Summer 8-19-2022

# Multiscale Anisotropy Analysis of Second-Harmonic Generation Imaging of Pancreatic Cancer

Joshua D. Hamilton
*University of Maine*, joshua.hamilton@maine.edu

**MULTISCALE ANISOTROPY ANALYSIS OF SECOND-HARMONIC GENERATION IMAGING OF PANCREATIC CANCER**

By

Joshua Hamilton

B.S. University of Maine, 2021

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(In Biomedical Engineering)

The Graduate School

The University of Maine

August 2022

Advisory Committee:

Andre Khalil, Professor of Biomedical Engineering, Co-Advisor

Karissa Tilbury, Assistant Professor of Biomedical Engineering, Co-Advisor

Peter Brooks, Faculty Scientist at Maine Health Institute for Research

**MULTISCALE ANISOTROPY ANALYSIS OF SECOND-HARMONIC GENERATION IMAGING OF PANCREATIC CANCER**

By Joshua Hamilton

Thesis Advisors: Dr. Andre Khalil and Dr. Karissa Tilbury

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(In Biomedical Engineering)
August 2022

Despite recent advancements in biomedicine, cancer is still the second leading cause of death in the United States. Early detection of cancer is critical to improving patient care, but there are risks of over screening caused by the need for surgical biopsies in many cancers for final diagnostics. Recent advancements in computer aided diagnostics for breast cancer screening has reduced the need for biopsies and resulted in earlier diagnoses which has lowered the mortality rate from breast cancer within the past two decades. Developing new computer aided diagnostic tools that can be applied to a vast majority of cancers would serve to improve quality of life worldwide. These tools could also help researchers target and understand biological markers that lead to more malignant cancers improving both our treatment and understanding of cancer progression. The novel combination of the label-free, collagen-specific microscopy technique known as second harmonic generation (SHG) and the 2D Wavelet Transform Modulus Maxima (2D WTMM) Anisotropy Method is a prime candidate to serve this role. The 2D WTMM Anisotropy Method, originally developed for galactic astronomy and then used in multiple biological studies, was further adapted for SHG imaging of cancer in this work by improving both the binning and normalization techniques. This improved method was then applied to forty slides from pancreatic ductal adenocarcinoma (PDAC) patients and eight images were captured per each tissue category on each slide. Cancer and fibrosis had greater anisotropy factors (Fa) at small wavelet scales

than normal and normal adjacent tissue. At scales larger than 21 μm this relationship changed with normal tissue having a higher Fa than all other tissue groups. This demonstrated that our developed method is sensitive to changes induced by PDAC. Our method was also compared to other open source SHG image analysis tools currently used by researchers in the field by generating 100 simulated fiber images at four different angle distributions of 0-180°, 30-150°, 60-120°, and 85-95°. The 2D WTMM Anisotropy method could differentiate the 0-180° and 30-150° groups at multiple scales whereas off-the-shelf tools could not. Four different levels of white noise were also added to the 60-120° angle distributions images to test each methods sensitivity to noise by comparing each noise convolved fiber image to pure white noise. The 2D WTMM Anisotropy Method was the only method capable of differentiating all added noise levels from white noise demonstrating its superior resistance to noise. This method will soon be applied to a larger breast cancer study and a breast cancer spheroid study. In both cases further developments to the method are planned, such as developing a version capable of analyzing 3D images and coupling the method with a machine learning technique.

**DEDICATION**

I dedicate this thesis to my stepfather, James Dunning Garland, who was taken from our family too soon by pancreatic cancer at the age of 52. He pushed me to pursue higher education and was there every morning to say I love you before I left for school at the University of Maine. I hope this work will be a step toward a world where no family has to suffer through the pain and grief that mine has suffered due to cancer.

**ACKNOWLEDGEMENTS**

I would first like to thank Dr. Karissa Tilbury who has been the biggest contributor in me having the resources and guidance needed to succeed in my research at the University of Maine, since she first mentored me in my honors thesis. Similarly, I would like to thank Dr. Andre Khalil for reaching out to me and giving me a project we both care deeply about. Without his mentorship I would not have found my passion for applying coding to real world biomedical problems. Dr. Peter Brooks deserves thanks for assisting this work by supplying his biological expertise and helping us with finding samples. I would like to thank all members of the CompuMAINE lab for providing helpful feedback and ideas along the way. I would like to thank my mother, Shelly Spear-Hamilton, for being the best support structure I could ever hope for and helping me in my times of need. My partner, Isabel Larsen, also deserves great recognition for always being there for me day in and day out. Lastly, I would like to thank any friends or family members who have spent any time simply just opening up and communicating to me. The most important thing in life is building positive relationships with those around us and I appreciate everyone who has done so with me.

**TABLE OF CONTENTS**

**LIST OF TABLES**

**LIST OF FIGURES**

**List OF Equations**

# CHAPTER 1

# INTRODUCTION

## Motivation

Current strategies for cancer diagnostics still rely heavily on surgical biopsies aided by tissue specific imaging modalities, if any [1-3]. In contrast to most cancers, the clinical pipeline for breast cancer diagnosis has recently introduced computer aided diagnostics (CAD) of mammography images which has coincided with a steady decline in mortality from the disease through the 90s and 2000s [4-6]. While CAD mammography does not directly cure breast cancer, it improves patient survival as early diagnosis results in early treatment which improves treatment success rates and minimizes the chance for metastasis [7-8]. In all cancers, this repeated screening has to be balanced with the invasiveness the screening requires. This invasive screening is the core problem in solid organ cancers, such as pancreatic or prostate cancer, which are extremely deadly with minimal symptoms before malignant growth [1, 3]. The current clinically used imaging modalities in solid organ cancers cannot differentiate cancerous growths from benign and can completely miss smaller growths resulting in the only effective screening method being a surgical biopsy [1, 3]. Surgical biopsy of benign growths is unnecessary and can harm patients both physically and monetarily [8-10]. Every excess or unnecessary surgical procedure should be avoided to minimize possibility of patient harm and in the end, patients are screened less for solid organ cancers than breast or colon cancer [11].

Despite the previously mentioned improvements in breast cancer screening and diagnostics it is still the second leading causes of cancer death in the United States [12]. This is due to some patients displaying a more aggressive cancer that exhibits both chemo and immunotherapy resistance [13]. This resistance is an active area of research and has been traced to genetic variations in what are known as

cancer associated fibroblasts (CAFs) [13-15]. These fibroblast cells are activated by and interact with the cancer cells causing overproduction of ECM biomolecules and more specifically, collagen [13-14]. A patient's resistance to therapy can be predicted using these variations and more fibrotic cancer can be traced to these variations [13-14]. More fibrotic cancers' directly correlate with worse patient prognosis preventing T-cell penetration and promoting tumor expansion [14-16]. This relationship between the patients' fibroblasts, the fibrosis, and the cancer cells is still largely unknown, but it can be hypothesized that the tumor microenvironment's collagen structure is becoming protumorigenic through interactions with the cancer cells and CAFs. Interestingly, in pancreatic ductal adenocarcinoma (PDAC) mouse models with collagen production by these CAFs limited, the cancer was deadlier and faster [15]. This signifies that the relationship is more complicated than fibrosis simply being a bad sign in the tumor microenvironment which leads to two overarching questions. The questions are if we can diagnose malignant growths earlier through changes in the microenvironment, and can we control these changes to be a non-protumorigenic microenvironment? Answering these questions first requires the development of a new tool that focuses on quantitatively understanding the tumor microenvironment which would assist research in both improving our understanding and diagnostic capabilities. This thesis proposes the novel combination of two-photon Second Harmonic Generation (SHG) microscopy and the 2D Wavelet Transform Modulus Maxima (2D WTMM) Anisotropy Method further described in this chapter as a tool to begin developing solutions to these problems.

**Second Harmonic Generation**

A microscopy tool already exists that is capable of label-free, collagen specific imaging allowing one to capture the information about the tumor microenvironment without manipulating or staining tissue [17-20]. This tool is a two-photon microscopy technique known as SHG that uses infrared pulsed lasers to elicit a non-linear response in the tissue allowing for a plethora of advantages compared to

other imaging techniques, such as the aforementioned label-free collagen imaging [17-20]. The typical excitation wavelength used for collagen is in the IR range and capable of penetrating tissue upwards of 500 µm [21]. This process works through excitation of a non-centrosymmetric structure at double the typical excitation wavelength by having two photons interact at the same moment, hence the name second harmonic (Figure 1). This is an improbable event and so excitation only occurs within areas of high photon counts allowing for axial sectioning by scanning the laser and stacking images [22]. The collagen selectivity results from the vector addition of the structure-based photon emission as any centrosymmetric structures result in a vector sum cancellation (Figure 2). Other non-centrosymmetric biological structures can also be imaged, such as myosin, microtubules, and cellulose [23-24]. In the case of the tumor microenvironment, the different types of collagen present are the only relevant non-centrosymmetric biological structures. This technique has been used in multitudes of studies for understanding ECM changes caused by cancer. The images of collagen generated by this tool

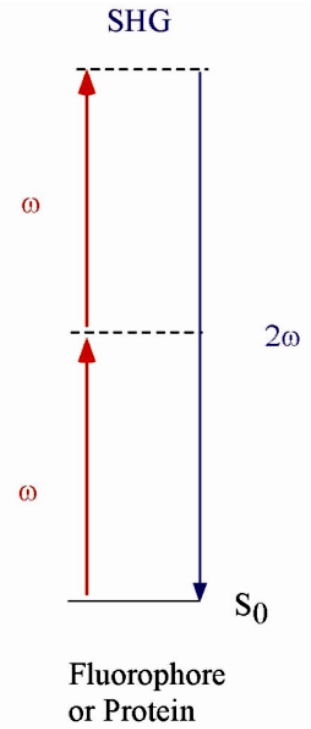


Figure 1: Jablonski diagram of two photon absorption with second harmonic generation

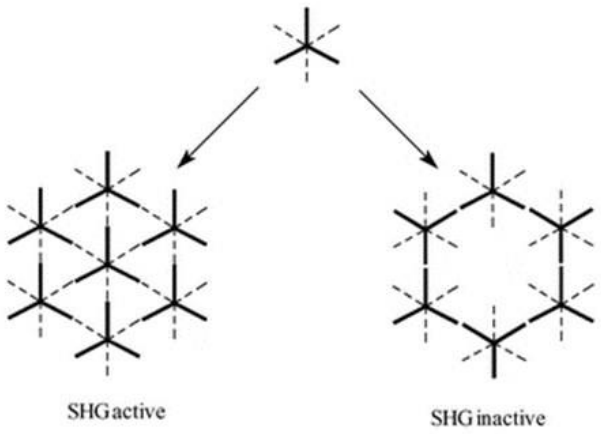


Figure 2: Example of non-centrosymmetric SHG active (left) and centrosymmetric SHG inactive (right) structures.

when combined with a powerful analytical technique could be the key for improved diagnostics and biological understanding of cancer.

**2D WTMM Anisotropy Method**

To understand the changing collagen structures, in the context of the tumor microenvironment, a tool capable of quantitative morphological analysis on SHG images is needed. A possible candidate for this tool is a wavelet-based anisotropy tool that was originally developed by Dr. Andre Khalil (Professor of Biomedical Engineering and Founding Director of the CompuMAINE Lab at UMaine) and applied in the context of Galactic astronomy [25] and then muscle morphogenesis [26-30] and artificial bone implants [31]. A version of this method was already specifically adapted for SHG microscopy images through a collaborative research effort between the University of Maine and Maine Health Institute for Research (MHIR) [32]. The method is called the 2D WTMM Anisotropy method and allows for a quantitative comparison of collagen alignment structure and directional preference between two images regardless of the number or density of collagen fibers. On each image, this analysis occurs across multiple scales by increasing the width of the Gaussian wavelet. The collaborative study involved the use of normal mouse skin, melanoma stimulated mouse skin, and a genetic knockout mouse skin that affects collagen fiber structure. These skin samples were imaged in a SHG microscope, and the aforementioned multiscale wavelet tool was applied. The statistically significant difference between the 3 skin types and overall trends changed with wavelet scale. At smaller scales melanoma was similar to normal tissue and the genetic knockout was different. This trend flipped at larger wavelet scales with melanoma now having the statistically significant difference and the genetic knockout did not. A probable interpretation, not possible without a multiscale technique, is that at smaller wavelet scales the genetic deformities in the collagen alignment are dominating whereas the melanoma is still normal at the fiber level. At larger scales the haphazard overproduction of the collagen in melanoma is dominating whereas the genetic

knockout collagen is still being produced and organized by cells whose deficiency is in collagen fibril assembly, not fiber bundle assembly. [32]. This study demonstrates the need for multiscale techniques to be available to researchers, and so the 2D WTMM Anisotropy Method will be formulated, updates to the method explored (Ch. 2), the updated method applied to PDAC (Ch. 3), and the method will be compared to those already used by researchers in the field (Ch. 5).

The method involves convolving an image with wavelets over a continuous range of scales allowing for multiscale analysis. The wavelets used for the analysis are the partial derivatives of a 2D Gaussian (isotropic) smoothing function, $\phi(x,y)$:

$$\psi_1(x,y) = \frac{\partial\phi(x,y)}{\partial x} \text{ and } \psi_2(x,y) = \frac{\partial\phi(x,y)}{\partial y} \tag{1}$$

The 2D wavelet-transform applied to the image, *f*, is calculated as follows:

$$\mathbf{T}_\psi[f](\mathbf{b},a) = \begin{pmatrix} T_{\psi_1}[f] = a^{-2}\int d^2\mathbf{x}\psi_1\left(a^{-1}(\mathbf{x}-\mathbf{b})\right)f(\mathbf{x}) \\ T_{\psi_2}[f] = a^{-2}\int d^2\mathbf{x}\psi_2\left(a^{-1}(\mathbf{x}-\mathbf{b})\right)f(\mathbf{x}) \end{pmatrix} \tag{2}$$

$$= \mathbf{T}_\psi[f](\mathbf{b},\mathrm{a}) = \boldsymbol{\nabla}\{\mathbf{T}_\phi[f](\mathbf{b},a)\}$$

from which the modulus and argument (angle) can be obtained:

$$M_\psi[f](\mathbf{b},a) = \left[\left(T_{\psi_1}[f](\mathbf{b},a)\right)^2 + \left(T_{\psi_2}[f](\mathbf{b},a)\right)^2\right]^{1/2} \tag{3}$$

$$A_\psi[f](\mathbf{b},a) = \mathrm{Arg}\left(T_{\psi_1}[f](\mathbf{b},a) + iT_{\psi_2}[f](\mathbf{b},a)\right)$$

At a given scale $a > 0$, the wavelet transform modulus maxima are the specific locations $\mathbf{b}$ within the image where the modulus $M_\psi[f](\mathbf{b},a)$ is locally maximum in the direction of the argument $A_\psi[f](\mathbf{b},a)$. At each size scale *a*, these WTMM are automatically organized as "edge detection maxima chains" [33]. A visual diagram of these mathematical operations is shown in Figure 3.

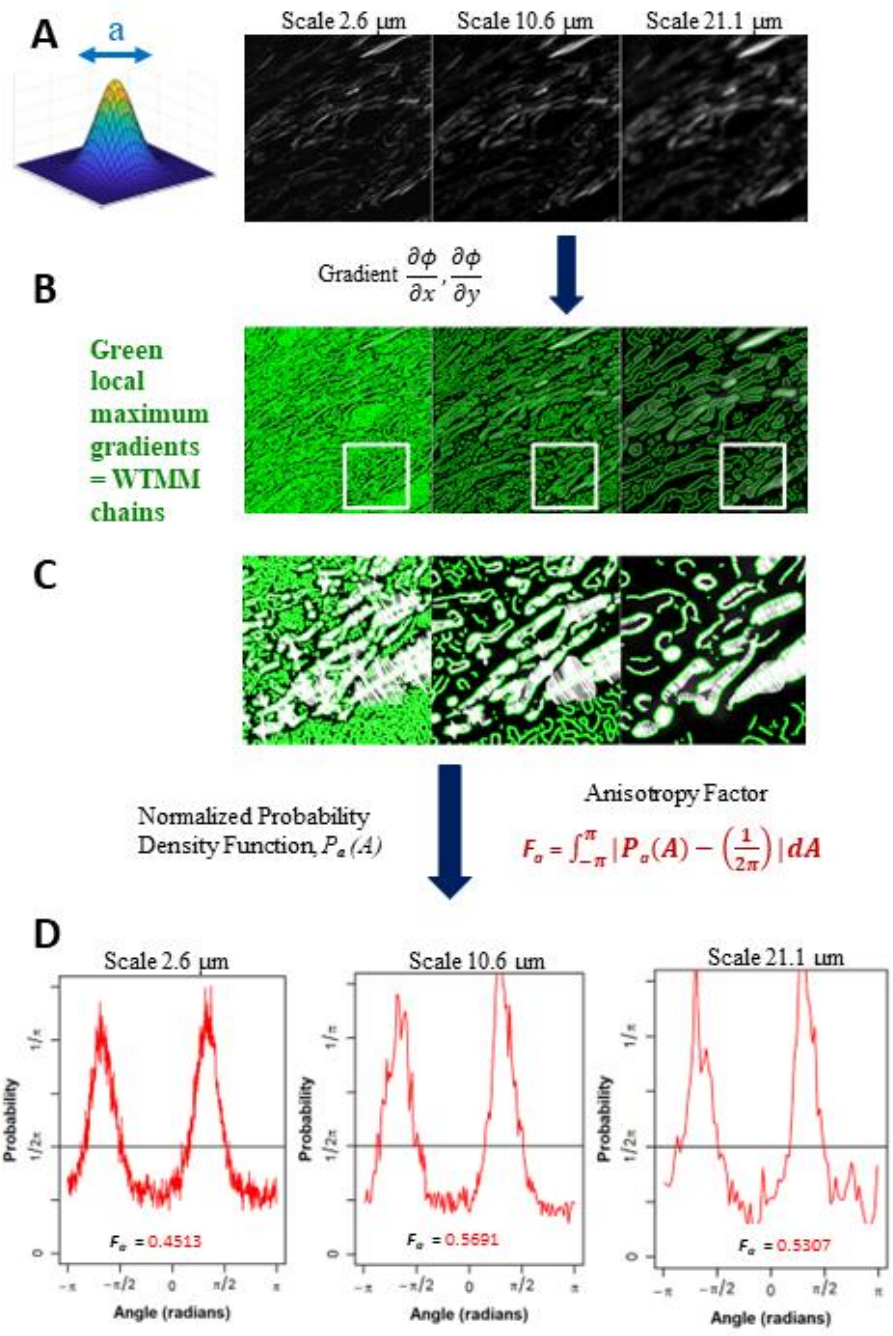


Figure 3: (A) Gaussian wavelet convolved SHG image at 3 representative wavelet size scales. (B) Maxima chains at 3 representative wavelet scales shown in green. (C) Zoomed inserts from white outlined area in panel B at 3 representative wavelet size scales to identify WTMM vectors (white arrows) showing magnitude and direction of local gradients of the maximum image density fluctuations. (D) Probability density functions (PDF), shown in red, of 3 representative wavelet scales generated from angles of local gradient vectors with the theoretical isotropic, $\frac{1}{2\pi}$, line shown in black. The anisotropy factor, Fa (Eq. 4), is shown for each PDF.

The angles from the WTMM chains are used to find the direction of changes in the image intensity. This allows for a definition of anisotropy based off the directional changes of image intensity across all size scales $a$. The angle distribution from the WTMM chains at each size scale are used to calculate probability density functions (PDFs), $P_a(A)$, of the angles. A flat PDF indicates a perfectly isotropic image with random direction of intensity variations. Therefore, deviations from a flat PDF are indicative of intensity variations at preferential angles (anisotropy). Numerically, binning of the PDF is fixed to a single number of histogram bins (i.e., 64) at each size scale regardless of the total number of maxima chains. To generate a quantitative measurement from the wavelet-transformed image at each size scale we used the anisotropy factor, $F_a$ where at each size scale $a$, the area between the curve of the image pdfs and a flat, $\frac{1}{2\pi}$, pdf is calculated:

$$F_a = \int_{-\pi}^{\pi} |P_a(A) - \frac{1}{2\pi}| dA \approx \sum_{i=0}^{\mathrm{N}_{bins}} \left| P_a(A_i) - \frac{1}{2\pi} \right| \Delta A \tag{4}$$

$$where\ \Delta A = \frac{2\pi}{\mathrm{N}_{bins}}\ , A_0 = -\pi\ , A_i = A_0 + \Delta A \cdot i$$

A value of $F_a = 0$ represents pure isotropy and greater values represent more anisotropy where the theoretical upper limit is 2 (See Appendix). This process is also shown in Figure 3. This results in a quantitative representation of image anisotropy which in turn provides information on collagen morphology.

## CHAPTER 2

## ANISOTROPY METHOD DEVELOPMENT

This chapter describes the changes to the previously published 2D WTMM Anisotropy Method (Chapter 1) that improve sensitivity, statistics, and introduces a new metric. This chapter also describes areas of improvement that still exist waiting to be explored.

### Variable Binning

The initial version of the 2D WTMM Anisotropy Method has a "fixed" number of bins per a wavelet scale, $a$. Each argument distribution histogram and its respective generated probability distribution function has 64 bins at each and every scale. Initially, this was done to minimize binning effects not from the 2D WTMM Anisotropy Method itself but, it obfuscates some of the available data. The number of maxima chains, $n$, is known to scale approximately as $a^{-2}$ [34] and an example of this scaling effect is shown by averaging the number of maxima chains at each wavelet scale for all PDAC SHG images (Figure 4). This also means the number of data points to bin is changing at each wavelet

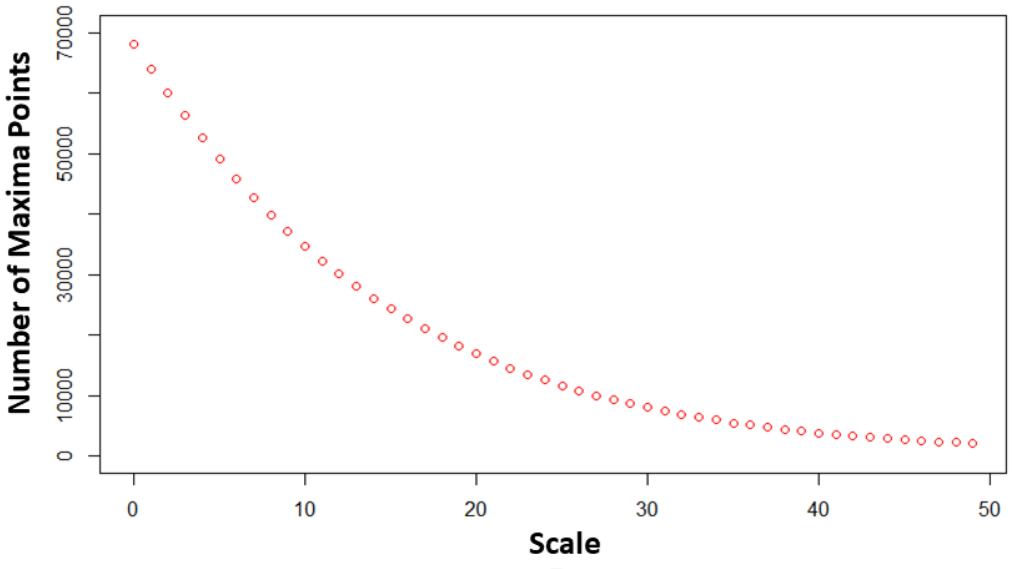


Figure 4: Average number of maxima points at each wavelet scale calculated from averaging every PDAC SHG image.

scale, with smaller scales having multiple orders of magnitude more points than the 64 bins, and larger scales having a few hundred data points. This results in oversampling at larger scales and under sampling at smaller scales. To remedy this, multiple histogram binning algorithms were explored such as Freedman-Draconis and Sturge's Rule, [35,36] but these are not designed to give equiprobable bins across multiple datasets or to vary bin widths within one dataset.

There is precedent for using variable bin widths within one histogram to maintain equiprobable frequency, so each bin is related by its samples per a bin or *n/k,* where *n* is number of samples and *k* is number of bins. This precedent could be applied across the wavelet scales, instead of within one histogram, to maintain a roughly equal number of samples per a bin across the scales. There is a general rule for variable binning based on the Pearson chi-squared test of $2n^{2/5}$ [37]. The number of bins

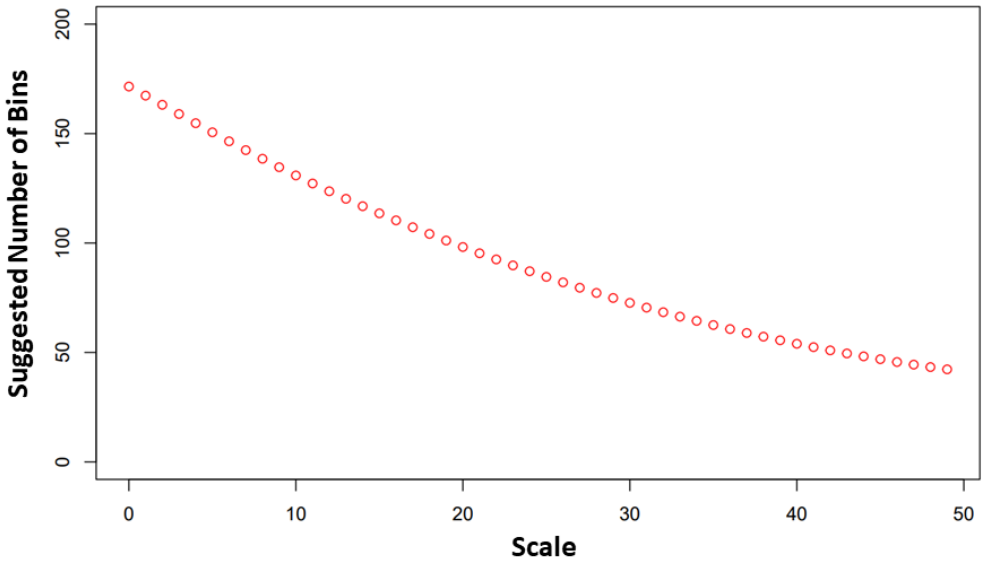


Figure 5: Suggested number of bins calculated from a Pearson chi-squared test derived number of groups algorithm ($2n^{2/5}$) per each wavelet scale.

suggested for each scale using this equation ranged from 172 to 42 (Figure 5). Following the suggested number of bins results in a fairly equiprobable samples per a bin graph ranging from 397 to 49 samples a bin (Figure 6). Ideally, there should be some sort of weight to the more statistically significant smaller wavelet scales, and less weight to the larger wavelet scales.

To try this, all the data was binned by *n/100,* which is the number of bins suggested from the equation for a middle wavelet scale. This causes our number of bins to range from 680 at the smallest scale to 20 at the largest scale and there are always 100 samples for each bin (Figure 7).

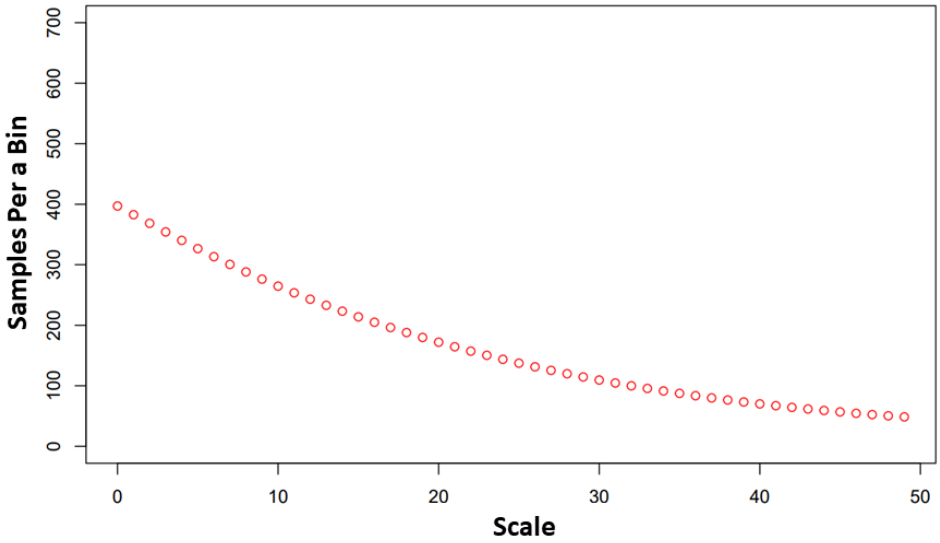


Figure 6: Number of samples per a bin when using the suggested number of bins shown in Figure 5

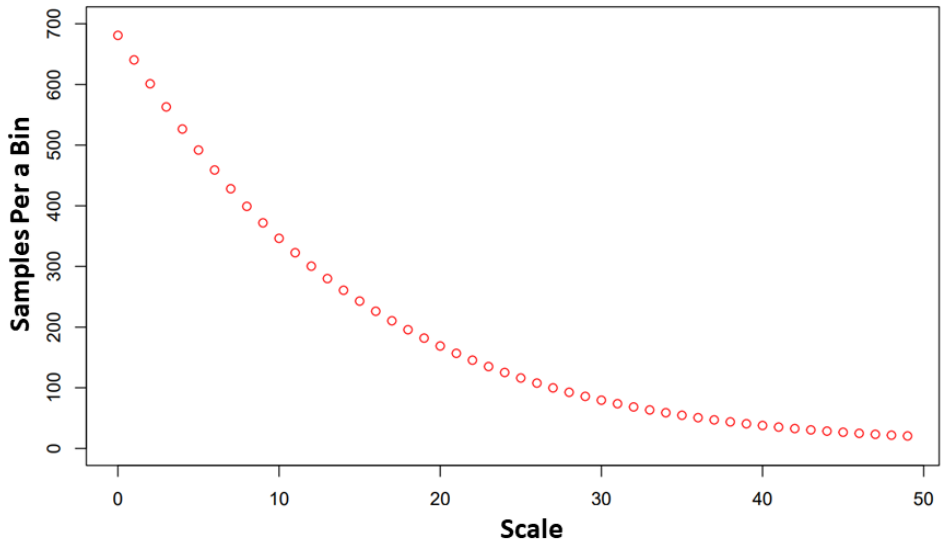


Figure 7: Number of samples per a bin when using the *n/100* binning strategy.

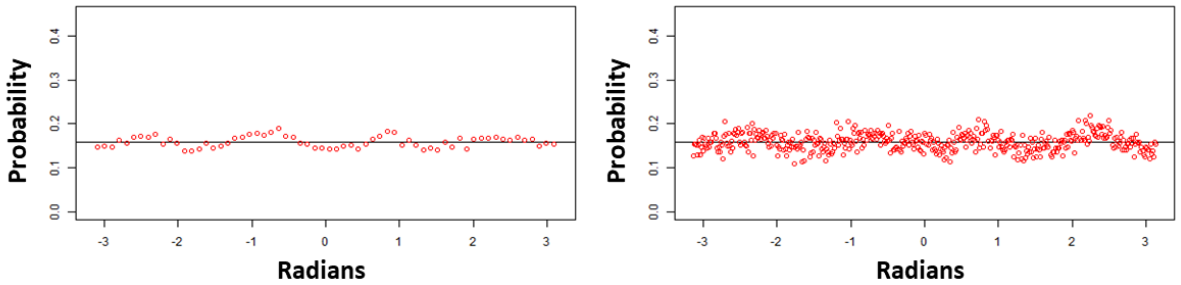


Figure 8: Example fixed binning (left) probability distribution function and variable binning (right) probability density function with $1/2\pi$ line shown in black.

To compare fixed vs variable binning, 100 white noise images of size 512x512 pixels were generated and processed using both binning strategies. An example probability density function from a white noise image at wavelet scale 10 of fixed and variable binning is shown in Figure 8. At each scale the mean anisotropy factor and standard deviation was recorded. A Wilcoxon ranked sum test was always used to compare fixed vs variable binning, but there was not a statistically significant difference. This is not a surprising result as the two methods were tested on the exact same dataset. Another metric to compare the methods called the coefficient of variation (CV) was used and is formulated as the following:

$$CV = \frac{\sigma}{\mu} \tag{5}$$

where $\sigma$ represents the standard deviation and $\mu$ represents the mean. This compares changes in variability and the effects of the aforementioned under sampling can clearly be seen by this test (Figure 9). The fixed binning method has a CV of 0.4 at the smallest wavelet scales where the method should statistically be the strongest. The variable binning method demonstrates these improved statistics at small size scales and eventually both methods have similar CVs at the larger size scales, but only variable binning can differentiate between the three white noise image sizes used. The updated 2D WTMM anisotropy method now consists of this binning strategy of *n/100*, but other variable binning strategies could still be explored in the future.

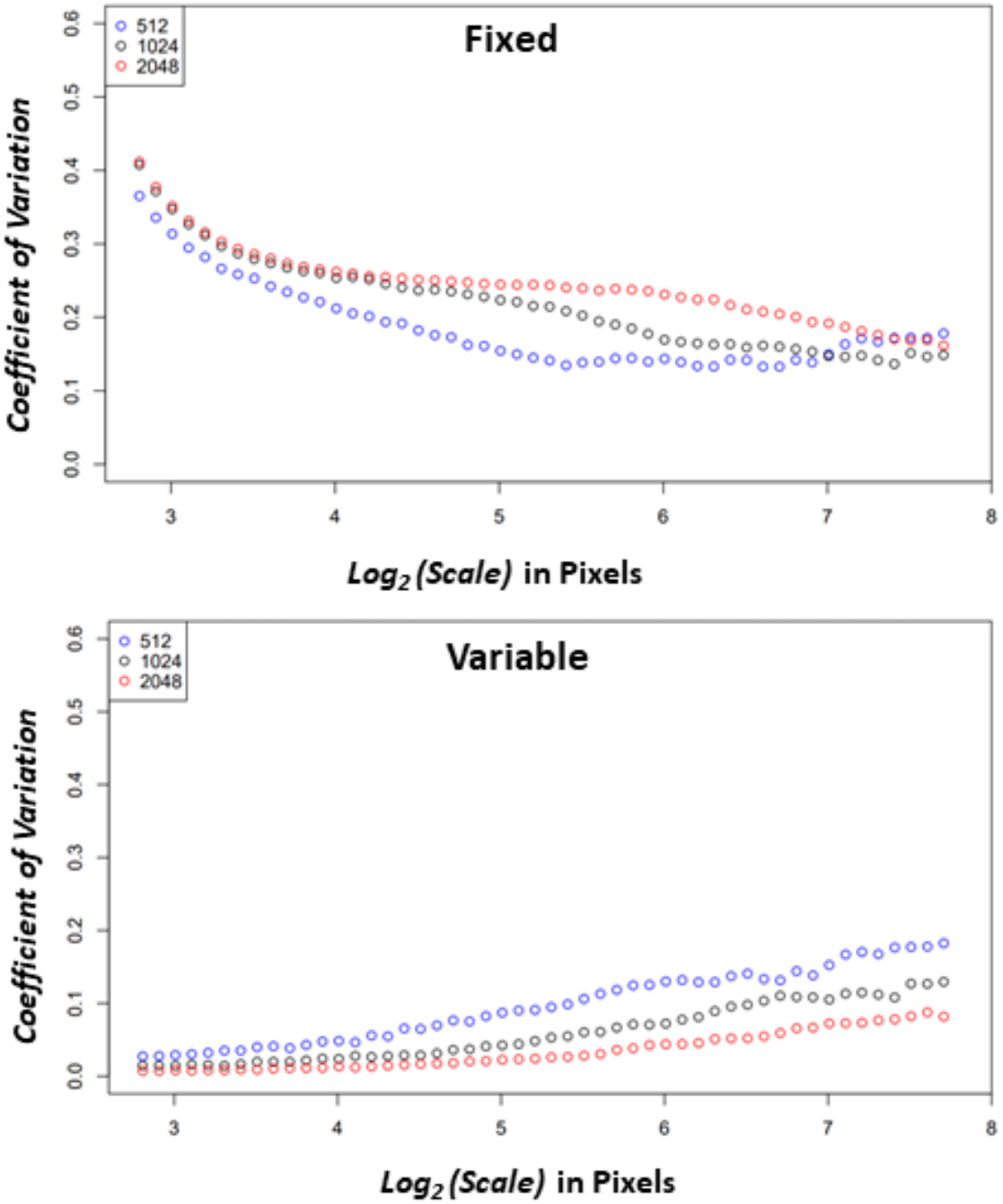


Figure 9: Using white noise calibration images of dimensions 512x512 (blue), 1024x1024 (black), and 2048x2048 (red) pixels for coefficient of variation calculation. The anisotropy factor across the 50 wavelet size scales was calculated for each white noise image and then all 136 image anistropy factors were used to calculate mean and standard deviation at each size scale. Coefficent of variaiton was calculated as standard deviation divided by mean (Eq. (1)) using both fixed binning anistropy factors (top) and variable binning anistropy factors (bottom).

**Normalization Strategies**

The originally formulated 2D WTMM Anisotropy Method uses Brownian surface motion images for normalization. This normalization was conducted by dividing the experimental data's median anisotropy factor by the median Brownian surface motion anisotropy factor at each wavelet scale. The goal for this specific normalization is to remove effects from discrete binning, image size, and the wavelet convolution itself, therefore a scale invariant isotropic image candidate such as a Brownian surface motion image is needed. Early on it was noticed that Brownian surface motion images generated by xmsurf have a large-scale sinusoidal pattern originating from where the low and high pixel intensity areas are located. This manifests in the anisotropy factor with Brownian surface motion images having similar anisotropy factors to some SHG image categories as can be seen in an early version of the SHG PDAC 2D WTMM Anisotropy analysis described later in chapter 3 (Figure 10). A scale invariant isotropic image candidate should not trend with images of biological collagen fibers and so other image types were explored. A quick test of generating Brownian sheets instead of surfaces was conducted in xsmurf and resulted in the same large scale anisotropy effects noticed in Brownian surface motion images. A promising candidate was soon found in white noise images, whose purely random pixel intensities should result in isotropy (Figure 11). An experimental comparison of white noise and Brownian surface motion images was conducted with 100 images of 512x512, 1024x1024, and 2048x2048 pixel dimensions which were generated in xsmurf and processed using the variable binning 2D WTMM Anisotropy Method (Figure 12).

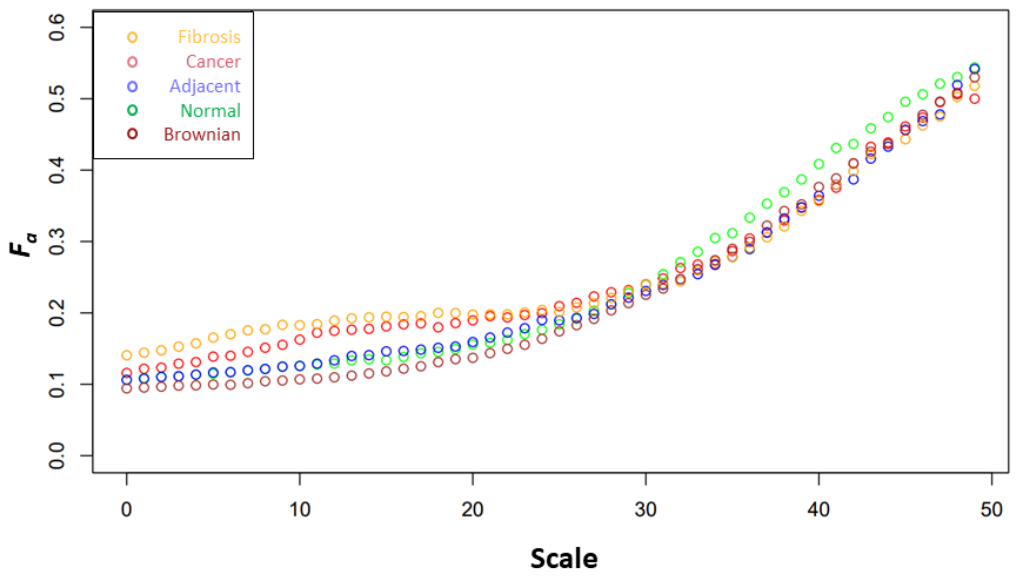


Figure 10: Median anisotropy factor across all images for fibrosis (orange), cancer (red), adjacent (blue), normal (green) tissue and Brownian surface motion (brown) plotted vs wavelet scale

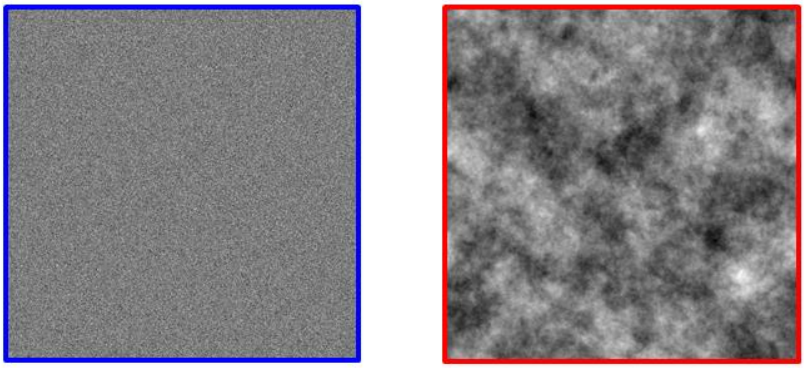

Figure 11: Example 512x512 images of white noise (blue) and Brownian surface motion (red) generated using xsmurf.

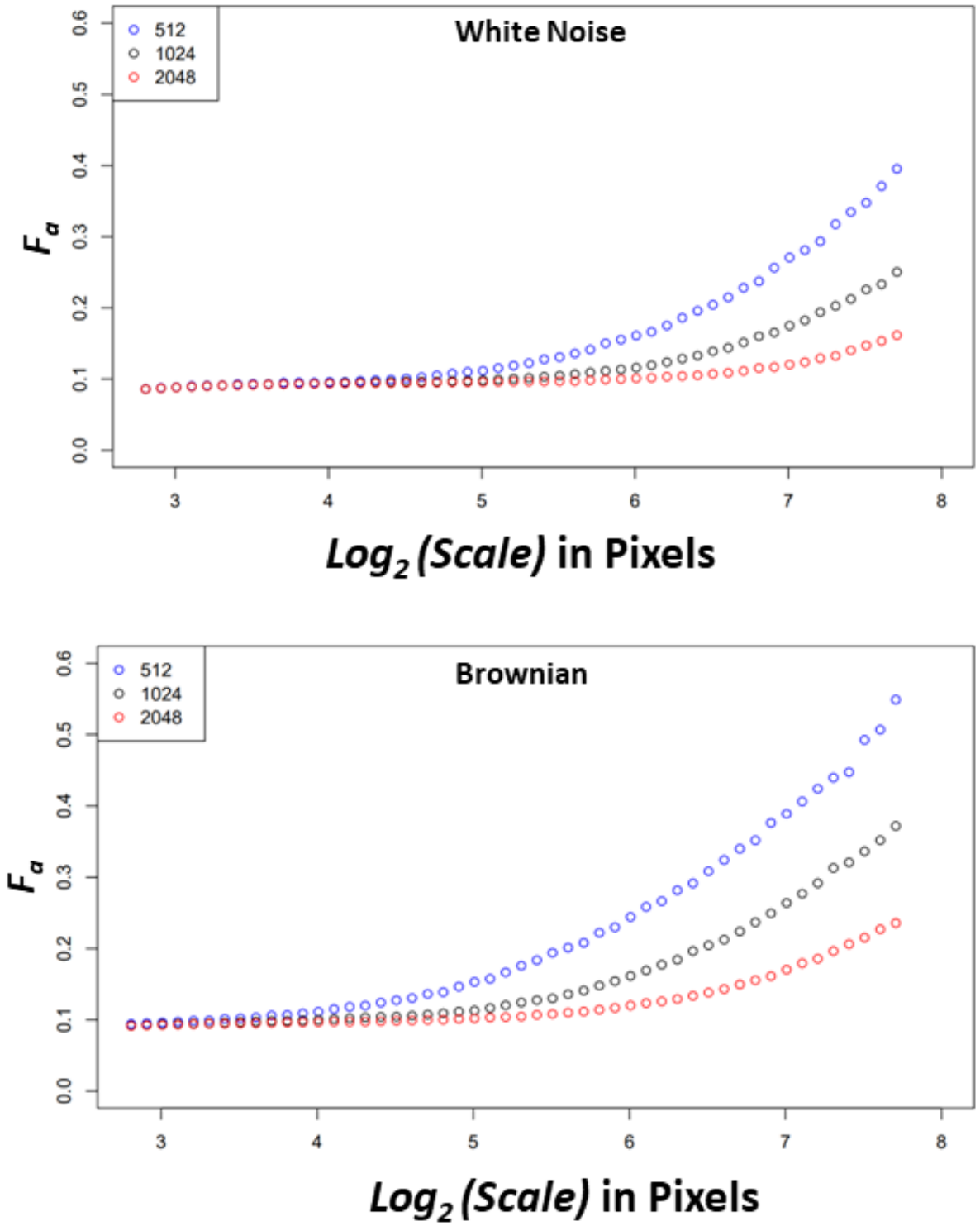


Figure 12: Using Brownian and white noise calibration images of dimensions 512x512 (blue), 1024x1024 (black), and 2048x2048 (red) pixels to study size effects. The median anisotropy factor across the 50 wavelet size scales was calculated from the full image for white noise (top) and Brownian (bottom).

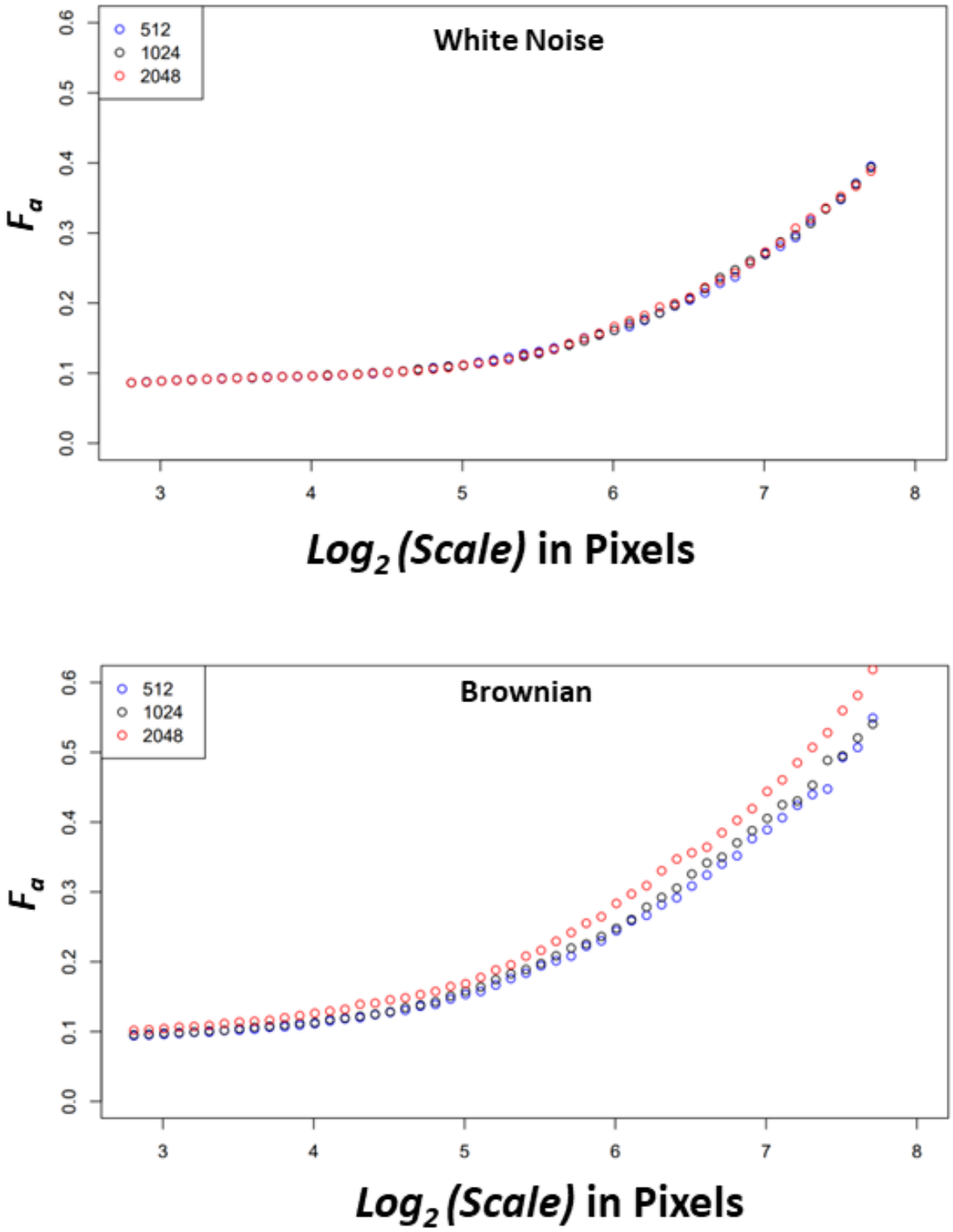


Figure 13: Using Brownian and white noise calibration images of dimensions 512x512 (blue), 1024x1024 (black), and 2048x2048 (red) pixels to study edge effects. The median anisotropy factor across the 50 wavelet size scales was calculated using only the 512x512 pixel center cutout of the wavelet convolved images from the 3 dimensions for white noise (top) and Brownian (bottom).

At any scale beyond the first few, white noise has a lower anisotropy factor, and the anisotropy factor remains scale invariant longer for every image size. To compare edge effects between the two image types, a center 512x512 square was cut-out of the post wavelet transform extrema image and only the cut-out maxima points were used in anisotropy analysis (Figure 13). This resulted in no differences in anisotropy factor between 512x512 images or cut outs for white noise, whereas Brownian motion surface images did experience differences in anisotropy factor. In both these experiments, white noise outperformed Brownian surface motion images and is now used as the isotropic normalization image for the updated 2D WTMM Anisotropy Method.

The other aspect of data normalization changed is how the information from the isotropic images are applied to actual experimental image data. Historically, it is the median anisotropy factor from the data divided by the median anisotropy factor from the isotropic data. The core issue with this approach is that an isotropic image should approach an anisotropy factor of zero, which would result in enlarging the anisotropy factor values, thereby exaggerating minor differences in anisotropy factors. Other approaches to normalization are being explored, with the current approach involving a simple subtraction of the white noise median anisotropy factor at each size scale. This automatically adapts by image size and wavelet scale removing discrete effects that appear from the method. In an ideal, perfect application of the method to an isotropic image the anisotropy factor would be zero, and so subtracting by the median anisotropy factor of zero would result in no normalization of the data and this is the main rationale for the new normalization approach.

**Oscillation Factor**

Other quantitative metrics from this approach would allow for improved multivariate analysis of experimental results and is a constant cogwheel turning in the back of the CompuMAINE team's minds. One quantitative metric known as the oscillation factor was first conceived of by looking at an isotropic probability density function hovering around the $1/2\pi$ line (Figure 14). If two images have an anisotropy factor of one, that describes only the area between the image's argument generated probability density function and the $1/2\pi$ line, not the shape of the function that generated the anisotropy factor. A probability density function could have four sharp peaks, such as a fiber image, to have an anisotropy factor of one, where another image could cross a dozen times, but also have an anisotropy factor of one. One way to compare these two probability density functions would be to count the number of times their function crosses the $1/2\pi$ line. This count is known as the aforementioned oscillation factor.

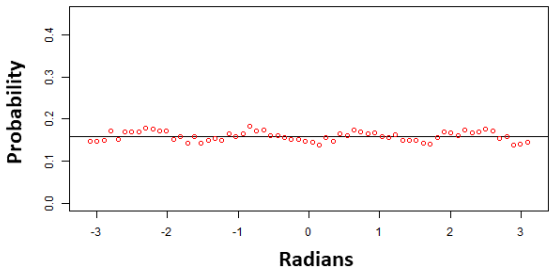


Figure 14: Example probability distribution function from a white noise image with $1/2\pi$ line shown in black.

The oscillation factor is less sensitive to minute changes in image structure when compared to the anisotropy factor but is sensitive to the macro structural origin of the anisotropy factor value itself. For example, the oscillation factor medians were calculated for 136 white noise and Brownian surface motion images, as well as 160 PDAC SHG images (see chap. 3) of normal and cancerous tissue (Figure 15). The white noise and Brownian surface motion images trend with each other, while the two SHG

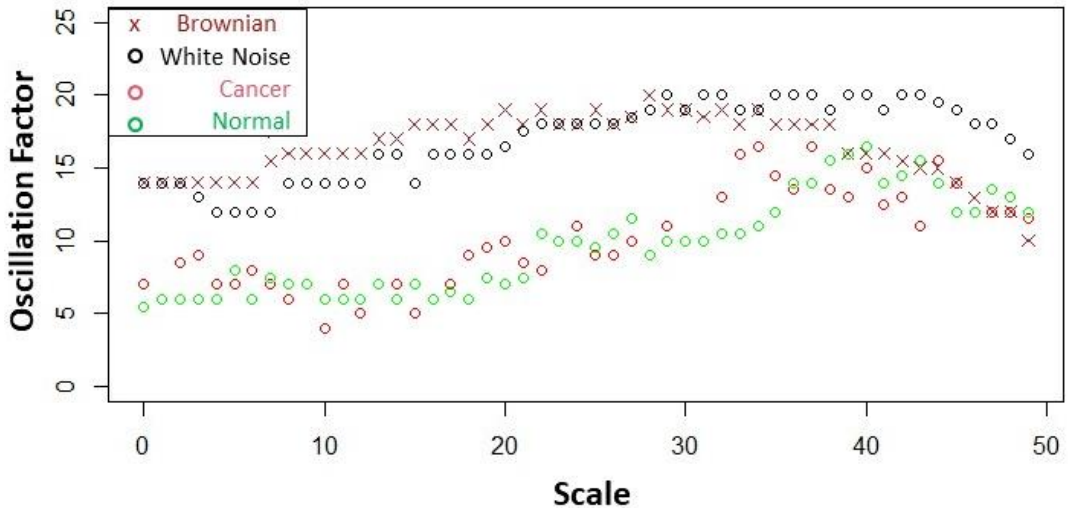


Figure 15: Median oscillation factor across the 50 wavelet scales for Brownian surface motion (brown), white noise (black), Cancer (red), and normal (green) image categories.

images of collagen fibers trend with each other. This is despite the statistically significant differences in anisotropy factor between the cancer and normal groups discussed later in chapter 3. The large-scale sinusoidal structure from the Brownian motion surface images can also be seen in this plot, where at larger scales the oscillation factor trends with the SHG fiber images. Since the oscillation factor is not a very sensitive metric to minute structural changes, it is not applied to the SHG study in chapter 3 but should be applied as a quantitative metric to determine if non-statistically significant anisotropy factors are caused by probability density functions with a similar number of peaks.

An exploration of the fundamental algorithmic principles of the oscillation factor was also conducted by comparing different random processes to the isotropic white noise and Brownian surface motion images. The first attempt at replicating an isotropic image's isotropy was by generating a

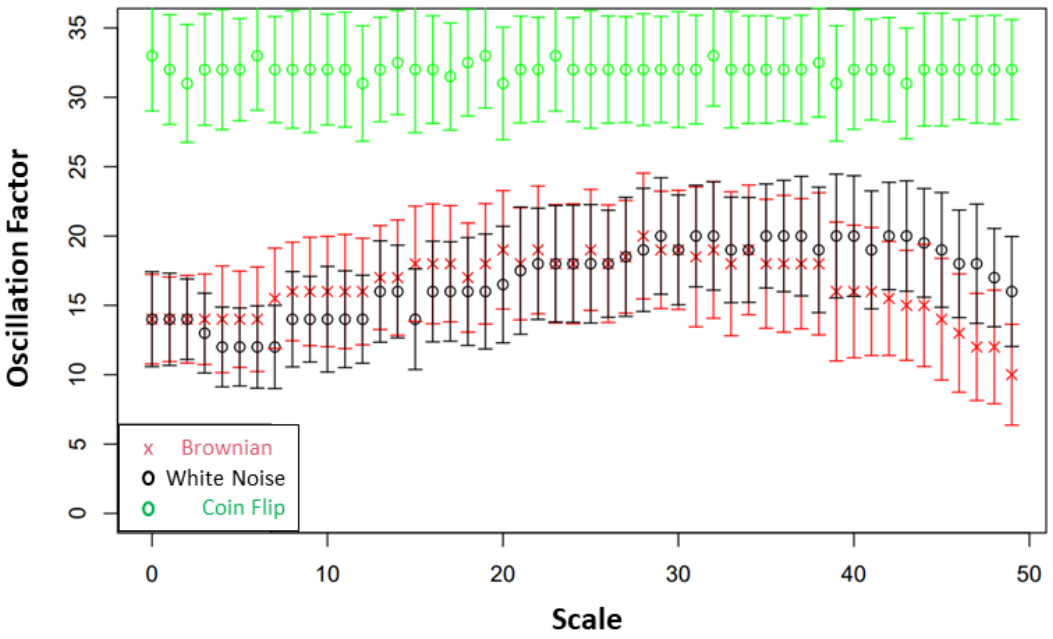


Figure 16: Median oscillation factor with standard deviation across the 50 wavelet scales for Brownian surface motion (brown), white noise (black) image categories. The coin flipping simulated oscillation factor's median and standard deviation are also plotted vs wavelet scale (green).

probability density function through a coin flip determining if its above or below the $^1/_{2\pi}$ line (Figure 16). This generated an oscillation factor far greater than the isotropic images' oscillation factor. The second attempt involved using a random walk to simulate an isotropic image's probability density function. The function was generated linearly for 64 bins starting on the $^1/_{2\pi}$ line, at each subsequent point a random number from 1-100 was generated and this number was used to determine if it takes one step towards the $^1/_{2\pi}$ line or away from the line.

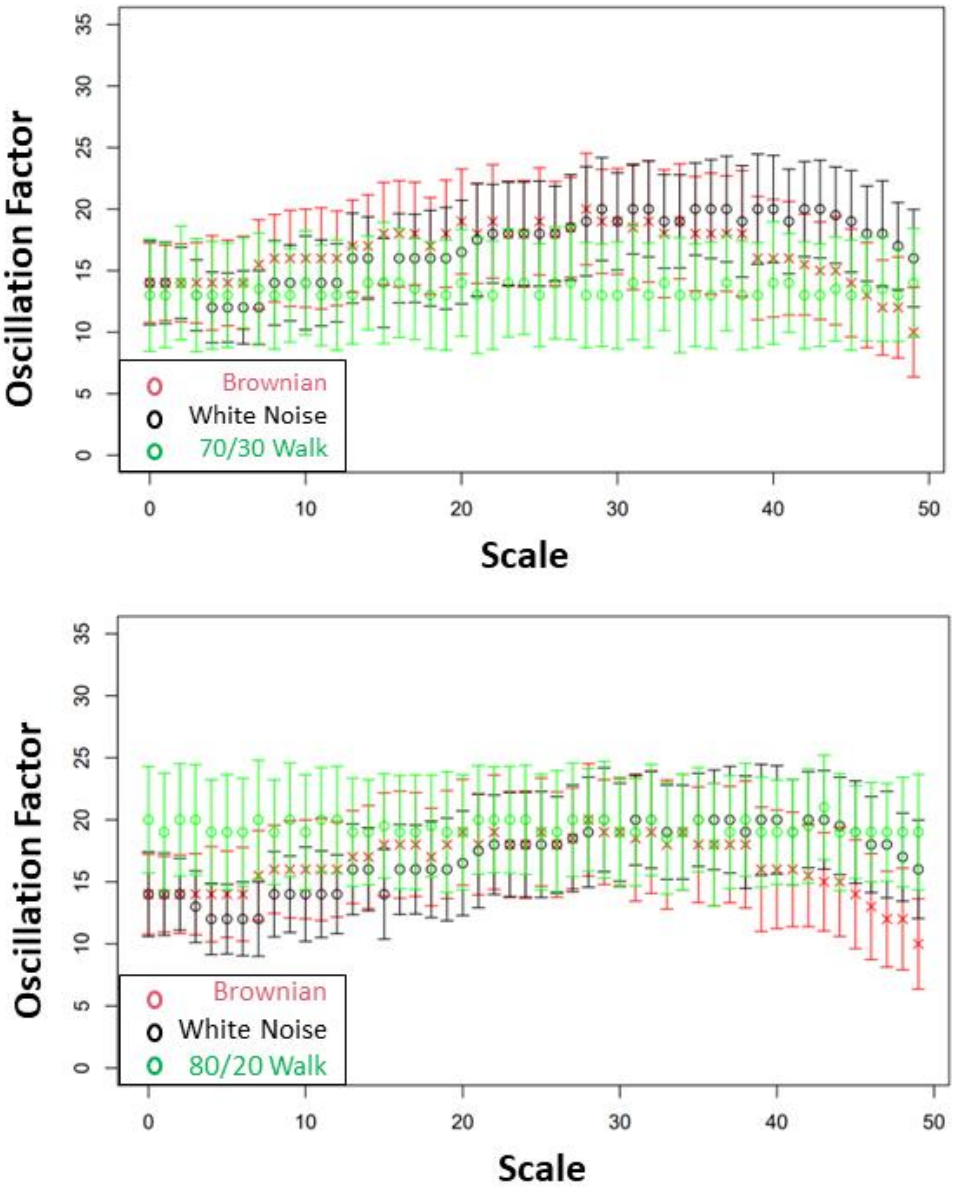


Figure 17: Median oscillation factor with standard deviation across the 50 wavelet scales for Brownian surface motion (brown), white noise (black) image categories. The oscillation factor's median and standard deviation from the 70/30 random walk (left) and 80/20 random walk (right) are also plotted vs wavelet scale (green).

If on the $^{1}/_{2\pi}$ line, the probability of moving up one step or down one step was 50/50. The probabilities of moving towards the line or away from the line were varied and oscillation factors calculated to determine if it replicated isotropic image oscillation factors. This is shown for a 70/30 probability of moving towards the $^{1}/_{2\pi}$ line versus away from the line, as well as an 80/20 probability of moving towards or away (Figure 17). These seem to replicate the oscillation factor, but not the multiscale effects. The 70/30 random walk follows the trend at the smallest and largest wavelet scales, where the 80/20 random walk trends with the middle group of wavelet scales' oscillation factor. Future work could be done to combine these two probabilities to replicate the oscillation factor almost perfectly in isotropic images and work is currently underway using variable binning probability density functions to calculate the oscillation factor.

## CHAPTER 3

## APPLICATION TO PANCREATIC DUCTAL ADENOCARCINOMA AND PREDICTIVE POWER

The importance of developing tools capable of distinguishing between benign and cancerous tissue microenvironments cannot be overstated when it comes to improving quality of patient care and positively impacting patient outcomes. To explore the efficacy of the updated 2D WTMM Anisotropy method as a quantitative analytical and diagnostic tool it was tested on SHG images of the deadliest solid organ cancer, pancreatic ductal adenocarcinoma (PDAC), known for its fibrosis and difficulty of diagnosis.

### Experimental Setup

We partnered with the Maine Health Institute for Research (MHIR) and obtained 40 H&E-stained pathology slides of stage III PDAC that were biopsied from forty unique patients. These slides were scanned with an Aperio2 slide scanner (Leica Biosystems, Wetzlar, Germany) and then annotated by a pathologist who highlighted areas of cancer, fibrosis, and normal tissue. Twenty slides were annotated as only normal tissue and the other twenty slides had areas annotated as cancer. Of the twenty cancer slides, 15 also had areas annotated as fibrosis and 13 had areas annotated as normal tissue. Normal tissue on the same slide as cancer tissue was categorized as normal adjacent tissue.

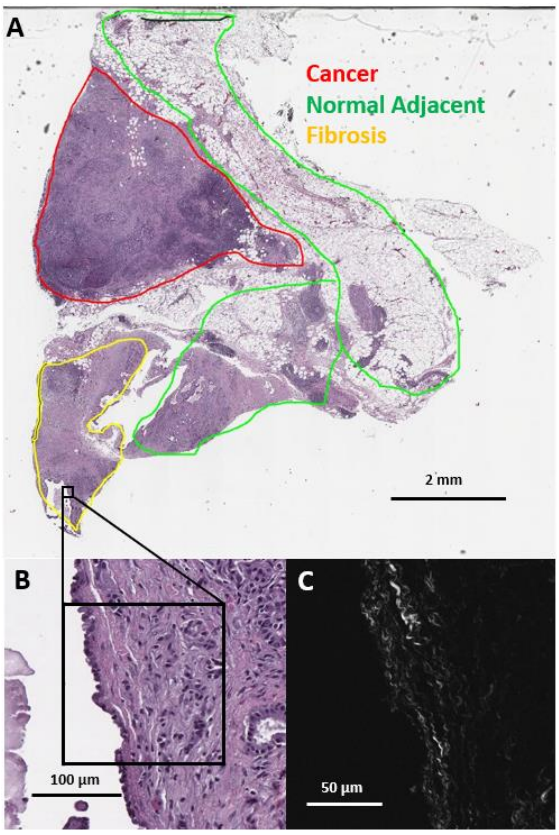


Figure 18: (A) Pathologist labeled H&E slide scan from a pancreatic tumor biopsy with an example of one SHG microscopy imaged area outlined in black. (B) Zoomed view of outlined area from H&E scan. (C) SHG microscopy image of outlined area.

This results in four tissue categories from the annotated slides: cancer, normal, fibrosis, and normal adjacent (see Table ch3t1). Eight distinct ductal structures within each normal, normal adjacent, fibrosis, and cancer were identified on the H&E slide and then imaged using SHG microscopy (Figure 18).

SHG images were acquired on a custom-built 2-photon microscope consisting of an upright microscope stand (Olympus BX50WI, Olympus, Center Valley, Pennsylvania), laser scanning unit (Fluoview300, Olympus), titanium sapphire femtosecond laser (Chameleon Ultra II, Coherent, Santa Barbara, California), and an electro optic modulator (ConOptics, Danbury, Connecticut) for laser power modulation. Circular polarization was used for SHG imaging; it was verified at the focal plane by rotating a polarizer and experiencing no change in laser power. Forward directed SHG images were acquired at 890-nm excitation, using a LUMPlanFLN 40 × 0.8 NA (Olympus, Center Valley, Pennsylvania) water immersion objective. Forward directed SHG was collected in a 0.9 NA condenser lens and filtered using a 448/20 bandpass filter and the forward directed SHG signal was collected using a 448/20-nm (Semrock Rochester, New York) prior to detection via a H7421 GaAsP PMT (Hamamastsu, Hamamstsu City, Japan). Each region of interest was imaged using 2x optical zoom (180 μm field of view) with a digital resolution of 512 × 512 pixels and using a laser scanning speed of 2.71 s/frame with the final image resulting from Kalman filtering using 4 state updates. The full list of images and tissue categories used in this study are shown in Table 1. The updated 2D WTMM anisotropy method was applied to all images listed, still organized by the same categories.

Table 1: Full Image Database Used in Study

| Tissue Category | Number of Slides | Number of Images |
|---|---|---|
| Cancer | 20 | 160 |
| Normal | 20 | 160 |
| Fibrosis | 15 | 120 |
| Normal Adjacent | 13 | 104 |

**PDAC SHG 2D WTMM Results and Discussion**

The new variable binned, white noise controlled, multi-scale anisotropy factor results of normal, normal adjacent, fibrosis, and cancer are plotted in Figure 19. There are significant differences in anisotropy factor values as can be seen between the cancer/fibrosis and normal/adjacent categories. Interestingly, at a wavelet scale of 21 µm all tissue types converge with only normal tissue diverging at larger scales. An off-the-shelf analysis tool a researcher might select to analyze this data would use a

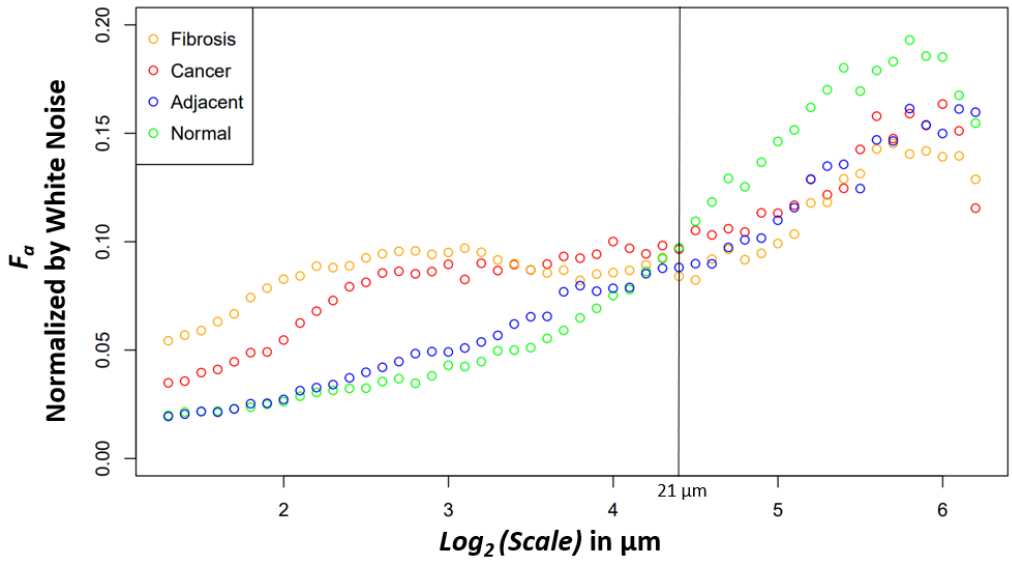


Figure 19: Median anisotropy factor across all images for fibrosis (orange), cancer (red), adjacent (blue), and normal (green) tissue normalized by white noise anisotropy factor from ~2.5 µm to 73.5 µm.

fixed scale and if that fixed analysis scale was 21 µm the researcher would conclude no difference in this data. This is a severe limitation of single scale analysis tools, and this limitation is further explored in chapter 4.

Four relevant pairwise comparisons are shown in Figure 20 with corresponding Wilcoxon ranked sum test p-values. Cancer vs normal, shown in Fig. 20A, are statistically significantly different at all size scales other than the convergence wavelet range. At smaller size scales, before the cross over scale, cancer is more anisotropic and at larger size scales, after the cross over scale, the normal tissue is more anisotropic. In Fig. 20B, the comparison between adjacent and cancer tissue can be seen where they are statistically significantly different at small wavelet size scales but are not at the larger wavelet size scales. The fibrosis and cancer are not statistically significantly different except at the first four wavelet size scales shown in Fig. 20C. In Fig. 20D, the normal tissue is more anisotropic than normal adjacent tissue at larger size scales but is not statistically significantly different at other size scales.

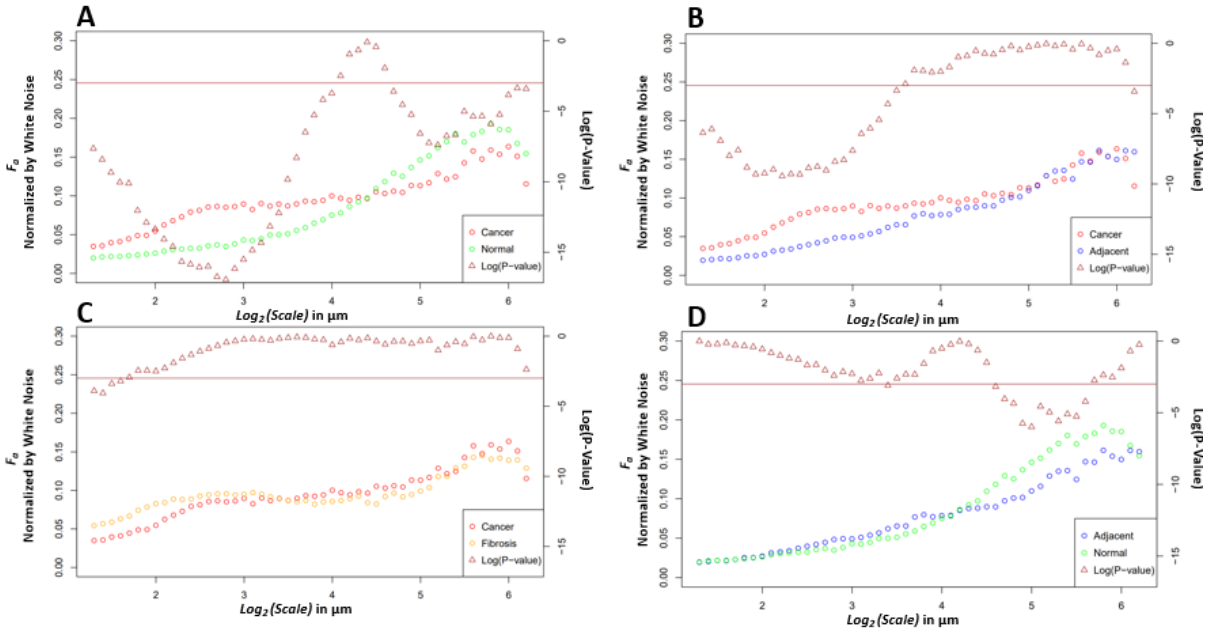


Figure 20: Pairwise comparisons of the median anisotropy factor across all images normalized by white noise anisotropy factor from ~2.5 µm to 73.5 µm (left *y*-axis) for cancer tissue *vs.* normal tissue (A), cancer tissue *vs*. adjacent tissue (B), cancer tissue *vs.* fibrotic tissue (C), and adjacent tissue *vs.* normal tissue (D). The *p*-values (right *y*-axis), shown in brown, from a Wilcoxon rank sum test ran at each scale between the tissue pair's anisotropy factors. Values below the horizontal brown line at 0.05 show statistical significance.

Overall, the results demonstrate that the updated 2D WTMM Anisotropy Method is sensitive to changes in collagen, furthermore the differences found here between the tissue types can elucidate information about ECM modification in both close and distant proximity from the tumor. The higher anisotropy of cancer at smaller scales could be caused through stellate cell over production of collagen and cancer cell mediated remodeling [38]. At larger scales, after an interesting ~21 μm wavelet size crossover point, the larger anisotropy in healthy tissue suggests that the cancer has disrupted the overall normal large-scale organization of collagen fiber bundles. This observation is further supported by the adjacent normal tissue results. The cancer and adjacent normal tissue are different at smaller scales, but this statistical significance starts dissipating at a wavelet scale of ~12 μm. At larger size scales the adjacent tissue cannot be differentiated from cancer suggesting again a large-scale collagen disruption due to the cancer even in areas distant from those annotated as cancer by a pathologist. An important note when comparing normal and adjacent normal tissue is the lack of 3D spatial organization as this study was prospective and careful 3D mapping of pancreatic tissue processing was not performed. Therefore, it is plausible that a slide containing only normal tissue is near-adjacent to cancerous tissues in a manner similar to normal adjacent tissue on a slide annotated with both cancerous and normal tissue. Despite this possibility, there was statistical significance between normal adjacent and normal tissues at larger scales suggesting that collagen organization is disrupted near cancerous tissue. Finally, the comparison of fibrosis vs cancer only found slight statistical significance over the first four wavelet scales which range from ~2.4 μm to ~3.0 μm. This difference could be due to the straightness of the fibers in the fibrotic tissue. Otherwise, the overall trend of anisotropy factor vs scale matched between the two tissue types. This suggests that there are minor differences in collagen structure between fibrosis and the cancer collagen. The pathologist labeled these tissue types using the gold standard, H&E histology, but it is highly subjective, and moreover, not highly sensitive to the collagen fibrillar aspects which seem to contain some valuable information regarding normal vs

cancerous vs fibrotic tissue. Overall, this method demonstrates the ability to differentiate changes in collagen structure induced by PDAC using a continuous multiscale suite of information allowing for deeper analysis than other methods.

**Preliminary Predictive Power**

The other goal of this study was to determine the diagnostic capabilities of the 2D WTMM Anisotropy Method combined with SHG on biopsied tissue. To accomplish this goal a slide level metric from the fifty wavelet scales needs to be developed. Statistically significant differences were noticed at both small and large scales before and after the 21 µm cross over point. These small- and large-scale differences could be used as a two-variable metric for slide level 2D clustering analysis. The same five small wavelet scale anisotropy factors and five large wavelet scale anisotropy factors for each tissue category were averaged to calculate an average small scale anisotropy factor and average large scale anisotropy factor per an image (Figure 21). These image anisotropy factors were then averaged at a slide level calculating an average small-scale anisotropy factor and large-scale anisotropy factor per a patient.

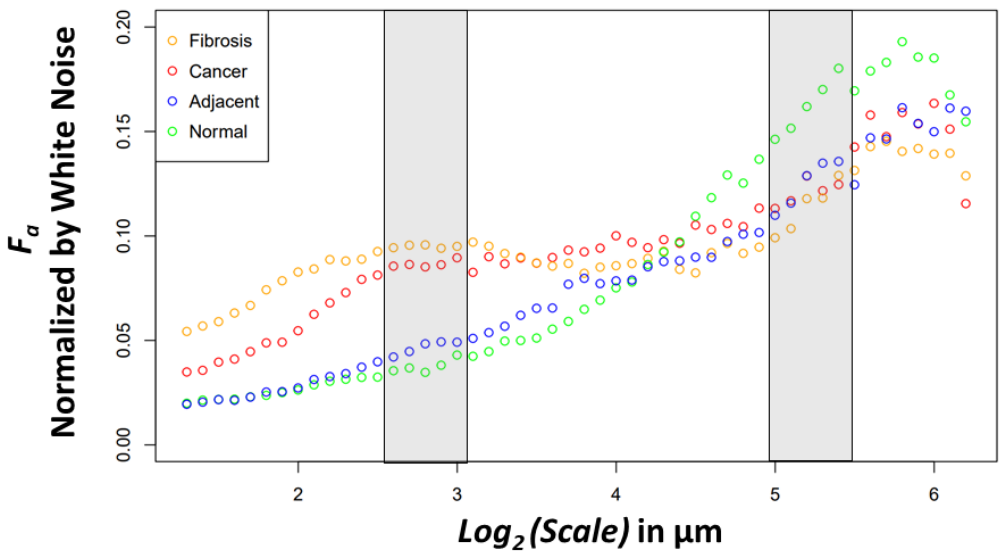


Figure 21: Median anisotropy factor across all images for fibrosis (orange), cancer (red), adjacent (blue), and normal (green) tissue normalized by white noise anisotropy factor from ~2.5 µm to 73.5 µm with small and large wavelet scale ranges highlighted.

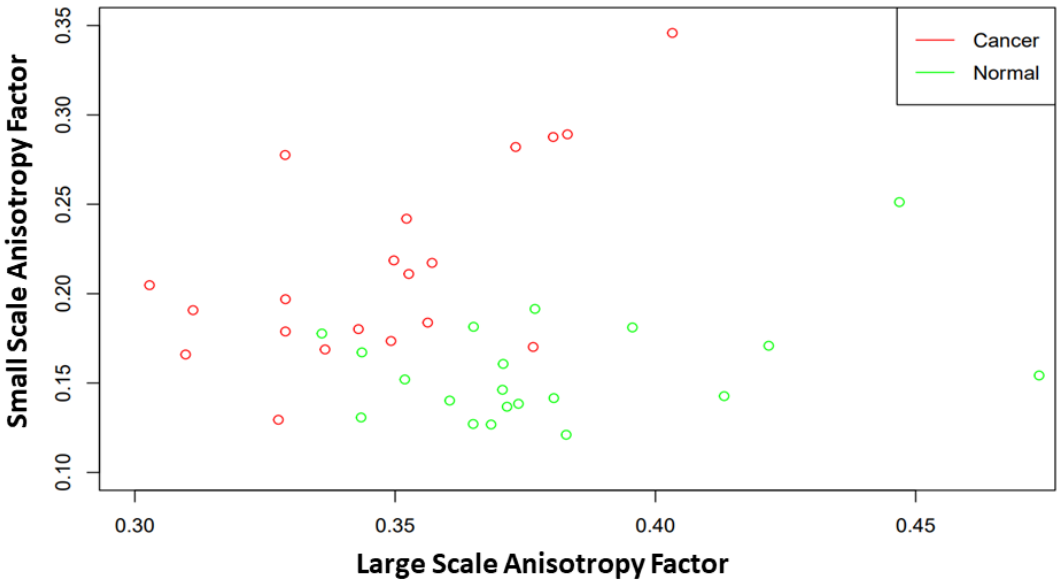


Figure 22: Plot of each cancer (red) and normal (green) slide by small- and large-scale anisotropy factor.

These were then plotted on a 2D plot of small vs large scale anisotropy value. The only categories with clear clustering were cancer and normal as all other category types were only statistically significantly different at one of the two wavelet ranges and therefore not shown (Figure 22). From this plot, healthy tissue tends to have less small-scale anisotropy and more large-scale anisotropy than cancer. Since clustering was noticed in the 2D plot it could be possible for a clustering algorithm to automatically categorize the slides into cancer and normal types based on the trends in anisotropy factor. For preliminary checking of this predictive power K-means clustering analysis was selected and conducted using the cluster and clusterR libraries in RStudio [39-41]. This clusters by placing two points randomly, known as K-means, on the plot and using an iterative minimization function that calculates the direction to move the K-mean based off minimizing distance from the other points. This is done until both K-means find local minimum and no longer move and then points are categorized based on the closest K-mean. Due to random initial K-mean placement, this is run 100 times to find the lowest minimum out of all local minima. The results of this analysis are shown in Figure 23 where a majority of slides are incorrectly categorized.

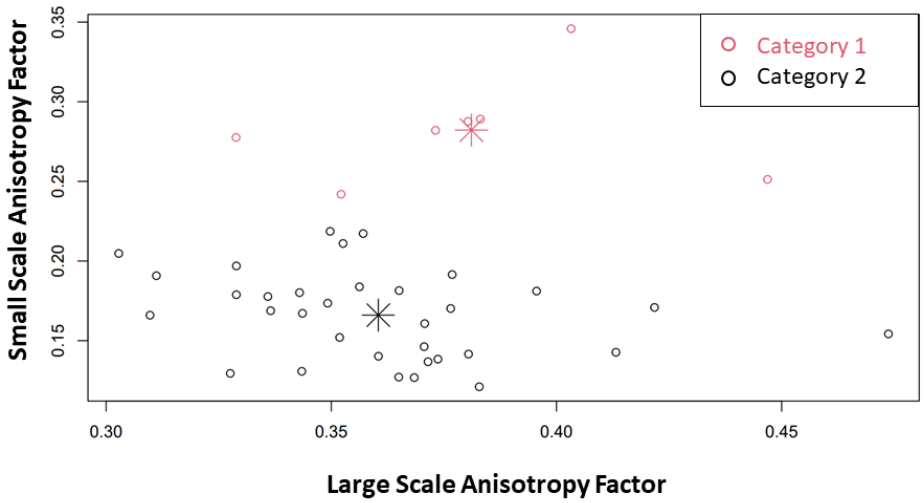


Figure 23: Plot of slides by small- and large-scale anisotropy factor categorized by nearest K-mean

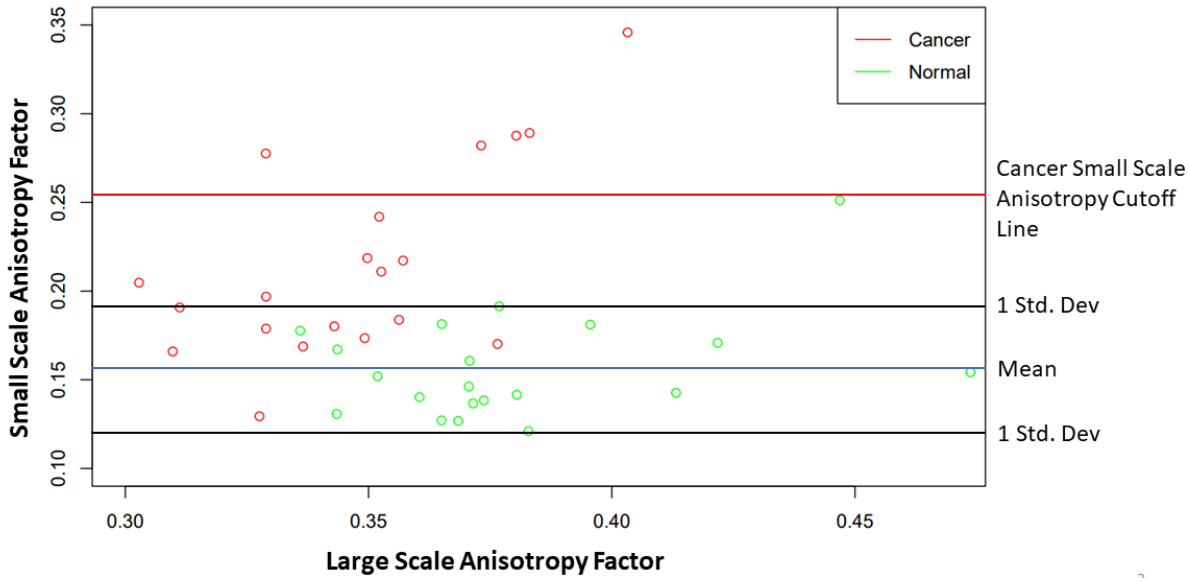


Figure 24: Plot of each cancer (red) and normal (green) slide by small- and large-scale anisotropy factor with a line (blue) at the mean small scale healthy anisotropy value. One standard deviation (black) away from the mean normal small scale anisotropy value and the cancer anisotropy factor cutoff value (red) are also shown with lines.

This seems to be caused by the group of high small scale anisotropy factors from the cancer slides. Since no healthy slides approach these values, the highest small-scale healthy anisotropy factor value was used as the cutoff point (Figure 24). This cut off point is statistically safe to use since the point's small-scale anisotropy factor is over 2.5 standard deviations away from the rest of the normal data and would be classified as an outlier using an outlier test. All slides with an anisotropy factor above the cutoff are categorized as cancer beforehand and removed from the K-means analysis. The analysis was re-run and the categorization capabilities greatly improved (Figure 25). Including the slides already categorized as

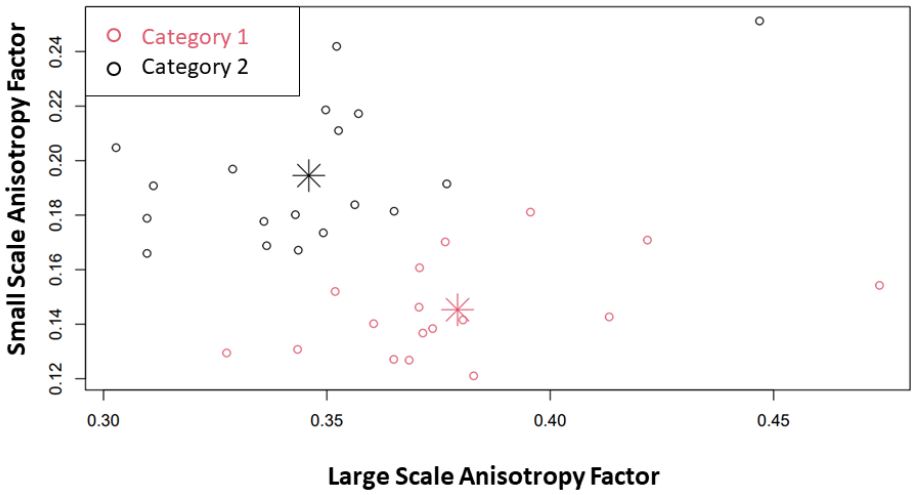


Figure 25: Plot of slides by small- and large-scale anisotropy factor categorized by nearest K-mean after removing cancer slides that passed the small-scale anisotropy cutoff.

cancer from the anisotropy factor cutoff, this method has 90% sensitivity and 75% specificity to diagnosing a slide as cancer. This is a promising result for use in the clinical environment as the metric is demonstrating good diagnostic sensitivity with decent specificity using metrics obtained from a small forty tissue sample dataset. Other clustering techniques should be explored and future work into what diagnostically relevant metrics to use will be explored in the future with a larger breast cancer dataset.

# CHAPTER 4

# COMPARISON TO OTHER SECOND HARMONIC GENERATION FIBER ANALYSIS METHODS

This chapter describes and applies two other computational second harmonic generation fiber analysis techniques to simulated images of fibers generated in MATLAB [32]. The calculation of the anisotropy factor on these simulated fibers using angle distribution outputs from these other analytical tools is compared to the 2D WTMM Anisotropy Method's anisotropy factor across different angle ranges and noise levels.

## Technique Comparison Across Different Angle Ranges

For tissue SHG analysis the tools commonly used by researchers are CT-FIRE [42] and OrientationJ [43]. CT-FIRE is a curvelet transform technique that also creates a fiber map. Its outputs include fiber diameter, number of fibers, fiber length, and fiber angle. The amount of information on collagen fibers specifically is superior to the current 2D WTMM Anisotropy Method, but it is not applicable to non-fiber images like 2D WTMM. OrientationJ uses a structure tensor sliding window approach to calculate the orientation, coherency, and energy of each pixel. For the comparison, only the fiber angle and pixel orientation information will be used to calculate an anisotropy factor in the same way as the WTMM anisotropy method. These comparisons will show if the WTMM anisotropy method is superior, comparable, or inferior to currently used SHG image analysis methods on simulated images.

For comparison of the ability of CT-FIRE, OrientationJ, and WTMM anisotropy method to differentiate fiber angle distributions from isotropic to anisotropic, 100 simulated fiber images (512x512 pixels) were generated at each of the following angle distributions 0-180°, 30-150°, 60-120°, and 85-95° (Figure 26) with a fiber diameter of two pixels. Other fiber diameters were generated and tested but results were similar to the two-pixel diameter results and not shown.

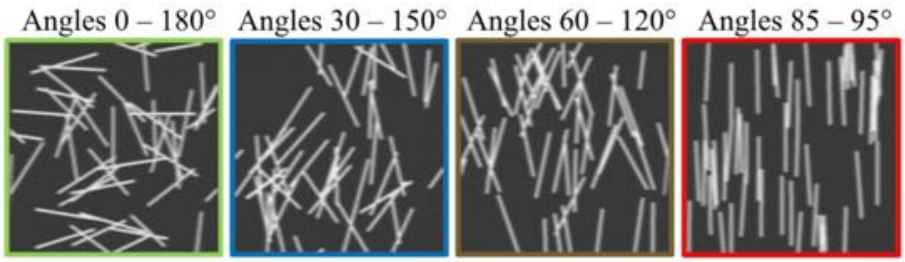


Figure 26: Examples of the simulated fiber images generated using an algorithm presented in reference 32, are shown here angle distributions ranging from 0-180°, 30-150°, 60-120°, and 85-95°

The 2D WTMM anisotropy method was applied using the variable binning technique without normalization and for all single scale comparisons, scale 25 was selected as for this data it had the greatest statistical significance. CT-FIRE was operated using its MATLAB UI by selecting images and then selecting the option to output fiber orientations into a csv file using default settings based on the fiber sizes present, which were then converted to probability density functions and used to calculate the anisotropy factor following the same methodology as the 2D WTMM Anisotropy Method. OrientationJ is a collection of FIJI plugins, and its distribution plugin was selected, and the cubic spline gradient was used. This outputs a pixel orientation histogram which was then converted to a probability density function and the anisotropy factor calculated. OrientationJ analysis was conducted at multiple structure tensor sizes with two pixels used as the first scale and increasing by powers of two until a size of 64 pixels was reached. The default tensor size OrientationJ suggests of four pixels was used for single scale comparisons. The image processing flowchart for both methods is shown in Figure 27.

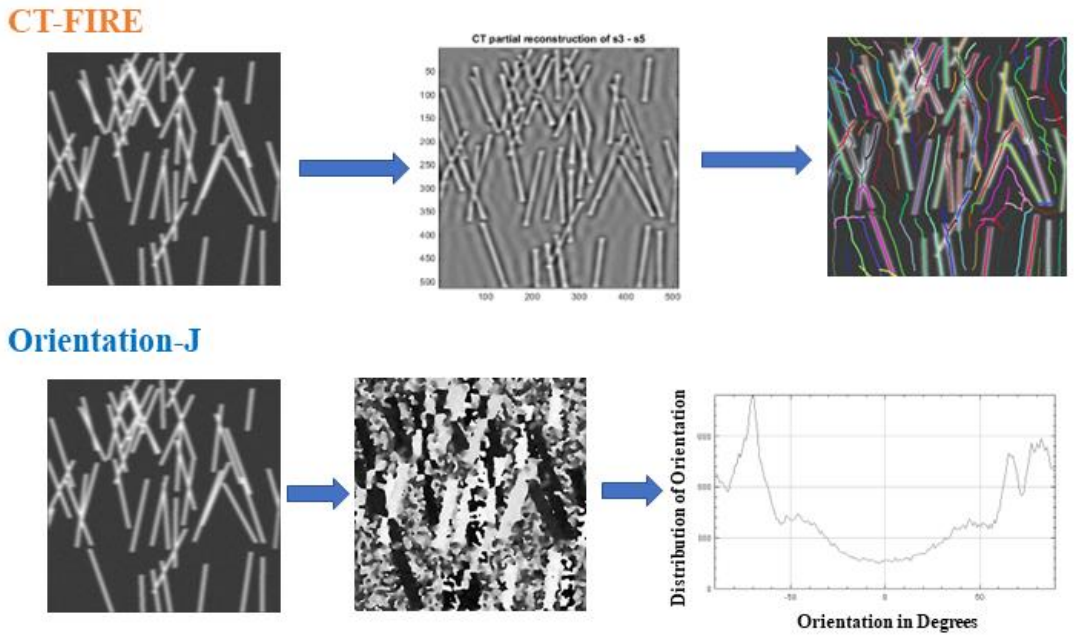


Figure 27: The CT-FIRE image processing pipeline (first row) with curvelet transform and fiber tracing is shown as well as the OrientationJ (second row) image processing pipeline with structure tensor calculated orientation map and orientation histogram.

The resulting 100 anisotropy factors for each angle distribution were box plotted by angle distribution and method (Figure 28). CT-FIRE and OrientationJ struggled to differentiate the two isotropic groups. CT-FIRE did not have a statistically significant difference using a Wilcoxon ranked sum test between the 0-180° and 30-150° angle distributions where OrientationJ did have a statistically significant result, but both have overlapping inter quartile ranges between the two isotropic fiber distributions. WTMM has statistically significant differences, and no overlapping inter quartile ranges across all four angle distributions. This result suggests that the 2D WTMM Anisotropy Method is more sensitive to subtle changes in image anisotropy of more isotropic fiber angle distributions.

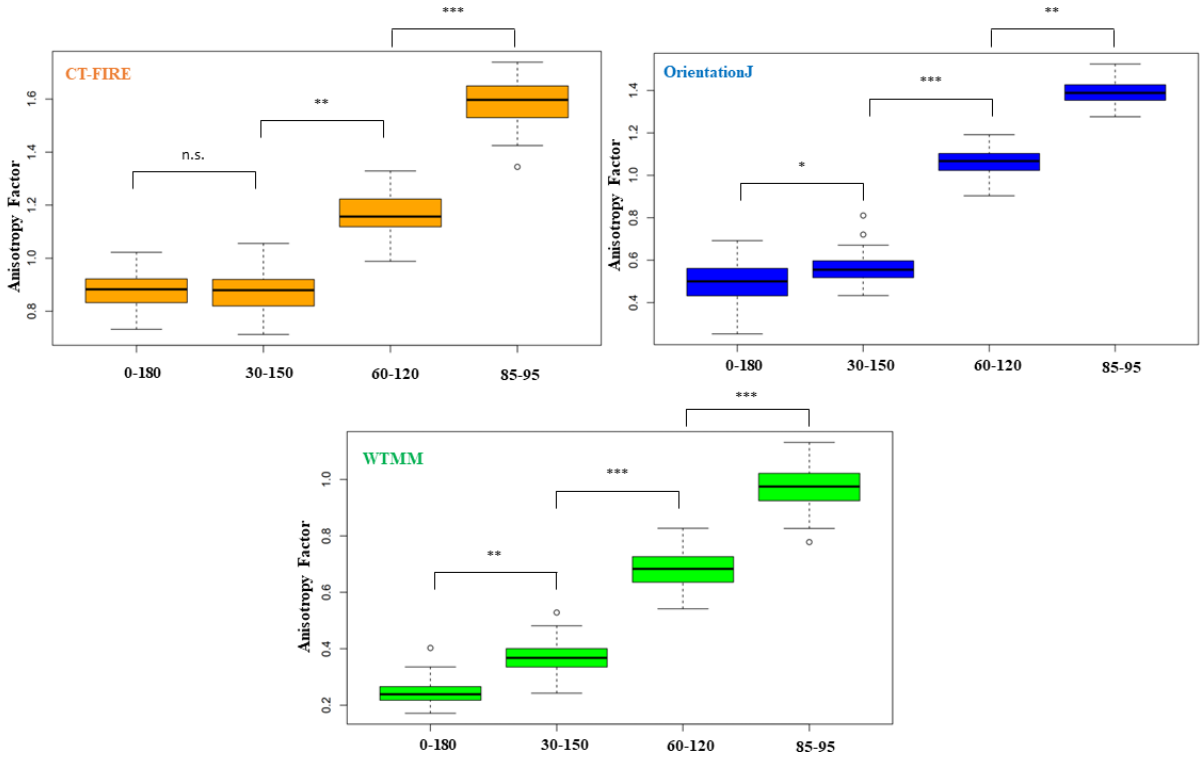


Figure 28: The anisotropy factor results from CT-FIRE (top left), OrientationJ (top right), and the 2D WTMM Anisotropy Method (bottom) are plotted by the angle distributions categories of 0-180°, 30-150°, 60-120°, and 85-95° with statistical significance shown.

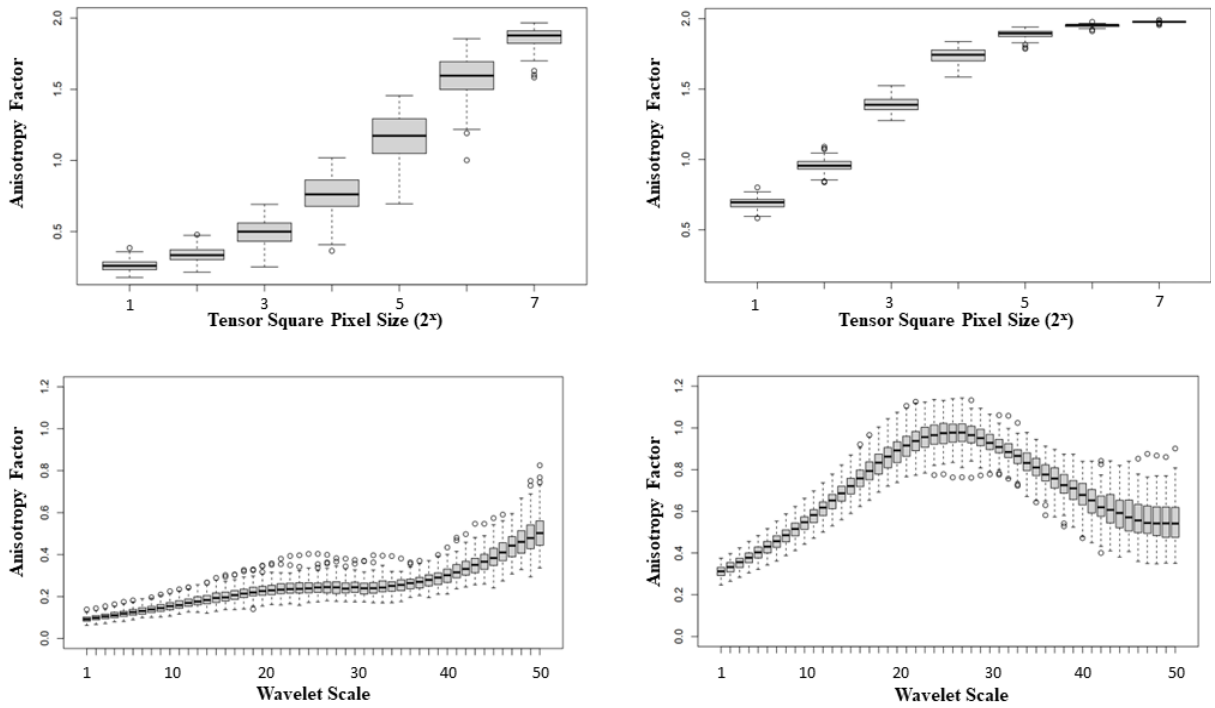


Figure 29: The multiscale anisotropy factor results from OrientationJ plotted by tensor size for the 0-180° (top left) and 85-95° (top right) angle distributions and the multiscale scale anisotropy factor results plotted by wavelet scale for the 0-180° (bottom left) and 85-95° (bottom right) angle distributions.

One key advantage of the 2D WTMM Anisotropy Method is its multiscale suite of information, and this multiscale information was compared to different tensor sizes in OrientationJ (Figure 29). Using the multiscale information from the WTMM anisotropy method, the 0-180° and 85-95° angle distributions are different at practically every wavelet size scale and also have different trends of anisotropy factor versus scale. OrientationJ's tensor size variation information also demonstrates differing trends between the 0-180° and 85-95° angle distributions, but the dynamic range is limited, approaching the maximum anisotropy factor value of two rapidly. This shows the superiority of the multiscale information that the 2D WTMM Anisotropy Method produces, allowing for a continuous range of comparisons to quantitatively determine structural differences between images.

**Technique Comparison Across Different Noise Levels**

This second experiment involves comparing the resistance to noise between the three different analysis methodologies. The 60-120° fiber images were added with four different levels of white noise generating five image categories consisting of the original image, noise level 1, noise level 2, noise level 3, and noise level 4 (Figure 30). The noise levels were simulated by first calculating the mean pixel value from the original simulated fiber image [32]. This mean value was then used to generate a white noise image with the same mean value, and this was added to the original image. Higher noise levels involve multiplying the white noise image by an integer before adding to the

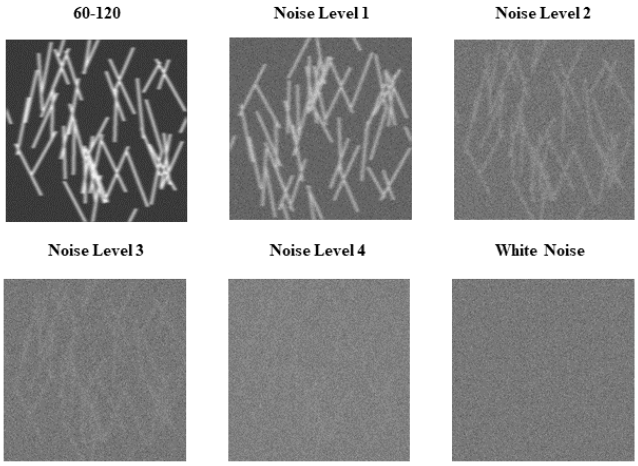


Figure 30: Example of 60-180° simulated fiber image with four added noise levels and an example white noise image.

original simulated fiber image. The multiples for noise levels one through four are 1, 5, 10, and 20. These were analyzed following the same image processing pipeline and anisotropy factor calculation as performed in the different angle distribution experiment described in the last section. A set of 100 pure white noise images (512x512 pixels) were also generated and anisotropy factors calculated following the same processing as the five image categories. These were used as a comparator to determine if the method could differentiate the white noise from the fiber images with noise added. To compare to white noise and to normalize anisotropy factor across methods, each image's anisotropy factor was divided by the median white noise anisotropy factor from its respective analysis technique. The closer the value is to one, the more similar it is to the white noise anisotropy factor. A Wilcoxon ranked sum test was also conducted for each image category's non-modified anisotropy factor distribution to the respective white noise anisotropy factor distribution. The white noise modified anisotropy factors were then plotted by noise level and color coded by analysis method with a line at an anisotropy factor of one to highlight similarity to white noise (Figure 31). Each boxplot with an asterisk above is statistically significantly different from its respective white noise distribution from a Wilcoxon ranked sum test.

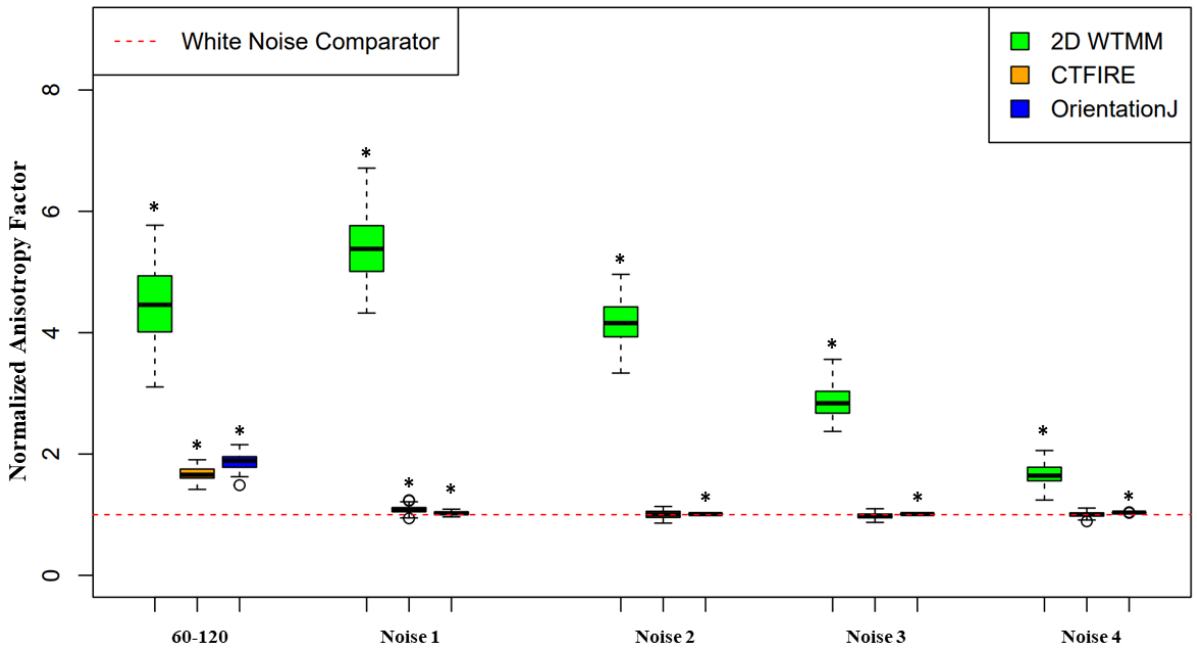


Figure 31: The anisotropy factor results from 2D WTMM anisotropy method (green), OrientationJ (blue), and CT-FIRE (orange), normalized by the respective methods white noise anisotropy factor results, applied on 100 60-120° simulated fiber images and the 100 images for each of the four added noise levels. The red dashed line is at an anisotropy factor of one and statistical significance from the methods respective white noise results are shown with an asterix.

Observing the 2D WTMM Anisotropy Method results it is clear that the method is extremely resistance to noise. The first two noise levels still have IQRs that overlap with the original images' anisotropy factor IQR. None of the noise levels have any data points that overlap with the white noise comparator. CT-FIRE barely survives the first noise level, but its anisotropy factor values are no longer similar to the non-noisy image. Past the first noise level it fails a Wilcoxon ranked sum test with its respective white noise anisotropy factor distribution and can no longer differentiate the noisy data from pure noise. OrientationJ has some interesting results, where similarly to CT-FIRE it survives the first noise level but is no longer similar to the original non-noisy image, but it maintains a statistically significant difference from white noise at all noise levels. This seems to be due to an artefact in the OrientationJ results. The collapsed boxplots, with no whiskers, from the higher-level noise plots is because all of the 100 images result in the same exact anisotropy factor value. This is not true for the pure white noise images who still maintain a distribution of values, and this is the source of the statistical significance. Why this occurs with OrientationJ is still unknown but is an interesting problem to be explored. Overall, these results show that the 2D WTMM Anisotropy Method is far more resilient to noise than current methods used in the field.

# CHAPTER 5

# FUTURE OUTLOOKS AND CONCLUSION

This chapter covers future areas of research for the 2D WTMM Anisotropy Method that are either currently underway or will soon be under way for usage of the method as a cancer diagnostics and research tool. These include an ongoing breast cancer study, a multivariate combination with other 2D WTMM techniques for diagnostics, and the expansion of the method to 2.5/3D for future research applications.

## Breast Cancer Study

A further application of the updated 2D WTMM Anisotropy Method is currently underway, still through a collaboration with MHRI. There are 67 breast cancer and benign slides from the MHRI. The proposed study is as follows: each slide will have eight forward SHG images acquired using a custom-built 2-photon microscope. The microscope uses a Ti:Sapphire femtosecond pulsed laser tuned to 890nm, with a 40x 0.8 NA water immersion objective and a 0.9 NA condenser. The SHG emission will be filtered through a bandpass filter (445/20nm) and detected using a 7421 Hamamatsu PMT. Images will be line scanned to generate a 512 by 512-pixel image with 0.35 μm a pixel. These SHG images will then be process through the updated 2D WTMM anisotropy method to calculate the multiscale anisotropy factors. Each slide will also be slide-scanned using a Cytation 5's brightfield slide scanning mode. A portion of these slide scans will be used as a training dataset for the machine learning algorithm. The rest will have areas on the slide labeled as cancer or benign from the algorithm, allowing for a ratio to be produced for each slide. Lastly, each slide has a matched set of mammograms that will be run through the 2D WTMM sliding-window method. This method will output which areas are dense, fatty, and disrupted tissue. From earlier CompuMAINE studies the ratio of disrupted over fatty tissue can be used

as a metric for breast cancer diagnosis [43]. These datasets will be combined to create a multiparametric diagnostics tool that will be stronger than the sum of its parts.

Currently, a majority of the slides (61) have been imaged by the two-photon microscope and all slides have been slide scanned, but none have been annotated by a pathologist. Pathology reports for each deidentified patient number have been received and used for preliminary analysis from the already gathered data. There have been two preliminary analyses conducted on SHG images using the 2D WTMM Anisotropy Method. The first study used the BIRADs assessment score from the patient's mammogram results. One patient selected had a BIRADs score of one, which means no cancer from the clinical mammogram results. Two of the patients selected had a score of 4 which is indicative of a 20 to 35% chance of cancer. The last two patients had a score of 5 which is a 95% chance of cancer. The eight SHG images from each patient were analyzed using the 2D WTMM Anisotropy Method and the anisotropy factor at each scale was averaged across all images in that category (BIRADS 1, 4, or 5). There

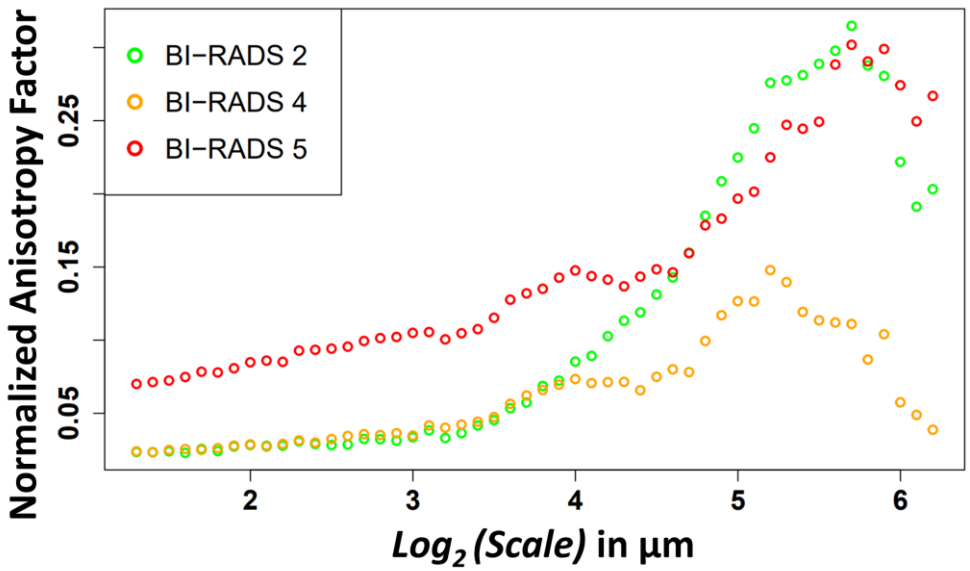


Figure 32: Median anisotropy factor across all images for BIRADs 2 (green), BIRADS 4 (orange), and BIRADS 5(red) normalized by white noise anisotropy factor from ~2.5 µm to 73.5 µm.

are clear trends apparent from this preliminary five slide dataset (Figure 32). At smaller wavelet scales both BI-RADS 2 and BI-RADS 4 are statistically significantly different from BI-RADS 5, but not each other. At larger wavelet scales this relationship swaps with BI-RADS 5 and BI-RADS 2 becoming statistically significantly different from BI-RADS 4, but not each other. It is a promising to see the 2D WTMM Anisotropy method trend with mammogram clinical data and a larger study will be conducted when all images are processed.

The second preliminary study involved comparing six patients who received a benign diagnosis versus six patients who received a cancer diagnosis. This was conducted using the same method as the previous preliminary study by using the 2D WTMM Anisotropy Method on the eight collected SHG images per a patient. The first plot generated was by averaging all patient's SHG image anisotropy factors in a category by scale (Figure 33). There was statistically significant difference using a Wilcoxon

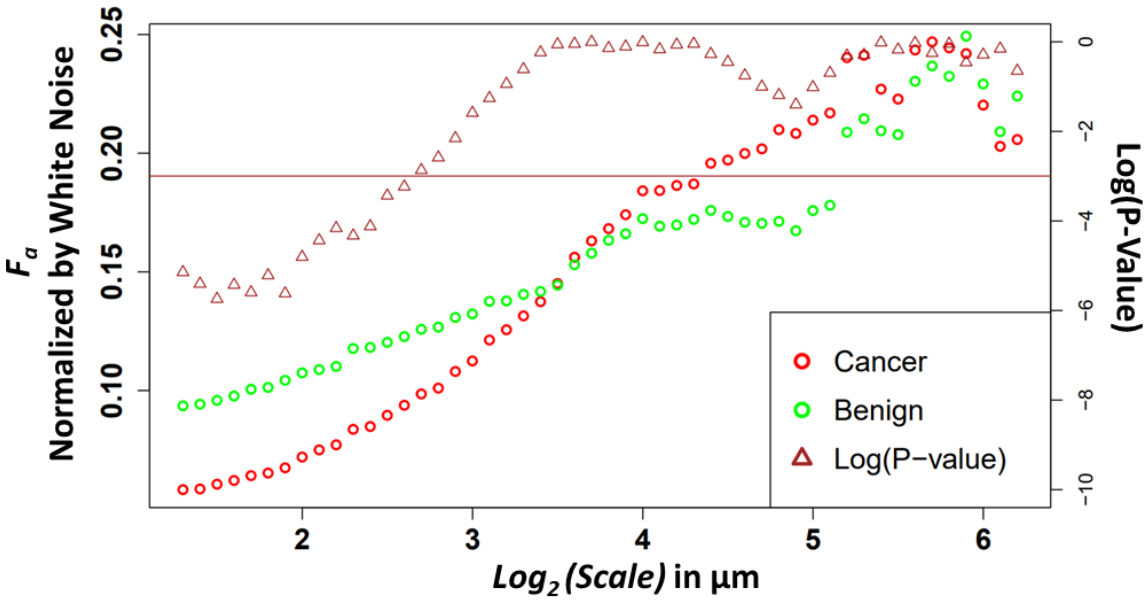


Figure 33: Median anisotropy factor across all images for benign (green) and cancer (red) normalized by white noise anisotropy factor from ~2.5 µm to 73.5 µm. The *p*-values (right *y*-axis), shown in brown, from a Wilcoxon rank sum test ran at each scale between the tissue pair's anisotropy factors. Values below the horizontal brown line at 0.05 show statistical significance.

ranked sum test at the first 14 wavelet scales. In this dataset, the benign has a higher anisotropy factor than cancer at smaller scales. This trend is opposite to the trend observed in the BIRADs data where a score of five had a higher anisotropy factor than the score two patient. Too delve deeper into this discrepancy, each patient was graphed separately and colored by their category (Figure 34). Interestingly, there seems to be two cancer trends particularly at the middle to larger scales where they group above or below the benign data. The smaller BIRADs study may be affected by this sampling of two different cancer groups as it has under half the patients than the 12-patient study. A deeper exploration of these trends by patient will be explored when all 67 patients' images are processed.

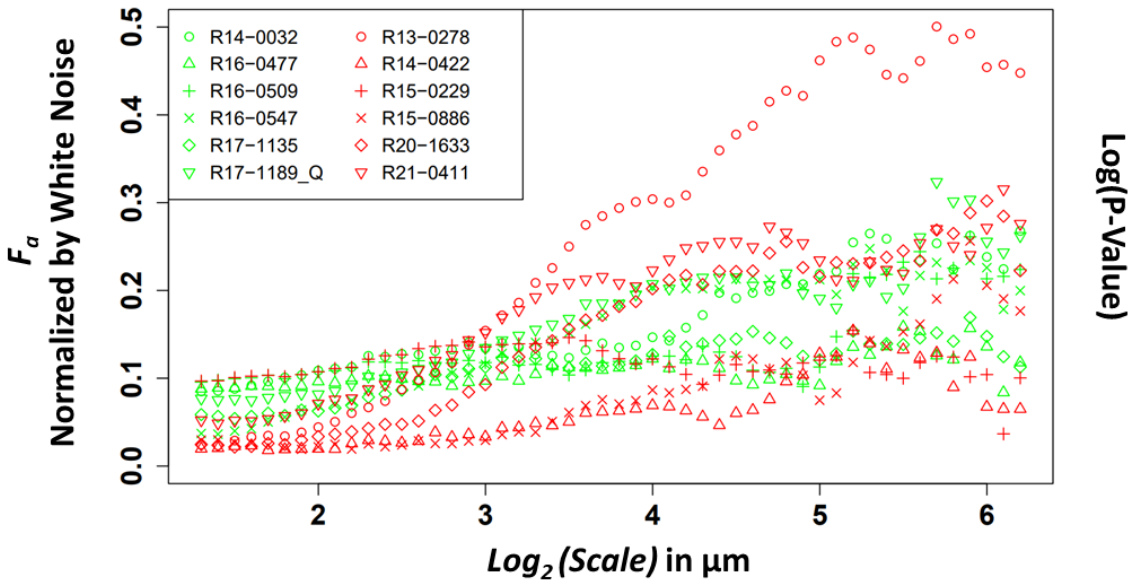


Figure 34: Median anisotropy factor across all patients for benign (green) and cancer (red) normalized by white noise anisotropy factor from ~2.5 µm to 73.5 µm. The *p*-values (right *y*-axis), shown in brown, from a Wilcoxon rank sum test ran at each scale between the tissue pair's anisotropy factors. Values below the horizontal brown line at 0.05 show statistical significance.

**Future Projects**

Jeremy Juybari, a PhD student in the CompuMAINE lab at the University of Maine, is leading the machine learning application to the bright field images of the H&E-stained slides. This is already underway for the PDAC slides where the annotations are used to generate a ground truth dataset. The model accepts multiple image inputs to generate a final overlay of cancer versus normal tissue. Currently these multiple inputs are different resolutions of the H&E-stained slides to mimic a multiscale approach. After a brainstorming session between Jeremy and I, we thought of using an SHG imaged that is matched to the same x-y location as the brightfield image as another input into the model to create the overlay. There could even be wavelet convolved SHG images to mimic the scaling effect of the bright field resolution. This machine learning approach when combined with the anisotropy analysis could create a powerful two variable analysis tool for cancer diagnostics. A preliminary literature review found only one paper similar to this idea published in nature [44]. They had SR-stained slides of rat liver fibrosis that they captured both traditional bright field and SHG x-y images of the same exact area. The brightfield were annotated by a pathologist and these annotations were used to generate ground truths in the SHG images. The machine learning was only conducted with SHG images as inputs where here we are suggesting using the brightfield as inputs as well. The two-photon microscopy system used in this study and under purview of Dr. Karissa Tilbury at the University of Maine is currently receiving upgrades to perform x-y scanning. Once this is complete this study can move forth using the same 67 slides for the breast cancer study mentioned above.

Another avenue of exploration is to adapt the 2D WTMM Anisotropy Method for analysis of breast cancer spheroids that Dr. Karissa Tilbury will soon be generating. This will remove the limitation of our current datasets consisting of 2D tissue slices. This allows for the capability of adapting the method into more dimensions. Currently, methods for 3D WTMM anisotropy analysis and tests for the

methods are being brainstormed by me, Dr. Andre Khalil, and Dr. Karissa Tilbury. The current idea is that a 2.5D anisotropy factor generation could calculate in two ways. First an SHG z-stack of the 3D structure of interest is captured. Each slice could be used to calculate an anisotropy factor and anisotropy factors across the z-stack can be compared. In the end, a summation of the entire stack could be used to generate a spheroid's total anisotropy factor. Another approach that will be compared to the prior is first taking a max projection of the z-stack and running the 2D WTMM Anisotropy Method on the max projection saving computational time. A third approach is by developing a 3D WTMM Anisotropy Method. The key issue will be from the point spread function (PSF) of the microscope as the axial versus lateral resolution will not be the same. This will generate a 3D image without dimensionally symmetric voxels. Multiple ideas have been proposed to tackle this issue. Simulated 3D fiber SHG images can be generated and then convolved with a model point spread function. Analysis can then be conducted between the two and the effect of the PSF on the 3D anisotropy factor can be understood. If the effect is too great, there are a couple suggested methods to attempt to fix this. One involves simply using wavelets at a scale where the axial versus lateral resolution difference is completely convolved out. Another could be to use a wavelet that is also skewed in the lateral direction and determine if that corrects for the effect of the PSF. Lastly, fully describing the PSF of the microscope we are using and algorithmically removing the PSF thereby generating the 3D SHG images. A comparison of the three methods (2D, 2.5D, 3D) is shown and work is underway to obtain funding for this project (Figure 35).

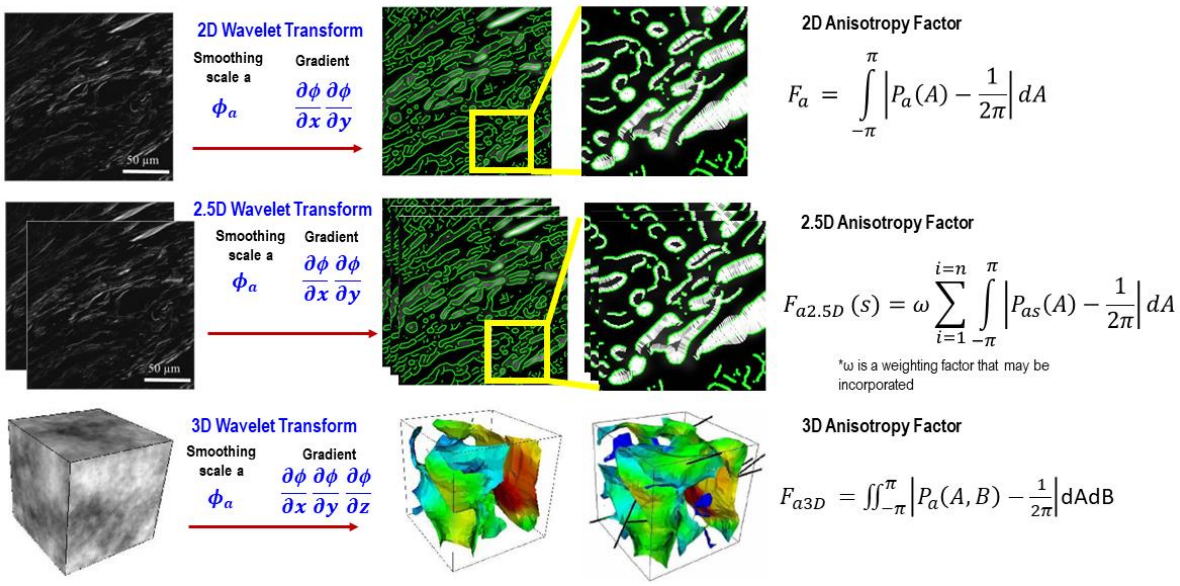


Figure 35: First row is current 2D WTMM anisotropy process, second row is proposed 2.5 process, and bottom row is 3D process. Green outlines in rows 1 and 2 represent local maximum gradients and can be used for image segmentation. White arrows indicate the direction in which the gradients are maximum and are used to calculate the anisotropy factor. In the 3rd row, surfaces are local maximum detecting edges, black lines correspond to local maxima.

**Conclusion**

As can be seen, there is still a lot of work to be done involving the combination of the 2D WTMM Anisotropy method and SHG two-photon microscopy. The method has been shown to be sensitive to collagen changes induced by cancer and now larger studies are underway to use this tool for morphological quantification in answering questions pertaining to breast cancer. This tool is also not limited to cancer, but any study involving collagen organization. The 2D WTMM Anisotropy Method itself can also be applied to any image where the quantification of directional organization is needed. The updates performed on this method during this thesis work will hopefully help many researchers within the scientific community.

**BIBLIOGRAPHY**


[1] K. Y. Elbanna, H. J. Jang, and T. K. Kim, "Imaging diagnosis and staging of pancreatic ductal adenocarcinoma: a comprehensive review," *Insights Imaging*, vol. 11, no. 1, 2020, doi: 10.1186/s13244-020-00861-y.

[2] S. Iranmakani *et al.*, "A review of various modalities in breast imaging: technical aspects and clinical outcomes," *Egypt. J. Radiol. Nucl. Med.*, vol. 51, no. 1, 2020, doi: 10.1186/s43055-020-00175-5.

[3] S. Sarkar and S. Das, "A Review of Imaging Methods for Prostate Cancer Detection," *Biomed. Eng. Comput. Biol.*, vol. 7s1, p. BECB.S34255, 2016, doi: 10.4137/becb.s34255.

[4] J. Shiraishi, Q. Li, D. Appelbaum, and K. Doi, "Computer-aided diagnosis and artificial intelligence in clinical imaging," *Semin. Nucl. Med.*, vol. 41, no. 6, pp. 449–462, 2011, doi: 10.1053/j.semnuclmed.2011.06.004.

[5] D. M. Ikeda, "Mainstream breast cancer radiology perspective," *Phys. Medica*, vol. 21, no. SUPPL. 1, pp. 4–6, 2006, doi: 10.1016/S1120-1797(06)80013-X.

[6] SEERS, "Cancer Stat Facts: Female Breast Cancer." https://seer.cancer.gov/statfacts/html/breast.html.

[7] Cancer Research UK, "Why is Early Diagnosis Important," https://www.cancerresearchuk.org/about-cancer/cancer-symptoms/why-is-early-diagnosis-important.

[8] E. Majestic, "Preventing chronic disease.," *Health Aff. (Millwood).*, vol. 22, no. 3, p. 253, 2003, doi: 10.1377/hlthaff.22.3.253.

[9] S. A. Kennedy, L. Milovanovic, and M. Midia, "Major bleeding after percutaneous image-guided biopsies: Frequency, predictors, and periprocedural management," *Semin. Intervent. Radiol.*, vol. 32, no. 1, pp. 26–33, 2015, doi: 10.1055/s-0034-1396961.

[10] K. Shyamala, H. Girish, and S. Murgod, "Risk of tumor cell seeding through biopsy and aspiration cytology," *J. Int. Soc. Prev. Community Dent.*, vol. 4, no. 1, pp. 5–11, 2014, doi: 10.4103/2231-0762.129446.

[11] R. A. Smith *et al.*, "Cancer screening in the United States, 2018: A review of current American Cancer Society guidelines and current issues in cancer screening," *CA. Cancer J. Clin.*, vol. 68, no. 4, pp. 297–316, 2018, doi: 10.3322/caac.21446.

[12] Centers for Disease Control and Prevention, "Basic Information About Breast Camcer." https://www.cdc.gov/cancer/breast/basic_info/.

[13] A. Costa *et al.*, "Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer," *Cancer Cell*, vol. 33, no. 3, pp. 463-479.e10, 2018, doi: 10.1016/j.ccell.2018.01.011.

[14] Calvo F. et al.; "Mechano-transduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer associated fibroblasts;" Nat. Cell Biol.; vol. 15 no. 6; 2013.

[15] B. C. Özdemir *et al.*, “Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival,” *Cancer Cell*, vol. 25, no. 6, pp. 719–734, 2014, doi: 10.1016/j.ccr.2014.04.005.

[16] B. Piersma, M. K. Hayward, and V. M. Weaver, “Fibrosis and cancer: A strained relationship,” *Biochim. Biophys. Acta - Rev. Cancer*, vol. 1873, no. 2, p. 188356, 2020, doi: 10.1016/j.bbcan.2020.188356.

[17] D. S. James and P. J. Campagnola, “Recent Advancements in Optical Harmonic Generation Microscopy: Applications and Perspectives,” *BME Front.*, vol. 2021, pp. 1–24, 2021, doi: 10.34133/2021/3973857.

[18] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, “Interactions between light waves in a nonlinear dielectric,” *Phys. Rev.*, vol. 127, no. 6, pp. 1918–1939, 1962, doi: 10.1103/PhysRev.127.1918.

[19] C. T. Rueden, M. W. Conklin, P. P. Provenzano, P. J. Keely, and K. W. Eliceiri, “Nonlinear optical microscopy and computational analysis of intrinsic signatures in breast cancer,” *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009*, pp. 4077–4080, 2009, doi: 10.1109/IEMBS.2009.5334523.

[20] J. S. Bredfeldt, Y. Liu, M. W. Conklin, P. J. Keely, T. R. Mackie, and K. W. Eliceiri, “Automated quantification of aligned collagen for human breast carcinoma prognosis,” *J. Pathol. Inform.*, vol. 5, no. 1, p. 28, 2014, doi: 10.4103/2153-3539.139707.

[21] V. Parodi, E. Jacchetti, R. Osellame, G. Cerullo, D. Polli, and M. T. Raimondi, “Nonlinear Optical Microscopy: From Fundamentals to Applications in Live Bioimaging,” *Front. Bioeng. Biotechnol.*, vol. 8, no. October, pp. 1–18, 2020, doi: 10.3389/fbioe.2020.585363.

[22] K. R. Campbell *et al.*, “3D second harmonic generation imaging tomography by multi-view excitation,” *Optica*, vol. 4, no. 10, p. 1171, 2017, doi: 10.1364/optica.4.001171.

[23] M. Rivard *et al.*, “The structural origin of second harmonic generation in fascia,” *Biomed. Opt. Express*, vol. 2, no. 1, p. 26, 2011, doi: 10.1364/boe.2.000026.

[24] O. Nadiarnykh, R. B. LaComb, P. J. Campagnola, and W. A. Mohler, “Coherent and incoherent SHG in fibrillar cellulose matrices,” *Opt. Express*, vol. 15, no. 6, p. 3348, 2007, doi: 10.1364/oe.15.003348.

[25] A. Khalil, G. Joncas, F. Nekka, P. Kestener, and A. Arneodo, “ Morphological Analysis of H i Features. II. Wavelet-based Multifractal Formalism ,” *Astrophys. J. Suppl. Ser.*, vol. 165, no. 2, pp. 512–550, 2006, doi: 10.1086/505144.

[26] C. J. Snow *et al.*, “Time-lapse analysis and mathematical characterization elucidate novel mechanisms underlying muscle morphogenesis,” *PLoS Genet.*, vol. 4, no. 10, 2008, doi: 10.1371/journal.pgen.1000219.

[27] C. J. Snow, M. T. Peterson, A. Khalil, and C. A. Henry, “Muscle development is disrupted in zebrafish embryos deficient for fibronectin,” *Dev. Dyn.*, vol. 237, no. 9, pp. 2542–2553, 2008, doi: 10.1002/dvdy.21670.

[28] M. F. Goody, M. W. Kelly, K. N. Lessard, A. Khalil, and C. A. Henry, “Nrk2b-mediated NAD+ production regulates cell adhesion and is required for muscle morphogenesis in vivo. Nrk2b and NAD+ in muscle morphogenesis,” *Dev. Biol.*, vol. 344, no. 2, pp. 809–826, 2010, doi: 10.1016/j.ydbio.2010.05.513.

[29] M. F. Goody, M. W. Kelly, C. J. Reynolds, A. Khalil, B. D. Crawford, and C. A. Henry, “NAD+ Biosynthesis Ameliorates a Zebrafish Model of Muscular Dystrophy,” *PLoS Biol.*, vol. 10, no. 10, 2012, doi: 10.1371/journal.pbio.1001409.

[30] E. C. Bailey et al., “NAD+ improves neuromuscular development in a zebrafish model of FKRP-associated dystroglycanopathy,” Skelet Muscle, vol. 9 no. 1, 2019.

[31] A. Khalil *et al.*, “Image analysis of soft-tissue in-growth and attachment into highly porous alumina ceramic foam metals,” *Med. Eng. Phys.*, vol. 31, no. 7, pp. 775–783, 2009, doi: 10.1016/j.medengphy.2009.02.007.

[32] K. Tilbury, X. Han, P. C. Brooks, and A. Khalil, “Multiscale anisotropy analysis of second-harmonic generation collagen imaging of mouse skin,” *J. Biomed. Opt.*, vol. 26, no. 06, pp. 1–19, 2021, doi: 10.1117/1.jbo.26.6.065002.

[33] A. Khalil, J. L. Grant, L. B. Caddle, E. Atzema, K. D. Mills, and A. Arneodo, “Chromosome territories have a highly nonspherical morphology and nonrandom positioning,” *Chromosom. Res.*, vol. 15, no. 7, pp. 899–916, 2007, doi: 10.1007/s10577-007-1172-8.

[34] A. Arnéodo, N. Decoster, and S. G. Roux, “A wavelet-based method for multifractal image analysis. I. Methodology and test applications on isotropic and anisotropic random rough surfaces,” *Eur. Phys. J. B*, vol. 15, no. 3, pp. 567–600, 2000, doi: 10.1007/s100510051161.

[35] Freedman, D., Diaconis, P. On the histogram as a density estimator:L 2 theory. Z. Wahrscheinlichkeitstheorie verw Gebiete 57, 453–476 (1981). https://doi.org/10.1007/BF01025868

[36] Herbert A. Sturges (1926) The Choice of a Class Interval, Journal of the American Statistical Association, 21:153, 65-66, DOI: 10.1080/01621459.1926.10502161

[37] National Institute of Standards and Technology, “Chi-square Goodness-of-fit Test.” https://itl.nist.gov/div898/handbook/prc/section2/prc211.htm.

[38] M. V. Apte, R. C. Pirola, and J. S. Wilson, “Pancreatic stellate cells: A starring role in normal and diseased pancreas,” *Front. Physiol.*, vol. 3 AUG, no. August, pp. 1–14, 2012, doi: 10.3389/fphys.2012.00344.

[39] Mouselimis L (2022). ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering. R package version 1.2.6, https://CRAN.R-project.org/package=ClusterR.

[40] Struyf A, Hubert M, Rousseeuw P (1997). “Clustering in an Object-Oriented Environment.” Journal of Statistical Software. doi: 10.18637/jss.v001.i04.

[41] Pham D, Dimov S, Nguyen C (2004). “Selection of K in K-means clustering.” Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. doi: 10.1243/095440605X8298, https://core.ac.uk/download/pdf/205410925.pdf.

[42] J. S. Bredfeldt, Y. Liu, C. A. Pehlke, M. W. Conklin, J. M. Szulczewski, D. R. Inman, P. J. Keely, R. D. Nowak, T. R. Mackie, and K. W. Eliceiri, “Computational segmentation of collagen fibers from second-harmonic generation images of breast cancer,” J. Biomed. Opt. 19, 016007–016007 (2014)

[42] R. Rezakhaniha, A. Agianniotis, J.T.C. Schrauwen, A. Griffa, D. Sage, C.V.C. Bouten, F.N. van de Vosse, M. Unser, N. Stergiopulos Experimental Investigation of Collagen Waviness and Orientation in the Arterial Adventitia Biomechanics and Modeling in Mechanobiology, vol. 11, 2012.

[43] Gerasimova-Chechkina, E., Toner, B. C., Batchelder, K. A., White, B., Freynd, G., Antipev, I., Arneodo, A., & Khalil, A. (2021). Loss of Mammographic Tissue Homeostasis in Invasive Lobular and Ductal Breast Carcinomas vs. Benign Lesions. Frontiers in physiology, 12, 660883. https://doi.org/10.3389/fphys.2021.660883

[44] Y. Yu *et al.*, “Deep learning enables automated scoring of liver fibrosis stages,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-34300-2.

**APPENDIX: Mathematical Proof of Anisotropy Factor Upper Limit**

For any scale $> 0, P_a(A)$ (where $A \in [-\pi, \pi]$ is an angle in radians) is a normalized probability distribution function, i. e.

$$\int_{-\pi}^{\pi} P_a(A) dA = 1 \tag{6}$$

The anisotropy factor, $F_a$, is defined as

$$F_a = \int_{-\pi}^{\pi} |P_a(A) - \frac{1}{2\pi}| dA \tag{7}$$

and the theoretical upper bound of $F_a$ can be determined using this conservative inequality

$$|x - y| \leq |x| + |y|. \tag{8}$$

When the inequality is applied to $F_a$, it follows as shown,

$$\begin{aligned} F_a = \int_{-\pi}^{\pi} |P_a(A) - \tfrac{1}{2\pi}| dA &\leq \int_{-\pi}^{\pi} \left( |P_a| + \left|\tfrac{1}{2\pi}\right| \right) dA \\ &\leq \int_{-\pi}^{\pi} P_a(A) dA + \int_{-\pi}^{\pi} \left|\tfrac{1}{2\pi}\right| dA \\ &= 1 + 1 = 2. \end{aligned} \tag{9}$$

## BIOGRAPHY OF THE AUTHOR

Josh Hamilton was born in Milo, Maine on December 18, 1998. He was raised in Alton, Maine and graduated from Old Town High School in 2017. He attended the University of Maine and graduated in the spring of 2021 with a bachelor's degree in Biomedical Engineering. He continued his education in a 4+1 M.S. program at the University of Maine in Biomedical Engineering in the summer of 2021. After receiving his degree, Josh will be continuing his work by pursuing a PhD in Biomedical Engineering through the University of Maine's Graduate School of Biomedical Science and Engineering. He hopes to one day become a professor to pursue both his passions of research and teaching. Josh is a candidate for the Master of Science degree in Biomedical Engineering from the University of Maine in August 2022.