

2016

## Imaging ductal carcinoma using a hyperspectral imaging system

Yasser M. Khouj

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Khouj, Yasser M., "Imaging ductal carcinoma using a hyperspectral imaging system" (2016). *Graduate Theses, Dissertations, and Problem Reports*. 5970.

<https://researchrepository.wvu.edu/etd/5970>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

# **Imaging Ductal Carcinoma Using a Hyperspectral Imaging System**

**Yasser M. Khouj**

**Thesis submitted to the  
Benjamin M. Statler College of Engineering and Mineral Resources  
at West Virginia University  
in partial fulfillment of the requirements  
for the degree of  
Masters in Science  
in  
Electrical Engineering**

**Jeremy M. Dawson, Ph.D., Chair**

**Yuxin Liu, Ph.D.**

**Linda Vona-Davis, Ph.D.**

**Lane Department of Computer Science and Electrical Engineering**

**Morgantown, West Virginia**

**2016**

**Keywords: Hyperspectral, Breast Cancer, Ductal Carcinoma, Spectral Reflectance, H&E,  
Unstained, K-means.**

**Copyright 2016 Yasser M. Khouj**

# Abstract

## Imaging Ductal Carcinoma Using Hyperspectral Imaging System

Yasser M. Khouj

Hyperspectral Imaging (HSI) is a non-invasive optical imaging modality that shows the potential to aid pathologists in breast cancer diagnoses cases. In this study, breast cancer tissues from different patients were imaged by a hyperspectral system to detect spectral differences between normal and breast cancer tissues, as well as early and late stages of breast cancer. If the spectral differences in these tissue types can be measured, automated systems can be developed to help the pathologist identify suspect biopsy samples, which will improve sample throughput and assist in making critical treatment decisions. Tissue samples from ten different patients were provided by the WVU Pathology Department. The samples from each patient included both normal and ductal carcinoma tissue, both stained and unstained. These cells were imaged using a snapshot HSI system, and the spectral reflectances were evaluated to see if there was a measurable spectral difference between the various cell types. Analysis of the spectral reflectance values indicated that wavelengths near 550nm show the best differentiation between tissue types. This information was used to train image processing algorithms using supervised and unsupervised data. K-Means and Support Vector Machine (SVM) approaches were applied to the hyperspectral data cubes, and successfully detected spectral tissue differences with sensitivity of 85.45%, and specificity of 94.64% with TNR of 95.8%, and FPR of 4.2%. These results were verified by ground truth marking of the tissue samples by a pathologist. This interdisciplinary work will build a bridge between pathology and hyperspectral optical diagnostic imaging in order to reduce time and workload on the pathologist, which can lead to benefit of lead reducing time, and increasing the accuracy of diagnoses.

# Acknowledgments

**My deepest gratitude and sincere to God for all the love, guidance, patience, and passion I was blessed with that helped me complete this work and through all my life.**

**I would like to thank my committee members starting with my academic and research advisor Dr. Jeremy Dawson, for giving me the opportunity to be one of his graduate research assistants, I thank him for his guidance, patience, encouragement, and the unlimited support. I also extend my appreciation, and gratefulness to Dr. Linda Vona-Davis who was generous enough to add me to Davis lab team to be one of WVU cancer institute researchers, and her undeniable help and support through my research journey, and Dr. James Coad for providing the tissue samples. I would also like to thank Dr. Yuxin Liu who has given me lots of support and advice in the field of bioengineering and biomedical engineering, since my first semester in WVU**

**It is also my honor to dedicate this work to my beloved family starting with my father Mohammed Khouj, my mother Wafa Khouja for their love, support, and patience through all my life. I dedicate this work to my other half and best friend Dr. Gofran Tarabulsi, and my man, the great Bakri Khouj (KJ) for being there for me every second in my life. I extend my love and appreciation to my brothers Saleh, Waseem and, my sister Afnan.**

**I thank WVU faculty members for helping me through my grad school. My colleagues, lab mates from Dawson lab, Davis lab, and friends for their support, and encouragement. Anand Kadiyala, Christo Robinson, Sinan Sabri, Rania Almohsen, Mohammed Jameel, Dustin Poster, Kyle Smith, Seth Leffel, Akshitha Gadde, Matt Paychol, Rupa, Loukhya, Kenney Hite, Laurel Pell, Kelsey Sadlek, Yuya Kodo, Ahmed Marah, and my best friend Yasser Al-Shehri, and all my friends.**



# Table of Content

Abstract.....	ii
Acknowledgments.....	iii
Table of Content.....	vi
List of Figures.....	vii
List of Tables.....	xiii
Chapter 1 Introduction.....	1
1.1 Motivation and Objectives.....	1
1.1.1 Problem Statement.....	2
1.1.2 Research Goals and Task Summary.....	2
1.1.3 Thesis Contributions and Impact.....	3
1.2 Breast Cancer Overview.....	4
1.2.1 Types and Stages of Breast Cancer.....	5
1.2.2 Diagnosis and Detection of Breast Cancer.....	8
1.3 Breast Cancer and Medical Imaging.....	10
1.4 Hyperspectral Imaging (HSI).....	11
1.4.1 Hyperspectral Imaging: History and Overview of Operation.....	11
1.4.2 How Hyperspectral Imaging Works.....	12
1.4.2.1 The development of hyperspectral imaging technology over time.....	14
1.4.3 Hyperspectral Imaging Applications.....	16
1.4.3.1 General applications.....	16
1.4.3.1 Hyperspectral Imaging Medical applications.....	17
1.4.4 Data Analysis and Image Processing of Hyperspectral Imaging.....	20
1.4.4.1 HSI Image processing.....	20
1.4.4.2 Image Processing Algorithms for Hyperspectral Imaging.....	20
1.5 Thesis Overview.....	22
Chapter 2 Theory.....	23
2.1 Introduction.....	23
2.2 Basic principles of light.....	23
2.3 The theory of reflectance.....	25
2.3.1 Electromagnetic waves.....	25
2.3.2 Fresnel equation and optical scattering.....	23
2.3.3 BRDF equation.....	27
2.4 Hyperspectral imaging system overview.....	27
2.4.1 Hyperspectral System Hardware.....	29
2.4.2 Spectrometer of imaging system.....	31

2.4.3 Hyperspectral imaging and spectrometer.....	32
2.4.4 Hyperspectral data cube.....	34
2.4.5 HSI Image processing.....	35
2.4.4.1 Examples of classification algorithms.....	39
2.5 Summary.....	39
Chapter 3 HSI System Evaluation, and Cellular Imaging .....	40
3.1 Introduction.....	40
3.2 HSI imaging Experiments.....	41
3.2.1 Preparation methods of the Pathology Samples .....	41
3.2.2 Experiments Result .....	43
3.3 H&E stained Normal and Ductal Carcinoma samples.....	44
3.4 Unstained Normal and Ductal Carcinoma samples.....	51
3.5 H&E Stained In Situ and Invasive Ductal Carcinoma.....	57
3.6 Summary.....	64
Chapter 4 Image Evaluation .....	65
4.1 Introduction.....	65
4.2 Spectral reflectance determination of manually marked cancer tissue .....	66
4.2.1. Hyperspectral imaging of H&E-stained normal and DCIS tissue samples from patients 1-3.....	66
4.2.2: Hyperspectral imaging of unstained normal and DCIS tissue samples of from Patient 1.....	74
4.2.3: Hyperspectral imaging of H&E-stained DCIS and IDC tissues from Mice 1-3.....	76
4.3 Selection of image data for training cancer detection algorithm.....	82
4.3.1 Supervised data for Support Vector Machine (SVM) classifiers using a Gaussian algorithm on human-female H&E stained breast tissue.....	82
4.3.2 Unsupervised data for K-means algorithm on human- female H&E stained breast tissue.....	84
4.3.3 Supervised data for Support Vector Machine SVM classifiers using a Gaussian algorithm on unstained human- female breast tissue samples.....	85
4.3.4 Unsupervised data for K-means algorithm on unstained human- female breast samples.....	86
4.3.5 Supervised data for Support Vector Machine SVM classifiers using a Gaussian algorithm on mice models with positive DCIS, and IDC.....	87
4.3.6 Unsupervised data for K-means algorithm on mice models with positive DCIS and IDC.....	89
4.5 Applying trained algorithms for semi-automated detection.....	89
4.5.1 Cancer detection in human sample.....	90

4.5.2 Mice samples detection.....	92
4.6 Results & Summary.....	93
Chapter 5 Conclusion and Future Work .....	95
Future work.....	98
References.....	100
Appendix A.....	111
Appendix B.....	134

# List of Figures

Figure 1.1: Concept diagram of how HSI can aid the pathologist in cancer detection and diagnosis. Pictures (left to right): pathology tech [60], hyperspectral imager system at WVU, image processing DCIS samples, pathologist diagnosing a sample on a monitor .....3

Figure 1.2: Anatomy of female breast.....5

Figure1.3: Anatomy of ductal carcinoma types.....6

Figure 1.4: Pathology images of different types of breast cancer. (Left): High grade DCIS, (Middle): Invasive Ductal Carcinoma. (Right): Invasive Lobular Carcinoma .....7

Figure 1.5: Abnormal duct tissue .....7

Figure 1.6: (a-c) left to right: surgical, core, and fine biopsy needle.(d-e) H&E dye, stained tissue .....9

Figure1.7: HSI remote sensing system in geographical and agricultural application//.....11

Figure 1.8: RGB image wavelength intensity spectrums.....13

Figure 1.9: Multispectral and Hyperspectral imaging number of bands .....13

Figure 1.10: (left) hyperspectral cube. (right) pixel spectrum in the spectral dimension.....14

Figure 1.11: Scanning and non-scanning to build the hypercube .....15

Figure 1.12: HSI remote sensing .....17

Figure 1.13: (left-top) RGB, and HSI images of healing diabetic foot showing a regular oxygen flow around the ulcer, which cannot be seen in the unhealed case in the (left-bottom) [42]. On the right side, HSI of the flat field calibration, and the image segmentation of the white blood cells count project [43]..... 18

Figure 1.14: spectrum of different pixels in the skin tissue showing the ability of the hyperspectral system of detection skin cancer .....19

Figure 1.15: a spectrum taken from a tissue showing the spectral reflectance value of normal and prostate cancer tissue .....19

Figure 2.1: light interaction with medium.....	24
Figure 2.2: light spectrum .....	24
Figure 2.3: The geometry of light reflection .....	27
Figure 2.4: Regular image of the breast tissue, hyperspectral image, and the spectrum of the marked pixel.....	28
Figure 2.5: Hyperspectral camera hardware.....	29
Figure 2.6 ImSpector Imaging Spectrograph.....	30
Figure 2.7. TF in hyperspectral systems.....	31
Figure 2.8: Schematic diagram of the basic elements of an imaging spectrometer.....	32
Figure 2.9: building the spectral data cube in both line scan, and snapshot systems.....	35
Figure 2.10: Hyperplane separating two classes (LSVM) [55].....	37
Figure 3.1: a- Hyperspectral imaging system setup Davis lab in WVU Cancer Institute. b - Nikon Optiphot-2.....	40
Figure 3.2: hyperspectral image of a ductal carcinoma tissue.....	42
Figure 3.3: the top three images: showing the three methods of preparing the tumor on the slides, H&E staining, Unstaining, and Unstaining with glass coverslip.....	42
Figure 3.4(a) Reflectance spectra comparison of three tissue preparation methods: H&E stain, unstained with no cover slip, and unstained with cover slip. (b) Reflectance spectra comparison of unstained samples (US) and unstained samples with cover slip (USC). (c) Reflectance spectra comparison of H&E and USC, note that the wavelength axis is the band number between 1 to 31 band and it is ranged from 460nm to 640 nm .....	43
Figure 3. 5: (a) Visible image of H&E-stained normal duct tissue from Patient 1. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....	45

Figure 3.6: (a) Visible image of tissue sample containing ductal carcinoma from Patient 1. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....46

Figure 3.7: (a) Visible image of H&E-stained normal duct tissue from Patient 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....47

Figure 3.8: (a) Visible image of tissue sample containing ductal carcinoma from Patient 2. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....48

Figure 3.9: (a) Visible image of H&E-stained normal duct tissue from Patient 3. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....49

Figure 3.10: (a) Visible image of tissue sample containing ductal carcinoma from Patient 3. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....50

Figure 3.11: (a) Visible image of unstained normal duct tissue from Patient 1. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....51

Figure 3.12: (a) Visible image of unstained tissue sample containing ductal carcinoma from Patient 1. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....52

Figure 3.13: (a) Visible image of unstained normal duct tissue from Patient 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....53

Figure 3.14: (a) Visible image of unstained tissue sample containing ductal carcinoma from Patient 2. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....54

Figure 3.15: (a) Visible image of unstained normal duct tissue from Patient 3. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....55

Figure 3.16: (a) Visible image of unstained tissue sample containing ductal carcinoma from Patient 3. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....56

Figure 3.17: (a) Visible image of H&E-stained DCIS tissue from mouse 1. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....58

Figure 3.18: (a) Visible image of H&E-stained IDC tissue from mouse 1. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....59

Figure 3.19: Visible image of H&E-stained DCIS tissue from mouse 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....60

Figure 3.20: (a) Visible image of H&E-stained IDC tissue from mouse 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....61

Figure 3.21: (a) Visible image of H&E-stained DCIS tissue from mouse 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....62

Figure 3.22: (a) Visible image of H&E-stained IDC tissue from mouse 3. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.....63

Figure 4.1: Spectral reflectance of each of the 4 marked areas of the 550nm spectral channel from top left to bottom right.....67

Figure 4.2: Average response showing spectral reflectance of the cancer and normal tissues.....68

Figure 4.3: SRCF of both normal duct (Left), and ductal carcinoma (Right), and the spectral reflectance normalized scale.....69

Figure 4.4: Spectral reflectance of each of the 4 marked areas of the spectral channel 550nm from top left to bottom right.....70

Figure 4.5: Average response clearly showing spectral reflectance of the cancer and normal tissues.....71

Figure 4.6: SRCF on both normal duct (Left), and ductal carcinoma (Right), and the RICF scale...72

Figure 4.7: Spectral reflectance of each of the 4 marked areas of the spectral channel 550nm from top left to bottom right.....72

Figure 4.8: Average response clearly showing spectral reflectance of the cancer and normal tissues.....73

Figure 4.9: SRCF of both normal duct (Left), and ductal carcinoma (Right), and the RICF scale...73

Figure 4.10: Spectral reflectance of each of the 4 marked areas of the spectral channel 550nm from top left to bottom right.....74

Figure 4.11: Average response clearly showing spectral reflectance of the cancer and normal tissues.....75

Figure 4.12: SRCF of both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.....75

Figure 4.13: Spectral reflectance of each of the 4 marked areas of the spectral channel 500nm from top left to bottom right.....76

Figure 4.14: Average response clearly showing spectral reflectance of the cancer and normal tissues.....77

Figure 4.15: SRCF of both DCIS (Left), and IDC (Right), and the RICF scale.....77



Figure 4.16: Spectral reflectance of each of the 4 marked areas of the spectral channel 500nm from top left to bottom right.....78

Figure 4.17: Average response clearly showing spectral reflectance of the cancer and normal tissues.....79

Figure 4.18: SRCF of both DCIS (Left), and IDC (Right), and the RICF scale.....79

Figure 4.19: Spectral reflectance of each of the 4 marked areas of the spectral channel 500nm from top left to bottom right.....80

Figure 4.20: Average response clearly showing spectral reflectance of the cancer and normal tissues.....81

Figure 4.21: SRCF on both normal duct (Left), and ductal carcinoma (Right), and the RICF scale..81

Figure 4.22: a- Scatter plot of the SVM algorithm, b- ROC curve of the prediction, c- Confusion matrix of the predicted model. ....83

Figure 4.23: a- Scatter plot of the SVM algorithm, b- ROC curve of the prediction, c- Confusion matrix of the predicted model. ....86

Figure 4.24: a- Scatter plot of the SVM algorithm, b- ROC curve of the prediction, c- Confusion matrix of the predicted model. ....88

Figure 4.25: hyperspectral imaging, training, and detection.....90

Figure 4.26: Dr. James Coad marking the ROI to compare the detection results of the algorithm...91

Figure 4.27: The blue boxes indicates the areas marked by the pathologist, the shaded areas shows the detection of the K-means algorithm detecting cancer tissues (Red), and normal tissues (blue) on the spectral channel 550nm.....92

Figure 4.28: The blue boxes indicates the areas marked by the pathologist, the shaded areas shows the detection of the K-means algorithm detecting IDC (Red), and DCIS (blue) on the spectral channel 500nm.....93

Figure 5.1: Concept diagram of how HSI can aid the pathologist in cancer detection and diagnosis. Pictures (left to right): pathology tech [60], hyperspectral imager system at WVU, image processing DCIS samples, pathologist diagnosing a sample on a monitor.....95

Figure 5.2: left to right, human normal and DCIS H&E tissues, human normal and DCIS unstained tissues, mice DCIS and IDC H&E tissues.....96

Figure 5.3: hyperspectral imaging, training, and detection.....97

Figure 5.4: an example of applying SVM trained algorithm on a random samples to detect normal and breast cancer.....98

Figure 5.5: a- a prototype of hyperspectral camera used to image the outer breast tissue (skin) of breast cancer mice models. b- the hyperspectral imaging system, c- WVU Erma Byrd Biomedical Research Center, d- new hyperspectral camera model of imec. Figure 4.34: spectral reflectance of each of the 4 areas from top left to bottom right.....99

# List of Tables

Table 4.1: SVM algorithm on 10 patients.....	82
Table 4.2: K-Means algorithm on 10 patients.....	84
Table 4.3: SVM algorithm on unstained sample.....	85
Table 4.4: K-means on unstained sample.....	86
Table 4.5: SVM algorithm on H&E mice samples.....	87
Table 4.6: K-means algorithm on Mice samples.....	89

# Chapter 1

## 1.Introduction

Breast cancer is one of the highest cause of cancer deaths among American women [1], [2], [3], [4]. According to the U.S. Breast Cancer Organization, statistics show that about 1 in 8 U.S. women will develop invasive breast cancer over their lifetime [1], [2], [3]. In 2016, about 246,660 new cases of invasive breast cancer will be diagnosed in women, 61,000 new cases of noninvasive carcinoma will be diagnosed in situ, and 40,450 women in the U.S. are expected to die from breast cancer. These numbers are large enough to note that, other than lung cancer, breast cancer death rates are higher for the U.S. women than any other cancer [1], [2], [3].

### 1.1 Motivation and Objectives

Hyperspectral Imaging (HSI) is a sophisticated non-invasive optical imaging modality that has the potential to be applied toward medical imaging research and clinical practice. From imaging labs, to National Institute of Health (NIH), to cancer institutes, hyperspectral imaging has been an optical imaging tool that is used for providing spectral information, which can help in material detection. Specifically, this research effort introduces a new tool that can help in breast cancer detection by providing pathologists with a tool that can potentially make their diagnosis process of ductal carcinoma under the microscope easier and more efficient [5]. This work is a collaboration between the Lane Dept. of Computer Science and Electrical Engineering, the WVU Pathology Department, and the WVU Medicine Mary Babb Randolph Cancer Center. The overall scope of the work is to image samples of breast cancer tissues from different patients in order to show the ability of the hyperspectral system to detect spectral differences between normal and breast cancer tissues, as well as early and late stages of cancer, which will also effect the treatment decision.

### 1.1.1 Problem Statement

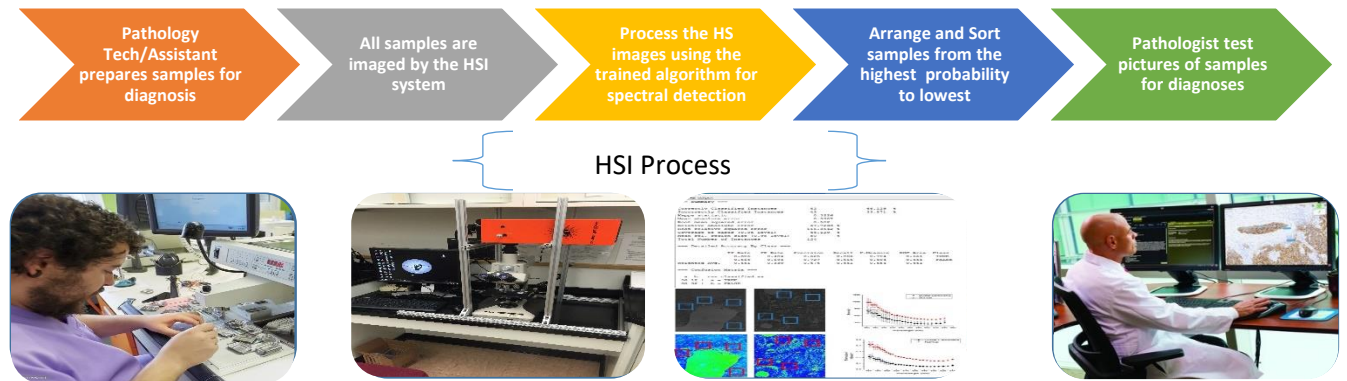
The role of the pathologist is undeniably important for cancer diagnosis because their job as physicians is to look at body fluids and tissues in order to help oncologists as well as other medical specialists make a diagnosis [6]. However, as the number of breast cancer cases increases, the pathologist's job will become more difficult due to the increasing number of samples they need to evaluate on a daily basis. Because the samples are provided via biopsy or surgical procedures, their schedule will have to be compatible with the surgeons' schedule. In addition, the pathologist will also need to spend time preparing the samples, stain them, and finally, examine them under a microscope for diagnoses.

### 1.1.2 Research Goals and Task Summary

The goal of this study is to evaluate the performance of a snapshot hyperspectral imager to see if measurable differences in spectral properties can be observed between normal and various stages of cancerous breast tissues fixed on biopsy slides, specifically in the case of ductal carcinoma. This interdisciplinary work will build a bridge between pathology and hyperspectral optical diagnostic imaging in order to reduce time and workload on the pathologist, with a secondary benefit of leading to more accurate diagnoses. The hypothesis of this work is that a hyperspectral imager can be used to spectrally determine the difference between healthy and cancerous cells on both stained and unstained slides, as well as spectrally determine the difference in cancer cells from varying stages of cancer. A secondary hypothesis is that, if these differences can be seen and verified by a pathologist, image processing techniques can be used to differentiate between cells in a semi-automated fashion. This technique will not replace the pathology work, however it will help improve the pathologist's workflow by pre-screening samples, allowing them to be prioritized for examination based on the likelihood of cancer cells being present in the image. The hyperspectral imaging system used for this work is a Rebellion Photonics snapshot HSI system. This camera has the ability to collect images of the tissue samples across a broad range of wavelengths simultaneously, creating what is known as a hyperspectral data cube. If spectral differences can be clearly identified and measured, trained algorithms can potentially sort through the wavelength-specific images contained in the data cube and identify cancerous and non-cancerous healthy tissue in both stained and unstained tissue samples. This HSI system and associated image processing tools can potentially be used to differentiate between the stages of the cancer cells being examined

based on spectral reflectance differences in breast cancer cells at different stages of progression, which may add additional useful information for the oncologist.

Figure 1.1 is a general concept diagram shows how the pathologist can use the HSI technology in order to make his or her job easier.



**Figure 1.1:** Concept diagram of how HSI can aid the pathologist in cancer detection and diagnosis. Pictures (left to right): pathology tech [60], hyperspectral imager system at WVU, image processing DCIS samples, pathologist diagnosing a sample on a monitor [59].

As Figure 1.1 explains the use of the hyperspectral imager system in the diagnosis process. Pathologist uses the hyperspectral system to image the prepared samples, then a trained algorithm is applied on the captured images, and sorts them in an order of priority (high-low) risk of breast cancer. Finally a pathologist will make a diagnosis decision by looking at the prioritized images of the biopsy samples.

### 1.1.3 Thesis Contributions and Impact

The tasks involved in this research are as follows:

*Task 1: Hyperspectral System Setup* – This task involves two experiments for the HSI system: calibration and understanding how the hypercube is built with different spectral information.

*Task 2: Cellular Imaging* - Stained and unstained tissue samples from normal and cancerous tissues from different patients were provided by the WVU Pathology Department. These cells were imaged using the HSI system connected to a microscope in lab of Dr. Linda Davis, and the spectral

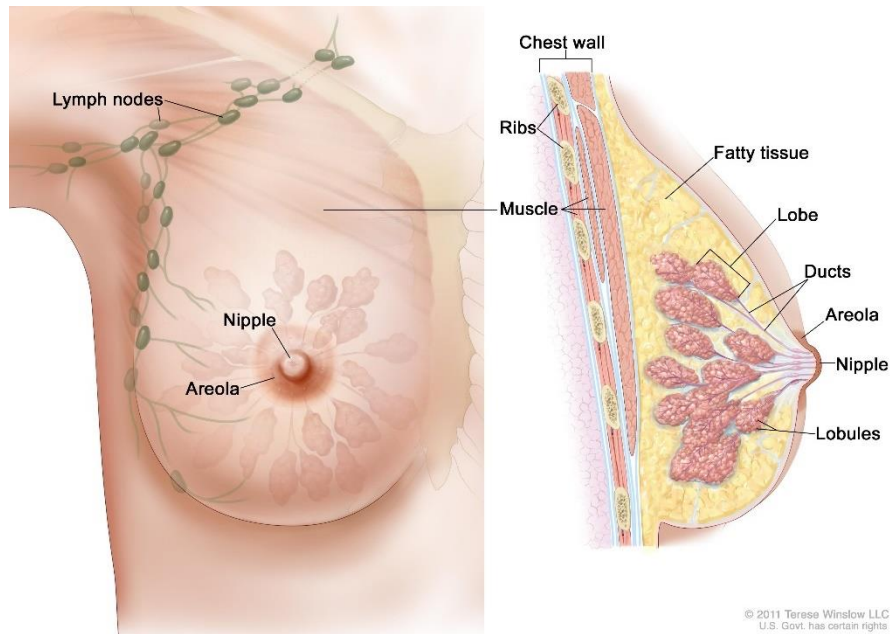
reflectances were evaluated to see if there was a measurable spectral difference between the various cell types.

*Task 3: Image Evaluation* - After imaging, the large amount of data contained in the three-dimensional hyperspectral data cube was evaluated using image processing techniques to look at specific regions of interest in order to determine the spectral reflectances over the visible light wavelength spectrum. The image processing and HSI toolboxes are used to analyze the data with custom algorithms. These algorithms are a collection of machine learning algorithms for data mining tasks that are contained in both MATLAB and Waikato Environment for Knowledge Analysis (WEKA). The algorithms are used to learn the about the nature of the hyperspectral data in order to understand the spectral reflectance value and how it reflects the tissue samples. An overview of this work is given in the last section of this chapter.

This study has shown a great potential for opening new avenues of exploration in the field of breast cancer optical imaging and diagnosis. Specifically, this research will add a new tool that may be used not only in research, but also used by pathology in breast cancer diagnoses. The interdisciplinary nature of the work has fostered a collaboration between the fields of Biomedical and Electrical Engineering, Cancer Pathology, Cancer Cell Biology, Medical Imaging and Image Processing, and has built a communication framework between those departments in West Virginia University and other educational institutes.

## 1.2 Breast Cancer Overview

Cancer appears in the body when cells start to grow in an abnormal way. It can grow in any part of the body, and spread to other areas of the body [1], [2]. Breast cancer, in fact, is considered a tumor that starts from an individual cell, which then starts growing and dividing to form what is known as malignant tumor. Learning about the anatomy of a normal breast (Figure 1.2) and how cancer effects normal cells helps in better understanding how cancer interact within the body. The female breast is made up of a variety of structures and tissues, including: lobules, which are glands that can produce breast milk, and ducts, which are small tubes that transfer milk from the lobules to the nipple, fatty and connective tissue, blood vessels, and lymph vessels [1], [2].



**Figure 1.2:** Anatomy of female breast [2].

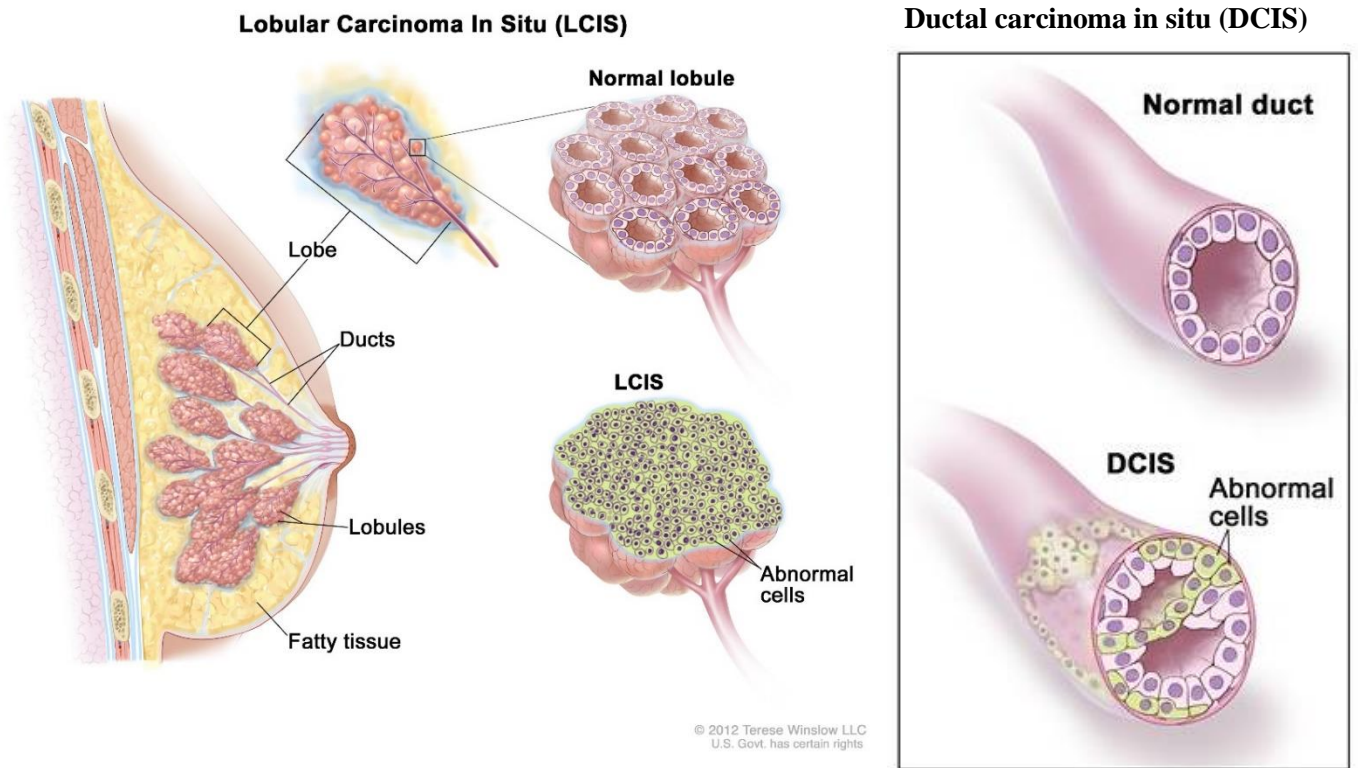
### 1.2.1 Types and Stages of Breast Cancer and

There are many different types of breast cancer, such as invasive and in situ ductal carcinoma, invasive, pre-invasive lobular carcinoma, and the less common cases of different breast sarcomas [1], [2], [4]. More than 80% of breast cancer cases are ductal carcinoma, which begins in the cells that line the ducts [2]. Few breast cancer cases start in the cells lining the lobules. In addition, a fewer number of breast cancers can also start in cells of the other tissues in the breast. These are called sarcomas and lymphomas; however, these are not as common as the first two. Carcinomas, the most common breast cancer, is also called adenocarcinoma, which is a type of cancer that starts in the epithelial cells which line organs and tissues. Sarcomas start in the cells of muscle, fat, or connective tissue, are considered less-common types of breast cancers [2]. Sometimes, a combination of different types of breast cancers may be found in one breast tumor. This research work has been focused on two types of breast cancer, both in situ and invasive which are: ductal carcinoma and lobular carcinoma [4]. Further explanation of these specific breast cancer types is provided in the following sections

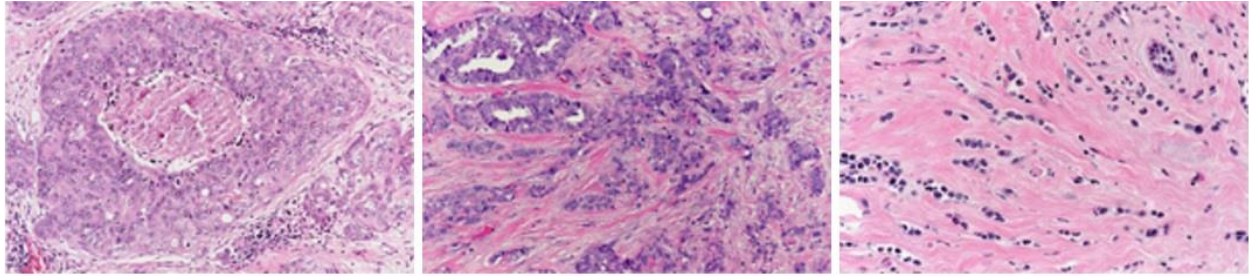


**Ductal carcinoma in situ (DCIS):** DCIS is known as noninvasive (or sometimes, pre-invasive) breast cancer [2]. In situ ductal carcinoma is first detected when the cells lining the walls of the ducts have changed to appear abnormal. (Figure 1.3 [7]), and (Figure 1.5 [1]) exhibit atypical biological, physical, and chemical behavior [7], or show drastic changes in cell-to-cell interaction with neighboring cells [7]. In the case of DCIS, the cancer cells stay confined inside the milk ducts. This characteristic is one clear difference between DCIS and invasive carcinoma; the cells have not spread (invaded) through the walls of the ducts into the surrounding breast tissue [1], [2], [4].

**Lobular carcinoma in situ (LCIS):** From the name, LCIS are tumor cells that have grown to fill the lobules. They grow over the cells that stay inside the milk-producing structures of the breast, which are lobules themselves. The in situ lobular carcinoma is not considered invasive; however, it is a considered a precursor for more invasive types of breast cancer [1], [2], [4].



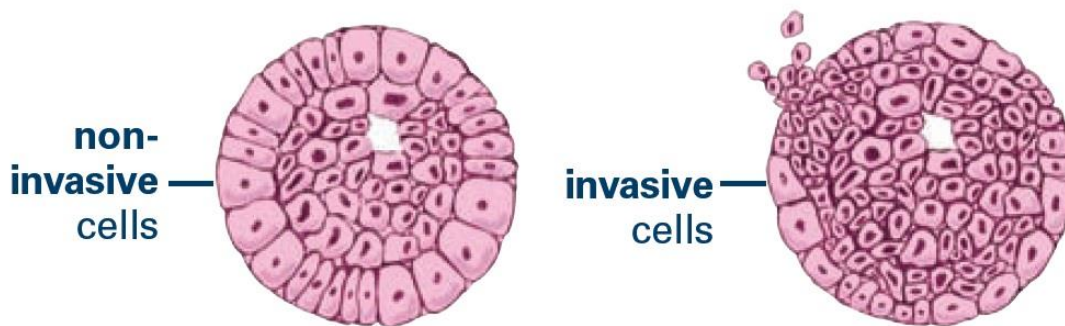
**Figure 1.3:** Anatomy of ductal carcinoma types [2]



**Figure 1.4:** Pathology images of different types of breast cancer. (Left): High grade DCIS, (Middle): Invasive Ductal Carcinoma. (Right): Invasive Lobular Carcinoma [7].

**Invasive ductal carcinoma (IDC):** Invasive ductal carcinoma is the most common breast cancer, comprising approximately 8 out of 10 invasive breast cancers diagnoses. It first appears in a milk duct of the breast, then breaks out the wall of the duct and progresses to infect the fat tissues of the breast. It can also grow into the surrounding normal tissue inside the breast [1]. In addition, it may also start to metastasize, which means to spread to other parts of the body through the lymphatic system and bloodstream [2], [7].

**Invasive lobular carcinoma (ILC):** Invasive lobular carcinoma occurs in the lobules which are the glands that produce milk. Being labeled as invasive, like IDC, it can also metastasize and spread to other parts of the body. Statistics show that about 1 in 10 invasive breast cancers is an ILC [2], [4], [7].



**Figure 1.5:** Abnormal duct tissue [1]

## 1.2.2 Diagnosis and Detection of Breast Cancer

Early detection of breast cancer reduces the mortality rate and increases the survival chance. If a patient is diagnosed with breast cancer, post-tests are performed for the purpose of confirming the diagnosis, determining tumor stage & location, and to help determine the options for treatment planning. There are several tests and procedures used for the purpose of diagnosing, detecting, and confirming whether or not the patient has breast cancer. Diagnosing cancer is often not the work of one individual. In most cases, the diagnosis is a team effort, involving of several specialists from oncology, pathology, radiology, and surgery. It is critical that the following tests and procedures are joined together to provide the patient the best and most appropriate diagnosis and treatment plan that can be offered.

**A) General Physical Examination and History:** A general physical exam is often performed to look at specific signs that could be considered abnormal, such as lumps or rashes. Also, the patient's long-term, short-term, and family health history are examined for previous history of disease or symptoms.

**B) Clinical Breast Examination:** A breast exam is performed by the primary care physician or other medical professional. In this procedure, the examiner carefully feels the breasts and the region under the arms, searching for lumps or other abnormal or suspicious conditions.

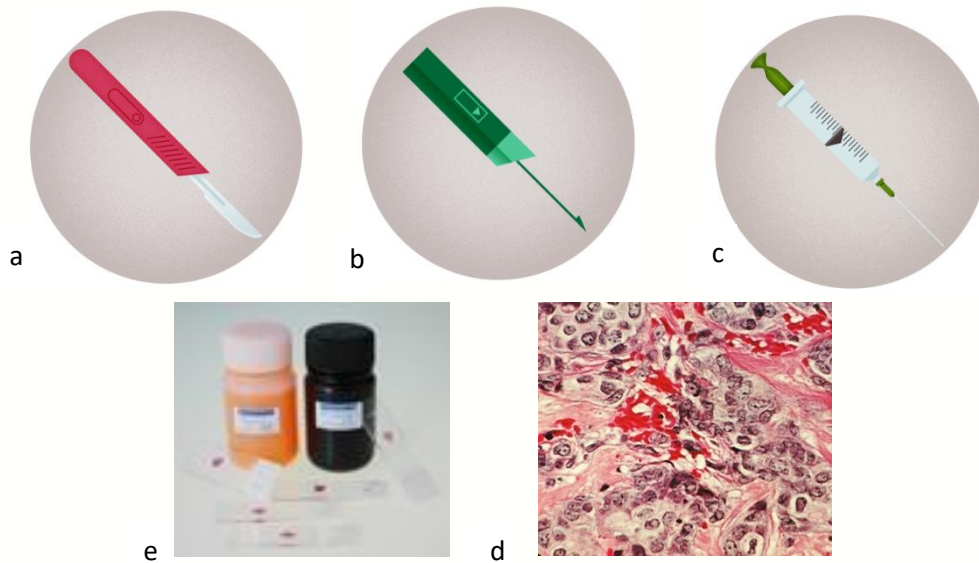
**C) Blood Chemistry Laboratory:** A lab test procedure is performed on the patient's blood sample which is acquired via a common venipuncture procedure. Blood is examined for abnormal numbers when compared to the patient's usual, baseline blood chemistry, such as an abnormal increase in the number of white blood cells, or reduced amounts of oxygen production.[3].

**D) Clinical Trials:** In addition to common office procedures and laboratory testing, the patient can opt into clinical trials aimed evaluating new diagnostic breast cancer procedures, including: evaluation and testing of nipple discharge, ductal lavage, and nipple aspiration exams [8].

**E) Biopsy Test:** A biopsy is a procedure whereby sample tissue is removed from a suspect region identified by one of the previously described procedures. These tissue samples are then viewed under a microscope by a pathologist to check for signs of cancer. Multiple elements could affect the doctor's decision of the best biopsy procedure. These elements, such as tumor size, location, and presence or extent of spreading or growth, can vary in weight when considered for decision making. The following are different types of biopsy procedures [3] [7]:

- *Surgical biopsy*: a surgical procedure that removes either part of a suspect lump, the whole lump, or just a sample of tissue. (Figure 1.6-a).
- *Core Needle Biopsy*: a wide needle (1/16 inch in diameter and 1/2 inch long) is used to puncture and remove a tissue sample using the cutting tip of the needle (Figure 1.5-b).
- *Fine Needle Biopsy*: a very thin, hollow needle attached to a syringe (smaller than core needle) is used to remove a tissue or fluid sample from the targeted area (Figure 1.5-c).

A key component to most biopsies that is relevant to this research effort is that dye is often added to the tissue sample. This dye enables a pathologist to easily view and evaluate the cells under a microscope. In many cases of breast cancer, pathologists use Hematoxylin and Eosin (H&E) dye (Figure 1.6 d-e).



**Figure 1.6:** (a-c) left to right: surgical, core, and fine biopsy needle.(d-e) H&E dye, stained tissue [9]

**F) Medical imaging:** For many decades, medical imaging has been one of the main, if not *the* main, diagnostic tools of many diseases. Development of a wide range of medical imaging modalities has enabled many application areas for these technologies, ranging from research to diagnostic imaging and treatment. Particularly, in breast cancer cases, several imaging modalities are used for detection, diagnosis, and the development of treatment plan to improve the quality of life of the breast cancer patient. These imaging methodologies are described in detail in the next section.

## 1.3 Breast Cancer and Medical Imaging

Different imaging modalities are used for breast cancer, ranging from ionizing and non-ionizing radiation to optical imaging. A short description to some of those modalities is provided as follow:

**A) Mammogram:** A mammogram is an x-ray of the breast. For this procedure, the breast is pressed between 2 plates to flatten and spread the tissue in order to image a large surface area of the breast tissue simultaneously. If the image shows any abnormal structures, the patient is urged to have other tests to acquire additional information that can be applied to form a diagnosis [10] [4].

**B) Ultrasound:** An ultrasound scan uses sound wave echoes to outline the imaged part of the breast tissue in order to create a picture. Ultrasound images are usually viewed in real-time to help locate the target for a physical biopsy [10], [11].

**C) Breast MRI:** High magnetic field and radio waves are used to look at very deep details of structures within the breast, which can help in diagnosis. This technique is not the primary imaging tool for all breast cancer patients, it is used for extracting deeper breast structural information about a specific patient [10], [12].

**D) PET/CT of the Breast:** This imaging method is an advanced imaging technology that combines Positron Emission Tomography (PET) with Computed Tomography (CT) to detect small tumors. This technique is also used to evaluate patients after breast cancer has already been diagnosed to monitor tumor activity, such as growth or spreading [13], [10]. This technology uses radiopharmaceutical drugs that are attracted to the cancer cells, after the patient is injected with the drugs, they can appear on the PET/CT that detects the emission photons and converts them to light signals that resulted by the radioactive decay [13].

**E) Optical Imaging:** Many optical imaging modalities have been involved in a variety of cancer applications including: optical coherence tomography, confocal microscopy, multiphoton microscopy, multispectral and hyperspectral endoscopy, and microscopy [14], [15], [16], [17]. Other optical imaging methods which take advantage of the optical properties of materials and apply optical imaging methodologies to diffuse reflectance imaging have recently emerged with

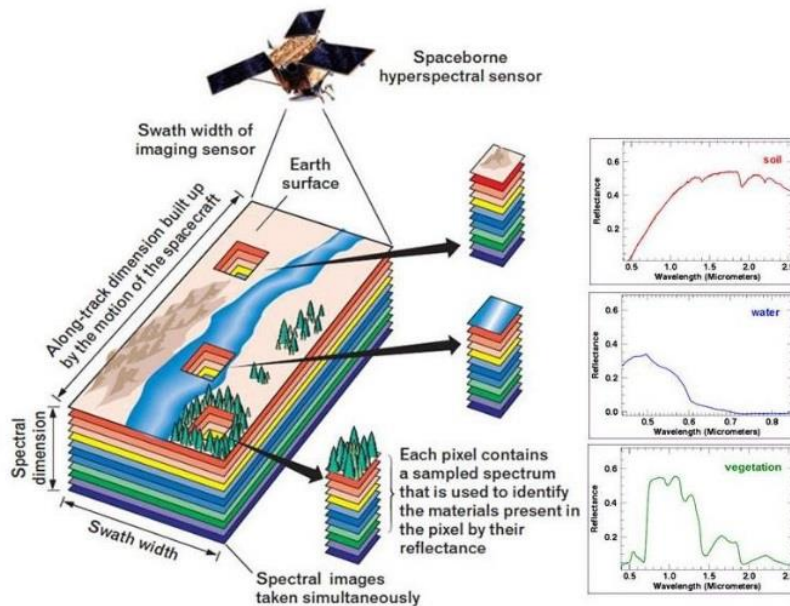


significant potential for non-invasive, portable, and cost-effective imaging for biomedical applications spanning tissue, cellular, and molecular levels [15], [16].

## 1.4 Hyperspectral Imaging (HSI)

### 1.4.1 Hyperspectral Imaging: History and Overview of Operation

Hyperspectral Imaging (HSI) is a well-studied optical imaging technology that was first developed for applications such as space exploration and Earth observation, satellite surveillance, agriculture, and industrial applications [17], [18], [19], [20]. NASA and U.S. Army scientists were the first in discovering and taking advantage of the spectral sensing technology. HSI technology, when properly exploited can provide the Army with additional and improved automated terrain analysis, image understanding, object detection, and material characterization capabilities, as shown in Figure 1.7 [17]



**Figure 1.7:** HSI remote sensing system in geographical and agricultural application [17].

HSI creates a 3-dimensional data structure known as a data cube which contains the spatial image detail (x & y coordinates) as well as detailed spectral information along the z-axis. The spectral information of each co-registered pixel allows investigator to trace, detect, and extract additional information from a target, such as a spectral signature.

### 1.4.2 How Hyperspectral Imaging Works

In order to understand the operation of an HSI system, it is helpful to compare the output of a HSI system to a typical Red, Green, and Blue (RGB) camera or imager. RGB represent the color sensitivity of the imaging pixels, where red, green, and blue are combined in various proportions to obtain any color in the visible spectrum [21], which will also be explained later in **Chapter 2** in the section that talks about the visible light spectrum ranging in wavelength between about 390 to 700 nm. Digital cameras have image sensors, such as CCD (charge-coupled device) and CMOS (complementary metal-oxide semiconductor) imaging arrays, which convert light into electrons to produce an image [21], [22]. As explained, the RGB images are a combination of different pixel intensities from discrete RGB pixels that combine to give our human eyes the appearance of covering the entire visible spectrum. As it is a discretized red green and blue image resulting from the input of a continuum of wavelengths, the image data is combined and smoothed in order to get one colorful image, and that will result of losing valuable data, which can only appear in by not combining the wavelength bands [23] [24]. However, in the HSI case, one image is broken down into tens or even hundreds of images with different wavelength channels, or bands. The images of the different bands will have details in each wavelength that might have been not seen by conventional RGB imagers. This is the main advantage of hyperspectral imaging, as shown in Figure 1.9. Hyperspectral imaging should not be confused with multispectral imaging. A simple way to differentiate between both terms is by the number of bands. Multispectral imaging often uses ~10 discrete spectral bands with center wavelengths separated by several nanometers of wavelength. Hyperspectral imaging uses a large number of spectral bands often separated by only 1-2 nm, allowing the construction of a hyperspectral data cube with rich, continuous spectral information. This defining characteristic gives the hyperspectral imagers the advantage of capturing very detailed spectral information for each pixel in an image, as seen in Figure 1.8.

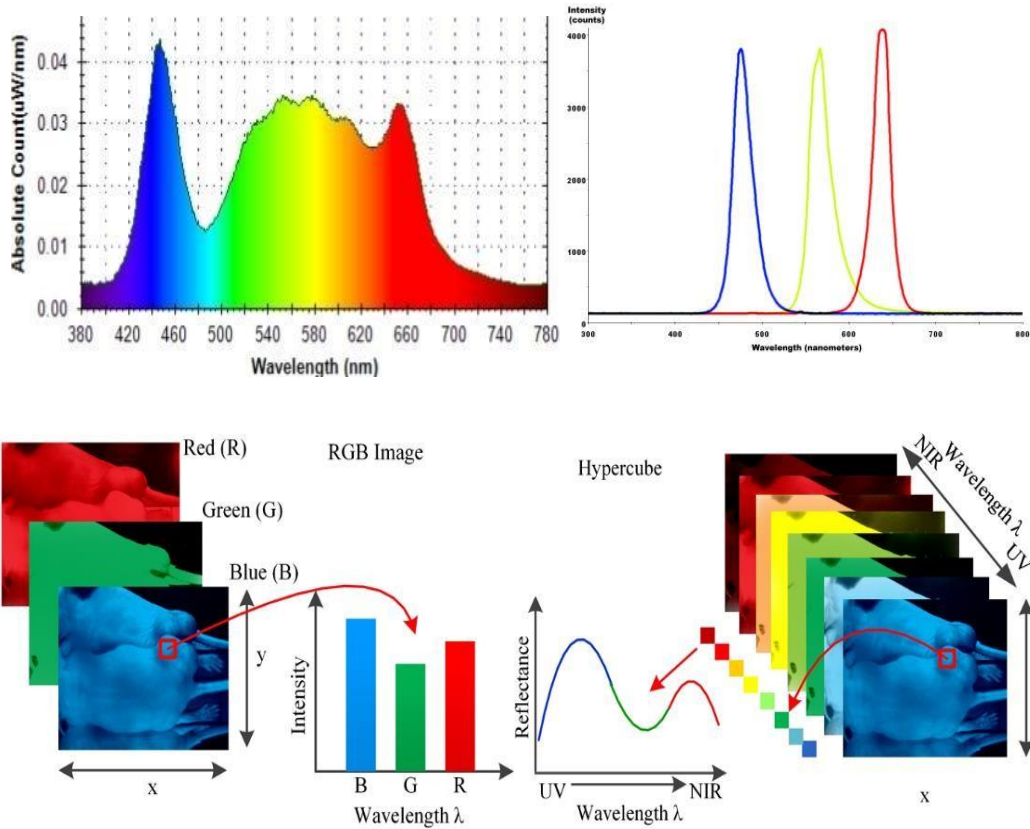


Figure 1.8: RGB image wavelength intensity spectrums [23]

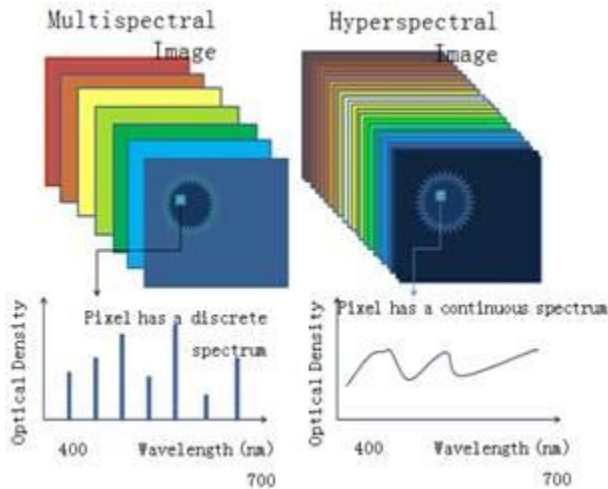
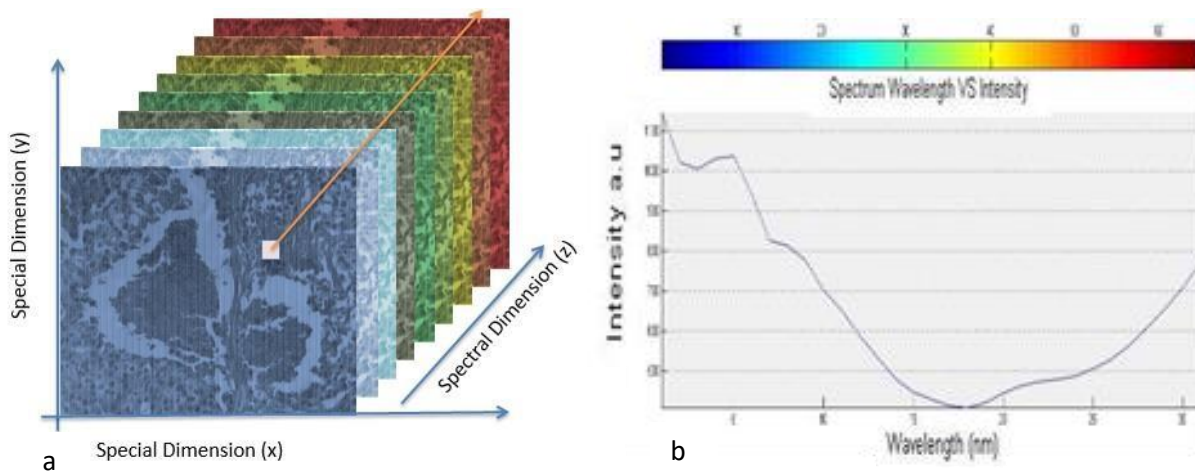


Figure 1.9 Multispectral and Hyperspectral imaging number of bands [23]

As HSI creates a three dimensional hypercube, depending on the capability of the HSI system, each pixel in the hypercube can be characterized by a spectral curve, which can range from the UV to IR regions as shown in Figure 1.10 [24] [25].





**Figure 1.10:** a) hyperspectral cube. b) pixel spectrum in the spectral dimension

#### 1.4.2.1 The development of hyperspectral imaging technology over time

Before Snapshot Hyperspectral technology was invented, the spectral hypercube was built and created by what is called a line-scan hyperspectral imager [25]. There are two different types of line-scan technology, described as follows:

**Spatial Scanning**, where two-dimensional scanning sensor output represents a full slit spectrum  $(x, \lambda)$ , and scan along the  $y$ -axis. The optical slit is basically used to provide a fixed or adjustable aperture and controls the angle of the light which enters through. Hyperspectral imaging (HSI) devices for spatial scanning obtain slit spectra by projecting a strip of the scene onto a slit and dispersing the slit image with a prism as shown in Figure 1.11-A [26].

**Spectral Scanning**, is also a form of two-dimensional scanning sensor. However, the slit spectrum is in  $(x, y)$ , the scan is performed along, and scan through different wavelengths  $(\lambda)$ , which is the spectral dimension, more information about how to scan through different wavelengths will be provided in Chapter 2 . However, each 2-D sensor output represents a single-color monochromatic image, as shown in Figure 1.11-B [26].

In both types of line-scanning system, optical filters, called line filters, are integrated with the camera sensor, [26], [27]. In the case of the spatial scanning, line scan hyperspectral cameras scan the image line per line [27]. The advantage of line scanning is being able to pick and choose spectral bands, and having a direct representation of the two spatial dimensions of the scene in the

spectral scanning case [26], [27]. Line scanning systems have also been used in remote sensing, where it is sensible to use mobile platforms. It has also been used to scan materials in mining and agriculture applications [27]. The disadvantage however, line scanning system require time to scan through all of the wavelengths, which requires a stable camera mount and nonmoving image target. Time of scanning vary depends on the size of the image, the integration time, and the acquisition speed of the camera. System vibrations and target movement have an impact on the accuracy and reconstruction of the image [27].

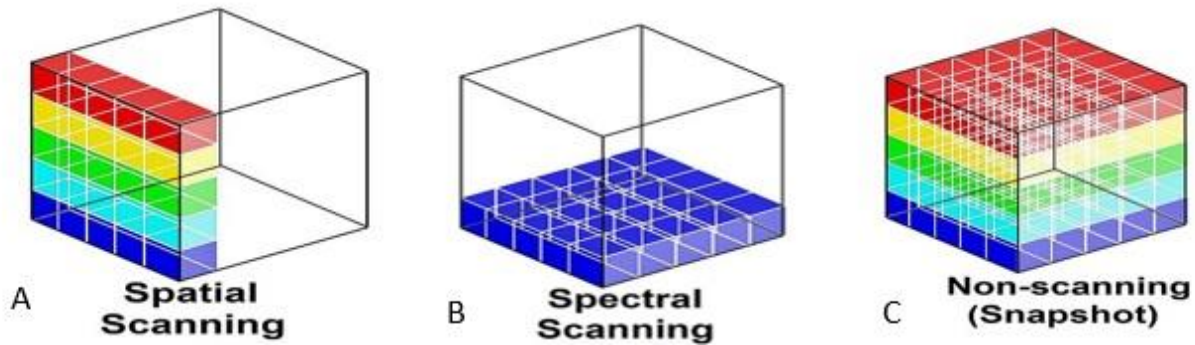


Figure 1.11: Scanning and non-scanning to build the hypercube [53]

### Snapshot (non-scanning) hyperspectral camera

The snapshot hyperspectral camera is able to capture an entire hyperspectral image, in which all of the wavelengths are captured instantly, then the hypercube is created. Snapshot hyperspectral technology is designed and built in different configurations than the line scan camera [27] [29]. It uses a prism to break up the light, and the diffracted wavelengths fall on different portions of the imaging sensor. Software is used to sort the varying wavelengths of light falling onto different pixels into wavelength-specific groups. While conventional line-scan hyperspectral cameras build the data cube by scanning through various filtered wavelengths, or an object, the snapshot HSI acquires an image, and the spectral signature at each pixel, simultaneously. Snapshot systems have an advantage of faster measurement, and also sensitivity improvement, so that measurements that were previously starved for light are now possible [53]. However, one drawback is that the resolution is limited by down-sampling the light falling onto the imaging array into a smaller number of spectral channels Figure 1.11-C shows the snapshot hypercube.

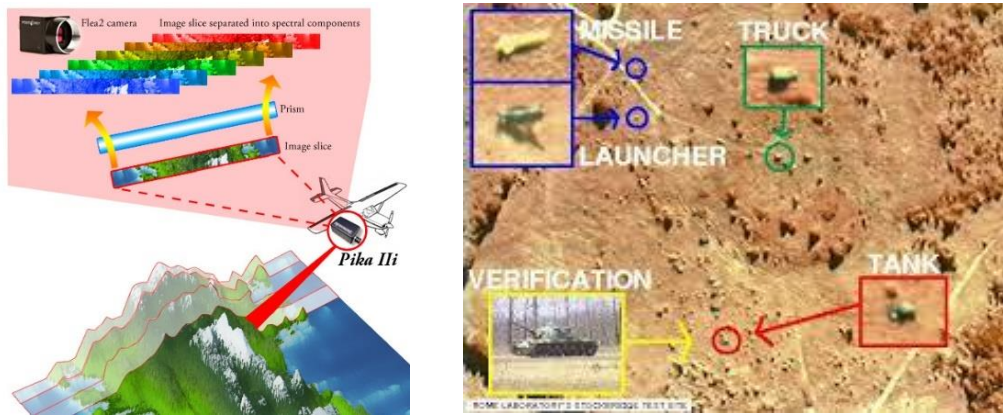
For the past two decades, different research has shown the potential and effectiveness of using HSI technology in medical applications, which HSI has been used as a noninvasive, and nonionizing imaging diagnostic tool [5], [15], [17], [23]. HSI has been used in many medical applications in the exploration of anatomy, physiology, and pathology. Also, beyond the visible spectrum, HSI can be used for optical biopsies which involve in vivo diagnosis of tissue without the need for sample excision and processing [5], [16], [30]. These applications take advantage of the spectral information contained in light that passes through (transmission) or reflects off of (reflection) different materials, which are fundamental properties associated with transmission and reflection spectroscopy [14], [15], [16]. Specific to reflection spectroscopy, spectral reflectance is defined as the ratio of reflected energy to incident energy as a function of wavelength [19], [20]. These spectral reflectance curves are used for the identification of different materials, which is why HSI has an advantage over the other imaging systems that can't reach such information [20]. More information will be provided in **Chapter 2** (Theory) on how hyperspectral imaging works, and how spectral reflectance curves are collected to provide a valuable spectral information that can be used for many applications.

### 1.4.3 Hyperspectral Imaging Applications

#### 1.4.3.1 General applications

HSI remote sensing systems sense the light that is reflected from objects that are usually some distance away from the sensing system. Many applications of HSI technology involve airborne and satellite-borne systems. The light that is sensed by a remote sensing system is affected by the distance between the system and by the source of the light. Any HSI system must account for these effects when comparing sensed spectra to spectral signatures. This technology has been used in many applications in geology, agriculture, and military target detection. As each material reflects different spectral signature, the HSI systems can identify the materials remotely. In geology, hyperspectral imagery has been used to study the geography of the different parts of the Earth's surface, which helps in the observation and study of changes that may take place over time due to natural or man-made causes. [32] [33]. In agriculture, the HSI system is also used to study characteristics such as soil quality, to determine the best regions for agricultural projects [34]. The

HSI systems are also used for military target detection to locate and identify camouflaged tanks or bases as shown in Figure 1.12. [35]



**Figure 1.12:** HSI remote sensing [35]

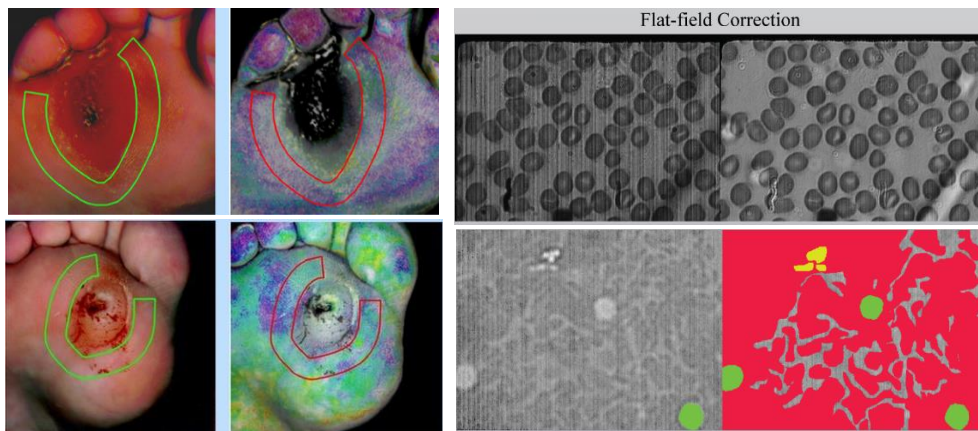
HSI systems are also used in the food industry. Automated HSI applications used for quality control on production lines is a promising application [36], specifically due to an improvement of acquisition time and data processing speeds enabled by modern CMOS and cameras and computer processors. HSI system can also be used to detect fungal diseases and bacterial contamination in fruits and vegetables by looking for variations in spectra caused by biochemical and physicochemical processes associated with the undesirable organisms [37]. HSI has also been applied to the detection of food freshness by the measuring the percentage of fat and sugar to determine how moisturized the product is [38].

#### 1.4.3.2 Hyperspectral Imaging Medical applications

In the last two decades, HSI has been used in several medical applications including: disease diagnoses and screening, real-time surgical monitoring, diabetes-related wound monitoring, and white/red blood cell differentiation [39], [40], [41]. The details of a selection of these applications most closely related to this research effort are provided in the following sections.

1.4.3.2.1 General Medical Applications: Researchers at the University of Texas Southwestern Animal Lab have been using the HSI technology in renal surgery. They use the HSI to monitor blood oxygenation levels in real-time before after and during surgery. This will help in preventing

the risk of or predicting the likelihood of renal failure after surgery. They also use HSI during the implementation of pharmacologic solutions to monitor kidney function to predict what might happen to patient during, and after surgery [44]. HSI has also been used in the cases of diabetic wounds to investigate hemoglobin saturation to help predict the ulcerations and track the healing of the diabetic foot ulcers as shown Figure 1.13 [42]. Ongoing research at WVU is aimed at using an HSI system to image and segment human white blood cells flowing in vessels to enable white blood cell counting without drawing blood as shown in Figure 1.13 [43].

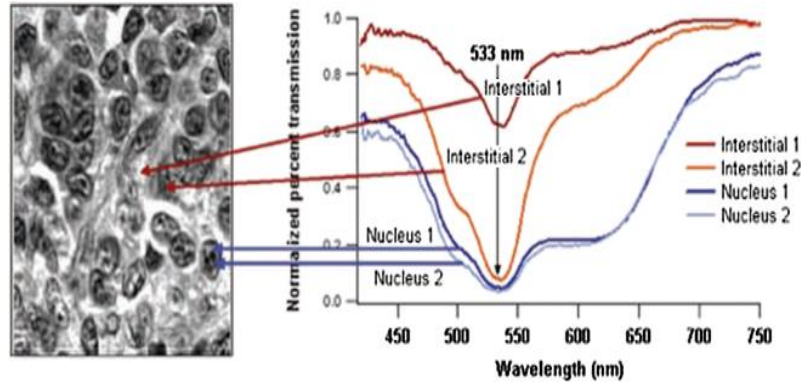


**Figure 1.13:** (left-top) RGB, and HSI images of healing diabetic foot showing a regular oxygen flow around the ulcer, which cannot be seen in the unhealed case in the (left-bottom) [42]. On the right side, HSI of the flat field calibration, and the image segmentation of the white blood cells count project [43].

#### 1.4.3.2.1 HSI and Cancer

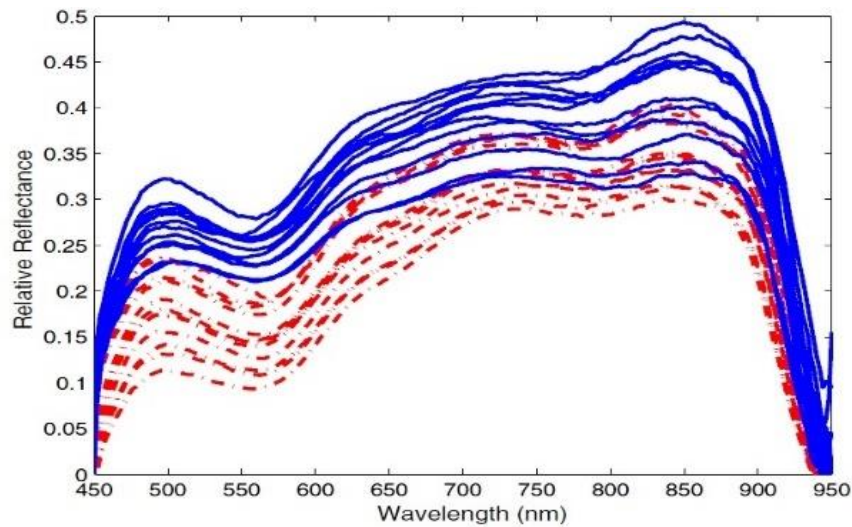
In general, optical imaging has also been widely used in oncology applications in both clinical and research settings. In fact, hyperspectral imaging has been used in detecting skin cancer in dermatology clinics [45]. In addition, many cancer research centers are exploring a variety of HSI-based research [46], which will enable a large expansion of clinical applications of HSI technology in the near future. Examples of different applications of HSI in the cancer field are provided below. HSI has been a very useful tool in skin cancer investigation [48]. Several melanoma and skin sarcoma cases have been confirmed using HSI system. The HSI noninvasive screening system will may increase the speed of diagnoses and melanoma detection as shown Figure 1.14 [47].





**Figure 1.14:** spectrum of different pixels in the skin tissue showing the ability of the hyperspectral system of detection skin cancer [47].

Research conducted by a team at Emory University has demonstrated HIS-based prostate cancer detection in mice. The idea is to use the spectral difference of tissues contained in the HSI data in order to enhance the detection of cancer tissue. Two different methods have been explored in this study. The first was to detect spectral differences, then create a method to highlight the differences of the reflectance properties of cancer versus those of normal tissue in tumor-bearing mice and on pathology slides. This imaging method may be able to help physicians to dissect malignant regions with a safe margin and to evaluate the tumor bed after resection [48]. This study may lead to advances in the optical diagnosis of prostate cancer using HSI technology [48].



**Figure 1.15:** a spectrum taken from a tissue showing the spectral reflectance value of normal and prostate cancer tissue [48]

#### 1.4.4 Data Analysis and Image Processing of Hyperspectral Imaging

Data Analysis and Image Processing are two main components of a complete imaging system that is used for analyze, and process images as well as material detection. These fields use standardized algorithms and functions to analyze, explore, and examine an image [49]. The images can also be visualized in specific conditions with focusing on regions of interest (ROIs). Image processing also allows for performing image analysis, image segmentation, image enhancement, noise reduction, geometric transformations, and image registration [50]. It can also be applied on specific data information per pixel, so image processing can be used to explore examine a region of pixels [49]. Image analysis is the process of extracting meaningful information from images such as finding shapes, counting objects, identifying colors, or measuring object properties [50].

1.4.4.1 HSI Image processing: In the application of HSI systems, image analysis enables the extraction of diagnostically-useful information that is taken from the hyperspectral data cube set at the tissue, cellular, and molecular levels [51], [52]. The data that is extracted from the hyperspectral images is not only considered a three-dimensional data set, it also provides specific spectral information per pixel, so image processing also enables examination of a specific spatial region of pixels [51]. Therefore, advanced image analysis, and classification methods for hyperspectral datasets are required to extract (unmix) spectral data and classify relevant spectral information [52]. Image analysis can also be defined as the process of extracting meaningful information from images such as shape detection, object counting, identifying colors, or measuring object properties [49] [50] [51]. In addition, the goal of applying hyperspectral imaging in this work is not only to identify clear spectral deference between healthy and malignant in breast tissues and provide pathology with an additional tool for strategic diagnosis, but also in advanced stages to break down the spectral mixtures into a spectral signature of molecular fingerprints with disease states and stages [45], [50], [52].

A general summary of the basic steps used for hyperspectral image analysis involve preprocessing, feature extraction and feature selection, and classification or unmixing, which applies different analysis methods to the data cube in order to get a better understanding of the image and spectral content and to extract some useful information that cannot be seen or observed using conventional RGB camera technology [53], [54]. Chapter 2 talks in more details about image processing, image

clustering, and segmentation, explaining why and how applying different algorithms will benefit the user. A short description of the two related image processing algorithms used in this research effort are provided in the following sections.

#### 1.4.4.2 Image Processing Algorithms for Hyperspectral Imaging.

K-means clustering, and Support Vector Machine (SVM), are two different data analysis algorithms that is used in this work to provide different ways of analyzing HSI data to help confirm diagnosis and detection.

1.4.4.2.1 K-means is a learning algorithm that can be applied to solve unsupervised data classification problems. It starts first by choosing a number of classes (clusters). The main idea is to define ' $k$ ' centers, one for each cluster [56]. These centers should be placed in a certain way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other [57]. The next step is to take each point belonging to a given data set and associate it to the nearest center. In this case, each class will have one centroid, and the many times the experiment is running the better stable result will be, also the more data the classifier is given, the closer better accurate centroid will be picked [56].

1.4.4.2.2 In the case of SVM, it is a classifier that can identifies and creates a hyperplane that separates between the different classes [58]. In other words, given labeled training data, the algorithm outputs an optimal hyperplane which categorizes new examples figure 2.3 in Chapter 2. This will happen only if the data is linearly separable. The optimal hyperplane is the best hyperplane that have the maximum separation distance [59], [60]. Different SVM algorithms were developed based on how to separate between classes in the data, and that is always depends on the data itself [58].

In summary, the aim of this study is to add a new tool and platform in breast cancer ductal carcinoma diagnosis, by evaluating the performance of the optical imaging snapshot hyperspectral imager to measure and explore the differences in spectral properties, which may be observed between normal and various stages of DCIS breast tissues fixed on biopsy slides. As a result, this interdisciplinary work will build a bridge between pathology and hyperspectral optical diagnostic imaging in order to reduce time and workload on the pathologist, which can lead to benefit of lead reducing time, and increasing the accuracy of diagnoses.



## 1.5 Thesis Overview

This work effort is aiming for applying Hyperspectral imaging technology to help pathology in breast cancer ductal carcinoma diagnosis. Specifically, the HSI system will be trained to image pathology biopsy samples from patients and then, in semi-automation method, the samples will be sorted by the priority of which samples appears more likely cancer from the others. This step will reduce time and workload on the pathologist, which will also lead to more accurate diagnoses. The thesis is organized as follows: **Chapter 2** describes a theory of light reflectance in different materials, then the fundamentals of how hyperspectral imaging works in details, and finally a summary of the mathematical approaches of the image processing algorithms used in this study. **Chapter 3** provides the setup and testing steps of the HSI system, which is the first step to understand the hyperspectral imaging system that is used in the work. Also, applying different experiments to test the system with different experiments in order to have a clear image of the spectrum depends on the light source. Two main experiments are done after the system calibration which are called: HSI leaves experiment, and three pathology slides preparation method. **Chapter 3** also provides the cellular imaging samples, where samples of stained and unstained tissues are imaged where studying the spectral difference between the normal and cancer tissues as well as In situ and invasive mammary cancer in mice. The samples were manually marked by pathologist, to show the differences in the spectral reflectance that the HSI system can show between the normal and the cancer tissues, as well as the in situ with the invasive cases. **Chapter 4** approaches the results of applying the image processing algorithms to train the system for the hyperspectral data detection. Two different algorithms are used, the first one is K-means algorithm that will be applied on unsupervised data and creates two different classes for detection. The second algorithm is support vector machine (SVM) that will be applied on a supervised data to test the detection ability. Finally, the last chapter (**Chapter 5**), a conclusion of the work with discussion and summary, then the future work that might be in continuation of this work

# Chapter 2

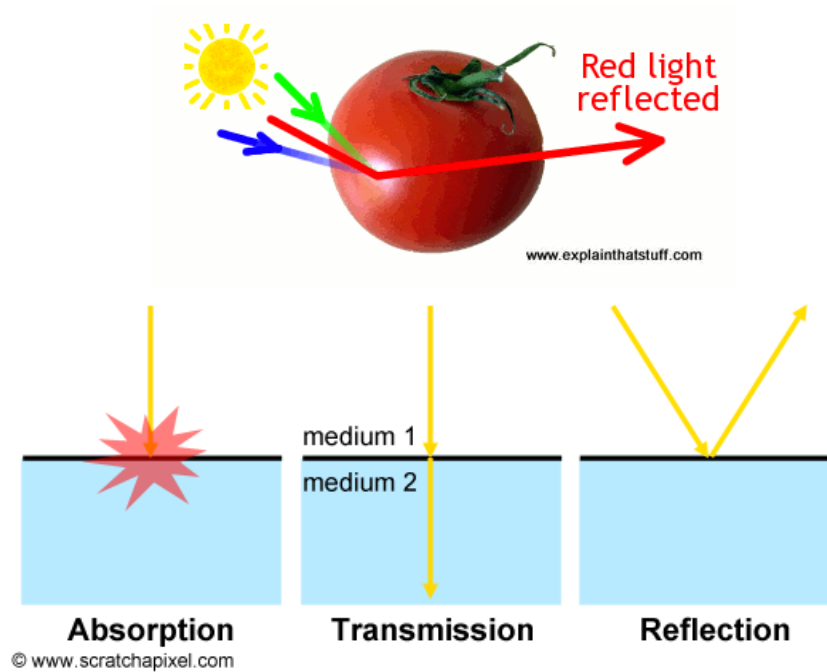
## 2. Theory

### 2.1 Introduction

This chapter provides a detailed description of the theory of light reflectance in different materials, how different shapes, structures, and components of materials can react to different light sources, and how it relates to hyperspectral imaging. The next sections will provide fundamentals of the hyperspectral imager, and how the reflected light travels through the optical components to build the hyperspectral image, the hyperspectral data cube. This Chapter also, provides the fundamentals of spectrometer, spectral reflectance, spectral signature, and how the overall hyperspectral imaging system works. Finally, the last section will include a summary of the mathematical approaches of the different image processing algorithms used in this study.

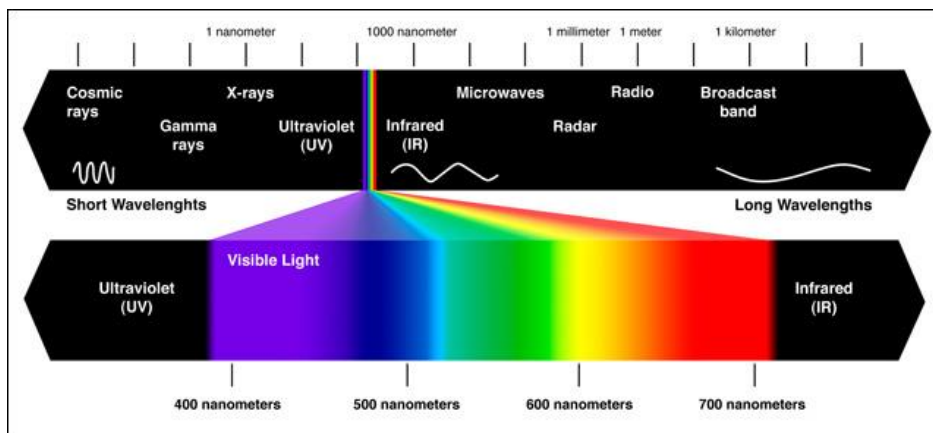
### 2.2 Basic principles of light

Light is a form of electromagnetic energy. As light travels through a medium, it interacts with the medium in different ways depending on the medium material and other properties. Three of the main interactions are: absorption, transmission and reflection as shown in Figure 2.1 [61], [62]. Light absorption occurs when the energy is transferred from electromagnetic waves to molecules of the medium. [61]. Transmission, however, occurs when electromagnetic waves of the light move through a material. Reflection occurs when electromagnetic waves of light reflect off a medium [62].



**Figure 2.1:** light interaction with medium [61]

Light waves range in wavelength from 400-700 nm as shown in Figure 2.2 [61]. Absorption, reflection, and transmission are all wavelength-dependent phenomena, a critical component to the operation of hyperspectral imagery.



**Figure 2.2:** light spectrum [61]

## 2.3 The theory of reflectance

Hyperspectral imaging spectroscopy has the capability to identify or detect specific materials based on their spectral reflectance. In order to get accurate and useful results, the hyperspectral imager system handles raw image data in several steps, ending by processing the captured data with a software by applying algorithms for material spectral reflectance for identification and detection [14]. However, spectral reflectance varies upon the nature of materials and the reflected surface that interacts with light. Reflectance of light can be clearly understood by understanding the Bidirectional Reflectance Distribution Function (BRDF), which quantifies the geometric radiance distribution which results from the reflected light [15]. To clearly understand BRDF, it is important to explain some of the fundamentals of electromagnetic waves, which leads to defining the Fresnel equations, and then the BRDF.

### 2.3.1 Electromagnetic waves

The electromagnetic plane wave that has the orthogonal x and y components of the electric field vector, which travels in the z direction may be expressed by the following equation:

$$\vec{\epsilon}(z, t) = (\epsilon_0 x i + \epsilon_0 y j) e^{i(\omega t - \vec{k} \cdot \vec{z})} \quad \text{Equation 2.1[62]}$$

where  $\vec{\epsilon}$  is the magnitude and direction of the electric field [V/m] in the x-y plane as a function of position, z, and time, t.  $\epsilon_{0x}$  and  $\epsilon_{0y}$ , are complex electric field projected onto the x- and y-axes.

$\vec{k}$  is the propagation direction vector, or wave vector, given by:

$$\vec{k} = \frac{2\pi}{\lambda} \quad \text{Equation 2.2}$$

where  $\lambda$  is the wavelength The angular frequency  $\omega$  can be found using :

$$\omega = \frac{2\pi C}{\lambda} \quad \text{Equation 2.3}$$

where  $c$  is the speed of propagation of light in a vacuum.

### 2.3.2 Fresnel equation and optical scattering

As light hits an object or surface, reflection, absorption, and transition are summed to unity. Fresnel equations provide the reflection and transmission magnitudes of an electromagnetic wave incident on a material [15]. The reflection and transmission are a function of the angle and polarization of the incident electromagnetic wave [16]. For the purposes of applying the equations, the orthogonal polarization components are defined relative to the plane of incidence. The component of the electric field perpendicular to the plane of incidence is called S-polarization. Reflection and transmission are given by:

$$r_S = \frac{\epsilon_r}{\epsilon_i} S = \frac{-\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \quad \text{Equation 2.4[15]}$$

$$t_S = \frac{\epsilon_t}{\epsilon_i} S = \frac{2\sin\theta_t \cos\theta_i}{\sin(\theta_i + \theta_t)} \quad \text{Equation 2.5[15]}$$

where  $\theta_t$  is the transmission angle given by Snell's law as:

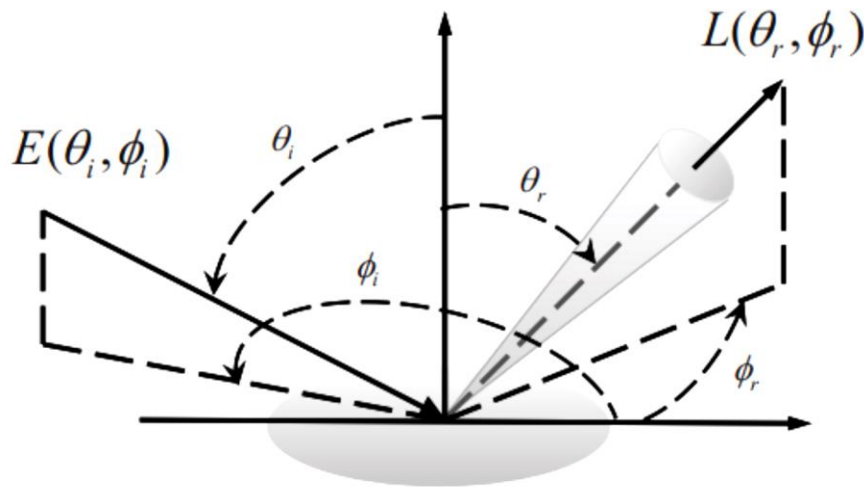
$$\theta_t = \sin^{-1} \left( \frac{\hat{n}_i}{\hat{n}_t} \sin \theta_i \right) \quad \text{Equation 2.6}$$

where  $\hat{n}_i$  and  $\hat{n}_t$  are the complex indices of refraction of the incident and transmitted mediums

### 2.3.3 BRDF equation

The optical property of a surface light reflection is usually described by the bidirectional reflectance distribution function (BRDF). It shows how light is reflected into different directions [17]. It is also defined as the ratio of the radiance scattered into the direction described by the orientation angles  $(\theta_r, \phi_r)$  to incident irradiance from  $(\theta_i, \phi_i)$ , as seen in Figure 2.3.

$$f = \frac{L(\theta_r, \phi_r)}{E(\theta_i, \phi_i)} [sr^{-1}] \quad \text{Equation 2.7[62]}$$

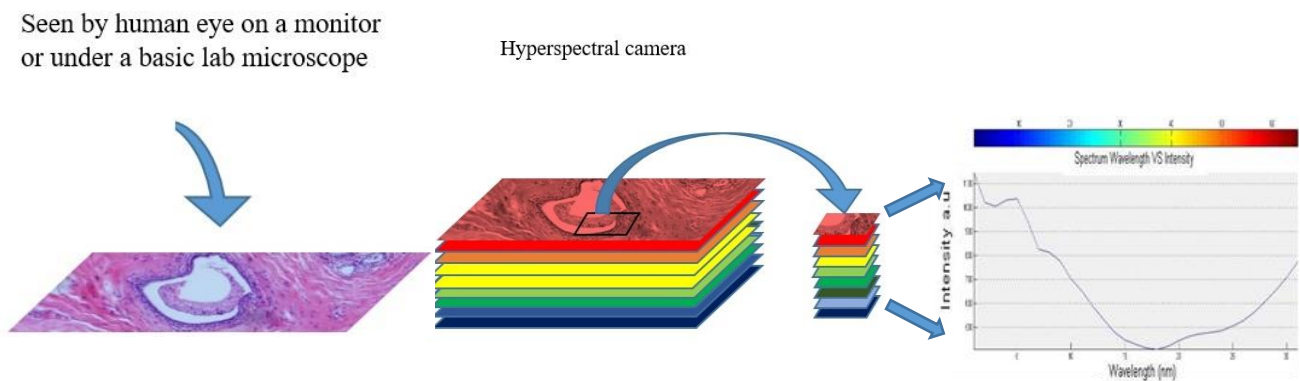


**Figure 2.3:** The geometry of light reflection [17]

## 2.4 Hyperspectral imaging system overview

Hyperspectral imaging is an optical imaging modality that collects and analyzes spectral information from a wide range across the electromagnetic spectrum. Hyperspectral imaging also has advantages over regular imaging in that it provides spectral reflection or absorption characteristics of the object being imaged in the form of spectral channels contained in an image data hypercube [17]. The hypercube is essentially a three-dimensional dataset, which means even a single pixel contains a feature vector with over 10 dimensions with an entire spectrum of

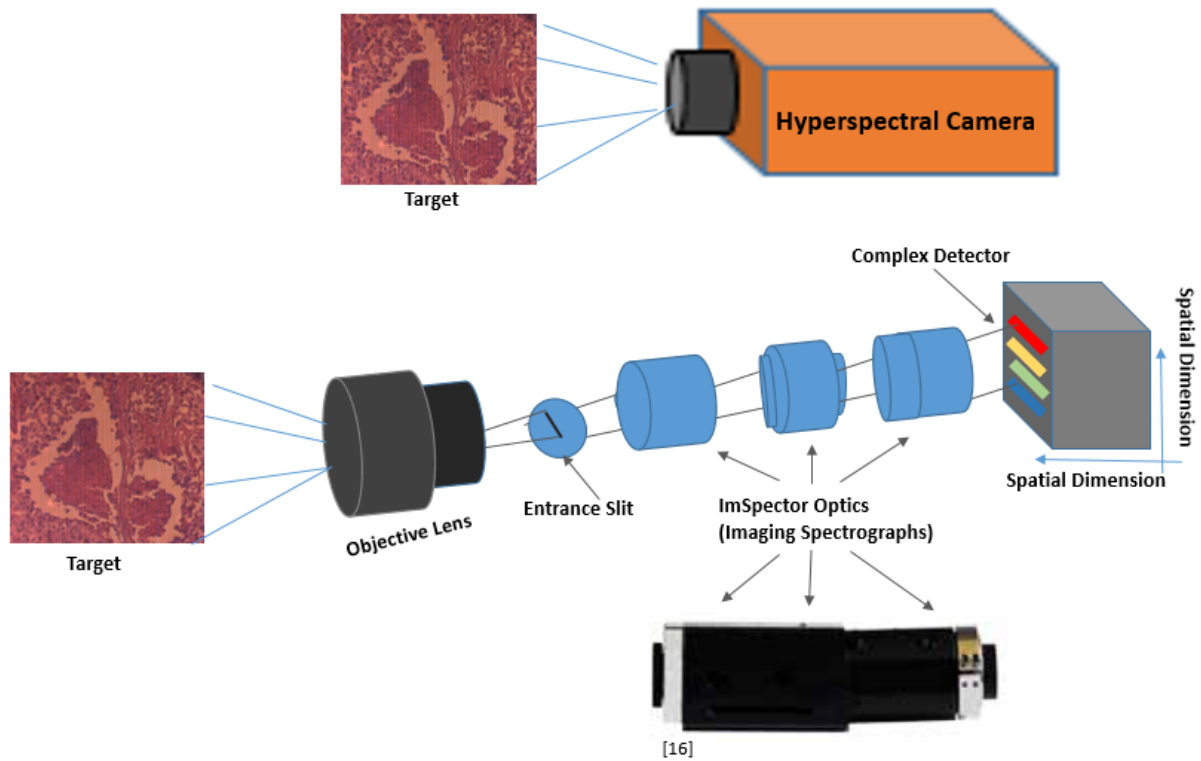
reflectance information [18]. The technology of the hyperspectral imager requires using optical sensors, which are adjusted to collect spectral information in several narrow bands or channels, and typically these channels are in range between 2-10nm wide. Hyperspectral imaging sensors produces image data in dozens or hundreds of narrow spectral bands [18], [19], [63], [64]. HSI technology, when properly exploited can be used in optical imaging applications such as satellite remote sensing, providing an additional and improved automated terrain analysis, image understanding, object detection, and material characterization capabilities, as shown in Figure 1.6 [17]. Hyperspectral imaging can also be used in optical medical imaging in clinical and research applications such as the work described here focused on breast cancer tissue detection (Figure 2.4). The information that is collected from the imaging system can be analyzed to obtain a continuous spectrum for each image pixel or cell. Areas targeted by the imaging system may have different biological, and chemical compositions that can be detected by the sensor due to differences in light reflection or absorption. The image spectra can be compared with field or laboratory reflectance spectra in order to recognize and map surface materials such as particular types of vegetation or diagnostic minerals associated with ore deposits [63]. Hyperspectral images have valuable data, which requires an understanding of exactly what properties of ground materials we are trying to measure, and how they relate to the measurements actually made by the hyperspectral sensor [63].



**Figure 2.4:** Regular image of the breast tissue, hyperspectral image, and the spectrum of the marked pixel.

## 2.4.1 Hyperspectral System Hardware

The optical and spectral characteristics of a hyperspectral imaging system are determined largely by the application requirements [26] [27]; however, all systems have the same basic components that are assembled to achieve a common goal: to image the object, and provide both a spectral and spatial dataset. The hyperspectral imaging system, basically built together as one optical system, consists of a suitable objective lens matched to the spatial and spectral requirements of the application [63]. Figure (2.5) shows how the reflected light of the imaged target travels onto the entrance slit in the spectrograph. This light is dispersed into different wavelengths by the prism in a spectrograph, such as an ImSpector (spectrograph) [64]. The optical slit, imaging spectrograph and proper detectors to collect the spectral, and spatial information are all required optical elements to build a proper hyperspectral imager. Figure (2.5) also describes how the ImSpector system works.



**Figure 2.5:** Hyperspectral camera hardware

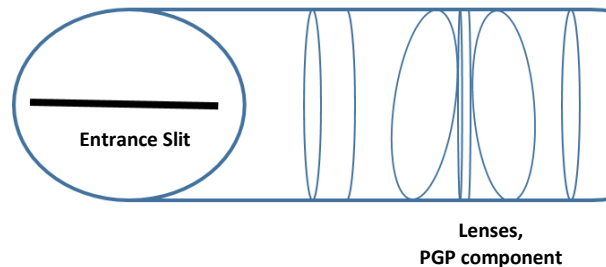
First the reflected light travels through the objective lens then to the slit of the main element of the spectrograph. The image that is captured by the camera sensor as a two dimensional image, the



spatial position and,  $\lambda$  as the spectral dimension. In order to capture the entire image, snapshots of every line across the target area must be captured [64]. The slit, grating and detector characteristics determine the spectral resolution as well as spatial resolution is determined by the pixel size. The imaging spectrograph elements, and the Prism Grating Prism (PGP) imaging component are explained in more details in the following section.

**A) ImSpector (Imaging Spectrograph):** The ImSpector - developed by SPECIM manufacturer - is an imaging spectrograph that has the capability to be attached and combined with a many monochrome imaging systems to form a spectral camera [66], [67]. Figure (2.6) shows the basic example of an ImSpector. Each ImSpector consists of an input slit, collimating optics, transmission grating and focusing optics as Prism Grating Prism (PGP) in a rugged package [67]. The ImSpector produces a flexible wavelength selection that can simultaneously cover a broad spectral range, and results of a high quality spectral and spatial resolution image [69].

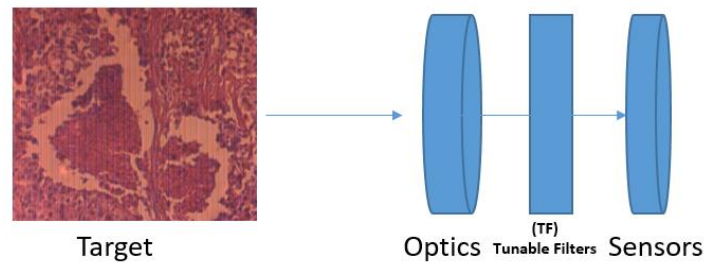
**B) Hyperspectral Prism Grating Prism (PGP) imaging component:** Prism Grating Prism is the main component that is located in the heart of the Imaging spectrograph. The PGP is designed as two identical prisms: short and long pass filters are placed in between grating and the prisms. The main objective controls the spectral response, and both filters within the PGP contain the aperture stop. PGP also allows the spectrum to be centered, in which when optimized, refracts the spectrum so that the light spots of both ends are similar.



**Figure 2.6** ImSpector Imaging Spectrograph

**C) Hyperspectral imaging optical filters:** Optical filters have been used in the hyperspectral imaging systems for the wavelength selection method. Specifically, Tunable Filters (TF) are normally used in the hyperspectral imager as shown in Figure 2.7. A TF is a device in which spectral transmission can be electronically controlled by applying voltage (in the case of liquid crystal Tunable Filter -LCTF-), or acoustic signal (in case of Acousto Optical Tunable Filter -

AOTF-). These two filters are the most common electronic filters that have been used in hyperspectral systems [66].



**Figure 2.7.** TF in hyperspectral systems

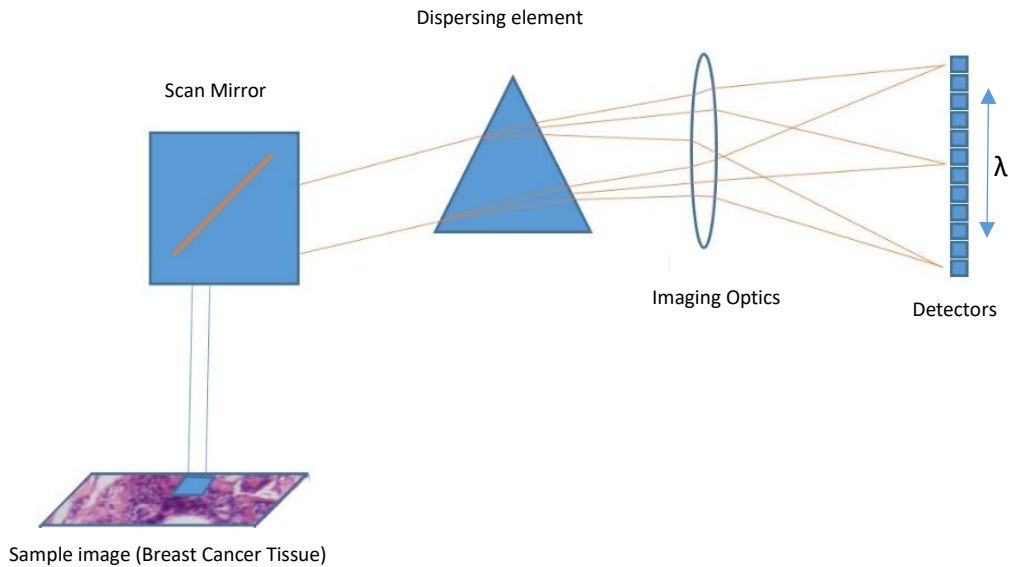
The main difference between LCTF and the AOTF is that LCTF is made with a stack of polarizers and tunable liquid crystal plates that work as band pass filter; whereas the AOTF consists of a crystal in which radio frequencies (RF) acoustic waves are used to separate a single wavelength of light from a broadband source. The wavelength of light selected is a function of the frequency of the RF applied to the crystal [15], [16].

#### 2.4.2 Spectrometer of imaging system

The two main structural components of the Hyperspectral imager system are the spectroscopy and the imaging systems. As pointed out, the imaging system can be attached to any optical device from satellites to microscopes to endoscopic medical systems. Spectroscopy is the study of light that is emitted by or reflected from materials and its variation in energy with wavelength [18]. Typically, applying light to the imaged target, which can be any type of light source in the laboratory or sunlight in the field, in order to diffuse the reflectance and the scattering of targeted materials at the surface [18] [19] [20]. In the hyperspectral imaging system, the spectrometer measures the reflected light intensity from different materials. The semi-automated algorithm in the imager system will then compare the collected information to a built-in spectral library for the detection purpose. The detection starts when an optical dispersing element such as a (grating) or prism (that can be seen in Figure 2.4) in the spectrometer splits this light into many wavelength bands and the energy in each band is measured by a separate detector [65]. The spectrometer then will be able to build a set of spectral measurements with using many detectors together, and get

measurements of spectral bands as narrow as 2 nanometers over a wide wavelength range [66] [67].

Depending on the application, and the optical component capability, the wavelength range includes visible to infrared. The hyperspectral imager that is used for this work covers the visible spectrum that corresponds to a wavelength range of 400 – 700nm. Hence, the work provided in this study cases is designed to focus on the measurements of the light reflected from many adjacent areas on the tissue sample surface on the pathology slides. The recent hyperspectral imaging (HSI) systems, uses a snapshot imager, which allows creating the whole hyperspectral image cube within one snapshot. However, in order to understand how the image is built, Figure 2.8 shows how one hyperspectral reflectance build each pixel in the hypercube.



**Figure 2.8:** Schematic diagram of the basic elements of an imaging spectrometer

### 2.4.3 Hyperspectral imaging and spectrometer

This section provides the different terms and keywords that are very important in the understanding of a hyperspectral system, from spectral reflectance to spectral signature to spectral library:

**A) Spectral Reflectance:** Spectral reflectance is the key to understanding how hyperspectral images are created and built. It is basically the physical and fundamental characteristics in spectroscopy of reflected light considering the reflected energy as a function of wavelength to measure the spectral information [66]. Reflectance differs within a wide range of wavelengths in almost every different material because energy at certain wavelengths is scattered or absorbed to different degrees [67]. These reflectance variations are evident when comparing spectral reflectance curves for different materials (Figure 2.3) [17]. The overall shape of a spectral curve and the position and strength of absorption bands, in many cases, can be used to identify and discriminate different materials [65].

**B) Spectral Signature:** As reflectance is a physical property of the object surface. It is defined as the percentage of incident electromagnetic radiation of each wavelength that is reflected by the object [20]. In addition, each object will have different spectral information depending on the physical and chemical properties. That spectral information is called the spectral signature, which means that every material will have a spectral signature value, or range of values [20].

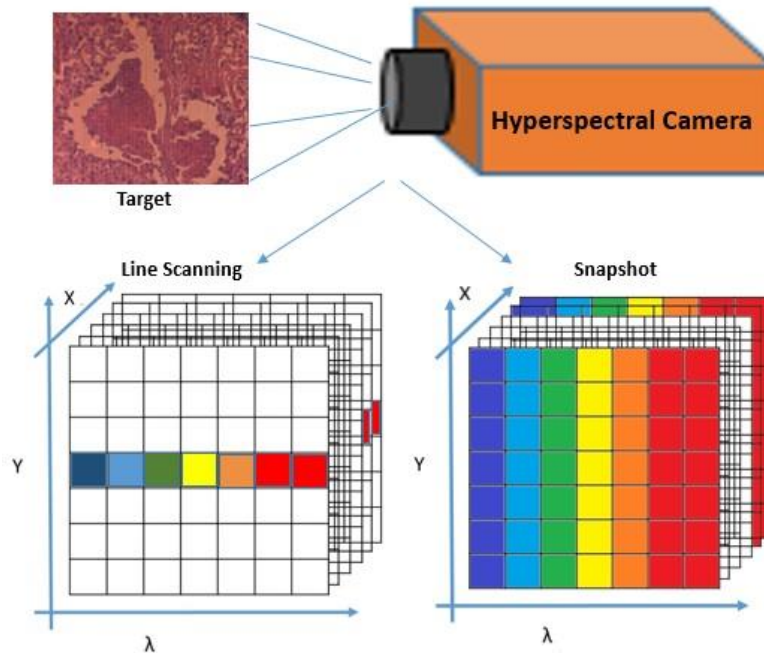
**C) Materials and Spectral Imaging:** Spectral reflectance that indicates the spectral signature can be observed in two different ways: 1) the wide range of the wavelength spectral dimension which includes visible light, near infrared, infrared and the rest of the spectrum and 2) the structural and chemical makeup of the material itself. That is the reason dealing with hyperspectral data from capturing the land surface, is different than the data in which was generated by captured human tissue. However, the basic principles of using the hyperspectral imaging will stay the same [64]. One of the main parameters that can show differences in the spectral information in different materials is chemical composition and crystalline structure, which has a very strong impact on the trend of the spectral curve [68]. In addition, wavelength, which is considered the spectral dimension, will reflect at different values of absorption of each particular material. This variation may be caused by the presence of particular chemical elements or ions, the ionic charge of certain elements, and the geometry of chemical bonds between elements, which is governed in part by the crystal structure of the material [68]. In this study, however, the work is applied on human and mice breast tissue, where the differences within the tissue is very slim because the combination of

cells and parameters within tissues have a lot in common that make a very close spectral curve. In contrast, cell behavior and some chemical elements such as protein, effects cell and tissue activity that cause a difference in the light absorbance, which result in different spectral reflectance values [69].

**D) Spectral Library and Spectral Detection:** A spectral signature is the spectral information that can be used to differentiate one material or groups of materials from another [70]. In general cases, the spectral signature of a certain material or element is the same, and the absorbance and reflectance that reflects the physical and chemical properties will not change. In addition, the spectral information of each material within a certain range is saved in what is called a spectral library. The spectral library is a digital library created by researchers at the Spectroscopy Lab by measuring the spectral reflectance of hundreds of materials in a laboratory setting, creating a useful reference resource for material identification in a variety of applications [70].

#### 2.4.4 Hyperspectral data cube

When comparing between hyperspectral line scan, and hyperspectral snapshot imaging methods, it is important to look at both imaging techniques, as well as advantages and disadvantages of both systems. In the case of line scan, each image collects one wavelength after the other, and the field of view of the imaging system is fixed in order to build the hypercube as shown in Figure 2.9 [26], [27]. However, in the case of the snapshot method, both spatial and spectral information of the target are captured with one exposure [29]. The snapshot method is an imaging technique that does not require scanning at all. As shown in Figure 2.9, the snapshot camera, has the capability to produce a complete spectral data cube in a single integration by directly imaging the target zones onto the spectral, and spatial detectors simultaneously [28], [29]. In the line scan method, time and stability are required, and it is necessary to wait until all wavelength images have been recorded, which that takes between seconds to minutes of measurement time plus few seconds between scans depending on the imaged target and the imager capability from illumination and integration conditions [25].



**Figure 2.9:** building the spectral data cube in both line scan, and snapshot systems.

Thus, the accuracy of the hyperspectral data cube may decrease if an object is moving. The snapshot method on the other hand, does not need the time that the line scan uses acquire to build the hyperspectral data cube, which is an advantage over using the line scan system, however, its spatial and spectral resolutions are limited as the total number of voxels cannot exceed the total number of pixels on the camera sensors [28]. In addition, both methods should produce spatial and spectral image quality across the output field of the spectrograph that is essentially the same at all points on the detector, to result in an image with high spectral resolution [27]. However, if the overall resolution is given by a certain number per channel, the line scan will produce an image per line to have the full resolution per spectral channel, while the resolution is divided by number of spectral channels, in case of snapshot, which then reduces the overall image resolution.

#### 2.4.5 HSI Image processing

Image processing is a method that is used to get more information and to characterize an image. It converts the image into digital form and applies various analysis methods in order to get a better understanding of the image and to extract some useful information that cannot be seen or observed. Usually, image processing systems treat images as two dimensional signals while applying already

set signal processing methods to them [49], [50]. Image processing can also be used to manipulate images, and apply many different operations in order to extract, detect, and distinguish new information. That however, will allow the user to visualize and present images in different and unique way [50], [73]. At the hyperspectral processing stage, geometric registration and calibration might be performed to make it possible to compare this data set with other data sets. Feature extraction may also be conducted to reduce the dimensionality of data and at the same time to increase accuracy by increasing the separability of various classes in the image. In this study, the data is normally divided into 2 classes, positive cancer, and negative cancer, which are referred to as 0 and 1. At last, the data goes through an interpretation stage depending on different applications, which means for different needs there are different processing methods to make full use of the data. As in this work, class recognition is only needed for detection, so the data is used to identify between classes. More details will be explained in the algorithm sections according to the type of data and algorithms used, whether the data is supervised or unsupervised.

**A) Image Clustering:** The clustering of an image, is an advanced recognition of the complex characterization of an image [71], [73]. It also describes an image from the dataset level, with certain information required to be able to categorize a certain image or set of images into multiple classes or clusters, also be described as image mapping. Image clustering can allow editing and dealing with the dataset as group of classes, which is very useful in many image processing applications such as feature imaging and image representation [71]. Organization of feature data is used to classify an image to a certain cluster.

**B) Image Classification:** Classification of an image creates multiple classes depending on the data contained in the image, so each class will have specifications to group them together as one class. Classification also refers to the task of extracting information classes from an image [72], [73]. The classifier creates a special environment for each dataset in order to classify each data. Classification is a process of multiple steps of studying the data and creating labels, training, and testing steps. Classification gets more complex with the size and complexity of the data. There are two types of classification: supervised and unsupervised [73].

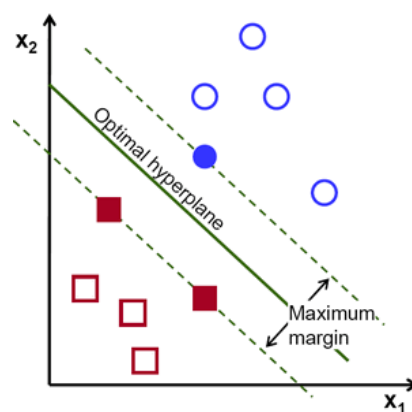
B-1) supervised classification: A very simple description of a supervised classification is when the classification is given to a dataset that is labeled based on a key feature, such as the spectral signatures obtained from training samples, to classify an image [73], [74]. Then the classifier can create a signature file from the training samples, which is then used by the multivariate classification to classify the image.

B-2) unsupervised classification: The opposite of supervised classification, unsupervised classification where classifier is given dataset without labels [73], [75]. The classifier then finds spectral classes in a multiband image without the any assigned values from the data provider. The Classification treats unsupervised data by providing access to the tools to create the clusters, capability to analyze the quality of the clusters, and access to classification tools [75].

#### 2.4.4.1 Examples of classification algorithms:

A) Support Vector Machine (SVM) algorithm for supervised data.

SVM can be described as a classifier that can identify and create a hyperplane that separates between the different classes [55], [56]. In other words, given labeled training data, the algorithm outputs an optimal hyperplane which categorizes an examples figure 2.8. This will happen only if the data is linearly separable. The optimal hyperplane is the best hyperplane that has the maximum separation distance, as shown in Figure 2.10.



**Figure 2.10:** Hyperplane separating two classes (LSVM) [55]



In the case of Linear SVM, the hyperplane an equation is applied to get the optimal hyperplane with the max margin. The SVM linear classifier relies on a dot product between data point vectors.

if.  $K(\vec{x}_i, \vec{x}_j) = \vec{x}_i^T \vec{x}_j$  Then the classifier is given by:

$$f(\vec{X}) = \text{sign}(\sum_i \alpha_i y_i K(\vec{x}_i, \vec{X} + b)) \quad \text{Equation 2.8[56]}$$

In MATLAB, the classifier that was used to analyze the dataset generates a nonlinear classifier with a Gaussian kernel function. First, it generates one class of points inside the unit disk in two dimensions, and another class of points in the annulus from radius 1 to radius 2. Second, it generates a classifier based on the data with the Gaussian radial basis function kernel. The default linear classifier is obviously unsuitable for this problem, since the model is circularly symmetric [57]. Then the classifier will identify the classified with the misclassified points by generating two sets of 100 points as follows: Generate 100 points uniformly distributed in the unit disk. To do so, generate a radius  $r$  as the square root of a uniform random variable, generate an angle  $t$  uniformly in  $(0, 2\pi)$ , and put the point at  $(r \cos(t), r \sin(t))$ , then generate 100 points uniformly distributed in the annulus. The radius is again proportional to a square root, this time a square root of the uniform distribution from 1 through 4. Finally, the classifier will compare between both [56], [57].

## B) K-means for unsupervised data.

K-means is a learning algorithm that is applied to solve unsupervised data classification problems [54]. It starts first by choosing number of classes (clusters). The main idea is to define  $k$  centers, one for each cluster. These centers should be placed in a cunning way because of different  $k$  center locations causes different result. So, the better choice is to place them as much as possible far away from each other [58]. The next step is to take each point belonging to a given data set and associate it to the nearest center. In this case each class will have one centroid, and the result becomes more stable with increase iterations Also the more data the classifier is given, the better the centroid accuracy will be [58], [59]. The equation is given by:

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2 \quad \text{Equation 2.9[58]}$$

where ' $\|x_i - v_j\|$ ' is the Euclidean distance between  $x_i$  and  $v_j$ , ' $c_i$ ' is the number of data points in  $i^{th}$  cluster, and ' $c$ ' is the number of cluster centers.

The process starts by randomly selecting ' $c$ ' cluster centers, then calculating the distance between each data point and cluster centers. After that, a data point is assigned to the cluster center whose distance from the cluster center is the minimum of all the cluster centers. The new cluster center is then recalculated using:

$$v_i = \left(\frac{1}{c_i}\right) \sum_{j=1}^{c_i} x_i \quad \text{Equation 2.10[59]}$$

where ' $c_i$ ' represents the number of data points in  $i^{th}$  cluster.

Then, again distance between each data point and newly obtained cluster centers is recalculated. Finally, if no data point was reassigned, the calculation loop is stopped. Otherwise, it is repeated until no changes in the centroid values occur [58].

## 2.5 Summary

In summary, Hyperspectral imaging is a very useful tool that can be applied in many different applications. As an imaging system, it can be adopted to be used in almost any application that can use optical imaging from the army to educational institutes, to hospitals and industries. Image processing is undeniably important to work with HSI, as the images needs to be analyzed, and processed. In this work the algorithms that were used, focused on both classification methods for the supervised and unsupervised data to measure the imaging system abilities to detect the spectral difference between cancer and normal. The next chapters will provide the different approaches of using both algorithms, the accuracy and results.

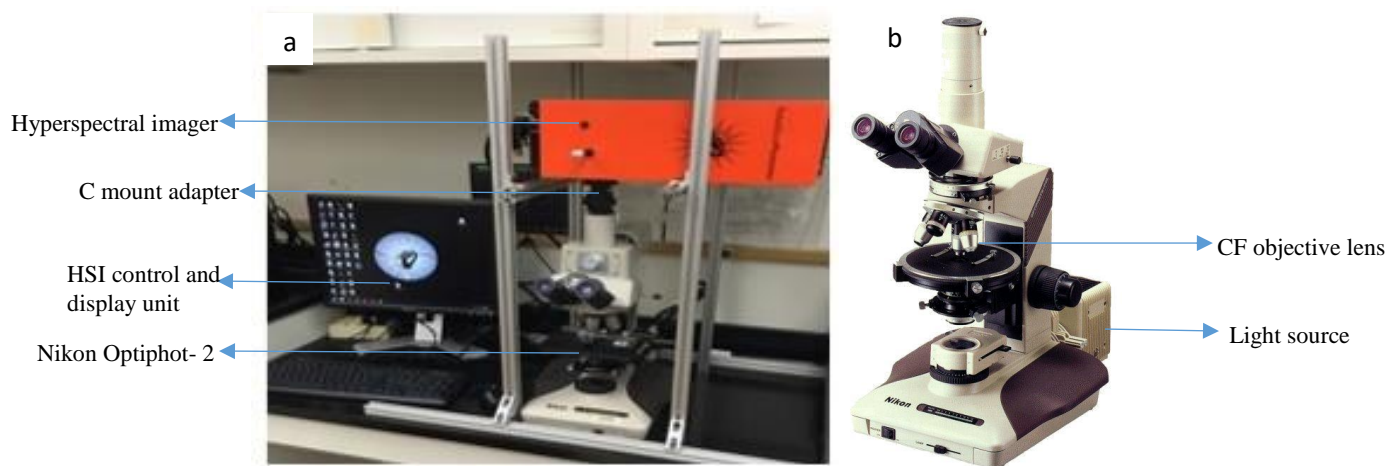
# Chapter 3

## 3. HSI System Evaluation, and Cellular Imaging

### 3.1 Introduction

This chapter describes the details of HSI system setup & testing experiment. Prior to experimentation, the hyperspectral imager was connected to a microscope and calibrated based on a standard process. Section 3.2 in this chapter provides the experiment applied on breast cancer tissue sample in order to make sure that the hyperspectral system is ready for the main study of this thesis. Sections 3.3 to 3.5 are the cellular imaging sections that provides more details about the main studies and the imaged samples.

The hyperspectral imager was connected to a microscope as shown in Figure 3.1-a. The upright microscope used in this setup is a Nikon Optiphot-2 (Figure 3.1-b) with a 12V-100W LL halogen lamp illumination source. The microscope has a three filters, such as a neutral-density filter, which were not used in this study. The objective lens that was used to capture all of the images in this study is CF Achromat P40x.



**Figure 3.1:** a- Hyperspectral imaging system setup Davis lab in WVU Cancer Institute. b - Nikon Optiphot-2

The Arrow snapshot hyperspectral imaging camera from Rebellion Photonics is capable of two imaging configurations with two different sets of spectral bands. The configuration that was used in this study is 443 x 313 pixel resolution in the spatial domain with 31 bands in spectral domain. The spectral range of this system is within the visible light spectrum between 461nm and 641nm. The field of view of the hyperspectral camera attached to the microscope with a 40x objective in its best resolution is 100 $\mu$ m X 80 $\mu$ m, allowing the capture of many images from a single tissue sample, which averaged 8.0x10<sup>-5</sup> cm<sup>2</sup> for this study,

## 3.2 Hyperspectral imaging experiment

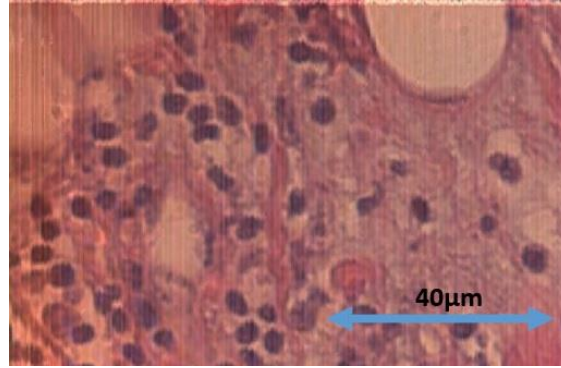
This section provides a description of the tissue sample imaging process. The samples are grouped into three main types:

1. H&E stained of normal and DCIS samples
2. Unstained samples of normal and DCIS tissues
3. In situ and invasive tissue samples of mouse mammary tumor

In the first imaging experiment, the hyperspectral imaging system was applied on 10 samples from patients that have tested positive for invasive ductal carcinoma breast cancer. Each patient has two H&E tissue samples marked by WVU Pathologist, Dr. James Coad. The West Virginia University Institutional Review Board approved this investigation [Protocol 1509816662 (Non-Human Subject Research)]. A copy of the approval letter can be found in Appendix C. The second imaging experiment includes set of breast tissue samples taken from the same 10 patients, but without staining in order to observe the spectral signature of both normal and cancer tissues in an unstained state. The last imaging experiment is performed on samples taken from mice that have tested positive for invasive ductal carcinoma (IDC) breast cancer and Ductal Carcinoma in Situ (DCIS). The West Virginia University Research Compliance Office and Institutional Animal Care and Use Committee approved the use of animals from which the mammary tissues were rendered [ACUC Protocol 10-1107]. Each mouse has two H&E tissue samples of IDC and DCIS marked by the pathologist. This chapter will show a few images of each samples set, with the remaining images included in Appendix A.

### 3.2.1 Preparation methods of the Pathology Samples

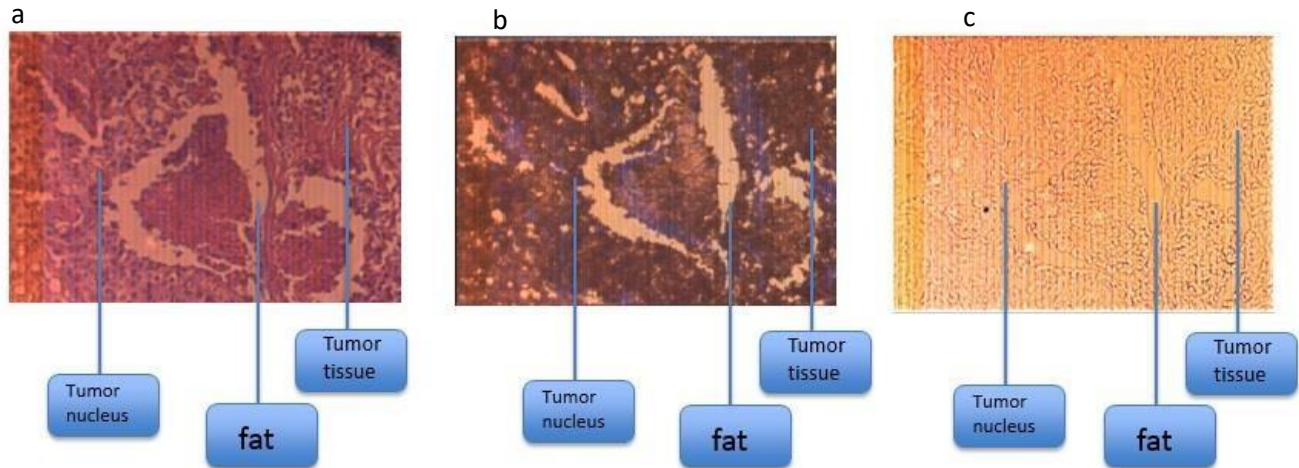
This study will apply hyperspectral imaging technology to provide an indication of the spectral reflectance differences of ductal carcinoma tissue samples. Figure 3.2 shows an example of H&E stained ductal carcinoma biopsy sample taken from a breast cancer patient.



**Figure 3.2:** Hyperspectral image of a ductal carcinoma tissue.

In this study, we have collaborated with the Pathology Department at WVU hospital to obtain breast ductal carcinoma tumor biopsies fixed on pathology slides prepared in three different as follow:

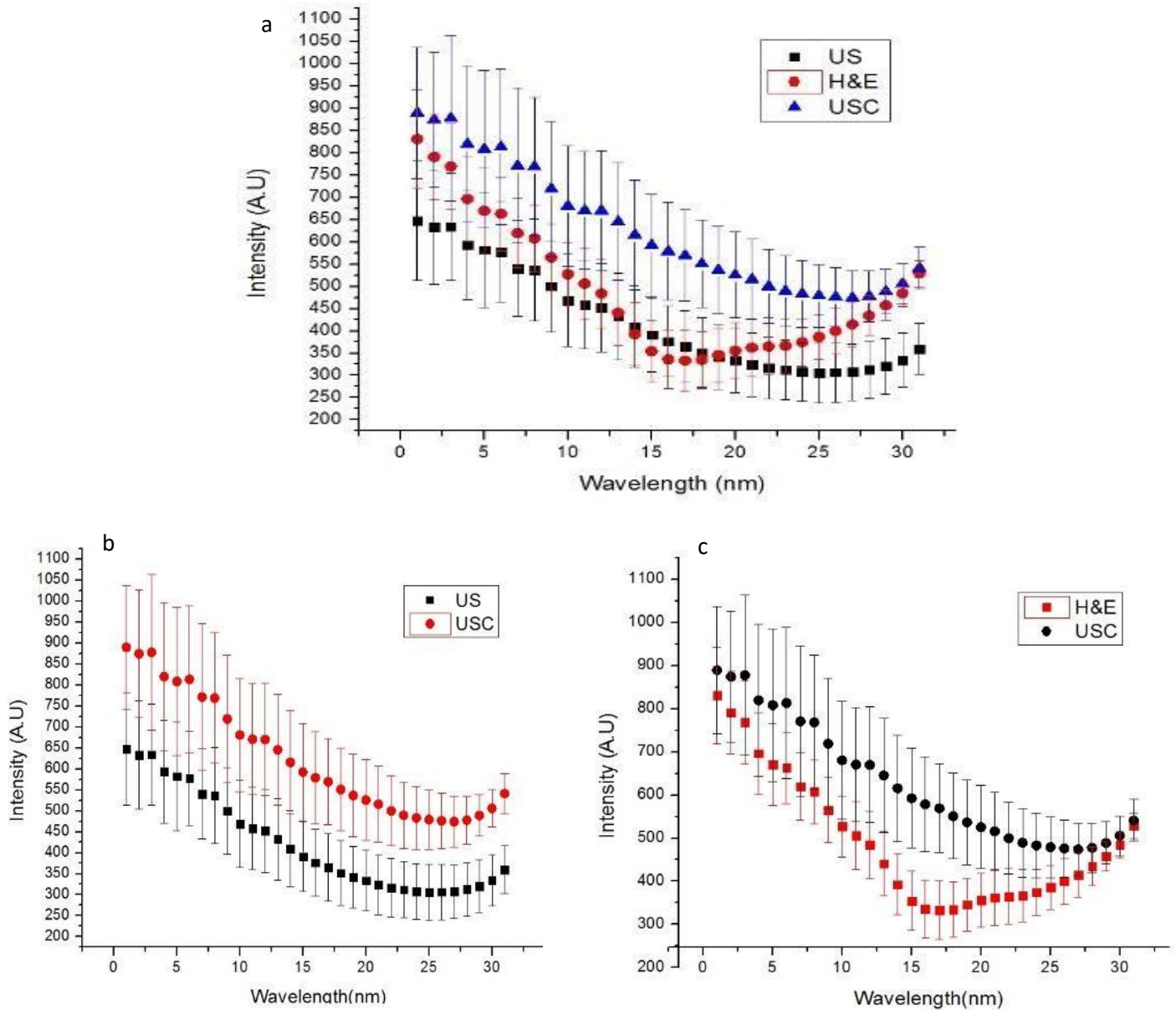
- 1) The commonly used method in pathology is hematoxylin and eosin (H&E) staining with glass coverslip shown in Figure 3.3-a.
- 2) Unstained tumor tissue directly on the slide shown in Figure 3.3-b, and
- 3) Unstained tumor on a slide with a glass coverslip shown in Figure 3.3-c.



**Figure 3.3:** The top three images: showing the three methods of preparing the tumor on the slides, (a) H&E staining, (b) Unstaining, and (c) Unstaining with glass coverslip.

### 3.2.2 Experiments Result

The main imaging experiment focused on the observations of three preparation methods of biopsy pathology slides. The spectra for each tissue sample type are shown in Figure 3.4.



**Figure 3.4:** (a) Reflectance spectra comparison of three tissue preparation methods: H&E stain, unstained with no cover slip, and unstained with cover slip. (b) Reflectance spectra comparison of unstained samples (US) and unstained samples with cover slip (USC). (c) Reflectance spectra comparison of H&E and USC, note that the wavelength axis is the band number between 1 to 31 band and it is ranged from 460nm to 640 nm

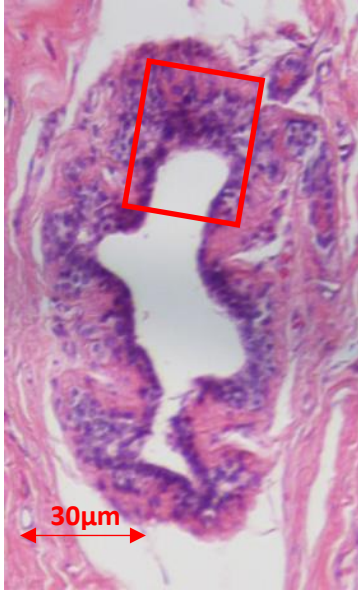


The error bars in Figure 3.4 are used to quantify uncertainty in the graph of based on the sample population of 50 pixels from each image focused on the same region of interest of the imaged tissue in order to show a clear spectral separation between the three preparation methods. Therefore the results of three plots from Figure 3.4-b shows a high degree of similarity in the spectra of the unstained samples, which means if we move the intensity of one, both will be on top of each other with no differences. Although the difference in the intensity values is clear enough for the algorithm that was used for analyzing the images to identify it, any change in the light source will result in change in the intensity. This result answers the first question posed in this initial study, which is: How does the glass coverslip affect the spectral value between two identical samples imaged in the exact same condition? Figure 3.4-c answers the second question, which is: Does the H&E stain affect the spectral reflectance value between two identical samples imaged in the exact same condition? The answer to this question is ‘yes,’ and the reason is because the interaction of light with the stain (absorption and reflection) will result in different spectral reflectance from the unstained samples. More details will be discussed in the next chapter as a part of main study of this work that is focusing on the spectral reflectance of the H&E stained as well as the unstained samples of normal, invasive, and pre-invasive ductal carcinoma breast tissues.

### 3.3 H&E stained Normal and Ductal Carcinoma samples

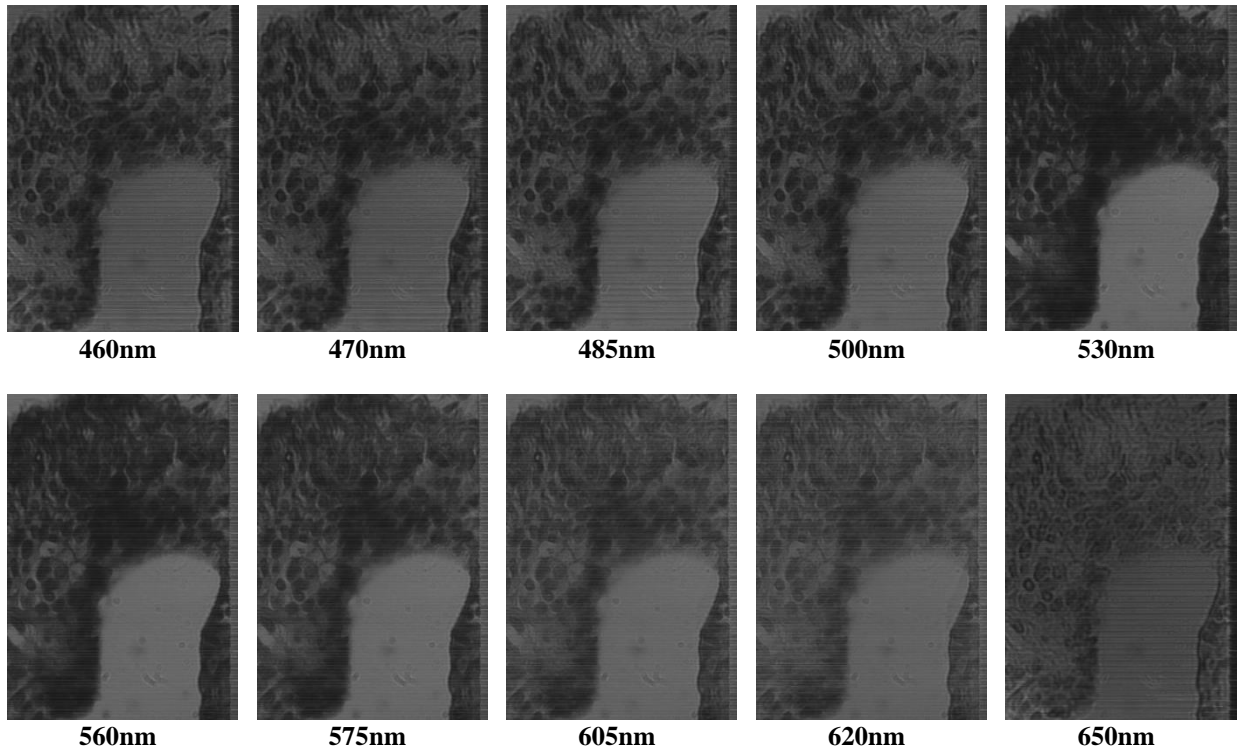
In this section, example images from biopsy slides shown from 3 out of 10 patients are displayed below for both of normal and DCIS tissue, the rest are included in Appendix A. Each sample was imaged by the hyperspectral imager, the analysis of which that will be discussed in Chapter 4. Figures 3.5, 3.7, and 3.9 show the H&E stained samples of the normal ducts of the three patients with a monochromatic HSI example images of different wavelength channels between 460nm and 650nm of the normal tissue images to show that the Arrow camera can produce images with clear detail of the tissue structure. Figures 3.6, 3.8 and 3.10 also show images of both H&E stained samples of the DCIS tissues of the three patients as well as the HSI images of the chosen wavelength channels within the same spectral range.



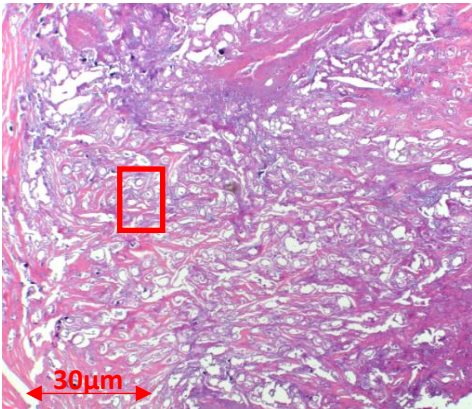


(a)

**Figure 3. 5:** (a) Visible image of H&E-stained normal duct tissue from Patient 1. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

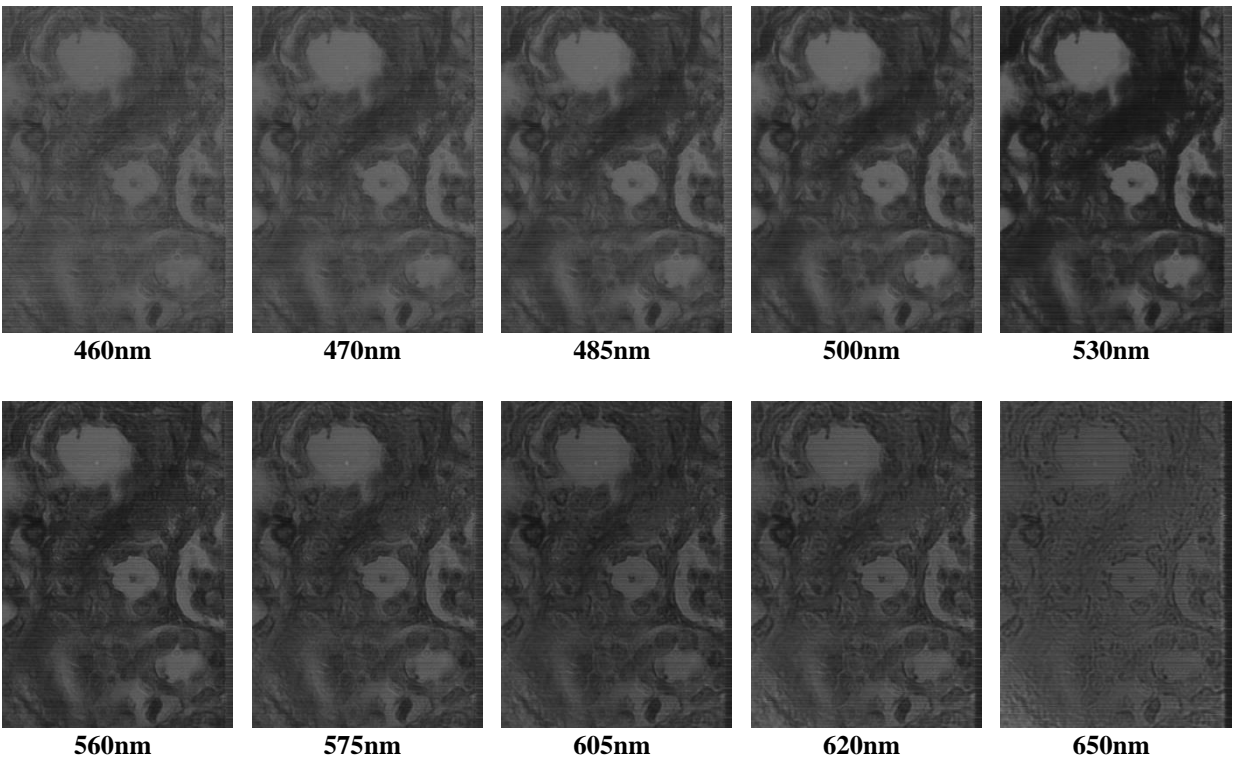


(b)

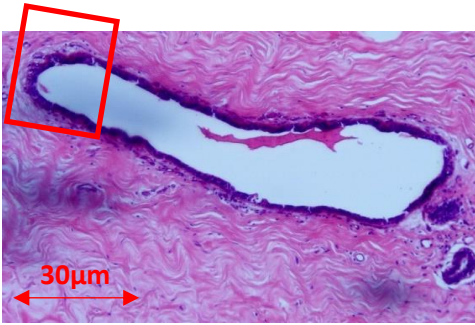


(a)

**Figure 3.6:** (a) Visible image of tissue sample containing ductal carcinoma from Patient 1. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

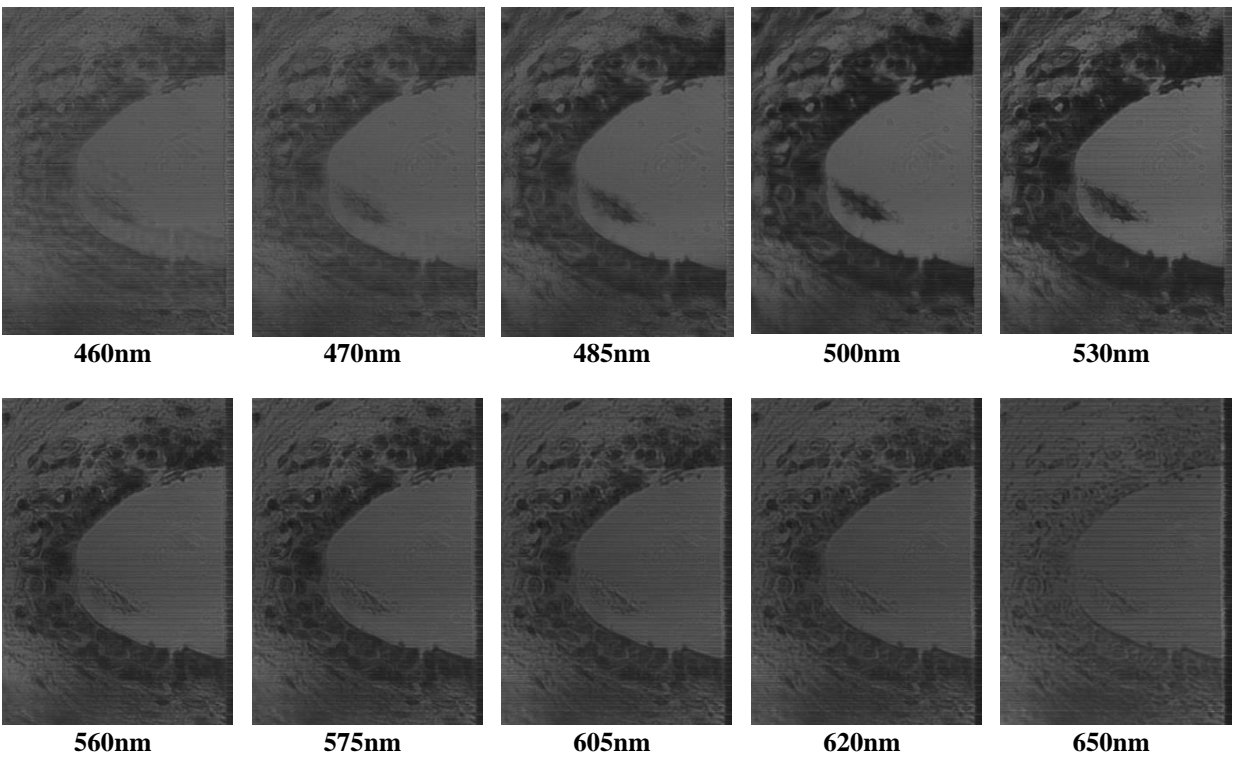


(b)



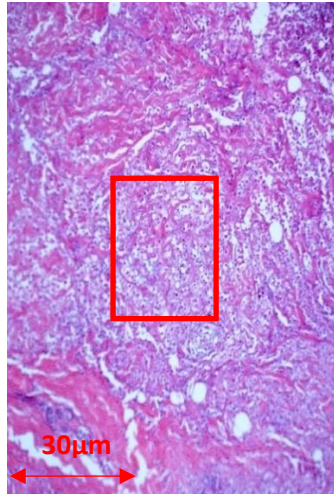
(a)

**Figure 3.7:** (a) Visible image of H&E-stained normal duct tissue from Patient 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.



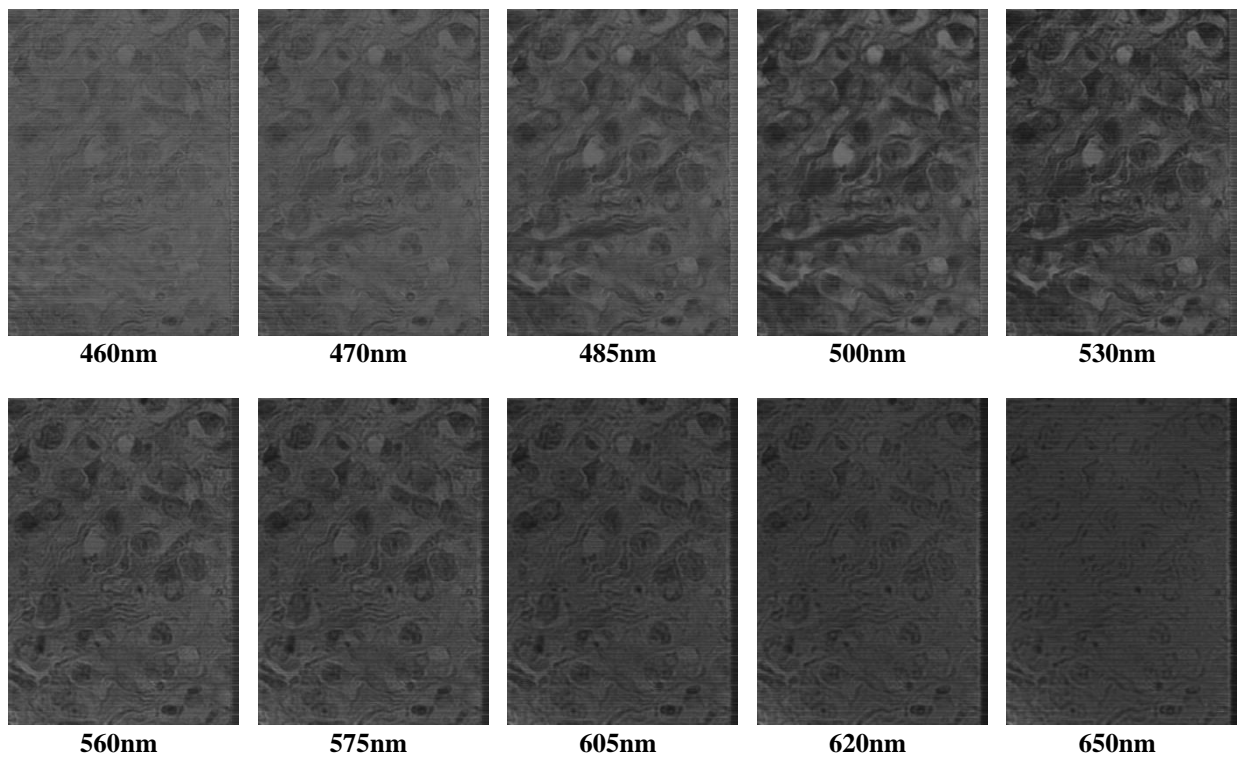
(b)



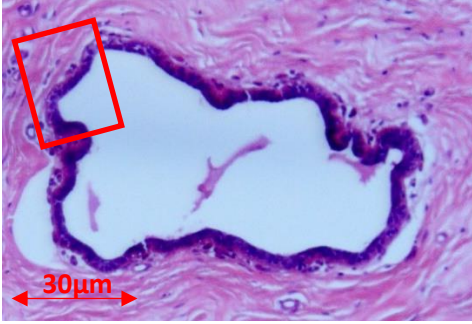


(a)

**Figure 3.8:** (a) Visible image of tissue sample containing ductal carcinoma from Patient 2. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

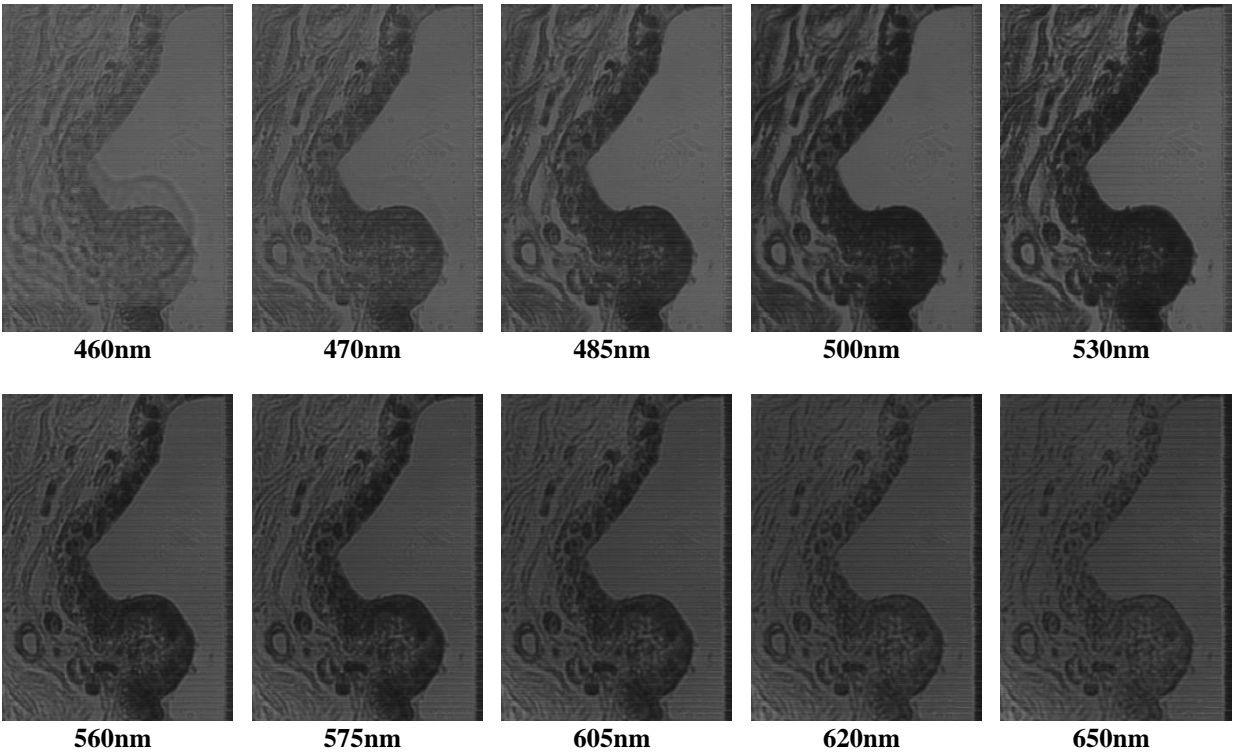


(b)

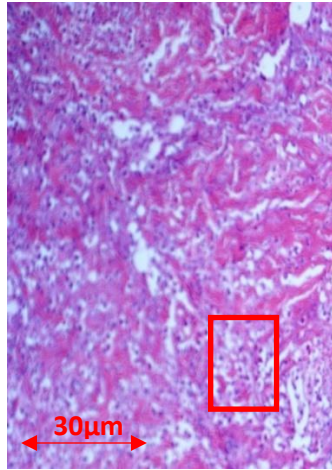


(a)

**Figure 3.9:** (a) Visible image of H&E-stained normal duct tissue from Patient 3. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

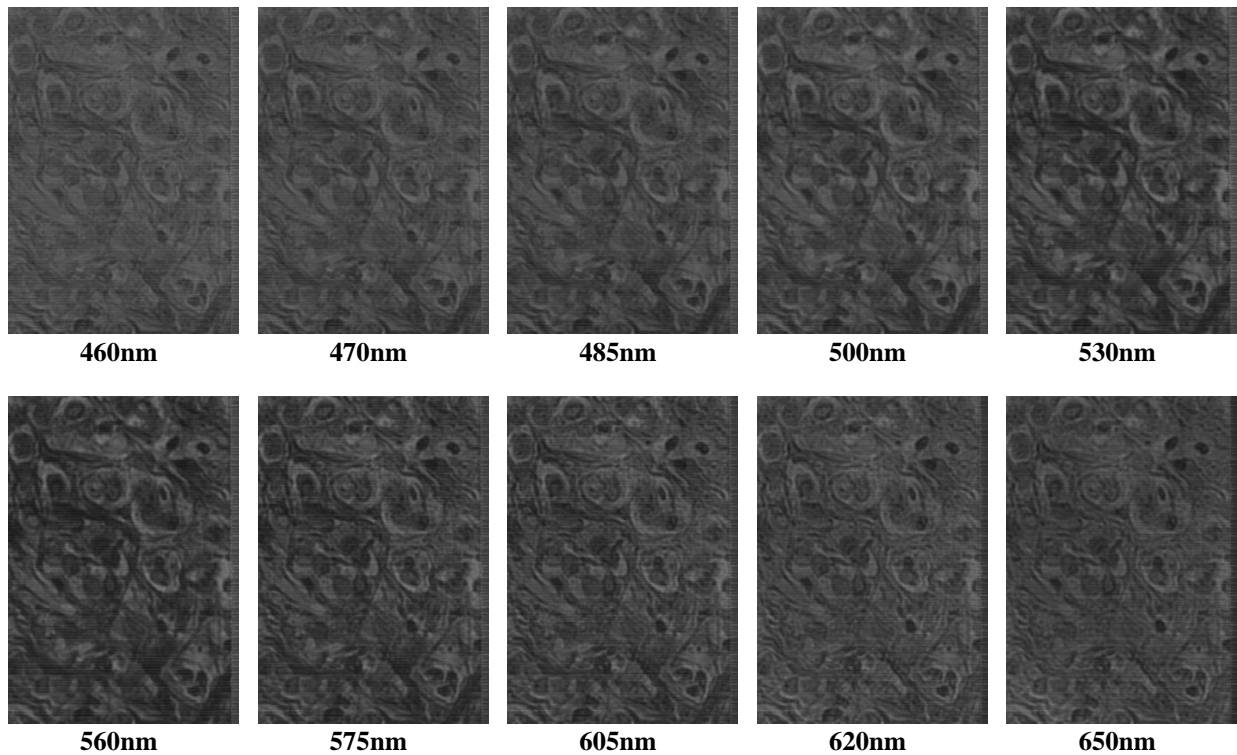


(b)



(a)

**Figure 3.10:** (a) Visible image of tissue sample containing ductal carcinoma from Patient 3. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

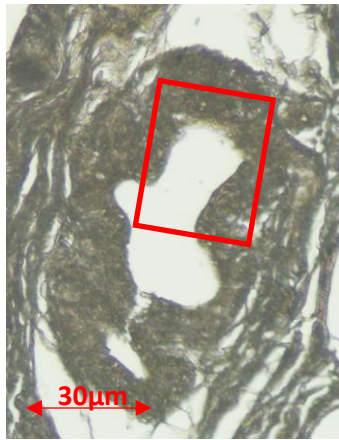


(b)

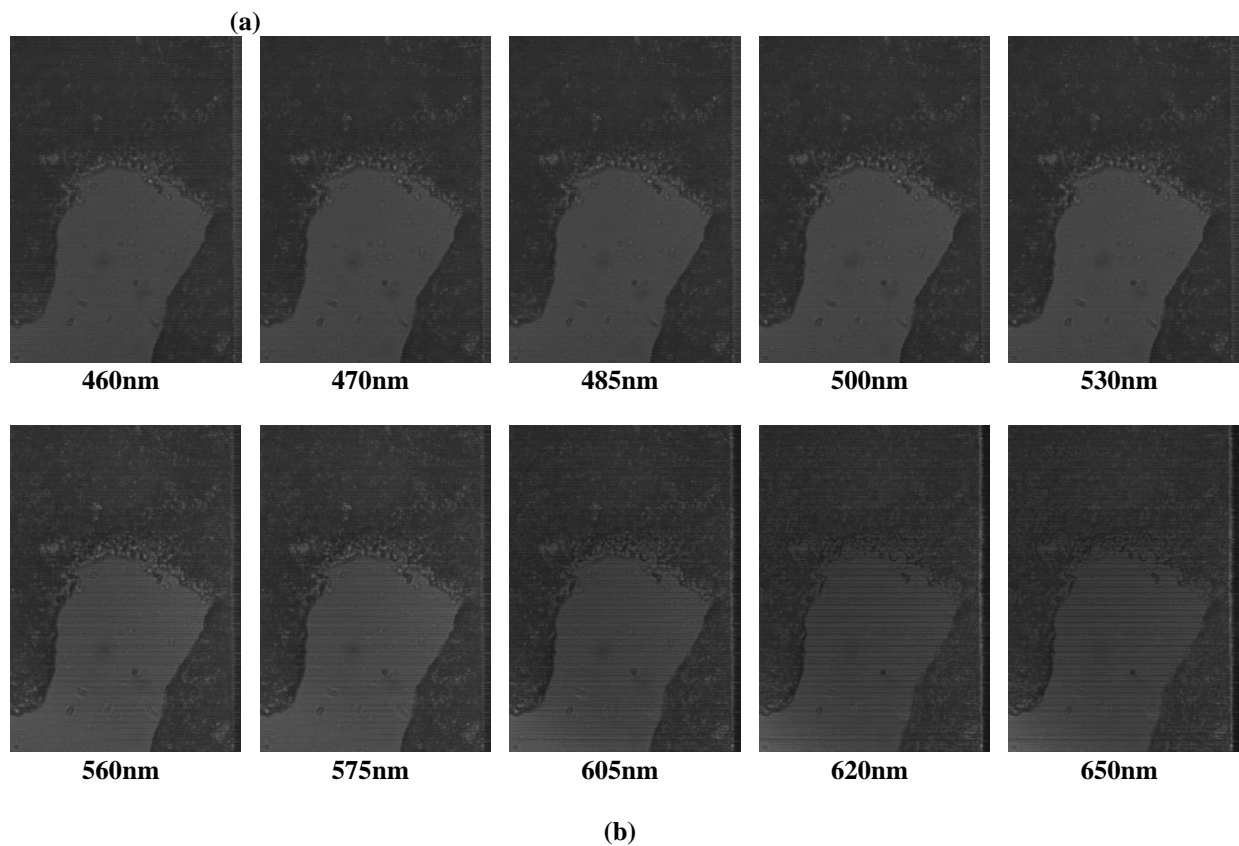
In this section, a visible images of H&E-stained normal duct, and DCIS tissues were shown from the three example patients, as well as the regions that were selected for hyperspectral imaging for select wavelengths ranging from 460-650nm. The next section will provide the images of the unstained tissue samples that was imaged and evaluated in this study.

### 3.4 Unstained Normal and Ductal Carcinoma samples

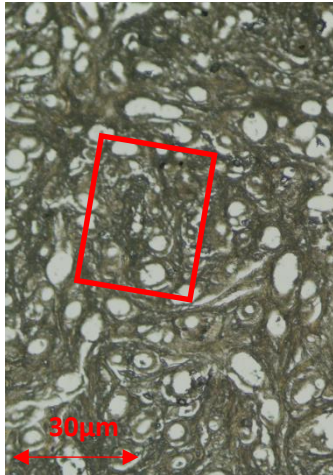
In this section also, using the same methodology described in Section 3.3, example images from 3 out of 10 patients are displayed below of both biopsy samples of normal and DCIS tissues, the rest are included in the appendix. However, the samples are unstained in this case. Each sample was imaged by the hyperspectral imager, the analysis of which that will be discussed in Chapter 4.



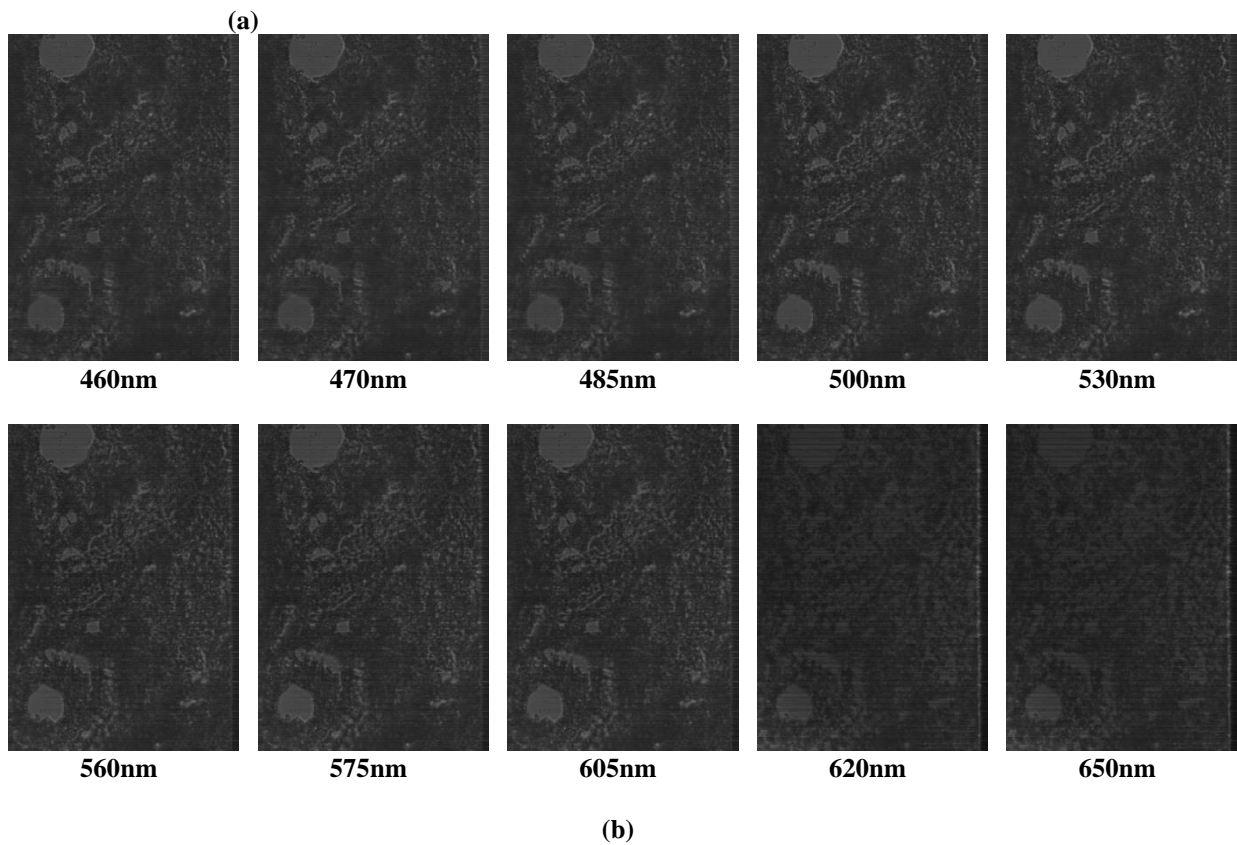
**Figure 3.11:** (a) Visible image of unstained normal duct tissue from Patient 1. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

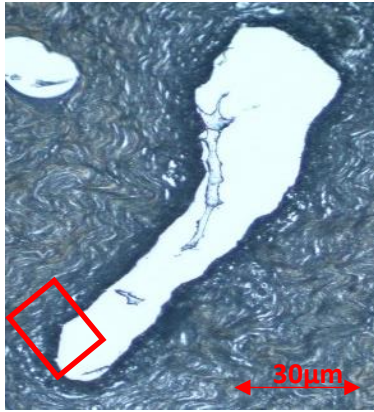






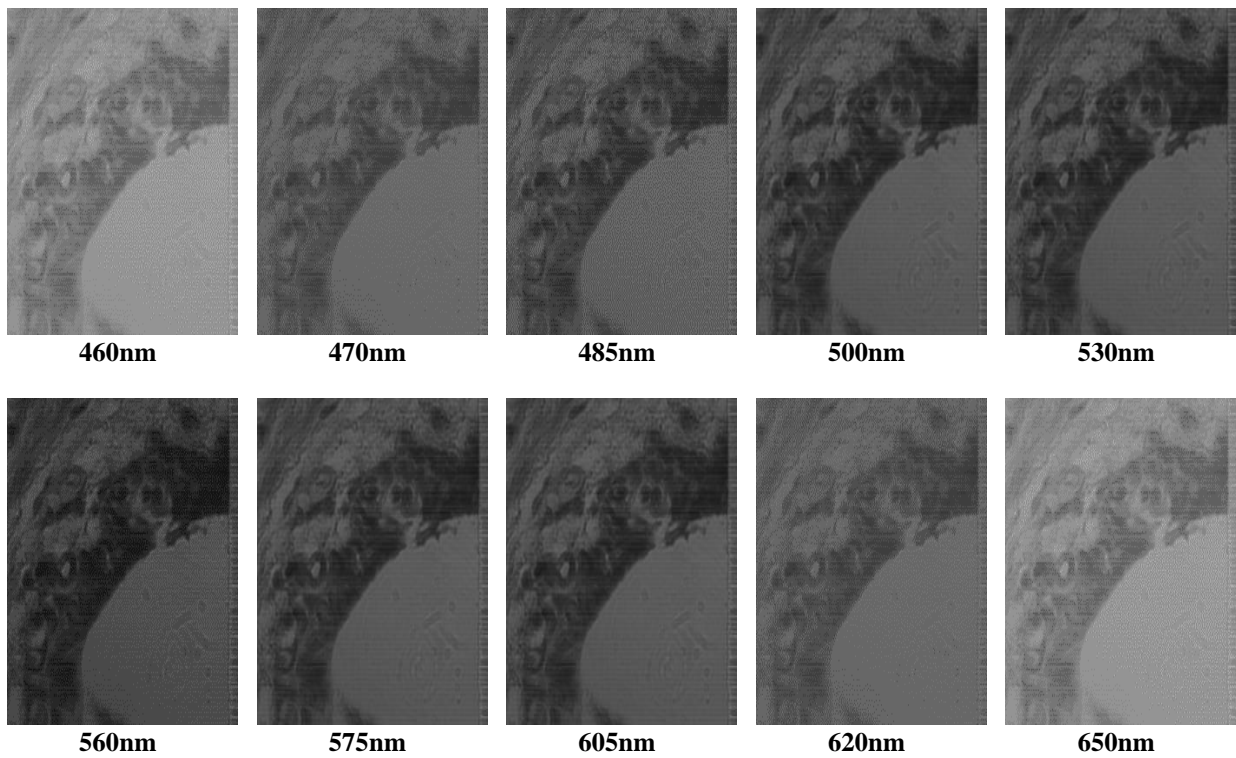
**Figure 3.12:** (a) Visible image of unstained tissue sample containing ductal carcinoma from Patient 1. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.



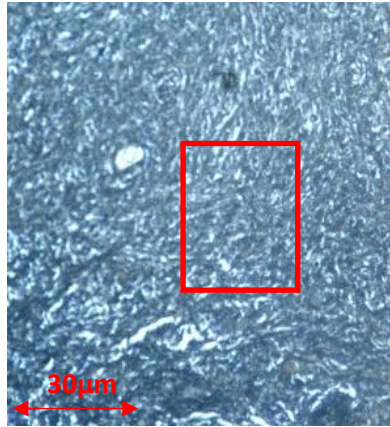


(a)

**Figure 3.13:** (a) Visible image of unstained normal duct tissue from Patient 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

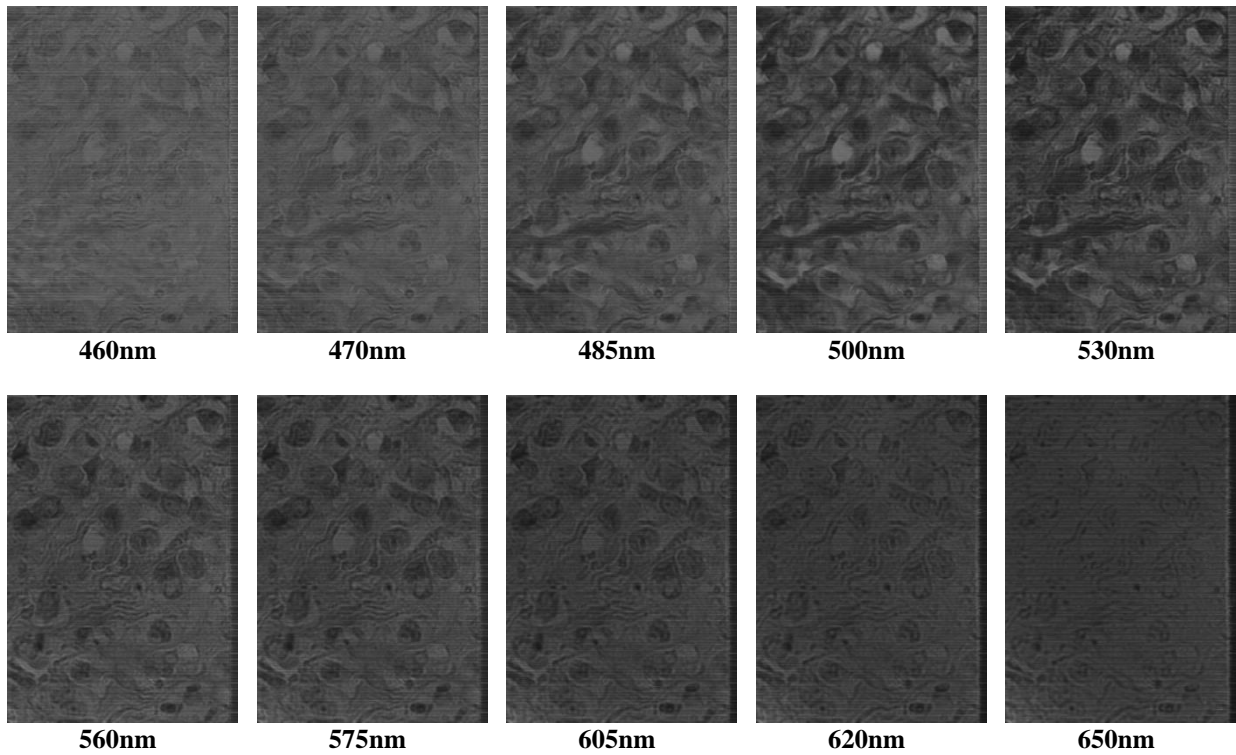


(b)

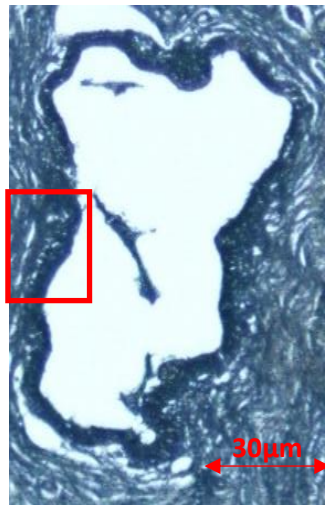


(a)

**Figure 3.14:** (a) Visible image of unstained tissue sample containing ductal carcinoma from Patient 2. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

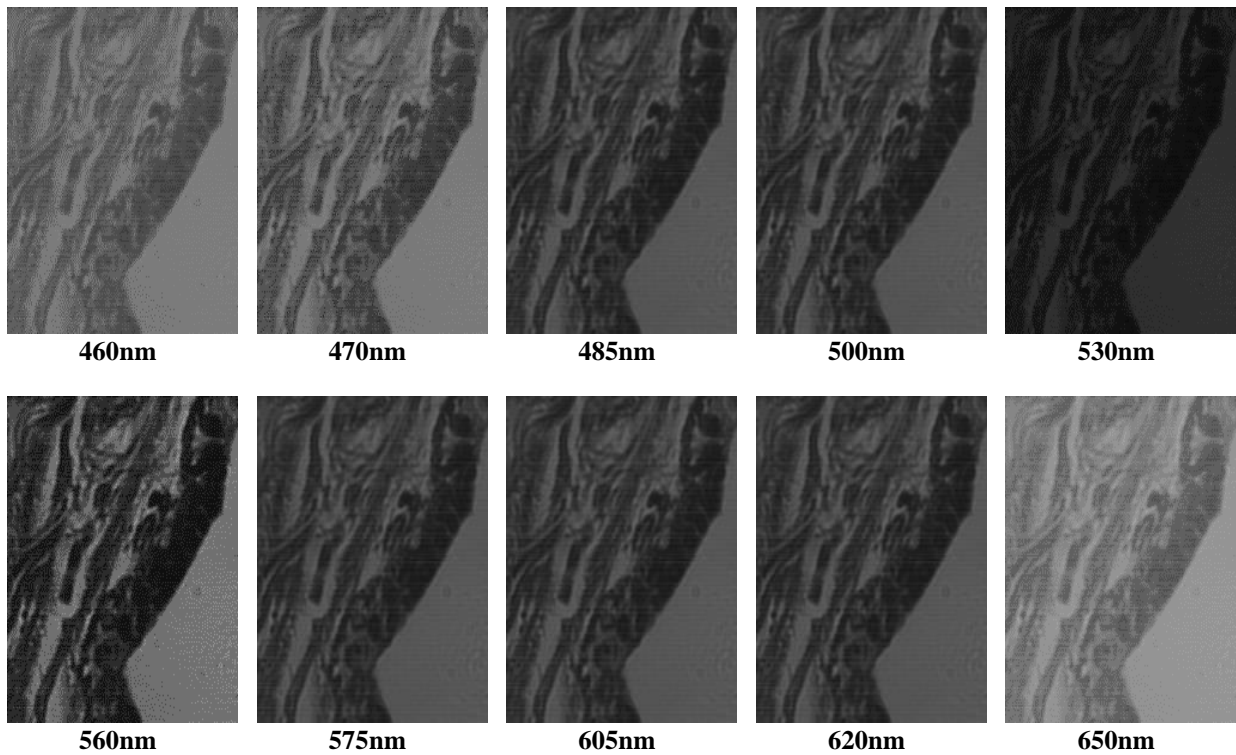


(b)



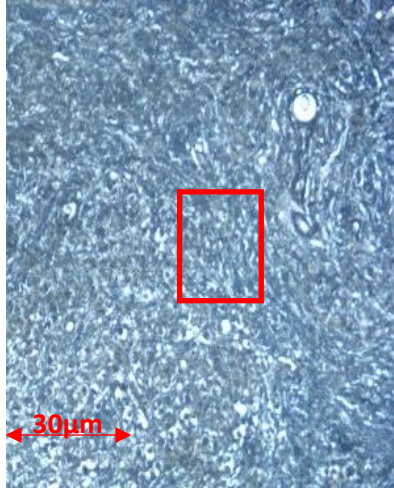
(a)

**Figure 3.15:** (a) Visible image of unstained normal duct tissue from Patient 3. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.



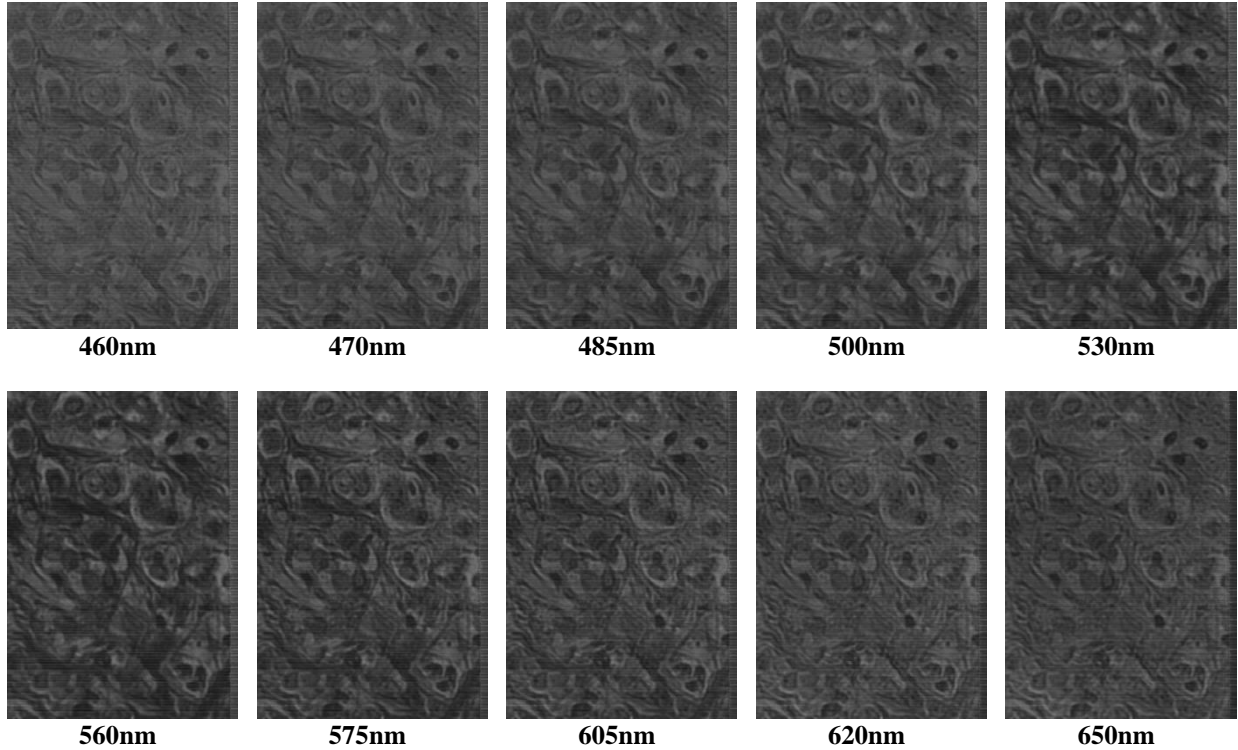
(b)





(a)

**Figure 3.16:** (a) Visible image of unstained tissue sample containing ductal carcinoma from Patient 3. Region inside red box a selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.



(b)

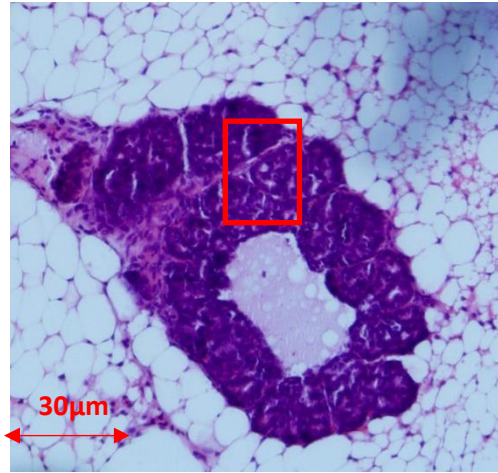
In this section, a visible images of unstained normal duct, and DCIS tissues were shown from the three example patients, as well as the regions that were selected for hyperspectral imaging for select wavelengths ranging from 460-650nm. Figures 3.11, 3.13, and 3.15 show the unstained samples of the normal ducts of the three patients with a monochromatic HSI example images of different wavelength channels between 460nm and 650nm of the normal tissue images to show

that the Arrow camera can produce images with clear detail of the tissue structure. Figures 3.12, 3.14 and 3.16 also show images of both unstained samples of the DCIS tissues of the three patients as well as the HSI images of the chosen wavelength channels within the same spectral range. The next section will provide the images of the unstained tissue samples that was imaged and evaluated in this study

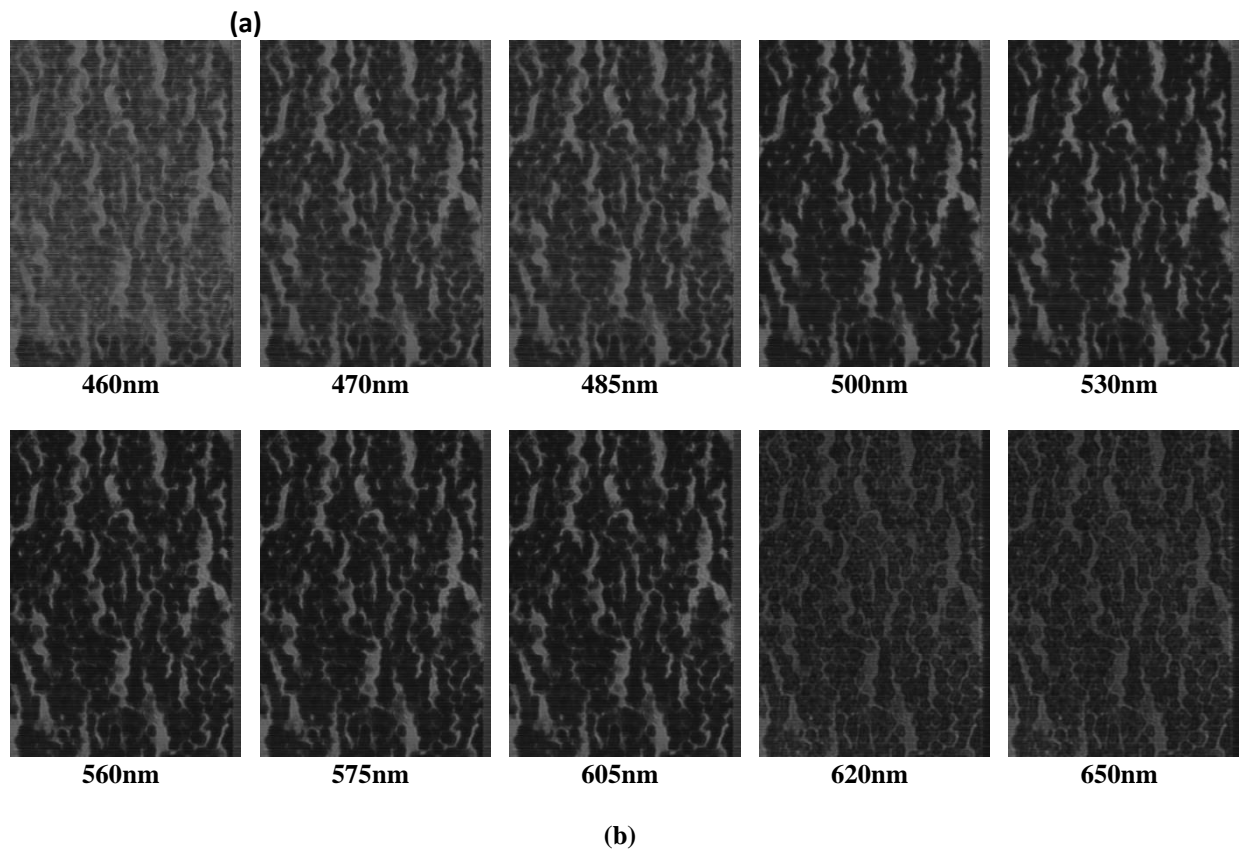
### 3.5 H&E Stained In Situ and Invasive Ductal Carcinoma

In this section, the methodology that was used for human breast tissue the previous sections is applied on 4 tissue samples from mice that have tested positive for invasive ductal carcinoma (IDC) breast cancer and Ductal Carcinoma in Situ (DCIS). Each mouse sample has two H&E marked pathology samples of IDC and DCIS tissue. Each sample was imaged by the hyperspectral imager, the analysis of which that will be discussed in Chapter 4.

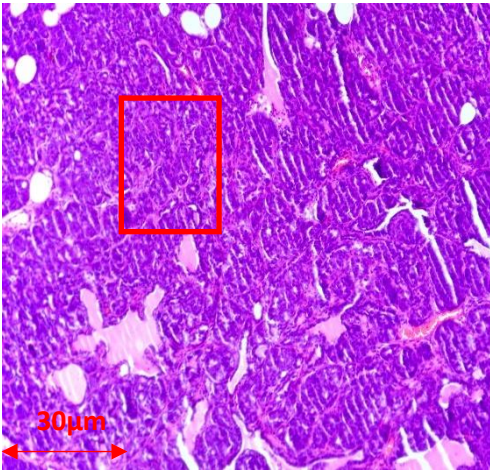
Figures 3.17, 3.19, and 3.21 show the H&E stained samples of the DCIS of the three mice with a monochromatic HSI example images of different wavelength channels between 460nm and 650nm of the normal tissue images to show that the Arrow camera can produce images with clear detail of the tissue structure. Figures 3.18, 3.20 and 3.22 also show images of both H&E stained samples of the IDC tissues of the three mice as well as the HSI images of the chosen wavelength channels within the same spectral range.



**Figure 3.17:** (a) Visible image of H&E-stained DCIS tissue from mouse 1. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

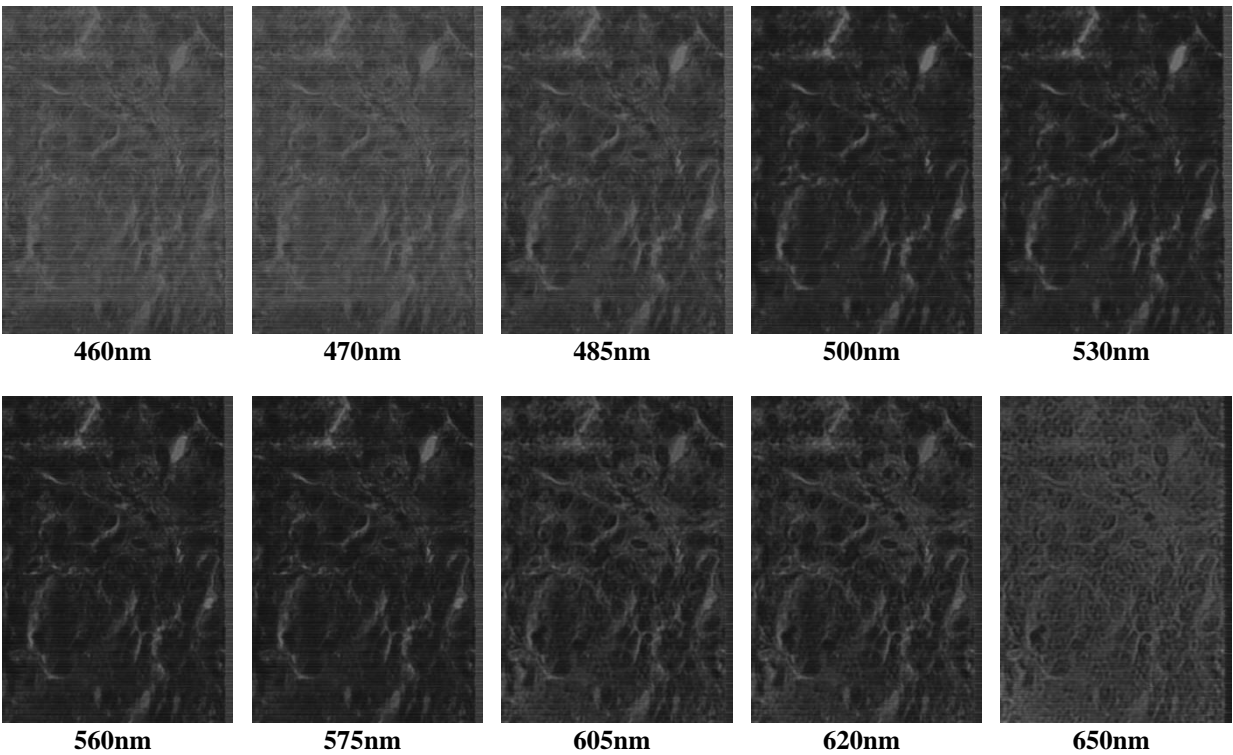




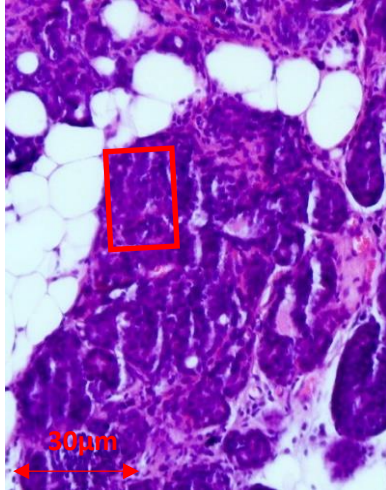


(a)

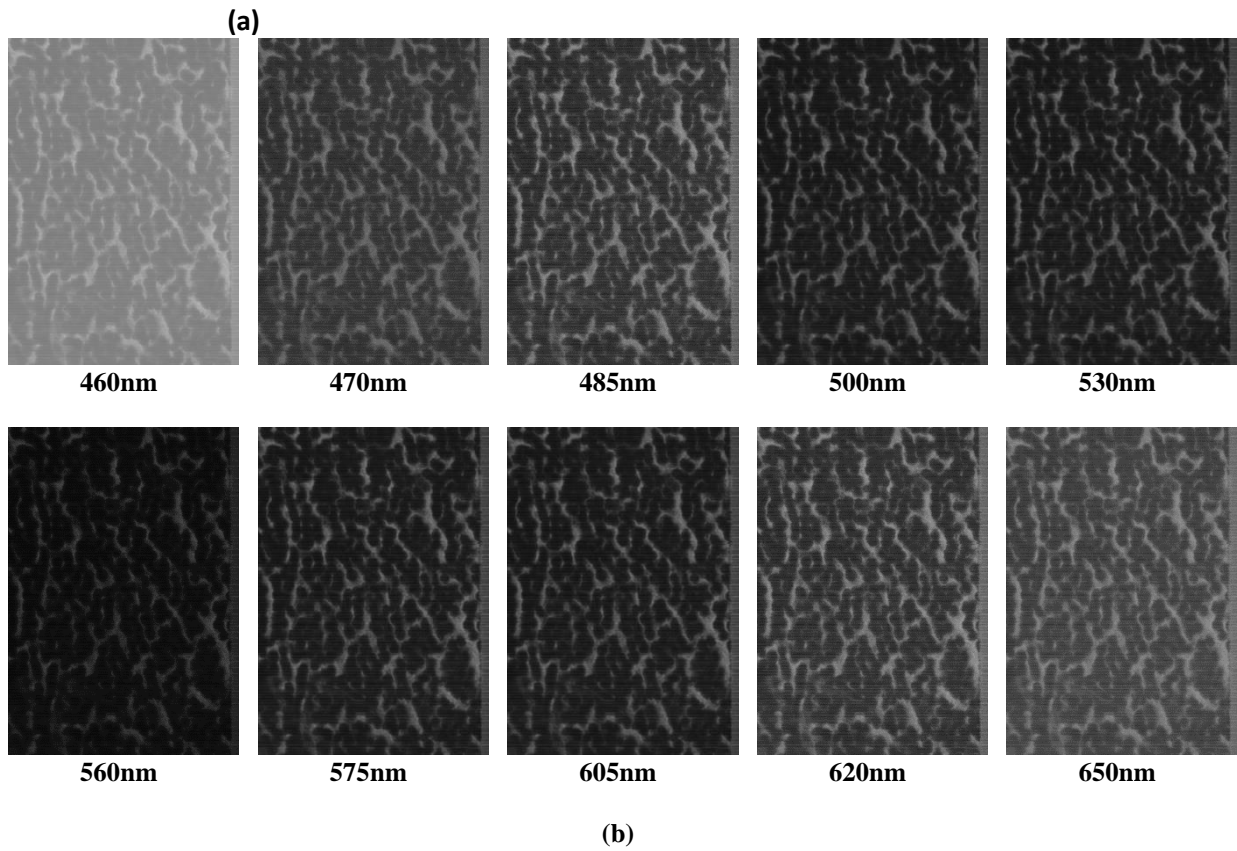
**Figure 3.18:** (a) Visible image of H&E-stained IDC tissue from mouse 1. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

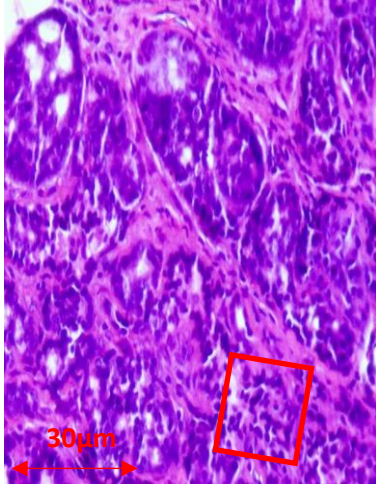


(b)

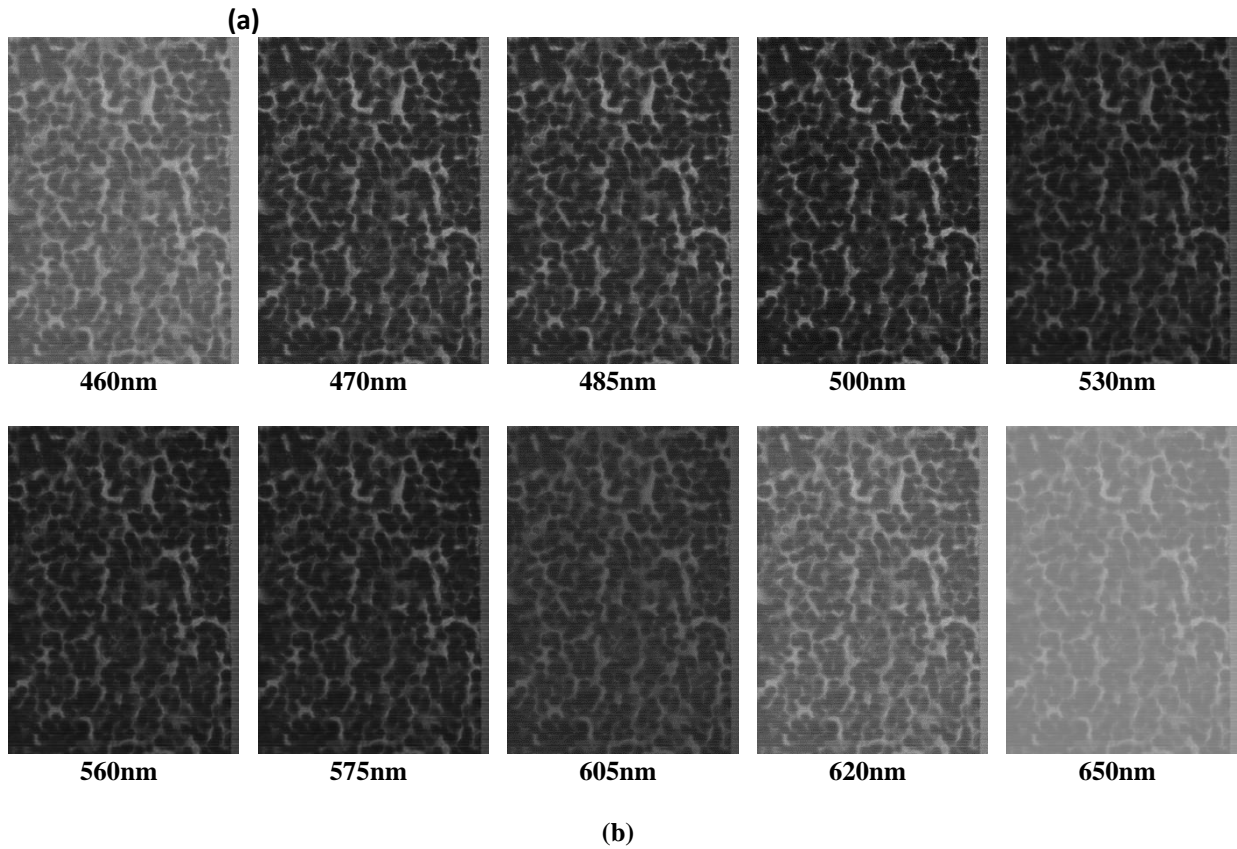


**Figure 3.19:** Visible image of H&E-stained DCIS tissue from mouse 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

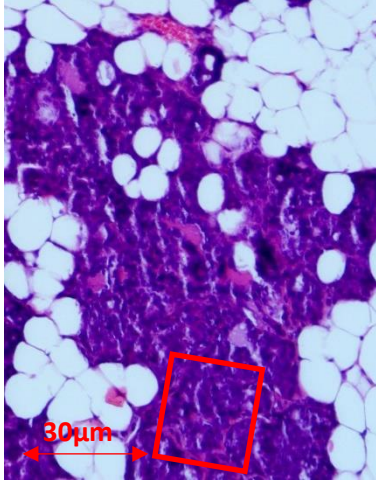




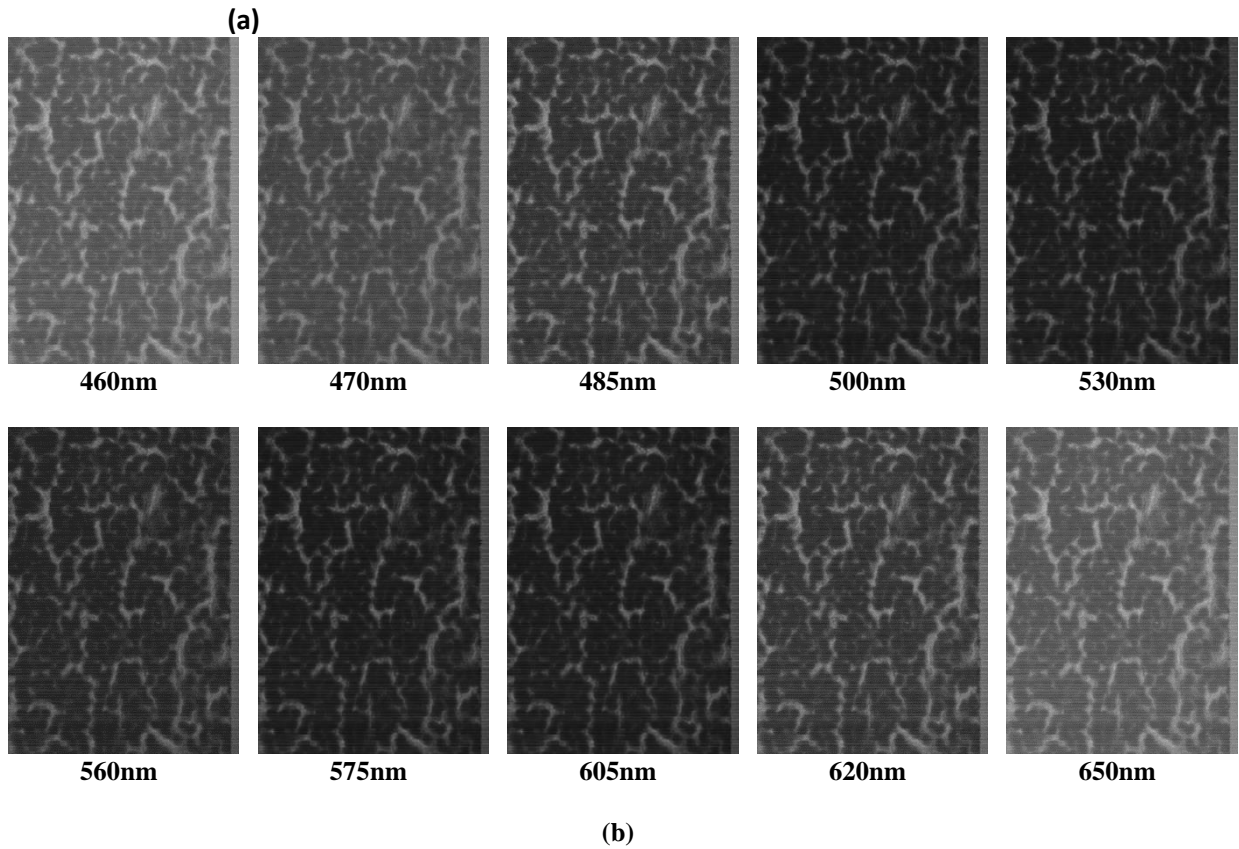
**Figure 3.20:** (a) Visible image of H&E-stained IDC tissue from mouse 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.

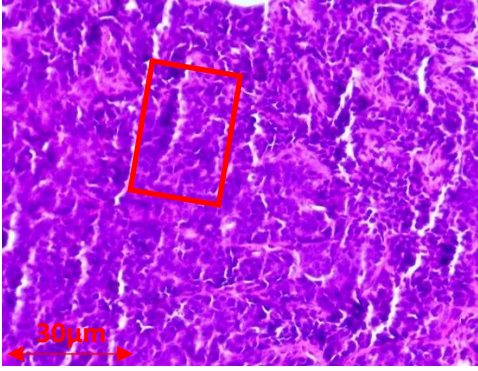




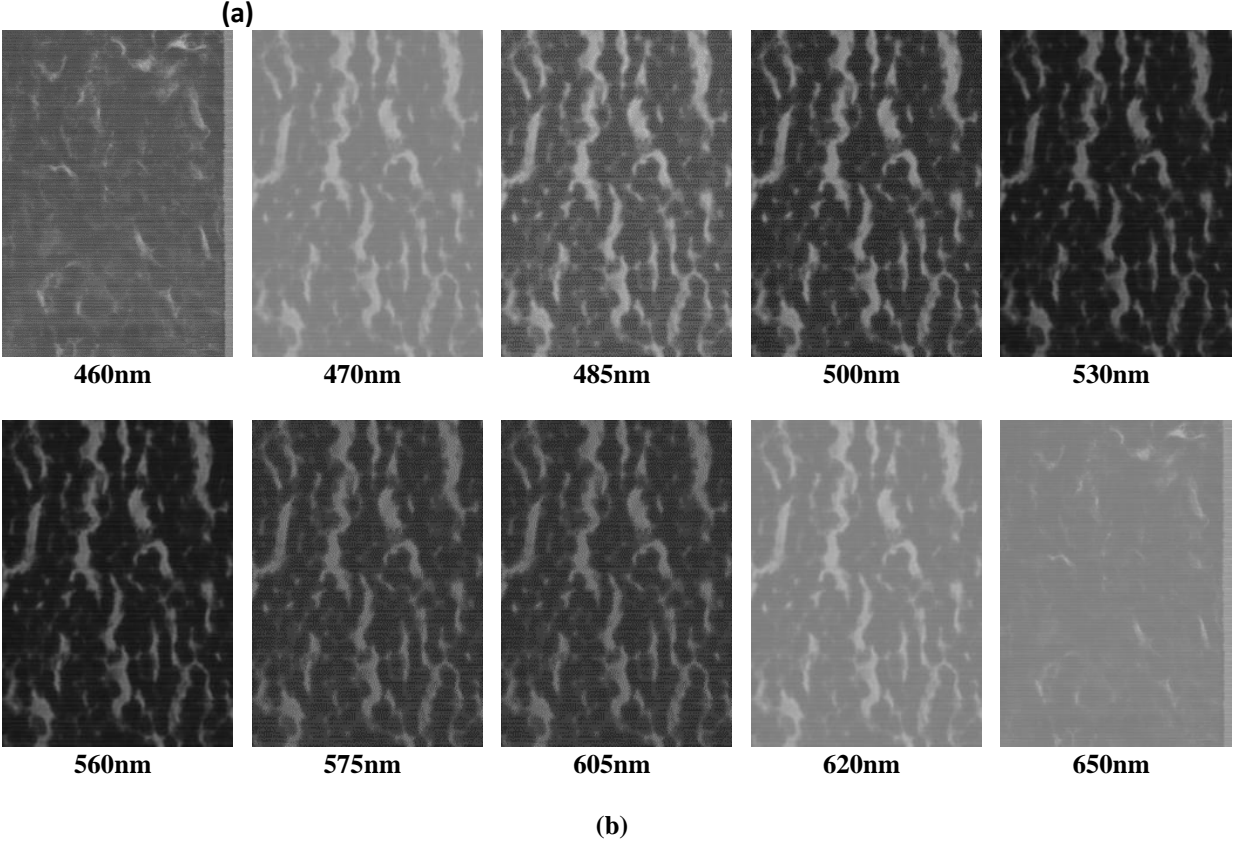


**Figure 3.21:** (a) Visible image of H&E-stained DCIS tissue from mouse 2. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.





**Figure 3.22:** (a) Visible image of H&E-stained IDC tissue from mouse 3. Region inside red box was selected for hyperspectral imaging. (b) Hyperspectral images of region highlighted in (a) for select wavelengths ranging from 460-650nm.



## 3.6 Summary

This chapter provides the first, and second tasks of this study that includes the HSI system setup, testing, and the cellular imaging. The first task includes a set of two experiments in order to have a clear image of the spectrum depends on our light source, as well as learning how the hypercube is built according to the hyperspectral imager that is used in this study. Two main experiments are done after the system calibration. These experiments are called: HSI leaves experiment, and three pathology slides preparation method. The first experiment shows the capability of the HSI system to identify, and differentiate between leaves, and their growing stages in the season depends on the spectral reflectance of the sun light. The second experiment focused on finding a spectral relationship between three different methods of preparing the biopsy slides by the pathologist,

The second task is the cellular imaging, which include includes three case studies provided in this work. Each section included three picked examples from the whole group of the imaged samples. The three imaged groups are as follow: H&E stained of normal and DCIS samples, unstained samples of normal and DCIS tissues, and finally in situ and invasive of mice mammary tumor. The aim of this chapter is to provide a set of examples to show how normal breast, DCIS, and IDC tissues would look like as a hyperspectral image, and as an image seen by pathologist under the microscope as well as the H&E stained and unstained samples. This chapter has also provided the hyperspectral images that includes the Region of Interest (ROI) that was marked by the pathologist to show the exact location that needed to be imaged and analyzed using the hyperspectral imager. The overall goal of this chapter is to prepare a set of hyperspectral images of the (ROI) to be analyzed and processed using two different algorithms for semi automation detection that will be shown in chapter 4.

# Chapter 4

## 4. Image Evaluation

### 4.1 Introduction

This chapter discusses image analysis and evaluation. The large amount of data contained in the three-dimensional hyperspectral data cube was evaluated in multiple steps. All images were initially manually marked by a pathologist for the purpose of providing ground truth description of the tissue samples showing evidence of cancer. This is an important step for training the auto detection algorithm, as well as evaluating the accuracy on tissue samples not included in the training set. All marked areas on the samples were identified by the pathologist to possess a high density of tumor tissue within the designated area. Image processing techniques were applied to look at the regions of interest identified by the pathologist and ultimately determine the spectral reflectance of tissues in these regions over the visible light wavelength spectrum. The image processing and HSI toolboxes contained in both MATLAB and Waikato Environment for Knowledge Analysis (WEKA) were used in image processing. WEKA is a collection of machine learning algorithms for data mining tasks that were developed at university of Waikato New Zealand). In this work, the algorithms are used to learn the about the nature of the hyperspectral data in order to understand the spectral reflectance value and how it can be used to differentiate between different regions in the tissue samples. The first section of this chapter is focused on analyzing the manually picked regions of the hyperspectral images, then plotting the spectral reflectance spectrum to compare between the tissues. Finally, spectral reflectance color filtering is applied to the images to evaluate the spectral response of the various regions and types of tissue visible in the images. In the latter sections of this chapter, a semi-auto algorithm was applied to both the supervised, and unsupervised (i.e. labeled, and unlabeled) hyperspectral data sets.



## 4.2 Spectral reflectance determination of manually marked cancer tissue

Because ductal carcinoma normally starts spreading from the duct, 4 square regions were randomly picked around each normal duct and in high density areas of the ductal carcinoma samples known to contain tumorous tissue. The results of hyperspectral images from each patient will consist of:

- 1) A comparison of spectral reflectance measurements from each area, and the average with error and standard deviation computed for 10 measurements from each of 4 separate regions per ample type.
- 2) Images with spectral reflectance color filtering applied on the spectral band that shows the highest degree of wavelength separation between normal tissue and ductal carcinoma. The spectral reflectance color filtering is presented as a color map that is representative of the spectral reflectance value.

### 4.2.1. Hyperspectral imaging of H&E-stained normal and DCIS tissue samples from patients 1-3

Spectral reflectance plots of the 4 marked areas of the H&E tissue samples of patient 1 from are shown in Figure 4.1.

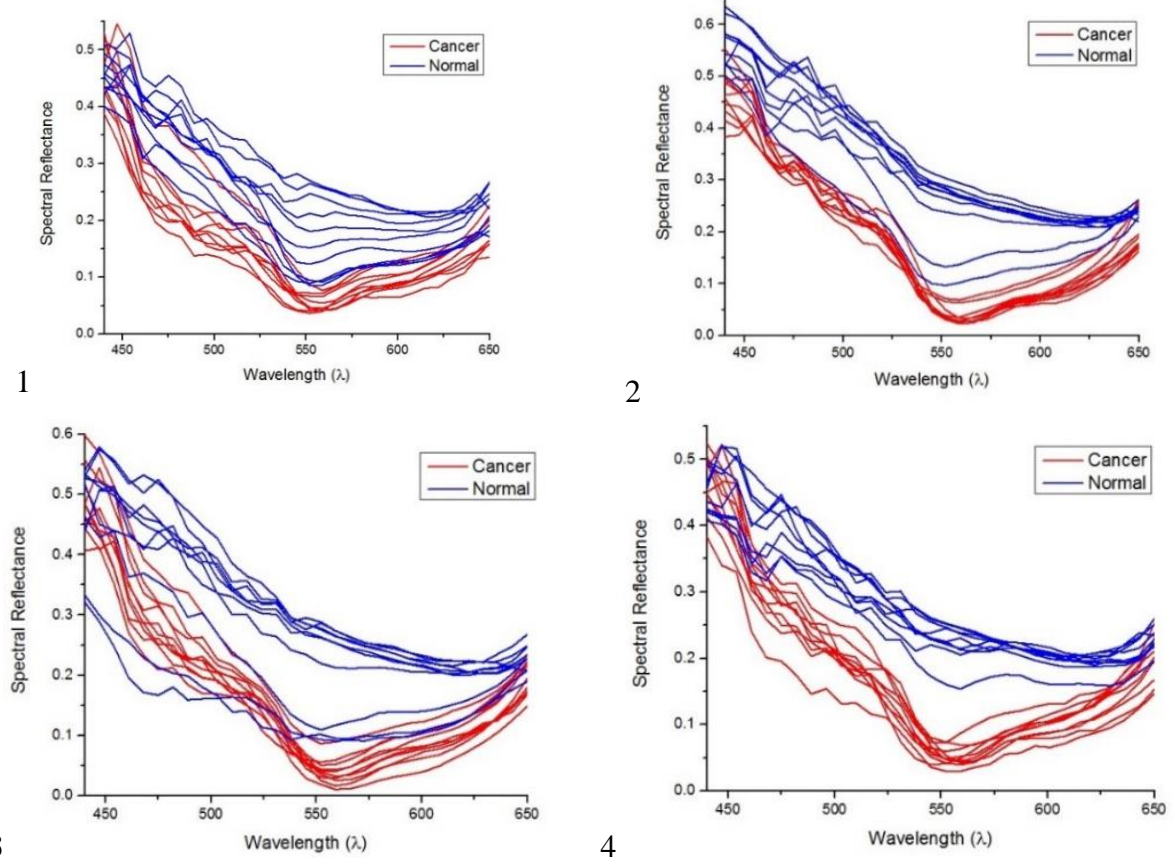
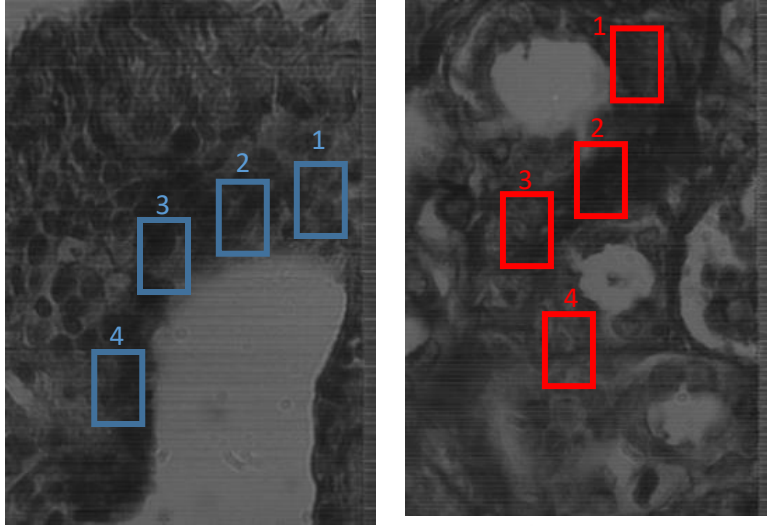
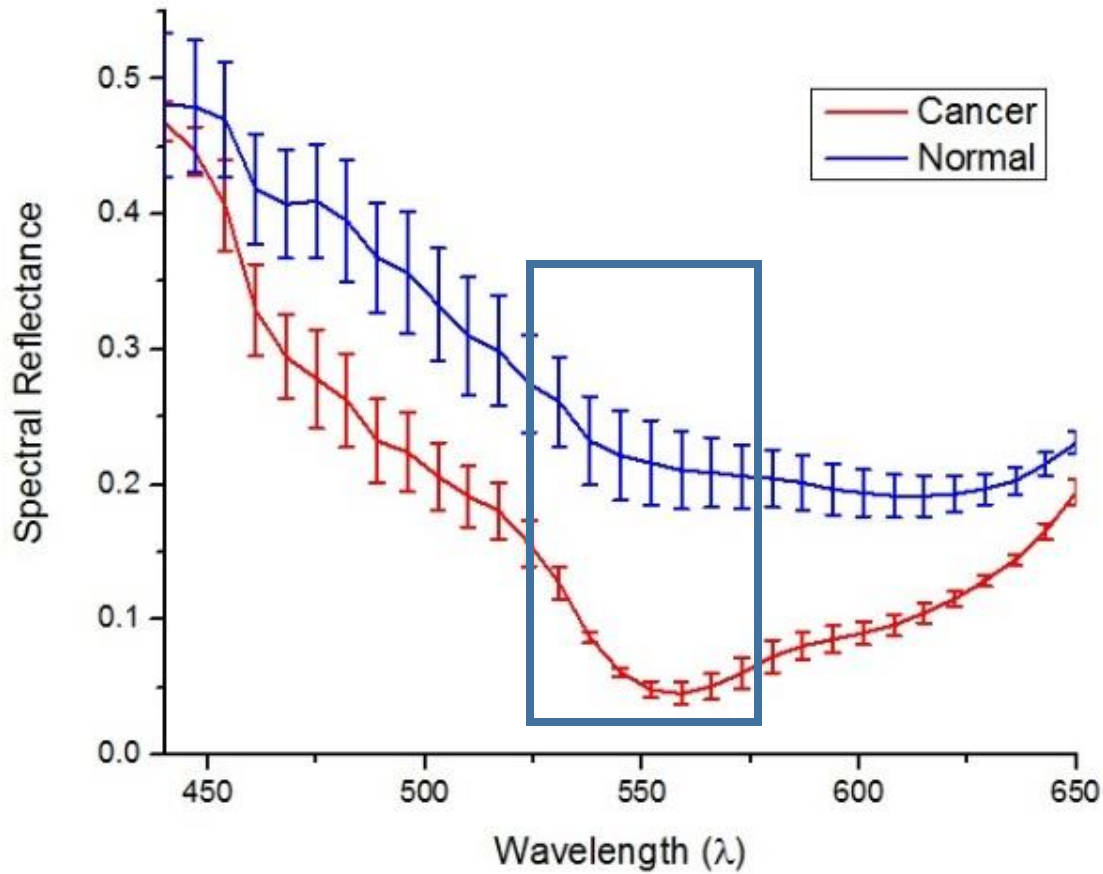


Figure 4.1: Spectral reflectance of each of the 4 marked areas of the 550nm spectral channel from top left to bottom right.

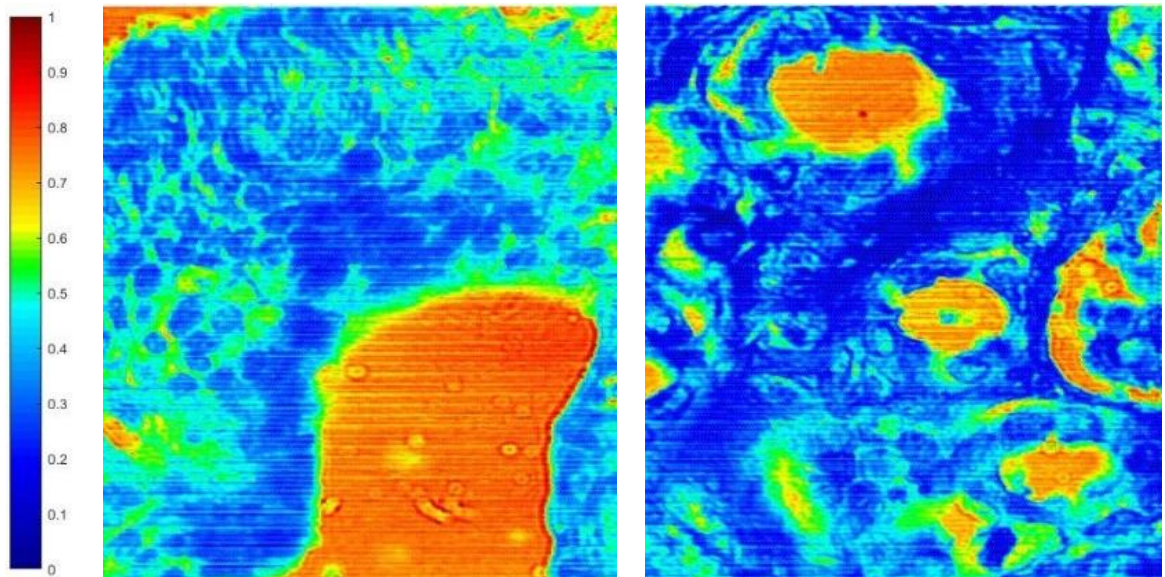


**Figure 4.2:** Average response showing spectral reflectance of the cancer and normal tissues.

The error bars in Figure 4.2 are used to quantify uncertainty in the graph based on an average of 10 measurements taken from each of the four marked region of interest in Figure 4.1 for both normal and cancer tissue. This was done in order to show a clear separation between both tissues at 550nm. This method is also applied on all data shown in the following sections 4.2.2, and 4.2.3.

The Spectral Reflectance Color Filtering (SRCF) method is applied to the hyperspectral images to visualize the difference in the spectral reflectance between tissue types present in one biopsy sample, as well as between tissues in different sample types (normal and DCIS). The scale in Figure 4.3 shows the normalized spectral reflectance intensity values for the normal and DCIS samples from Patient 1. The high-intensity regions, denoted by the red/orange color, represents the location of the milk ducts as marked by the pathologist. The lower intensity regions denoted by the dark blue color, which are the regions of interest for this research effort, represent normal and DCIS tissues surrounding the ducts. It can be seen that both normal and DCIS tissue have similar low SCRF values, which corresponds to the spectral plot shown in Figure 4.2. This small

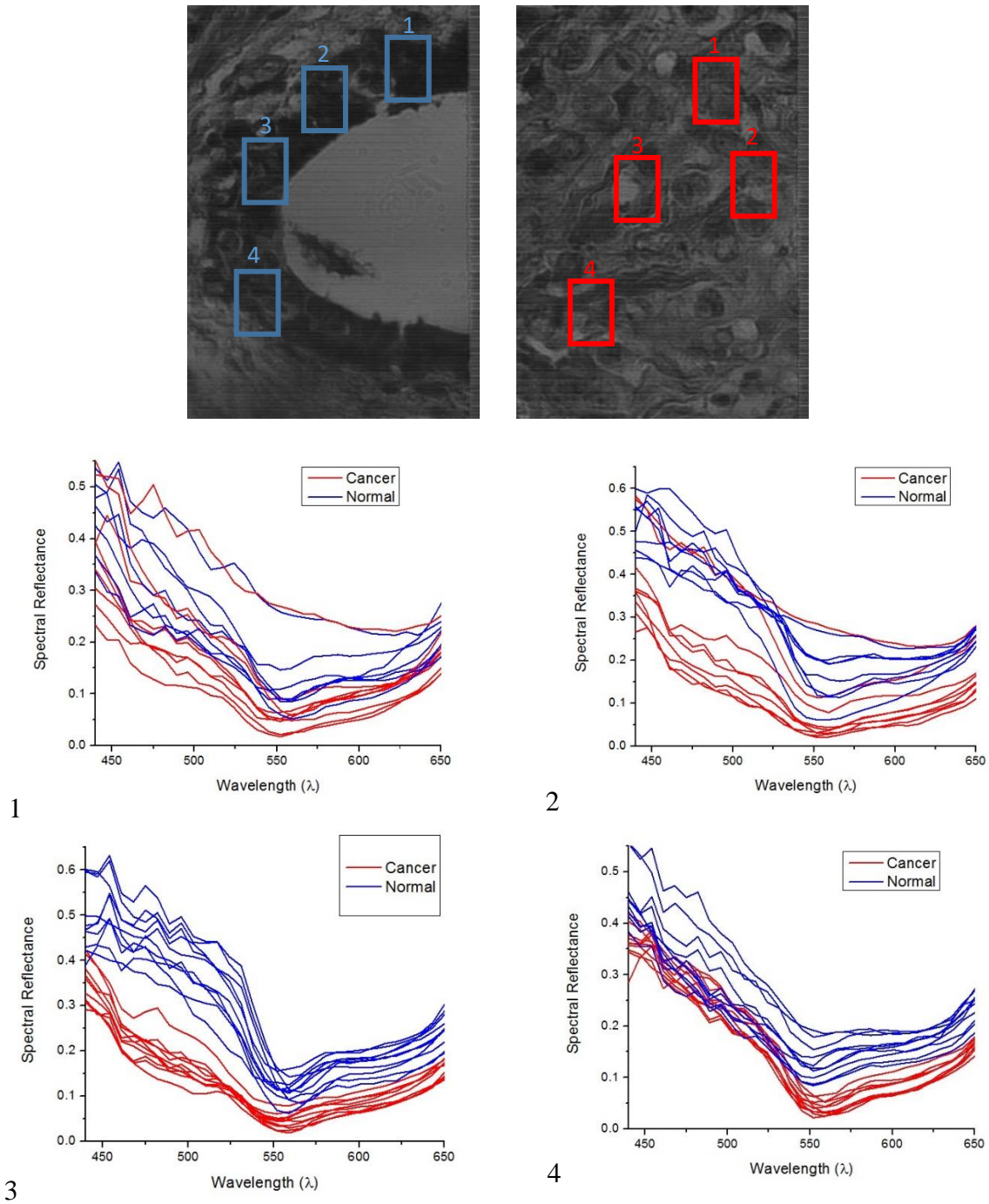
difference makes it challenging to perform accurate detection by just comparing the difference between 'light' and 'dark' blue in the SCRF images. Additional image analysis must be applied to achieve accurate cancer detection. To further illustrate this, results from the H&E-stained samples from two additional patients are shown in Figures 4.4-4.7.



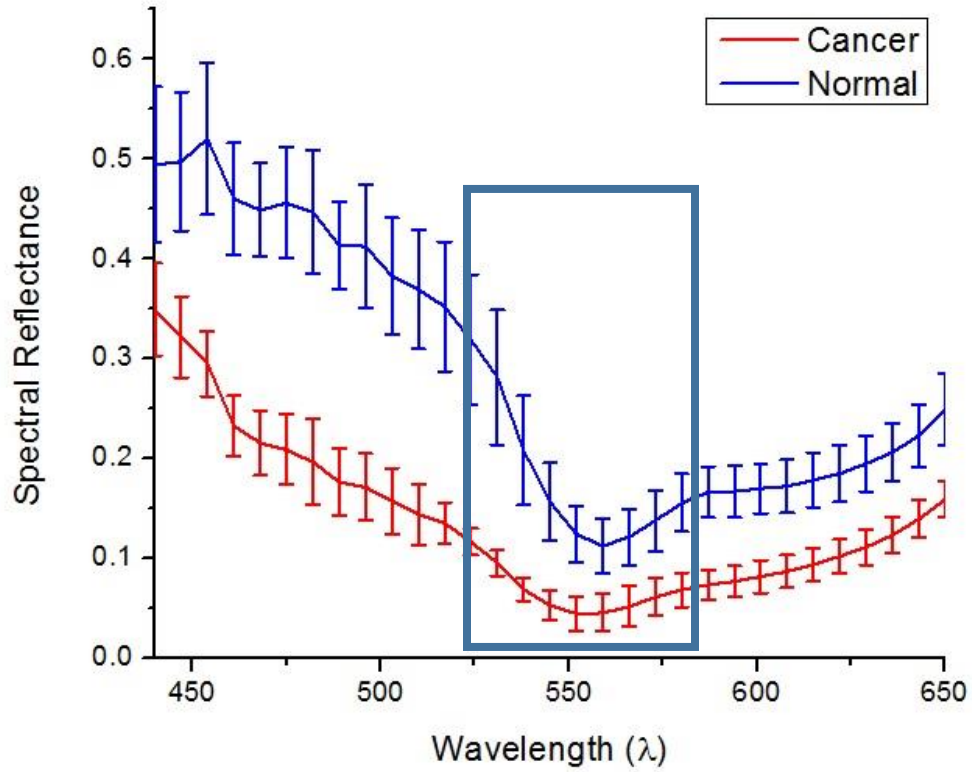
**Figure 4.3:** SRCF of both normal duct (Left), and ductal carcinoma (Right), and the spectral reflectance normalized scale.



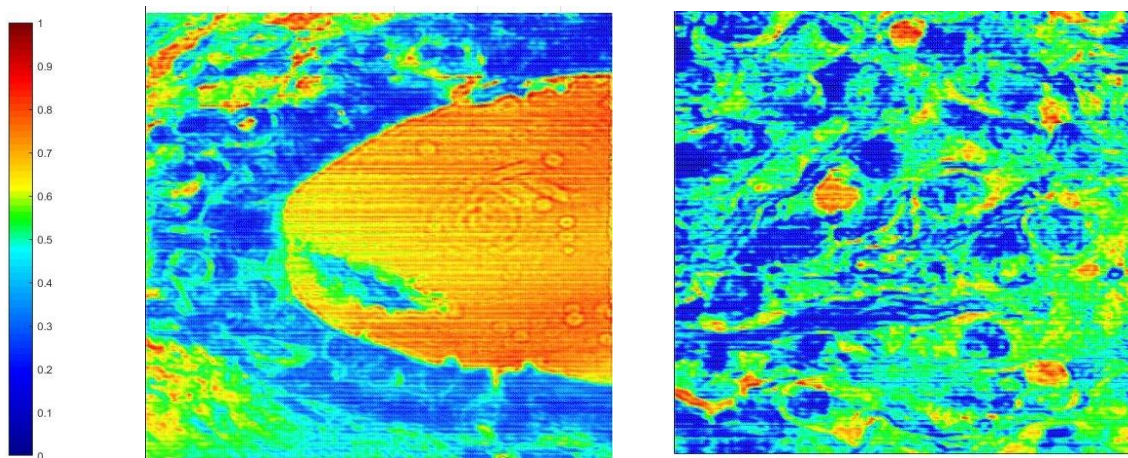
The spectral reflectance of the 4 marked areas of the H&E-stained normal and DCIS tissue samples from Patient 2 are shown in Figure 4.4 & 4.5, with the SCRF results shown in Figure 4.6



**Figure 4.4:** Spectral reflectance of each of the 4 marked areas of the spectral channel 550nm from top left to bottom right.

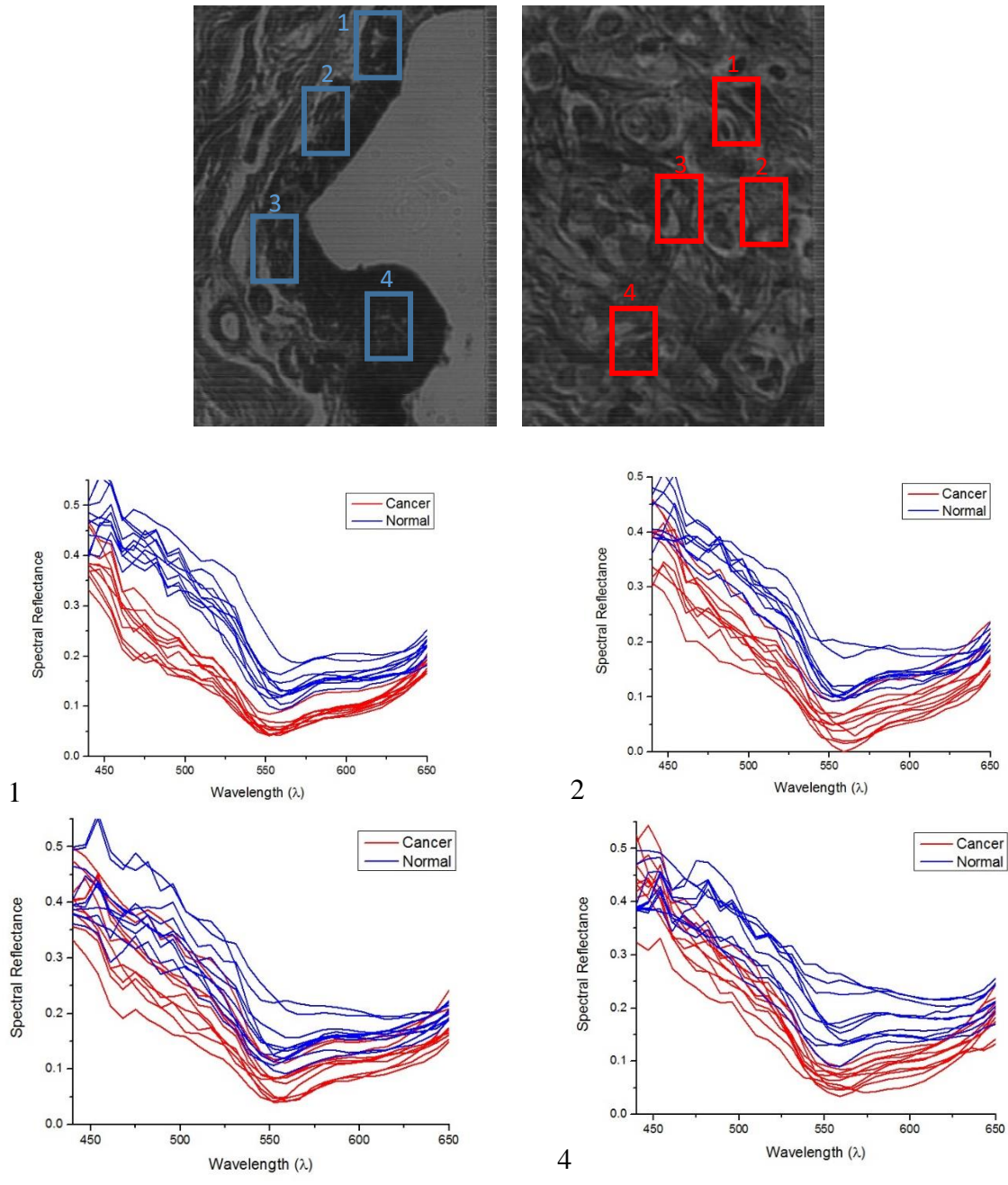


**Figure 4.5:** Average response clearly showing spectral reflectance of the cancer and normal tissues.



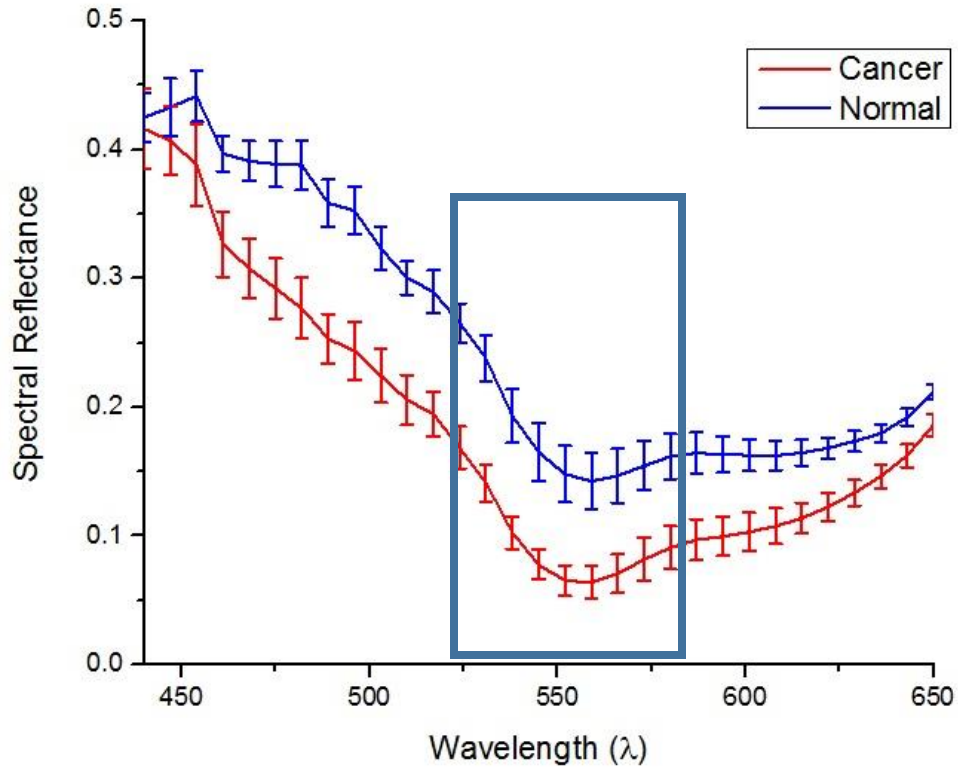
**Figure 4.6:** SRCF of both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.

The spectral reflectance plots for the 4 marked areas of the H&E-stained normal and DCIS tissue samples from Patient 3 are shown in Figures 4.7 & 4.8, with the SCRF results shown in Figure 4.9.

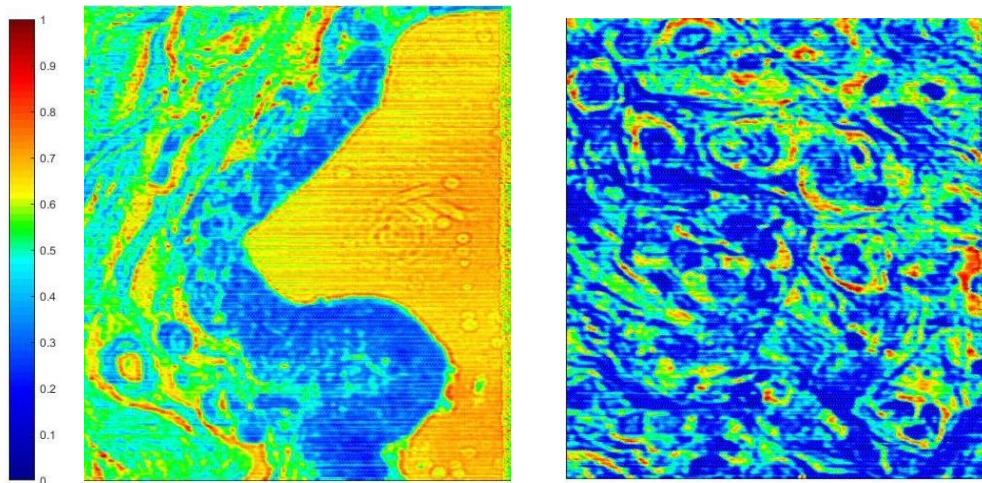


**Figure 4.7:** Spectral reflectance of each of the 4 marked areas of the spectral channel 550nm from top left to bottom right.





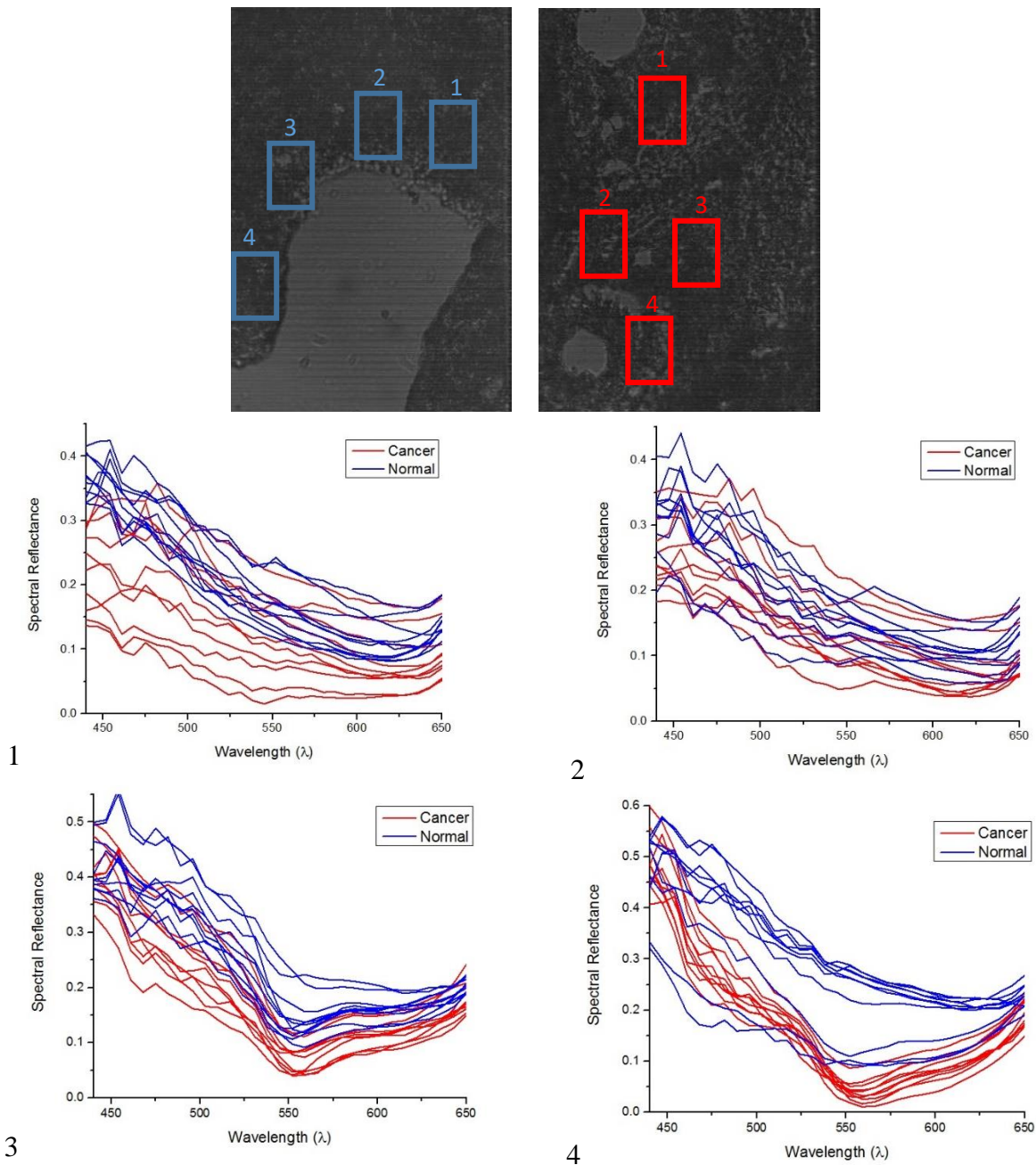
**Figure 4.8:** Average response clearly showing spectral reflectance of the cancer and normal tissues.



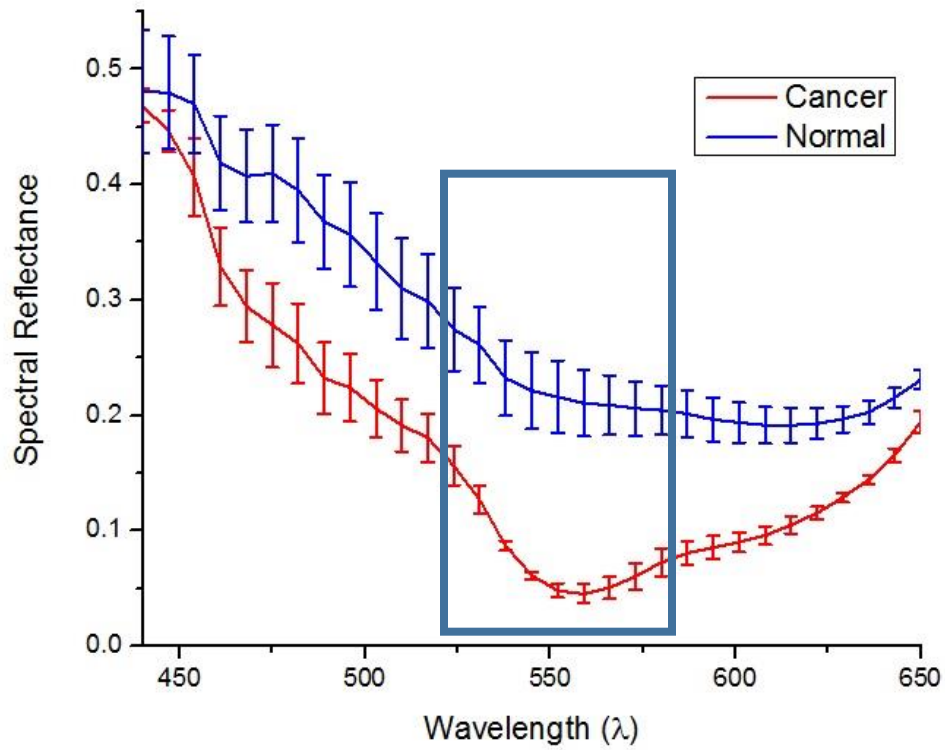
**Figure 4.9:** SRCF of both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.

## 4.2.2: Hyperspectral imaging of unstained normal and DCIS tissue samples of from Patient 1

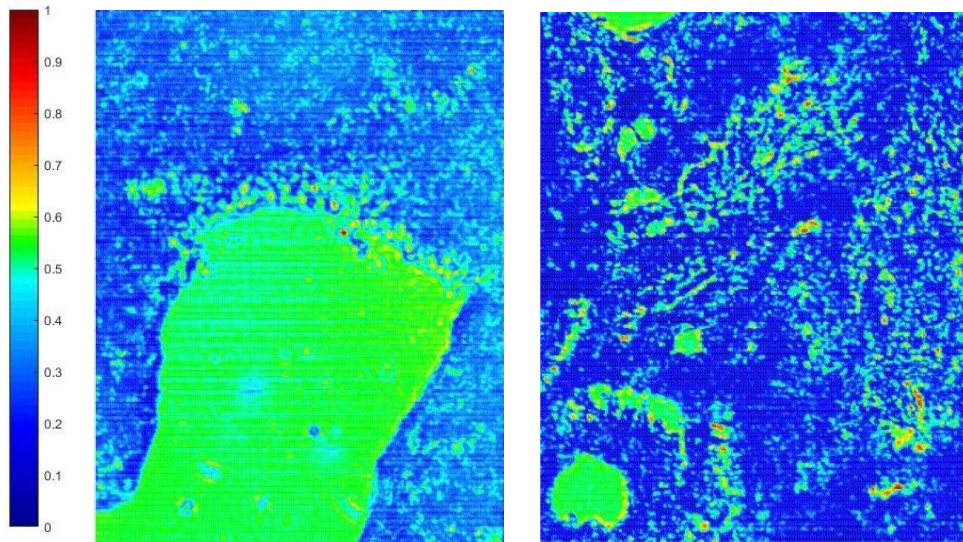
The spectral reflectance plots of 4 marked areas of unstained normal and DCIS tissue samples from Patient 1 are shown in Figure 4.10 & 4.11, with the corresponding SCRF images shown in Figure 4.12.



**Figure 4.10:** Spectral reflectance of each of the 4 marked areas of the spectral channel 550nm from top left to bottom right.



**Figure 4.11:** Average response clearly showing spectral reflectance of the cancer and normal tissues.

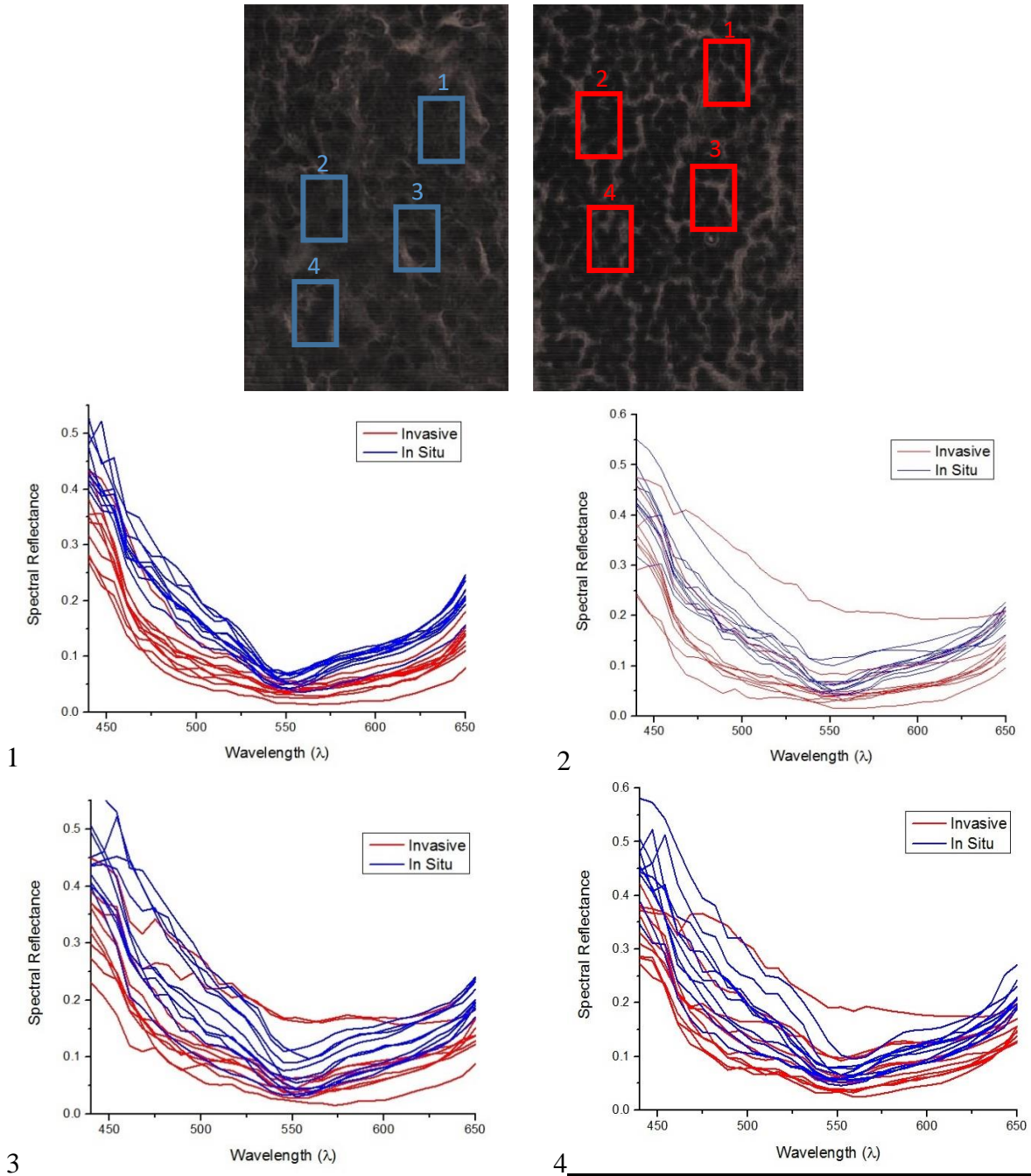


**Figure 4.12:** SRCF of both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.

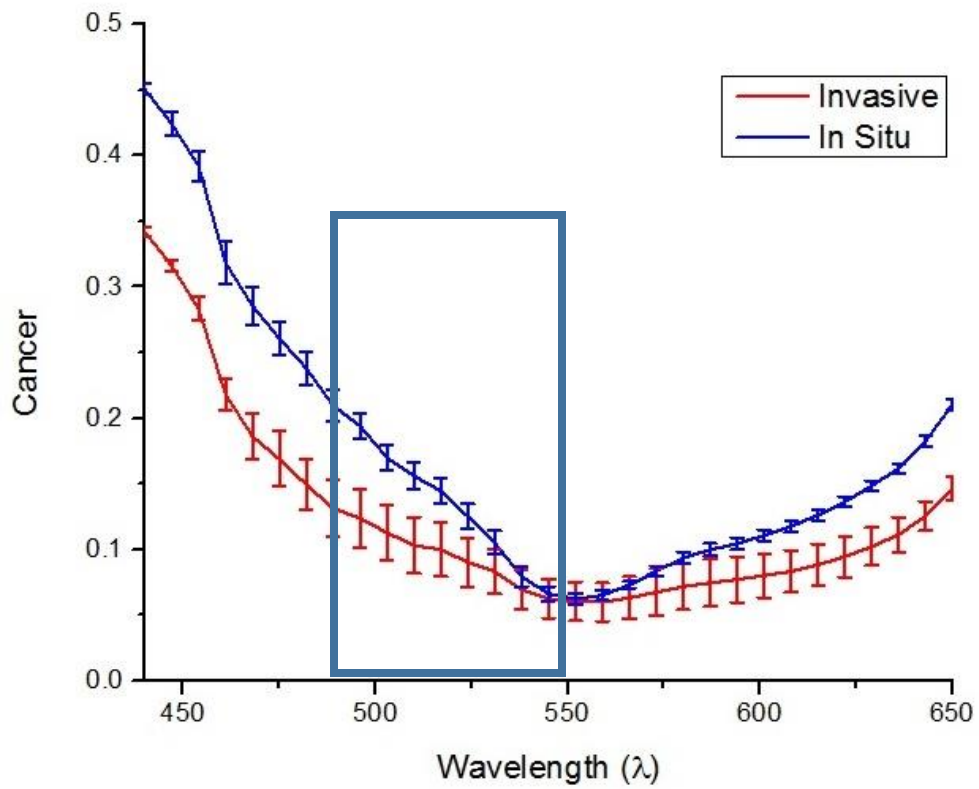


### 4.2.3: Hyperspectral imaging of H&E-stained DCIS and IDC tissues from Mice 1-3

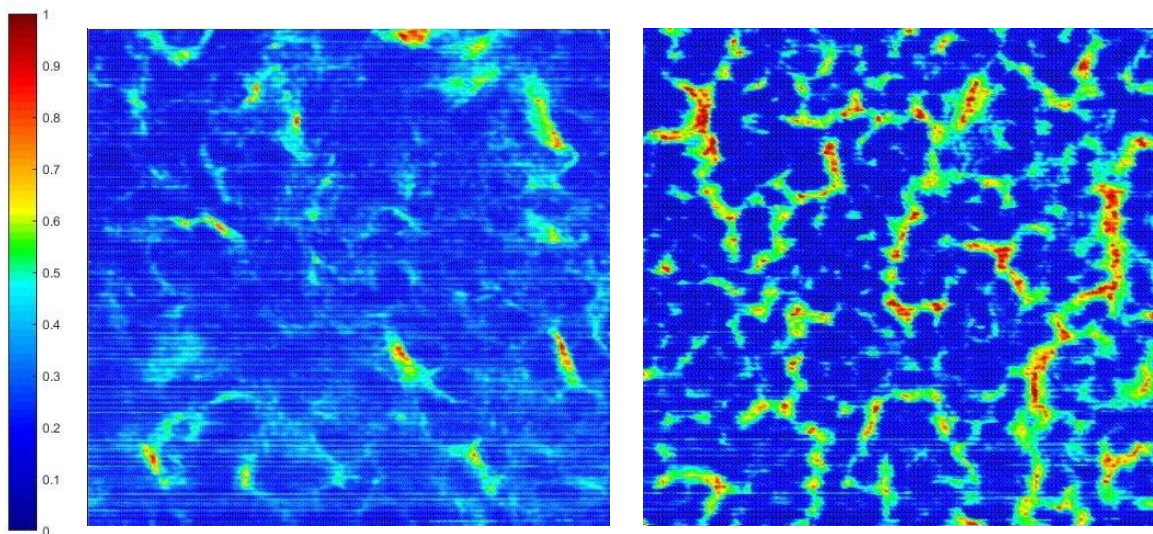
The spectral reflectance plots of 4 marked areas H&E-stained DCIS and IDC tissue samples from Mouse 1 are shown in Figures 4.13 & 4.14, with the corresponding SCRF images shown in Figure 4.15.



**Figure 4.13:** Spectral reflectance of each of the 4 marked areas of the spectral channel 500nm from top left to bottom right.

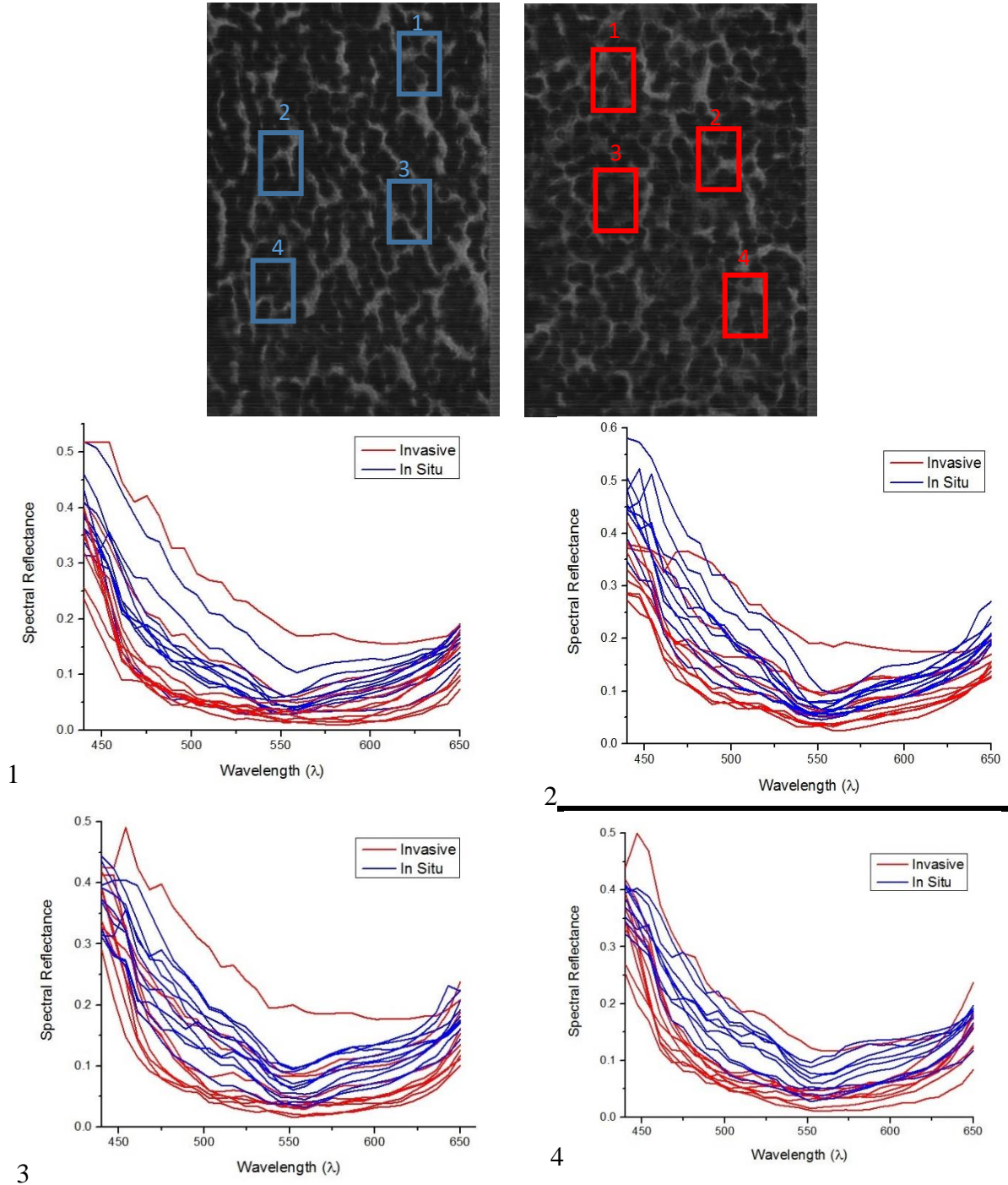


**Figure 4.14:** Average response clearly showing spectral reflectance of the cancer and normal tissues.



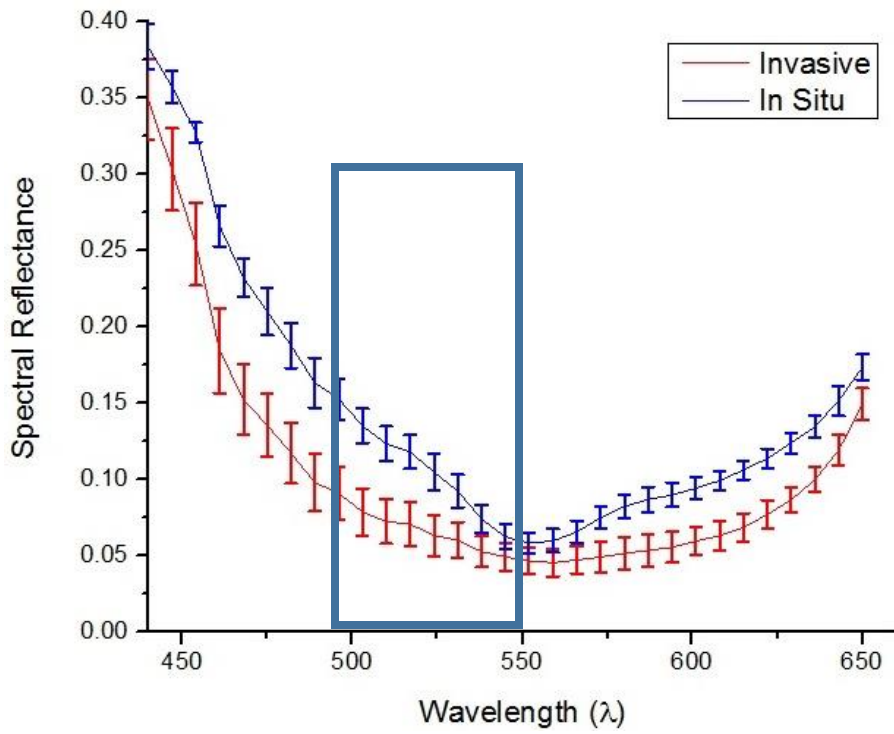
**Figure 4.15:** SRCF of both DCIS (Left), and IDC (Right), and the RICF scale.

The spectral reflectance plots of 4 marked areas H&E-stained DCIS and IDC tissue samples from Mouse 2 are shown in Figures 4.16 & 4.17, with the corresponding SCRF images shown in Figure 4.18.

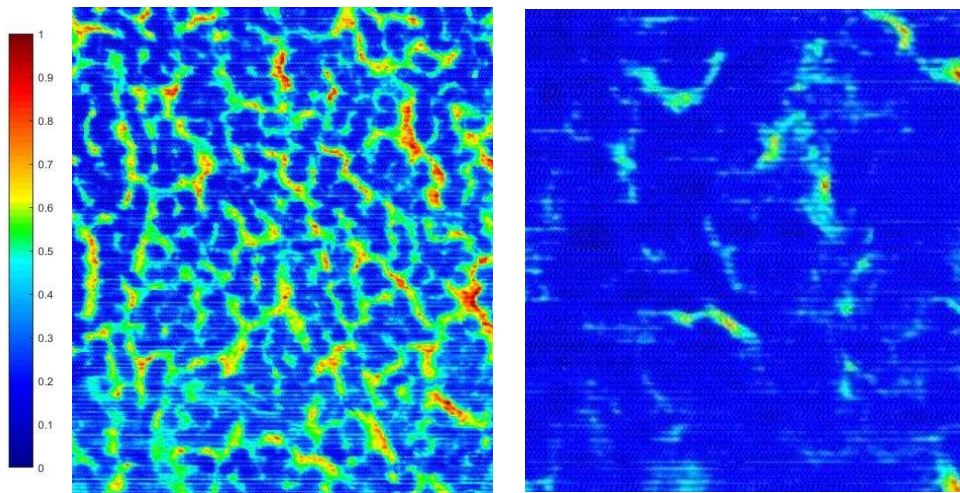


**Figure 4.16:** Spectral reflectance of each of the 4 marked areas of the spectral channel 500nm from top left to bottom right.



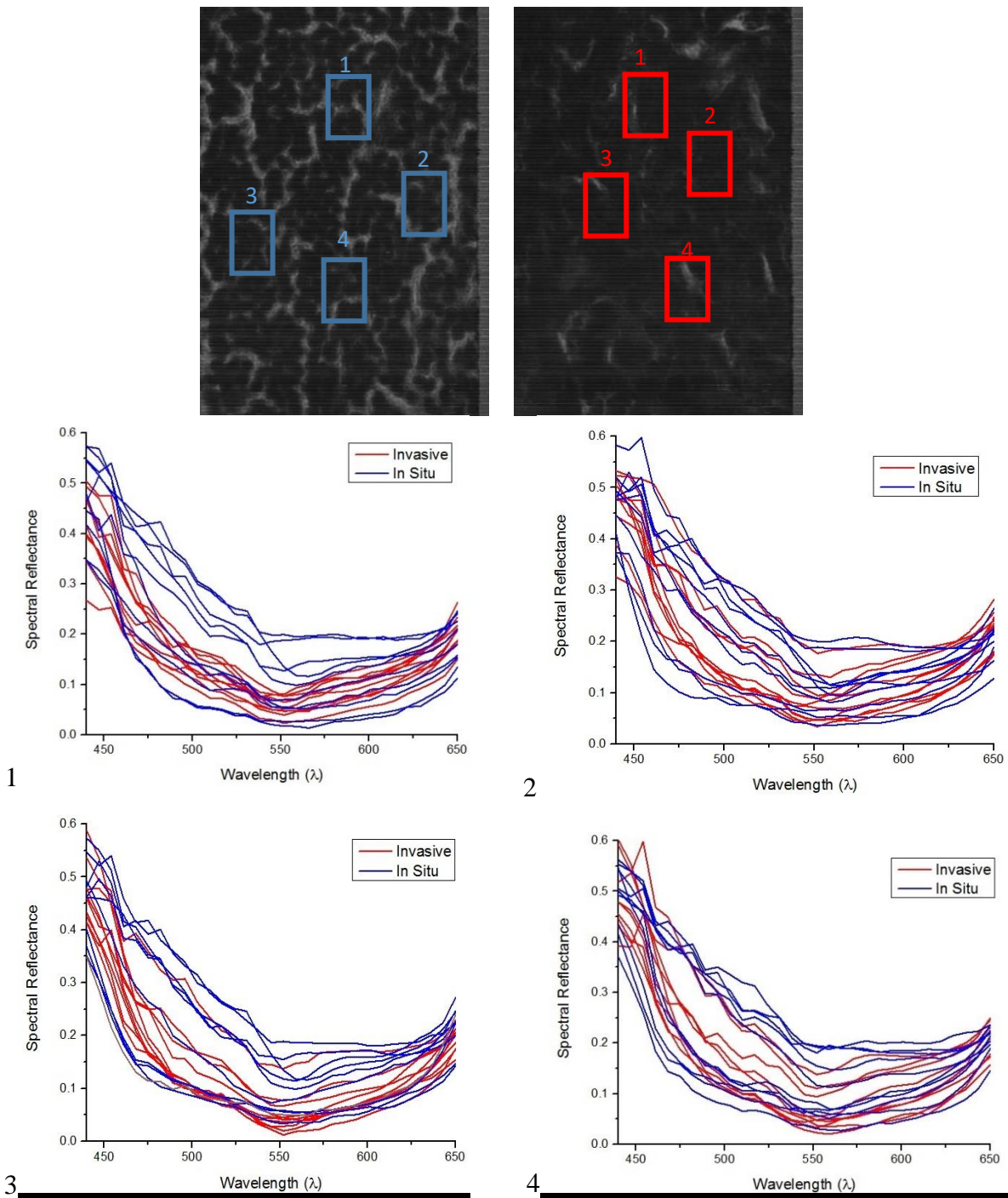


**Figure 4.17:** Average response clearly showing spectral reflectance of the cancer and normal tissues.

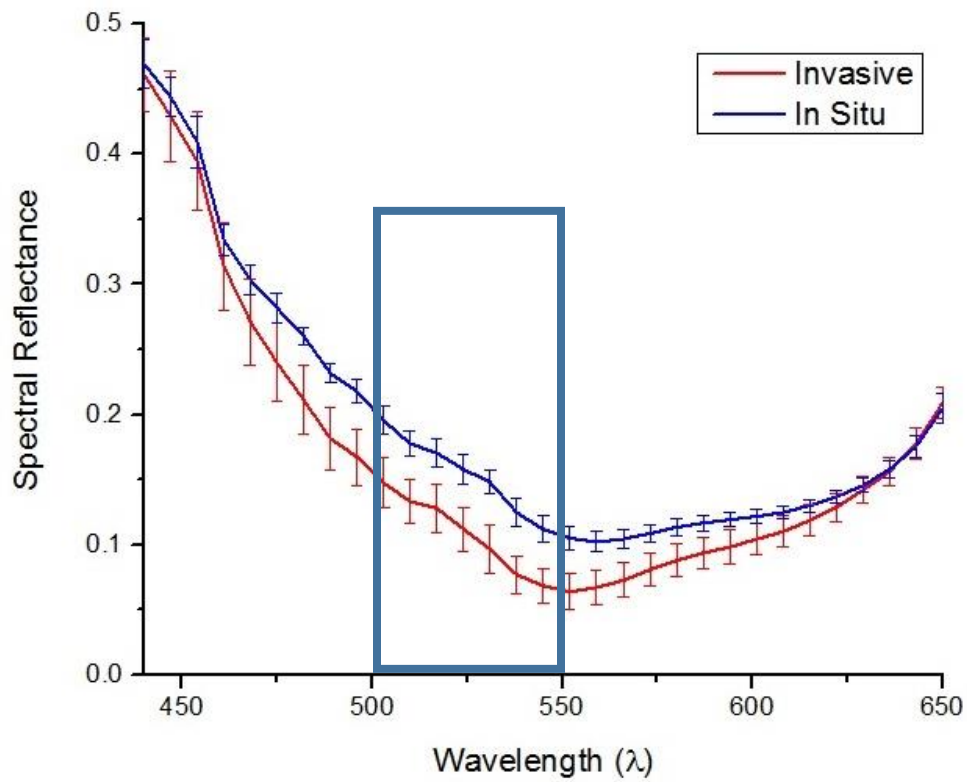


**Figure 4.18:** SRCF of both DCIS (Left), and IDC (Right), and the RICF scale.

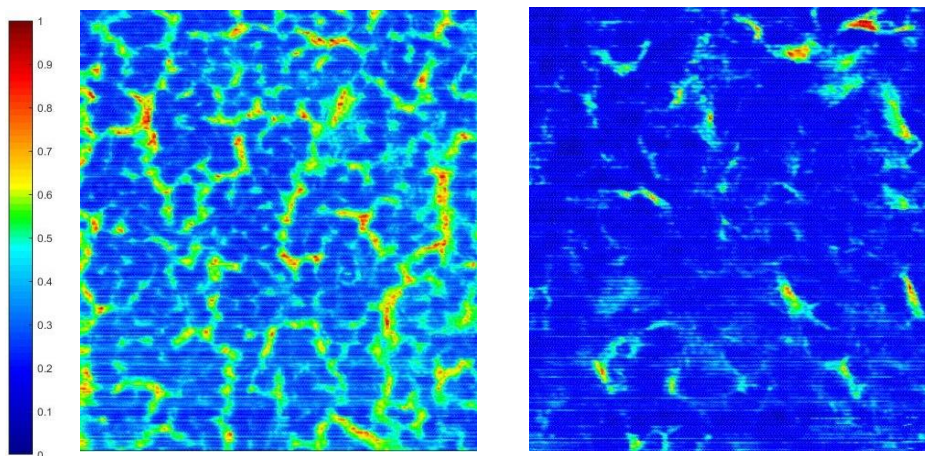
The spectral reflectance plots of 4 marked areas H&E-stained DCIS and IDC tissue samples from Mouse 3 are shown in Figures 4.19 & 4.20, with the corresponding SCRF images shown in Figure 4.21.



**Figure 4.19:** Spectral reflectance of each of the 4 marked areas of the spectral channel 500nm from top left to bottom right.



**Figure 4.20:** Average response clearly showing spectral reflectance of the cancer and normal tissues.



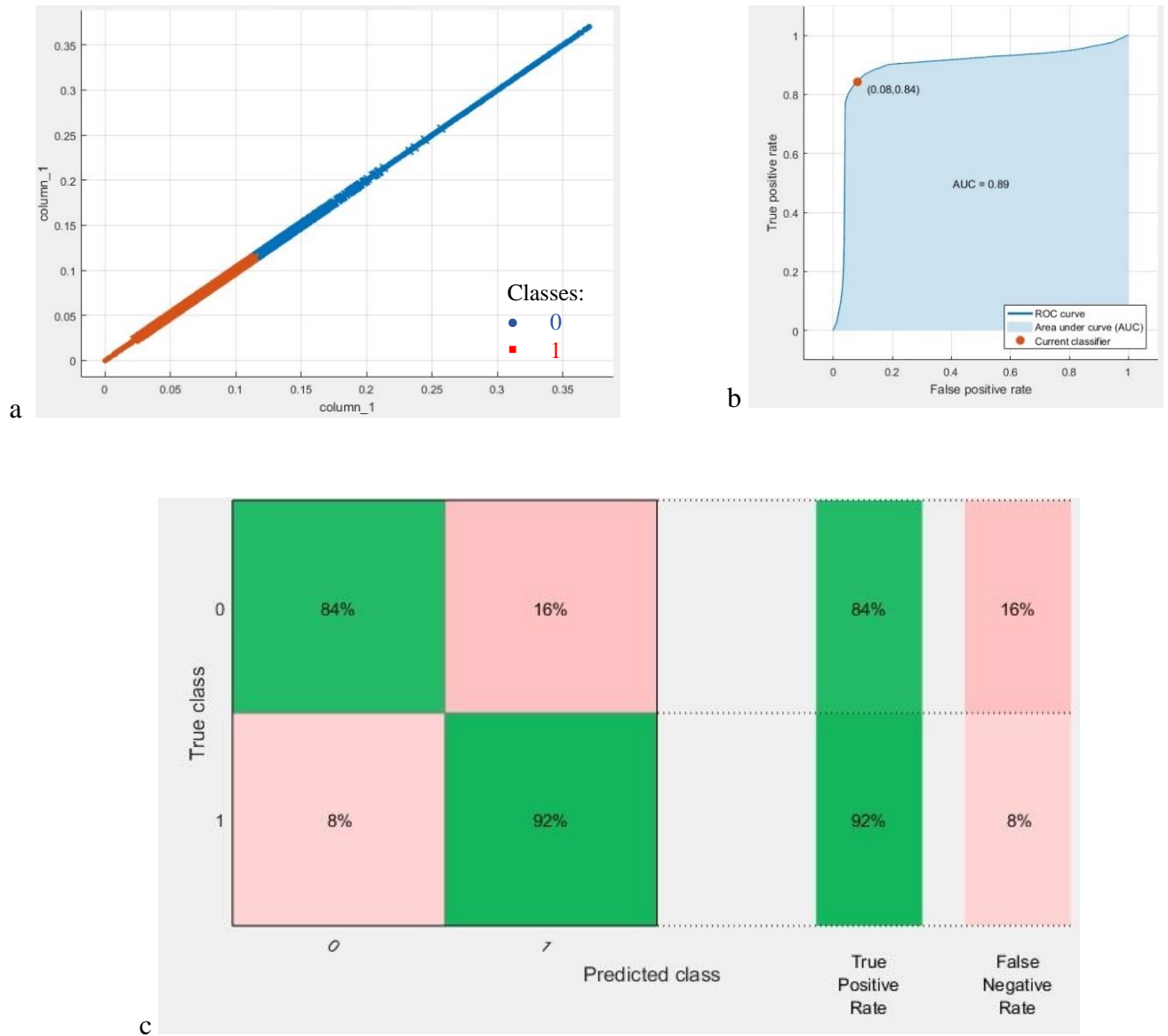
**Figure 4.21:** Spectral reflectance of each of the 4 marked areas of the spectral channel 500nm from top left to bottom right.

## 4.3 Selection of image data for training cancer detection algorithm

### 4.3.1 Supervised data for Support Vector Machine (SVM) classifiers using a Gaussian algorithm on human-female H&E stained breast tissue

In this section, the method used to process and analyze images using the HSI toolboxes contained in both software MATLAB and WEKA is described. The SVM methods were applied on supervised (labeled) data, and a K-means algorithm were applied on the unsupervised (unlabeled) data. As described in Chapter 2, a Gaussian SVM is applied on the data set after preparing the data for classification. The data is transferred into 1D and then labeled with two classes, 0 and 1, for the system to be able to train and test for testing the classifier accuracy. This classifier is applied for the purpose of detection cancer tissue, so the answer that is required is either positive cancer or negative cancer, allowing the data to be transferred into 1D for simplicity. Table 4.1, shows the classifier accuracy for each patient, and the diagrams follow that describe the prediction and classifier accuracy of the tested model for detection. More details are given in the discussion following the diagrams. **Table 4.1:** SVM algorithm on 10 patients

Patient #	Accuracy	Patient #	Accuracy
1	91.0%	6	98.7%
2	90.2%	7	91.7%
3	88.5%	8	85.0%
4	90.4%	9	87.8%
5	99.1%	10	79.3%
Average Accuracy:		90.17%	



**Figure 4.22:** a- Scatter plot of the SVM algorithm, b- ROC curve of the prediction, c- Confusion matrix of the predicted model.



After applying the trained algorithm on the samples from the patients, the following step is to apply the trained algorithm on the testing data for prediction, and detection accuracy. Figure 4.22-a shows the results given by a scatter plot of the values for both classes, followed by the ROC curve showing the area under the curve= 0.89 in Figure 4.22-b which is supported by the confusion matrix in Figure 4.22-c, which shows the confusion matrix of the tested model showing a true positive rates of 92% and 84%, false negative rates of 8% and 16%.

#### 4.3.2 Unsupervised data for K-means algorithm on human- female H&E stained breast tissue

The K-means algorithm described in Chapter 2 was applied to unsupervised (i.e. unlabeled) data. The user chooses the number of classes, which is 2 for this study (cancer, and non-cancer), and then the algorithm picks the best centroid for each class. The class centroids changes each time the algorithm runs until the program stops, i.e. when the numbers of each class do not change, and each centroid is considered the best pick. Table 4.2 shows the results of applying K-means separately on each set of data from 9 out of 10 patients for training, iterating to find the two best values of centroids, and finally, using the known labels of cancer and non-cancer to measure the accuracy of each run. The last step is to run the K-means trained algorithm on the 10th patient for testing the cancer detection scheme in order to determine if the classifier can identify and detect both classes, and then compare the detected results with the ground-truth pathologist diagnosis in order to support the algorithm accuracy. The detection results will be shown in the Section 4.5. Table 4.2 also shows the prediction accuracy of the algorithm by applying the labels on the data by using the centroids that the trained algorithm created.

**Table 4.2:** K-Means algorithm results from tissues obtained from Patients 1-9.

Patient #	Centroid1	Centroid2	Accuracy	Patient#	Centroid1	Centroid2	Accuracy
1	<b>0.251</b>	<b>0.072</b>	<b>85.38%</b>	6	<b>0.287</b>	<b>0.066</b>	<b>98.34%</b>
2	<b>0.182</b>	<b>0.062</b>	<b>82.87%</b>	7	<b>0.270</b>	<b>0.096</b>	<b>80.59%</b>
3	<b>0.209</b>	<b>0.088</b>	<b>79.41%</b>	8	<b>0.223</b>	<b>0.099</b>	<b>82.22%</b>
4	<b>0.170</b>	<b>0.070</b>	<b>87.38%</b>	9	<b>0.265</b>	<b>0.075</b>	<b>87.84%</b>
5	<b>0.296</b>	<b>0.046</b>	<b>98.5%</b>	1-9	<b>0.207</b>	<b>0.091</b>	<b>80.27%</b>
<b>Average Accuracy</b>				<b>85.47%</b>			

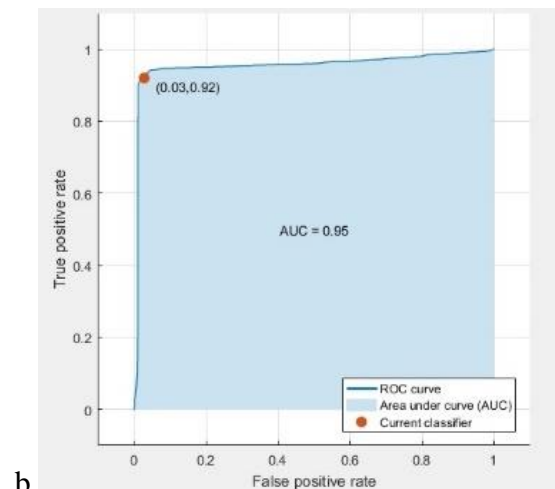
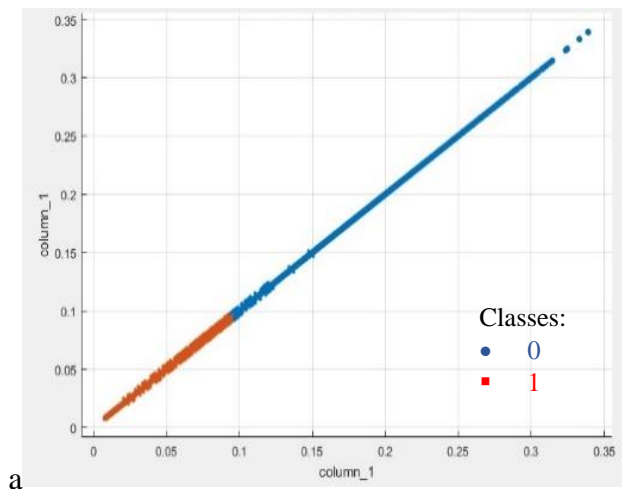


### 4.3.3 Supervised data for Support Vector Machine SVM classifiers using a Gaussian algorithm on unstained human- female breast tissue samples.

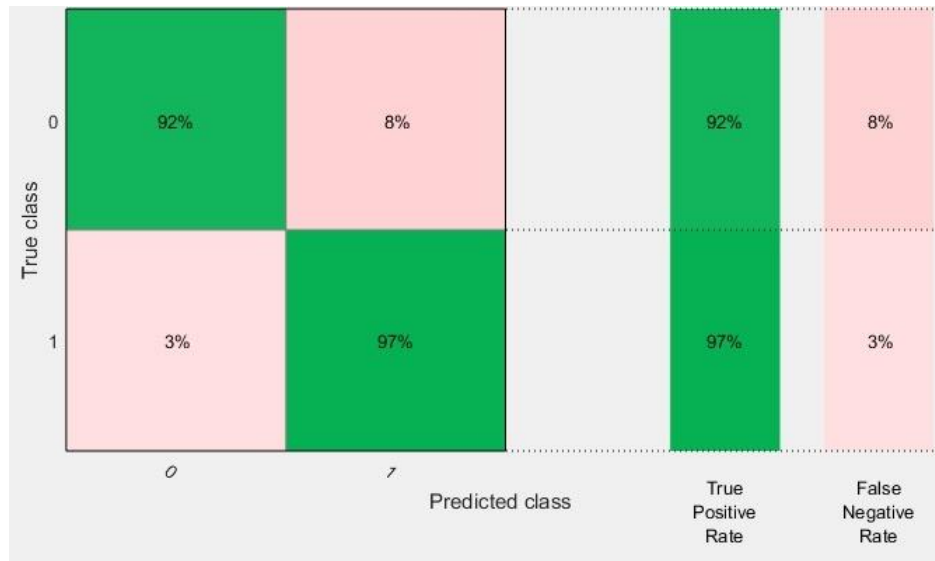
The same method applied in section 4.3.1 to the hyperspectral images from the H&E stained samples was then applied to the unstained samples. On the sample, 1600-pixel samples were tested using the SVM algorithm. Every pixel is labeled for system training, which is followed by testing. 10-fold cross validation was applied for accurate results. Table 4.3 presents the accuracy predicted detection of the trained algorithm, and the diagrams follow that describes the prediction and classifier accuracy of the tested modal for detection. More details are given in the discussion follows the diagrams

**Table 4.3: SVM algorithm on unstained sample**

Patient 1-10	Accuracy
AVG	<b>94.4%</b>



c



**Figure 4.23:** a- Scatter plot of the SVM algorithm, b- ROC curve of the prediction, c- Confusion matrix of the predicted model.

After analyzing the data by training the algorithm on the samples from patients, the following step is to apply the trained algorithm on the testing data for prediction and detection accuracy. Figure 4.23-a shows the results given by a scatter plot of both classes with the prediction, followed by the ROC curve showing the area under the curve= 0.95 in Figure 4.23-b. Finally 4.23-c shows the confusion matrix of the tested model showing a true positive rates of 97% and 92%, false negative rates of 8% and 3%.

#### 4.3.4 Unsupervised data for K-means algorithm on unstained human- female breast samples.

The same K-means algorithm method that was applied in section 4.3.2 on the stained samples was applied and ran on the unstained samples on the average data from the patients for training to create 2 centroids as seen in Table 4.4.

**Table 4.4:** K-means results from unstained sample images.

Patient#	Centroid1	Centroid2	Accuracy
<b>AVG</b>	<b>0.208</b>	<b>0.085</b>	<b>93.66%</b>

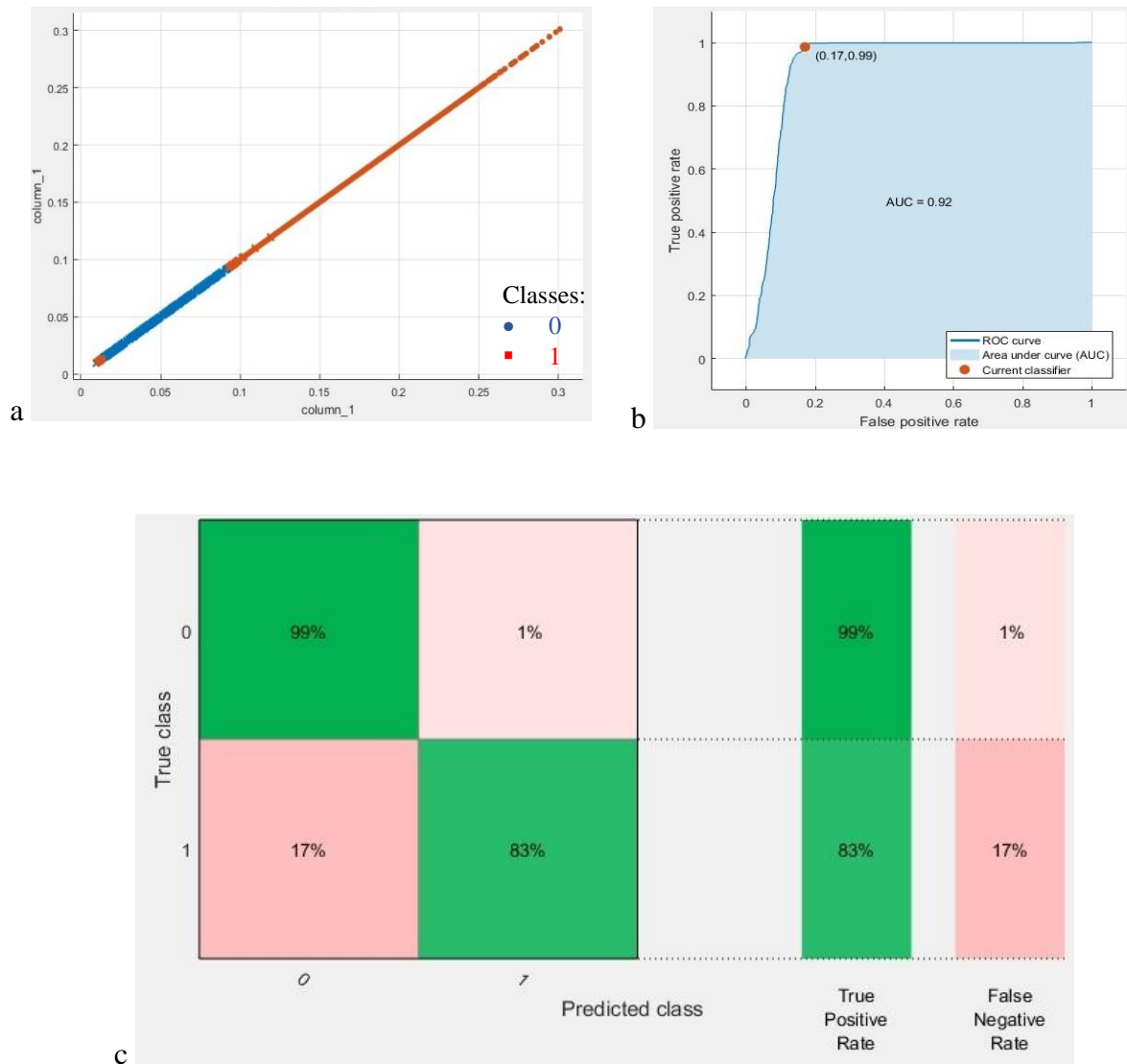
As Table 4.4 shows the result of applying the K-means algorithm on the data that was extracted from the unstained samples, the trained algorithm created two centroids one for each class as explained on the previous section. The trained algorithm then applied on the testing data for detection to get an accuracy of 93.66%.

#### 4.3.5 Supervised data for Support Vector Machine SVM classifiers using a Gaussian algorithm on mice models with positive DCIS, and IDC

The same SVM method that was applied on sections 4.3.1 and 4.3.3 to analyze the H&E-stained and unstained tissue samples of the human patients was applied to the unstained mouse tissue samples. The predicted classes that are labeled 0, and 1, representing the presence of DCIS IDC respectively. Again, 1600-pixel samples were tested using the SVM algorithm, with every pixel labeled for system training then testing. 10-fold cross validation was applied for accurate results: Table 4.5 presents the accuracy of the detection using the trained algorithm, and the diagrams that follow describe the prediction and classifier accuracy of the tested modal for detection. More details are given in the discussion following the diagrams

**Table 4.5:** SVM algorithm on H&E mice samples

Mouse#	Accuracy	Mouse #	Accuracy
1	90.9%	3	92.8%
2	94.6%	4	84.2%
<b>Average Accuracy:</b>		<b>90.17%</b>	



**Figure 4.24:** a- Scatter plot of the SVM algorithm, b- ROC curve of the prediction, c- Confusion matrix of the predicted model.

After analyzing the data by training the algorithm on the samples from mice models, the following step is to apply the trained algorithm on the testing data for prediction, and detection accuracy. Figure 4.24-a shows the results given by a scatter plot of both classes with the prediction, followed by the ROC curve showing the area under the curve= 0.92 in Figure 4.24-b. Finally 4.24-c shows the confusion matrix of the tested model showing a true positive rates of 99% and 83%, false negative rates of 1% and 17%.

### 4.3.6 Unsupervised data for K-means algorithm on mice models with positive DCIS and IDC

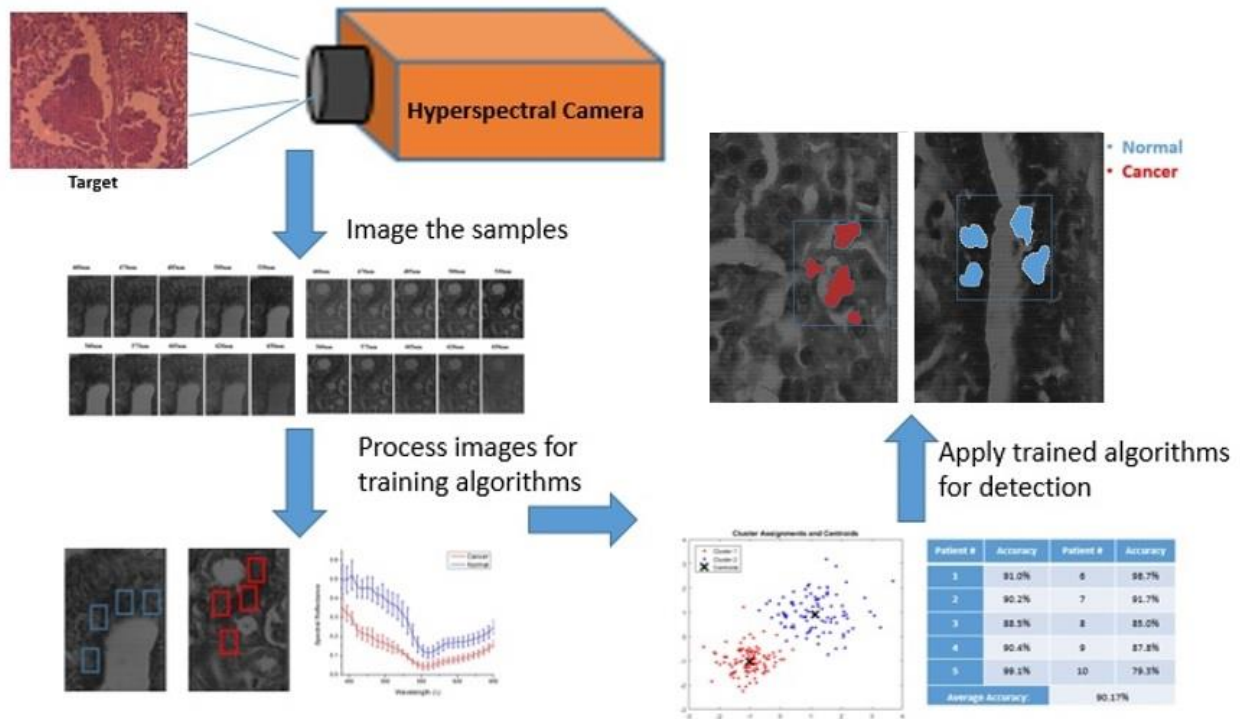
The K-means algorithm was ran on the unstained samples to create 2 centroids, as seen in Table 4.2. The K-means algorithm method described in Section 4.3.2 was applied to the stained mice samples. Table 4.6 shows the results of applying K-means separately on each set of data from 3 out of 4 mice modals for training and run to create two best values of centroids and finally, use the labels to measure the accuracy of each run. The last step is to run the trained K-means algorithm on the 4th mouse for detection in order to determine if the classifier can identify and detect both classes, and then compare the detected results with the ground truth pathologist diagnosis in order to support the algorithm accuracy. The detection results will be shown in the next section 4.5. Table 4.6 also shows the prediction accuracy of the algorithm by applying the labels on the data by using the centroids that the trained algorithm created.

**Table 4.6:** K-means algorithm on Mice samples

Patient #	Centroid1	Centroid2	Accuracy
1	<b>0.168</b>	<b>0.050</b>	<b>89.5%</b>
2	<b>0.185</b>	<b>0.047</b>	<b>93.47%</b>
3	<b>0.174</b>	<b>0.039</b>	<b>88.84%</b>
1-3	<b>0.149</b>	<b>0.056</b>	<b>83.03%</b>

## 4.5 Applying trained algorithms for semi-automated detection

This section provides a description of the effectiveness of the application of the trained detection algorithm for the detection of normal, DCIS, and IDC in random tissue samples from human and mice models that were not used in the previous testing or training sample sets. An overview of this application, from image acquisition, training, and application of the detection scheme is shown in Figure 4.25.

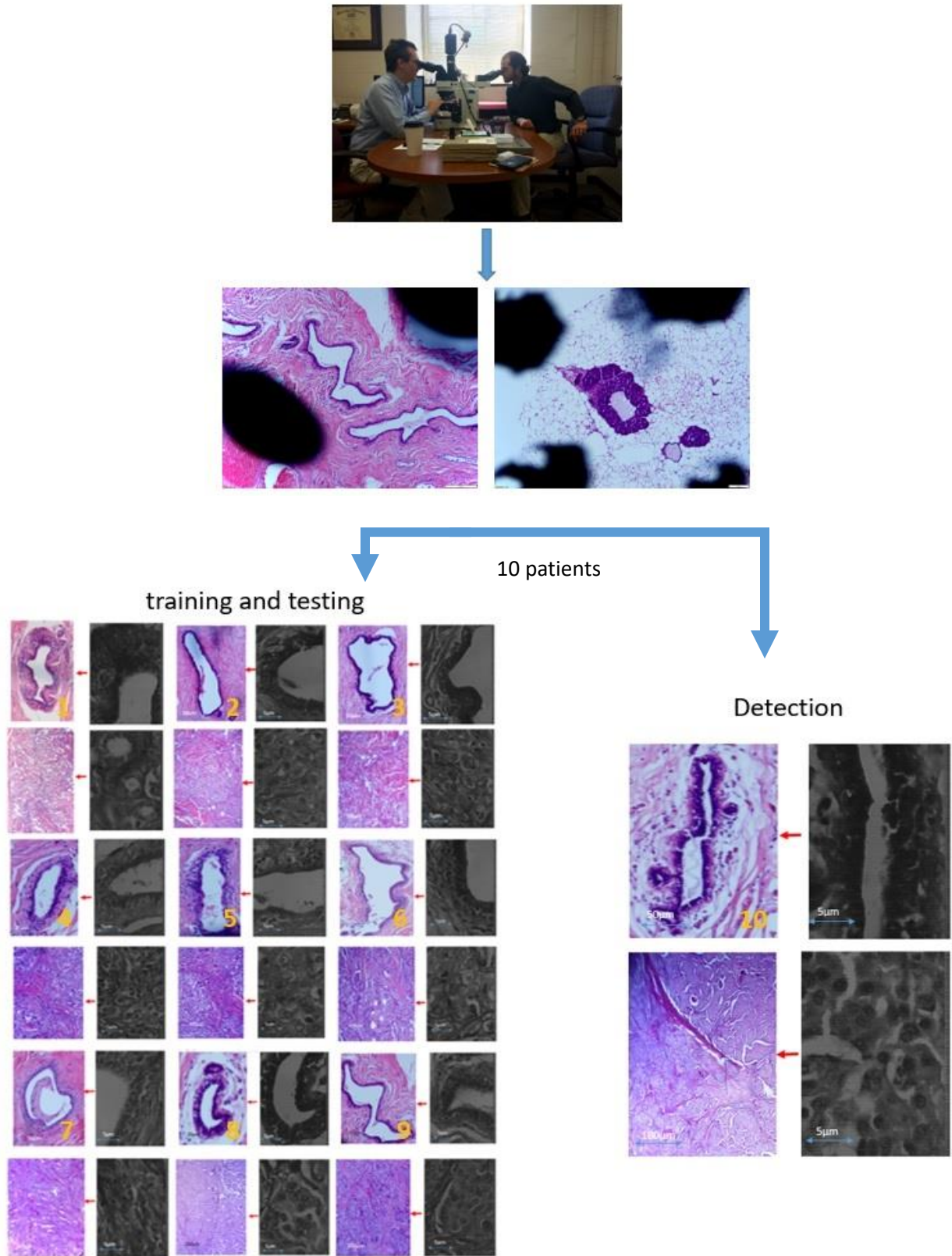


**Figure 4.25:** hyperspectral imaging, training, and detection.

#### 4.5.1 Cancer detection in human sample

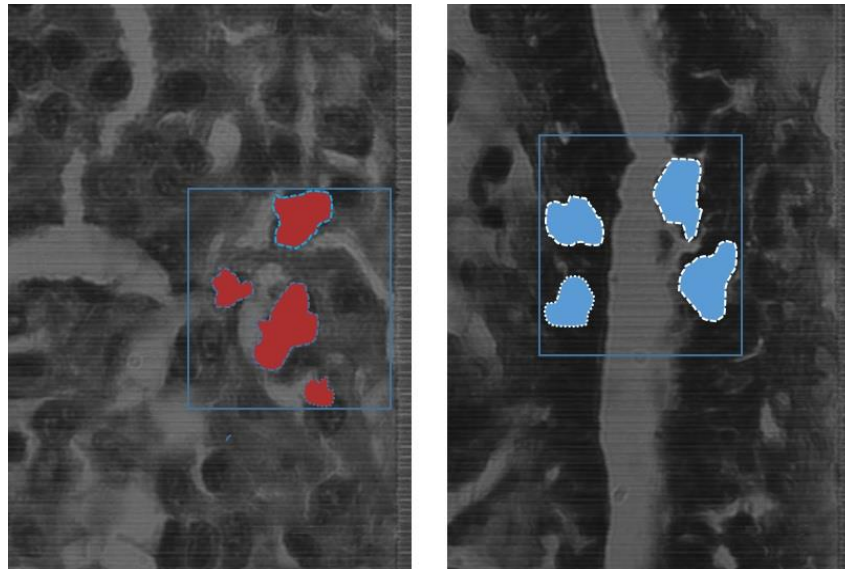
The trained algorithms were applied samples from patient that was diagnosed with DCIS. The samples were marked, and prepared by a pathologist, and then images were captured and processed by the hyperspectral imaging system. Figure 4.26 shows the steps taken to provide a ground truth that the trained algorithms were successfully able to match the pathologist's performance at detecting difference between normal and cancer. After the samples were prepared, Dr. James Coad the Vice-Chair of WVU Pathology diagnosed and marked regions of interest on the samples within the entire region of the tissue on the slide. Figure 4.26 also shows the samples from the 9 patients that were used in training the K-means algorithm as well as the 10<sup>th</sup> patient data used for detection.





**Figure 4.26:** Dr. James Coad marking the ROI to compare the detection results of the algorithm

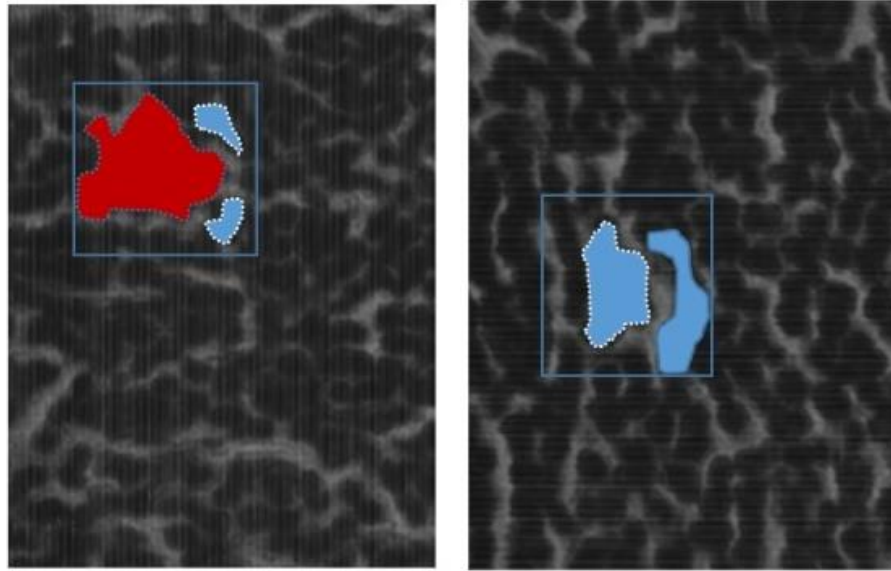
The result of training the K-means algorithm on the data that was extracted from the first 9 patients, then the trained algorithm was applied on the untrained data that was extracted from the 10<sup>th</sup> patient. The algorithm and successfully detected normal and cancer tissue as shown in Figure 4.27. In this figure, the blue boxes indicate the areas containing cancer marked by the pathologist, and the red and blue shaded regions are the regions with cancer identified by the hyperspectral detection algorithm.



**Figure 4.27:** The blue boxes indicates the areas marked by the pathologist, the shaded areas shows the detection of the K-means algorithm detecting cancer tissues (Red), and normal tissues (blue) on the spectral channel 550nm.

#### 4.5.2 Mice samples detection

The method described in Section 4.5.1 was repeated in on the mouse tissue samples. Three out of four samples were used for training and the 4<sup>th</sup> sample was used for testing and detection. The trained algorithms were applied on random tissue samples from mice that were diagnosed with both IDC, and DCIS. The samples were marked and diagnosed by Dr. Majed Pharaon from WVU pathology. Images were then captured and processed by the hyperspectral imaging system. Figure 4.28 shows that the trained algorithm were successfully able to detect a spectral difference between IDC and DCIS with an accuracy of 90%.



**Figure 4.28:** The blue boxes indicates the areas marked by the pathologist, the shaded areas shows the detection of the K-means algorithm detecting IDC (Red), and DCIS (blue) on the spectral channel 500nm.

## 4.6 Results & Summary

Image evaluation and analysis were applied on three sample groups in order to develop semi-automated methods to identify the spectral differences between:

- 1) normal and DCIS cancer tissues, and
- 2) DCIS and IDS tissues, which indicate different stages of breast cancer

The HSI system was used to image and analyze breast biopsy tissue from human patients, as well as mammary tumor tissue extracted from mouse models. The biopsy tissues were fixed and stained on pathology paraffin-embedded slides at the WVU Health Sciences Center. Each slide with either normal, invasive, or in situ ductal carcinoma was marked by the pathologist. In this study, two samples, consisting of normal and DCIS tissue, were obtained from each of 10 patients. There are two tissue samples from each mouse as well, but one is invasive and the other one is in situ ductal carcinoma. Each slide was imaged by the hyperspectral system, and a manual region of interest was chosen for testing the spectral differences, between normal, DCIS, and invasive and DCIS tissues. For more accurate results, an algorithm was built to choose random pixels from both regions of interest. The region of interest spectrum for chosen pixels within the images of the normal, invasive, and DCIS tissues shows spectral separation between both tissues, which at

~550nm in most cases. This wavelength was chosen for testing the spectral reflectance by color filtering. The spectral reflectance color filtering results show that the spectral reflectance values clearly identifies, and separates normal, and DCIS in the spectrum, and also a slight difference between invasive and DCIS, as DCIS tissue has higher spectral reflectance value within the chosen wavelength band.

In the second part, two image processing algorithms have been applied on the hyperspectral data of the tissue sample images for training the system, and then testing whether or not the system is capable of detecting the spectral difference. Each algorithm has a different approach. The first algorithm was a supervised Support Vector Machine (SVM) algorithm, where the algorithm is trained on image data labeled by a pathologist. The system then is trained according to the provided labels. Finally, the system will randomly pick data that has not been trained and test it. The detection of normal, and DCIS in human samples result reached a true positive rates of 97% and 92%, false negative rates of 8% and 3%, while, the detection of IDC, and DCIS in mice result reached a true positive rates of 99% and 83%, false negative rates of 1% and 17%. The second algorithm, K-means, was applied on unsupervised data, where the system trained itself to divide the data into two classes depending on the Euclidean distance. After the system classified two classes, it creates two centroids, one for each class. The value of each centroid keeps on changing while the algorithm is running until a minimum (or zero) change in distance is reached. Finally, labels were used to measure the algorithm accuracy. In the case of H&E-stained samples, the trained algorithm was applied on untrained data for testing and detection, and the results were compared with the pathologist diagnoses to show that the trained algorithm can differentiate between normal and cancer tissue, as well as between different stages of cancer. In summary, the hyperspectral imager system shows a potential for moving forward to create a baseline in order to make this system ready for clinical applications.

# 5. Conclusion and Future Work

Hyperspectral Imaging (HSI) is a sophisticated non-invasive optical imaging modality that has the potential to be applied toward medical imaging research and clinical practice. From imaging labs, to National Institute of Health (NIH), to cancer institutes, hyperspectral imaging has been an optical imaging tool that is used for providing spectral information, which can help in material detection. Hyperspectral imaging brings a new hope in breast cancer research and diagnostic imaging. Specifically, this research effort introduces a new tool that can help in breast cancer detection by providing pathologists with a tool that can potentially make their diagnosis process of ductal carcinoma under the microscope easier and more efficient. This work is a collaboration between the Lane Dept. of Computer Science and Electrical Engineering, the WVU Pathology Department, and the WVU Medicine Mary Babb Randolph Cancer Center.

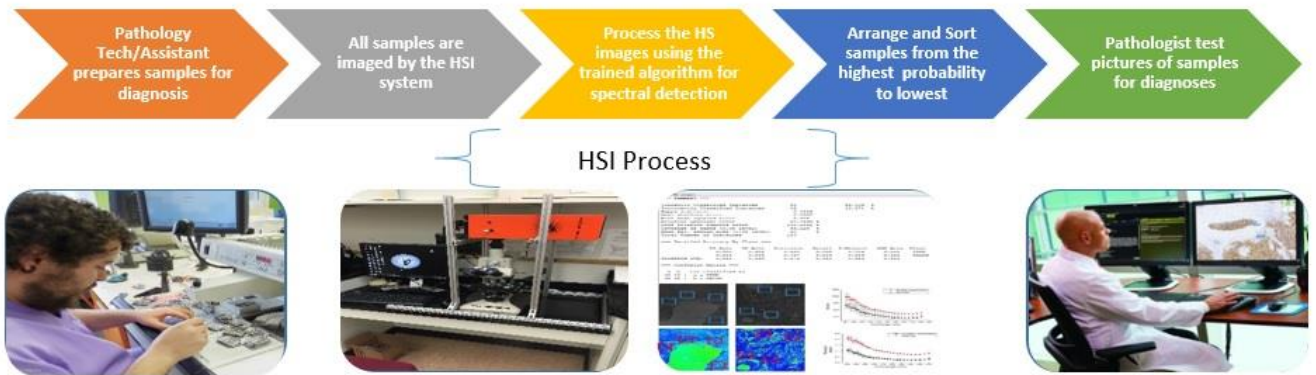


Figure 5.1: Concept diagram of how HSI can aid the pathologist in cancer detection and diagnosis. Pictures (left to right): pathology tech [60], hyperspectral imager system at WVU, image processing DCIS samples, pathologist diagnosing a sample on a monitor [59].

The overall scope of the work is to image samples of breast cancer tissues from different patients in order to show the ability of the hyperspectral system to detect spectral differences between normal and breast cancer tissues, as well as early and late stages of cancer, which will also effect the treatment decision. Figure (5.1), shows how hyperspectral imaging is introduced to pathology

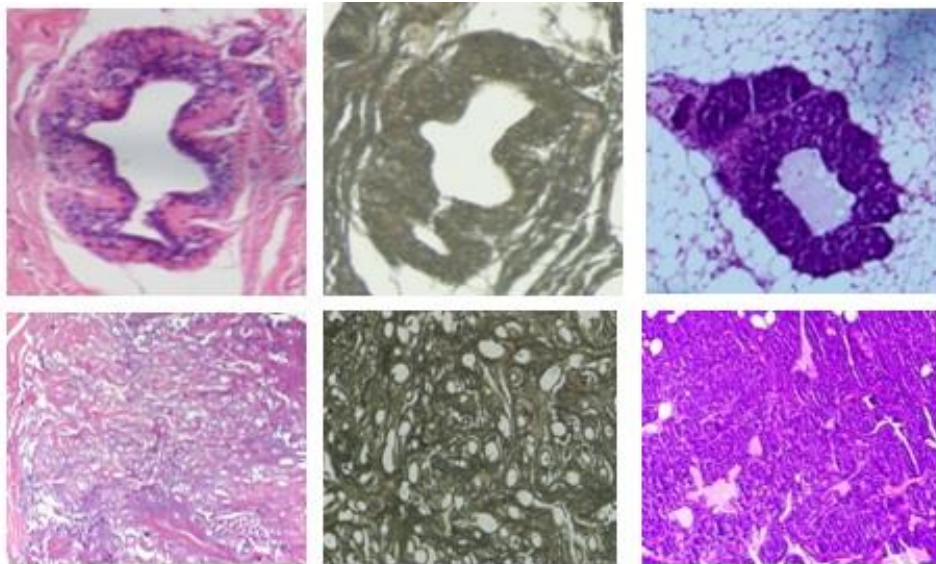


to be a bridge connects between pathology assistant or pathology tech, who prepare the samples for diagnosis, and the pathologist to make a diagnosis decision.

The HSI technology was applied on three case studies as follow:

- 1) H&E stained tissue samples of normal and DCIS from breast cancer patients
- 2) Unstained tissue samples of normal and DCIS from breast cancer patients
- 3) H&E stained tissue samples of the DCIS, and IDC of mice models.

Each of the three cases represents two tissue samples from each patient or mouse model. Each sample was imaged by the hyperspectral system and a manual region of interest was chosen for testing. The system results showed clear spectral reflectance separation between each of the two tissues that was tested in the three case studies, which are the H&E stained normal and Ductal carcinoma tissues. The unstained normal and ductal carcinoma tissues. And lastly, invasive and in situ mice mammary tumor to show spectral reflectance between breast cancer stages as shown in Figure (5.2).

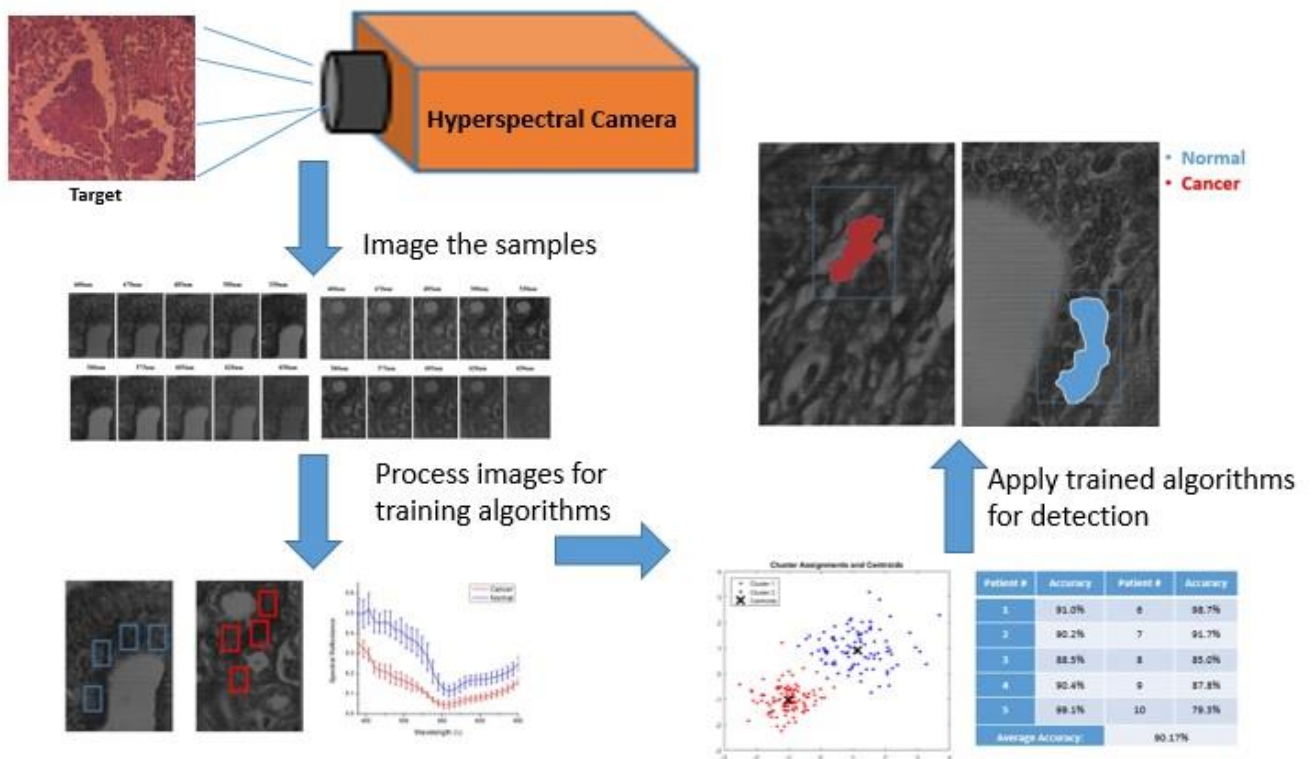


**Figure 5.2):** left to right, human normal and DCIS H&E tissues, human normal and DCIS unstained tissues, mice DCIS and IDC H&E tissues.

In some cases, the results that showed the largest separation appeared in the 550nm wavelength channel of the snapshot hyperspectral camera, making it a good wavelength to pick for testing the spectral reflectance color filtering method. The spectral reflectance color filtering method was



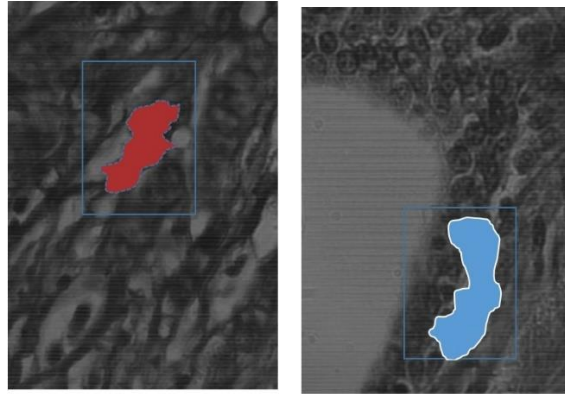
applied on the 550nm channel, which reflects a scale of colors depends on the spectral reflectance value. The results show a slight difference between normal and cancer tissues, while invasive and DCIS as DCIS tissue has higher spectral reflectance value for the chosen wavelength channel. In the hyperspectral image analysis chapter, two image processing algorithms, SVM and K-means, show the potential of building a semi-atomized system that can identify and sort between the samples and show the degree of the spectral reflectance difference. Figure (5.3), shows a diagram to explain the steps of applying the hyperspectral imaging system starting with imaging the samples, to analyzing the images, to train the algorithms on the data, and finally applying the trained algorithms for detection.



**Figure 5.3:** hyperspectral imaging, training, and detection.

The detection result reached the highest accuracy of 94.6% with an average accuracy of 90.17% by using the supervised SVM algorithm as detection can be seen in Figure 5.4. The second algorithm K-

means, was applied on unsupervised data, where the system trained itself to divide the data into two classes depending on the Euclidean distance.

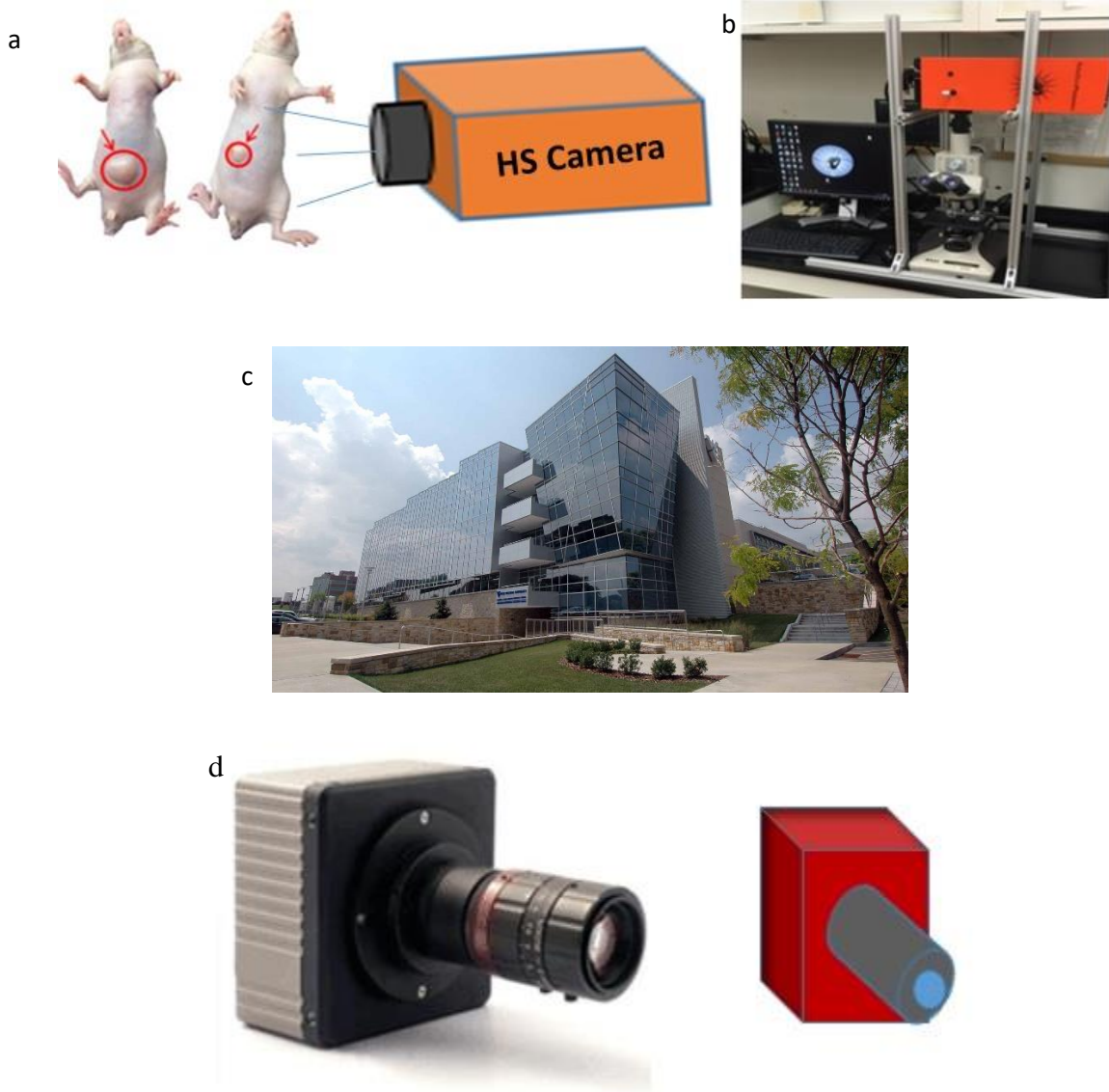


**Figure 5.4:** an example of applying SVM trained algorithm on a random samples to detect normal and breast cancer.

Finally, ground truth labeling was used to measure the algorithm accuracy. The highest accuracy of this algorithm was 93.47%, where the average accuracy was 88.71%, and applying the algorithm on the whole data joined together resulted in a higher accuracy than the average of 76.01%.

## Future Work

The future work that this project is moving toward will be taking two separate paths. The first one is to continue testing the hyperspectral system on normal, Ductal Carcinoma in Situ (DCIS), and invasive pathology samples but in larger number, which means the patient sample group will be expanded. In addition, for more accurate results as well as building a spectral library, approximately pure samples are more desired. Working with pathology, normal breast tissues will be acquired from the patient, which can be considered negative cancer in pathology report. The second path, would be a new cooperation with the animal facility and imaging facility by building a new animal study. The study plan includes detaching the hyperspectral camera from the microscope and using it as a regular camera and take pictures of whole mice models with breast cancer as shown in Figure (5.1), focusing on the breast tissues.



**Figure 5.5:** a- a prototype of hyperspectral camera used to image the outer breast tissue (skin) of breast cancer mice models. b- the hyperspectral imaging system, c- WVU Erma Byrd Biomedical Research Center, d- new hyperspectral camera model of imec.

In addition, a new hyper4spectral imager could be used, as seen in Figure (5.5). The new camera should have more capabilities to capture near IR and IR to be able to go deeper in the tissues in addition to the visible spectrum. This will open a new prospective and expand the research to include a variety of research ideas with a potential of clinical applications such as directly imaging the breast cancer mice modals without the need of imaging the biopsy tissues under the microscope, which may lead to a new optical diagnostic imaging system.

# References

- [1] R. Siegel et al., “Cancer statistics,” *Cancer J. Clin.* 64(1), 9–29 (2014).
- [2] American Cancer Society annual report (2016)
- [3] G. Lu, and B. Fei, “Medical hyperspectral imaging: a review,” *J Biomed Opt*, 19(1), 10901 (2014).
- [4] B. Saleh., ‘Fundamental of Photonics’
- [5] G. Lu, L. Halig, D. Wang et al., “Spectral-Spatial Classification Using Tensor Modeling for Cancer Detection with Hyperspectral Imaging,” *Proc SPIE*, 9034, 903413 (2014).
- [6] Tabar L, Vitak B, Chen HH, Yen MF, et al. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001; 91:1724-1732
- [7] D. BRADY R., “OPTICAL IMAGING AND SPECTROSCOPY”
- [8] H. Akbari, L. V. Halig, D. M. Schuster et al., “Hyperspectral imaging and quantitative analysis for prostate cancer detection,” *J Biomed Opt*, 17(7), 076005 (2012).
- [9] H. Akbari, L. V. Halig, H. Zhang et al., “Detection of Cancer Metastasis Using a Novel Macroscopic Hyperspectral Method,” *Proc SPIE*, 8317, 831711 (2012).

[10] M. S. Chin, B. B. Freniere, Y. C. Lo et al., “Hyperspectral imaging for early detection of oxygenation and perfusion changes in irradiated skin,” *J Biomed Opt*, 17(2), 026010 (2012).

[11] B. S. Sorg, B. J. Moeller, O. Donovan et al., “Hyperspectral imaging of hemoglobin saturation in tumor microvasculature and tumor hypoxia development,” *J Biomed Opt*, 10(4), 44004 (2005).

[12] J. Galeano, R. Jolivot, F. Marzani et al., “Unmixing of human skin optical reflectance maps by Non-negative Matrix Factorization algorithm,” *Biomedical Signal Processing and Control*, 8(2), 169-175 (2013).

[13] A. A. Fawzi, N. Lee, J. H. Acton et al., “Recovery of macular pigment spectrum in vivo using.

[14] D. Roblyer et al., “Multispectral optical imaging device for in vivo detection of oral neoplasia,” *J. Biomed. Opt.* 13(2), 024019 (2008).

[15] S. M. Ismail et al., “Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia,” *Br.Med. J.* 298(6675), 707–710 (1989).

[16] G. Lu et al., “Spectral-spatial classification using tensor modeling for cancer detection with hyperspectral imaging,” *Proc. SPIE* 9034, 903413 (2014).

[17] R. Pike et al., “A minimum spanning forest based hyperspectral image classification method for cancerous tissue detection,” *Proc. SPIE* 9034, 90341W (2014).



- [18] G. Lu et al., “Hyperspectral imaging for cancer surgical margin delineation: registration of hyperspectral and histological images,” Proc. SPIE 9036, 90360S (2014).
- [19] K. Masood and N. Rajpoot, “Texture based classification of hyperspectral colon biopsy samples using CLBP,” in IEEE Int. Symposium on Biomedical Imaging: From Nano to Macro, 2009. ISBI '09, Boston, Massachusetts, pp. 1011–1014 (2009).
- [20] K. Masood, “Hyperspectral imaging with wavelet transform for classification of colon tissue biopsy samples,” Proc. SPIE 7073, 707319 (2008).
- [21] X. Qin et al., “Automatic segmentation of right ventricle on ultrasound images using sparse matrix transform and level set,” Proc. SPIE 8669, 86690Q (2013).
- [22] S. V. Panasyuk et al., “Medical hyperspectral imaging to facilitate residual tumor identification during surgery,” Cancer Biol. Ther. 6(3), 439–446 (2007).
- [23] E. Claridge and D. Hidovic-Rowe, “Model based inversion for deriving maps of histological parameters characteristic of cancer from ex-vivo multispectral images of the colon,” IEEE Trans Med Imaging 33(4), 822–835 (2014).

- [24] X. Qin and B. Fei, “Measuring myofiber orientations from high frequency ultrasound images using multiscale decompositions,” *Phys. Med. Biol.* 59(14), 3907–3924 (2014).
- [25] X. Qin et al., “Extracting Cardiac Myofiber Orientations from High Frequency Ultrasound Images,” *Proc. SPIE* 8675, 867507 (2013).
- [26] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Rev.* 51(3), 455–500 (2009).
- [27] T. Vo-Dinh, Ed., *Biomedical Photonics Handbook*, Boca Raton, FL: CRC, 2003.
- [28] T. Vo-Dinh, M. Panjehpour, B.D. Overholt, C. Farris, and R. Sneed, “In vivo cancer diagnosis of the esophagus using differential normalized fluorescence (DNF) indices,” *Laser in Surgery and Medicine*, vol. 16, pp. 41–47, 1995.
- [29] H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, “Hyperspectral imaging and diagnosis of intestinal ischemia,” in *Proc. 30th Annu. Int. Conf. IEEE EMBC*, Vancouver, BC, Canada, Aug. 2008, pp. 1238–1241.
- [30] Y. Kosugi and H. Akbari, “Hyperspectral imaging: A new modality in surgery,” in *Recent Advances in Biomedical Engineering*, G. R. Naik, Ed. Austria: In-Tech publisher, 2009, pp. 223–240.
- [31] S. Friedland, D. Benaron, S. Coogan, D. Y. Sze, and R. Soetikno, “Diagnosis of chronic mesenteric ischemia by visible light spectroscopy during endoscopy,” *Gastrointest. Endosc.*, vol. 65, no. 2, pp. 294–300, Feb. 2007.
- [32] M. E. Martin, M. B. Wabuyele, K. Chen, P. Kasili, M. Panjehpour, M. Phan, B. Overholt, G. Cunningham, D. Wilson, R. C. Denovo, and T. Vo-dinh, “Development of an advanced hyperspectral imaging (HSI) system with

applications for cancer detection,” *Ann. Biomed. Eng.*, vol. 34, no. 6, pp. 1061–1068, Jun. 2006.

[33] Tuan V D, David L S, Nusundi B W, et al. A hyperspectral imaging system for in vivo optical diagnostics. *IEEE Eng Med Biol*, 2004, 23(5): 40–49

[34] Li H B, Shu R, Xue Y Q. Pushbroom hyperspectral imager and its potential application to oceanographic remote sensing. *J Infrared Millim Waves (in Chinese)*, 2002, 21(6): 429–433

[35] Xiao G H, Shu R, Xue Y Q. Design of microscopic hyperspectral imaging system. *Opt Prec Eng (in Chinese)*, 2004, 12(4): 367–372

[36] Martin A A, James B C, David M H. Multispectral imaging of burn wounds: A new clinical instrument for evaluating burn depth. *IEEE T Bio-Med Eng*, 1988, 35 (10): 842–850

[37] Martinez L. A non-invasive spectral reflectance method for mapping blood oxygen saturation in wounds. In: *Proceedings of the 31st Applied Imagery Pattern Recognition Workshop*. Washington D.C.:IEEE Computer Society, 2002. 112–116

[39] Seong G K, Zheng D, Matthew M, et al. Hyperspectral fluorescence image analysis for use in medical diagnostics, advanced biomedical and clinical diagnostic systems III. *Proc SPIE*, 2005, 5692: 21–28

[40] Matt E M, Wabuyelea M B, Panjehpourb M, et al. Dual modality fluorescence and reflectance hyperspectral imaging: Principle and applications. *Proc SPIE*, 2005, 5692: 133–139

[41] R. O. Green, M. C. Helmlinger, J. E. Conel, and J. van den Bosch, “Inflight validation of the calibration of Airborne Visible/Infrared Spectrometer in 1993,” *Proc. SPIE* 2231, 177-185 (1994).

- [42] D. K. Clark, M. Feinholz, M. Yarbrough, B. C. Johnson, S. W. Brown, Y. S. Kim, and R. A. Barnes, "Overview of the radiometric calibration of MOBY," Proc. SPIE 4483, 64 (2002).
- [43] S. Richtsmeier, R. Sundberg, A. Berk, S. Adler-Golden, and R. Haren, "Full spectrum scene simulation," Proc. SPIE 5425, 530-537 (2004).
- [44] Y. Zong, S. W. Brown, B. C. Johnson, K. R. Lykke, and Y. Ohno, "Simple spectral stray light correction method for array spectrometers," Appl. Opt. 45, 1111-1119 (2006).
- [45] Abbott, J. A., R. Lu, B. L. Upchurch, and R. L. Stroshine. 1997. Technologies for non-destructive quality evaluation of fruits and vegetables. Horticultural Reviews 20: 1-120.
- [46] Albers, B., J. DiBenedetto, S. Lutz, and C. Purdy. 1995. More efficient environmental monitoring with laser-induced fluorescence imaging. Biophotonics Int. Nov.: 42-54.
- [47] Borregaard, T., H. Nielsen, L. Norgaard, and H. Have. 2000. Crop-weed discrimination by line imaging spectroscopy. J. Agri. Eng. Res. 75: 389-400.
- [48] Chappelle, E. W., F. M. Wood, J. E. McMurtrey, and W. W. Newcomb. 1984. Laser induced fluorescence of green plants: 1. A technique for the remote detection of plant stress and species differentiation. Appl. Optics 23: 134-138.
- [49] Goetz, Alexander F.H., and Boardman, J.W. (1997). Atmospheric Corrections: On Deriving Surface Reflectance from Hyperspectral Imagers. In Descour, Michael R. and Shen, S.S. (eds.), Imaging Spectrometry III: Proceedings of SPIE, 3118, 14-22.
- [50] Van der Meer, Freek (1994). Calibration of Airborne Visible/Infrared Imaging Spectrometer Data (AVIRIS) to Reflectance and Mineral Mapping in Hydrothermal Alteration Zones: An Example from the "Cuprite Mining District". Geocarto International, 3, 23-37.

[51] Adams, John B., Smith, M.O., and Gillespie, A.R. (1993). Imaging Spectroscopy: Interpretation Based on Spectral Mixture Analysis. In Pieters, Carle M. and Englert, Peter A.J. (eds.), Remote Geochemical Analysis: Elemental and Mineralogic Composition. Cambridge, UK: Cambridge University Press, pp. 145-166.

[52] Clark, R.N., Gallagher, A.J., and Swayze, G.A. (1990). Material absorption band depth mapping of imaging spectrometer data using a complete band shape least-squares fit with library reference spectra. Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop, JPL Publication 90-54, pp. 176-186.

[53] G. Lu, L. Halig, D. Wang et al., "Spectral-spatial classification for noninvasive cancer detection using hyperspectral imaging," J Biomed Opt, 19(10), 106004 (2014).

[54] G. Lu, L. Halig, D. Wang et al., "Hyperspectral Imaging for Cancer Surgical Margin Delineation: Registration of Hyperspectral and Histological Images," Proc SPIE, 9036, 90360s (2014).

[55] Pike, S. K. Patton, G. Lu et al., "A Minimum Spanning Forest Based Hyperspectral Image Classification Method for Cancerous Tissue Detection," Proc SPIE, 9034, 90341w (2014).

[56] American Cancer Society. Detailed Guide: Breast Cancer. 2014. Accessed at [www.cancer.org/Cancer/BreastCancer/DetailedGuide/index](http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/index) on September 3, 2015.

[57] Centers for Disease Control and Prevention. National Breast and Cervical Cancer Early Detection Program. Accessed at [www.cdc.gov/cancer/nbccedp/about.htm](http://www.cdc.gov/cancer/nbccedp/about.htm) on September 3, 2015.



[58] Kushi LH, Doyle C, McCullough M, et al. American Cancer Society guidelines on nutrition and physical activity for cancer prevention: Reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin. 2012;62:30-67.

[59] Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005;353:1773-1783.

[60] Landarebe, D. Hyperspectral image data analysis. IEEE Signal Processing Magazine(2002),17-28.

[61] Li, Z., Li, L., Zhang, R., and Ma, J. An improved classification method for hyperspectral data based on spectral and morphological information. International Journal of Remote Sensing(2011), vol.32, 2919-2929.

[62] Gamps-Valls, G., Tuia, D., Bruzzone, L., and Atli Benediktsson, J. Advances in Hyperspectral Image Classification: Earth monitoring with statistical learning methods. Signal Processing Magazine, IEEE Journal (2014), 1007-1011.

[63] Starr, C. Biology: Concepts and Application, Thomson Brooks/Cole, 2005.

[64] Han, J., Kamber, M., and Pei, J. Data Mining: Concepts and Techniques., Morgan Kaufmann Publishers, 2011.

[65] D. Allen ; S. Maxwell ; Karel J. Zuzak.; “Hyperspectral image projection of a pig kidney for the evaluation of imagers used for oximetry”

[66] Kumar, Abbas, Aster; “Robbins Basic Pathology”

[67] D. Kopans; “Breast Imaging”

- [68] K. Heichman and J. M. Roberts, Cell 79, 557 (1994); J. Wuarin and P. Nurse, *ibid.* 85, 785 (1996).
- [69] T. Hunter and J. Pines, *ibid.* 79, 573 (1994).
- [70] M. Hall and G. Peters, Adv. Cancer Res. 68, 67 (1996).
- [71] Johns Hopkins Pathology Sidney Kimmel Cancer Center at Johns Hopkins.
- [72] R. Weinberg; "Biology of Cancer"
- [73] Jemal A, et al. Cancer statistics, 2005. CA Cancer J Clin. 2005;55:10-30.68:208-12.
- [74] Duffy SW, et al. The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties. Cancer. 2002;95:458-69.
- [75] Thomas DB, et al. Randomized trial of breast self-examination in Shanghai: final results. J Natl Cancer Inst. 2002;94:1445-57.
- [76] Semiglazov VF, et al. Evaluate the significance of self-examination for the early detection of breast cancer]. Vopr Onkol. 2003;49:434-41
- [77] Ellman R, et al. Breast self-examination programs in the trail of early detection of breast caner: ten year findings. Br J Cancer. 1993.
- [78] Smith RA, et al. American Cancer Society guidelines for breast cancer screening: update 2003. CA Cancer J Clin. 2003;53:141-69
- [79] M. Born and E. Wolf, Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light, 1999, Cambridge Univ.
- [80] L. V. Wang and H.-i. Wu, Biomedical Optics: Principles and Imaging, 2007, Wiley
- [81] C. Bohren and D. Huffman, Absorption and Scattering of Light by Small Particles, 1983, Wiley-Interscience

- [82] A. Ishimaru, Wave Propagation and Scattering in Random Media, 1999, IEEE
- [83] R. R. Anderson and J. A. Parrish, "The optics of human skin", J. Investigative Dermatol., vol. 77, no. 1, pp. 13-19, 1981
- [84] A. N. Bashkatov, "Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm", J. Phys. D: Appl. Phys., vol. 38, no. 15, pp. 2543-2555, 2005
- [85] S.Yoon, K. Lawrence, W. Windham, G. Heitschmidt,; "Line-scan hyperspectral imaging system for real-time inspection of poultry carcasses with fecal material and ingesta" 2011
- [86] Chein-I Chang (31 July 2003). Hyperspectral Imaging: Techniques for Spectral Detection and Classification. Springer Science & Business Media. ISBN 978-0-306-47483-5.
- [87] Hans Grahn; Paul Geladi (27 September 2007). Techniques and Applications of Hyperspectral Image Analysis. John Wiley & Sons. ISBN 978-0-470-01087-7.
- [88] S.M. Kay, Fundamentals of Statistical Signal Processing. Englewood Cliffs, NJ: Prentice Hall, 1998.
- [89] J.B. Adams, M.O. Smith, and A.R. Gillespie, "Remote geochemical analysis: Elemental and mineralogical composition," in Imaging spectroscopy: Interpretation Based on Spectral Mixture Analysis, C.M. Pieters and P.A.J. Englert, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1993, pp.145-166.
- [90] A.D. Stocker and A. Schaum, "Application of stochastic mixing models to hyperspectral detection problems," SPIE Proc., vol. 3071, April 1997.
- [91] A. Schaum and A. Stocker, "Spectrally-selective target detection," in Proc.ISSSR, 1997.

[92] C. Robison, C. Kolanko, T. Bourlai, J. Dawson, "Imaging of blood cells based on snapshot Hyper-Spectral Imaging systems", in Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXI, Miguel Velez-Reyes; Fred A. Kruse, Editors, Proceedings of SPIE Vol. 9472 (SPIE, Bellingham, WA 2015), 94721L.

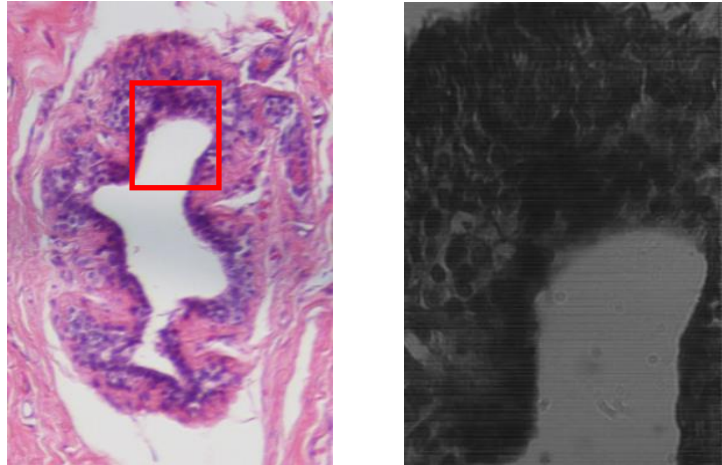
[93] D. Allen; S. Maxwell; J. Rice; R. Chang; M. Litorja; J. Hwang; J. Cadeddu; E. Livingston; E. Wehner; K. Zuzak; 'Hyperspectral image projection of a pig kidney for the evaluation of imagers used for oximetry' 2011

[94] D. T. Dicker et al., "Differentiation of normal skin and melanoma using high resolution hyperspectral imaging," Cancer Biol. Ther. 5(8), 1033–1038 (2006).

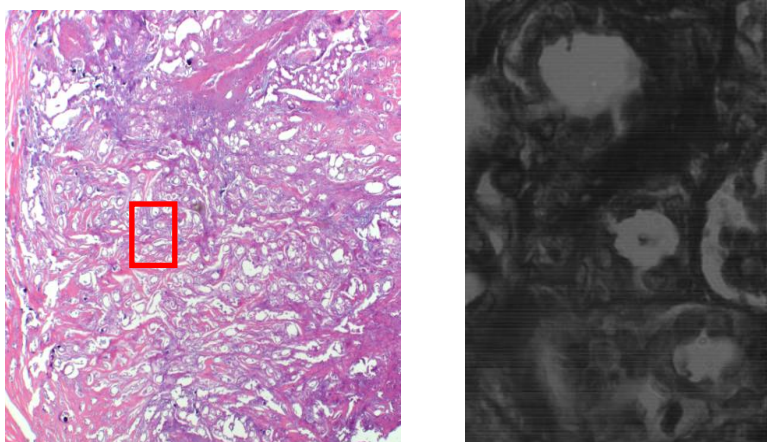
# Appendix A

Images of H and E Stained normal and DCIS samples with the manually picked spectral reflectance regions:

A-1: patient 1



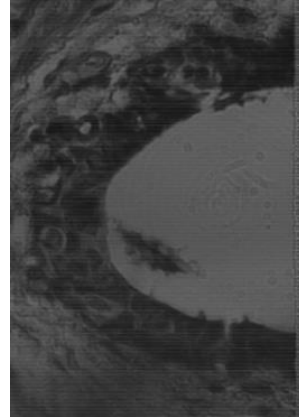
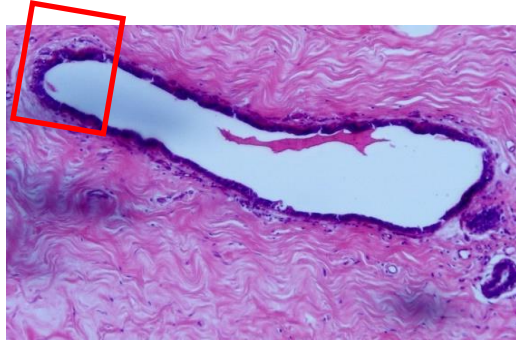
H&E stained sample of normal duct (Left), HSI (Right)



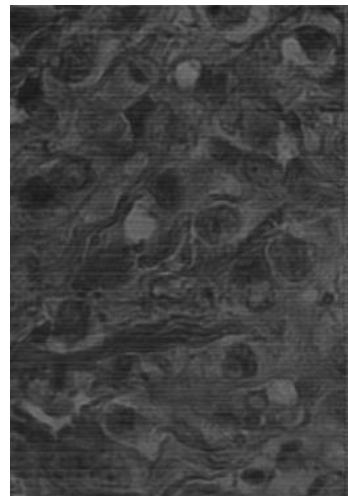
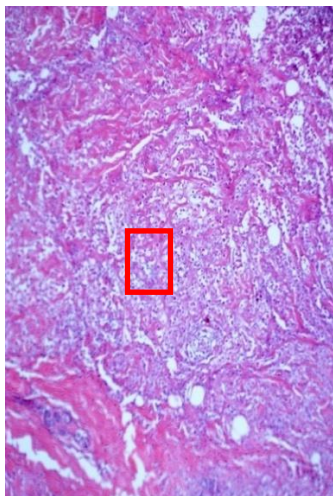
H&E stained sample of ductal Carcinoma (Left), HSI (Right)



**A-2: Patient 2**

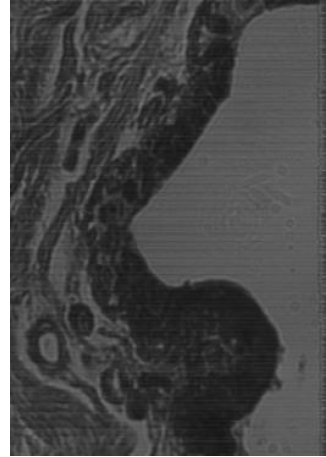
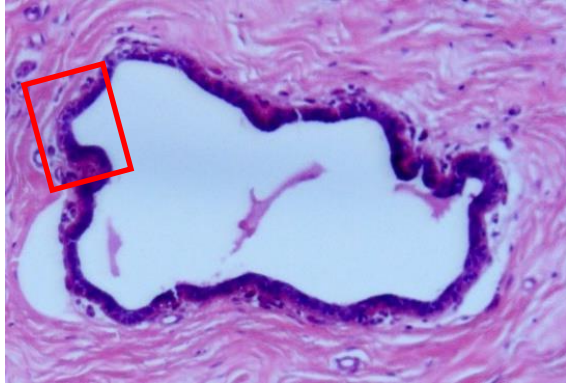


H&E stained sample of normal duct (Left), HSI (Right)

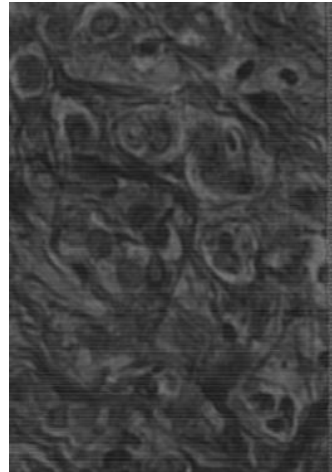
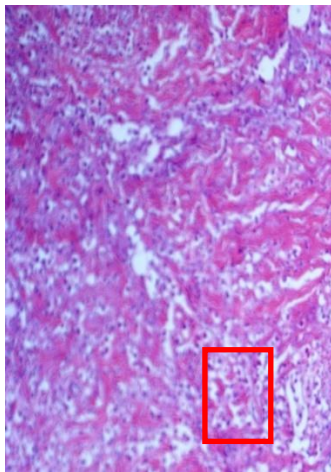


H&E stained sample of ductal Carcinoma (Left), HSI (Right)

**A-3: Patient 3**

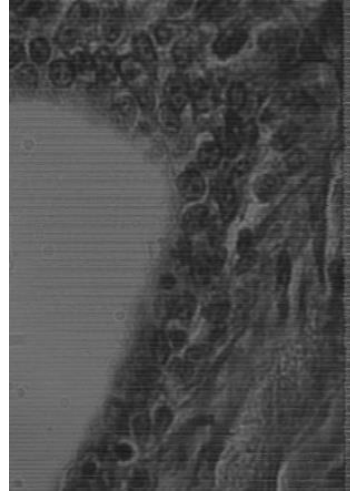
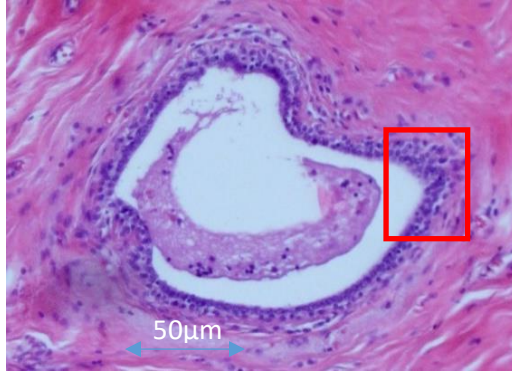


H&E stained sample of normal duct (Left), HSI (Right)

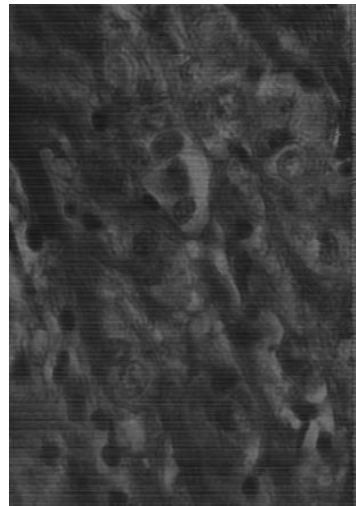
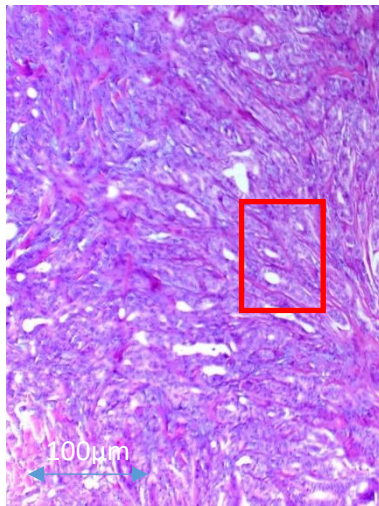


H&E stained sample of ductal Carcinoma (Left), HSI (Right)

**A-4: Patient 4**

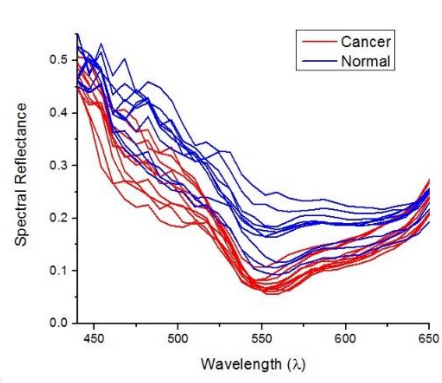
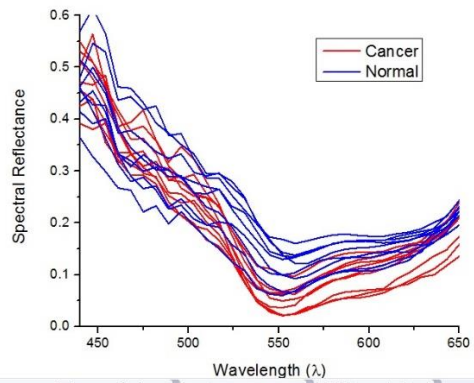
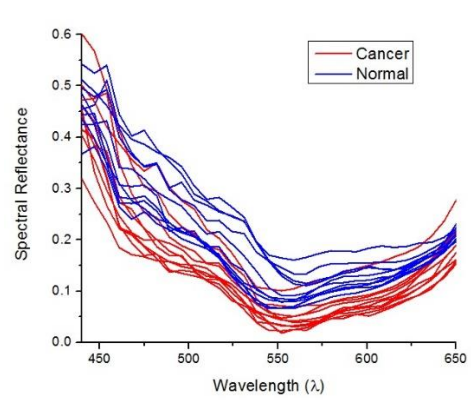
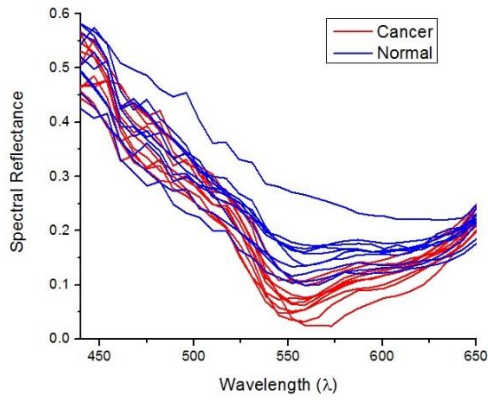
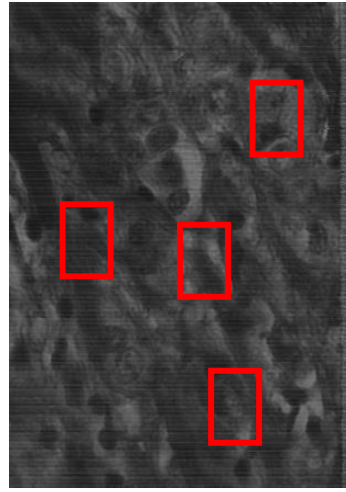
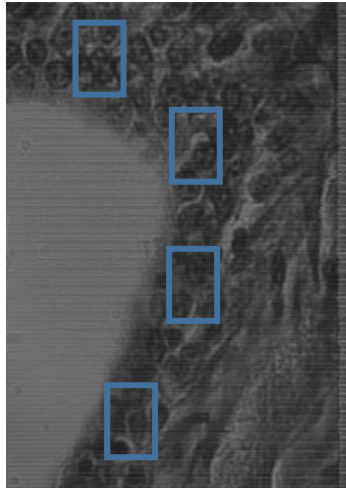


H&E stained sample of normal duct (Left), HSI (Right)



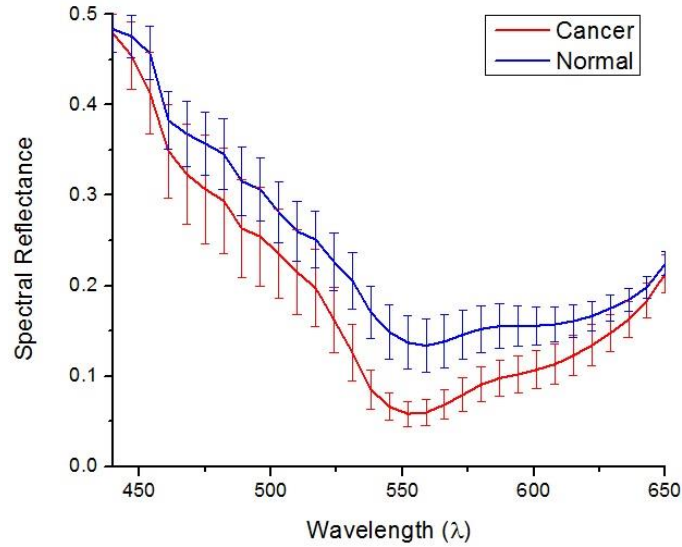
H&E stained sample of ductal Carcinoma (Left), HSI (Right)

Spectral Reflectance of the 4 marked areas:



Spectral reflectance of each of the 4 areas from top left to bottom right.

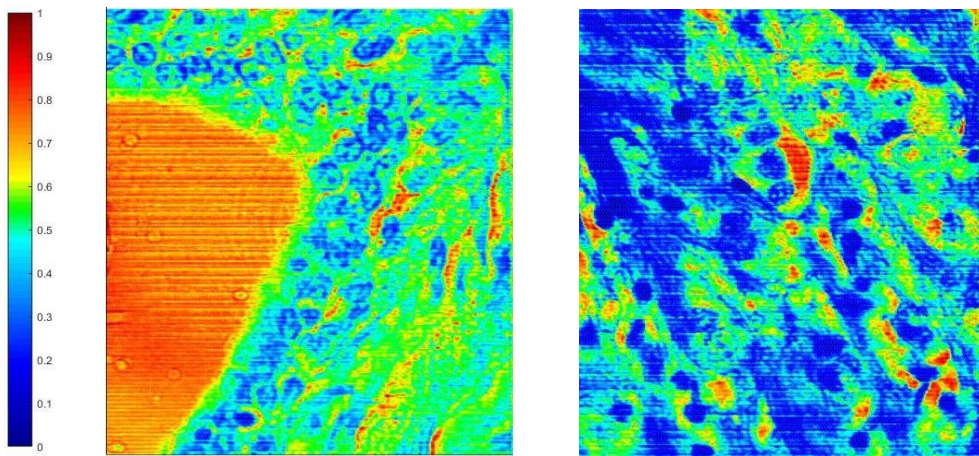




Average spectral reflectance of the 4 areas with standard deviation applied.

Apply Spectral Reflectance Color Filtering (SRCF) method.

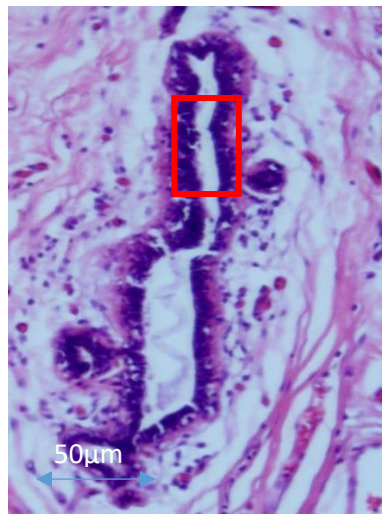
SRCF method is applied on the samples to visualize the difference in the spectral reflectance between both samples where the dark blue reflects the lowest reflectance intensity value



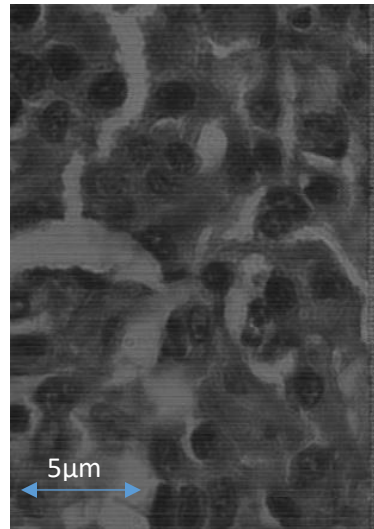
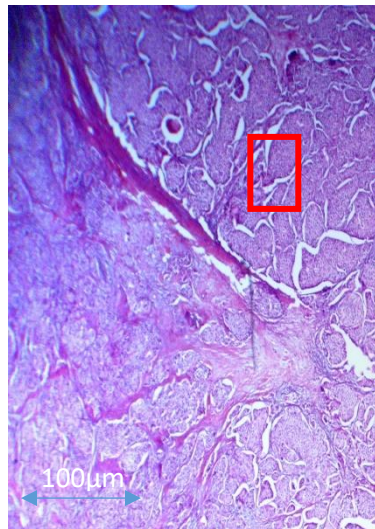
SRCF on both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.



**A-5: Patient 5**

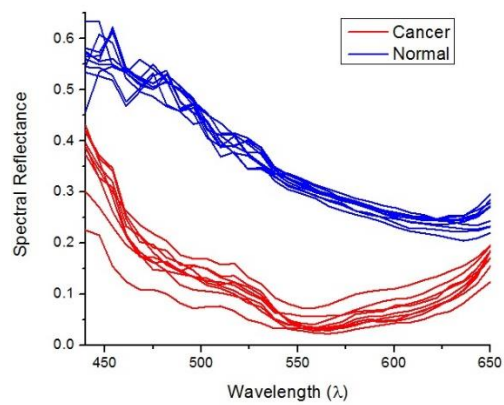
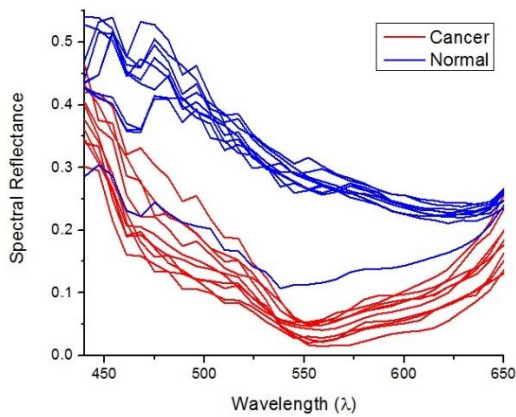
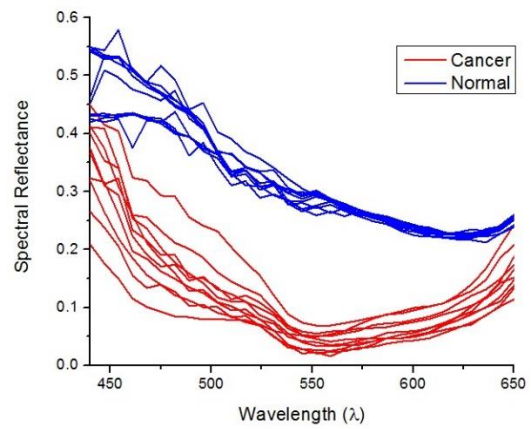
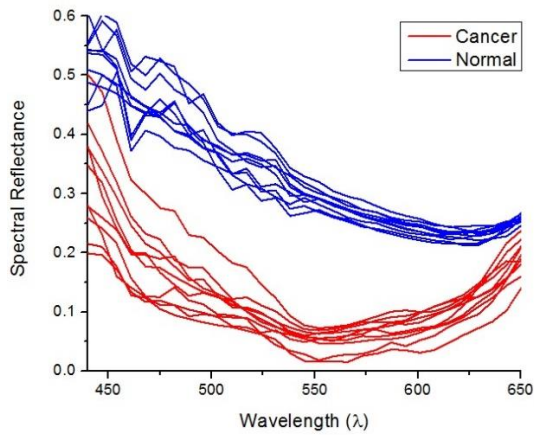
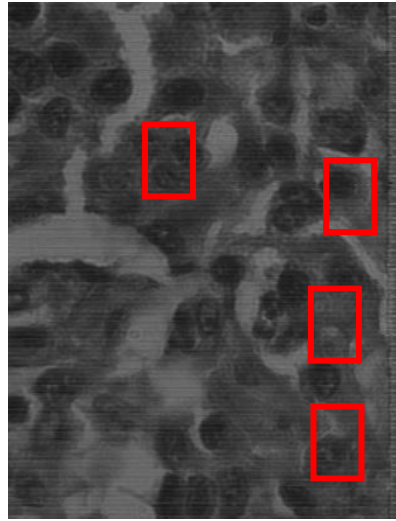
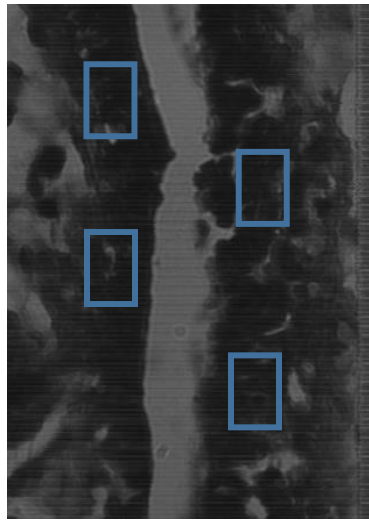


H&E stained sample of normal duct (Left), HSI (Right)

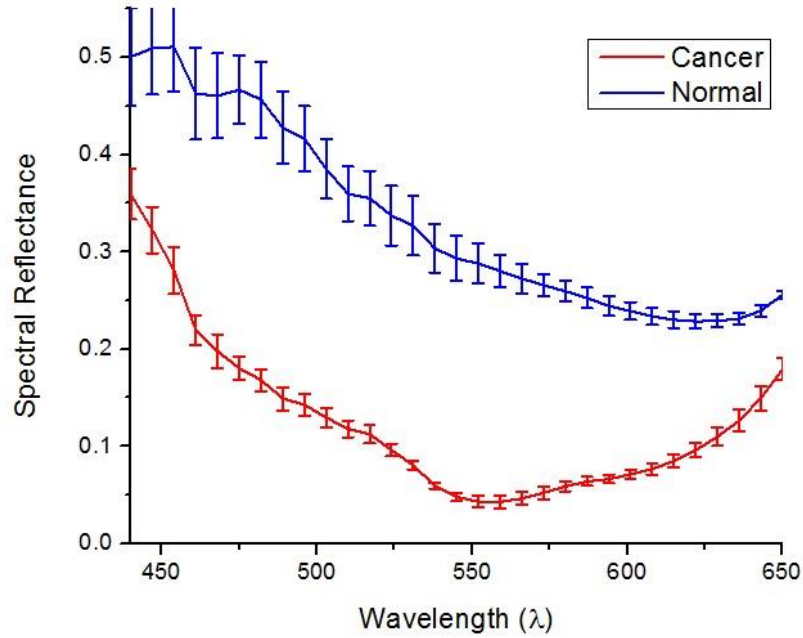


H&E stained sample of normal duct (Left), HSI (Right)

Spectral Reflectance of the 4 marked areas:



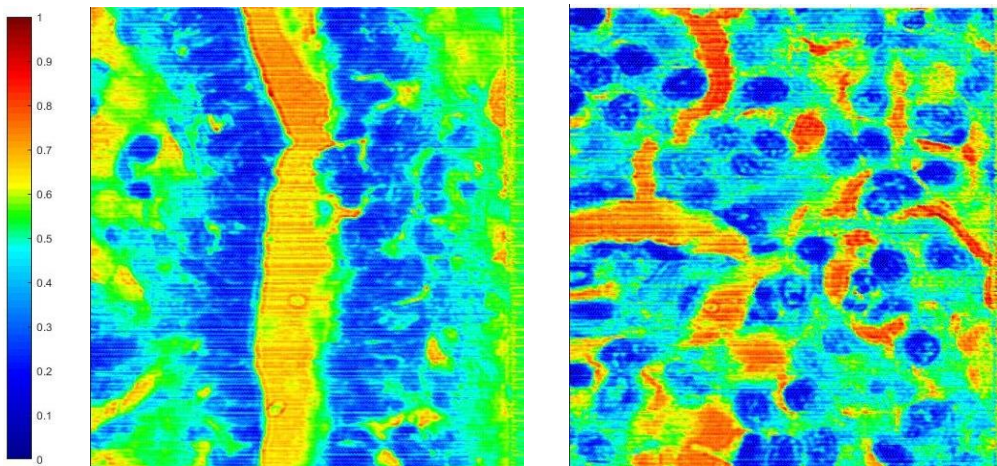
Spectral reflectance of each of the 4 areas from top left to bottom right.



Average spectral reflectance of the 4 areas with standard deviation applied.

Apply Spectral Reflectance Color Filtering (SRCF) method.

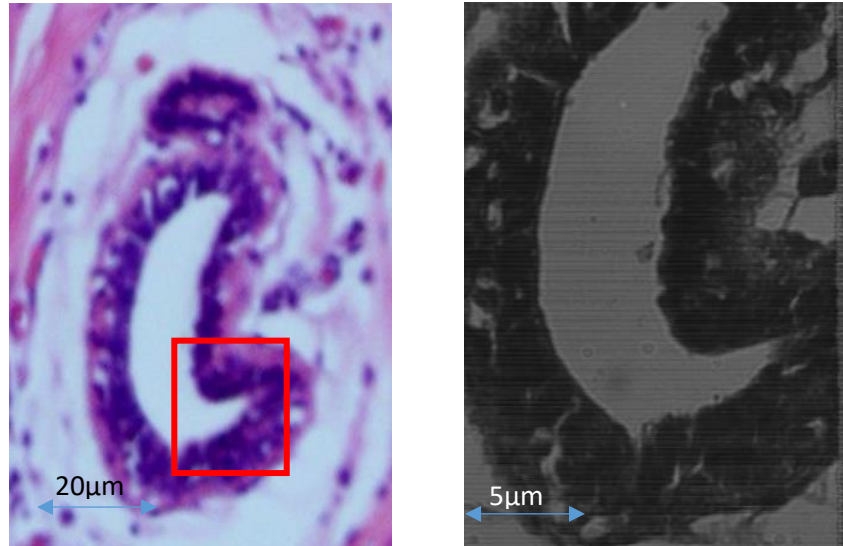
SRCF method is applied on the samples to visualize the difference in the spectral reflectance between both samples where the dark blue reflects the lowest reflectance intensity value



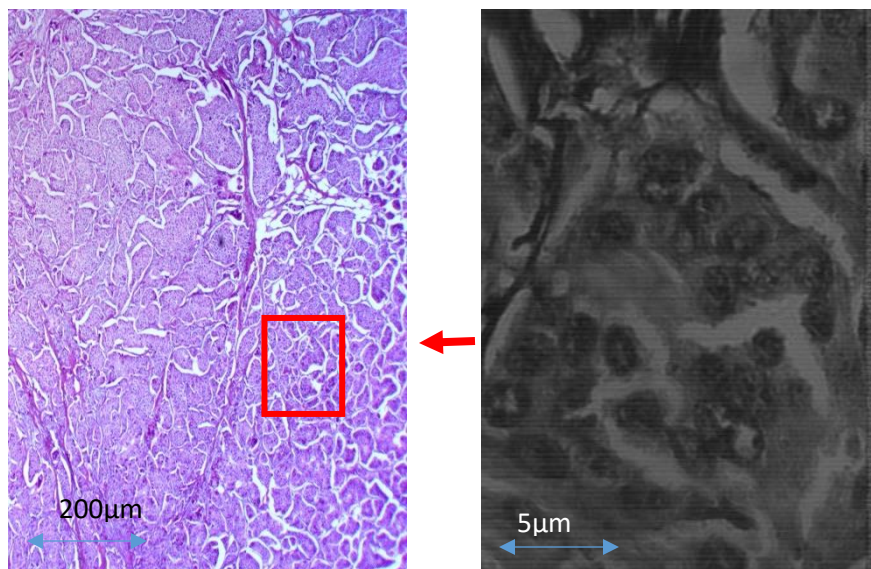
SRCF on both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.



**A-6: Patient 6**

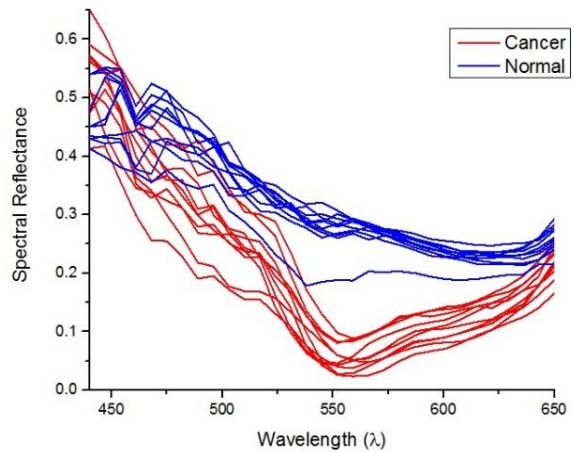
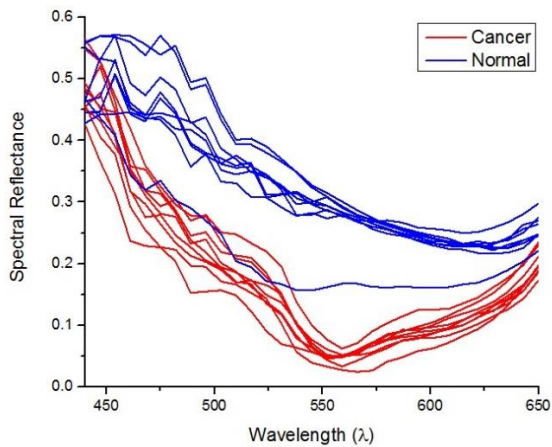
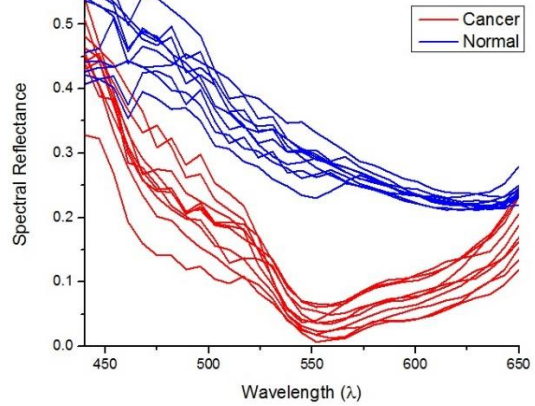
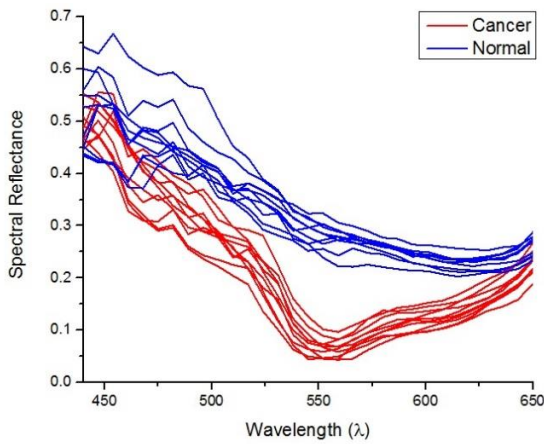
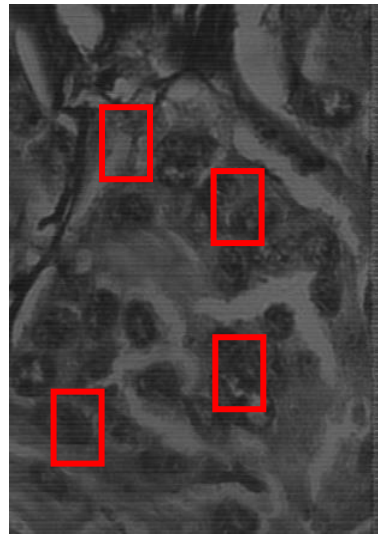
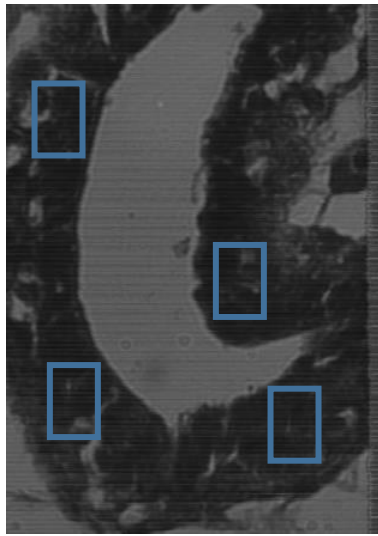


H&E stained sample of normal duct (Left), HSI (Right)



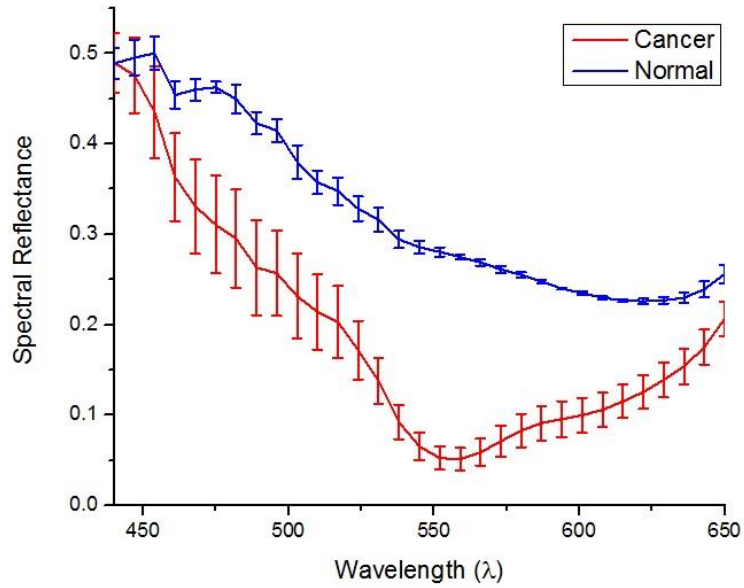
H&E stained sample of ductal carcinoma (Left), HSI (Right)

Spectral Reflectance of the 4 marked areas:



spectral reflectance of each of the 4 areas from top left to bottom right.

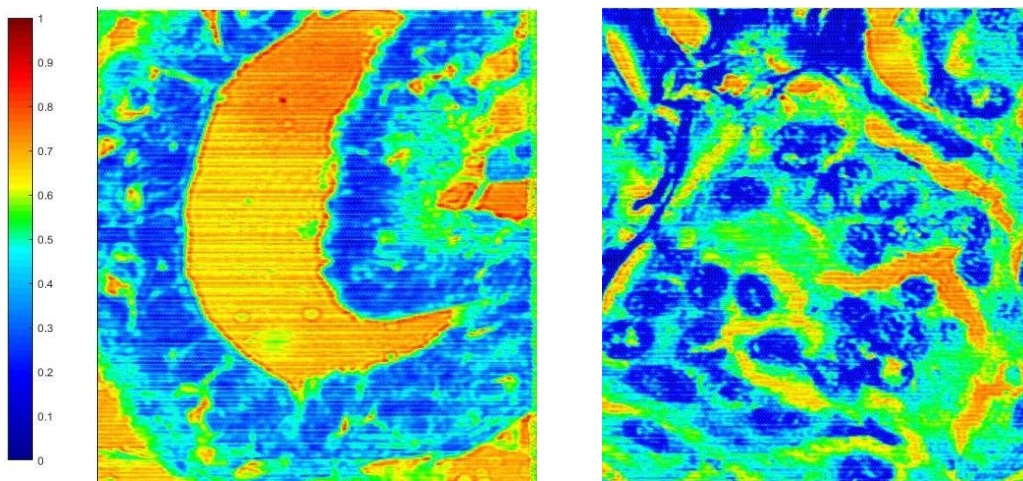




Average spectral reflectance of the 4 areas with standard deviation applied.

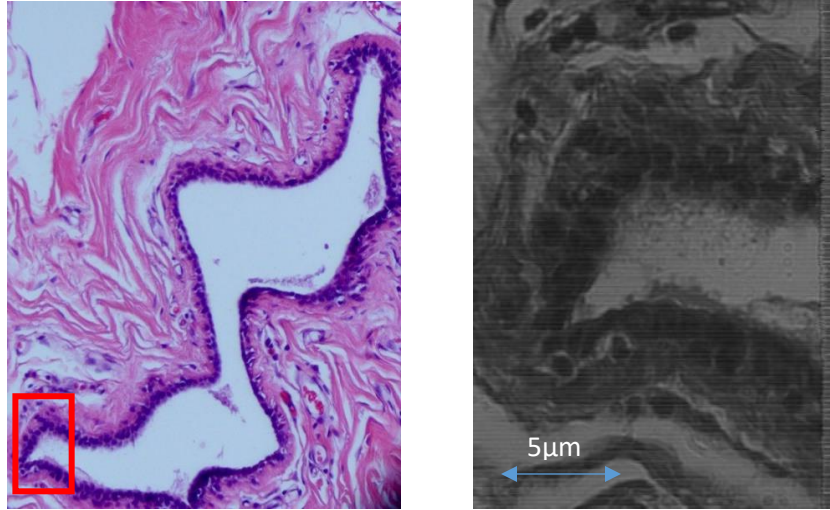
Apply Spectral Reflectance Color Filtering (SRCF) method.

SRCF method is applied on the samples to visualize the difference in the spectral reflectance between both samples where the dark blue reflects the lowest reflectance intensity value

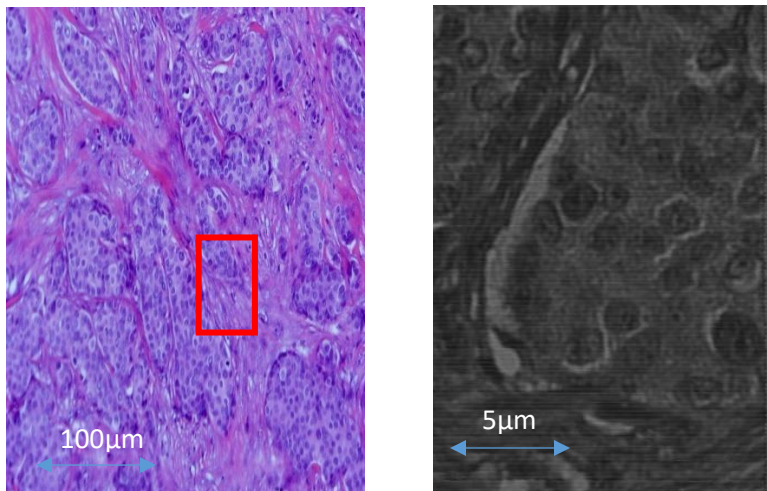


SRCF on both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.

**Patient 7:**

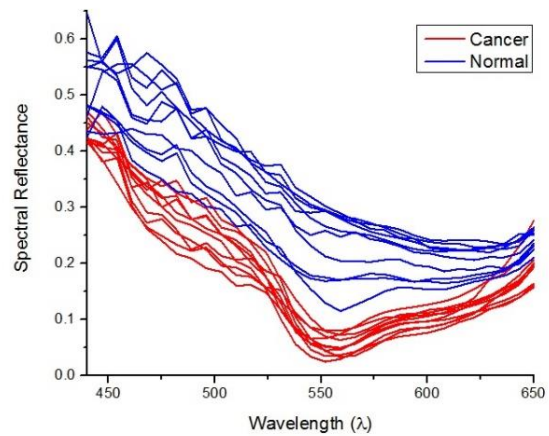
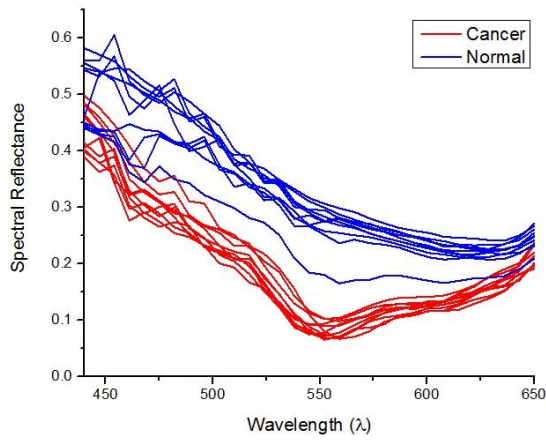
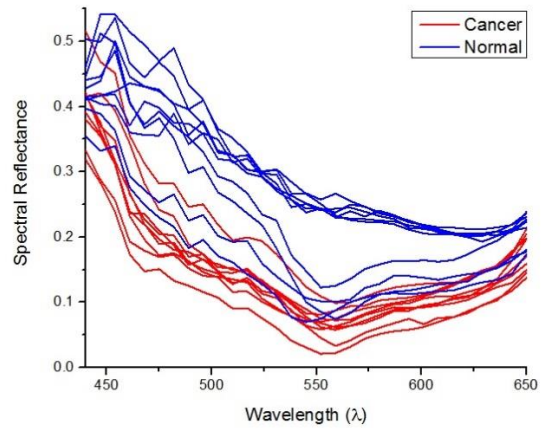
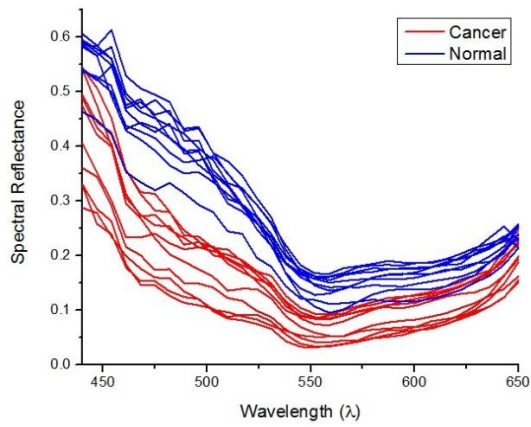
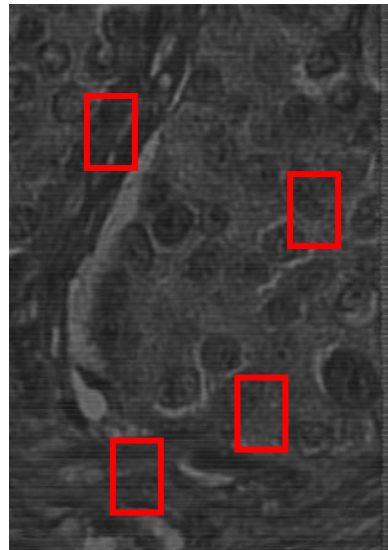
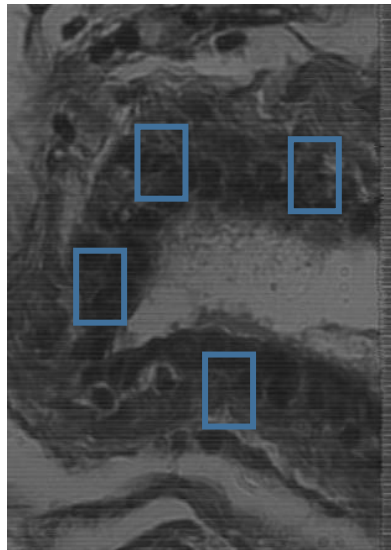


H&E stained sample of normal duct (Left), HSI (Right)



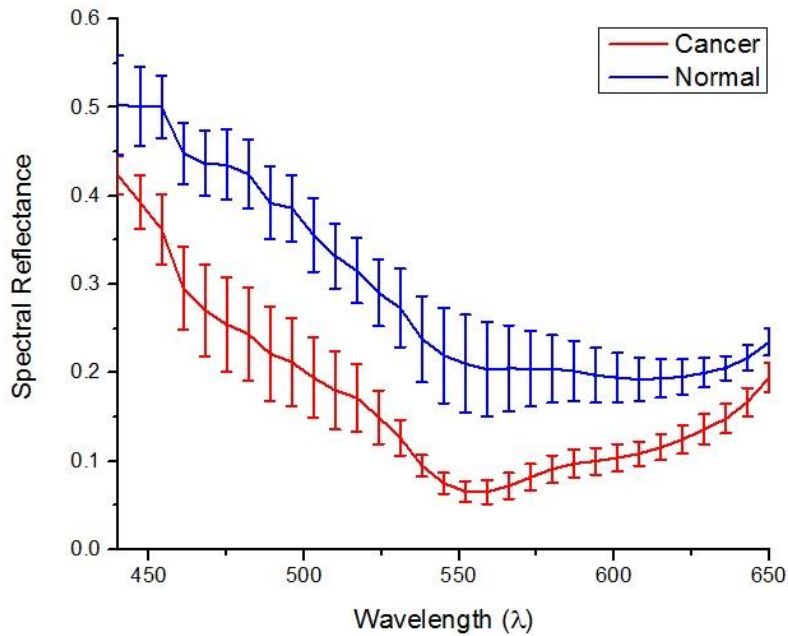
H&E stained sample of ductal carcinoma (Left), HSI (Right)

Spectral Reflectance of the 4 marked areas:



:spectral reflectance of each of the 4 areas from top left to bottom right.

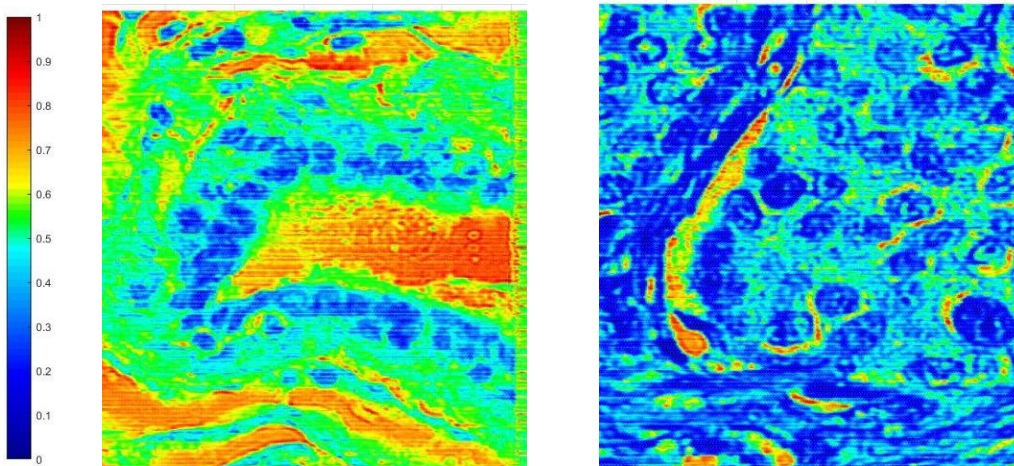




Average spectral reflectance of the 4 areas with standard deviation applied.

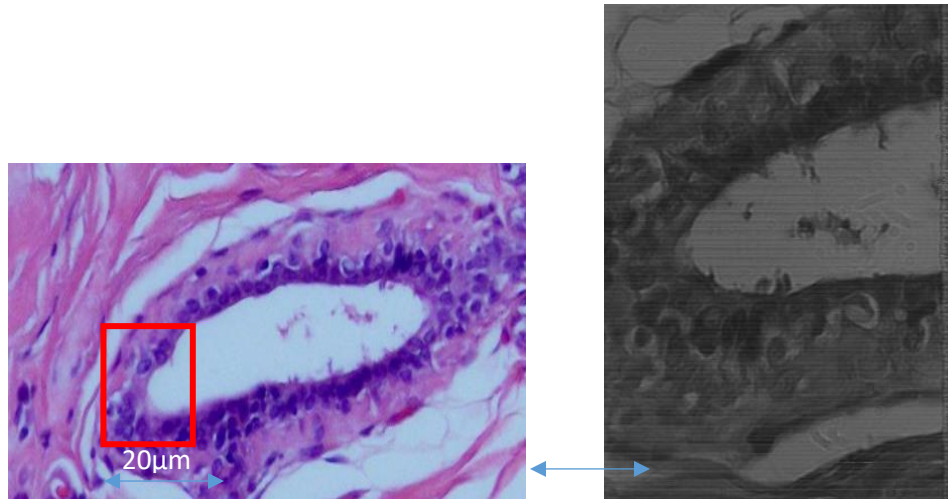
Apply Spectral Reflectance Color Filtering (SRCF) method.

SRCF method is applied on the samples to visualize the difference in the spectral reflectance between both samples where the dark blue reflects the lowest reflectance intensity value

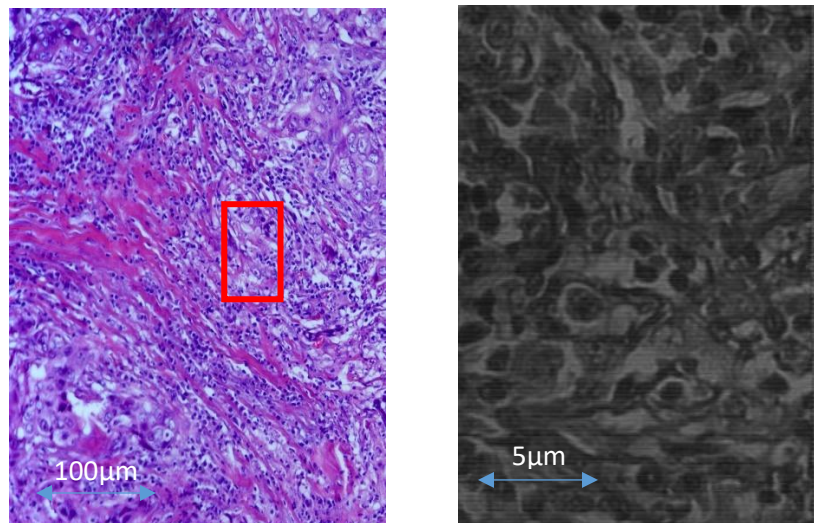


SRCF on both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.

**A-8: Patient 8**



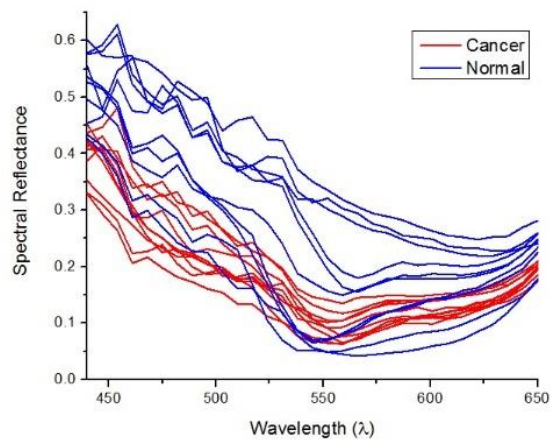
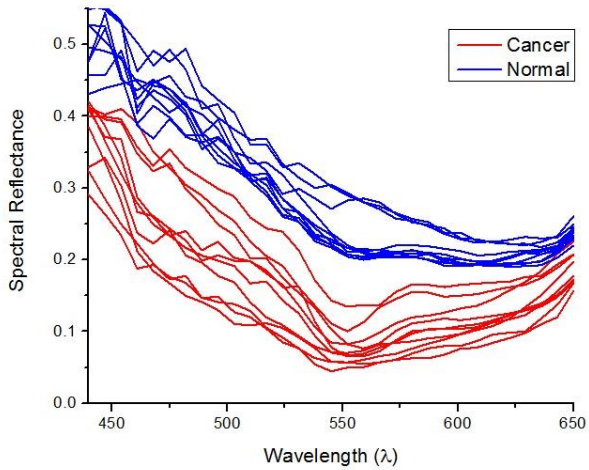
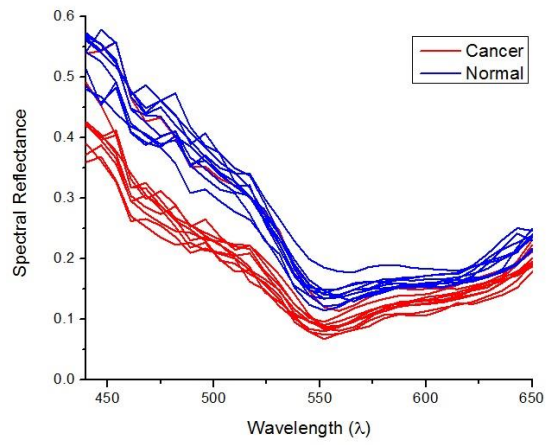
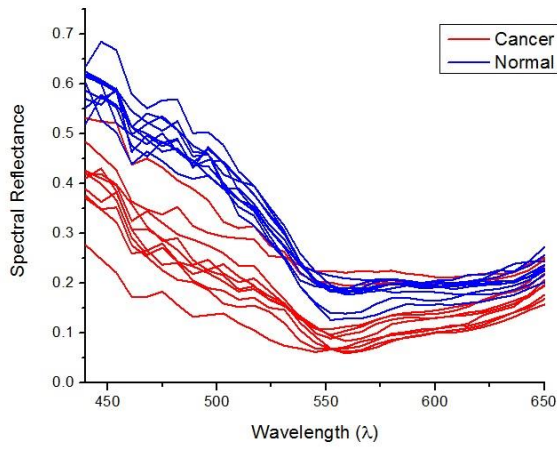
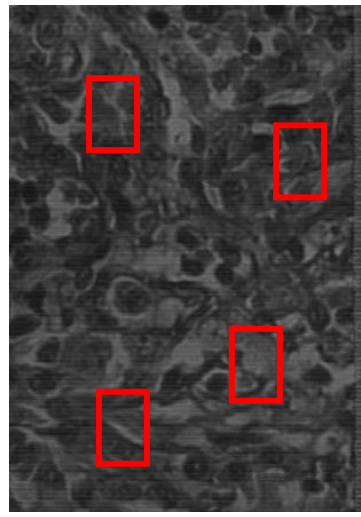
H&E stained sample of normal duct (Left), HSI (Right)



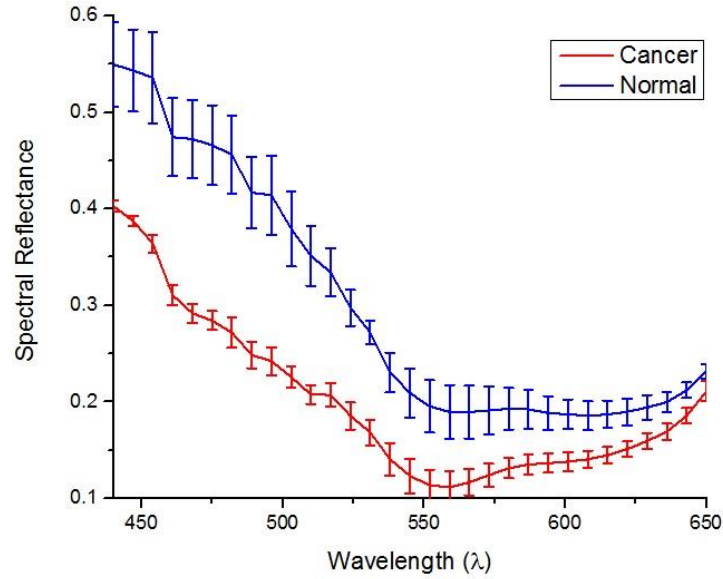
H&E stained sample of ductal carcinoma (Left), HSI (Right)



Spectral Reflectance of the 4 marked areas:



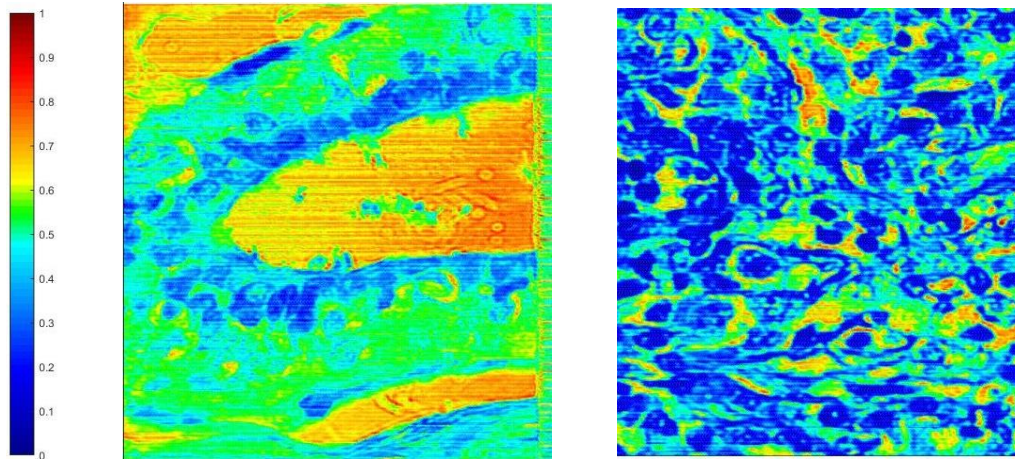
Spectral reflectance of each of the 4 areas from top left to bottom right.



Average spectral reflectance of the 4 areas with standard deviation applied.

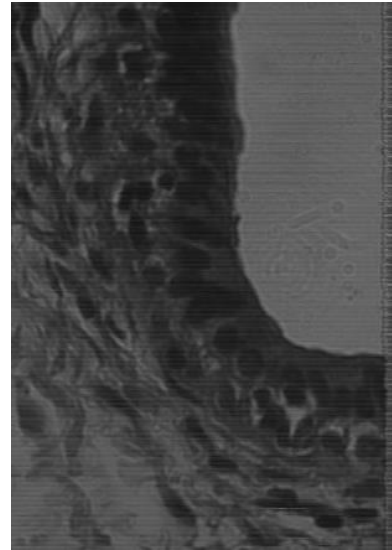
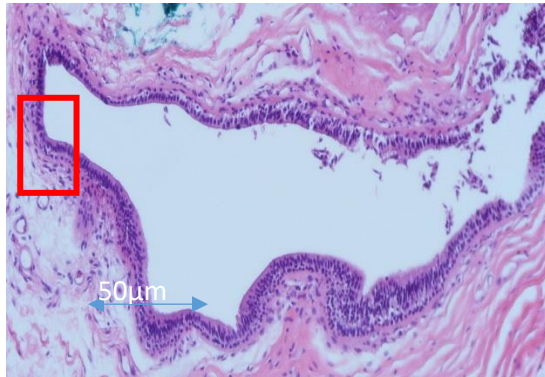
Apply Spectral Reflectance Color Filtering (SRCF) method.

SRCF method is applied on the samples to visualize the difference in the spectral reflectance between both samples where the dark blue reflects the lowest reflectance intensity value



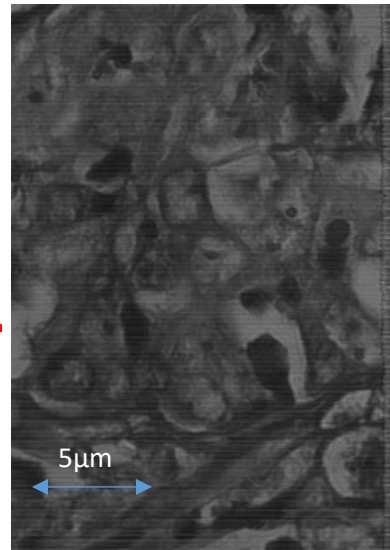
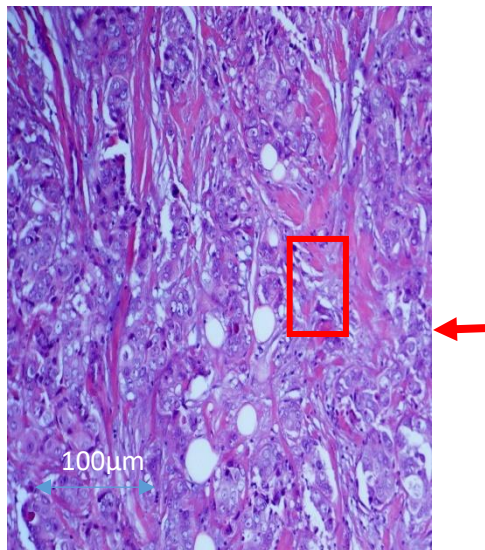
: SRCF on both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.

**A-10 Patient 10**



H&E stained sample of normal duct (Left), HSI (Right)

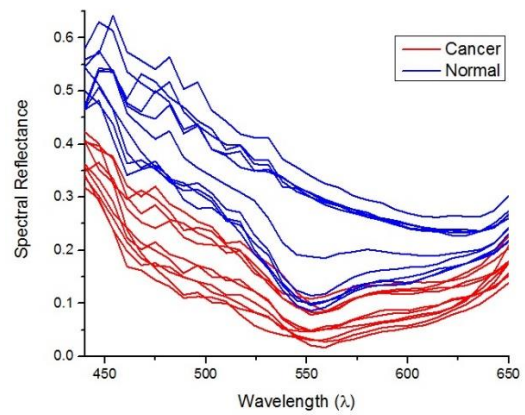
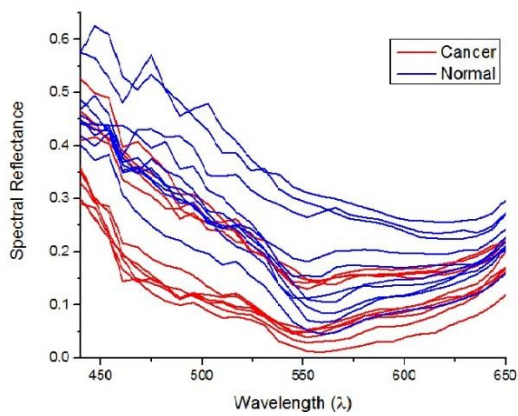
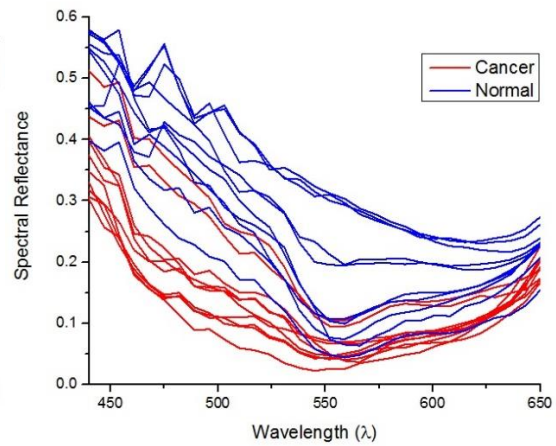
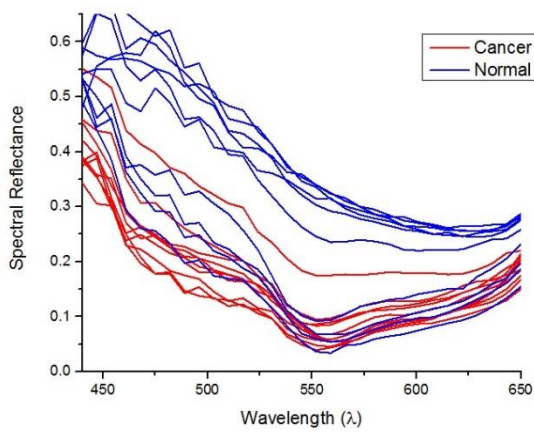
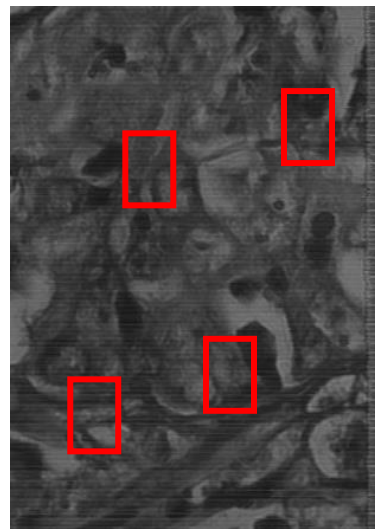
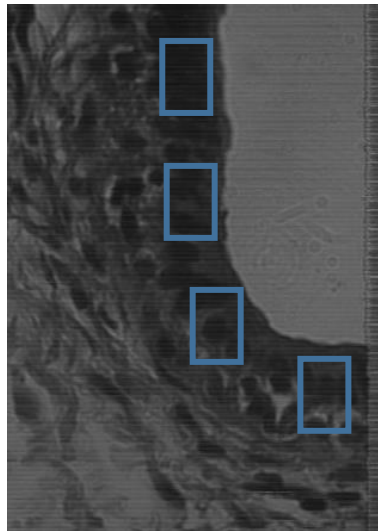
:



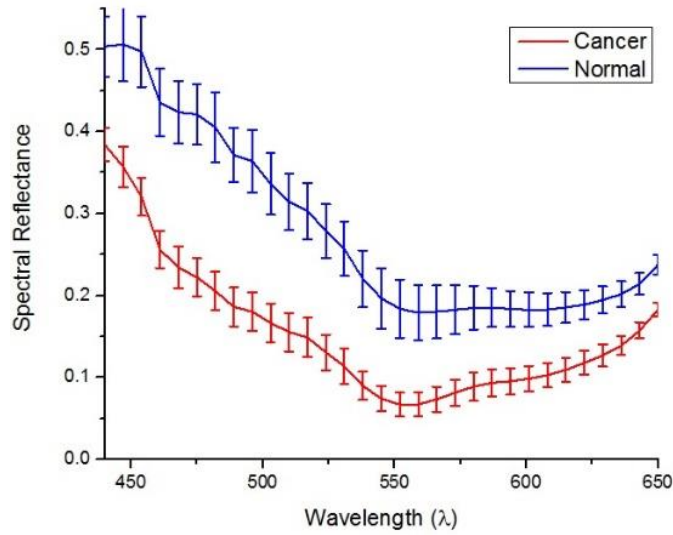
H&E stained sample of ductal carcinoma (Left), HSI (Right)



Spectral Reflectance of the 4 marked areas:



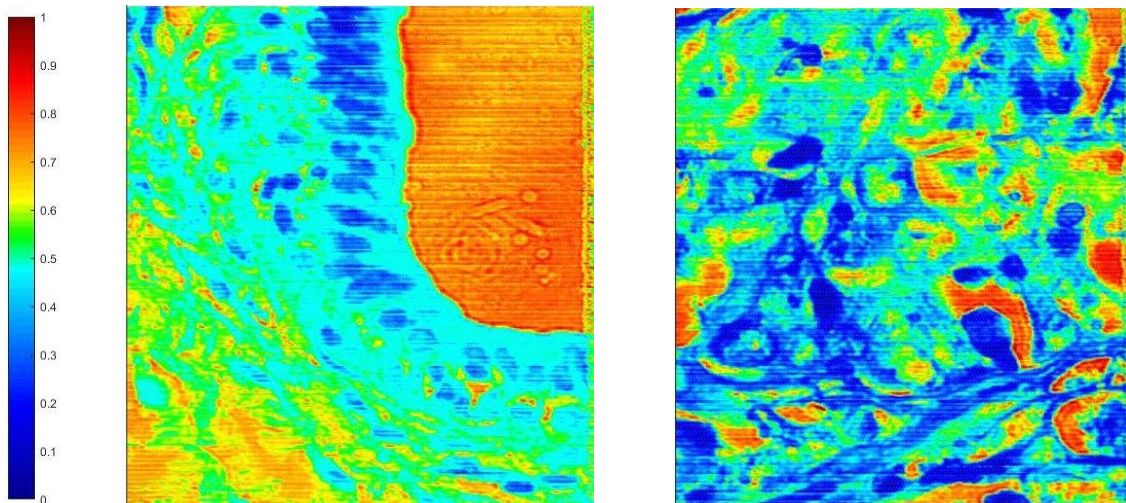
spectral reflectance of each of the 4 areas from top left to bottom right.



Average spectral reflectance of the 4 areas with standard deviation applied.

Apply Spectral Reflectance Color Filtering (SRCF) method.

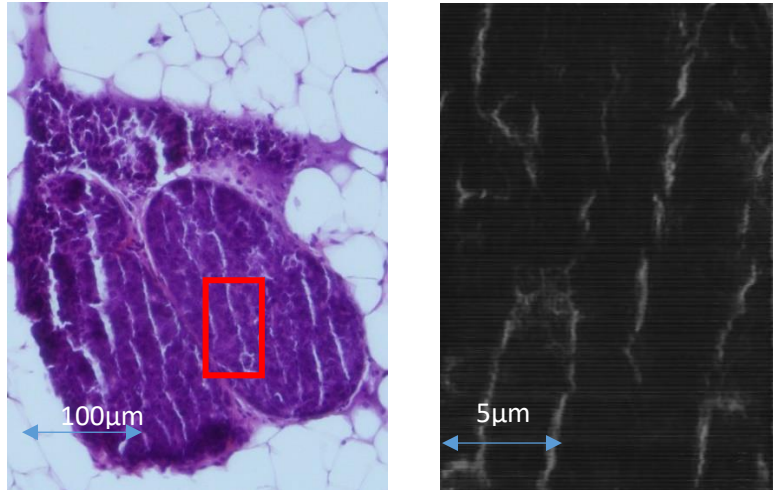
SRCF method is applied on the samples to visualize the difference in the spectral reflectance between both samples where the dark blue reflects the lowest reflectance intensity value



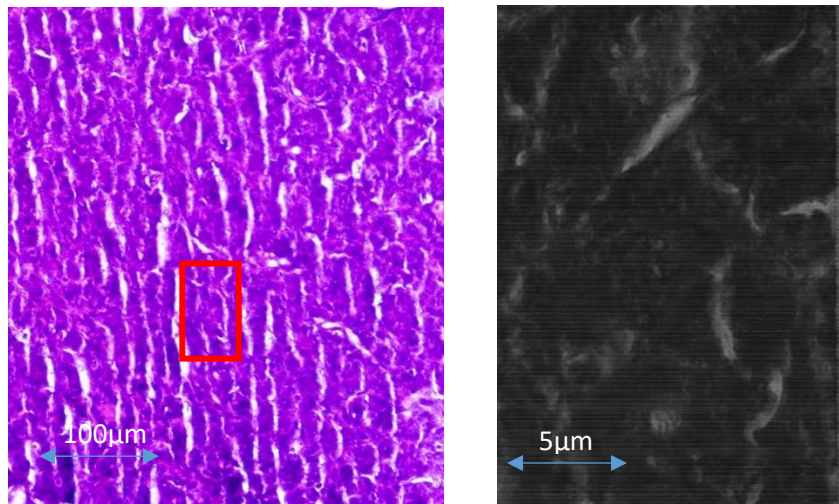
SRCF on both normal duct (Left), and ductal carcinoma (Right), and the RICF scale.



A-11: Mouse 4

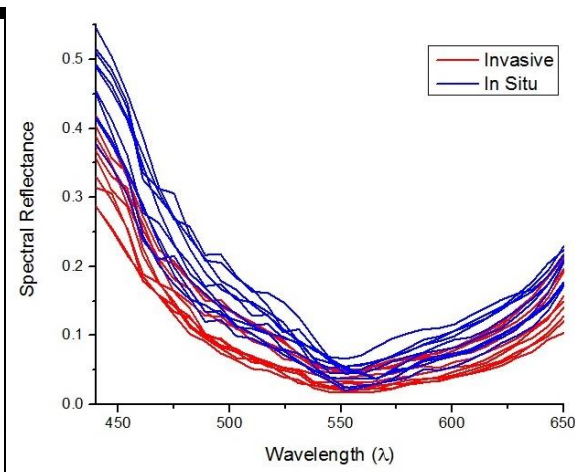
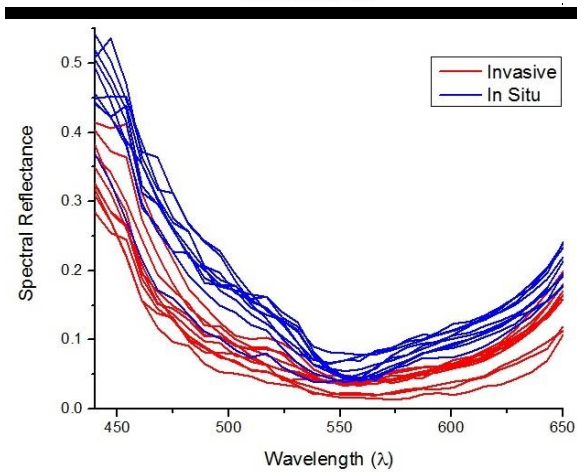
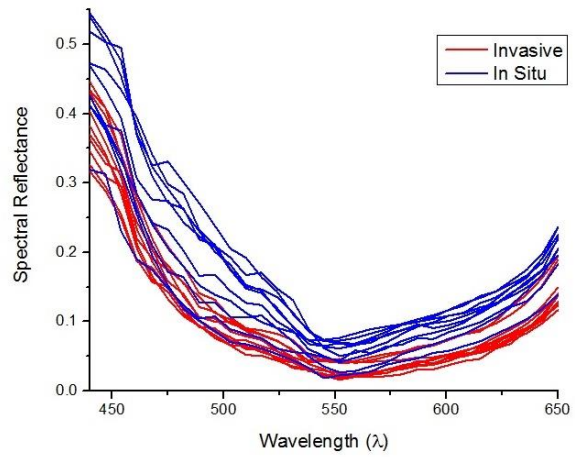
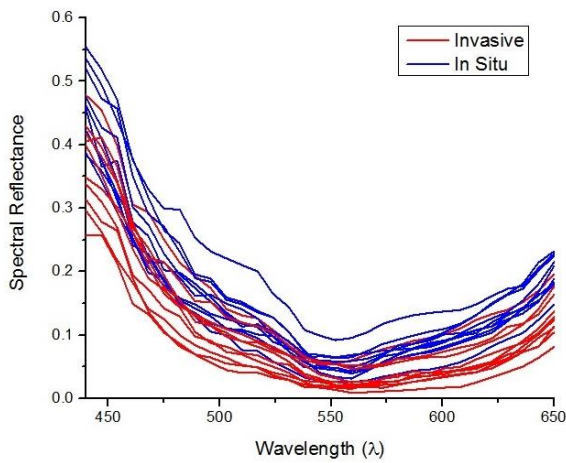
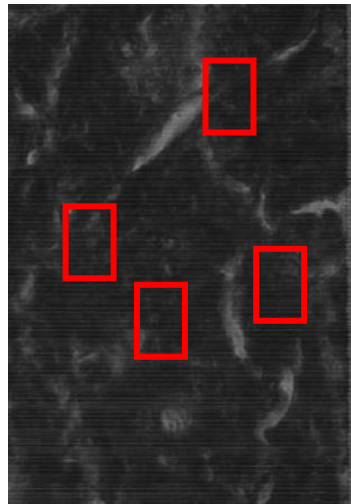
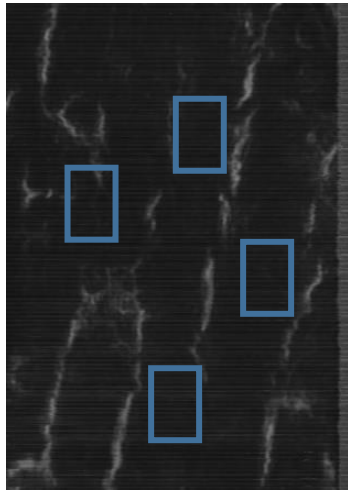


H&E stained sample of normal duct (Left), HSI (Right)

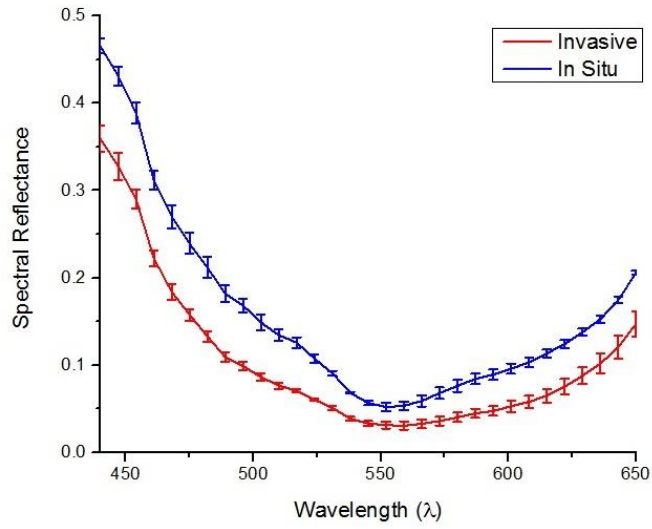


H&E stained sample of ductal Carcinoma (Left), HSI (Right)

Spectral Reflectance of the 4 marked areas:



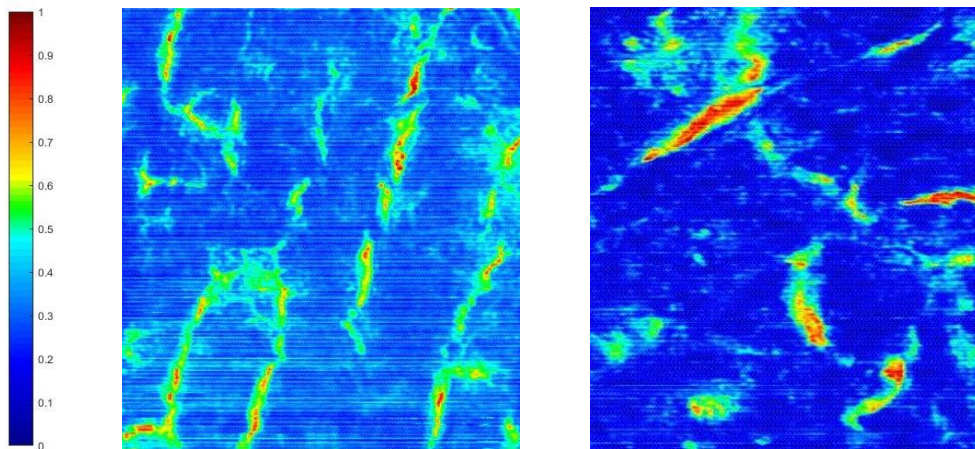
spectral reflectance of each of the 4 areas from top left to bottom right.



Average spectral reflectance of the 4 areas with standard deviation applied.

Apply Spectral Reflectance Color Filtering (SRCF) method.

SRCF method is applied on the samples to visualize the difference in the spectral reflectance between both samples where the dark blue reflects the lowest reflectance intensity value



SRCF on both DCIS (Left), and IDC (Right), and the RICF scale.

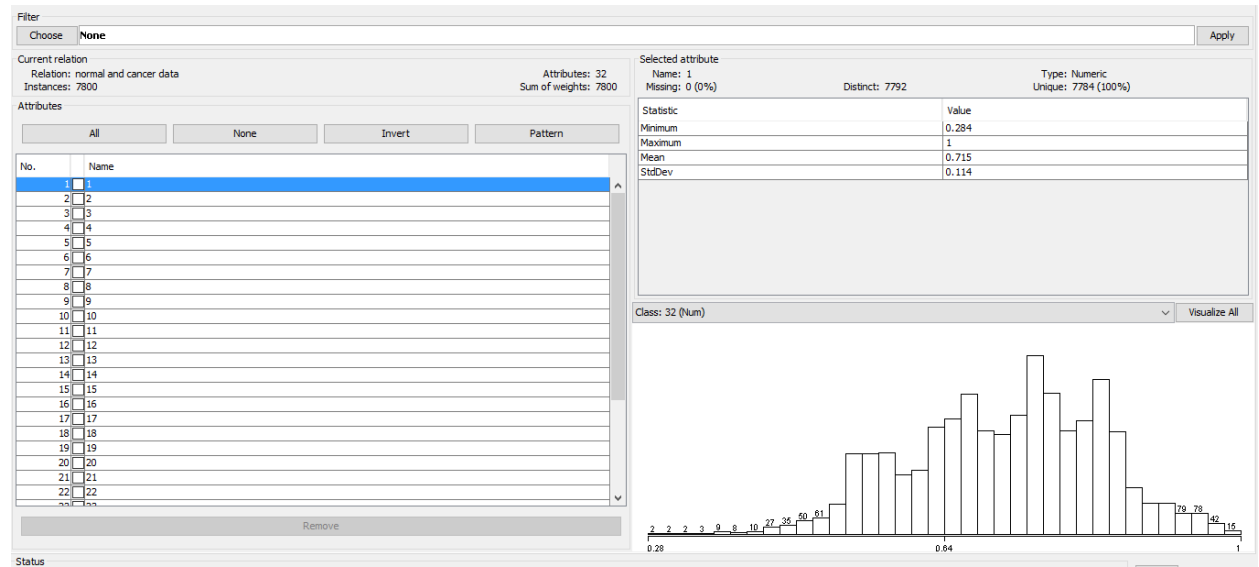
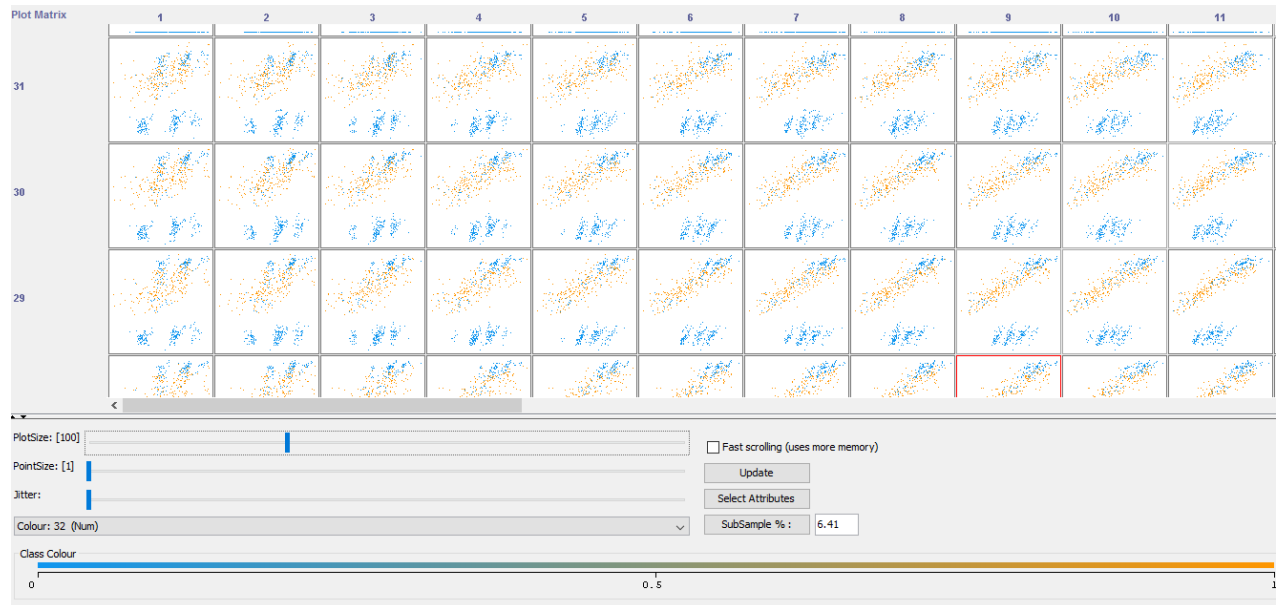
# Appendix B

Weka Software for SVM data detection:

Key_Dataset	Key_Run	Key_Fold	Key_Scheme	Key_Scheme_options	Key_Scheme_version_ID	Date_time
test4	1	1	weka.classifiers.functions.LibSVM	'-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model C:\\Users\\yasser\\Downloads -seed 1'	14172	2.02E+07
test4	1	2	weka.classifiers.functions.LibSVM	'-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model C:\\Users\\yasser\\Downloads -seed 1'	14172	2.02E+07
test4	1	3	weka.classifiers.functions.LibSVM	'-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model C:\\Users\\yasser\\Downloads -seed 1'	14172	2.02E+07
test4	1	4	weka.classifiers.functions.LibSVM	'-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model C:\\Users\\yasser\\Downloads -seed 1'	14172	2.02E+07
test4	1	5	weka.classifiers.functions.LibSVM	'-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model C:\\Users\\yasser\\Downloads -seed 1'	14172	2.02E+07
test4	1	6	weka.classifiers.functions.LibSVM	'-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model C:\\Users\\yasser\\Downloads -seed 1'	14172	2.02E+07

Number_of_traini ng_instances	Number_of_testi ng_instances	Number_c orrect	Number_i ncorrect	Number_ unclassifi ed	Percent _correct 61.5384	Percent_i ncorrect	
111	13	8	5	0	6	38.46154	
111	13	7	6	0	5	46.15385	
111	13	8	5	0	6	38.46154	
111	13	8	5	0	6	38.46154	
112	12	7	5	0	3	41.66667	
112	12	9	3	0	75	25	
Percent_u nclassified	Kappa_ statisti c	Mean_abs olute_erro r	Root_mean _squared_er ror	Relative_a bsolute_er ror	Root_relative_ squared_error	SF_prior _entrop y	SF_sche me_ent ropy
0	0.2696	0.384615	0.620174	76.87075	123.9455	13.0135	5370
0	0.1333	0.461538	0.679366	92.2449	135.7755	13.0135	6444
0	0.2696	0.384615	0.620174	76.87075	123.9455	13.0135	5370
0	0.1975	0.384615	0.620174	76.87075	123.9455	13.0135	5370
0	0.1666	0.416667	0.645497	83.33333	129.0994	12	5370
0	0.5	0.25	0.5	50	100	12	3222





# Appendix C



## Acknowledgement Letter Not Human Subject Research NHR

<b>Action Date</b>	09/03/2015
<b>To</b>	Linda Vona-Davis
<b>From</b>	WVU Office of Research Integrity and Compliance
<b>Approval Date</b>	09/03/2015
<b>Expiration Date</b>	09/02/2020
<b>Subject</b>	Not Human Subject Research Acknowledgment
<b>Protocol Number</b>	1509816662
<b>Title</b>	Spectral analysis of breast tumors

Thank you for your submission to the West Virginia University Institutional Review Board IRB.

It has been determined that your project does not meet the definition of human subject research for the following reasons:

- " In order to be considered human subject research, individually identifiable private information must be obtained or used in the research. If there is no individually identifiable private information involved, the project is not human subject research and does not require being submitted to the Office of Research Integrity & Compliance. Private information must be individually identifiable (i.e., the identity of the subject is or may be readily ascertained by the investigator or someone else associated with the information) in order to constitute research involving human subjects.

Documents reviewed and/or approved as part of this submission:

**HIPAA De-Identification Form.docx:** 2015-09-01-04:00

**Letter from IBC safety RE IRB Protocol Notice Protocol 1509816662 Returned to PL.msg:**  
2015-09-02-04:00

The Office of Research Integrity and Compliance is here to provide assistance to you from the initial submission of an IRB protocol to its approval and all subsequent activity. Please feel free to contact us by phone at 304.293.7073 with any question you may have. Thank you.

WVU Office of Research Integrity and Compliance

Date: 09/03/2015

Signed:

A handwritten signature in blue ink that reads 'Afton Wagner'.

Afton Wagner  
IRB Administrator