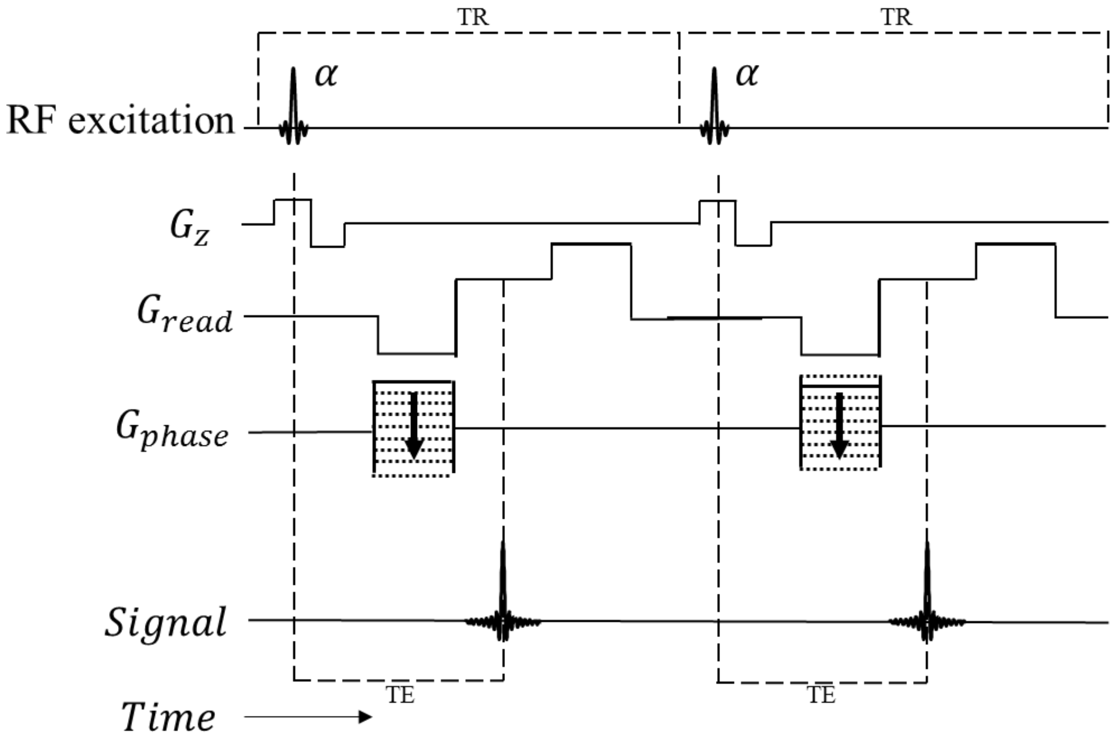


Figure 1.5.1: Example of a gradient echo MRI pulse sequence. The pulse sequence is repeatad using several RF excitations to record the adequate frequency data to reconstruct an image.

MRI data are collected by using magnetic field gradients which linearly change the frequency of the spins with locations. The signal collected in MRI contains the frequency content (K-space) of the object and the magnetic field gradient waveforms describe what K-space points are collected in time. A typical 2D gradient echo pulse sequence begins with an RF pulse that is calibrated to flip the spins to the prescribed FA as measured from the z-axis defined by the start portion of the magnetic field. This pulse is played while applying a slice-selective gradient, G_z, to select the slice plane. To encode spatial frequency data in $k_{\mathbf{x}}, k_{\mathbf{y}}$ in the slice plane, the G_{read} and G_{phase} magnetic field gradients are used. G_{read}, G_{phase}, are first turned on simultaneously to move to the (-$k_{x_{max}}$, $k_{y_{max}}$) position. $G_{phase}$is then turned off and the polarity of G_{read} is reverse to read-out a $k_{\mathbf{y}}$ line in K-space. After readout a strong gradient is applied to diphase any remaining signal in the transverse plane before the next FA excitation. This is repeated several times until all $k_{\mathbf{y}}$ lines are read. Figure 1.5.1 shows the two main pulse sequence parameters, *TE* and *TR*. The *TE* time defines the halfway point of the RF excitation to where half the signal is recorded to. *TR* is the time between FA excitations. Once all K-space data are obtained inverse Fourier methods are often employed to reconstruct the image. There is extensive research in the field of reconstruction and we will refer the reader to.[34–36] For this dissertation we use reconstruction methods provided by Siemens on its Skyra scanner. The sequence in Figure 1.5.1 is for a 2D-scan, the scans in this dissertation are performed with a 3D-gradient echo pulse sequence. The major differences between a 2D and 3D pulse sequence is a wide bandwidth for the RF pulse to excite the 3D volume of interest and a phase encoding gradient applied along G_Z.

While it is possible to retrieve contrast from a sample by use of manipulating MRI pulse sequences, it is also possible to temporarily enhance the contrast of the properties of the sample by use of contrast agents which are injected into a patient. Gadolinium (Gd), a rare earth metal ion, chelated with a stable molecule is commonly used in MRI contrast agents and was introduced in 1984.[31,33] Gd based contrast agents have an effect of shortening the T_1 of water molecules that are in contact with it. Gd based agents also shorten the T_2 of the molecules in the surrounding volume. However, since T_1 is usually much greater than T_2 in tissue and the difference in T_1 and T_2 for Gd is smaller, a greater percent change is observed in T_1 values. In HF MRI applications Gd based contrast agents are used increase contrast in the T_1 realm. The inverse of the observed T_1 relaxation is equal to the sum of the inverses of the relaxation of the components,

$$\frac{1}{T_{1_{Observed}}} = \frac{1}{T_{1_{Tissue}}} + \frac{1}{T_{1_{Gd}}}.^{31} \tag{1.5.9}$$

1.6 MRI as an Alternative to Biopsy

Delayed-phase Gd-enhanced MRI (DE-MRI) is capable of producing significant contrast between fibrotic and normal liver tissue, presumably due to accumulation of contrast agent in the extracellular space.[37,38]

Clinical MRI with breath holds and advanced pulse sequences are capable of resolving features near the apparent size of lobules associated with HF. Figure 1.6.1 shows one comparison of data collected from DE-MRI at the University of Arizona Medical Imaging center at 1.5T. It is clear from these images that DE-MRI is capable of producing contrast between healthy and diseased patients, which an experienced radiologist can detect at the late-stages cases of severe HF.

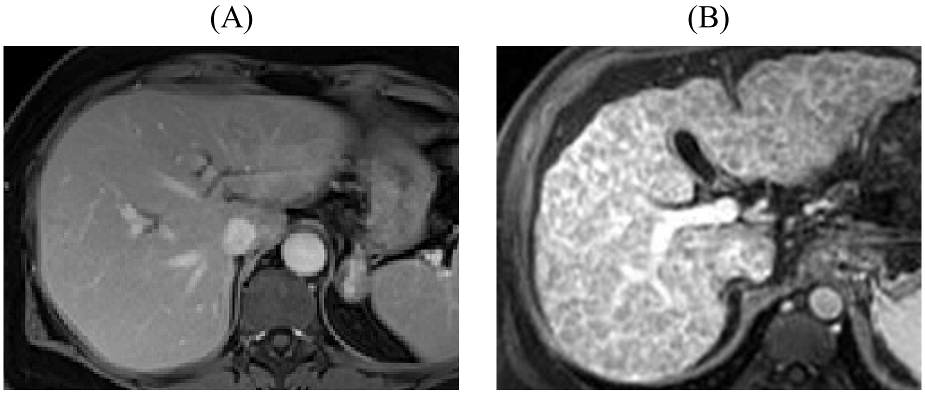


Figure 1.6.1: Examples of liver DE-MRI from *in vivo* scans. (A) Image from a healthy patient and no ECM is present. (B) Image from a patient with cirrhosis, an ECM is across the entire organ.

However, staging of HF based on DE-MRI images in the early stages of disease is much more difficult and requires specialized training of the radiologist.[14,39] As with pathologists, radiologists are subject to variability between readers. Radiologists look for gross versions of the same features pathologists look for in biopsy slides.[39,40] MRI provides the advantage over biopsy of sampling the entire liver volume, but without quantitative statistical analysis, the results are limited by many of the same drawbacks pathologists face in deciding on a score that describes the overall liver state.

Despite its invasive nature and potential complications biopsy the gold standard for diagnosing HF. Due to the ability to treat early HF and CLD, a globally sampled non-invasive test for fibrosis is valuable to both physicians and patients. MRI is capable of producing images with soft tissue contrast the is sensitive to the infiltration of the ECM associated with HF; it is the quantitative image-science tools that still need development.[14]

1.7 Magnetic Resonance Elastography

One approach to using MRI to assess HF quantitatively is via magnetic resonance elastography (MRE). MRE is based on the observation that as the ECM spreads throughout the liver, the stiffness of the liver increases. This is why external-palpation has traditionally been used as a basic detection of disease liver tissue.[41] MRE was developed by Ehman *et al.* at the Mayo to display liver stiffness data in a map (often displaying values estimated above a 95% confidence interval). Displaying the data as a map similar to an MRI image allows a radiologist to combine their knowledge of anatomy with observations made from stiffness maps. A map highlighting likely regions of HF with quantitative data are a potentially useful tool for the MRI clinic that illustrates CAD, computer-aided diagnosis.

MRE uses MRI to analyze an externally-generated shear wave that is administered via the patient's abdomen to estimate the stiffness of the liver.[16,42,43] The stiffness is estimated from the deformation in response to the pressure waves and is temporally measured with a phase-sensitive MRI technique.[44–47]

MRE-derived liver stiffness and extent of fibrosis has been shown to be positively correlated, however a study by Castéra *et al.* observed that one in five liver stiffness measurements fail to produce reliable results.[48] The technique is also difficult to implement, sometimes leads to inconclusive images, and can be expensive to add to an MR clinic's workflow. While MRE is often effective at quantifying liver stiffness, it is insensitive to the change of texture in early stages of HF, less than or equal to F2 disease. [46,45,48,49]

MRE has made many advances, yet is still seen as a developing method[14]. MRE only observes a feature that correlates with the structure present. It is not a direct measurement of presence of a texture that pathologists and radiologists can view with imaging approaches. Therefore, a quantitative, local measurement of HF texture with broad sampling across the liver would represent a major advance towards early detection and would allow monitoring of response to therapies and lifestyle changes designed to reverse collagen infiltration.

1.8 Statistical Decisions for Imaging

In order to aid a radiologist in staging HF, a well-defined method to make unbiased decisions is desirable. Statistical decision theory is useful to provide a framework to develop mathematical observers to perform the task of staging liver fibrosis from MRI images. We review first the concepts of the optimal linear observer, the quadratic observer, and classification tasks that are necessary for the work presented in following chapters.

An imaging system, optical or otherwise, is designed to produce information in which features are extracted by an observer, human or mathematical. An imaging system can be represented as,

$$\vec{g} = \boldsymbol{\mathcal{H}}\vec{f} \tag{1.8.1}$$

Where vector $\vec{g}$ is an image collected from object vector $\vec{f}$ from imaging system $\boldsymbol{\mathcal{H}}$, represented as a matrix in the case of a voxelized representation of an object or an integral transformation in the case of a continuous object. Often a desired task to perform is to use $\vec{g}$ to classify the object as a member of one category or another.

In the simplest classification task the single question asked of an observer is if a signal present or not. If a signal is not present, this is often described as the *signal absent*, noted as state or hypothesis H_0, while *signal present is noted as* H_1. The goal in this dissertation is to develop a model observer to perform the task of analyzing images collected from MRI. When constructing an observer, the goal is to maximize task performance using available statistical information.[50–52]

To build many of the arguments for the multidimensional problem, we begin with a two dimensional demonstrational problem consisting of just two pixels. A mathematical observer will have a form $t(\vec{g})$ and produce a scalar result known as a *test statistic*. Now, t is the test statistic and has units related to the image data in $\vec{g}$. The maximum performance of an observer is to fully separate images into the correct hypothesis groups based on imposing the test statistics to a threshold t_c.[50] A simplified picture of this task is shown in Figure 1.8.1.

The result of a single experiment based on Figure 1.8.1 has four possible outcomes. An object will either be a member of the signal absent or signal present classes, and is classified by $t(\vec{g})$ as a member of one or the other. The four possible outcomes are highlighted in Table 1.8.1. A perfect observer will only report *true positive* and *true negatives*. However, some fraction of *false positives* and *false negatives* are almost always unavoidable in real medical imaging data[50].

	H_1 is true	H_0 is true
Decide H_1 is true	True Positive	False Positive
Decide H_0 is true	False Negative	True Negative

Table 1.8.1: Possible decision outcomes

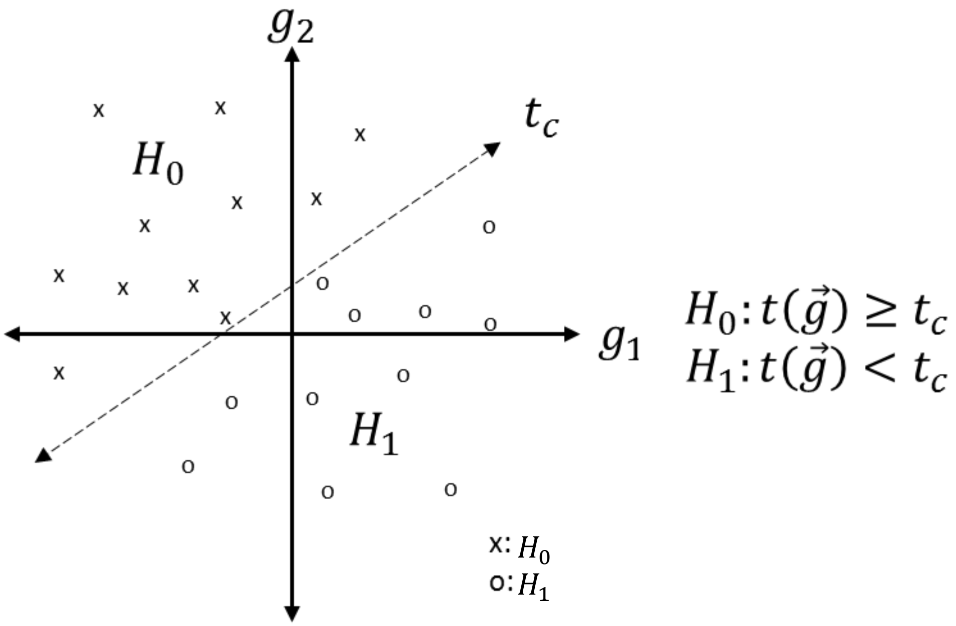


Figure 1.8.1: A two pixel system separated by test statistic $t(\vec{g})$

statistic

A threshold for $t(\vec{g})$ must be chosen to separate results into the categories D_0 and D_1. The statistics of D_0 and D_1 relative to H_0 and H_1will determine the *sensitivity* and *specificity* of the observer. The total number of experiments is equal to the sum of all *true positive, false positive, true negative,* and *false negative* decisions. If the number of experiments is large, the true- and false- positive fractions approach to the underlying probabilities are given by:

$$TPF = \Pr(D_1|H_1) = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{1.8.2}$$

$$TNF = \Pr(D_0|H_0) = \frac{N_{TN}}{N_{TN} + N_{FP}} \tag{1.8.3}$$

$$FPF = \Pr(D_1|H_0) = \frac{N_{FP}}{N_{TN} + N_{FP}} = 1 - TNF \tag{1.8.4}$$

$$FNF = \Pr(D_0|H_1) = \frac{N_{TN}}{N_{TP}+N_{FN}} = 1 - TPF\,.^{50} \tag{1.8.5}$$

The *TPF* is the *sensitivity,* which is appropriate, since a high *sensitivity* is associated with the ability to effectively detect the presence of the non-normal state. *TNF* is the *specificity.* A test with high specificity is one with a low fraction of false positive outcomes. How one sets a threshold to choose an acceptable *sensitivity* and *specificity* to operate an observer at is not the focus of this dissertation. That decision relies on information not always available, such as the relative “costs” associated with over-diagnosing versus under-diagnosing disease.

Using the results of repeated experiments, the probability density functions are generated as function on $t(\vec{g})$. An example of this is shown in Figure 1.8.2 where it is clear the TPF, FPF, TNF, and FNF are all fractional areas under the probability density curves:

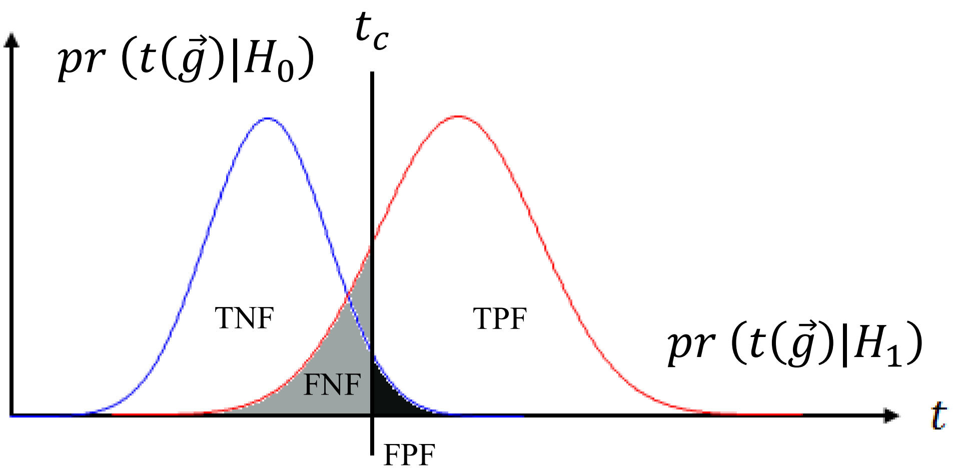


Figure 1.8.2 Example of the probability density functions for a test statistic with two hypothesis and one threshold. The TNF, FNF, FPF, and TPF are labeled based on the threshold.

$$TPF = \Pr(t \geq t_c | H_1)$$

(1.8.6)

$$FPF = \Pr(t \geq t_c | H_0)$$

(1.8.7)

$$TNF = 1 - \text{FPF}$$

(1.8.8)

$$TPF = 1 - \text{TPF}.$$

(1.8.9)

Equations (1.8.6) and (1.8.7) are dependent on t_c, a decision threshold for the observer. Plotting the TPF as a function of FPF at every possible t_c will generate a plot known as the receiver operator characteristic curve (ROC curve). The ROC curve is a measure of the performance of the observer and is one possible figure of merit. An example ROC curve based on Figure 1.8.2 is given in Figure 1.8.3. The separation of the two probability distribution functions results in the curve's shape.

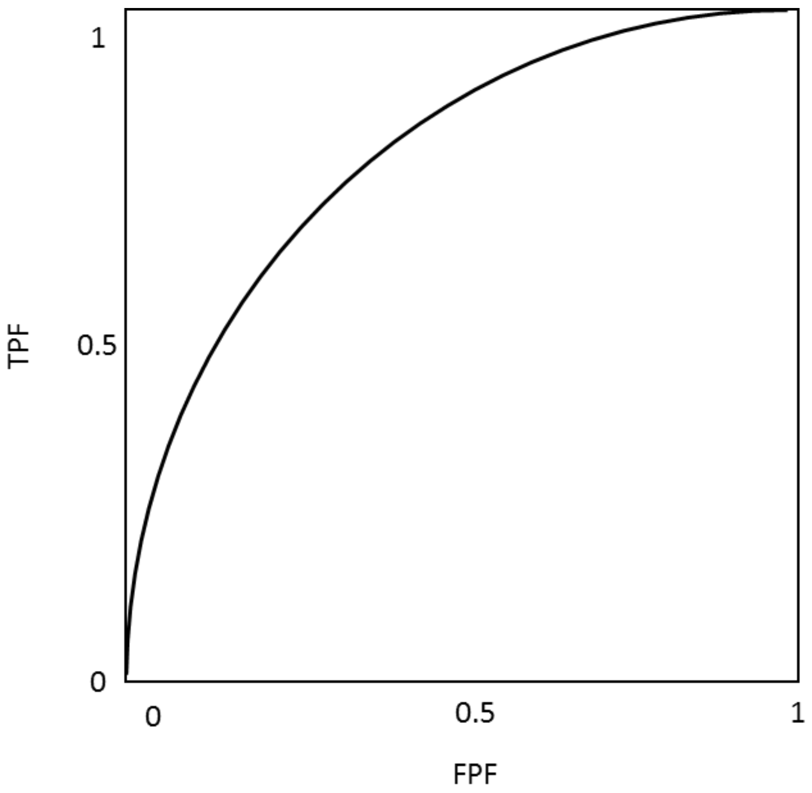


Figure 1.8.3 An Example ROC Curve

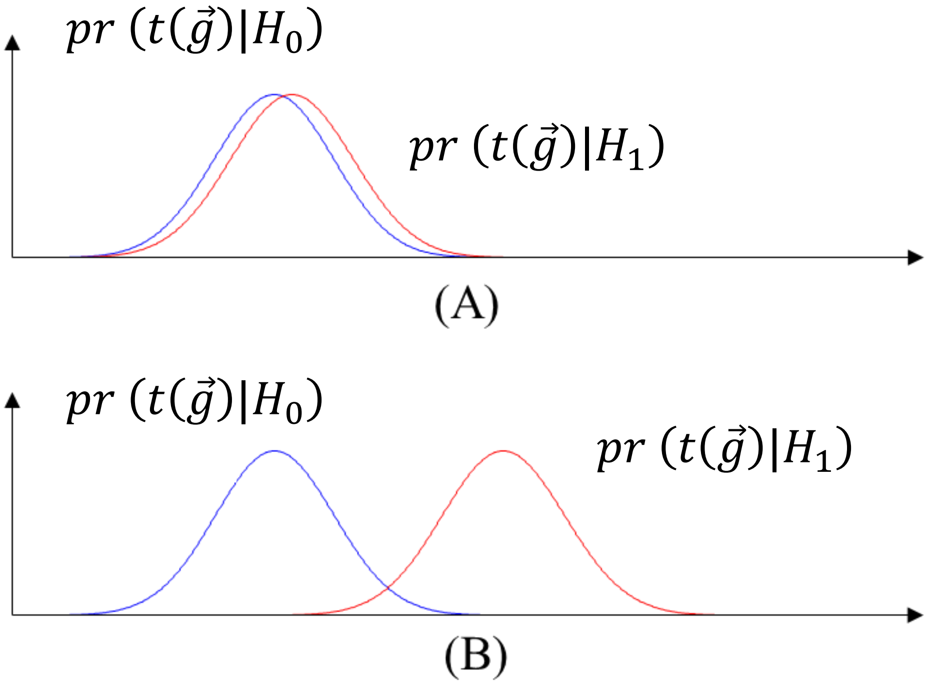


Figure 1.8.4 (A) The images are nearly identical with an observer resulting in an AUROC ≈ 0.5 (B) The images are significantly different, resulting in an almost perfect separation with an AUROC ≈ 1

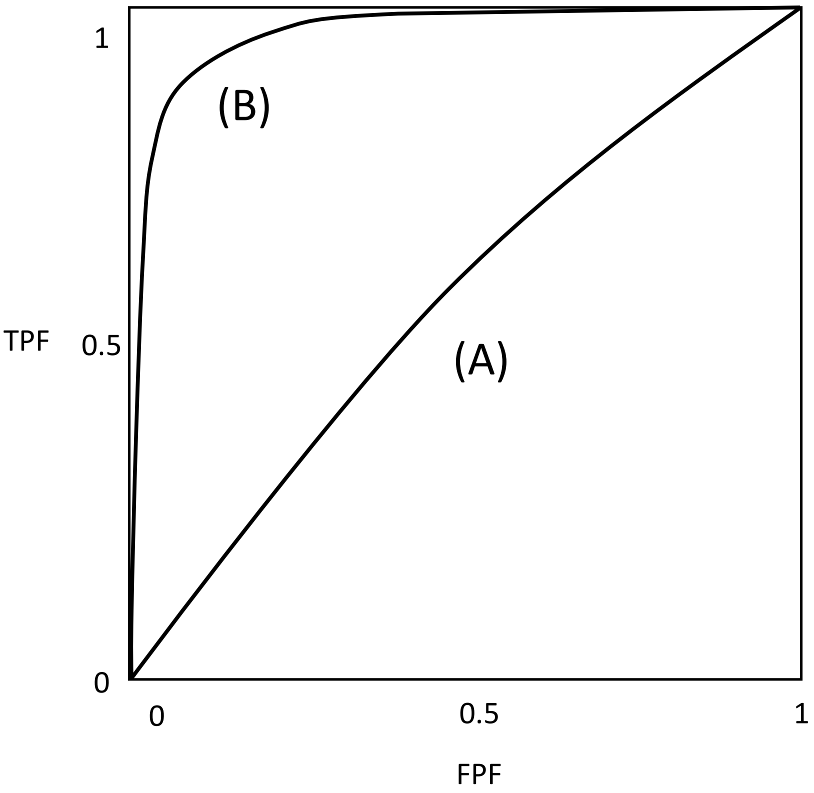


Figure 1.8.5 (A) ROC curve resulting from the probability density functions illustrated in Figure 1.8.4 (A), (B) ROC curve resulting from the probability density functions illustrated in Figure 1.8.4 (B)

A figure of merit function that characterizes the ROC curve and thereby the efficacy of the imaging method and observer is measure of the area-under-the-receiver-operator-characteristic-curve (AUROC or AUC for short). The AUROC is the integral of the ROC curve:

$$AUROC = \int_0^1 TPF(FPF)dFPF$$

(1.8.10)

Based on the integral and limits of the ROC curve it is clear the maximum value of the AUC is 1.0 and minimum value is 0. An AUROC equal to 1 represents a complete separation of the two probability density functions and perfect performance. Figure 1.8.4 (A) is an example of an imaging method and observer that are performing no better than a "coin flip" or "random guess." The two distribution functions are nearly on top of one another. For each decision threshold, the observer finds equal probability of whether the object in one class or the other. The ROC curve for this experiment is shown in Figure 1.8.5 (A) and the AUROC value is approximately 0.5.

An optimal imaging method and observer would have results similar to the probability density functions shown in Figure 1.8.4 (B) where the PDF's have almost no overlap, resulting in an ROC curve similar to Figure 1.8.5 (B). If an AUROC below 0.5 is recovered, there is a sign error in the observer function i.e. one should change the decision from $t(\vec{g}) \geq t_c$ to $t(\vec{g}) \leq t_c$ or vice versa.

1.9 Mathematical Observers

If complete statistics are known about the *signal absent* and *signal present* cases, which is usually never true, an optimal decision strategy is possible. Is has been shown to be that a ratio of the probability densities on the data given the signal absent and signal present hypotheses,

$$\Lambda(\vec{g}) \underset{D_0}{\overset{D_1}{\gtrless}} \frac{\Pr(t(\vec{g})|H_0)}{\Pr(t(\vec{g})|H_1)},$$

(1.9.1)

leads to an optical decision strategy.[50,52] This is known as the likelihood ratio and requires entire knowledge of the system statistics including variability in objects and the noise characteristics of the imaging system. The *maximum-likelihood criteria, Neyman-Pearson criterion, minimum average cost,* and *minimum error detector* are all decision strategies that lead to the conclusion to choose H_1 if,

$$\Lambda(\vec{g}) > \Lambda_c$$

(1.9.2)

where Λ_c is the threshold associated with the test statistic[50]. A test that takes the form of 1.9.2 is a likelihood-ratio test and is the form of the theoretically ideal observer.

Recovering the ideal observer for a classification task is a difficult procedure and often requires knowledge that is generally impossible to measure or estimate accurately.

Since the AUC is a measure of the overlap and separation of the probability density functions describes the observer acting on the data, it can be used as a basis for choosing the observer. The ideal linear observer is one that maximizes the signal-to-noise-ratio of the test statistic, defined as,

$$SNR = \frac{\langle t(\vec{g})\rangle_1 - \langle t(\vec{g})\rangle_0}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_0^2}}. \tag{1.9.3}$$

We will show in the next section that the Hotelling observer is the ideal linear observer when the data have normal statistics. The true SNR of a task cannot be known, unless the true statistics of the system are known, but comparing the SNR of two observers is one method to determine which performs closer to the ideal observer.

1.10 The Hotelling Observer as the Ideal Linear Observer

In the previous section, we discussed the figures of merit for observers where full knowledge of the statistics is known. However, in practice there is limited knowledge of the variability between patients and statistics of the MRI imaging system, and the initial hypothesis is that the data are linear. A linear discriminant generates an observer that converts multidimensional data into a scalar test statistic via an inner product,

$$t(\vec{g}) = \vec{w}^{\dagger}\vec{g}.$$ [50,53]

(1.10.1)

The square of the SNR of this observer is governed by equation (1.9.3):

$$SNR^2 = \frac{(\langle t(\vec{g})\rangle_1 - \langle t(\vec{g})\rangle_0)^2}{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_0^2}.$$

(1.10.2)

The numerator of 1.1010.2 is charecterized by a difference in the means of the *signal absent* and the *signal present* image data:

$$\langle t(\vec{g})\rangle_0 = \langle \vec{w}^{\dagger}\vec{g}\rangle_0 = \vec{w}^{\dagger}\langle \vec{g}\rangle_0$$

$$\langle t(\vec{g})\rangle_1 = \langle \vec{w}^{\dagger}\vec{g}\rangle_1 = \vec{w}^{\dagger}\langle \vec{g}\rangle_1$$

$$\langle t(\vec{g})\rangle_1 - \langle t(\vec{g})\rangle_0 = \vec{w}^{\dagger}[\langle \vec{g}\rangle_1 - \langle \vec{g}\rangle_0] = \vec{w}^{\dagger}\overline{\Delta g}$$

$$\overrightarrow{\overline{\Delta g}}_m = \langle \overrightarrow{g_m}\rangle_1 - \langle \overrightarrow{g_m}\rangle_0.$$ [53]

(1.10.3)

Where the brackets represent the expectation value of the mth pixel in the image g. The denominator is recognized as the average of the variances. The variance of an image is characterized by the covariance matrix ***K***,

$$K_{mn} = \langle (g_m - \langle g_m\rangle_0)(g_n - \langle g_n\rangle_0)\rangle_0.$$ [53]

(1.10.4)

Where K_{mn} is the covariance matrix element of the mth row and nth column. The SNR, defined by terms of equations (1.10.3) and (1.10.4) is,

$$SNR^2 = \frac{\left(\vec{w}^{\dagger}\overrightarrow{\Delta g}\right)^2}{\vec{w}^{\dagger}\boldsymbol{K}\vec{w}}. \tag{1.10.5}$$

The linear observer that maximizes the SNR is the Hotelling observer. To maximize the SNR, one initially takes the natural logarithm, a monotonic transformation that does not move the location of the minima or maxima,

$$\ln(SNR_{\tau}^2) = \ln(\vec{w}^{\boldsymbol{t}}\overrightarrow{\Delta g})^2 - \ln \vec{w}^{\boldsymbol{t}}\left(\frac{1}{2}(\boldsymbol{K_0} + \boldsymbol{K_1})\right)\vec{w} \tag{1.10.6}$$

where the average covariance is presented as the average of the two separate covariance matrices from hypothesis data. To maximize SNR, the gradient is taken with respect to $\boldsymbol{w}$ and setting the it to 0,

$$\left(\frac{\boldsymbol{1}}{\boldsymbol{2}}(\boldsymbol{K_0} + \boldsymbol{K_1})\right)\vec{w} = \left[\frac{\vec{w}^{\boldsymbol{t}}\left(\frac{1}{2}(\boldsymbol{K_0}+\boldsymbol{K_1})\right)\vec{w}}{\vec{w}^{\boldsymbol{t}}\Delta\vec{g}}\right]\overrightarrow{\Delta g}\boldsymbol{.} \tag{1.10.7}$$

The coefficient to $\overline{\Delta \boldsymbol{g}}$ in equation 1.10.8 is a scalar that is set to 1,

$$\left[\frac{\vec{w}^{\boldsymbol{t}}\left(\frac{1}{2}(\boldsymbol{K_0} + \boldsymbol{K_1})\right)\vec{w}}{\vec{w}^{\boldsymbol{t}}\Delta\vec{g}}\right] = \boldsymbol{1},$$

$$\vec{w}^{\boldsymbol{t}}\left(\frac{\boldsymbol{1}}{\boldsymbol{2}}(\boldsymbol{K_0} + \boldsymbol{K_1})\right)\vec{w} = \vec{w}^{\boldsymbol{t}}\overrightarrow{\Delta g},$$

$$\left(\frac{\boldsymbol{1}}{\boldsymbol{2}}(\boldsymbol{K_0} + \boldsymbol{K_1})\right)\vec{w} = \overrightarrow{\Delta g},$$

and the Hotelling observer is defined as

$$\vec{w} = \left[\frac{1}{2}(\boldsymbol{K}_1 + \boldsymbol{K}_0)\right]^{-1} \overrightarrow{\overline{\Delta g}}.$$

(1.10.8)

A benefit of the Hotelling observer is only first-order (means) and second-order (covariance) statistics are necessary to define the observer. This is beneficial, since often one does not have knowledge of the underlying probability densities of the hypothesis data and most instead estimate them from training data. The Hotelling observer is also known as the pre-whitened matched filter in the field of electrical engineering, due to the step involving the matched filter, $\overrightarrow{\overline{\Delta g}}$, and what is referred to as a pre-whitening step, that uses the average covariance matrix as weights for the matched filter.[50,51,53] The Hotelling observer is the ideal observer when the data are Gaussian, there is a difference in the means, and the covariance matrices are equal. The Hotelling observer maximizes the SNR, equation 1.9.2 for these types of images. If image data does not meet these requirements, then a different observer may perform better.

1.11 The Quadratic Observer

One method to recover an observer is through the likelihood ratio of the conditional probability of two classes. We continue to assume the data are Gaussian, but now we consider the situation that the covariance matrices are not equal.

$$pr(t|H_0) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{\frac{-(t-\mu_0)^2}{2\sigma_0^2}}$$

$$pr(t|H_1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{\frac{-(t-\mu_1)^2}{2\sigma_1^2}}$$

(1.11.1)

These PDF's are used to write a likelihood ratio as from 1.9.1

$$\frac{\frac{1}{\sigma_1\sqrt{2\pi}} e^{\frac{-(t-\mu_1)^2}{2\sigma_1^2}}}{\frac{1}{\sigma_0\sqrt{2\pi}} e^{\frac{-(t-\mu_0)^2}{2\sigma_0^2}}} \underset{<}{>} \Lambda\,.$$

(1.11.2)

To find a threshold, t, the natural logarithm, of both sides is taken,

$$\ln\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{(t-\mu_1)^2}{2\sigma_1^2} + \frac{(t-\mu_0)^2}{2\sigma_0^2} \underset{<}{>} \ln(\Lambda),$$

$$\ln\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{t^2 - 2t\mu_1 + {\mu_1}^2}{2\sigma_1^2} + \frac{t^2 - 2t\mu_0 + {\mu_0}^2}{2\sigma_0^2} \underset{<}{>} \lambda\,,$$

$$\ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{t^2}{2\sigma_0^2} - \frac{t^2}{2\sigma_1^2} + \frac{2t\mu_1}{2\sigma_1^2} - \frac{2t\mu_0}{2\sigma_0^2} - \frac{{\mu_1}^2}{2\sigma_1^2} + \frac{{\mu_0}^2}{2\sigma_0^2} \underset{<}{>} \lambda\,,$$

$$\ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{t^2}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) + t\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2}\right) + \left(\frac{{\mu_0}^2}{2\sigma_0^2} - \frac{{\mu_1}^2}{2\sigma_1^2}\right) \underset{<}{>} \lambda\,.$$

(1.11.3)

The ROC curve is insensitive to monotonic transformations, allowing assumptions to be made that simplify the observer[50]. If we normalize and shift such that $\sigma_0 = 1$ and $\mu_0 = 0$, equation (1.11.3 simplifies to,

$$\ln\left(\frac{1}{\sigma_1}\right) + \frac{t^2}{2}\left(1 - \frac{1}{\sigma_1^2}\right) + t\left(\frac{\mu_1}{\sigma_1^2} - \frac{0}{1}\right) + \left(\frac{0}{2} - \frac{{\mu_1}^2}{2\sigma_1^2}\right) \underset{<}{>} \lambda\,,$$

$$\ln\left(\frac{1}{\sigma_1}\right) + \frac{t^2}{2}\left(1 - \frac{1}{\sigma_1^2}\right) + t\,\frac{\mu_1}{\sigma_1^2} - \frac{{\mu_1}^2}{2\sigma_1^2} \begin{array}{c} > \\ < \end{array} \lambda.$$

(1.11.4)

A term in (1.11.4 with no dependence on *t* is equivalent to a shift, and has no effect on the ROC. The decision by a quadratic observer is,

$$\frac{t^2}{2}\left(1 - \frac{1}{\sigma_1^2}\right) + t\,\frac{\mu_1}{\sigma_1^2} \begin{array}{c} > \\ < \end{array} \lambda.$$

(1.11.5)

The quadratic observer template is found with similar steps, starting by assuming data are Gaussian, with its own covariance matrix,

$$pr(\vec{g}|H_n) = [(2\pi)^M \det(\boldsymbol{K}_n)]^{-1/2} \exp\left[\frac{-1}{2}\left(\vec{g} - \bar{\vec{g}}_n\right)^t \boldsymbol{K}^{-1}\left(\vec{g} - \bar{\vec{g}}_n\right)\right],$$

(1.11.6)

and takes the natural logarithm of the ratio of the two likelihoods,

$$t(\vec{g}) = ln\left(\frac{pr(\vec{g}|H_1)}{pr(\vec{g}|H_0)}\right) =$$

$$\ln\left(\det(\boldsymbol{K_0})^{-1/2}\right) - \ln\left(\det(\boldsymbol{K}_1)^{-1/2}\right) +$$

$$\tfrac{1}{2}\left(\vec{g} - \overline{\vec{g}_0}\right)^t \boldsymbol{K_0}^{-1}\left(\vec{g} - \overline{\vec{g}_0}\right) - \tfrac{1}{2}(\vec{g} - \overline{\vec{g}_1})^t \boldsymbol{K_1}^{-1}(\vec{g} - \overline{\vec{g}_1}).$$

(1.11.7)

The determinant of the inverse of $\boldsymbol{K}_0$ and $\boldsymbol{K}_1$ are both assumed to be small compared to the remainder of the observer. These terms are ignored, resulting in the final form of the quadratic observer,

$$t(\vec{g}) = \frac{1}{2}\left(\vec{g} - \overline{\vec{g}_0}\right)^t \boldsymbol{K_0}^{-1}\left(\vec{g} - \overline{\vec{g}_0}\right) - \frac{1}{2}(\vec{g} - \overline{\vec{g}_1})^t \boldsymbol{K_1}^{-1}(\vec{g} - \overline{\vec{g}_1}). \tag{1.11.8}$$

As with the Hotelling observer, the quadratic observer only needs knowledge of the first- and second- order statistics to compute a test statistic. It is thus efficient to test the quadratic observer at the same time as testing a Hotelling observer, since both require the same first- and second-order statistics recovered from training data.

If the data are Gaussian, have different means, and different covariance matrices, then the quadratic observer will maximize the SNR, and will be the ideal observer. It is often a bad assumption that the image data are Gaussian, but it is a good starting point. If these methods do not work, then exploring alternative observers with simpler statistics is a valid option. Examples of these are the channelized Hoteling observer or channelized quadratic observer.[54]

1.12 Local Texture Analysis

The ECM associated with HF has a repeating cellular structure with characteristic dimensions on the order of 1-5mm. The liver is a three dimensional object and the lobules are locally close to a crystalline formation, but without long range order. The ECM defines this space and results in contrast between the tissue and the ECM in DE-MRI. Normal liver has little structure and images appear less textured.

The Hotelling and quadratic observers depend on sampling the mean of the two classes to train the observers. To perform local texture analysis, one starts at a random location in the object and select independent, non-overlapping, adjacent ROI's. The texture from HF also has a random size and distribution. The texture across ROI's may not stay in phase across the entire liver, which will lead to the mean of texture washing out across many observations.

Local texture analysis with autocorrelation or wavelet analysis emphasizes changes in texture in the ROI and removes the effects of the local shifts and rotations in the image data. Texture is identified by the local cellular structure, not a global arrangement of the collagen "lattice".

1.12.1 2D Autocorrelation

The first analysis method tested is the 2D discrete autocorrelation (2DAC), a special case of the 2D discrete cross-correlation. The 2DAC is a measure of a function's self-similarity. For a MxN image vector $\vec{f}$, the 2DAC is given as:

$$R'_{k,l} = \sum_{m=1}^{M}\sum_{n=1}^{N} f_{m,n}\, f^{*}{}_{m-k+1,n-l+1} \qquad \begin{matrix} 1 \le k \le 2M-1 \\ 1 \le l \le 2N-1 \end{matrix}$$

$$R_{k,l} = \frac{R'_{k,l}}{R'_{max}}. \tag{1.12.1}$$

The autocorrelation function yields a larger output vector than the input vector. A NxN image yields a $2N-1x2N-1$ 2DAC. For example, the autocorrelation function for a

7x7 pixel image yields an output with a size of 13x13 pixels and a characteristics 2-D pyramidal background.

1.12.2 The 2D Discrete Circular Autocorrelation

Even though it has symmetry, the training requirement for a 2DAC is large compared to the initial image it starts with. One method to reduce the training requirement size is to use a local texture analysis operator whose output is the same size as the input. The 2D discrete circular autocorrelation (2DCC) provides this result,

$$S'_{k,l} = \sum_{m=1}^{M} \sum_{n=1}^{N} f_{m,n}\, f^{*}_{m-k+1\sigma(M),n-l+1\sigma(N)} \qquad \begin{array}{c} 1 \leq k \leq M \\ 1 \leq l \leq N' \end{array}$$

$$S = \frac{S'_{k,l}}{S'_{max}}.$$

(1.12.2)

The 2DAC is also measure of self-similarity as a function of 2D offset, but the measure wraps, due to the modulo shift also removing the lag-dependent background. The result is a NxN image has an NxN 2DCC and will have a smaller training requirement than the analysis performed by the 2DAC.

1.12.3 Wavelet Analysis

While an autocorrelation is effective for detecting a cellular texture in an image, the infiltration of HF also changes the frequency spectrum of the image. One start to this approach would begin with a Fourier transform. However, a Fourier transform frequency analysis generally applies effectively to an entire image. In a clinical MRI image, the 2D

Fourier transform would include information from any location in the plane of the image. Instead, a method sensitive to a local change in frequency is needed to detect the presence of the ECM.

The wavelet transform (WT) converts the signal into a function of two variables, one associated with frequency and the other with location, $f(\vec{x})$ is transformed into $\vec{\omega}(a,b)$ in which a is associated with frequency content and b contains spatial information. The 1-D wavelet transform is given by:

$$\vec{\omega}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(\vec{x}) h^{*}\left(\frac{\vec{x}-b}{a}\right) dx$$

(1.12.3)

where $f(\vec{x})$ is the analyzed image and $h(x)$ is the chosen wavelet. Valid wavelet functions have limiting conditions. First, the wavelet and function must not have infinite integrals. Second, it is preferred that$\vec{f}(x)$ is recoverable by an inverse wavelet transform. This limits the family of wavelets to an orthonormal set:

$$\int_{-\infty}^{\infty} h_{mn}{}^{*}(\vec{t})\, h_{m'n'}{}^{*}(\vec{t}) dt = \delta_{mm'} \delta_{nn'}.$$

(1.12.4)

Due to the complexity of the problem of analyzing liver tissue with MRI images, a simple wavelet is desired to ensure the features extracted from the wavelet transform are properties of the image and not of the chosen wavelet. With this consideration, the Haar wavelet was chosen to analyze the target images. The data analyzed is real and positive and so are the wavelets. The sampling points must be chosen as:

$$a_j = 2^{-j},$$

(1.12.5)

$$b_{jk} = 2^{-j}k.$$

(1.12.6)

The wavelet is defined as:

$$h_{jk}(\vec{t}) = \frac{1}{\sqrt{a_j}} h\left(\frac{\vec{t}-b_{jk}}{a_j}\right) = 2^{\frac{j}{2}} h\left(2^j \vec{t} - k\right).$$

(1.12.7)

The wavelet transform is then:

$$\omega_{jk} = w\left(a_j, b_{jk}\right) = \int_{-\infty}^{\infty} f(\vec{t})\, h_{jk}^{*}(\vec{t}) dt.$$

(1.12.8)

However, this is not useful unless $\vec{f}(t)$ is in vector form. A low pass filter to be applied to $\vec{f}$ to down sample the function. A high pass filter representing the wavelet is applied to retrieve the wavelet coefficients for the same level,

$$f_k^{j-1} = \sum_{n=0}^{2N-1} f_{2k+n}^{j}\, l_n,$$

(1.12.9)

$$w_k^{j-1} = \sum_{n=0}^{2N-1} f_{2k+n}^{j} \, h_n.$$

(1.12.10)

Where the Haar wavelet filters are defined as:

$$l = \frac{1}{\sqrt{2}}[1 \quad 1] \, , h = \frac{1}{\sqrt{2}}[1 \quad -1]$$

(1.12.11)

The DWT is a separable transform, meaning the expansion into 2D involves performing the 1D DWT sequentially to the rows and then the columns of the image $\vec{f}$.

The 2D wavelet coefficients are often associated with the low pass, horizontal, vertical, and diagonal coefficients. The filters are applied to the matrix in both directions, and much more efficient methods are used to actually retrieve the coefficients. However, each group of coefficients can be illustrated through four separate operations:

$$(1) \; w_{l,l} = f_{k,m}^{j-1} = \sum_{n=0}^{2N-1} \sum_{m=0}^{2N-1} l_n l_m f_{2k+n,2k+m}^{j}$$

$$(2) \; w_{l,h} = \sum_{n=0}^{2N-1} \sum_{m=0}^{2N-1} l_n h f_{2k+n,2k+m}^{j}$$

$$(3)\ w_{h,l} = \sum_{n=0}^{2N-1} \sum_{m=0}^{2N-1} h_n l_m f^{j}_{2k+n,2k+m}$$

$$(4)\ w_{h,h} = \sum_{n=0}^{2N-1} \sum_{m=0}^{2N-1} l_n l_m f^{j}_{2k+n,2k+m}\,.$$

(1.12.12)

N is the number of pairs in the wavelet, which for the Haar wavelet is $N = 2$. Many wavelet transforms will refer to each group of coefficients as the approximation, horizontal, vertical, and diagonal coefficients respectively. j is associated with the number of times a function can be down sampled before reaching a DC value. For a vector of length 8, $j = 3$, therefore it is possible to recover 3 levels.

1.13 Goal of the Following Work

We have presented the tools necessary to discuss the experiments and results of training and testing mathematical observers to detect texture in the liver from DE-MRI images. The primary goal is to train an observer and test the 2DAC and 2DCC as viable texture analysis methods. For computer-aided detection of HF.

2. Designing a Phantom

The hallmark of hepatic fibrosis is the emergence of collagen that progresses from portal veins, and bridges to form a *network* of fibrous tissue. As the collagen burden increases, the liver tissue becomes dominated by the introduced texture. The microscopic structure underlying this texture is the feature pathologists use to assess liver fibrosis in biopsy samples. Developing mathematical (computer) observers that perform texture analysis to detect hepatic fibrosis requires training and testing data that includes known-normal and known-diseased livers. There are two options to obtain that data. One option is to collect patient data from biopsy-confirmed patients, and build training and testing data from libraries of clinical scans. The other option is to use phantoms that mimic the images of *in-vivo* Gd-enhanced delayed-phase MRI in the liver. The first option is expensive, time consuming, and requires that at least some biopsies are carried out on healthy livers. The second option is much less expensive, and allows repeated experiments without the tissue changing over time.

2.1 Observation of Livers from Autopsy

Early in the project, clinical scans were collected, and eventually it was suggested to recover a liver from autopsy and place it in the scanner to collect images. Tissue collected from autopsy is fixed in formalin to prevent the tissue from deteriorating without the blood supply of the body.

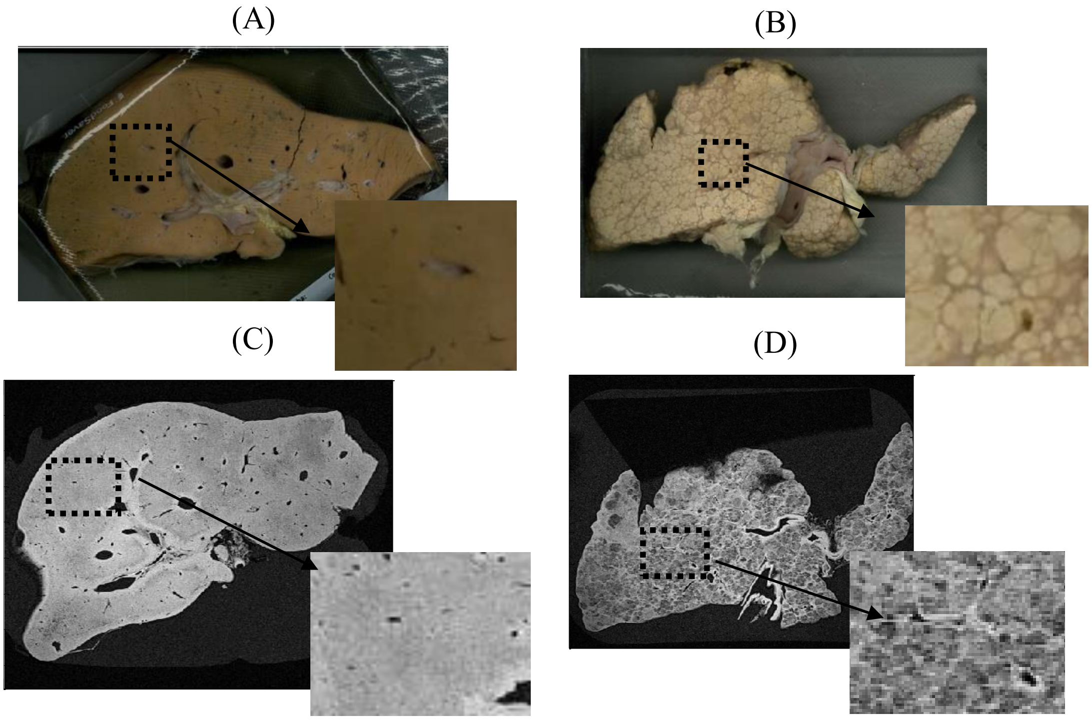


Figure 2.1.1: (A) & (B)Photos of phantom recovered from autopsy. (C) & (D) MRI of phantoms. The images highlight the preservation of contrast of fibrosis associated with HF.

When the formalin-fixed tissue is imaged in the MRI, we find contrast between fibrosis and tissue, similar to contrast observed in *in-vivo* DE-MRI. Figure 2.1.1 shows two of the first tissues collected from autopsy. Figure 2.1.1 (A) is a photo of a liver confirmed with biopsy as stage *F1* on the METAVIR scale. Figure 2.1.1 (B) is a photo of a liver confirmed with biopsy as METAVIR stage *F4*, and is thus cirrhotic. Figure 2.1.1 (C) and (D) are slices from 3D gradient echo sequences at *TR/TE/FA* 9.79ms/4.44ms/15° with fat suppression SPAIR. The resolution in these images are 0.35x0.35x0.35mm^3 which is possible due to the lack of motion and possibility of for long scan times.

Visually these results are promising. A radiologist confirmed the MRI images of formalin-fixed tissue, shown in Figure 2.3.1, do replicate the contrast associated with HF in CE-MRI. A quantitative comparison is necessary to confirm these findings on the way to training an observer sensitive to a change in local texture associated with HF.

2.2 Validation of Texture in Formalin-Fixed Tissue Phantoms

A valid phantom consistently replicates results of *in vivo* imaging. With initial phantoms recovered, we designed a test based on work done by *Burgess et al.* to confirm that the texture in *ex vivo* phantoms is consistent with and substantially similar to images from *in* vivo scans. [55,56]

We recover formalin-fixed, one-inch-thick slices of liver from autopsies performed at the University of Arizona Department of Pathology. Figure 2.3.1 shows a comparison of typical images from liver phantoms imaged with a 3D gradient echo sequence (*TR/TE*/FA =9.79ms/4.44ms/43°) to *in vivo* images of patient (*TR/TE/FA*= 3.51ms/2.20ms/10°).

Burgess showed that radially-averaged 2D power spectra of breast tissue images are repeatable and have a characteristic slope in a log-log plot of the power versus spatial frequency. [56] The method utilizes the discrete 2D Fourier Transform (DFT) of a ROI, calculation of the power via the square modulus, and an averaging procedure along radial directions. This technique is also implemented by Metheany *et al*, and similar techniques are used in ultrasound by Wagner *et al.*, who use the power spectra to describe local texture in tissue. [57,58] This technique has also been extended to ultrasound images of liver tissue.[59]

The textures of the images of the formalin fixed liver phantom to *in-vivo* Gd-enhanced liver images are compared using power spectral analyses of the pixelated images. A 41x41 pixel ROI, x, in each liver is selected to ensure only image data fully inside the liver is analyzed. This size is selected to sample short and long range power spectra frequencies. As suggested by Burgess, a 2D Hanning taper window with a width of 41 pixels is applied to the image. The Hanning window is given as:

$$q_{n,m} = \frac{1}{4}\left(1 - cos\left(2\pi\frac{n}{N}\right)\right)\left(1 - cos\left(2\pi\frac{m}{M}\right)\right), 0 \leq n \leq N \;\; 0 \leq m \leq M.^{56} \tag{2.2.1}$$

The power spectra of the images were computed as the square moduli of the discrete Fourier transform (DFT),

$$\boldsymbol{\chi}_{k,l} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} x_{n.m} q_{n,m} \mathrm{exp}\left(-\frac{2\pi ink}{N}\right) \mathrm{exp}\left(-\frac{2\pi imk}{M}\right).$$

(2.2.2)

The Hanning window reduces the effect of high ringing frequency terms introduced by truncation at the ROI boundaries in the DFT.

Radial averaging is performed with a grid of 400 lines originating at the DC component of a 41x41 $\boldsymbol{\chi}$ and covering the angular range from 0 to 2π, by calculating the nearest Cartesian point neighbor to each of a set of regularly spaced points along the radial line.

2.3 Results of the Phantom Texture Validation Test

Figure 2.3.2, which shows four *ex vivo* tissue phantom samples and the respective ROI's for which the power spectra is calculated. Figure 2.3.3 is the collection of *in vivo* images and ROI's that were analyzed. The radially averaged power spectra are plotted on a log, log scale in shown in Figure 2.3.4, of the formalin-fixed *ex vivo* phantoms to the Gd-enhanced *in vivo* patient power spectra follow very similar trends. Comparing the mean of the fitted slopes and their standard deviation we conclude that the image data of the liver tissue phantoms have similar textural content and contrast as the Gd-enhanced *in vivo* images.

With this conclusion, we can perform local textural analysis of liver fibrosis using the formalin-fixed phantoms as a surrogate task, and later translate the methods to images acquired on patients *in vivo*.

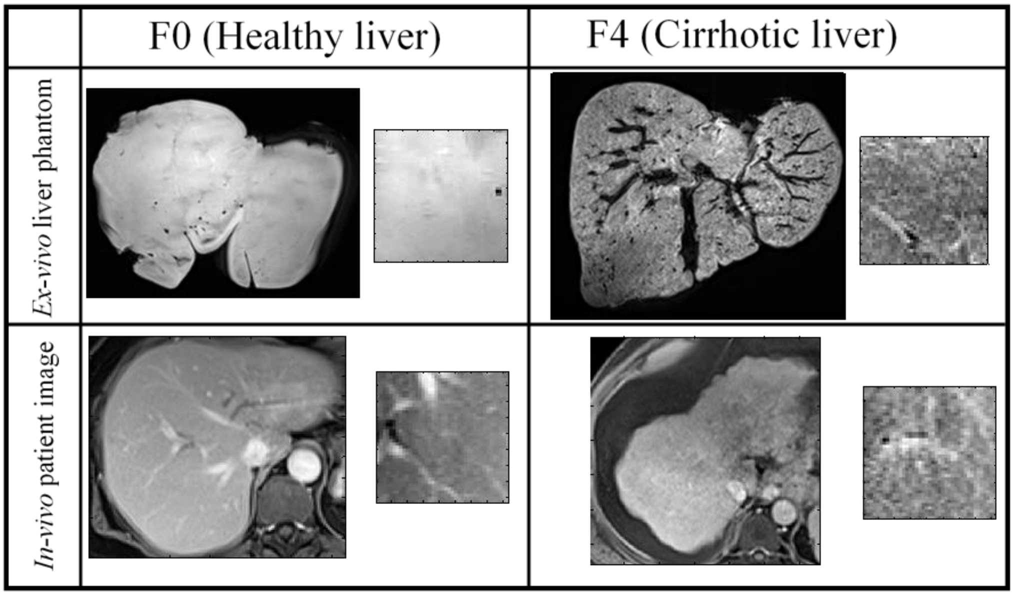


Figure 2.3.1: A comparison highlighting the comparable contrast between ex-vivo phantoms and in-vivo patient images. The formalin fixed F4 phantom has an extensive collagen network that has contrast between the tissue and collagen similar to the in-vivo image shown

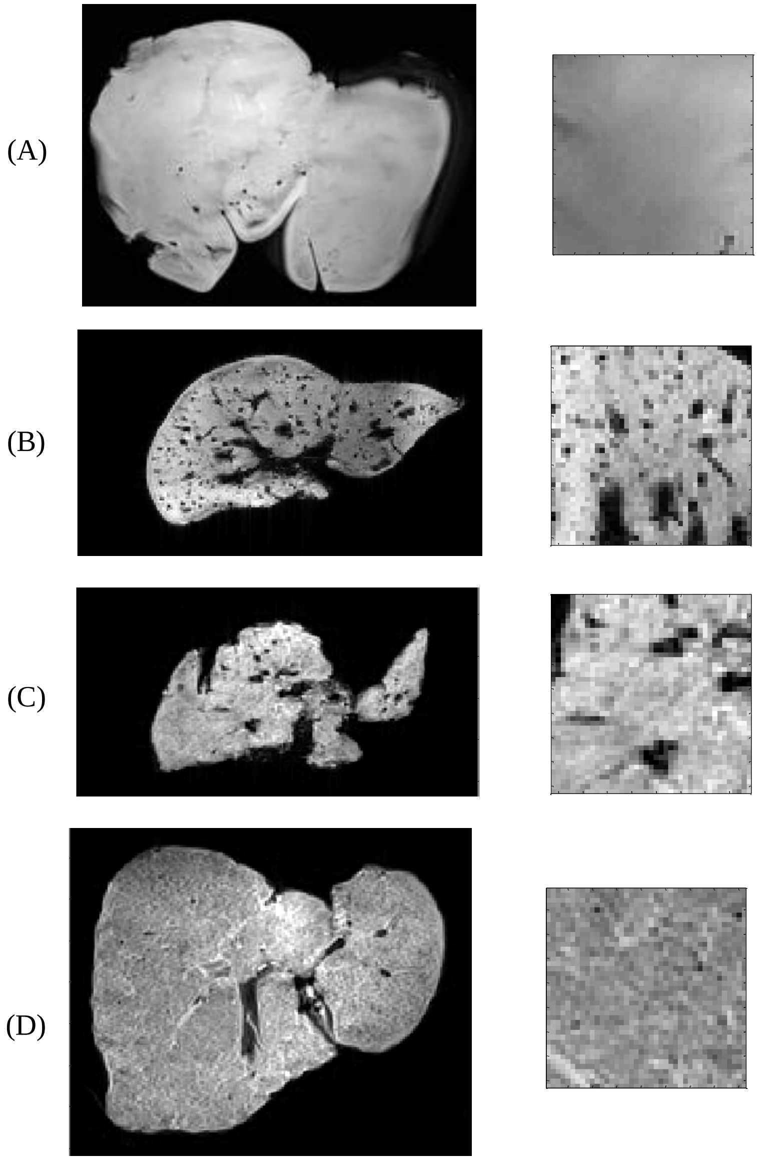


Figure 2.3.2: A & B are in plane images of an ex vivo F0 and F1 livers respectively. C & D are in plane images of an ex vivo F4 liver. Next to each image is the respective ROI for which radial power spectra were calculated.

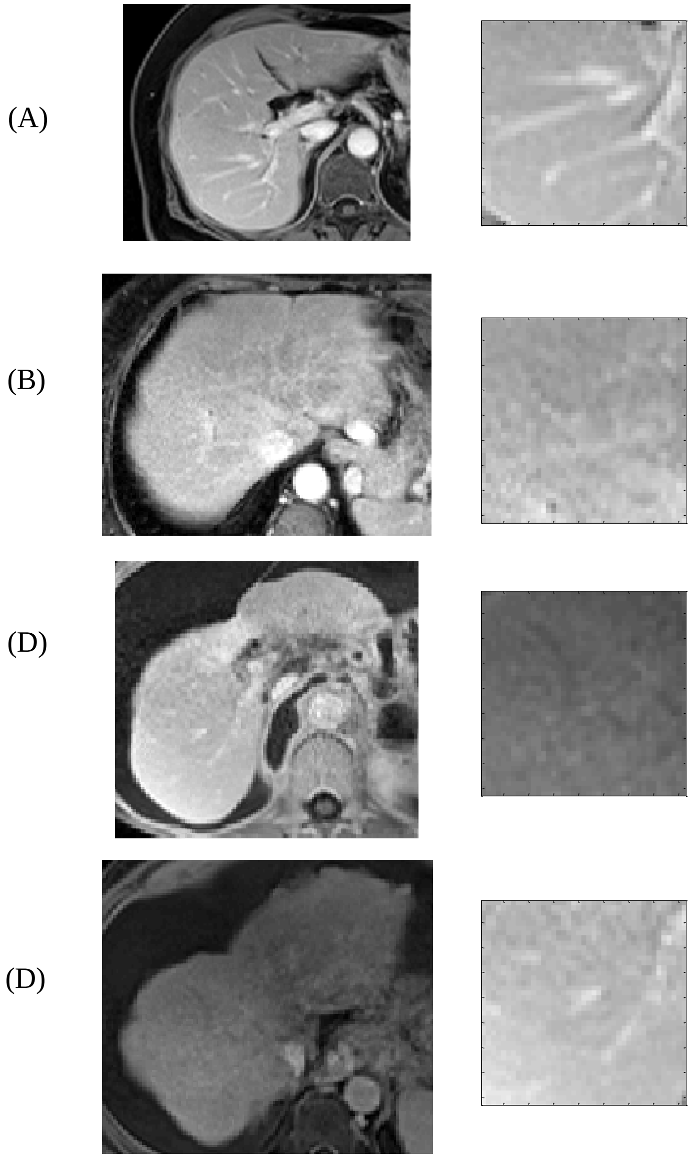


Figure 2.3.3:A & B are in plane images of an in vivo F0 patient liver. C & D are in plane images of an in vivo F4 patient liver. Next to each image is the respective ROI for which radial power spectra were calculated.

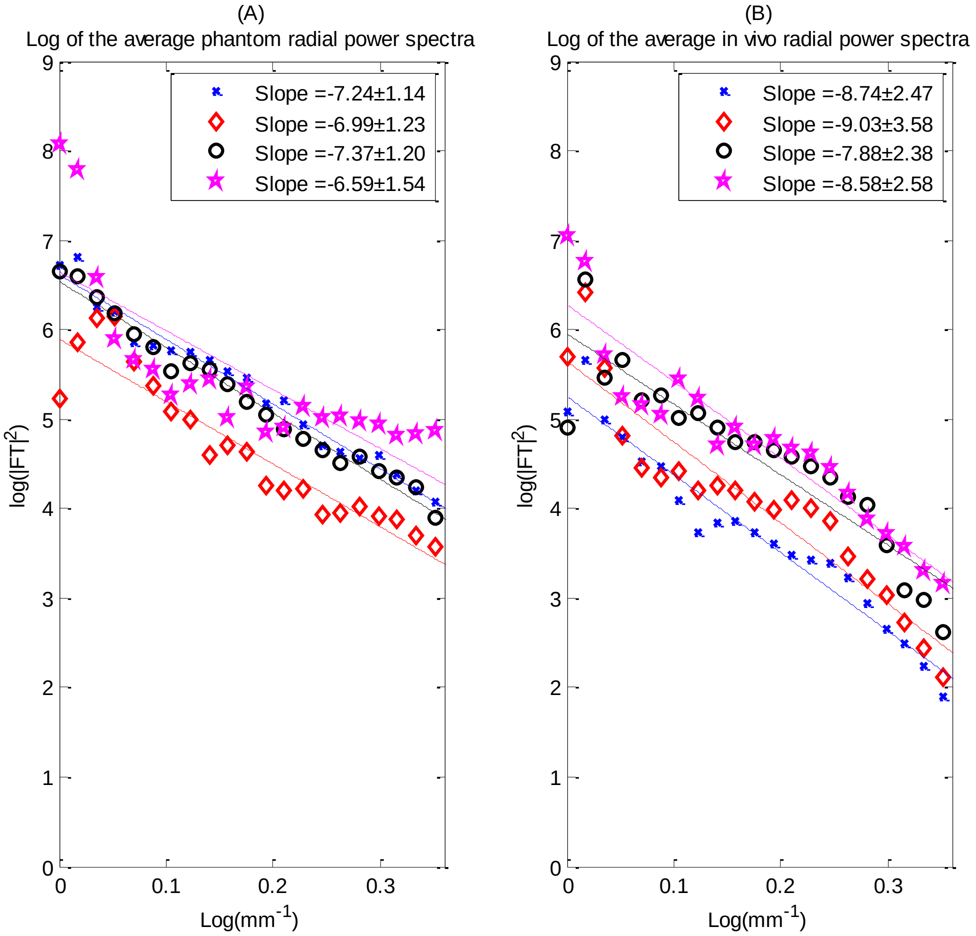


Figure 2.3.4: (A) Log of the average power spectra ±2σ of in vivo and (B) ex vivo healthy and diseased liver tissue, plotted against a log frequency scale with ± 2σ.

3. Training and Testing a Linear Observer

With a valid phantom confirmed from experiments covered in Chapter 2, a Hotelling template based on local texture analysis described in sections 1.12.1 and 1.12.2 is trained and tested on phantom images. We present the requirements to take images, prepare the data for analysis, train the observer with first and second order statistics and present experimental results.

3.1 MRI of Phantoms

An MRI sequence similar to *in vivo* gadolinium enhanced MRI (DE-MRI) is used in order to replicate the *in vivo* HF features and in order to ease translation to the clinic. The phantoms were imaged on a SIEMENS 3T Skyra with a *TR/TEI/*FA = 9.79ms/4.44ms/15°. With these parameters, images are collected at an isotropic resolution of $0.35mm^3$. The grid necessary for this scan is 768x760 pixels in-plane and 96 slices in the z-direction. All images were collected at room temperature (22º C). The total scan time at one FA is approximately 25 minutes. All phantoms were imaged with the same scan parameters to ensure that the contrast between the samples is consistent. A model observer cannot easily be trained with data taken at different resolutions, so the experiment was set up to ensure there is no change in resolution between scans. The phantoms range in thickness from 1 to 2.5 cm. Imaging with a small slice thickness prevents errors from partial volume entering the image. Figure 3.1.1 provides representative slices from each phantom. The sequence is almost capable of reproducing texture associates with fibrosis that is seen in biopsy results.

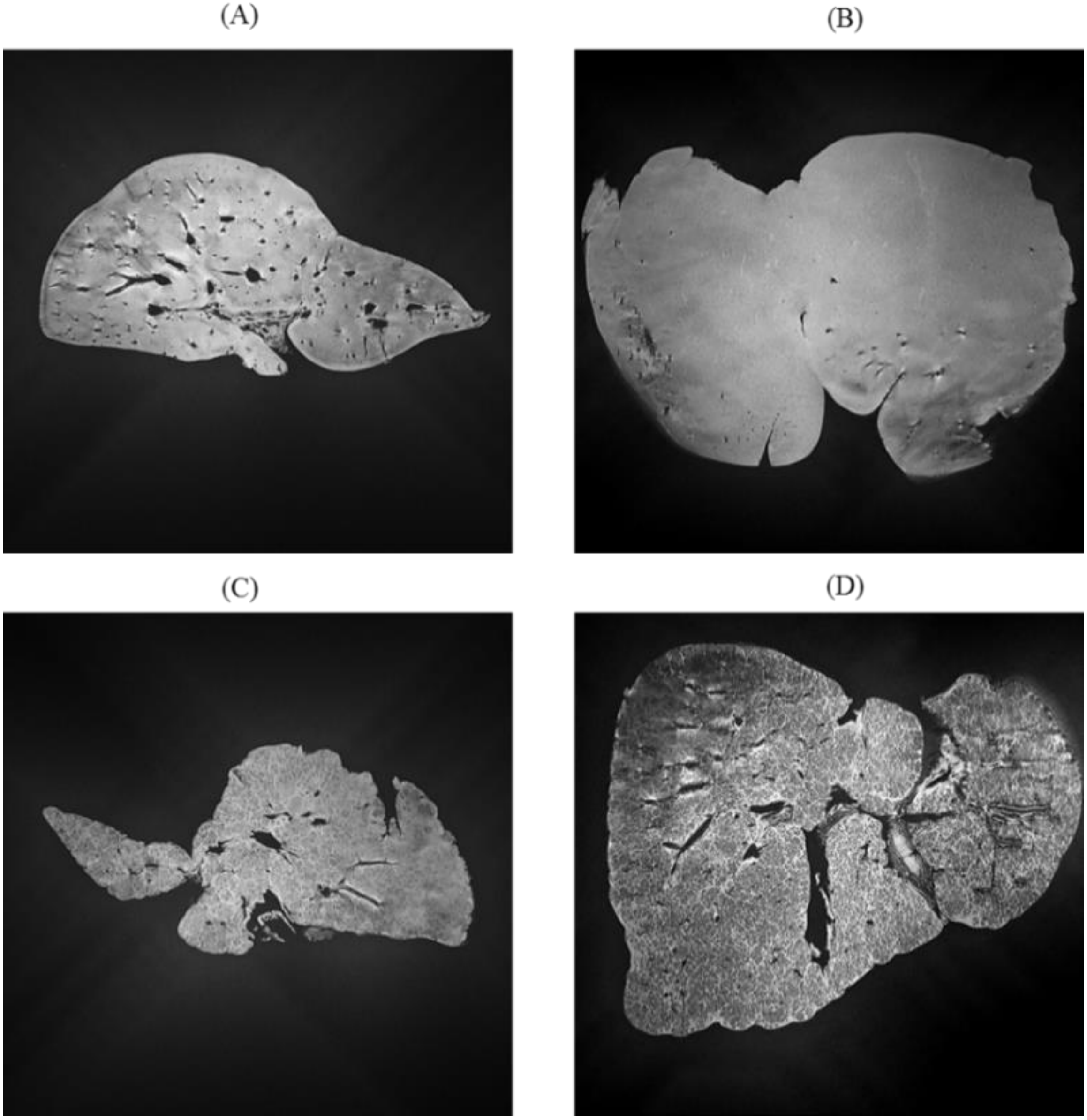


Figure 3.1.1: A and B are the F0 and F1 phantoms collected for the null hypothesis, C and D are the F4 phantoms. These are selections of the image data sets used to train and test the Hotelling observers.

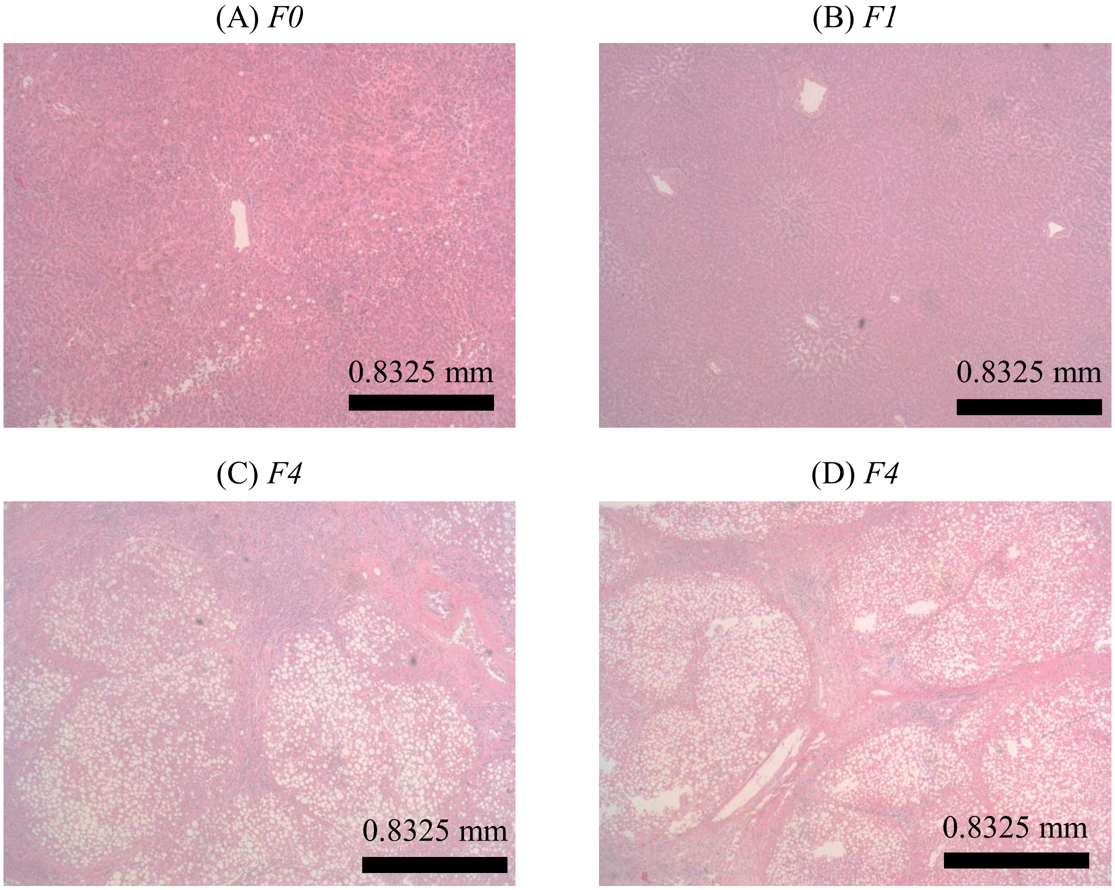
3.2 Biopsy Results

Dr. Bhattacharyya, MD, the head of the Pathology Department at the University of Arizona College of Medicine collaborated with us in this research and assisted with the project by staging the biopsies on the METAVIR scale. To train the normal class of the observer, a *F0* and *F1* liver were recovered from autopsy. For the diseased case, two livers recovered from autopsy had all biopsies diagnosed as *F4.* Only livers with homogenous results were selected for training and testing observers. Eight biopsies were collected from each phantom.

Optical images from biopsy samples from each of the four phantoms are shown in Figure 3.2.11. Images are collected on a video microscope. Figure 3.2.11 A is a slide taken from an *F0* liver and has no fibrosis in the tissue as observed by our radiologist. Figure 3.2.11 B is from an *F1* liver and has mild fibrosis around the portal veins. Figure 3.2.11 C and D are from cirrhotic livers, and were identified as *F4*. The ECM is seen creating large boundaries around the lobules with the hepatocytes inside.

The current method for securing phantoms first requires consent of the attending physician(s) of an autopsy that the liver tissue can be released to us. During the autopsy, a 1-inch-thick slice of the liver is cut from the organ and fixed in 10% neutral buffer formalin until the tissue is fixed, requiring about 3-4 days. After the tissue is fixed, eight biopsies are collected as triage specimens and taken to produce histological slides. The turn-around for slides is at least two weeks. After slides are picked up, we can schedule a read with Dr.

Bhattacharyya, whose expertise was critical to this project. The time between recovering a phantom and the read to determine the class can take about 1 month. This has been sufficient for this project, but may require additional resources in continuing work.



(A) *F0*

(B) *F1*

(C) *F4*

(D) *F4*

Figure 3.2.1: (A) Biopsy from 3.2.1 (A), (B) Biopsy from 3.2.1 (B), (C) Biopsy from 3.2.1 (C), (D) Biopsy from 3.2.1 (D), liver phantoms diagnosed by the pathologist

3.3 Thresholding to Remove Effects of Veins

Images of both *ex vivo* and *in vivo* tissue contain vasculature features of the liver. In both of these cases, the vessels will add texture to the image, however, it is not texture associated with fibrosis. Large vessels in *F4* tissue completely contained in the ROI image will register as images with no texture, looking like *F0* tissue to an observer. Smaller vessels in *F0* tissue could be confused for strands of collagen associated with *F4*.

In *in vivo* imaging, Gd accumulates in the blood and, if a scan is taken early, there is high signal in the blood vessels. When working with *ex vivo* phantoms, vessels are dark, either containing air or formalin, depending on how the phantom is prepared. In either case, removing these features is a crucial step to ensure that the observer is trained on statistics from liver tissue and not vasculature.

The scope of this dissertation does not cover automatic segmentation. While there are many effective methods to complete the task of automatic segmentation, a simple approach was used to segment major blood vessels and background from phantom images. The solution selected is basic thresholding with visual confirmation by a trained human. Results for healthy phantoms are shown in Figure 3.3.1 and diseased livers in Figure 3.3.2. Some pixels not part of the phantom do survive the thresholding, but will not survive a second stage threshold which is applied during texture analysis.

Every ROI is confirmed to contain data from the liver by rejecting data containing values equal to the background of each respective slice. The images show the removal of low signal background from the MRI scan. Features due to fibrosis are not detectable in these images due to the high brightness setting of the images.

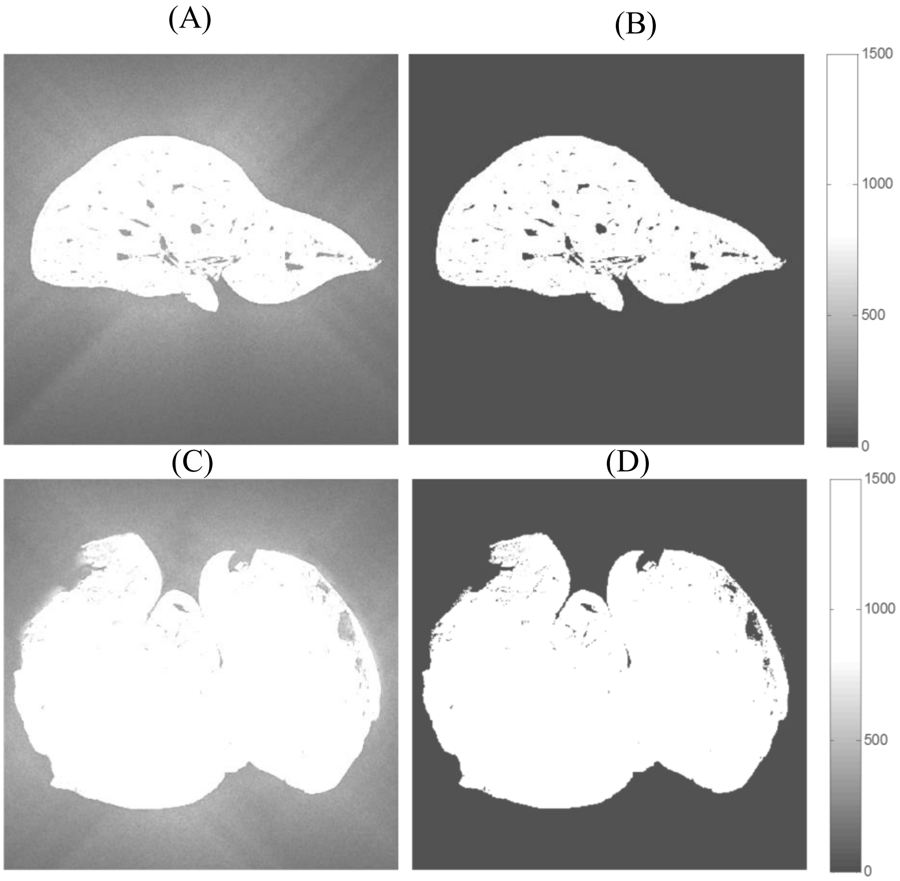


Figure 3.3.1: (A) and (C) are *F1* and *F0* phantoms before thresholding, (C) and (D) are the results of applying the threshold.

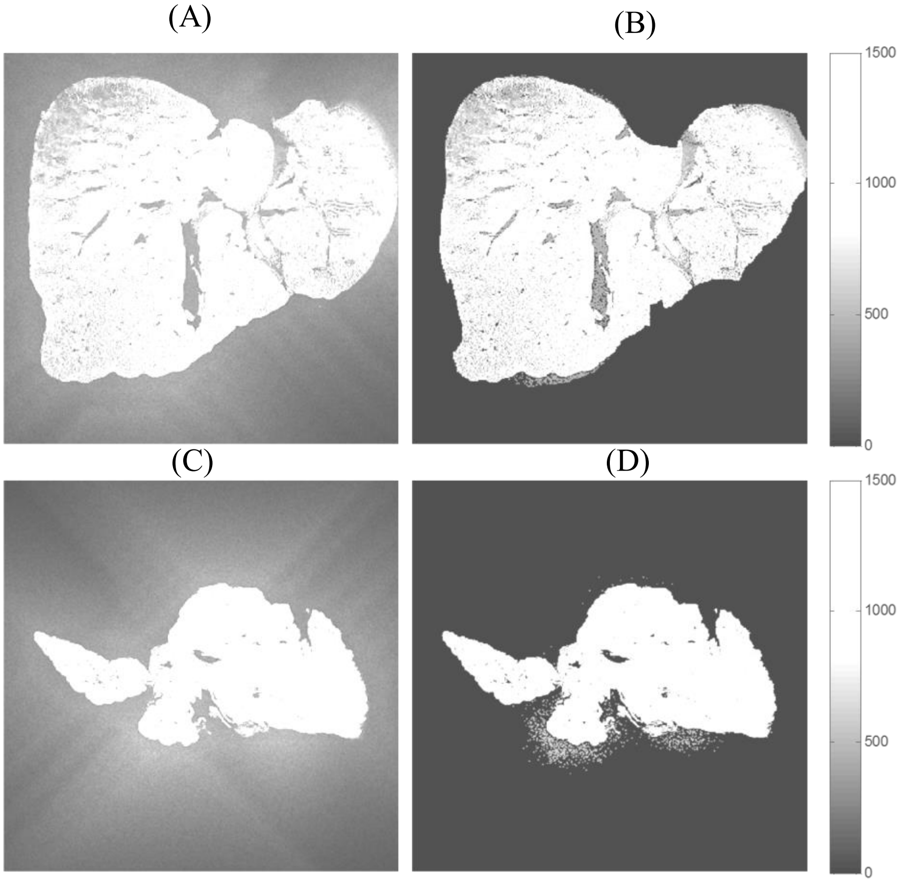


Figure 3.3.2: (A) and (C) are *F4* phantoms before thresholding, (C) and (D) are the results of applying the threshold.

3.4 Normalization

The Hotelling observer described in section 1.10 is dependent on the matched filter,

$$\vec{s} = \overline{\overrightarrow{g_1}} - \overline{\overrightarrow{g_0}},$$

(3.4.1)

which suggests different DC offsets or gains in data acquisition would greatly impact the performance of the observer. To remove bias of the DC offsets and gains, a normalization method was applied before an observer is trained or tested.

MRI suffers from artifacts from partial volume, and the effects of limited resolution. There are also coil sensitivities that will impact the recovered signal on a slice-by-slice basis. In fact, coil placement can heavily impact the final image recovered and there is an inevitable patient to patient variability in the amount of Gd contrast agent injected and accumulated in the liver. The observation of slice-by-slice variation suggests that the normalization needs to be applied on a slice-by-slice basis.

The calculation of the observer depends on measuring the mean and covariance of the image data associated with the object. To remove the systematic effects, strategy of slice-by-slice, zero-mean, unit-variance was chosen to act as our normalization method. A threshold was applied to the DICOM image vector $\vec{G}$ of size m x n x z as described in section 3.3 and it has the mean of each slice removed from it. The mean is calculated from pixels that are not the background,

$$\overline{G_z} = \frac{\sum_{m=1}^{M}\sum_{n}^{N} G_{mnz},\ G_{mnz} \neq min(G_{mnz})}{M'}, M' = \sum G_{mnz} \neq min(G_{mnz}).$$

(3.4.2)

The same data used to calculate the mean is used to calculate the variance. This information is used to create the normalized data set,

$$\overrightarrow{G'}_{z,Norm} = \frac{\vec{G}_z - \overline{\vec{G}_z}}{\sigma_{\vec{G}_z}},$$

(3.4.3)

where σ_{G_z} is the standard deviation of the liver pixel values**.** To see the impact of this normalization, we compare the histograms of the data collected in a phantom before and after normalization. Results are shown in

Figure 3.5.1: A single slice from a *F4* phantom and it's pixel values before normalization

and Figure 3.5.2. The histograms omit the data not normalized, i.e. the background. We use this method to normalize all data, whether it is used to train or test an observer.

3.5 Training an Observer

Training a linear Hotelling observer requires careful consideration to ensure the template is only trained on features in local texture content. The observer is linear in, $\vec{g}$, putting requirements on how the DICOM data are handled.

The size of the observer ROI is important to consider. A large observer samples more tissue and therefore might include more HF, but this is balanced with the training requirements on the observer. For that reason, we chose to limit our local texture analysis to 7x7 and 9x9 pixel regions.

Once the phantoms are divided by class, the normalized DICOMs are used to train the observer. What follows is a detailed account of training the null or healthy class of the observer.

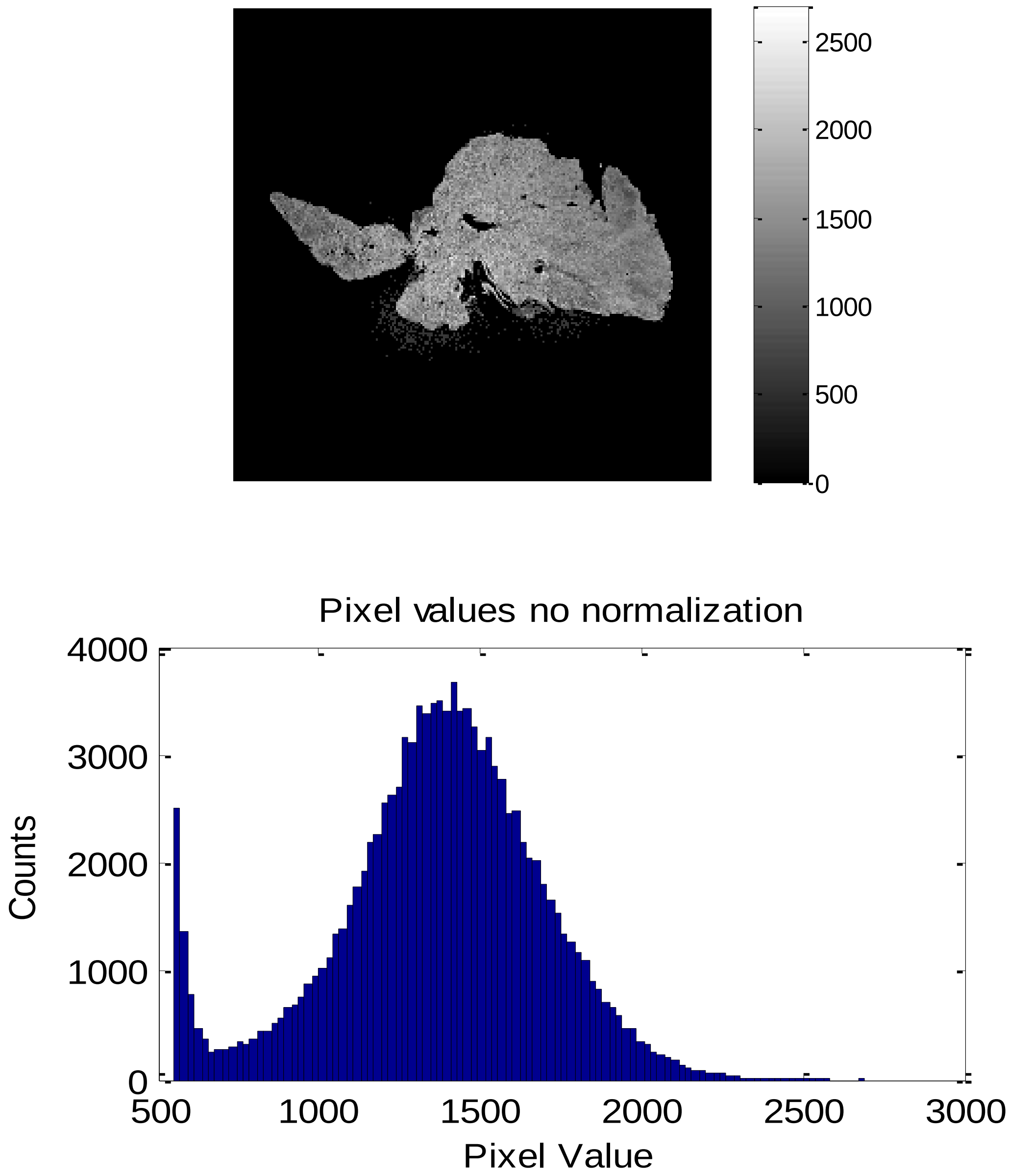


Figure 3.5.1: A single slice from a *F4* phantom and it’s pixel values before normalization

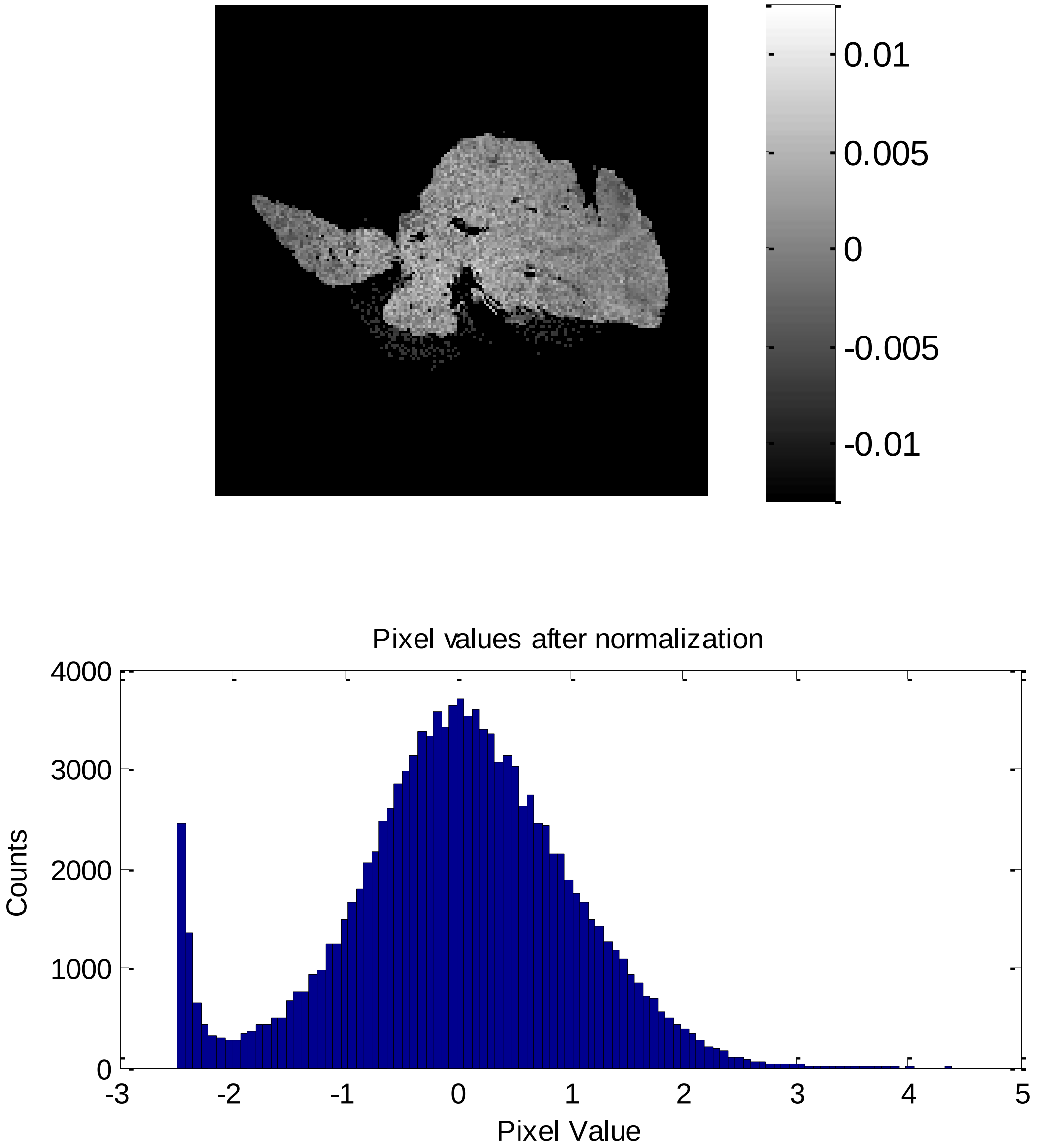


Figure 3.5.2 A single slice from a *F4* phantom and it's pixel values after normalization

For the *F0* phantom, there is a 3D image data set of size PxQxZ. Across each slice z, an image PxQ is normalized to zero mean, unit variance. Any regions including background data are eliminated via thresholding. This ensures only liver data trains the observer.

Training images are sampled with independently gridded ROI's to avoid bias from overlapping. A value, N, is chosen for the size of the template for which B training data sets are selected from the training data:

$$\begin{gathered}\vec{g}_{0_1} = \overrightarrow{G'}_{0_{p:p+(\mathrm{N}-1),q:q+(\mathrm{N}-1)}} \\ \vec{g}_{0_2} = \overrightarrow{G'}_{0_{p+\mathrm{N}:p+\mathrm{N}+(\mathrm{N}-1),q+\mathrm{N}:q+p+\mathrm{N}+(\mathrm{N}-1)}} \\ \vec{g}_{0_\mathrm{b}} = \overrightarrow{G'}_{0_{[p+N(b-1)]:[p+N(b-1)+(N-1)]:[q+N(b-1)]:[q+N(b-1)+(N-1)]}}. \end{gathered} \tag{3.5.1}$$

To test either analysis method, each gridded region is analyzed with either the 2DAC or 2DCC as introduced in 1.12 .

To train the observer it is important to ensure that the covariance matrices meet minimum sampling requirements. The requirement for a 7x7 ROI for a 2DAC is 3,528, and the 2DCC requires 288 independent samples per covariance matrix. The phantom in Figure 3.1.1A has 248,439 independent grid ROIs, Figure 3.1.1B has 56,865, Figure 3.1.1C has 127,150, and Figure 3.1.1D has 49,858 regions. Thus, each phantom contains more than enough data to train the covariance matrix with either local texture analysis for a 7x7 region. The requirement for a 9x9 ROI for a 2DAC is 10,368 and the 2DCC requires 800 independent grid ROI's. There is still adequate training data with a 9x9 ROI.

Once the training set for class 0 is collected, the training set of class 1 is collected repeating equations 2.9-2.11. With the training sets, we can calculate the components of the linear Hotelling observers for each analysis method,

$$\hat{R}_{k,l} = \frac{\sum_{y=1}^{Y} R_{k,l_y}}{Y} \qquad \hat{S}_{k,l} = \frac{\sum_{y=1}^{Y} S_{k,l_y}}{Y}$$

(3.5.2)

The matrix is flattened from a two dimensional matrix to a one dimensional vector:

$$\vec{R}_\varepsilon = \hat{R}_{k,l} \quad \begin{matrix} 1 \le k \le K \\ 1 \le l \le L \end{matrix} \qquad \vec{S}_\varepsilon = \hat{S}_{k,l} \quad \begin{matrix} 1 \le k \le K, \\ 1 \le l \le L' \end{matrix}$$

(3.5.3)

Where K and L are selected to only collect data before the center point of the autocorrelation analysis and ε is the component of the texture analysis ranging from one to the final unique element. The training data are also used to calculate the covariance matrix for each class

$$K_{0_{i,j}} = \frac{1}{Y-1}\sum_{i=1}^{Y} \left|\vec{R}_i - \vec{\bar{R}}\right|^t \left|\vec{R}_j - \vec{\bar{R}}\right|.$$

(3.5.4)

With the measurement of the mean vector and covariance matrix for each class and with either the 2D discrete autocorrelation or the 2D discrete circular correlation analysis, a linear Hotelling observer is calculated.

With these results a Hotelling observer is calculated based on either local 2D autocorrelation, 2D circular autocorrelation, or wavelet coefficients,

$$w_{2DAC} = \left(\frac{\boldsymbol{K}_{0_R} + \boldsymbol{K}_{1_R}}{2}\right)^{-1} \left(\vec{\bar{R}}_1 - \vec{\bar{R}}_0\right), \tag{3.5.5}$$

$$w_{2DCC} = \left(\frac{\boldsymbol{K}_{0_S} + \boldsymbol{K}_{1_S}}{2}\right)^{-1} \left(\vec{\bar{S}}_1 - \vec{\bar{S}}_0\right), \tag{3.5.6}$$

$$w_{WT} = \left(\frac{\boldsymbol{K}_{0_\omega} + \boldsymbol{K}_{1_\omega}}{2}\right)^{-1} \left(\vec{\bar{\omega}}_1 - \vec{\bar{\omega}}_0\right). \tag{3.5.7}$$

3.6 Hotelling Observer Results

To confirm that a trained Hotelling observer is effective when applied to testing data, only image data from one liver is allowed to represent the training class for training an observer. With four phantoms, it is possible to train four independent Hotelling observers, each phantom corresponds to phantoms found in Figure 3.1.1.

Both the 2DAC and 2DCC are normalized and symmetric, leading to the possibility to leave out the 0,0 point in the data and rejecting the second half of the 2DCC and 2DAC result. For a 7x7 ROI, the 2DAC is 13x13 and 2DCC is 7x7. This results in the need for only the first 84 points for the 2DAC template and the first 24 points of the 2DCC.

Each local texture analysis is performed on selected ROIs. To train the observer, the mean of the local texture analysis data are calculated. Figure 3.6.1 and Figure 3.6.2 show the mean of the 2DAC and 2DCC data from each phantom, respectively. The data that is used to calculate mean also is used to compute covariance matrices, shown in Figure 3.6.3 and Figure 3.6.4 for the 2DAC and 2DCC, respectively. A human observer can see a difference between the mean data computed from the healthy and diseased livers.

1D representation of templates calculated from training data are shown in Figure 3.6.5 and Figure 3.6.6 for the 2DAC and 2DCC respectively. For each combination of livers used to train a template, the template is applied to the other two phantoms as testing data. The templates from both texture analysis methods exhibit rotational symmetry as expected from texture analysis methods. For all training data, we find the templates to be highly reproducible.

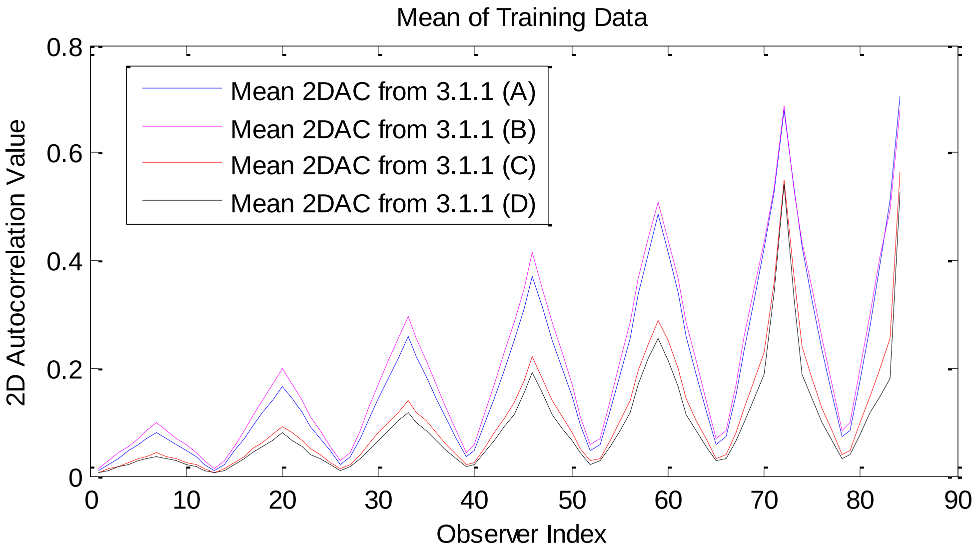


Figure 3.6.1: Mean data from 2DAC texture analysis from each phantom. 3.1.1 (A) and (B) are the *F0* and *F1* livers and 3.1.1 (C) and (D) are diseased livers

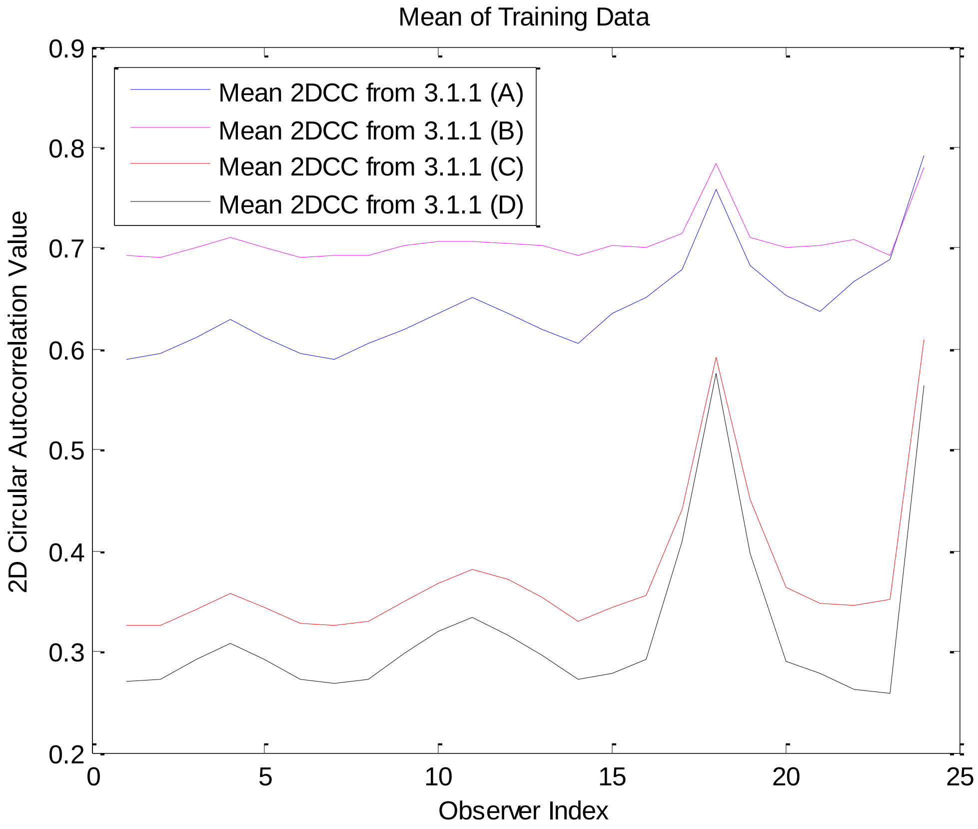


Figure 3.6.2: Mean data from 2DCC texture analysis from each phantom. 3.1.1 (A) and (B) are the *F0* and *F1* livers and 3.1.1 (C) and (D) are diseased livers

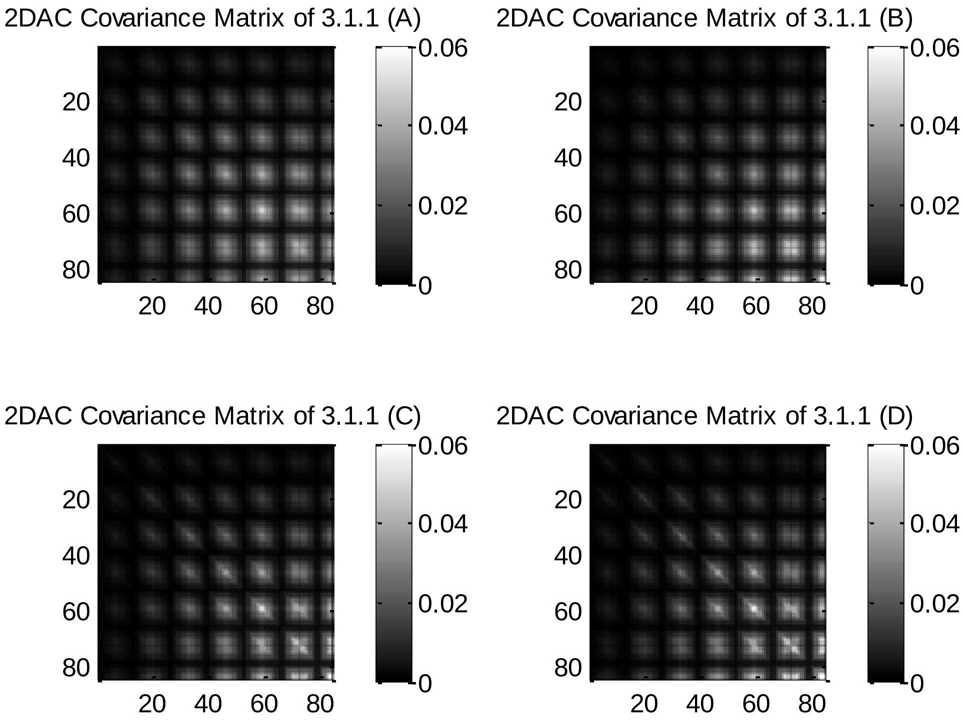
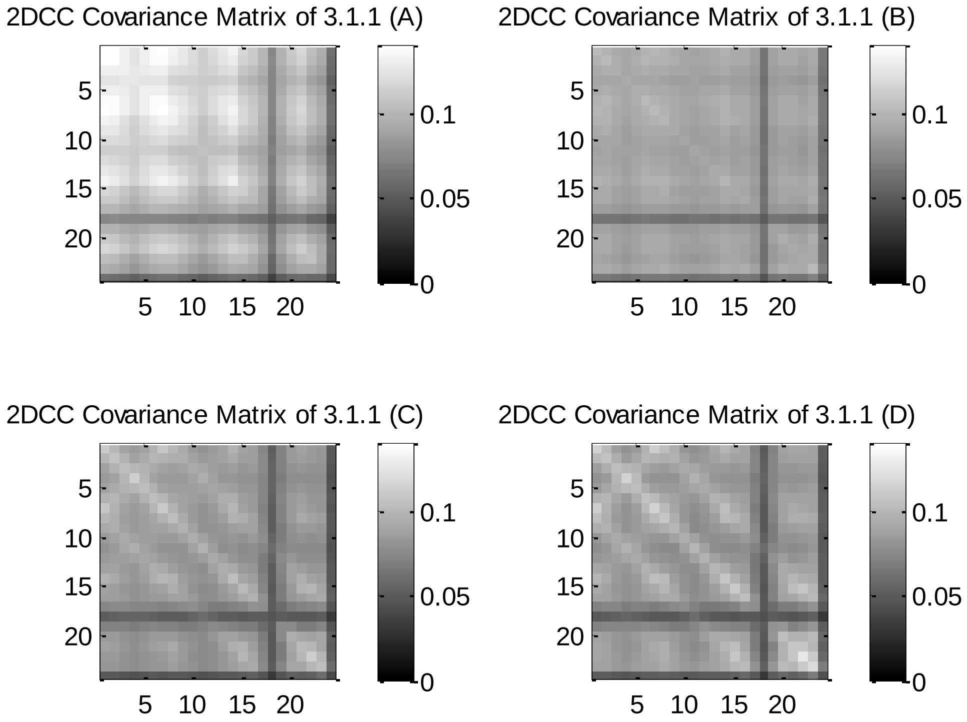


Figure 3.6.3: Covariance data from 2DAC texture analysis from each phantom



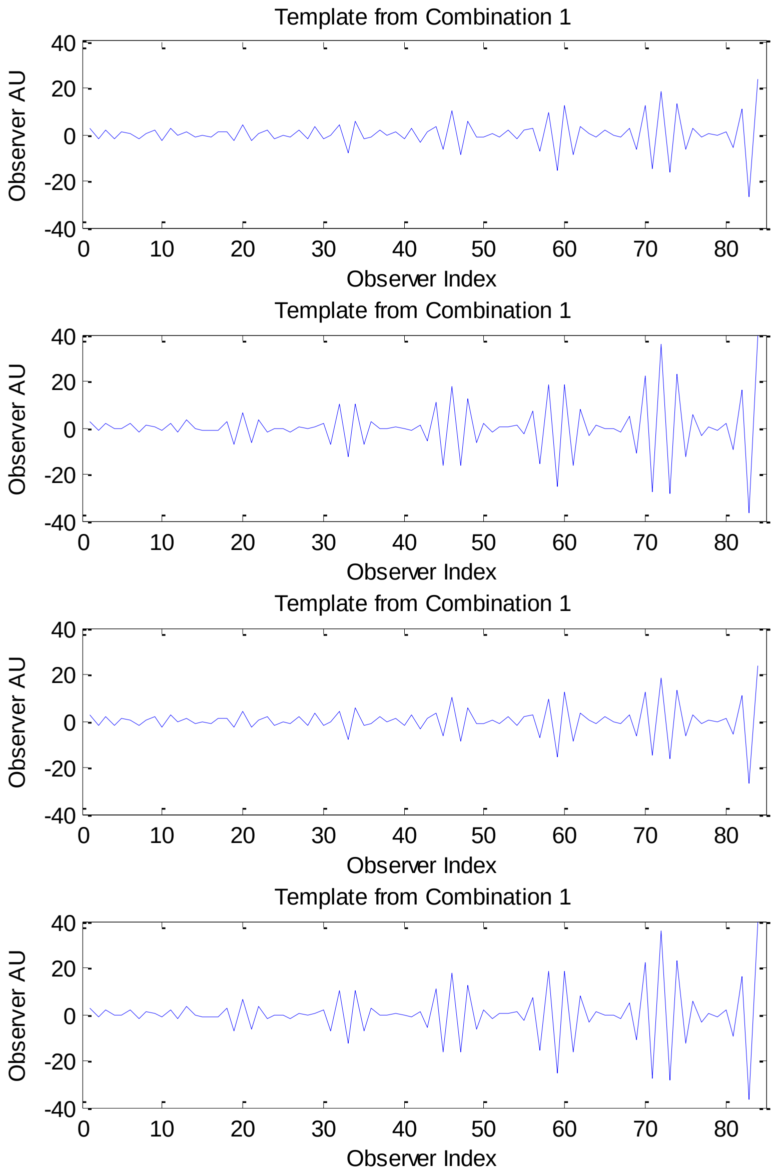
2DCC Covariance Matrix of 3.1.1 (A)

2DCC Covariance Matrix of 3.1.1 (B)

2DCC Covariance Matrix of 3.1.1 (C)

2DCC Covariance Matrix of 3.1.1 (D)

Figure 3.6.4: Covariance data from 2DCC texture analysis from each phantom



Template from Combination 1

Template from Combination 1

Template from Combination 1

Template from Combination 1

Figure 3.6.5: 1D representation of the Hotelling template of each combination of training data resulting from a 2DAC texture analysis

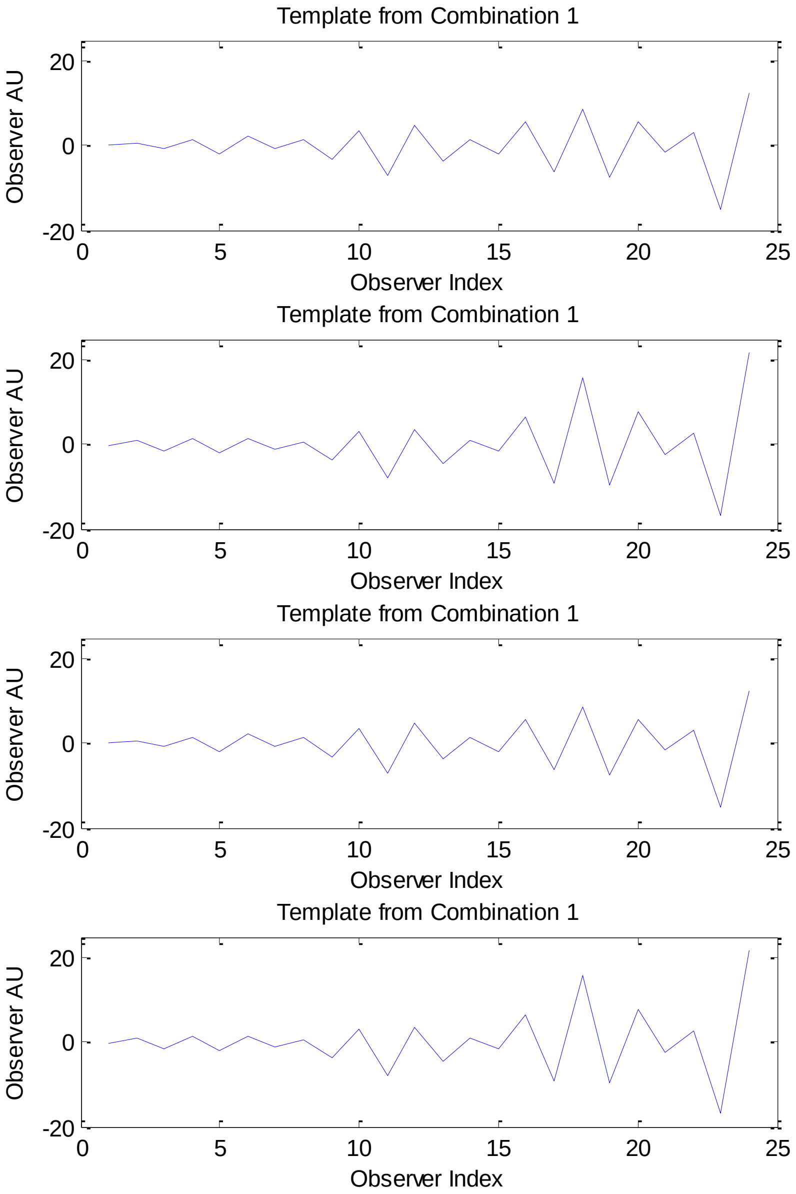


Figure 3.6.6: 1D representation of the Hotelling template of each combination of training data resulting from a 2DCC texture analysis

The third texture analysis method applied to our data is a local wavelet transform. A 7x7 ROI will have edge effects from the wavelet transform, so an 8x8 ROI was analyzed for the Hotelling Observer. The Haar wavelet is sensitive to boundaries, and the visual interpretation of the means and templates are more difficult to understand. Also to only be sensitive to changes in local texture, the template was trained on the magnitude of the wavelet transform coefficients. The 8x8 ROI results in a 64-pixel observer.

The means are shown in Figure 3.6.7. One notices immediately that the data from the normal livers, from Figure 3.1.1 (A) and 3.1.1 (B) have a gap between them. The fibrotic livers from Figure 3.1.1 (C) and Figure 3.1.1 (D) appear very similar. The covariance matrices from each data set are shown in Figure 3.6.8. Finally, the four independent templates are shown in Figure 3.6.9. The templates appear to have different characteristics, suggesting limited reproducibility.

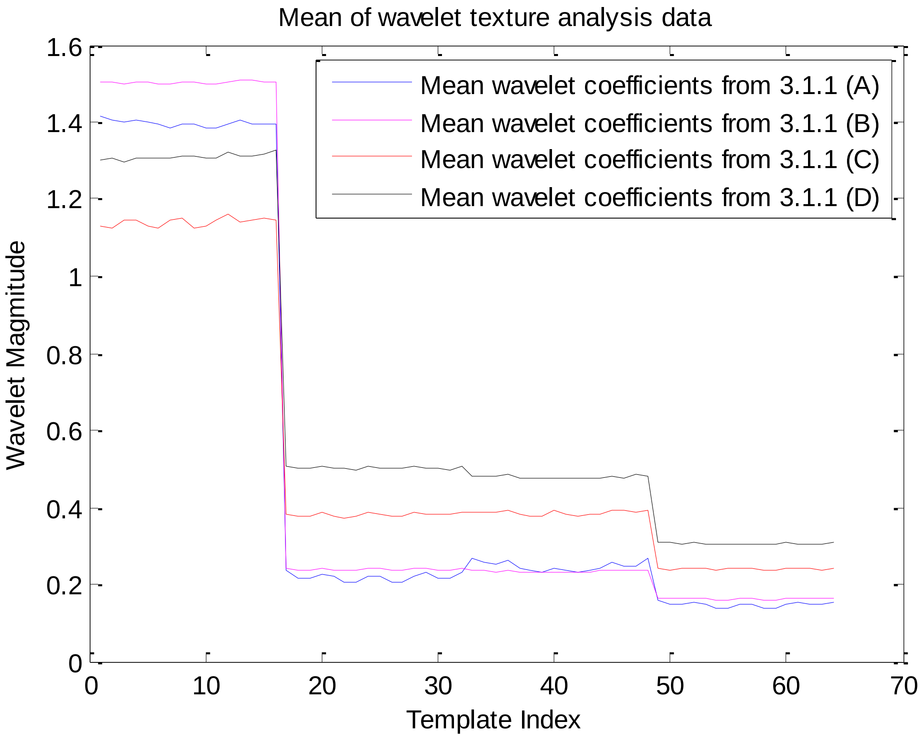
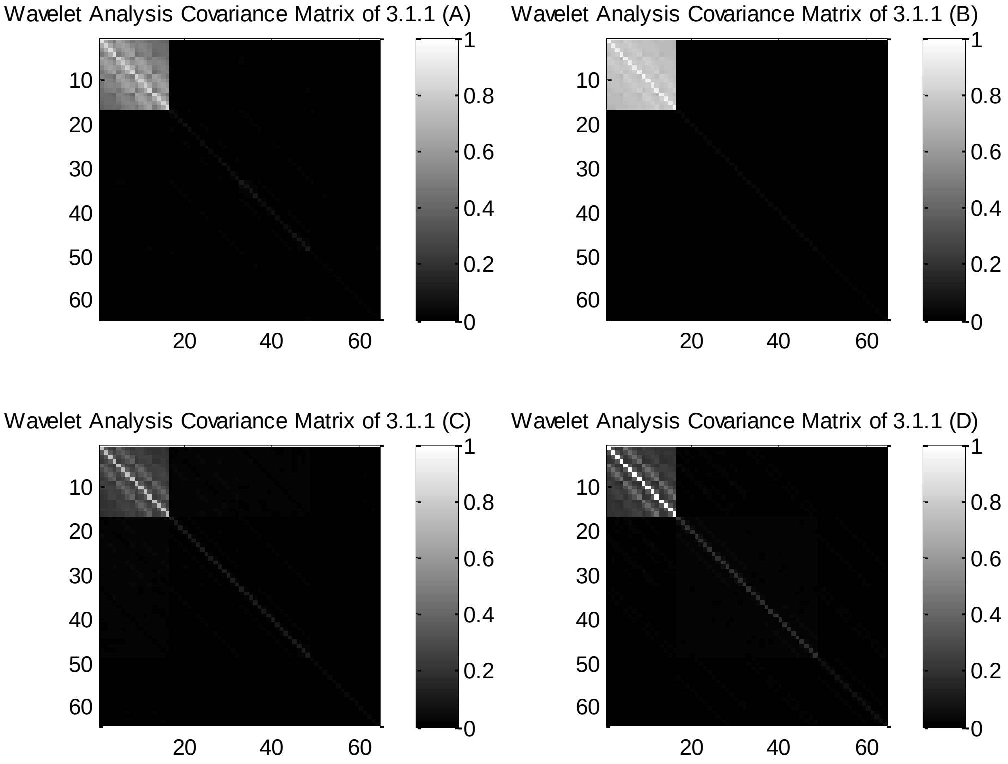


Figure 3.6.7: Mean wavelet analysis data from the four phantoms



Wavelet Analysis Covariance Matrix of 3.1.1 (A)

Wavelet Analysis Covariance Matrix of 3.1.1 (B)

Wavelet Analysis Covariance Matrix of 3.1.1 (C)

Wavelet Analysis Covariance Matrix of 3.1.1 (D)

Figure 3.6.8: Covariance matrices for the four phantoms based on a local wavelet transform

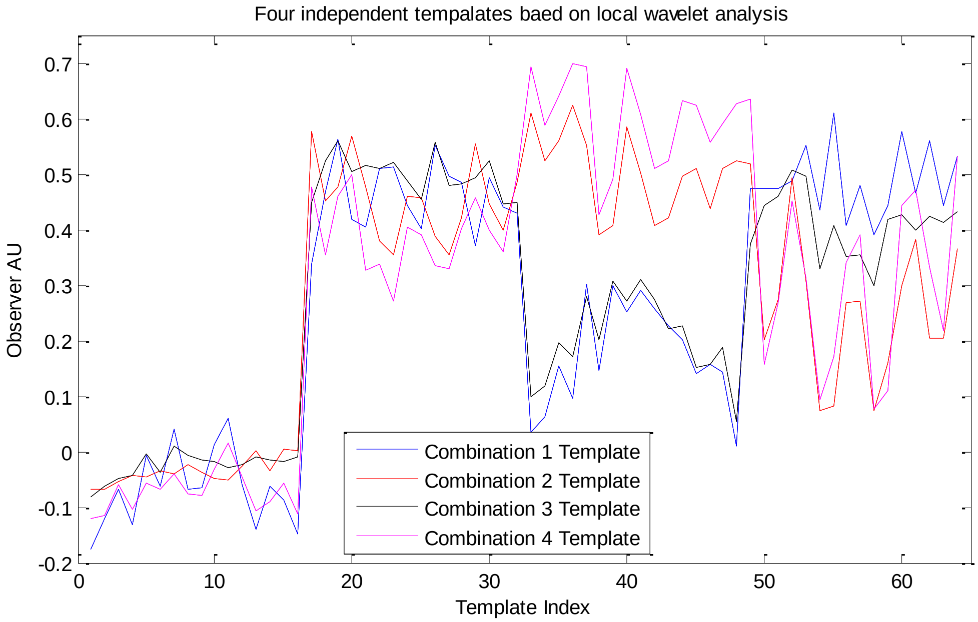


Figure 3.6.9: Four independent templates based on local wavelet analysis for the collected phantoms

3.7 Testing an Observer

With a trained linear Hotelling observer $\vec{w}$, the steps to test the ability to separate two classes are very similar to the training process. Recovering a test statistic requires taking the inner product of the template with image data collected and processed in the same method $\vec{w}$ is trained:

$$t(\vec{g}) = \vec{w}'\vec{g}. \tag{3.7.1}$$

To evaluate a local texture analysis such as the 2D discrete autocorrelation or 2D discrete circular autocorrelation, the vector is replaced with the respective analysis method,

$$t(\vec{R}) = \vec{w}'_{2DAC}\vec{R}, \qquad t(\vec{S}) = \vec{w}'_{2DCC}\vec{S}. \tag{3.7.2}$$

To test the Hotelling observer, an independent testing set drawn from the individual classes is necessary. F0 and F4 livers that were not part of the training set test were used as the testing data sets.

Following similar steps to training the observer for the $F0$ phantom, there is a 3D image data set G' of size PxQxZ. For each slice z, there is an image G of size PxQ on which the local texture analysis is performed. An ROI the same size as the training ROI is stepped across the testing data. A sliding window has advantages, first, more test statistics are calculated to estimate the ROC curve. Second, the sliding window shows local changes in the test statistic across the liver which will allow us to build a computer-aided diagnostic tool, which will be described in section 3.9. Finally, while training the observer requires the data to be statistically independent, this constraint is not necessary for testing a mathematical observer. The test statistics of consecutive ROI's are independent of one another. The testing data set is therefore much larger than the training data set. Applying the sliding window is as follows:

$$\vec{g}_{0_1} = \overrightarrow{G'}_{0_{p:p+(\mathrm{N}-1),q:q+(\mathrm{N}-1)}}$$

$$\vec{g}_{0_2} = \overrightarrow{G'}_{0_{p+1:p+1+(\mathrm{N}-1),q+1:q+1+(\mathrm{N}-1)}}$$

$$\vdots$$

$$\vec{g}_{0_\mathrm{b}} = \overrightarrow{G'}_{0_{p+\mathrm{b}-1:p+\mathrm{b}+\mathrm{N}-2,q+\mathrm{b}-1:q+\mathrm{b}+\mathrm{N}-2}}. \tag{3.7.3}$$

For each local region, the image matrix is flattened and both cross-correlation analysis methods are applied and the test statistic is extracted:

$$t(\vec{R}_{0_1}) = \vec{w}'\vec{R}_{0_1}$$

$$t(\vec{R}_{0_2}) = \vec{w}'\vec{R}_{0_2}$$

$$\vdots$$

$$t(\vec{R}_{0_\mathrm{b}}) = \vec{w}'\vec{R}_{0_\mathrm{b}}.$$

(3.7.4)

After evaluating each region across the testing data set for class 0, the steps are repeated for a class 1 testing data set.

Equation 3.7.3 provides steps for an NxN pixel area, however, only a 7x7 and 9x9 pixel regions were tested with cross-correlation techniques.

The two test statistic distributions, corresponding to HF present and absent, are plotted in a histogram on the same linear scale. ROC analysis is performed by moving a threshold and determining what fraction of the test statistic is less than the threshold. If it is, than that local region is determined to be healthy or part of the null hypothesis[50,60]. Sliding the threshold across the full range of the test statistics allows the measurement of false positive fraction (FPF) and true positive fraction (TPF) and the receiver operator characteristic (ROC) curve to be plotted.[50]

3.8 Results of Testing the Observer

As mentioned above, to apply the observer, an inner product is computed between the template and the testing data taken from the liver phantoms not used to train the observer. Data from phantoms used for testing data are selected through a sliding window ROI. As a result, the phantom from Figure 3.1.1A has 7,686,034 separate 7x7 regions a template based on either local texture analysis can analyze. The phantom from Figure 3.1.1B has 2,952,176 regions; the phantom from Figure 3.1.1C has 6,053,726 regions and the phantom from Figure 3.1.1D has 2,503,899 regions.

The test statistics from each ROI are placed in a histogram according to the training data set they originate from. Each histogram is normalized to an area of one to recover a probability density function (PDF). The PDF's for the 2DAC and 2DCC are shown in Figure 3.8.1 and Figure 3.8.2, respectively. It is clear that the healthy and diseased groups have significant differences in the locations, shapes and widths of their PDFs.

Using the test-statistic density functions, ROC analysis was used to calculate the true positive and false positive fractions, resulting in an ROC curve. Figure 3.8.3 shows the AUROC and ROC's for the four experiments utilizing a local 2DAC. Figure 3.8.4 shows the results for a 2DCC. The AUROC's for both experiments are consistent with each other and the 2DCC performs as well as the 2DAC even though the template is only a third of the size and requires less training data.

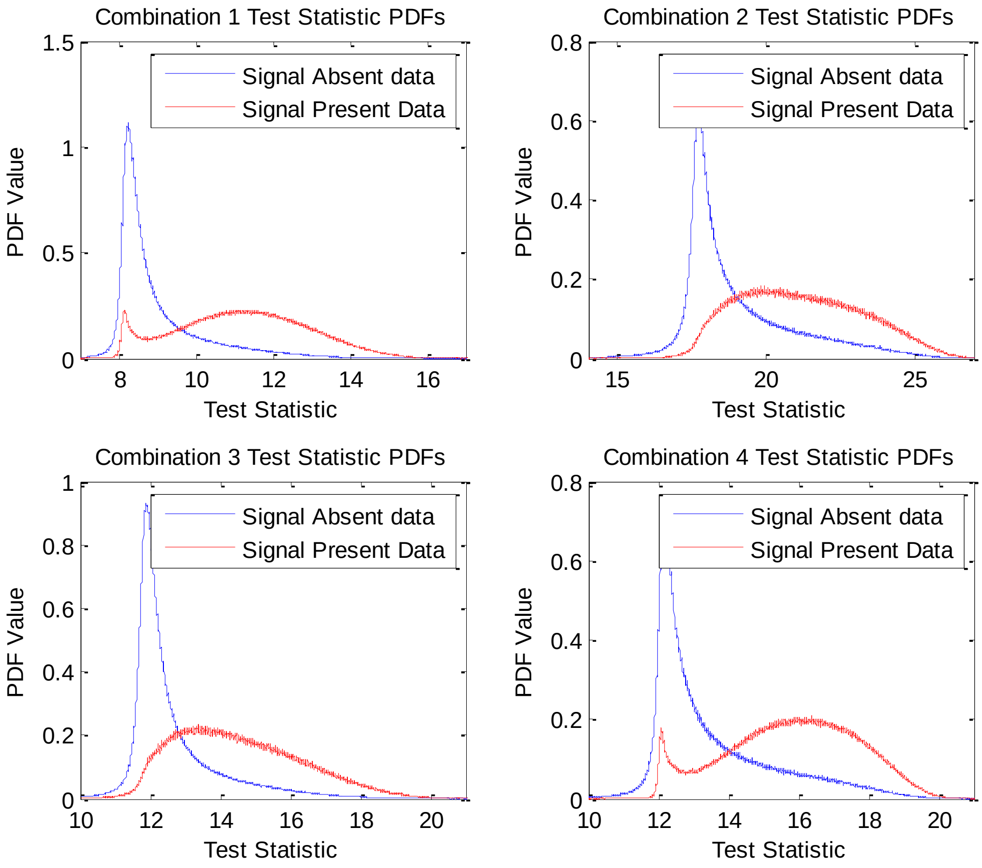


Figure 3.8.1: Histograms of test statistic PDFs from testing data resulting from each Hotelling template applied to testing data for 2DAC texture analysis

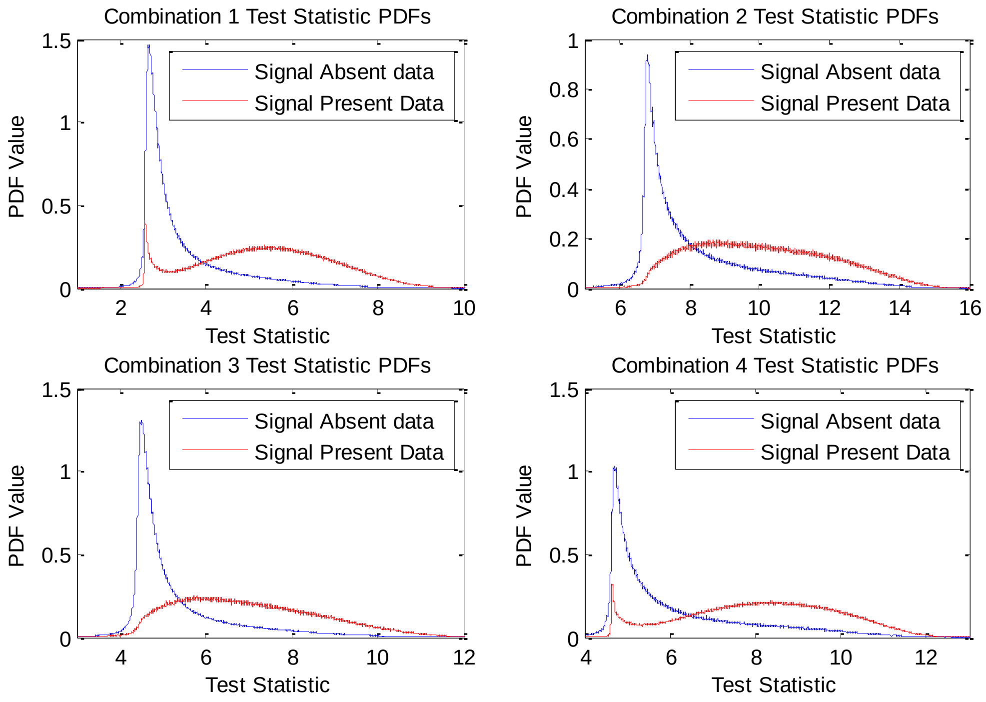


Figure 3.8.2: Histograms of test statistic PDF's from testing data resulting from each Hotelling template applied to testing data for 2DCC texture analysis

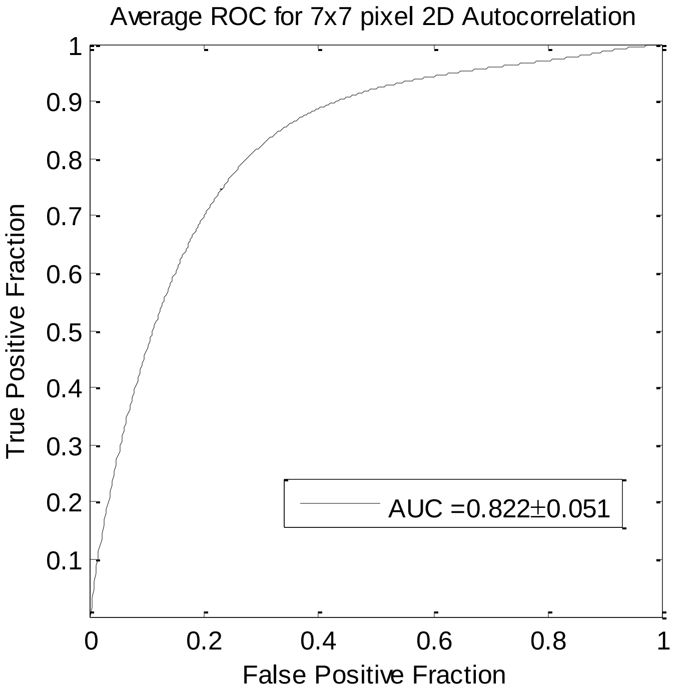


Figure 3.8.3 Mean ROC and mean AUC±2σ using a 2DAC template

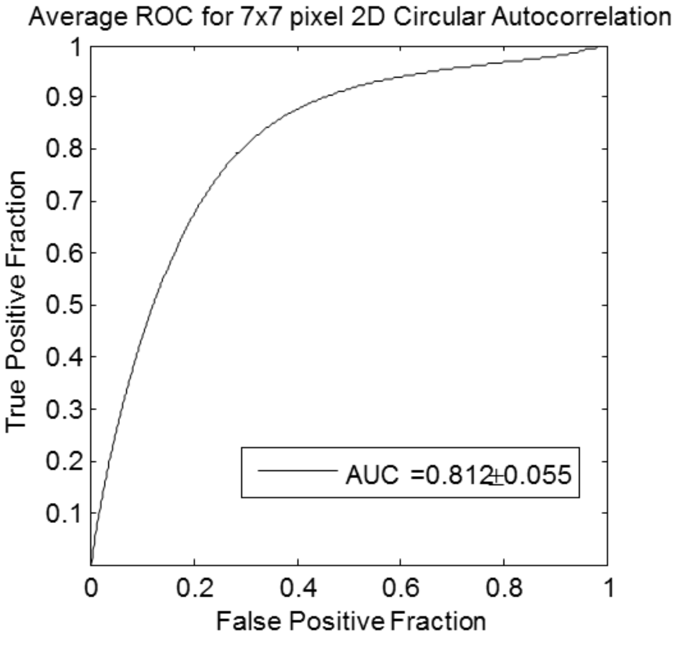


Figure 3.8.4: Mean ROC and mean AUC±2σ using a 2DCC template

The AUC values for the observer based on the wavelet transform performed about as well as the observers based on autocorrelation analysis. The histograms for the four combinations of training data are shown in Figure 3.8.5, the average ROC curve is shown in Figure 3.8.6. The AUC value is higher than the other two methods of texture analysis, but the templates are not reproducible, which will require more investigation into the improved performance in future work.

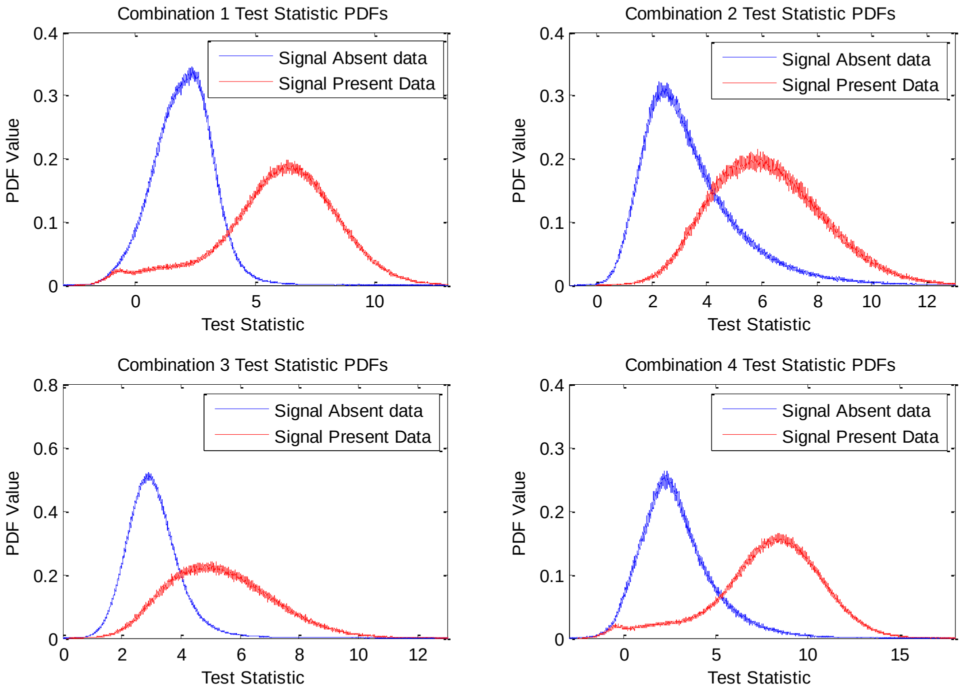


Figure 3.8.5: Histograms of test statistics from testing data resulting from each Hotelling template applied to testing data for wavelet transform analysis

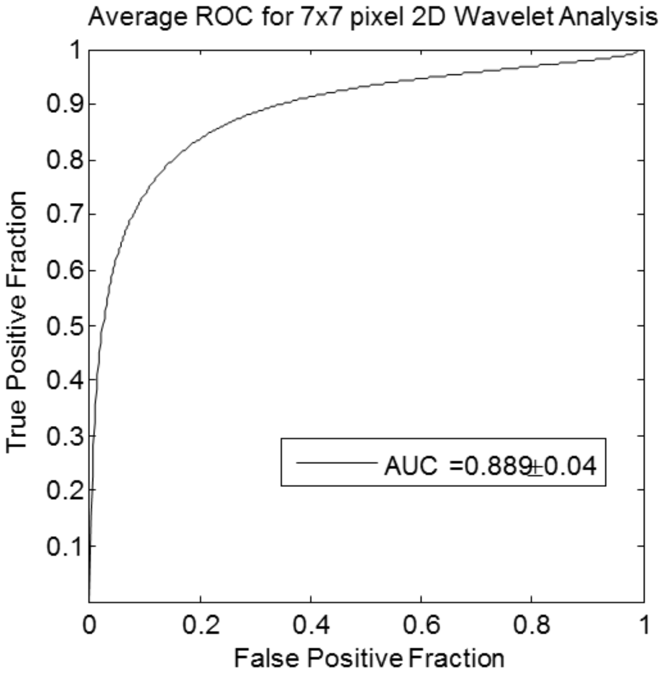


Figure 3.8.6: Mean ROC and mean AUC ±2σ using a wavelet transform template

3.9 Test Statistic Maps

Traditionally, a Hotelling observer is applied to testing data and ROC analysis is performed to determine the sensitivity and specificity of the observer. Our observer works on an ROI much smaller than the area of the image, allowing the creation of maps similar to what an MRE exam provides to a radiologist.

The sliding window for testing an observer retrieves an image which has a local texture analysis performed on it. The observer is applied to the analyzed image:

$$t\left(\vec{R}_{0_{\mathrm{b}}}\right) = \vec{w}'\vec{R}_{0_{\mathrm{b}}}.$$

(3.9.1)

The image $\boldsymbol{g}_{0_{\mathrm{b}}}$ is associated with $\vec{R}_{0_{\mathrm{b}}}$. The center of the ROI of has an index *p, q, z*. The indices are collected into an array:

$$I_b = [p_b\ q_b\ z_b].$$

(3.9.2)

A null image, $\vec{\mathrm{T}}$ that is the same size as $\overrightarrow{G'}$is created, and pixels are filled in with test statistics associated with the index they were collected from:

$$\vec{\mathrm{T}}(p_b\ q_b\ z_b) = t\left(\vec{R}_{0_{\mathrm{b}}}\right).$$

(3.9.3)

The test statistic map $\vec{\mathrm{T}}$ shows the value of all test statistics at the location to where they are computed from. These maps highlight areas where the observer detected a high probability that an ROI is part of the signal present hypothesis. The collection of high test statistic values highlight areas of suspected fibrosis in an image, guiding a human observer to apply extra focus in indicated regions of the liver.

We suggest that textural data can aid a radiologist through what we refer to as test statistic maps. Figure 3.9.1 and Figure 3.9.2 show the two different texture analysis techniques applied to the same liver slices across the same 7x7 pixel regions using the templates from the first combination of training and testing data. Figure 3.9.4 and Figure 3.9.5 show a comparison of the 2DAC and 2DCC techniques using the template from combination 2. These maps, combined with a radiologist's knowledge of the delayed phase MRI image, create a powerful method to locate and identify liver fibrosis. The holes in the maps are primarily a result of local vasculature structure. In these regions the template was neither trained with the data nor used to test a model observer.

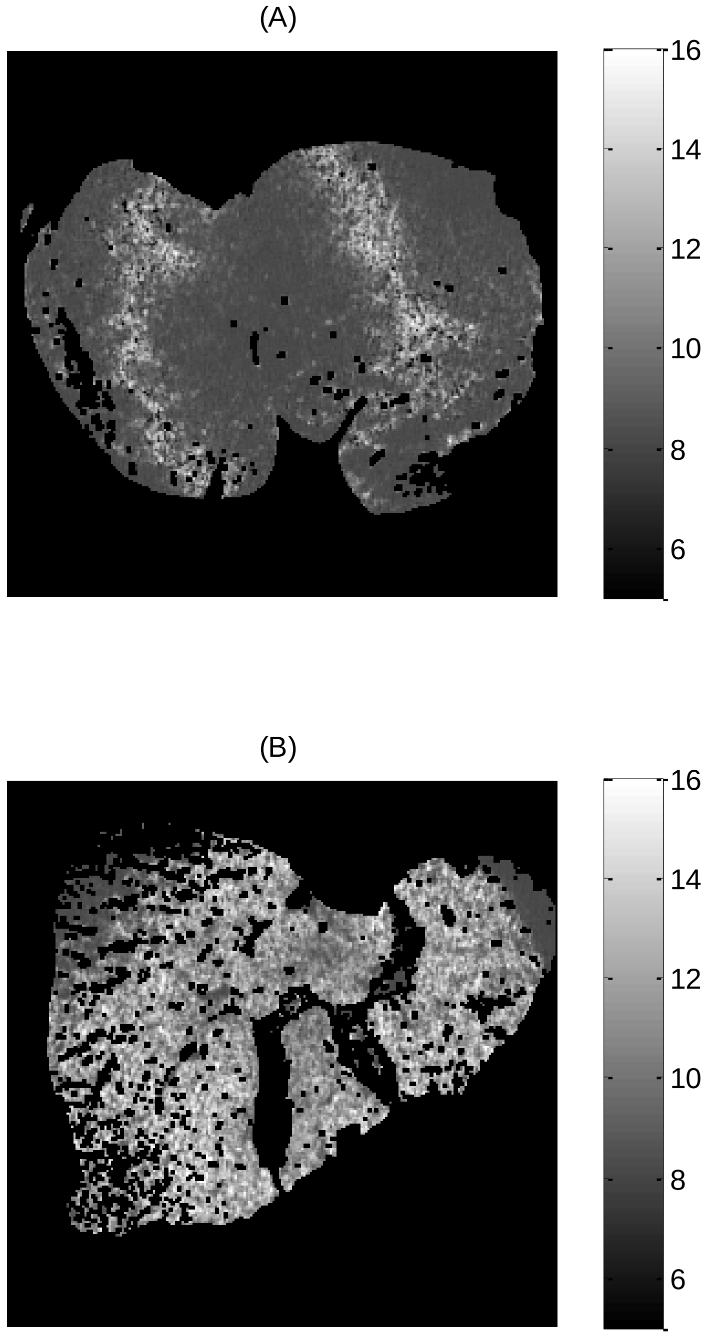


Figure 3.9.1 Test statistic maps generated from the Hotelling observer from Combination 1 utilizing a 2DAC texture analysis for (A) an F0 liver and (B) F4 liver

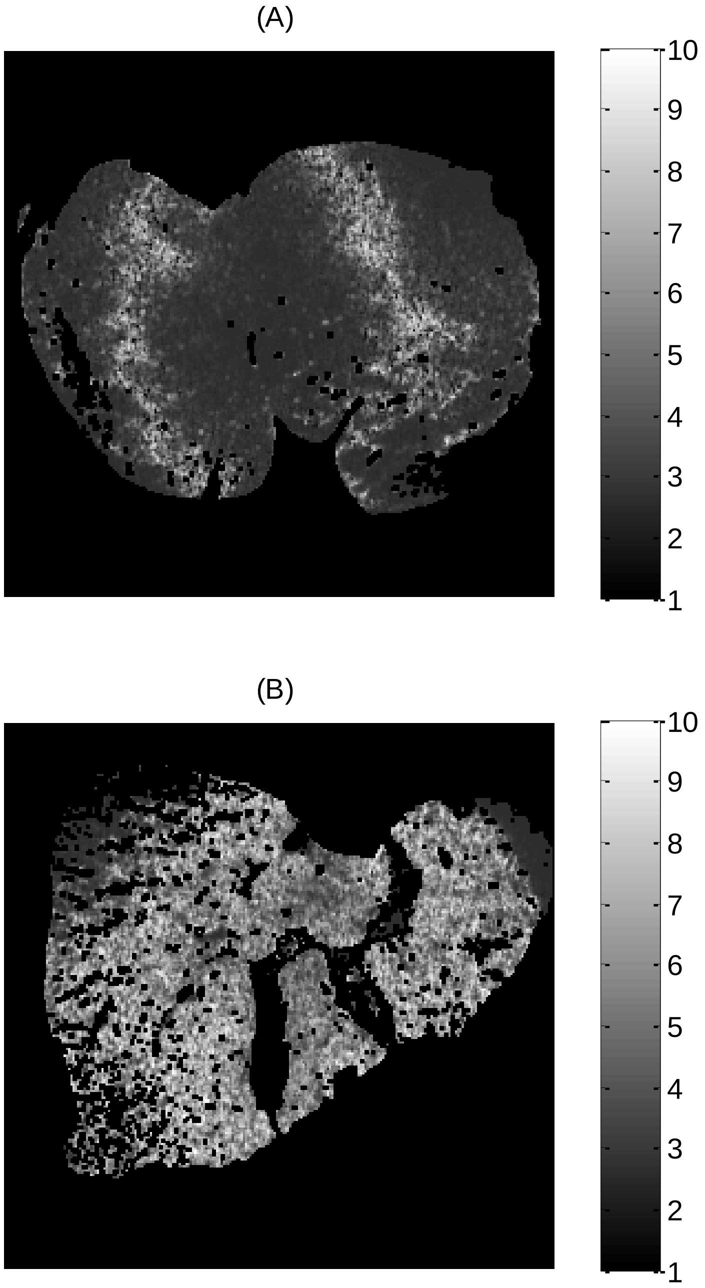


Figure 3.9.2: Test statistic maps generated from the Hotelling observer from Combination 1 utilizing a 2DCC texture analysis for (A) an F0 liver and (B) F4 liver

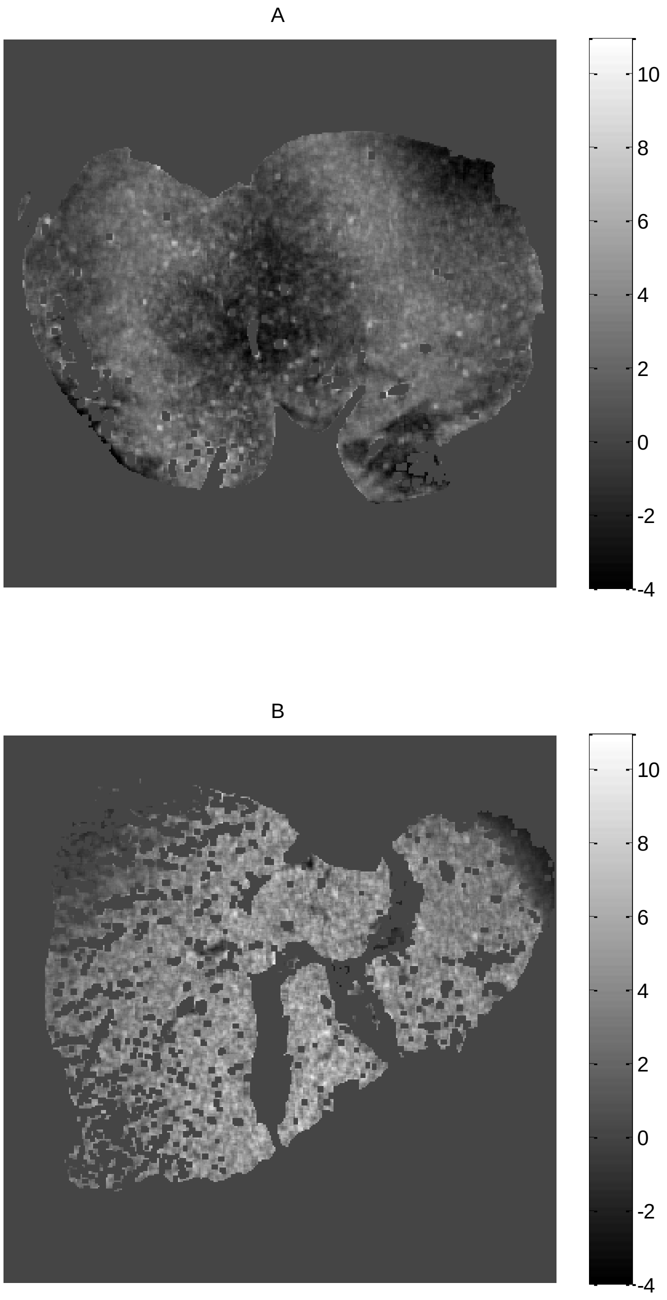


Figure 3.9.3 Test statistic maps generated from the Hotelling observer from Combination 1 utilizing a wavelet texture analysis for (A) an F0 liver and (B) F4 liver

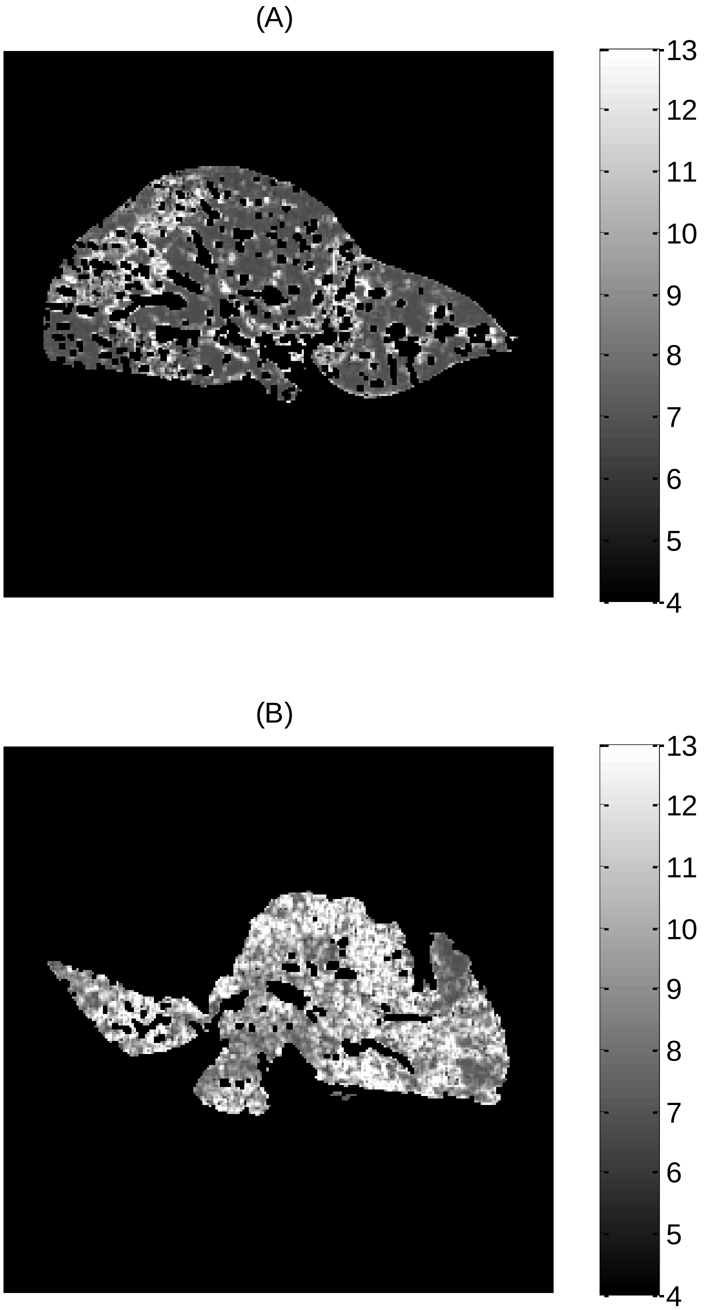


Figure 3.9.4: Test statistic maps generated from the Hotelling observer from Combination 2 utilizing a 2DAC texture analysis for (A) an F1 liver and (B) F4 liver

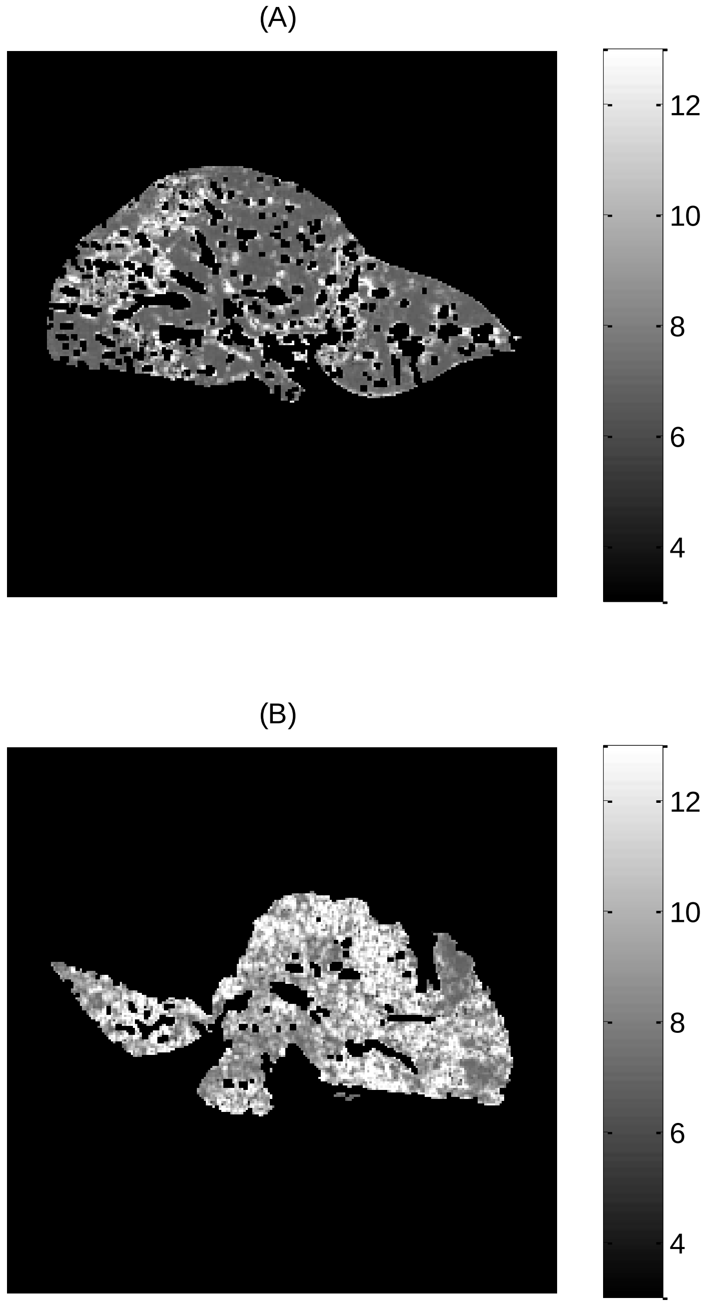


Figure 3.9.5: Test statistic maps generated from the Hotelling observer from Combination 2 utilizing a 2DCC texture analysis for (A) an F1 liver and (B) F4 liver

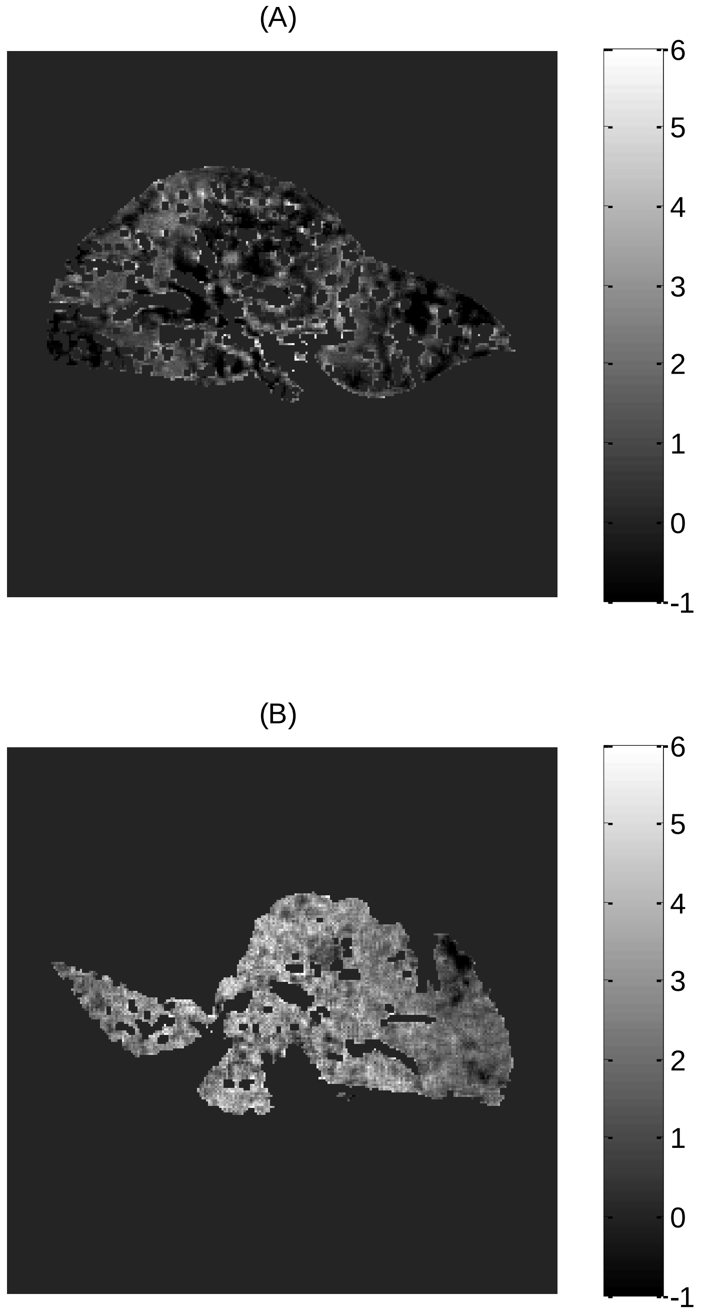


Figure 3.9.6 Test statistic maps generated from the Hotelling observer from Combination 2 utilizing a wavelet texture analysis for (A) an F1 liver and (B) F4 liver

3.10 Limited Training for an Observer

The development of a mathematical observer requires sampling the data to measure the first and second order statistics of the data. The covariance matrix is a crucial component of the trained Hotelling observer. An NxN image has a covariance matrix N^2xN^2. The covariance matrix is also symmetrical, requiring $N^2/2$ independent measurements to sample the matrix effectively. When building an experiment, there may be shortages of training data.

Initial experiments are conducted on ROI's 7x7 pixels. We then expanded the ROI area from 7x7 to 9x9 in order to test the effect of including more of the textured signal. With a 9x9 ROI, the training requirements increase as well. For a 2D autocorrelation, the covariance matrix is 144x144 elements, requiring 10,368 independent measurements, and the 2DCC results in a covariance matrix 40x40, which requires 800 independent measurements to train a covariance matrix. With the samples available, the phantom in Figure 3.1.1 A has 144,798 9x9 ROIs, Figure 3.1.1 B has 27,496 ROIs, the phantom in Figure 3.1.1 C has 63,521 independent ROIs, and Figure 3.1.1 D has 28,133 independent ROIs.

Before testing a covariance matrix for a 9x9 ROI, we test observers based on a 7x7 ROI with under-sampled covariance matrices. Using the 2DCC, the performance of the

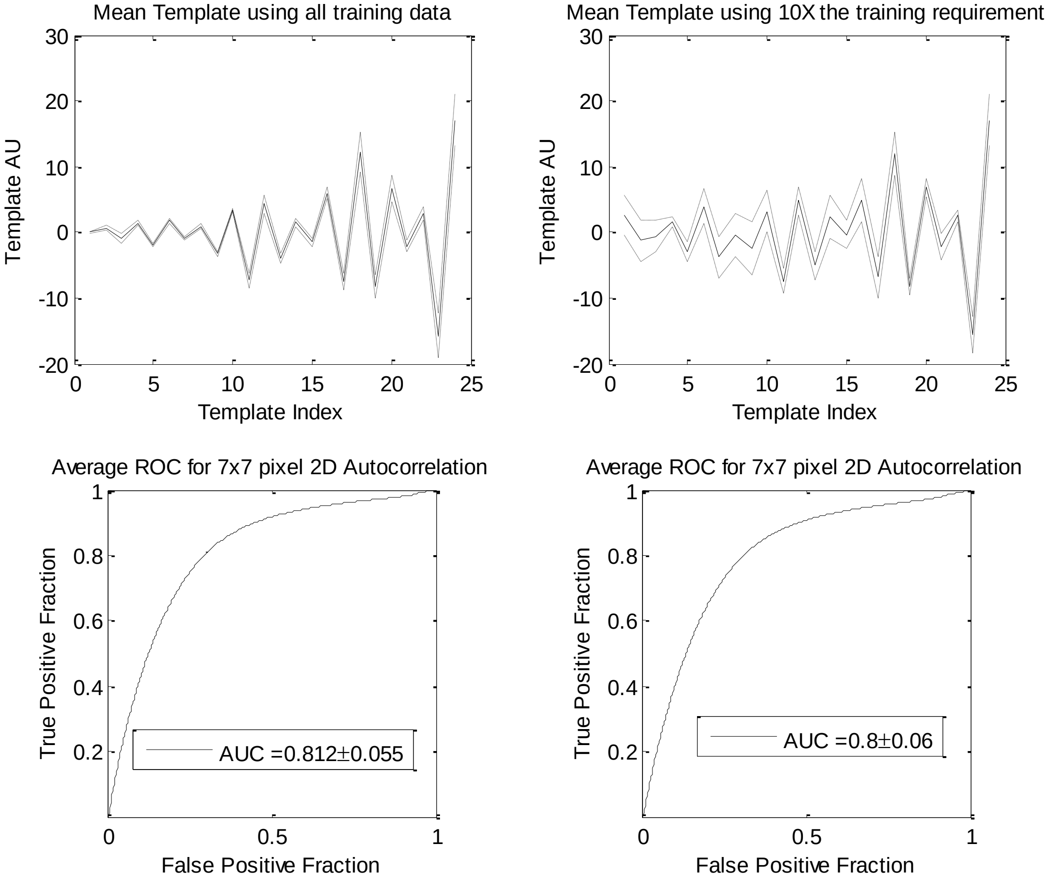


Figure 3.10.1: Mean Template ±2σ based on different amounts of training data. The observer does not degrade as the sample size decreases and mean ROC's ±2σ from respective covariance sampling sizes.

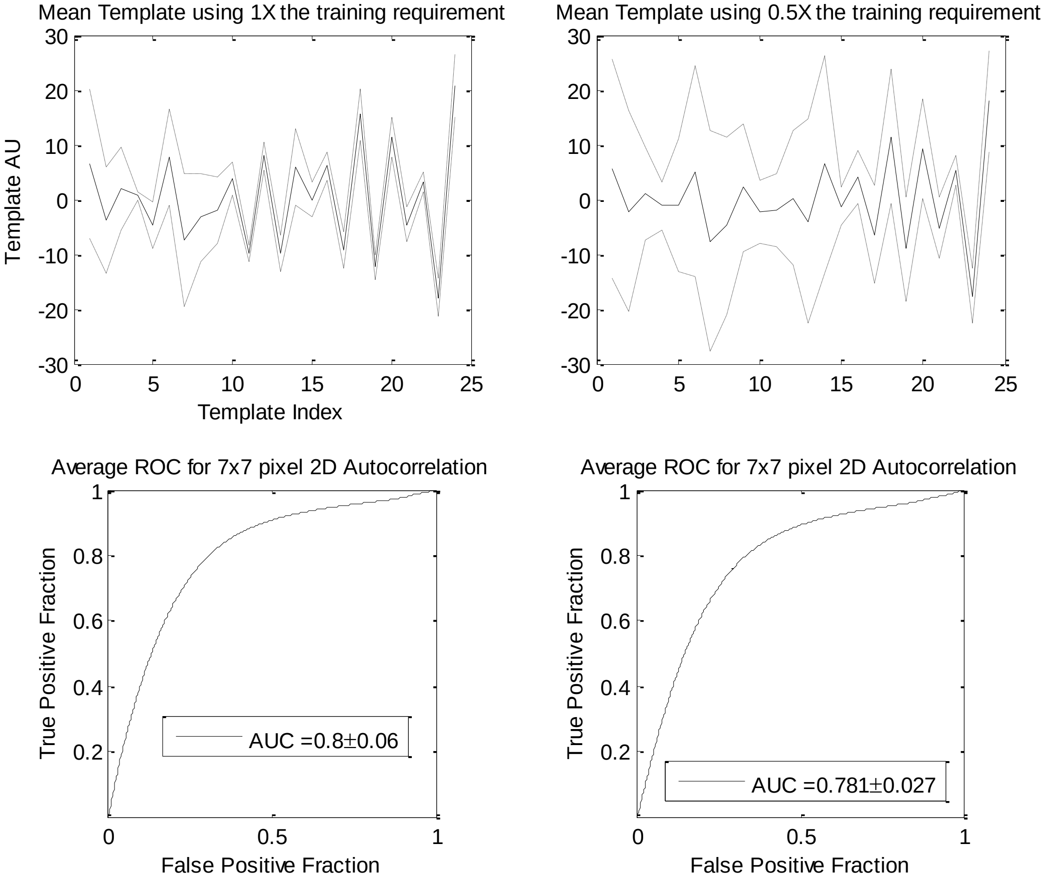


Figure 3.10.2 Mean Template ±2σ based on different amounts of training data. The observer does not degrade as the sample size decreases and mean ROC's ±2σ from respective covariance sampling sizes.

templates using 10x the sampling requirement, the minimum training requirement and half the training requirement were evaluated.

The templates shown in Figure 3.10.1 and Figure 3.10.2 are from each sampling population. The templates have some variance, but appear sensitive to similar features. Each template is applied to testing data sets and ROC analysis recovers an AUC in Figure 3.10.1 and Figure 3.10.2. The AUC's have an error range of less than 0.1. A well trained covariance matrix is required for observer work.

Following this result, the four combinations of training data were used to calculate four observers that were tested on the four independent testing groups for a 9x9 ROI. The ROC results are shown in Figure 3.10.3 for the 2DAC analysis, and Figure 3.10.4 for the 2DCC analysis.

While the covariance matrix for some combinations is under trained for the 2DAC, the 2DCC was always appropriately trained. The AUC results show a modest improvement in observer performance with a larger ROI. This result is expected in the larger ROI because more data from the fibrotic structure is sampled by the observer and the feature is repeated throughout the liver. If the ROI continues to increase, eventually, long range features will influence the observer. Also, if too large an ROI is used, larger areas of the image will be ignored surrounding the vasculature of the liver.

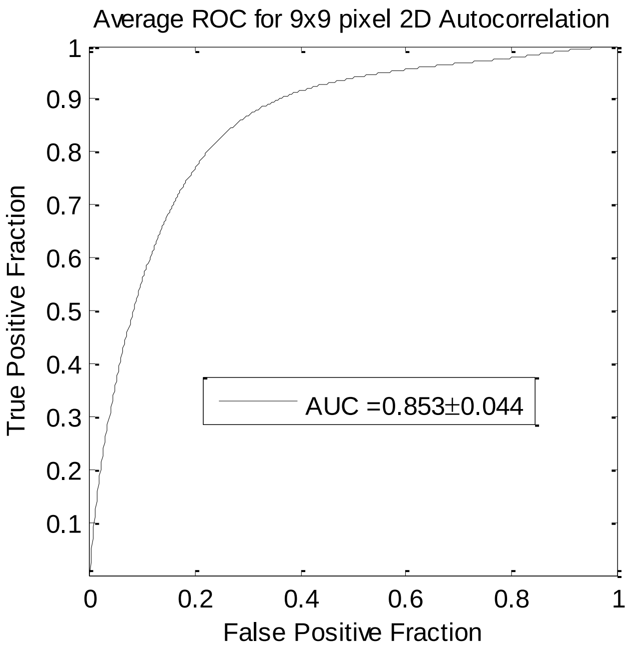


Figure 3.10.3: Mean ROC ±2σ results for a 9x9 local analysis region using a 2DAC

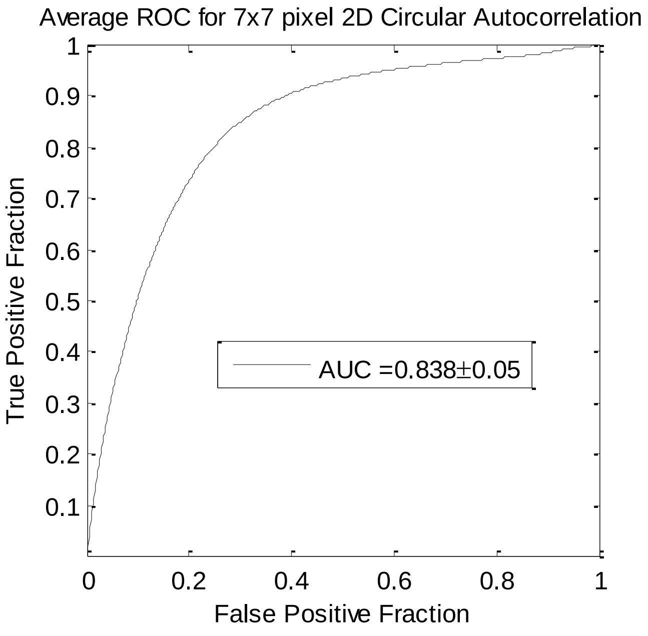


Figure 3.10.4: Mean ROC ±2σ results for a 9x9 local analysis region using a 2DCC

3.11 Conclusions from Chapter 3

In this chapter we describe how MRI can be used to image phantoms that replicate images containing HF in delayed phase DE-MRI. The phantoms are autopsy liver sections fixed in formalin and have multiple tissue samples evaluated by a trained pathologist. We developed a method to remove background and normalize the data.

We discussed the details of how to train and test an observer and showed results with strong AUC values. Finally, we introduced a novel approach to mapping test statistic values with the intent of guiding a radiologist to identify local HF. We tested the reproducibility of observers based on the amount of training data.

The template based on the wavelet transform performs as well as the autocorrelation templates. One drawback is the additional time required for this analysis, which is currently the slowest analysis method with respect to total computation time. It is possible to decrease the computation time, but since the autocorrelation technique is performing equally well, there is no need to spend resources trying to improve the code. Finally, increasing the ROI size increased the performance of the texture analysis observer, but requires additional training data that is not always available.

4. Optimizing the MRI Sequence

In clinical MRI, the operator has control over the *TR*, *TE*, and FA. The operators receive radiologist feedback to confirm if the sequence is collecting images with diagnostically acceptable contrast.[16,38] The image sequences used by radiologists are not necessarily ideal for performing the quantitative task of separating images of *F0* and *F4* liver images with a mathematical observer.[38,61] Our goal is to use task-based performance assessment using the linear ideal (Hotelling) observer to determine the ideal parameters for maximizing sensitivity to fibrotic structures in MR imaging.

The *TR*, *TE*, and FA parameters directly contribute to the contrast of an MR image. However, *TR* and *TE* also directly impact the scan time of the image sequence. Increasing the duration of the MR acquisition in the abdomen is not desirable due to increased artifacts from motion associated with the patient's breathing. Changing the FA also affects overall contrast, but without significantly impacting the length of the sequence. For this reason, we focused on determining the ideal FA for the MR sequence that will be used to assess HF.

We use area-under-the-receiver-operator-characteristic curve (AUROC) as the figure of merit, and present results of a study to find the optimal FA for detecting HF in liver phantoms. This optimization method is translatable to the clinical setting.

4.1 MRI of the Phantoms

All images in this optimization study are collected on a Siemens 3T Skyra MRI using the Siemens flex body imaging coil with a 3D gradient-echo T1-weighted imaging sequence (3D VIBE, Siemens) with *TR/TE*=9.79ms/4.44ms and a range of FAs available from ~10° to ~50°. FAs available are based on hardware limitations. A field-of-view (FOV) of 26.5x26.2x3.36cm with a sampling matrix of 768x760x96 was selected, resulting in images with isotropic resolution of 0.35mm^3. All images were collected at room temperature (22^o C). The total scan time at one FA was approximately 25 minutes.

Before training the mathematical observer to perform the task of HF-detection on liver tissue, a basic threshold was implemented to remove areas of the image that contained blood vessels from the analysis as shown in section 3.3 . We found this to be a necessary step in developing the observer technique.

Each phantom was imaged at five FA's: 8°, 15°, 19°, 30°, and 45° in order. Selected slices from each phantom at 19° are shown in Figure 4.1.1 The images suggest that there is visible contrast between the ECM and liver tissue in the *F4* livers that appears at the expected length scale associated with fibrosis. The images from the *F0* and *F1* phantoms appear to lack the texture associated with ECM in the images and the liver tissue appears uniform in signal throughout a majority of the tissue.

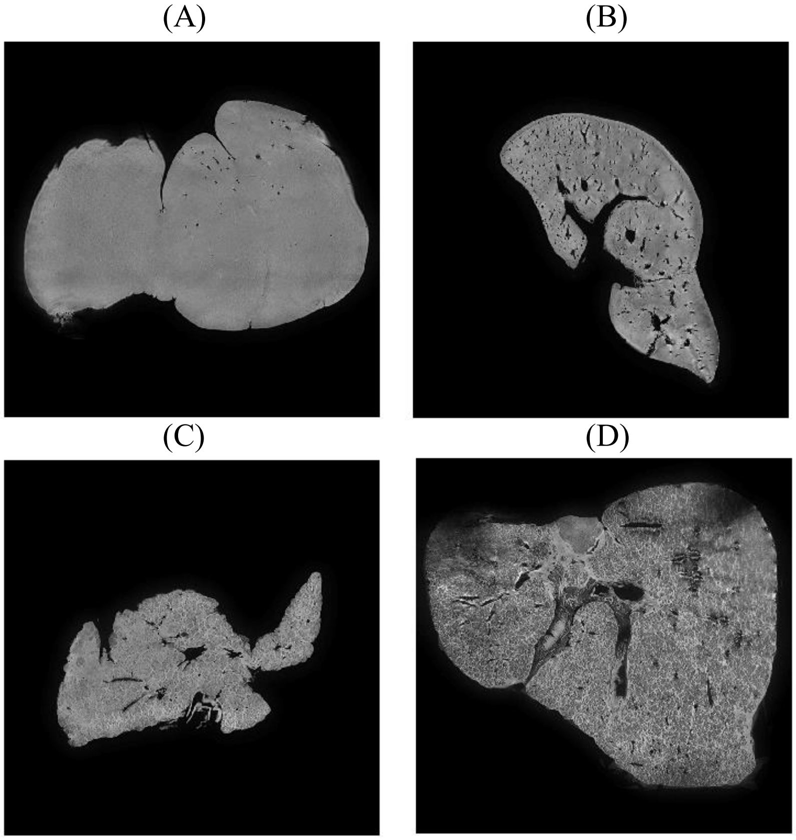


Figure 4.1.1 Representative slice images of (A) F0, (B) F1, (C) and (D) F4 phantoms at a 19° FA. Image resolution is 0.35mm3 isotropic.

4.2 Training Mathematical Observers

The set of local 2DCCs from the *F0* and *F1* phantoms comprise our null hypothesis, or signal absent data, for training a mathematical observer. The 2DCCs from the two F4 phantoms comprise the signal present data. To avoid bias in the results, only one phantom was used to train the observer; the other phantom was selected as the testing data. With four phantoms, two in each hypothesis class, we can derive and test 4 independent observers to check for reproducibility. The linear observers for each FA are shown in flattened 1D form in Figure 4.2.1. We found that the templates all detected the same features, regardless of choice of training and testing data and FA – namely the peaks in the 2DCC function associated with the ECM cell size.

The Hotelling observer has a template form $\vec{w}$ which one can visualize, whereas the quadratic observer does not. The quadratic observer is an extension of the linear observer and performs better when $\boldsymbol{K}_0 \neq \boldsymbol{K}_1$. The sample covariance matrices for each FA for a representative signal absent and signal present training combination are shown in

Figure 4.2.2. It is clear from the figures that the covariance matrices have similar features that one can visually distinguish, but they are not strictly equal to one another between the two classes.

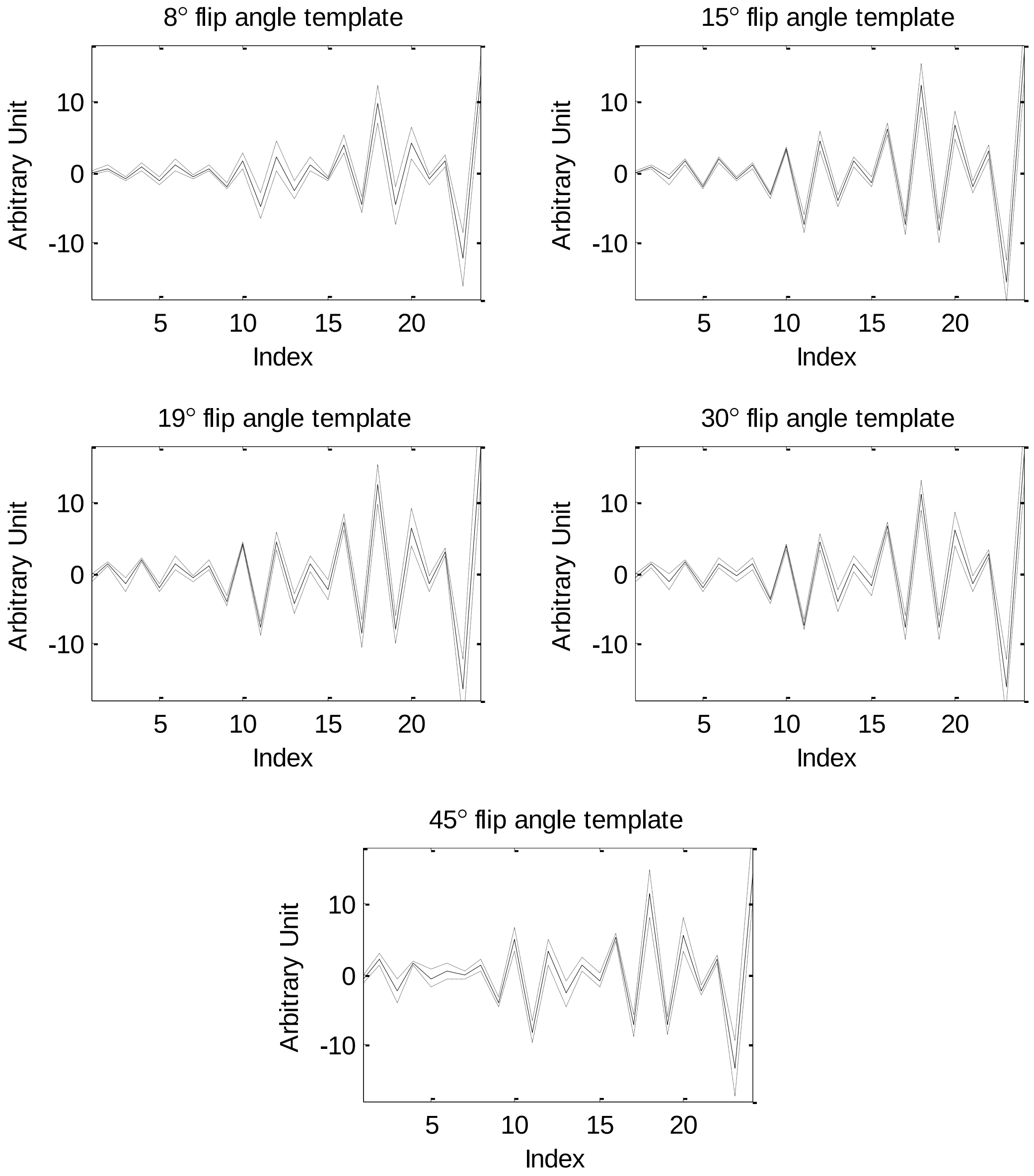


Figure 4.2.1: Average template ± σ variation between different training and testing data combinations for each acquired FA.

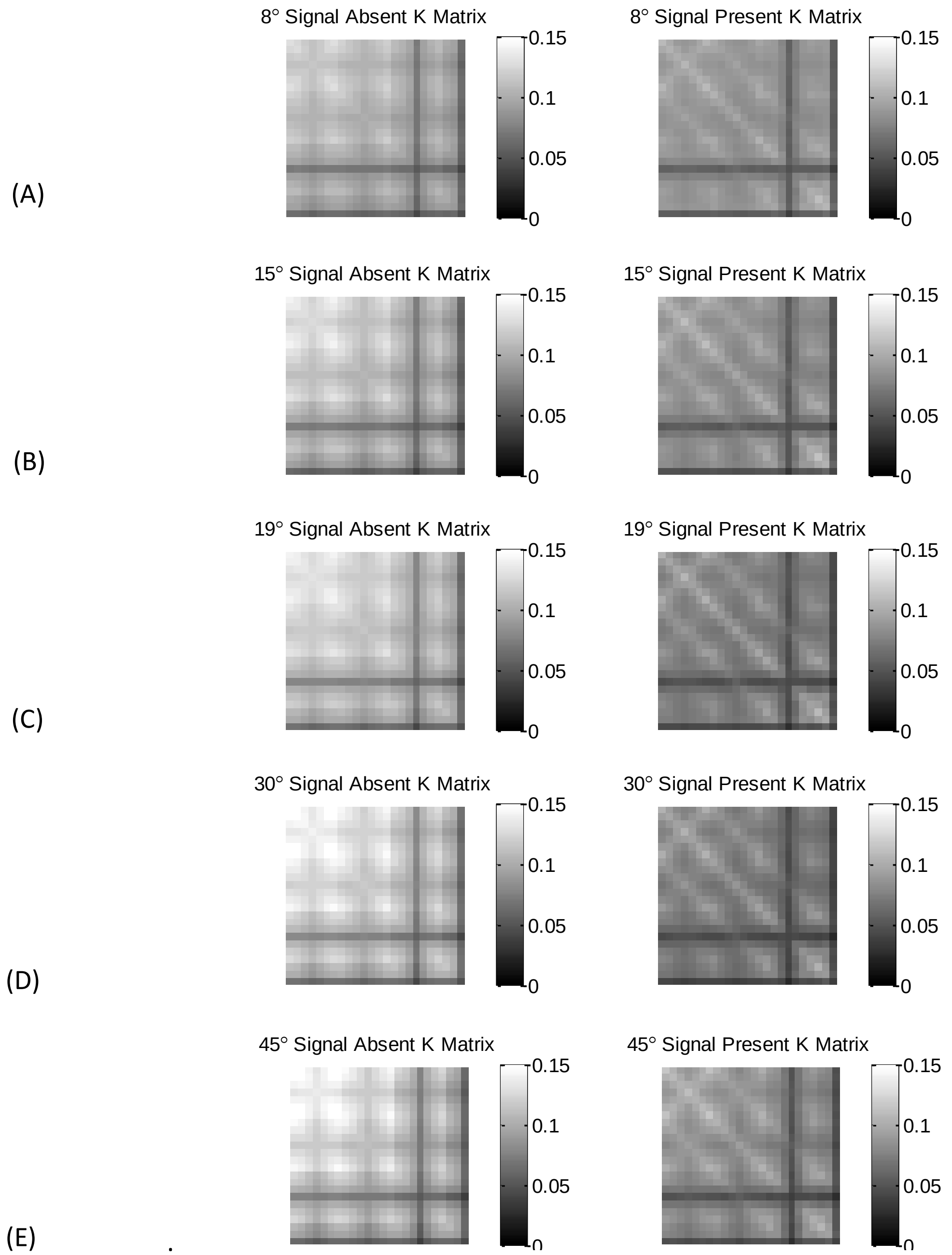


Figure 4.2.2: A-E shows the sample covariance matrices for a representative set of training data at each FA

4.3 Results and Optimization

The four phantoms allow for four different combinations of training and testing data. For each combination, ROC analysis was performed and the AUROC for each combination was calculated as a function of FA. Figure 4.3.3 shows the AUC as a function of FA for each combination of training and testing data. Each combination has similar behavior with a peak around 23°. The mean relative AUROC was plotted as a function of FA in Figure 4.3.3, after a minimal least squares error shift to remove systematic variation between training and testing combinations. The plot of these values for the quadratic observers are shown in Figure 4.3.4. The optimal FA was chosen based on maximizing the AUROC. The optimal FA for both the linear and quadratic observers was found to be near 24°.

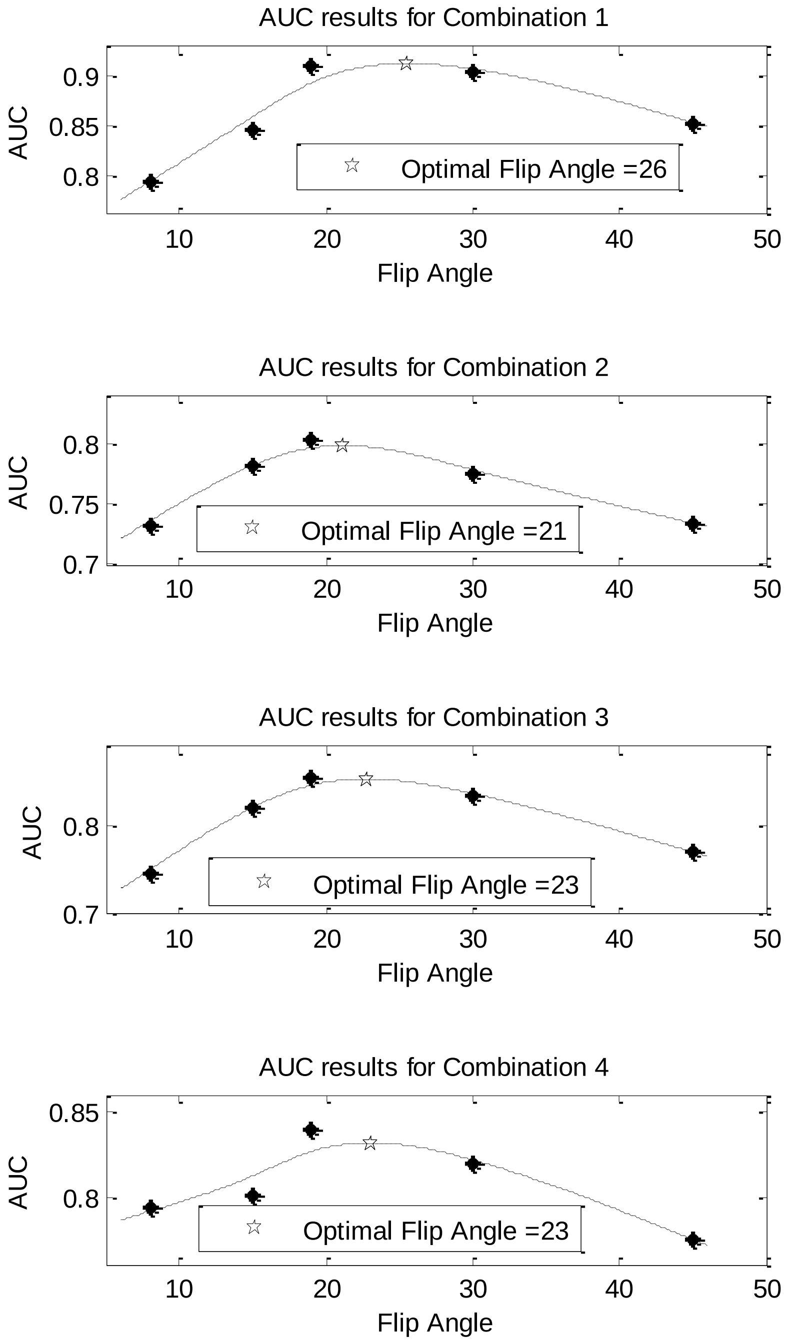


Figure 4.3.1: AUC as a function of FA for the four independent combinations of training and testing data using the linear observer

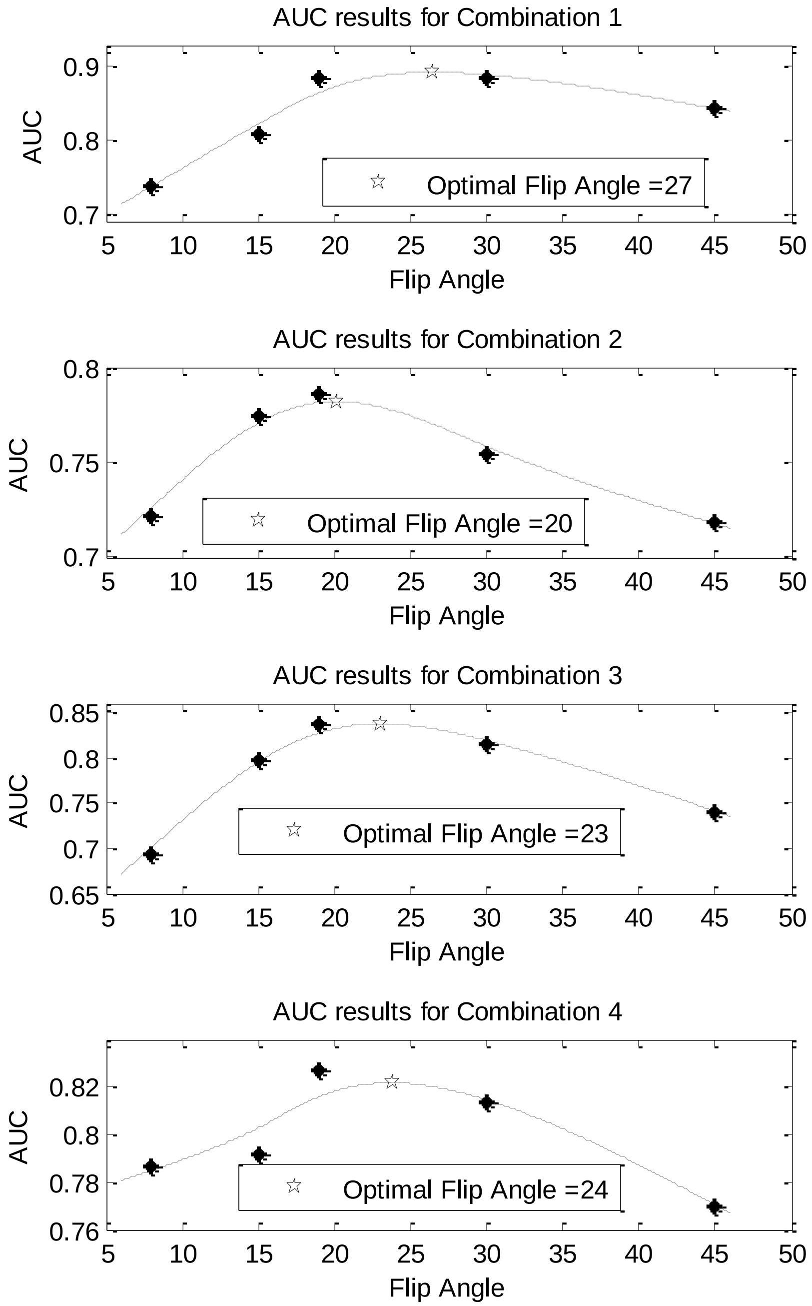


Figure 4.3.2: AUC as a function of FA for the four independent combinations of training and testing data using the quadratic observer

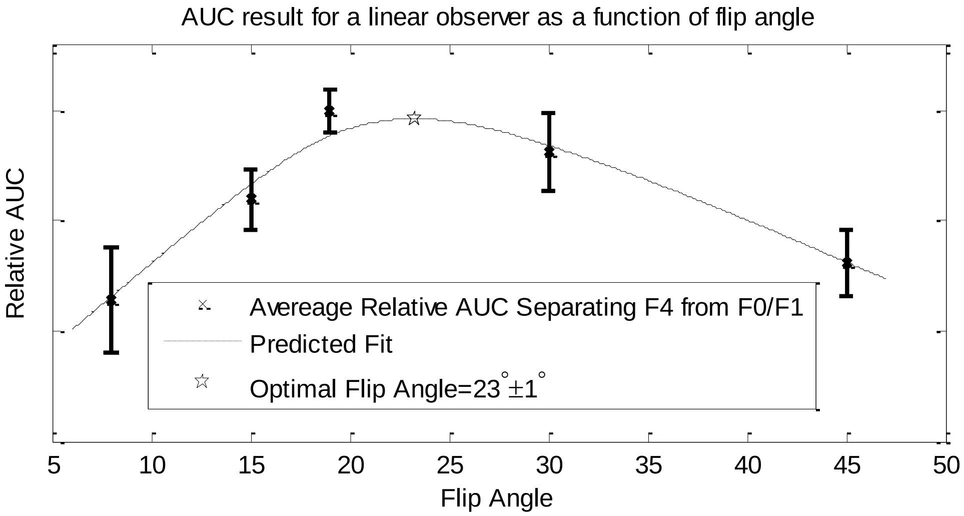


Figure 4.3.3: Plot of the relative AUROC for a linear observer as a function of FA.

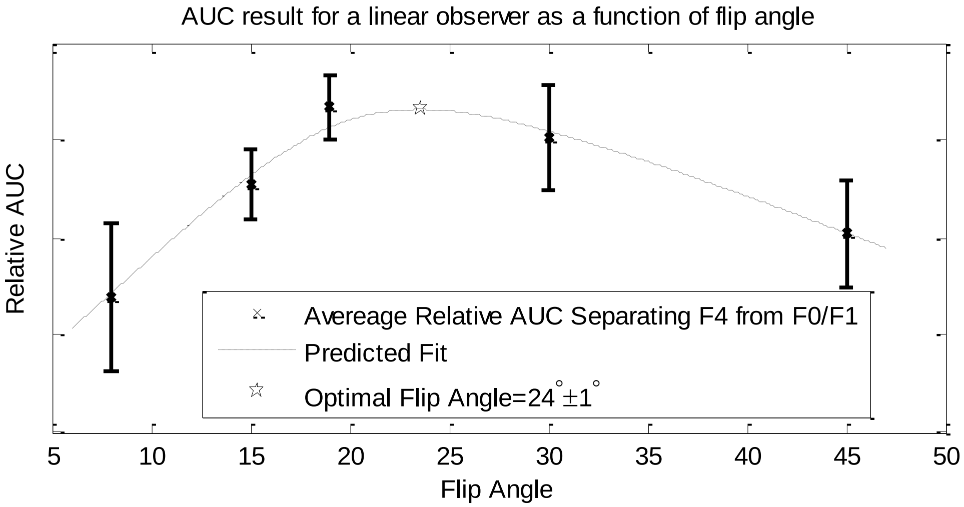


Figure 4.3.4: Plot of the relative AUROC for a quadratic observer as a function of FA.

4.4 Quadratic and Linear Observer Performance Comparison

An unexpected result of this experiment was a decrease of the AUC when the quadratic observer was applied to the data. This is unexpected because when the covariance matrices of the classes are compared in

Figure 4.2.2, they are not equal as required by the assumptions for the Hotelling observer to be the idea observer.

To confirm this result, a linear and quadratic observer were tested using the same testing and training data. The observers were trained and tested on independently gridded ROI's. One healthy and one diseased liver were used as signal absent and signal present imaging sets. The trained means and covariance matrices are shown in Figure 4.4.1 and Figure 4.4.2. These means and covariances are similar to previously discussed data.

The PDF's on the test statistics from the same training data are highlighted in Figure 4.4.3 and Figure 4.4.4 for the linear and quadratic observer respectively. The SNR of the linear observer is 1.15, and 0.63 for the quadratic observer. The linear Hotelling observer performs better than the quadratic observer, confirming the difference in the AUC between the two observers noted in the parameter optimization experiment.

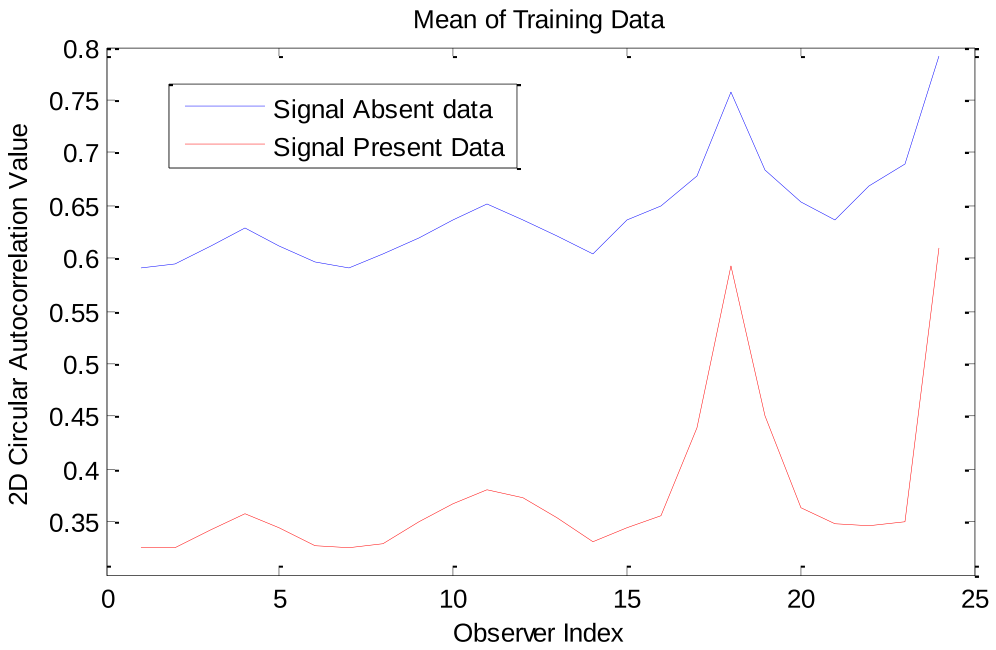


Figure 4.4.1: Trained means for signal absent and signal present data

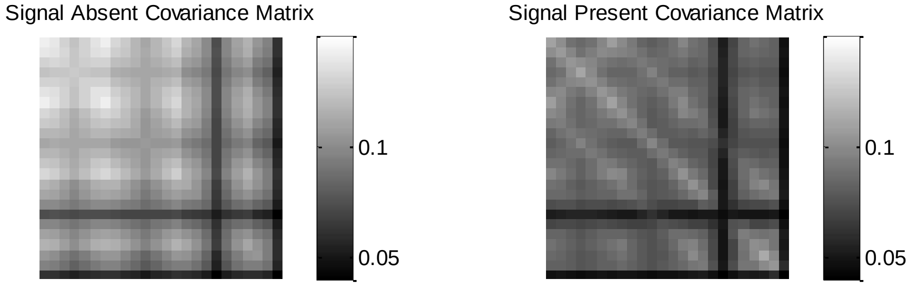


Figure 4.4.2: Trained covariance matrices for signal absent and signal present data

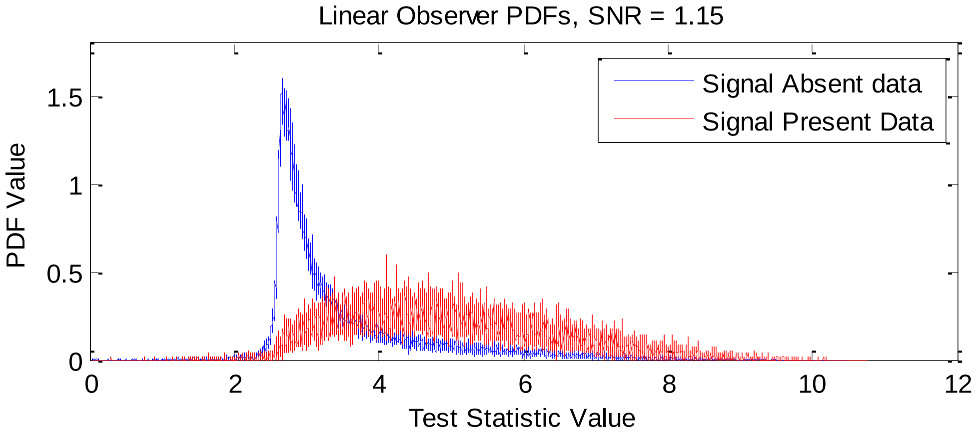


Figure 4.4.3: PDF's collected from the linear observer. The SNR is equal to 1.16, all the test statistics are shown in this plot.

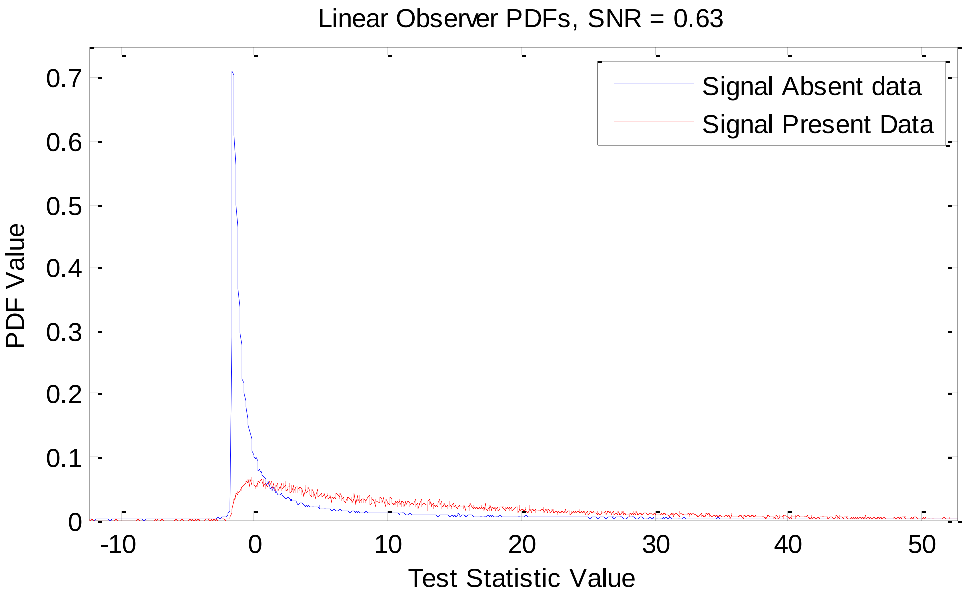


Figure 4.4.4: PDF's collected from the quadratic observer. The SNR is equal to 0.63.

4.5 Conclusions Regarding Optimization

Task-based optimization of MRI acquisition sequence parameters can be carried out whenever a model observer is applied to the MRI images and an AUC value can be computed. This method has general utility for a variety of clinical applications where a choice of a particular sequence is largely histological and not based on quantitative assessment.

The method of isolating an imaging parameter and comparing AUC's based on the same observer is also transferrable to many other imaging modalities.

5. A Two Stage Observer

5.1 Introduction to the Two Stage Observer

In previous chapters a model observer was developed to determine the HF class of a small ROI in the liver. This is shown for 7x7 and 9x9 pixel ROI's in Chapter 3. The results are sensitive to regional coil sensitivities and heterogeneity of structures in the liver due to natural variation. Test statistic maps create a feature map for a radiologist to determine areas of possible fibrosis, but this is still a set of local evaluations.

This chapter develops the method to stage larger regions cirrhotic livers based on the whole set of test statistics recovered from a liver. The first stage observer assesses local texture analysis, while the second stage observer makes decisions based on global data.

5.2 Training a Two Stage Observer

Training a two stage observer requires the collection of enough images to meet training requirements for HF and non-HF covariance matrices. In Chapter 3, the 2DCC was chosen as the easiest observer to work with, given its successful AUC results and light computational requirements. We start this experiment with four phantoms that provide all of the training and testing data by separating slices.

A *F0* and *F4* liver 3D dataset is selected to train both the first and second stage observer. The 3D image is a matrix of size MxNxZ and is divided into two sets by Z. Slices were normalized using methods in Chapter 3.4 and a local 2DCC is applied to adjacent 7x7 pixel ROIs as outlined in Chapter 3.5 The local texture data are used to train the first stage observer,

$$\overrightarrow{w_1} = \left(\frac{\boldsymbol{K}_{1_S} + \boldsymbol{K}_{0_S}}{2}\right)^{-1} \left(\bar{\vec{S}}_1 - \bar{\vec{S}}_0\right).$$

(5.2.1)

The first stage observer is applied to the second half of the training data set and a new image $g'_{z,q}$, where z is the image slice and q is the known class of the training data. The new image is generated from application of the first stage observer to the training data,

$$t\left(\overrightarrow{g'_{z,q}}\right) = \overrightarrow{w_1}'\overrightarrow{S_{m,n,z}}.$$

(5.2.2)

$g'_{z,q}$ exists in τ-space and is normalized to an area of one to recover a probability density function as a function of the test statistic.

Channelizing the image is one method to reduce the dimensionality of the problem. Ten channels were selected across τ-space and all values outside the channels were ignored. t_{min} and t_{max} were selected based on the results of equation 5.2.2. The 10 channels are selected at even intervals between the t and t_{max},

$$\overrightarrow{g''_{z,q}} = \overrightarrow{w_1}'\overrightarrow{S_{m,n,z}} \in \left[t_{1,}t, \ldots, t\right], \tau_{min} = t, t_{max} = t_{10}.$$

(5.2.3)

Image $\overrightarrow{g''_{z,q}}$ is used to train the second-stage Hotelling observer,

$$\overrightarrow{w_2} = \left(\frac{\boldsymbol{K_0''} + \boldsymbol{K_1''}}{2}\right)^{-1} \left(\overrightarrow{\overline{g_1''}} - \overrightarrow{\overline{g_0''}}\right).$$

(5.2.4)

5.3 Testing a Two Stage Observer

Testing $\overrightarrow{w_1}$ and $\overrightarrow{w_2}$ is more straightforward than training $\overrightarrow{w_1}$ and $\overrightarrow{w_2}$. An independent 3D data set MxNxZ of an *F0* and *F4* liver were collected. Using the method in Chapter 3.7 each MRI slice is evaluated by $\overrightarrow{w_1}$ to create a new image $\overrightarrow{g'_{z,q}}$. $\overrightarrow{g'_{z,q}}$ is divided into the same channels between τ_{min} and τ_{max} to generate $\overrightarrow{g''_{z,q}}$. The final scalar test statistic represents the evaluation of the local texture analysis results across an entire slice of an MRI image,

$$t_q\left(\overrightarrow{g''_{z,q}}\right) = \overrightarrow{w_2}'\overrightarrow{g''_{z,q}}$$

(5.3.1)

ROC analysis was performed over τ_q to recover the AUC as the figure of merit for the overall classification of the slice.

5.4 Results of Training the First Stage Observer

Training data was based on one liver of size MxNxZ slices. The initial observer was trained on half the slices available, making the training data size, M x N x Z/2. Slices were normalized using the methods in Chapter 3.4 4. The local 2DCC is applied to adjacent ROI's, and a Hotelling observer was trained using the same methods as in Chapter 3.5 The data were used to train the first stage observer,

$$\overrightarrow{w_1} = \left(\frac{\boldsymbol{K}_{\boldsymbol{1}_{\boldsymbol{S}}} + \boldsymbol{K}_{\boldsymbol{0}_{\boldsymbol{S}}}}{2}\right)^{-1} \left(\bar{\vec{S}}_1 - \bar{\vec{S}}_0\right)$$

(5.4.1)

Figure 5.4.1 shows the template, based on a 2DCC, recovered from half the training data recovered from one F0 and one F4 phantom. Visually comparing this template to ones previously calculated, there is no change evident due to the smaller training, but still full rank data set.

The template was then applied to the other half of the phantom images, treating them as testing data. The technique for testing the observer was outlined in Chapter 3.7 . For each slice a histogram of the test statistics was recovered. These histograms are shown in Figure 5.4.2. Each slice has similar number of analyzed ROIs.

To use the curves as data, additional normalization is needed. The observer must be independent to the number of ROI's in the region. Also, outliers must be eliminated from the data. To account for these two issues, each curve associated with a slice of MRI data was normalized to have an area equal to one. Also, outliers were eliminated based on results of the training histograms. The results for each slice, z, are plotted in histograms across 10,000 bins in Figure 5.4.2. We selected t_{min} and t_{max} based on these results. For this data, τ_{min} was selected as 4, and t_{max} was selected as 14.

The training requirement for the second stage observer is based on the number of bins selected for the histogram after the application of the first template. We cannot collect enough training data for an observer based on 10,000 points. To avoid this, the first stage testing data was binned into 10 channels of equal width along the test statistic axis.

The histograms for the 16 slices are shown in Figure 5.4.3. These curves were used to train the second stage Hotelling observer. Figure 5.4.4 shows the means of the two classes, $\overrightarrow{g_0'}$ and $\overrightarrow{g_1'}$. Figure 5.4.5 shows the covariance matrices of each class, $\boldsymbol{K_0'}$ and $\boldsymbol{K_1'}$. These data are used to train the second stage Hotelling observer from equation 5.4.2.

The channelized template is shown in Figure 5.4.6. With the observers in Figure 5.4.1 and Figure 5.4.6, testing data are used to determine if the two classes are separable on an slice-by-slice bases.

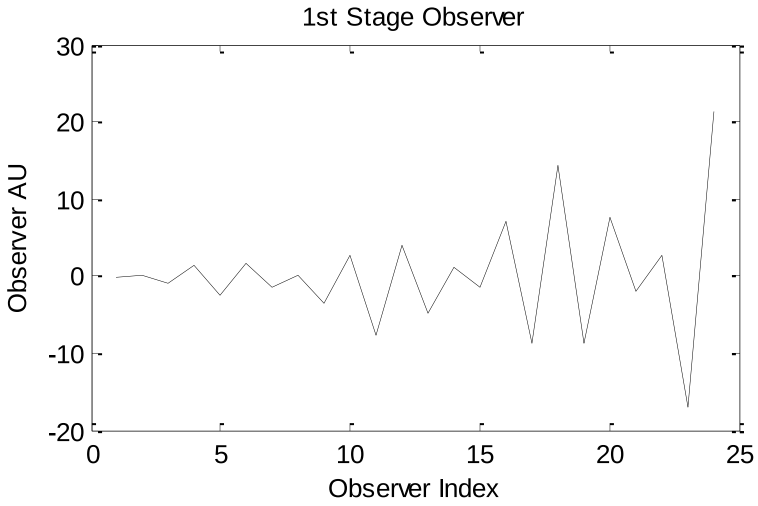


Figure 5.4.1: 7x7 ROI 2DCCTemplate recovered using half of the training data available.

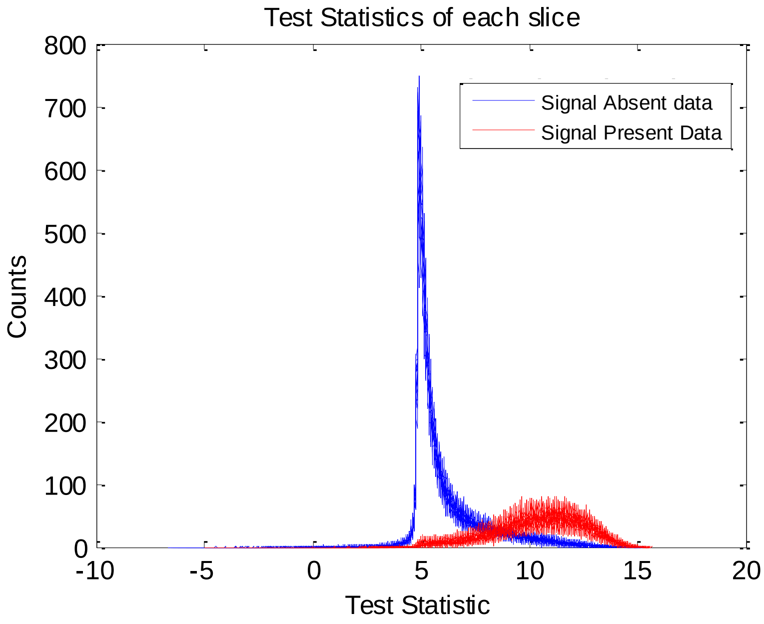


Figure 5.4.2: Histogram of test statistics for 16 slices of testing data.

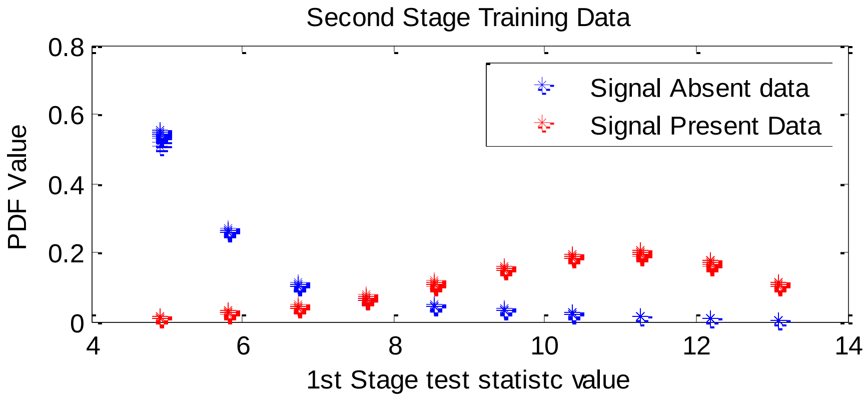


Figure 5.4.3: Histograms of each slice from first stage observer training results. These curves are used to train the second stage observer

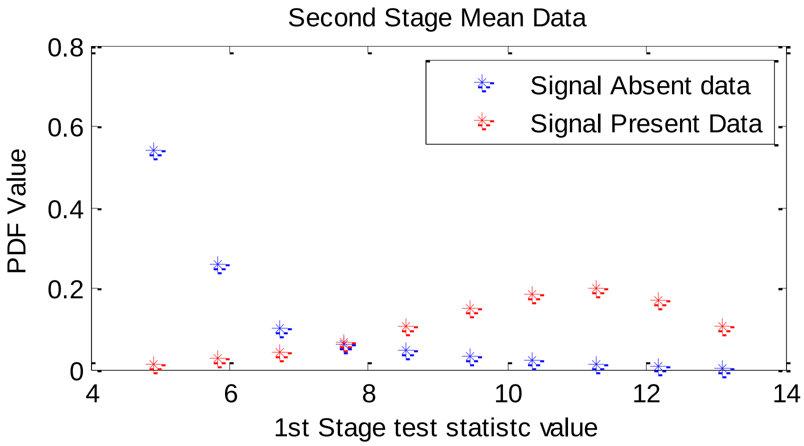


Figure 5.4.4: Means of normalized slice data

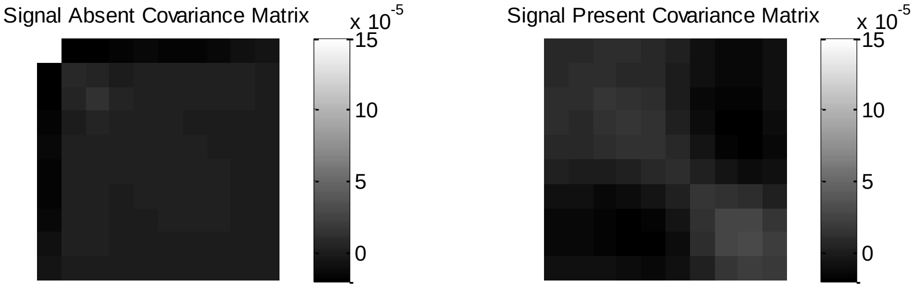


Figure 5.4.5: Covariance matrix data for each class for the second stage observer

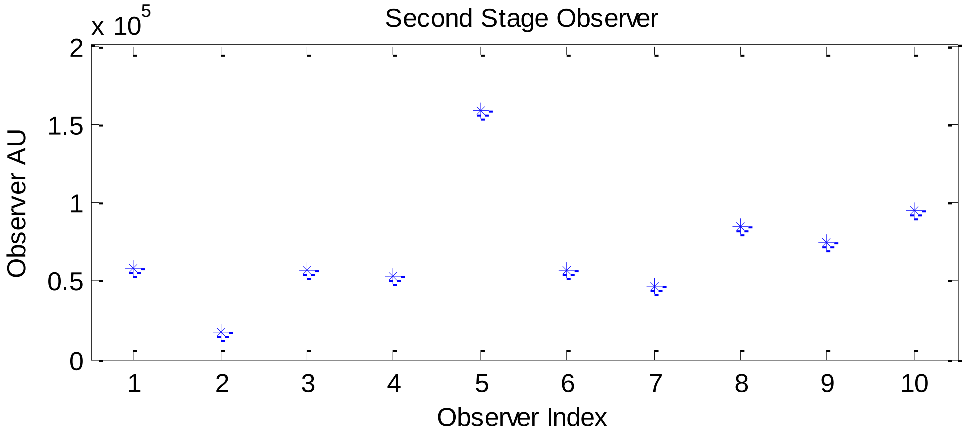


Figure 5.4.6: The second stage observer for one combination

5.5 Results of Testing the Two Stage Observer

Trained observers were tested on independent liver phantoms. Observers were applied to all available the slices in the healthy and diseased phantoms, which is phantom dependent. The slice data was normalized to an area of one, and the outlying test statistics were ignored. The result of applying the first stage observer, in Figure 5.4.1, are shown in Figure 5.5.1.

The second stage observer is applied to each histogram curve, resulting in a single scalar test statistic that represents the HF status of the entire slice. These results are shown for Combination 1-4 if Figure 5.5.2. ROC analysis was performed on the results of all the combinations, the AUC was equal to one for all the combinations of data. This means full separation between F0/F1 and F4 tissue is achievable by the two stage model observer in tissue phantoms.

This technique is expandable to not only slices of MRI data, but regions selected by a radiologist. Each slice is normalized based on area, making various size regions compatible with the two stage observer. A two stage observer is a promising tool to call an entire slice of the liver or entire liver as healthy or diseased.

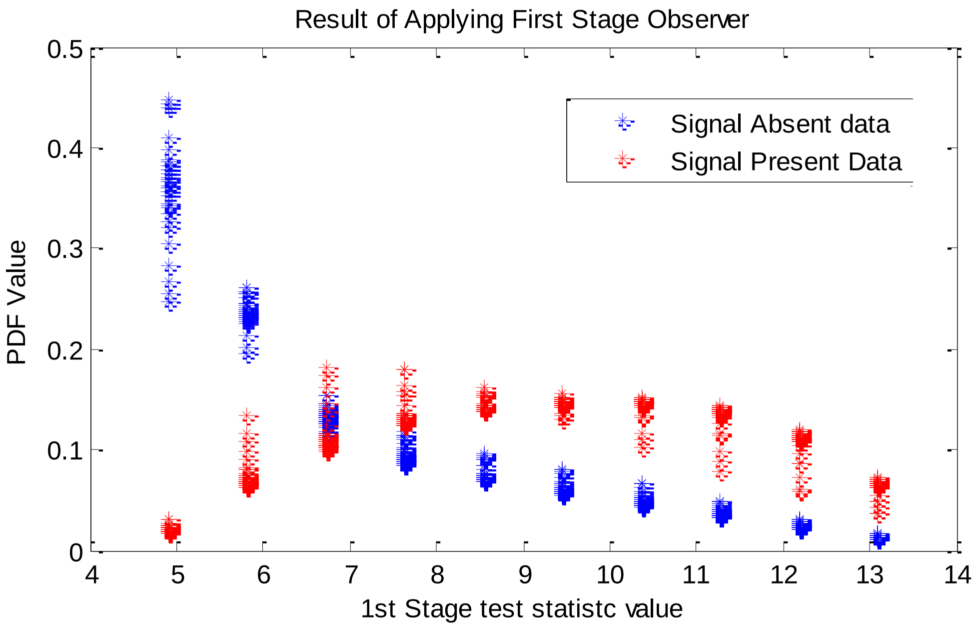


Figure 5.5.1: Application of first stage observer to independent testing data

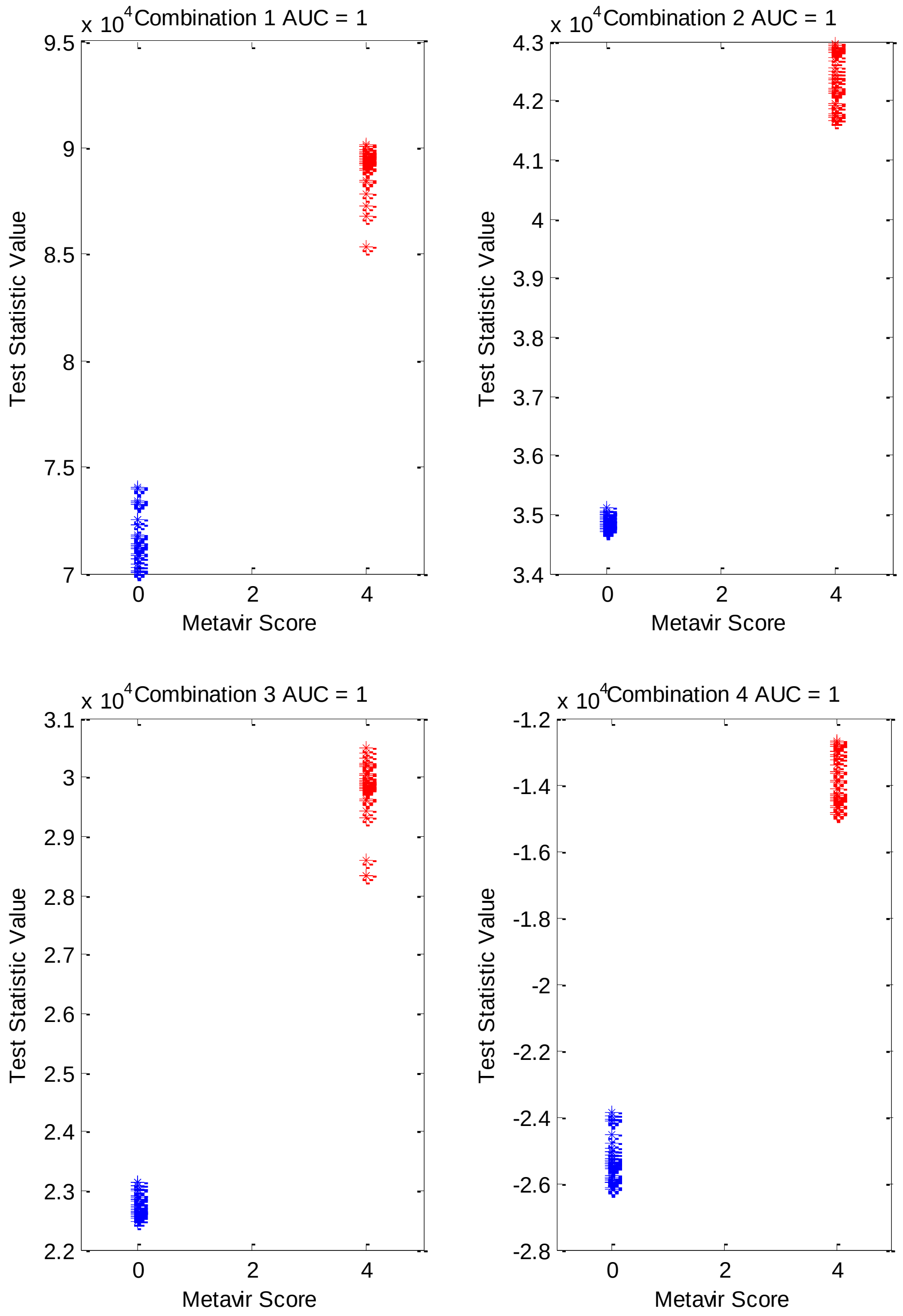


Figure 5.5.2: Results for four combinations of independent training and testing data. All four AUC's equal 1.0.

6. Translation of the Observer Technique to the Clinic

In the final part of this dissertation, necessary steps are outlined to translate the mathematical observers developed in the previous chapters to clinical images. This chapter includes an exploration of the minimum resolution required to perform the task of detecting fibrosis and the steps necessary to carry out imaging in the clinic and collecting training data from patient volunteers

6.1 Minimal Acceptable Resolution

Training and testing observer with the phantoms allows the use of high resolution imaging. Indeed, phantoms were imaged at 0.35mm^3 isotropic resolution. This resolution is not achievable in abdominal MRI in the clinic at this time. The extracellular matrix in cirrhosis is observable in lower resolution images, but there is noise and loss of sensitivity. To assess the impact of the reduced resolution, images of the phantoms with various in-plane resolutions and slice thicknesses were acquired and observers were trained and tested.

The resolution of the images collected were at slice thicknesses of one, two, and three millimeters. The in-plane resolutions tested were 0.35mm^2, 0.70mm^2, and 1.40mm^2. Figure 6.1.1, Figure 6.1.2, and Figure 6.1.3 show the average templates based on slice thickness for each in plane resolution setting. The 0.35mm^2 in-plane resolution templates are not affected by slice thickness. However, as the resolution is degraded, the templates are increasingly affected. An in-plane resolution of $0.70mm^2$ suffers some sensitivity to slice thickness. The templates trained at $1.4mm^2$ in plane resolution have the most change with respect to an increase in slice thickness.

The templates were also trained for 9x9 ROI 2DCC and 8x8 ROI wavelet analyses. Similar effects on templates were seen with the increase in slice thickness for reduced in-plane resolutions.

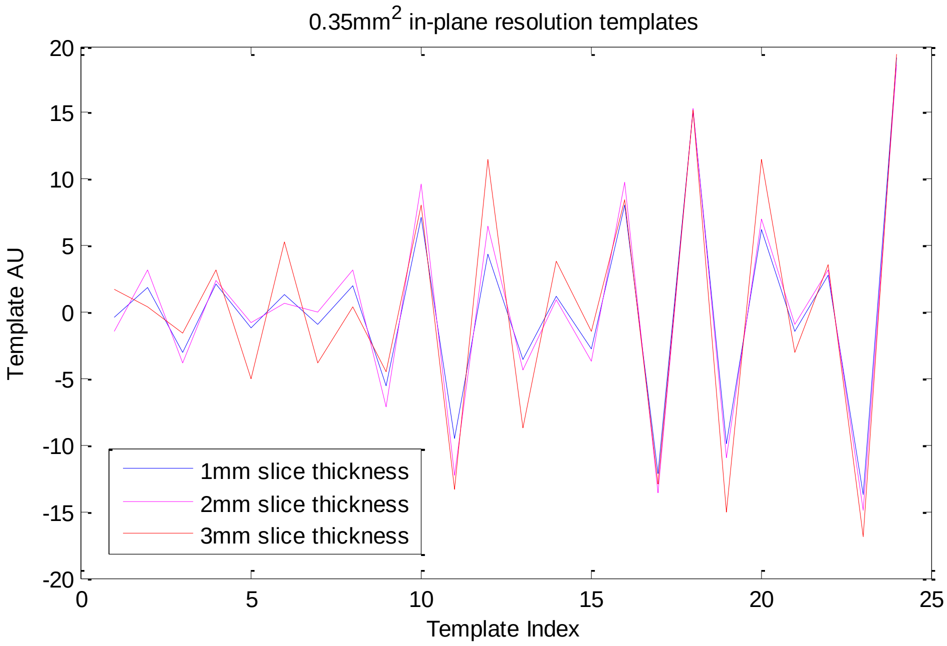


Figure 6.1.1: 7x7 ROI Templates from scans with 0.35mm^2 in plane resolution

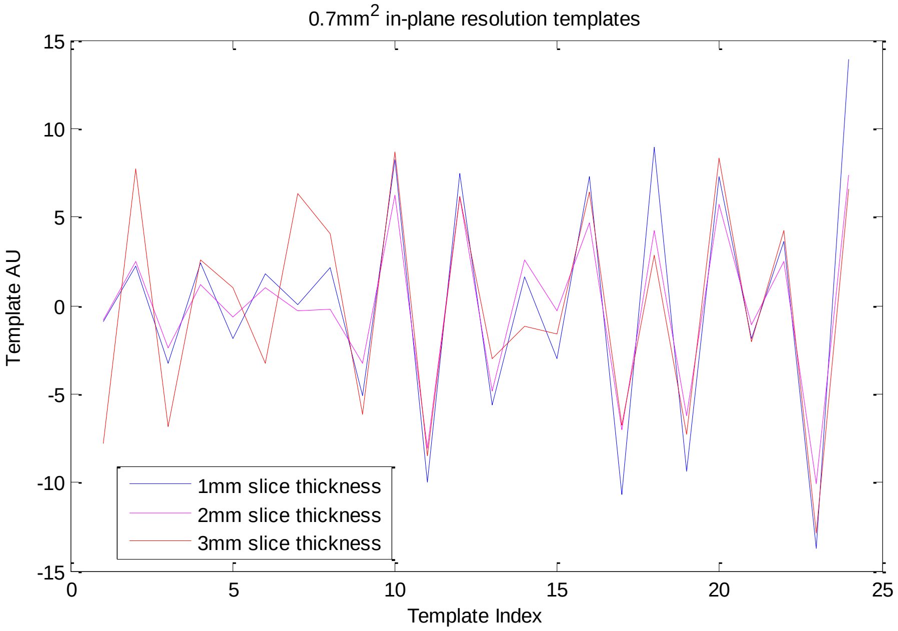


Figure 6.1.2: 7x7 ROI templates from scans with 0.70mm^2 in plane resolution

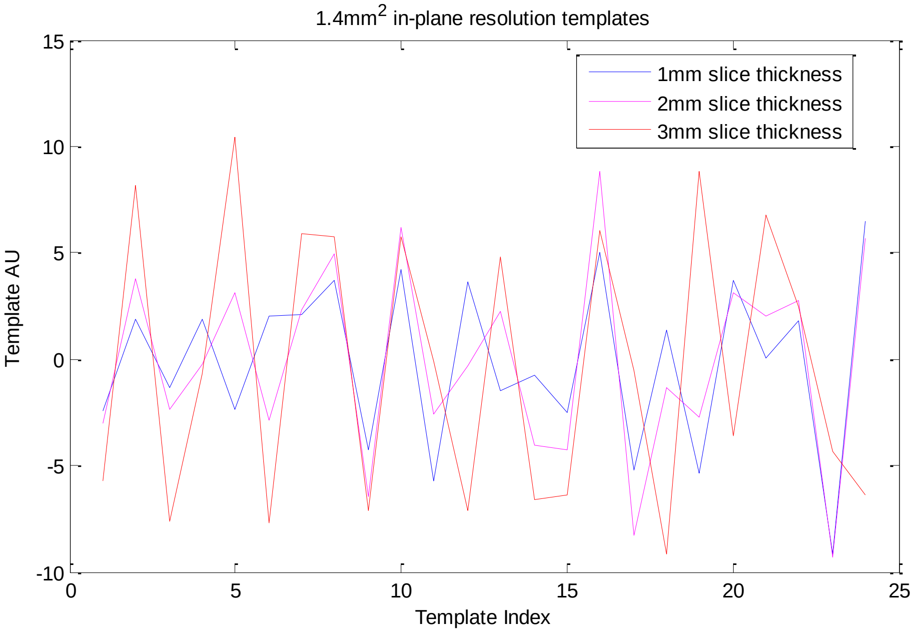


Figure 6.1.3: 7x7 ROI templates from scans with 1.40mm^2 in plane resolution

6.2 Testing Results

At each resolution at which images were collected, four sets of training and testing data were used and AUC's were measured as the figure of merit. Average AUC values for low resolution observer experiments. Each table shows a different texture analysis method. Table 6.2.1 shows the average AUC and standard deviation of the training and testing data for different texture analysis methods. Two texture analysis methods were used: 2D circular autocorrelation and wavelet transform. The 2DCC was applied to two different ROI sizes: 7x7 and 9x9. The wavelet transform was only performed with one template size. To avoid down sampling an odd number of pixels, the next possible template size for a wavelet based texture analysis was 16x16 pixels, resulting in a 256x256 covariance matrix. This is a large training requirement for the limited amount of data that is collected at the time of the experiment. The results confirm that the AUC performance decreases drastically as the in-plane resolution is degraded.

7x7 Circular Autocorrelation Template	Slice Thickness		
In Plane Resolution	1mm	2mm	3mm
0.35mm x 0.35mm	0.76±0.04	0.76±0.08	0.63±0.08
0.70mm x 0.70mm	0.67±0.11	0.53±0.01	0.49±0.08
1.40mm x 1.40mm	0.49±0.02	0.35±0.12	0.39±0.10

9x9 Circular Autocorrelation Template	Slice Thickness		
In plane resolution	1mm	2mm	3mm
0.35mm x 0.35mm	0.79±0.07	0.75±0.07	0.66±0.11
0.70mm x 0.70mm	0.71±0.13	0.58±0.04	0.51±0.08
1.4mm x 1.4mm	0.50±0.11	0.41±0.08	0.39±0.16

8x8 Wavelet Linear Template	Slice Thickness		
In plane resolution	1mm	2mm	3mm
0.35mm x 0.35mm	0.87±0.03	0.85±0.04	0.73±0.05
0.70mm x 0.70mm	0.68±0.02	0.61±0.04	0.52±0.06
1.40mm x 1.40mm	0.53±0.06	0.43±0.03	0.43±0.09

Table 6.2.1: Average AUC values for low resolution observer experiments. Each table shows a different texture analysis method.

In conclusion, for an observer based on autocorrelation texture analysis to be effective, the resolution of the images must be close to the size of the features of the texture. Improving the volumetric resolution will increase the observer’s ability to separate classes.

6.3 *In Vivo* MRI Parameters

Results from section 6.2 show a severe decline in observer performance as resolution decreases. The need for a diagnostically-acceptable SNR limits the resolution of a clinical MRI image along with breath hold time and other factors. An axial delayed phase, no GRAPPA sequence with $TR/TE/\alpha$ of 3.13ms/1.23ms/9° was used. The FOV was 38.0x29.8cm with an acquisition matrix of 352x276 pixels. The scan was taken over 32 slices with 3mm thickness. The in-plane resolution was $1.08 \times 1.08 \text{mm}^2$ and the scan time is approximately 20 seconds.

Normally an MR tech will select the ROI, which sets the number of pixels for the clinical scan. The mathematical observer is trained to sample features at the resolution of the images. If the observer is trained at one resolution and images are collected at a different resolution, either the image data or observer must be re-sampled to the limiting image set. It is desirable to avoid this situation; therefore, the images should be acquired with the same parameters for all patients. This will limit potential test subjects if their abdomen is larger than the ROI, which would result in aliasing artifacts. Despite these challenges, *in-vivo* images were collected from 4 patients. The gold standard for this study was the radiologist report. A biopsy for a healthy patient is an unnecessary procedure and is not possible to collect. Cirrhotic patients have confirmation from the written reports which also have patient history. While this is not ideal for an observer study, it is the best possible *in*-*vivo* data set we can currently collect.

6.4 *In Vivo* Imaging Results

Patients were either selected from previous examinations and volunteered for additional scanning, or the sequence was added to existing clinical examinations. We collect images from two patients where the report confirmed that the liver was healthy, and two patients whose reports confirmed cirrhosis. A selected slice from each patient is shown in Figure 6.4.1. Figure 6.4.1 (A) and (B) constitute the signal absent group and Figure 6.4.2 (A) and (B) constitute the signal present group.

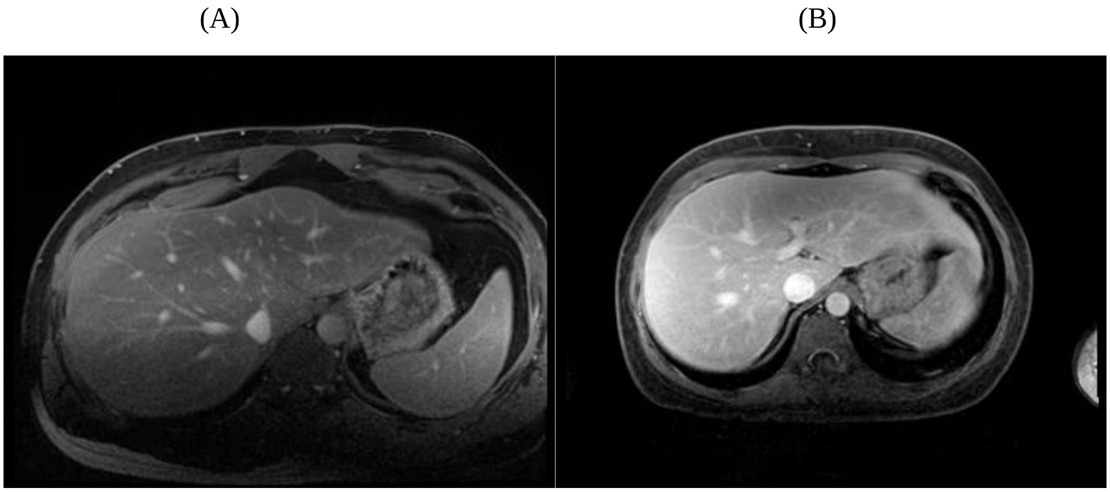


Figure 6.4.1: (A) and (B) are MRI slices from patients that were diagnosed with healthy liver tissue.

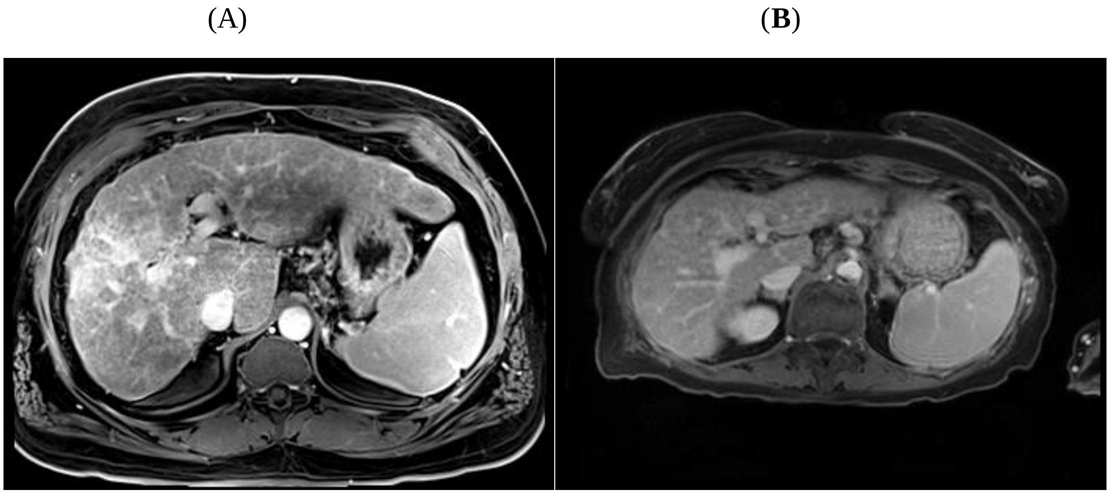


Figure 6.4.2: (A) and (B) are MRI slice images from patients diagnosed as having cirrhosis or chronic liver disease

6.5 Training the Observer

In vivo contrast enhanced MRI results in images with signal due to contrast agent in the vasculature and ECM. Healthy tissues in Figure 6.4.1 have contrast agent in the vasculature. Cirrhotic tissues in Figure 6.4.2 have contrast agent in the ECM of the cirrhotic livers, and have contrast agent in the blood vessels. The template is based on a change in texture of the liver, and boundaries between vessels and tissue will lead to confusion in the observer. Manual segmentation was used to remove the vessels from each image slice before training the observer. This method requires training by a radiologist. It is a successful method for initial experiments to show the capabilities of local texture analysis of HF *in vivo*, but an automated method will be required for computer-aided diagnosis.

The result of manual segmentation for one slice in each patient data set is shown in Figure 6.5.1. With the segmented livers, the same process of training the observer as described in Section 3.5 to 3.6 was used to train the observer for the *in vivo* images. After segmentation, the gridded ROI selection was used to measure the means and covariance matrices of the signal absent and signal present data for a 7x7 pixel 2D circular autocorrelation. The resulting means are shown in Figure 6.5.2 and resulting covariance matrices in Figure 6.5.3. These data were used to compute the Hotelling observer for this system. The 1D representation is highlighted in Figure 6.5.4. With a trained observer, independent testing data are used to validate the observer.

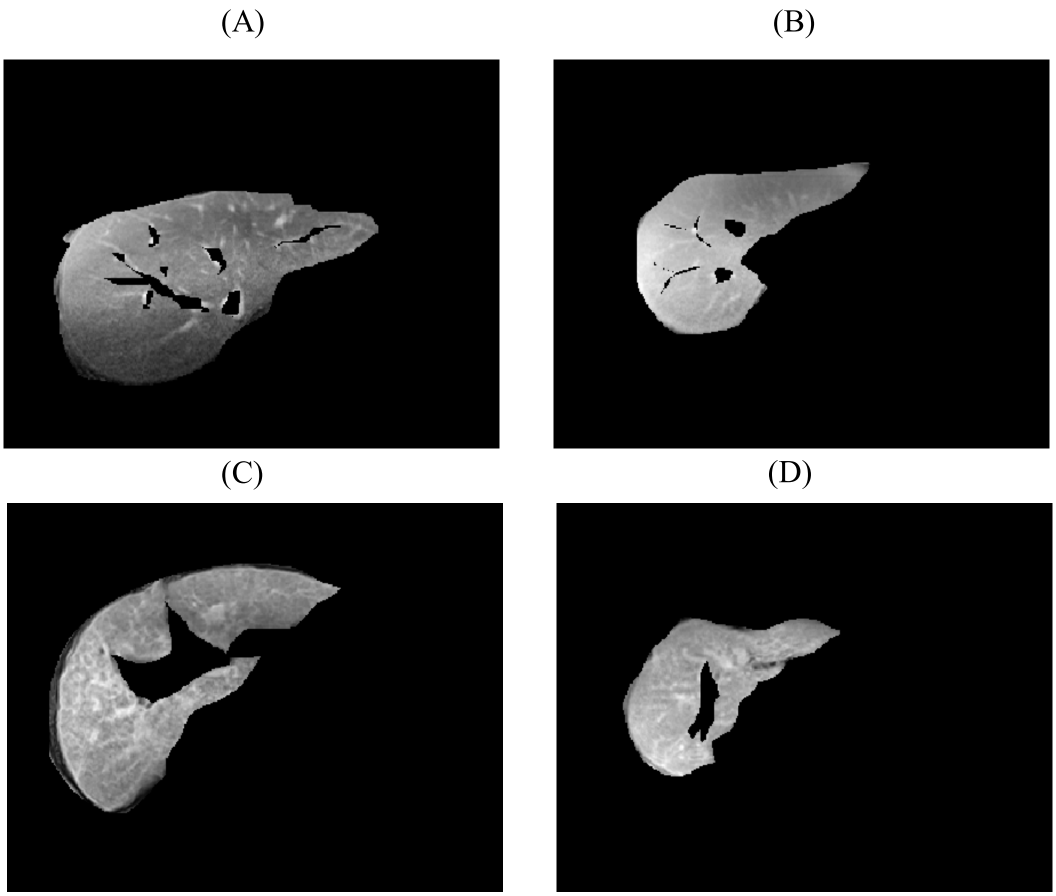


Figure 6.5.1: Results of a manual segmentation to remove vasculature from four patient data sets

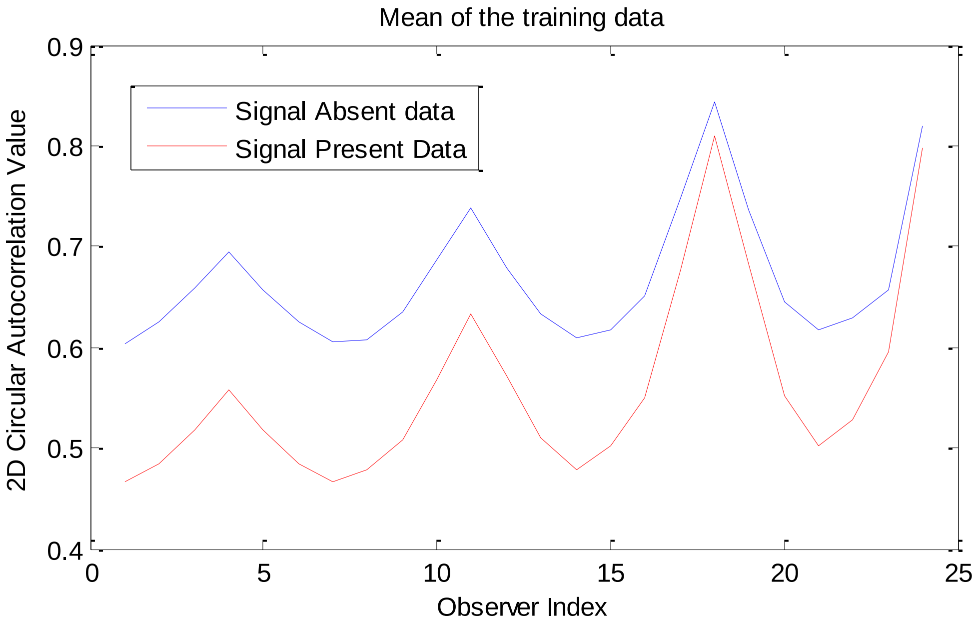


Figure 6.5.2: Signal absent and signal present means for a 2D circular autocorrelation local texture analysis

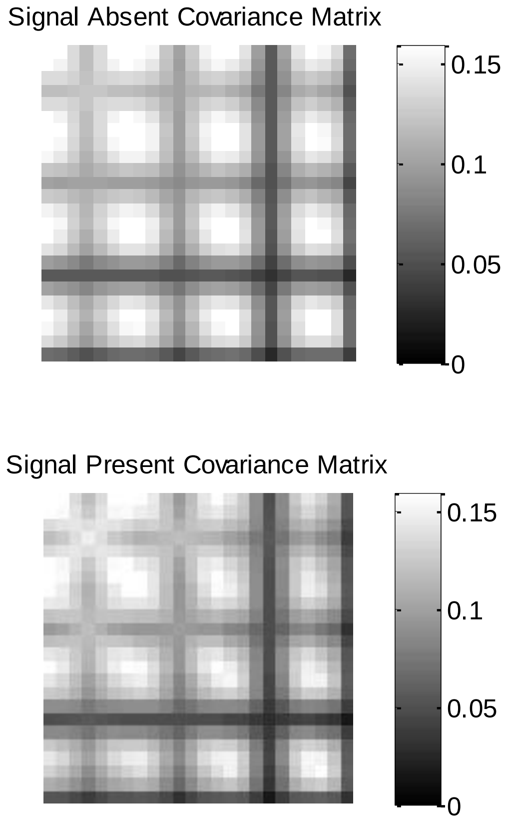


Figure 6.5.3: Signal absent and signal present covariance matrices

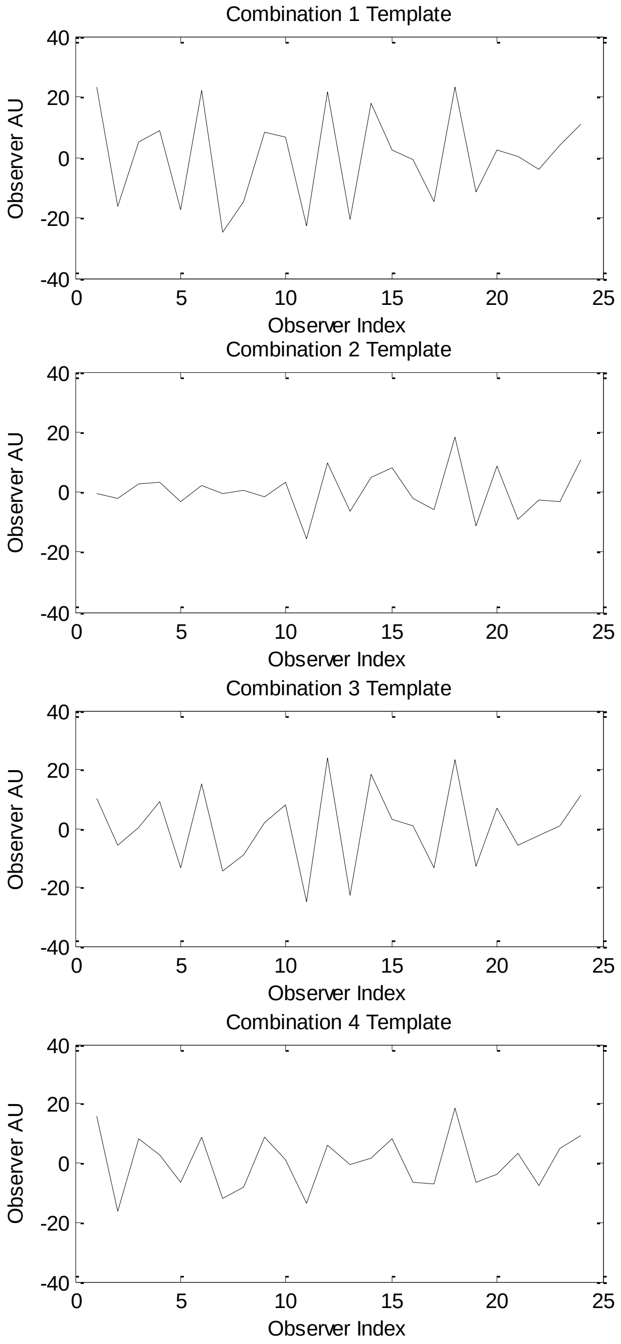


Figure 6.5.4: 1D representation of the Hotelling observer based on a local 2D circular autocorrelation texture analysis of *in vivo* data

6.6 Testing the Observer with *In Vivo* Data

With four livers, four combinations of independent training and testing data were possible. One liver from each class is responsible for either training or testing. The templates were applied to testing data and four ROC curves were recovered. Figure 6.7.1 shows the results from all four combinations along with an average ROC curve.

We also created test statistic maps for these data using the method outlined in section 3.9. Figure 6.7.2, Figure 6.7.3, Figure 6.7.4 and Figure 6.7.5 show the test statistic maps from each combination of training and testing data.

6.7 Discussion of *In Vivo* Results

The AUC values for the *in vivo* results are lower than the phantom experiments. There are multiple factors that impact this result. First, the segmentation is a very difficult task to perform by hand. If segmentation is poor, the observer will train on undesired features, such as blood vessels.

The test statistic maps still may be of use to a human reader, but are expected to improve further with additional training data.

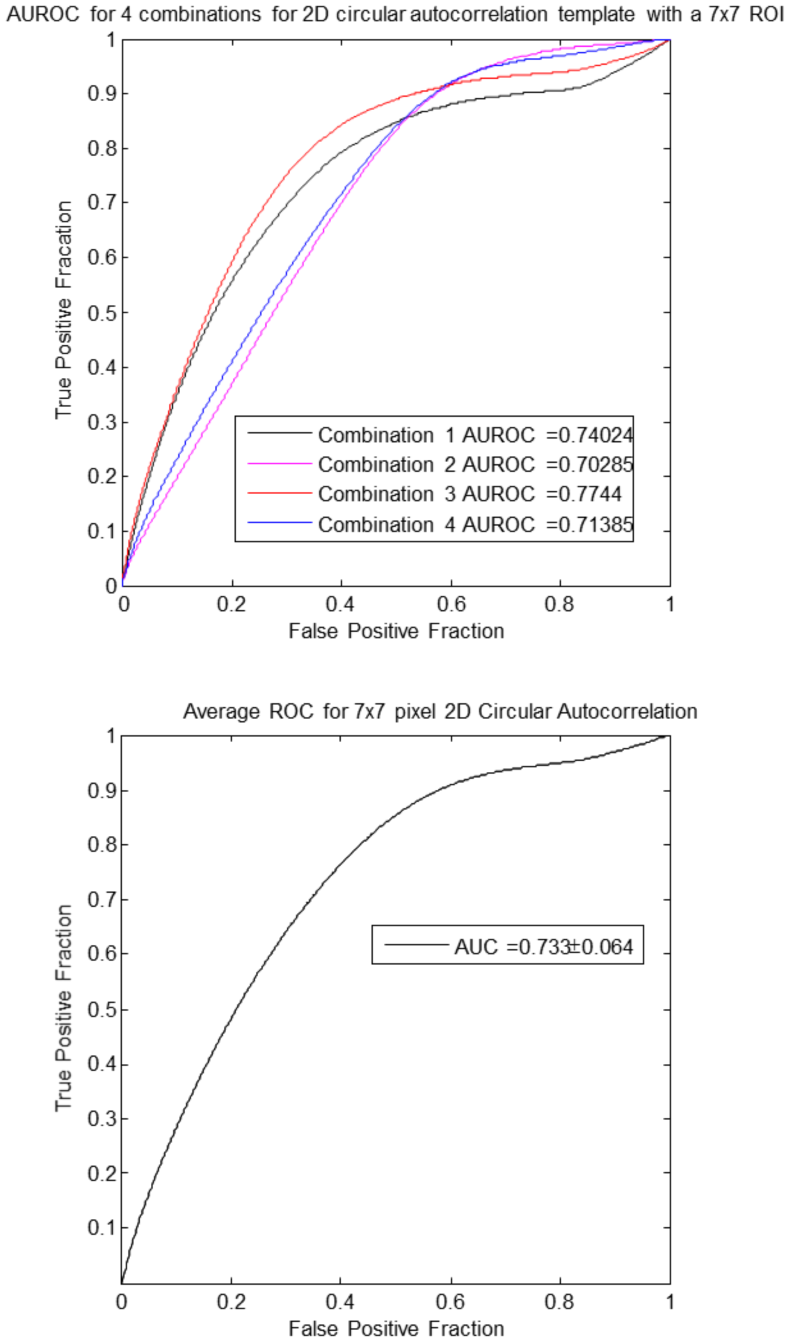


Figure 6.7.1: ROC results for individual trials and combined

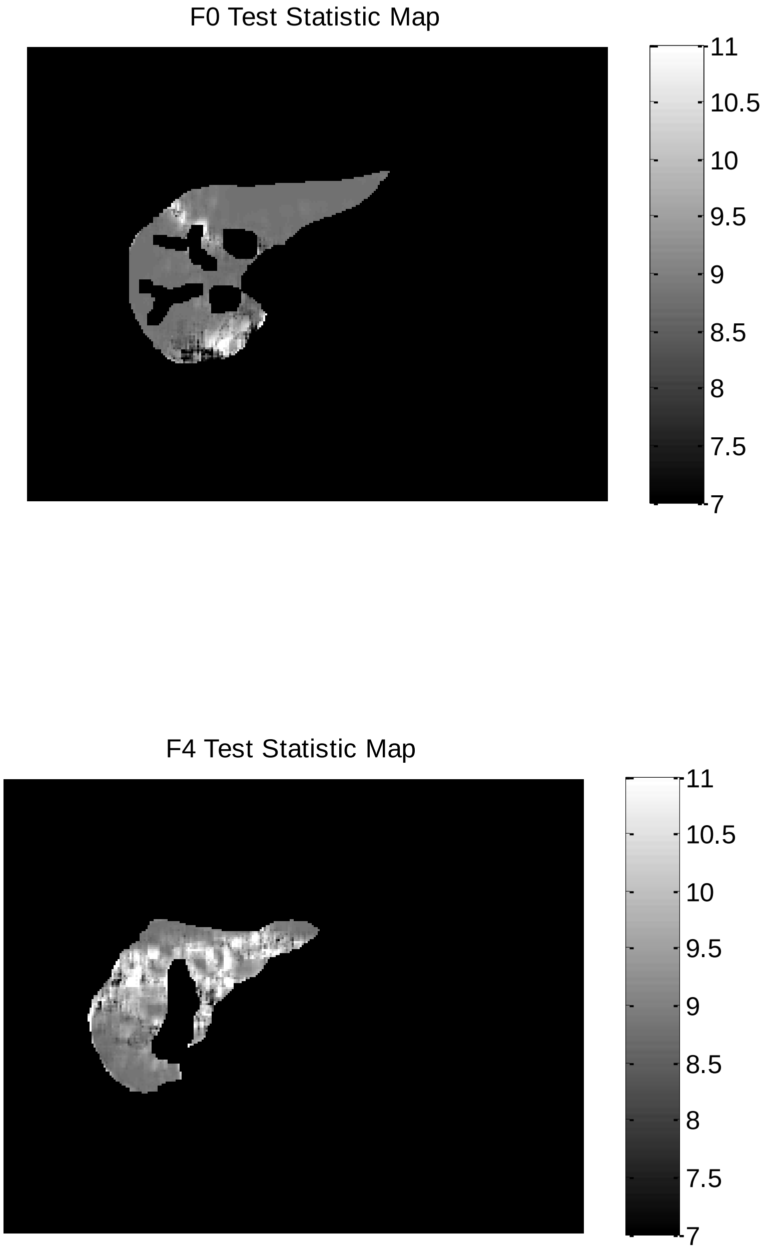
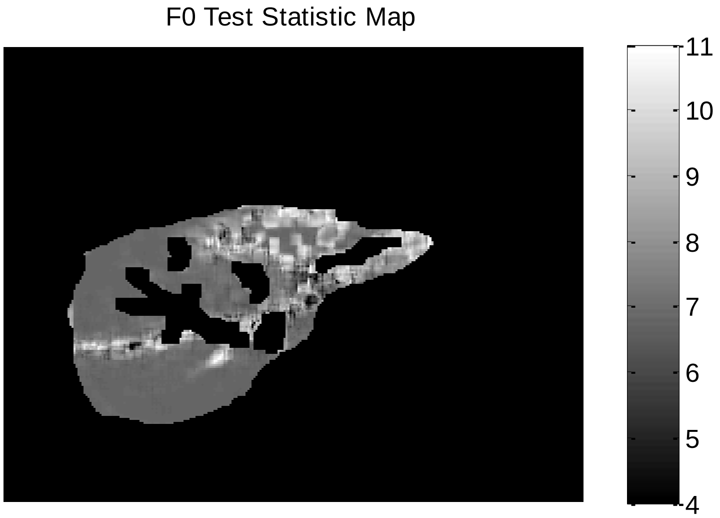
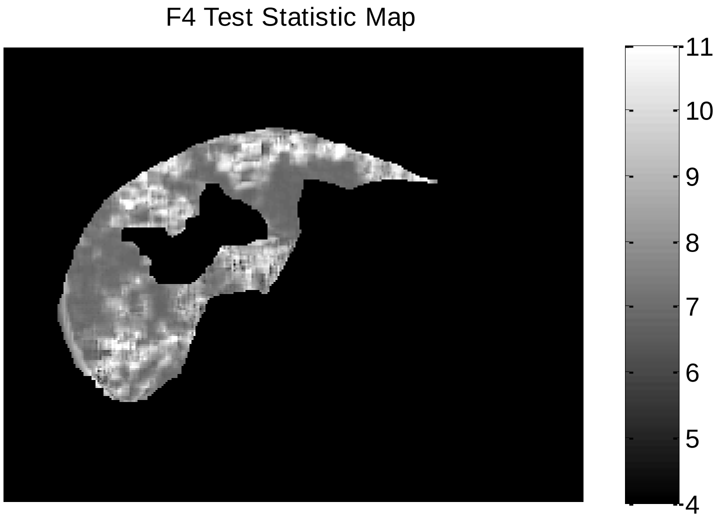


Figure 6.7.2: Test statistic maps for combination 1

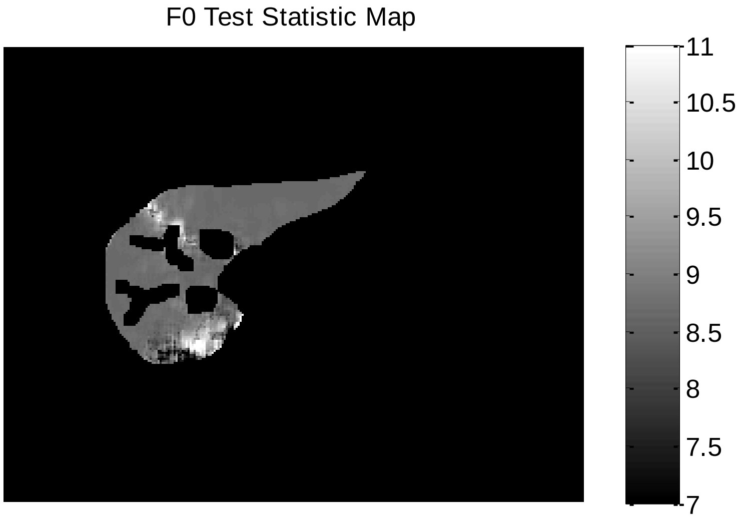


F0 Test Statistic Map



F4 Test Statistic Map

Figure 6.7.3: Test statistic maps for combination 2



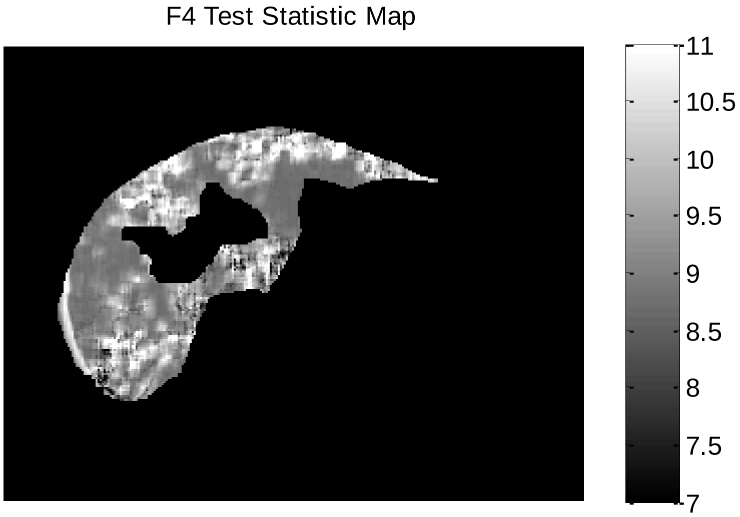
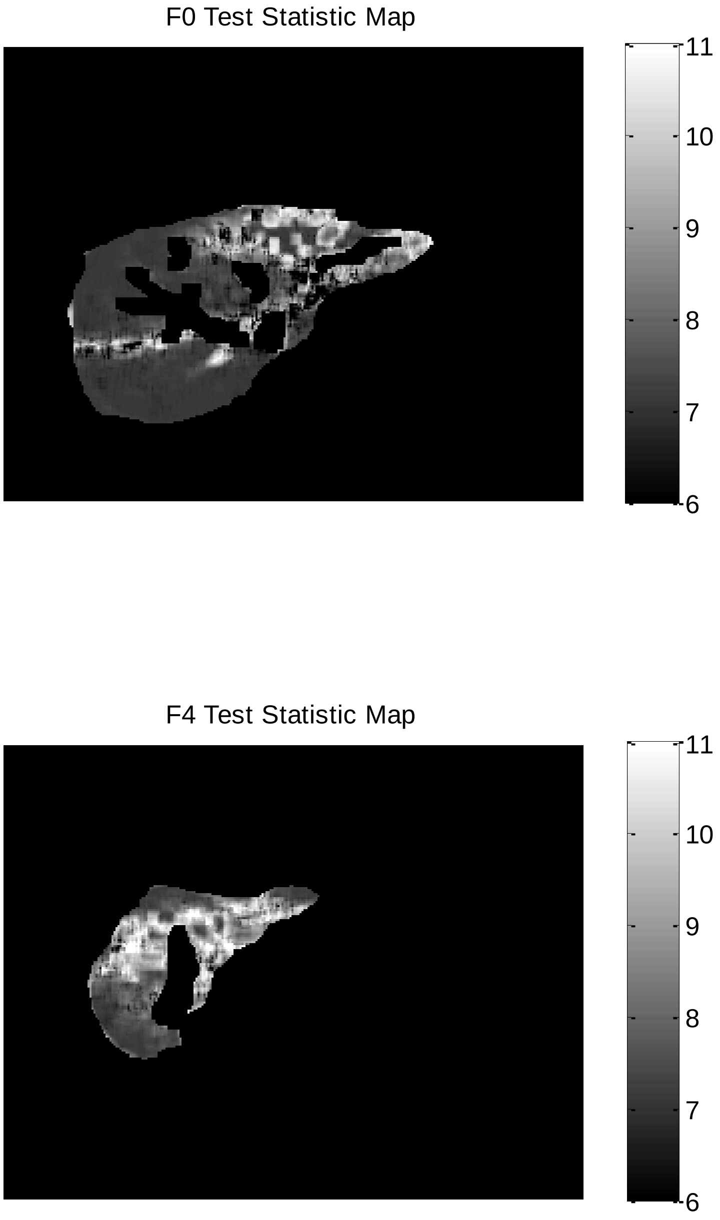


Figure 6.7.4 Test statistic maps for combination 3


F0 Test Statistic Map

F4 Test Statistic Map

Figure 6.7.5 Test statistic maps for combination 4

6.8 Training and Testing a Two-Stage Observer with *In Vivo* Data

Inspired by the initial success of the phantom results, the two stage observer method from Chapter 5 was applied to the same *in-vivo* data in the previous section. To train the two observers, the 3D image training data was divided into two groups. The first half of the data was used to train the first stage observer and was tested on the second half of the image set. The results of the testing of the first stage observer were used to train the second stage channelized observer. As before, the channel boundaries were selected manually. The first and second stage observers are shown in Figure 6.8.1 for all four combinations.

The trained templates were applied to the independent testing data in each of the four combinations. For each MRI slice, a scalar test statistic is recovered and plotted with respect to the fibrosis stage of the liver. Figure 6.8.2 shows the results for the four combinations. The minimum AUC value is 0.931 and two combinations achieve almost perfect separation. This is a strong suggestion that the methods outlined in Chapter 5 can have success in clinical applications.

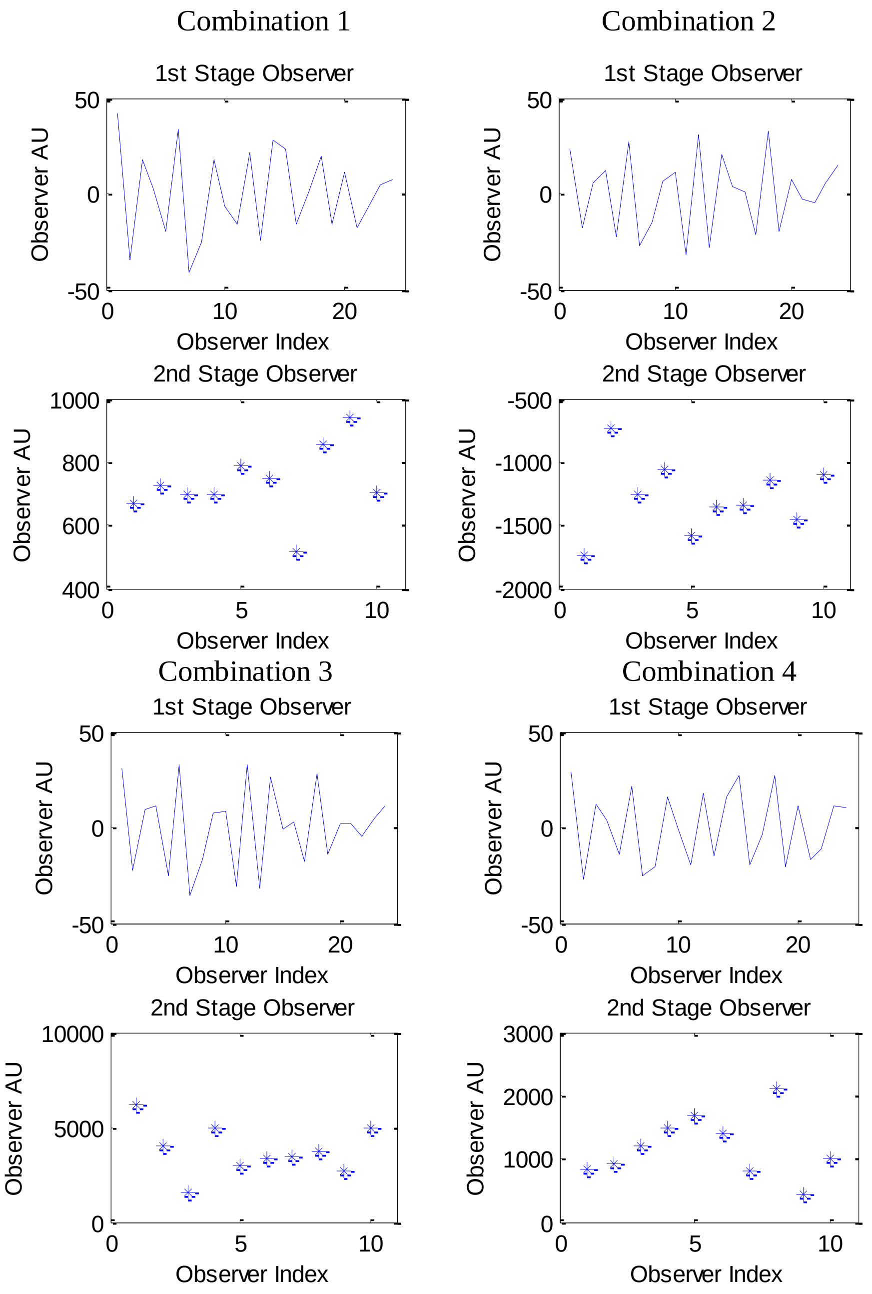


Figure 6.8.1: First and second stage Hotelling observers recovered for four combinations of *in vivo* testing and training data

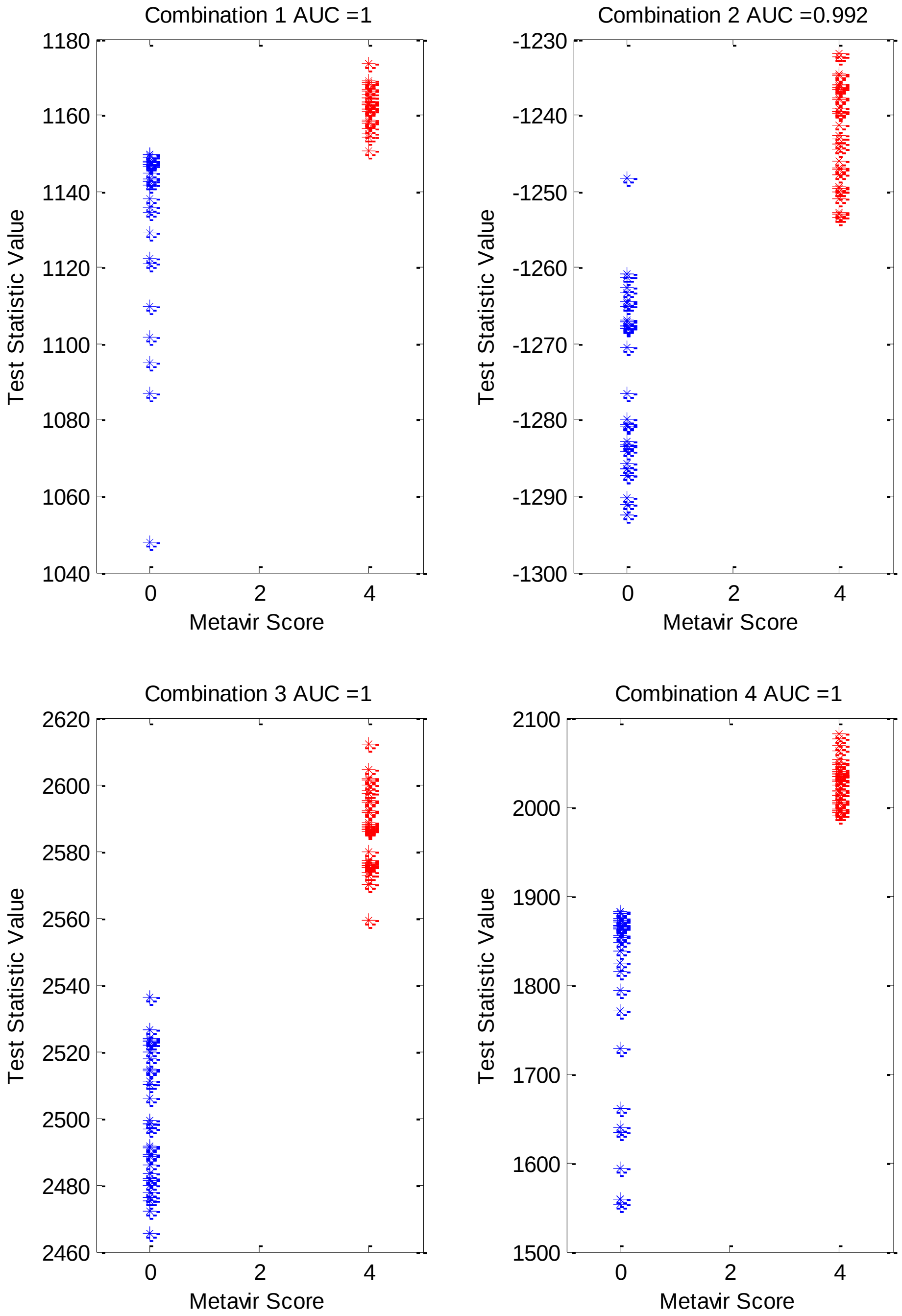


Figure 6.8.2: Two-stage observer results for four combinations of *in vivo* testing and training data with their AUC values

6.9 Conclusions for Chapter 6

Completing the translation of these techniques from phantoms to clinical application requires more data to train and test observer methods. These first experiments were completed with limited training and testing data, and compromised were made concerning how to handle the training and testing of the observers.

With that consideration, these observers performed at a high success rate with *in vivo* data. The local texture analysis observer performance decreased relative to the phantoms and the two stage observer performance decreased slightly. Two factors impacting these results are the decrease in image resolution achievable in *in vivo* images and imperfect segmentation. Lower resolution reduces sensitivities to the presence of fibrotic content, while poor segmentation leaves textured data in healthy tissue as diseased tissue that hurts specificity. Improved segmentation will increase the effectiveness of these observers. We conclude, the experiments performed in phantom studies in Chapters 3 through 5 are worth repeating in a larger *in vivo* study.

7. Conclusions

7.1 Summary of Results

In this dissertation, a mathematical observer technique was explored as a method to stage hepatic fibrosis in MRI images. MRI is emerging as a tool to diagnose chronic liver disease, with MRE currently in use as a an FDA-approved method to detect change in liver stiffness. However, MRE is only sensitive to the overall stiffness, which may not be sensitive to the presence of an early ECM. Radiologists who read MR images would like to detect the ECM to perform the same task pathologists carry out when working with biopsy slides.

Both radiologists and pathologists are attempting to detect a change in texture. In this dissertation, a tool was developed and used to train a linear Hotelling observer. Experiments in texture analysis were carried out with phantoms that were found to replicate the contrast of *in vivo* abdominal images. The phantom used is formalin fixed liver tissue taken from autopsy. A gradient echo T1 weighted sequence was used and images were analyzed with local texture analysis.

Local texture analysis performed by a 2D circular autocorrelation and wavelet analysis were found to be effective methods to detect the presence of HF. Both methods separate the early and late stage diseases effectively. The 2D circular autocorrelation has the smallest training data requirement, which may be beneficial to those who work with limited data sets. Researchers must balance the training requirements for sample covariance matrices to achieve texture analysis methods that are effective enough for a radiologist to make a reliable call.

We used task-based performance evaluation to select an optimal MRI parameter by repeating measurements of the objects multiple times while varying the acquisition parameter, the FA. MRI offers many additional parameters that can be optimized to increase the performance of clinical tasks. Combining quantitative image analysis with optimization will result in better MRI sequences, as these techniques move into clinical applications.

Test-statistic maps are introduced as a means to visualize the result of an observer assessing local texture. Our images correlate fibrotic structure with spatial location, allowing a radiologist to locate possible areas of fibrosis quickly. These results were validated through the use of a two stage observer.

The two-stage observer histograms the test statistic values from the local texture analysis into channels for a second observer to make a decision about an entire slice in an MRI image. This is a powerful tool to evaluate a large area of tissue and will be what is necessary to detect *F2* and *F3* HF.

In the final part of this dissertation we applied all of the techniques we tested on phantoms on *in vivo* data. A minimal acceptable resolution was determined and an MRI sequence was added to clinical examinations to collect patient data. The performance of the local texture analysis observer decreased when *in vivo* data was used.

Nevertheless, when the two stage observer was applied to *in vivo* data, the AUC of the experiment ranged from 0.996 to 1.0. This level of separation supports our technique, but is also a sign that separation of *F0* and F4 is much easier than finding early stage disease. While it will be beneficial to collect more data the number of eligible patients is low and no additional *in vivo* studies are available at the conclusion of this dissertation.

7.2 Future Work

The continuation of the translation of these techniques from phantom experiments to clinical techniques should be the main goal for the next step in this project. A major drawback of the observer technique is the fact that resolution between patients can change if a consistent sequence implementation is not adhered to by the MR technicians. Scans that will use texture analysis need the best possible resolution. MRI clinics will have to train the technicians and incorporate appropriate sequences in order to build on these techniques.

Additionally, investigation into an acceptable gold standard for identifying healthy patients is necessary, since collecting a biopsy is not possible from a healthy volunteer. One suggestion is to attach a liver MRI at delayed phase to patients undergoing kidney scans. Another possible gold standard is to use young patients, under 40, who are identified as relatively healthy. A specially trained radiologist could, perhaps, confirm these findings with an evaluation of the liver MRIs.

Higher resolution MR sequences will only improve the texture analysis techniques, providing samples of the ECM in early and mid-stages of hepatic fibrosis.

Chapter 4 laid out a procedure for finding optimal MRI settings in phantoms. This can be expanded to clinical practice with the help of volunteers. With injections, scans at various delay times can determine the optimal time delay after Gd injection. While testing for optimal delay, scans can have various FA's. One could plot AUC as a function of FA and time after injection. This would lead to an improved protocol for task-based classification of HF.

Much of the continuing work involves identifying and collecting images from patient volunteers. Another crucial step to improve the data analysis is automated effective segmentation of major vessels and their removal from the liver images. Without this tool, the technique is dependent on manual segmentation, a time-consuming process subject to user variability.

Works Cited

1. O. College, *Anatomy & Physiology*, Rice University 2011, Houston, TX.
2. W. F. Boron and E. L. Boulpaep, *Medical physiology: a cellular and molecular approach*, Saunders/Elsevier, Philadelphia, PA (2009).
3. National Center for Health Statistics, "FastStats," Leading Causes of Death, 2013, <http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>.
4. R. Bataller and D. Brenner, "Liver fibrosis," J. Clin. Invest. **115**(2), 209–218 (2005).
5. J. Jiao, S. L. Friedman, and C. Aloman, "Hepatic fibrosis.," Curr. Opin. Gastroenterol. **25**, 223–229 (2009) [doi:10.1097/MOG.0b013e3283279668].
6. A. Regev et al., "Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection.," Am. J. Gastroenterol. **97**(10), 2614–2618 (2002).
7. L. J and V. Tr, "Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease.," Mayo Clin. Proc. **55**(7), 434–438 (1980).
8. B. Coco et al., "Transient elastography: A new surrogate marker of liver fibrosis influenced by major changes of transaminases," J. Viral Hepat. **14**(5), 360–369 (2007).

9. H. Juergen Nord, “Biopsy diagnosis of cirrhosis: blind percutaneous versus guided direct vision techniques—a review,” Gastrointest. Endosc. **28**(2), 102–104, Elsevier (1982).

10. L. B. Seeff et al., “Complication Rate of Percutaneous Liver Biopsies Among Persons With Advanced Chronic Liver Disease in the HALT-C Trial,” Clin. Gastroenterol. Hepatol. **8**(10), 877–883, Elsevier Inc. (2010).

11. T. F. M. C. S. Group, “Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group.,” Hepatology **20**(1 Pt 1), 15–20 (1994) .

12. R. A. Standish et al., “An appraisal of the histopathological assessment of liver fibrosis.,” Gut **55**(4), 569–578 (2006) .

13. J. a. Carrión et al., “Efficacy of Antiviral Therapy on Hepatitis C Recurrence After Liver Transplantation: A Randomized Controlled Study,” Gastroenterology **132**(5), 1746–1756 (2007).

14. D. R. Martin et al., “Liver MRI and histological correlates in chronic liver disease on multiphase gadolinium-enhanced 3D gradient echo imaging,” J. Magn. Reson. Imaging **36**(2), 422–429 (2012).

15. P. Marcellin et al., “Non-invasive assessment of liver fibrosis by stiffness measurement in patients with chronic hepatitis B.,” Liver Int. **29**(2), 242–247 (2009).

16. J. A. Talwalkar et al., “Magnetic resonance imaging of hepatic fibrosis: emerging clinical applications.,” Hepatology **47**(1), 332–342 (2008).

17. D. R. Martin et al., “Reversible heterogeneous arterial phase liver perfusion associated with transient acute hepatitis: Findings on gadolinium-enhanced MRI,” J. Magn. Reson. Imaging **20**(5), 838–842 (2004).

18. S. M. Martinez et al., “Longitudinal Liver Stiffness Assessment in Patients with Chronic Hepatitis C Undergoing Antiviral Therapy,” PLoS One **7**(10), e47715 (2012).

19. F. Degos et al., “Diagnostic accuracy of FibroScan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: A multicenter prospective study (the FIBROSTIC study),” J. Hepatol. **53**(6), 1013–1021 (2010).

20. Z. D. Goodman, “Grading and staging systems for inflammation and fibrosis in chronic liver diseases,” J. Hepatol. **47**(4), 598–607 (2007).

21. L. Huwart et al., “Magnetic Resonance Elastography for the Noninvasive Staging of Liver Fibrosis,” Gastroenterology **135**(1), 32–40 (2008).

22. S. L. Friedman, “Liver fibrosis -- from bench to bedside.,” J. Hepatol. **38 Suppl 1**, S38–S53 (2003).

23. M. Pinzani, K. Rombouts, and S. Colagrande, “Fibrosis in chronic liver diseases: diagnosis and management,” J. Hepatol. **42**(1), 22–36 (2005).

24. H. Yoshiji, S. Kuriyama, and H. Fukui, “Angiotensin-I-converting enzyme inhibitors may be an alternative anti-angiogenic strategy in the treatment of liver fibrosis and hepatocellular carcinoma. Possible role of vascular endothelial growth factor.,” Tumour Biol. **23**(6), 348–356 (2002) .

25. J. G. McHutchison et al., “Measurement of serum hyaluronic acid in patients with chronic hepatitis C and its relationship to liver histology,” J. Gastroenterol. Hepatol. **15**(8), 945–951 (2000).

26. P. Marcellin et al., “Regression of cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: A 5-year open-label follow-up study,” Lancet **381**(9865), 468–475 (2013) [doi:10.1016/S0140-6736(12)61425-1].

27. P. Muriel and V. Castro, “Effects of S-adenosyl-L-methionine and interferon-alpha2b on liver damage induced by bile duct ligation in rats,” J Appl Toxicol **18**(2), 143–147 (1998).

28. P. Bedossa et al., “Transforming growth factor beta 1: in situ expression in the liver of patients with chronic hepatitis C treated with alpha interferon,” Gut **34**(0017-5749), S146–S147 (1993).

29. C. Raetsch et al., “Pentoxifylline downregulates profibrogenic cytokines and procollagen I expression in rat secondary biliary fibrosis,” Gut **50**(2), 241–247 (2002).

30. R. G. Romanelli et al., “Effect of pentoxifylline on the degradation of procollagen type I produced by human hepatic stellate cells in response to transforming growth factor-beta 1.,” Br. J. Pharmacol. **122**(6), 1047–1054 (1997).

31. M. NessAiver, *All you really need to know about MRI Physics*, Simply Physics, Baltimore, MD (1997).

32. D. J. Griffiths, *Introduction to Quantum Mechanics*, in Quantum **1**, Pearson Prentice Hall, Upper Saddle River, NJ (2005).

33. J. L. Prince and J. M. Links, *Medical Imaging Signals and Systems*, 2nd ed., Pearson Education, 2014 (2014).

34. A. Bilgin et al., “Parallel magnetic resonance imaging using compressed sensing,” Proc. SPIE **7073**, 70731G – 70731G – 10 (2008) [doi:10.1117/12.797206].

35. G. McGibney et al., “Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI.,” Magn. Reson. Med. **30**(1), 51–59 (1993).

36. R. Chartrand, “Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data,” in Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, pp. 262–265 (2009).

37. S. C. Faria et al., “MR imaging of liver fibrosis: current state of the art.,” Radiographics **29**(6), 1615–1635 (2009).

38. D. R. Martin and R. C. Semelka, “Magnetic resonance imaging of the liver: Review of techniques and approach to common diseases,” Semin. Ultrasound, CT MRI **26**(3), 116–131 (2005).

39. R. C. Semelka et al., “Chronic hepatitis: Correlation of early patchy and late linear enhancement patterns on gadolinium-enhanced MR images with histopathology initial experience,” J. Magn. Reson. Imaging **13**(3), 385–391 (2001).

40. D. A. Aguirre et al., “Liver fibrosis: noninvasive diagnosis with double contrast material-enhanced MR imaging.,” Radiology **239**(2), 425–437 (2006).

41. M. G. Ghany and E. Doo, “Assessment of liver fibrosis: Palpate, poke or pulse?,” Hepatology **42**(4), 759–761 (2005).

42. M. Yin et al., “Assessment of hepatic fibrosis with magnetic resonance elastography.,” Clin. Gastroenterol. Hepatol. **5**(10), 1207–1213.e2 (2007).

43. D. Kim et al., “Advanced fibrosis in nonalcoholic fatty liver disease: noninvasive assessment with MR elastography.,” Radiology **268**(2), 411–419 (2013).

44. S. Singh et al., “Diagnostic Performance of Magnetic Resonance Elastography in Staging Liver Fibrosis: A Systematic Review and Meta-analysis of Individual Participant Data,” Clin. Gastroenterol. Hepatol. **13**(3), 440–451, Elsevier Ltd (2015).

45. S. K. Venkatesh et al., “Correlation of MR elastography with morphometric quantification of liver fibrosis (Fibro-C-Index) in chronic hepatitis B,” in Magnetic Resonance in Medicine **1129**, pp. 1123–1129 (2013).

46. S. K. Venkatesh et al., “Non-invasive detection of liver fibrosis: MR imaging features vs. MR elastography,” Abdom. Imaging **40**(4), 766–775 (2015).

47. J. Braun, K. Braun, and I. Sack, “Electromagnetic actuator for generating variably oriented shear waves in MR elastography.,” Magn Reson Med **50**(1), 220–222 (2003) .

48. L. Castéra et al., “Pitfalls of liver stiffness measurement: A 5-year prospective study of 13,369 examinations,” Hepatology **51**(3), 828–835 (2010).

49. O. Rouvière et al., “MR elastography of the liver: preliminary results.,” Radiology **240**(2), 440–448 (2006).

50. H. H. Barrett and K. J. Myers, *Foundations of Image Science*, Wiley-Interscience, Hoboken, NJ (2003).

51. H. H. Barrett et al., “Model observers for assessment of image quality.,” Proc. Natl. Acad. Sci. U. S. A. **90**(21), 9758–9765 (1993).

52. X. He and S. Park, “Model observers in medical imaging research,” Theranostics **3**(10), 774–786 (2013).

53. R. N. Strickland, *Image-Processing Techniques for Tumor Detection*, Marcel Dekker, New York, NY (2002).

54. M. Kupinski and E. Clarkson, “A New Method for Optimizing Channelized Quadratic Observers for Binary Classification of Large-Dimensional Image Datasets,” J. Opt. Soc. Am. A **x**(4), 1–26 (2015).

55. A. E. Burgess, “Mammographic structure: data preparation and spatial statistics analysis,” SPIE Conf. Image Sciennce Proc. **3661**(February), 642–653 (1999).

56. A. E. Burgess, “Bach , breasts , and power-law processes,” SPIE Image Percept. Perform. Proc. **4324**(617), 103–113 (2001).

57. K. G. Metheany et al., “Characterizing anatomical variability in breast CT images.,” Med. Phys. **35**(10), 4685–4694 (2008).

58. R. F. Wagner, M. F. Insana, and D. G. Brown, “Unified approach to the detection and classification of speckle texture in diagnostic ultrasound,” Opt. Eng. **25**(6), 738–742 (1986).

59. M. Insana and R. Wagner, “Analysis of ultrasound image texture via generalized Rician statistics,” Opt. Eng. **25**(6), 743–748 (1986).

60. H. H. Barrett, C. K. Abbey, and B. D. Gallas, “Stabilized estimates of hotelling-observer detection performance in patient-structured noise,” SPIE Proc. **3340**, 27–43 (1998).

61. S. Naganawa et al., “Rapid MR imaging of the liver: comparison of twelve techniques for single breath-hold whole volume acquisition,” Radiat Med **12**(6), 255–261 (1994).