HYPERSPECTRAL IMAGING TO DISCERN MALIGNANT AND BENIGN CANINE MAMMARY TUMORS

A Thesis

Submitted to

The Temple University Graduate Board

In Partial Fulfillment

of the Requirement for the Degree

MASTER OF SCIENCE

in ELECTRICAL ENGINEERING

By
Amrita Sahu
August, 2013

Thesis Approval(s):

Chang-Hee Won, PhD, Thesis Adviser, Electrical and Computer Engineering Department.

Joseph Picone, PhD, Electrical and Computer Engineering Department.

Nancy Pleshko, PhD, Bioengineering Department.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ABSTRACT

Hyperspectral imaging is an emerging technology in the field of biomedical engineering which may be used as a non-invasive modality to characterize tumors. In this thesis, a hyperspectral imaging system was used to characterize canine mammary tumors of unknown histopathology (pre-surgery) and correlate the results with the post-surgical histopathology results. The system consisted of a charge coupled device (CCD) camera, a liquid crystal tunable filter in the near infrared range (650-1100 nm), and a controller. Spectral signatures of malignant and benign canine mammary tumors were extracted and analyzed. The reflectance intensities of malignant tumor spectra were generally lower than benign tumor spectra over the wavelength range 650-1100nm. Previous studies have shown that cancerous tissues have a higher hemoglobin and water content, and lower lipid concentration with respect to benign tissues. The decreased reflectance intensity observed for malignant tumors is likely due to the increased microvasculature and, therefore, higher blood content of malignant tissue relative to benign tissue. Second derivative method was applied to the reflectance spectra. Peaks at 700, 840, 900 and 970 nm were observed in the second derivative reflectance spectra. These peaks were attributed to deoxy-hemoglobin, oxy-hemoglobin, lipid and water respectively. A Tissue Optical Index (TOI) was developed that enhances contrast between malignant and benign canine tumors. This index is based on the ratio of the reflectance intensity values corresponding to the wavelengths associated with the four chromophores. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were also applied on the canine spectral dataset and the method was cross-validated. Preliminary results from 22 canine mammary tumors showed that the sensitivity and specificity of the PCA-LDA is method is 86% and 86% respectively. The sensitivity and specificity of the TOI model is 86% and 95% respectively. These results show promise in the non-invasive optical diagnosis of canine mammary cancer.

TABLE OF CONTENTS

ABSTRACT i

LIST OF FIGURES v

LIST OF TABLES ix

CHAPTER 1 INTRODUCTION 1

1.1 Motivation 1
1.2 Research Objectives 2
1.3 Organization of the Thesis 3
1.4 Contributions 3

CHAPTER 2 BACKGROUND AND LITERATURE REVIEW 5

2.1 Infrared Hyperspectral Imaging 5
2.2 Mammary Tumor Detection using Infrared Hyperspectral Imaging 7
2.2.1 Breast Cancer Detection 8
2.2.2 Canine Mammary Tumor Detection 10
2.3 Other Diseases Diagnosed By Infrared Hyperspectral Imaging 11
2.3.1 Prostate Cancer Detection 12
2.3.2 Tongue Cancer Detection 13
2.3.3 Gastric Cancer Detection 14
2.3.4 Skin Cancer Detection 16
2.3.5 Intestinal Ischemia and Lung Emphysema Detection 16

CHAPTER 3 DIFFERENTIATION OF MALIGNANT AND BENIGN TUMORS 18

3.1 Introduction 18
3.2 Tissue Optical Indices 18
3.3 Principal Component Analysis 20
3.4 Linear Discriminant Analysis 22
3.5 Normalization and Preprocessing 25
3.5.1 Data Normalization by the Radiometric Reflection Calibration Process 25
3.5.2 Other Normalization Methods 26
3.5.3 Smoothing and Differentiation 30

CHAPTER 4 EXPERIMENTS AND RESULTS 33

4.1 Hyperspectral Imaging System Description 33
4.2 Experiments to Characterize the Hyperspectral Imaging System 34
4.2.1 Testing the Repeatability of the Hyperspectral Imaging System 34
4.2.2 Depth of Penetration of NIR Light into Chicken Breast Tissue 38
4.2.3 Experiment to Demonstrate Effect on Camera to Sample Distance on the Reflectance Spectra of Chicken Tissue 41
4.3 Canine Mammary Tumor Experiment 42
4.3.1 Data Acquisition From Hospital 43
4.3.2 Extracting the Reflectance Spectra of the Tumors 44
4.3.3 Smoothing and Normalization 45
4.3.4 The Reflectance Spectra of Canine Patients 47
4.3.5 Finding the wavelength characteristic of four chromophores 61
4.3.6 Data analysis using TOI method and PCA-LDA method 62
4.3.6.1 Tissue Optical Indices Method 62
4.3.6.2 PCA-LDA model and cross validation 67

CHAPTER 5 CONCLUSIONS AND FUTURE WORK 70

5.1 Discussion 70
5.2 Conclusion 73
5.3 Future Work 74

REFERENCES 76
APPENDIX A.. 78

LIST OF FIGURES

Figure 1: Needle biopsy for breast cancer. The method is invasive and time-consuming. 2

Figure 2: A hyperspectral three dimensional imaging cube where the three coordinates represent the horizontal axis, vertical axis and the wavelength respectively 6

Figure 3: The absorption spectra of tumor and normal tissue. Tumor tissue has increased absorption compared to normal tissue 9

Figure 4: The Diffuse Optical Spectroscopy System. 10

Figure 5: Instrumentation for the multipixel imaging device. 11

Figure 6: *In-vivo* spectral signature of benign and malignant lesion. The dashed red lines represent the malignant lesion and the continuous line represents the benign regions 12

Figure 7: Detection of prostate cancer tissue (green area) using the classification method with least square support vector machine 12

Figure 8: The hyperspectral imaging setup used by the authors for the gastric cancer experiment 15

Figure 9: An illustration of PCA. a) A data set given as 3-dimensional points. b) The three orthogonal Principal Components (PCs) for the data, ordered by variance. c) The projection of the data set into the first two PCs, discarding the third one. 21

Figure 10: In LDA, the line that maximizes the separability of scalars is selected. 24

Figure 11: Flowchart showing PCA-LDA used for validation of model. Here 44-fold cross validation is used. 24

Figure 12: The Spectralon reflectance standards SRS-99 and SRS-02 26

Figure 13: The smoothed area normalized spectra of normal, benign, and malignant ROIs of all canine patients 27

Figure 14: The smoothed unit vector normalized spectra of normal, benign, and malignant spectra of all canine patients. 28

Figure 15: The smoothed mean normalized spectra of normal, benign, and malignant spectra of all canine patients.....29
Figure 16: The smoothed maximum normalized spectra of normal, benign, and malignant spectra of all canine patients. 30
Figure 17: Schematic view of the hyperspectral imaging acquisition system.. 34
Figure 18: Graph showing the raw spectral data (wavelength vs. reflectance intensity) measurement of neoprene rubber sheet over five days. 36
Figure 19: Graph showing the smoothed spectral data (wavelength vs. reflectance intensity) measurement of neoprene rubber sheet over five days. 37
Figure 20: Second derivative reflectance spectra of smoothed spectral data of neoprene rubber. 37
Figure 21: Graph showing the reflectance intensity vs. the wavelength for neoprene rubber sheet, chicken slice and varying width of chicken slices on the neoprene sheet.....40
Figure 22: The reflectance spectra of the chicken sample with the distance between the camera and the sample varying each time. 42
Figure 23: The canine patient with multiple mammary tumors held by the doctor as images are being captured by the hyperspectral system. 43
Figure 24: The digital image of canine mammary tumors. 44
Figure 25: The hyperspectral image of canine mammary tumors at 760nm wavelength 44
Figure 26: The reflectance spectra of one of the canine patients. The raw reflectance spectra have a low signal-to-noise ratio. To mitigate the effect of noise, we smoothen the signal. 45
Figure 27: The smoothed reflectance spectra of the all canine patient tumors. 46
Figure 28: The range normalized reflectance spectra of the all canine patient tumors. 46
Figure 29: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Maxine. 47
Figure 30: The spectral plot of the malignant, benign and normal tissue of canine patient Maxine. 48

Figure 31: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Chelsey 48
Figure 32: The spectral plot of the benign and normal tissue of canine patient Chelsey. 49
Figure 33: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Hanna. 50
Figure 34: The spectral plot of the benign and normal tissue of canine patient Hanna. 51
Figure 35: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Lady. 51
Figure 36: The spectral plot of the benign and malignant lesion of canine patient Lady. 52
Figure 37: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Suzanna. 53
Figure 38: The spectral plot of the benign lesion and normal tissue of canine patient Suzanna. .. 54
Figure 39: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Queenie. 55
Figure 40: The spectral plot of the malignant lesion and normal tissue of canine patient Queenie. 55
Figure 41: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Tinkerbell. 56
Figure 42: The spectral plot of the malignant and normal tissue of Tinkerbell 57
Figure 43: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Gretchen. 57
Figure 44: The spectral plot of the benign lesion and normal tissue of canine patient Gretchen .. 58
Figure 45: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Kiki. 59
Figure 46: The spectral plot of the malignant and normal tissue of canine patient Kiki 59

Figure 47: The second derivative absorption spectra of the reflectance data of all canine patients. Four negative peaks were observed at the wavelengths of 700, 840, 900 and 970 nm. 62

Figure 48: The whole tumor ROIs of one of the canine patients 64

Figure 49: Graph showing that the malignant tumors generally have a higher TOI than benign tumors. TOI threshold of 2 units was used. 66

Figure 50: Multiple ROIs (4 in each tumor) taken to increase the overall number of ROIs. 67

Figure 51: The 500W quartz tungsten halogen lamp used in the canine experiment. 71

Figure 52: The 300W Sunbeam Xenon Lightsource to which a fiber optic cable can be attached. 71

LIST OF TABLES

Table 1: Weather, Temperature and Humidity Measurement…….…....…….36

Table 2: Clinical Results of Maxine……....…….47

Table 3: Clinical Results of Chelsey……...……....…….49

Table 4: Clinical Results of Hanna.. ……...……....…….52

Table 5: Clinical Results of Lady……....……....…….53

Table 6: Clinical Results of Suzanna……...……....…….....54

Table 7: Clinical Results of Queenie……...……....…….....56

Table 8: Clinical Results of Tinkerbell……...……....…….58

Table 9: Clinical Results of Gretchen……...……....……. 58

Table 10: Clinical Results of Kiki…….....……....……... 58

Table 11: The reflectance intensity values all ROI at 700, 840, 900 and 970 nm…….60

Table 12: Tissue Optical Indices Value of the Reflectance Spectra..……....……... 64

Table 13: Cross Validation Results using PCA-LDA……...……....…….....68

Table 13: Sensitivity and specificity of TOI and PCA-LDA method applied for the canine spectral dataset compared to the results of other algorithms used for different types of cancer…….……...……....…….....72

CHAPTER 1

INTRODUCTION

1.1 Motivation

Hyperspectral imaging (HSI) in the near infrared region is a non-invasive modality used to extract spectral information from samples. This system is being used in the field of biomedical engineering and medical informatics. Hyperspectral imaging is capable of providing both the spectral and spatial information of tissues over a hundred or more spectral bands. The spectral information helps us identify the composition of the material. Hyperspectral imaging may aid a surgery and allow the assessment of tissue continuously without interrupting surgery. In this project, near infrared hyperspectral imaging will be used for canine cancer characterization.

The incidence of mammary neoplasia varies widely across the mammalian class. For example, among domestic species, dogs have the highest prevalence of mammary neoplasia, which is approximately three times higher than that of women [1]. Mammary cancer in female dogs accounts for 70% of all cancer cases [2]. Once a tumor reaches the metastatic stage, the chances of successful treatment are low. Therefore, early detection of malignant tumors is necessary for a full recovery. Currently, a surgical biopsy is the gold standard for cancer diagnosis, in dogs with mammary tumors. The standard approach involves an excisional biopsy (removing the entire tumor) and the subsequent examination by a pathologist. It is an invasive, time-consuming, and expensive procedure that requires highly trained surgeons and pathologists. Moreover, in a biopsy, the whole tumor is typically not analyzed; only a few representative sections are evaluated, which may leave room for sampling error. To avoid these disadvantages, we investigated the use of a hyperspectral imaging system for the characterization of canine mammary tumors.

HSI has been used to detect prostatic, gastric and tongue cancer [3] [4] [5]. HSI was also used by Medina and co-workers for the characterization of iris reflectance to non-invasively diagnose ocular diseases [6] and by Panasyuk and co-workers for the detection of tumor and normal breast tissues [7]. This imaging technology has also been used for the detection and analysis of intestinal ischemia [8], for detection and characterization of emphysema from healthy lung tissue [9] and for *in-vivo* detection and grading of cervical pre-cancers and of pigmented skin lesions [10]. All these experiments were conducted for research purposes.

.

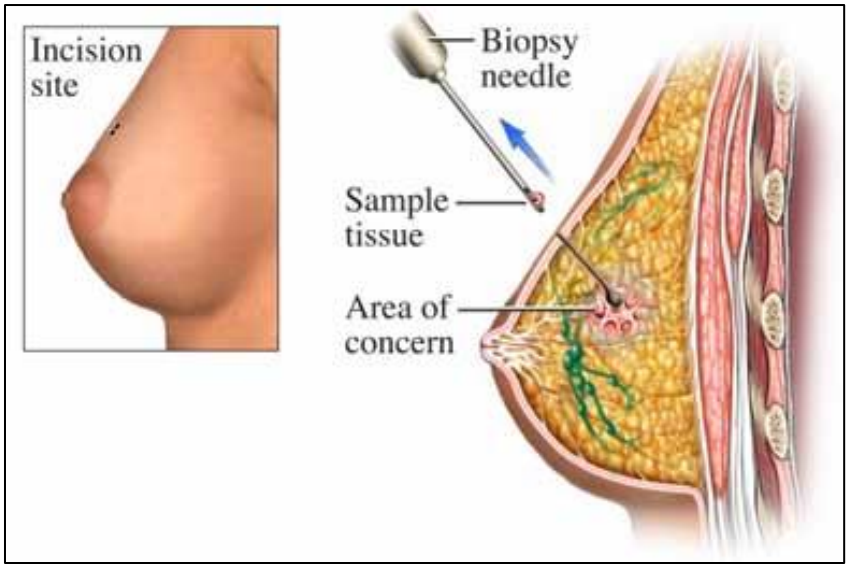


Figure 1: Needle biopsy for breast cancer. The method is invasive and time-consuming [16].

1.2 Research Objectives

The research objectives of this thesis are:

a) Characterizing a hyperspectral imaging system and using it for malignant tumor identification.

b) Normalize and preprocess the spectral data.

c) Develop an algorithm to discern the malignant tumors from the benign tumors.

d) Design an experiment to acquire the hyperspectral images of the canine mammary tumors, and analyze the results.

1.3 Organization of the Thesis

This thesis is organized as follows. In Chapter 2, background and literature reviews on infrared hyperspectral imaging of tissues are given in order to lay the foundation for this work. Cancer detection using hyperspectral imaging is also introduced. The method of breast, canine mammary tumor, gastric, prostrate, skin and tongue cancer characterization using near-infrared hyperspectral imaging is described. In Chapter 3, the theoretical approach used to analyze the hyperspectral canine data is discussed. Also, various normalization and preprocessing methods that would be applied on the canine spectral data are detailed. Chapter 4 is broadly divided into 3 parts. Section 4.1 describes the hyperspectral imaging system, Section 4.2 discusses the experiments performed to characterize the hyperspectral system, and the results obtained . Section 4.2 details the canine mammary tumor acquisition followed by normalization and preprocessing of the spectral data and then application and validation of the TOI model and the PCA-LDA model on the canine spectral data. The sensitivity and specificity of the proposed model is also given. Chapter 5 consists of a discussion of various issues faced during the project, followed by the concluding remarks and a plan for future work.

1.4 Contributions

The contributions of this thesis are:

a) Use of a hyperspectral imaging system to characterize malignant and benign canine mammary tumors. The method is non-invasive and less time-consuming.

b) Identification of four wavelengths that characterize four important tissue chromophores; deoxyhemoglobin, oxyhemoglobin, lipid and water. Second derivative method was applied on the reflectance spectra to identify the characteristic wavelengths. This can

significantly decrease the time required for imaging the sample; since now images at 4 wavelengths are required to be taken compared to conventional procedure of capturing hyperspectral images from 650 nm to 1100 nm at an increment of 10 nm.

c) A Tissue Optical Indices (TOI) method was used to discern malignant and benign canine tumors. Preliminary results from 22 canine mammary tumors showed that the sensitivity and specificity of the TOI model is 86% and 95% respectively

d) Principal Component Analysis method and Linear –Discriminant Method was used to discern malignant and benign tumors. Preliminary results from 22 canine mammary tumors showed that the sensitivity and specificity of the PCA-LDA is method is 86% and 86% respectively.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

2.1 Infrared Hyperspectral Imaging

Conventional color cameras acquire intensity information from the visible electromagnetic spectrum, that is, red, green and blue. Hyperspectral imaging measures and collects intensity information over more than one hundred spectral bands across the electromagnetic spectrum. Hyperspectral images are acquired by 'hyperspectral sensors' that have a variety of applications in the fields of agriculture, physics, remote sensing and surveillance. This technique was used originally for the purpose of remote sensing, a field pioneered by NASA [15]. But recently, hyperspectral imaging is also being used in the field of medicine, particularly cancer detection. This is because this kind of novel technology can quantify several biomarkers found in tissue, and thereby can be used for non-invasive tissue analysis.

During hyperspectral imaging, the data produced by the sensor consists of a collection of images. The set of images can be represented by a three-dimensional cube of image (Figure 2), where the first two coordinates represent the spatial coordinate of a pixel and the third coordinate gives the wavelength of a particular spectral band. Thus it gives both the spectral and spatial information at the same time.

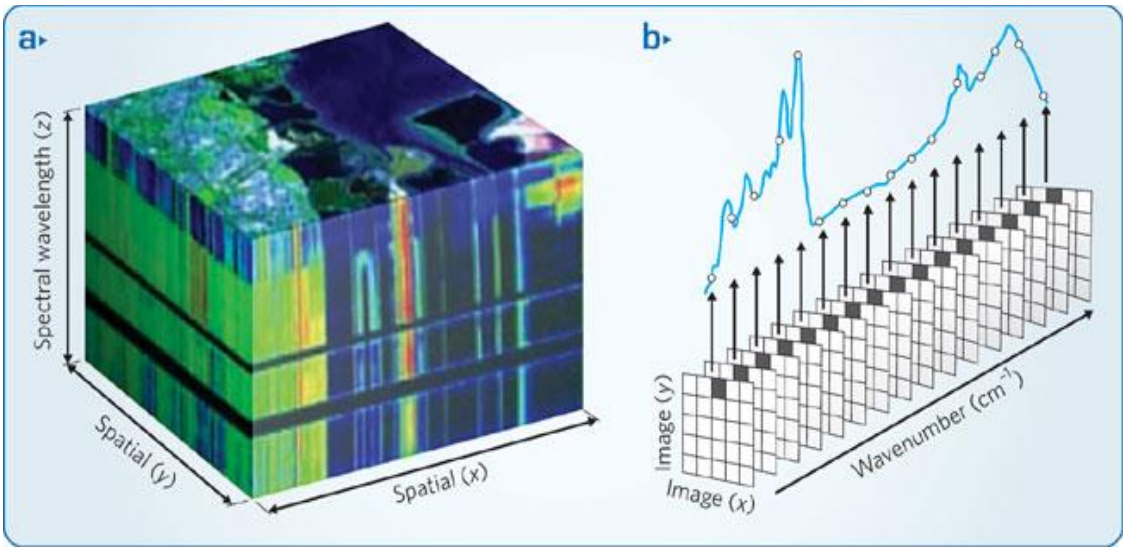


Figure 2: A hyperspectral three dimensional imaging cube where the three coordinates represent the horizontal axis, vertical axis and the wavelength respectively [17].

There are three different types of hyperspectral imaging techniques. The first type uses a prism or a grating as the dispersive medium in the optical system, the second type uses band pass filter, which can be tunable or fixed, and the third type implements a technique called the Fourier transform spectroscopy. In this technique, Fourier transform is applied to obtain the spectral decomposition of the light entering the hyperspectral sensor. In this project, the second class of hyperspectral imaging technique will be used. Our hyperspectral camera consists of a liquid crystal tunable filter, which scans the near-infrared spectrum (650-1100 nm) in increment of 10 nm.

In agriculture, hyperspectral imaging is used to monitor the health of crops and to detect chemical composition, nutrient and water status and disease outbreak for plants. Hyperspectral imaging is used for obtaining the spectral signatures of geological samples like feldspar, silica, calcite groups to differentiate the minerals of interest from the surrounding region. Some minerals like gold and diamond are also identified from hyperspectral airborne images. Hyperspectral surveillance is used for military purposes, and it is able to detect the presence of opposition even if countermeasures are taken to avoid airborne surveillance. Nowadays, hyperspectral cameras

that incorporate thermal infrared imaging are used for this purpose. Chemical Imaging is another application for which hyperspectral imaging is used. At war, soldiers are exposed to chemical agents and harmful plumes which are difficult to detect. But using hyperspectral imaging, detection and identification of such harmful chemicals have now become possible. This technology is also used in many countries for the continuous monitoring of oil drills. If oil leaks from an oil drill, it could cause extensive damage to the ecosystem. So hyperspectral sensors can be deployed to build detection and warning system for oil drills.

In this project, we will focus on the use of hyperspectral imaging for non-invasive canine cancer detection. Near-infrared hyperspectral imaging has been used for the detection of various kinds of cancer; breast, gastric, prostate, tongue. In this project, we focus on the near-infrared region (NIR) that extends from 650 nm to 1100 nm. This is because NIR light is absorbed by certain chromophores that are biochemically significant, namely, hemoglobin, water and lipid. Thus using NIR spectroscopy, we can quantify the concentration of these important biomarkers. Literature shows that NIR light penetrates into tissue farther than light in any other spectrum, because tissue has low absorptivity in this region [12]. Thus weakly absorbed NIR light that can penetrate several centimeters is used for the hyperspectral imaging of thick tissue such as brain, breast and muscle [12]. Literature shows some favorable results in the detection of breast cancer using NIR spectroscopy [12]. Since a canine mammary tumor is physiologically similar to human breast tumor, we use NIR spectral range (650-1100 nm) as the light transmission range for our project.

2.2 Mammary Tumor Detection using Infrared Hyperspectral Imaging

In this section, breast cancer and canine cancer characterization using near-infrared spectroscopy are discussed.

2.2.1 Breast Cancer Detection

Diffuse optical spectroscopy using the near-infrared light spectrum has been investigated for the detection and clinical management of breast cancer [11] [12]. In these papers, instrumentation consists of a broadband steady-state spectrometer system (650-1100 nm) and a six wavelength frequency domain instrument which consists of six laser diodes (661, 686, 786, 808, 822, and 852 nm). A fiber optic cable delivers the laser light to the tissue. A hand held probe which consists of an avalanche photodiode detector, records the diffused light signals from the tissue. Although this system is different from the hyperspectral imaging system used in this thesis, the principle is the same; both investigate non-invasive tissue analysis using laser scanners and NIR light.

Currently, the most common non-invasive procedure for breast cancer screening is mammography. However this method suffers from low sensitivity and specificity and does not work well for dense breast masses. Because of these disadvantages, many research groups have tried to use optical methods for breast cancer screening. Optical methods are non-invasive, less time-consuming, do not use ionizing radiation and is less expensive. In these papers, the authors have used NIR laser diodes to enhance physiological and biochemical contrast between the malignant and normal breast tissues.

Paper by Cerussi [12] presents the clinical results from 58 malignant breast tumors. Physiological properties like hemoglobin, water and lipid content was significantly different between malignant and normal breast tumors. Malignant tumors showed higher concentration of deoxy-hemoglobin, oxy-hemoglobin and water and lower contents of tissue lipid compared to normal tissue. A tissue optical index was defined which increased the contrast between the malignant and normal tissue:-

$$TOI = \frac{[H_2O][HbT]}{[Lipid][StO_2]} \quad (1)$$

Higher total hemoglobin content [HbT] indicates higher tissue blood volume and angiogenesis. Higher water content [H_2O] suggests abnormal accumulation of interstitial fluid in the tissue. Decreased tissue oxygen saturation which is the ratio of oxygenated hemoglobin to total hemoglobin [StO_2] indicats tissue hypoxia (a pathological condition in which tissue is deprived of adequate oxygen supply) and was caused by metabolically active tumor cells. A decreased lipid content [Lipid] suggests that the parenchymal adipose have been displaced. These changes can be grouped together to enhance contrast through the formation of the TOI, where elevated TOI values suggest high metabolic activity and malignancy [12]. Figure 3 shows the absorption spectra of normal and tumor tissue.

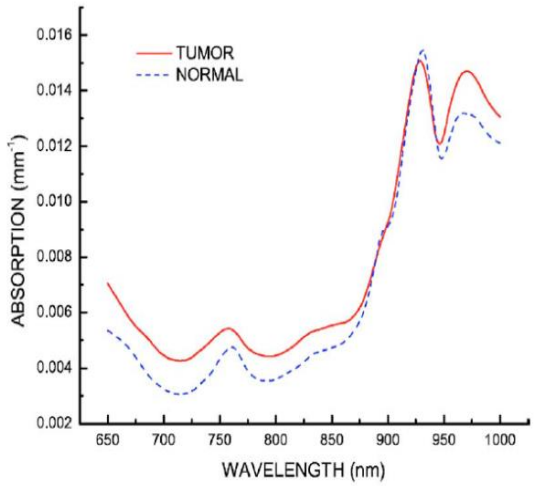


Figure 3: The absorption spectra of tumor and normal tissue. Tumor tissue has increased absorption compared to normal tissue [12].

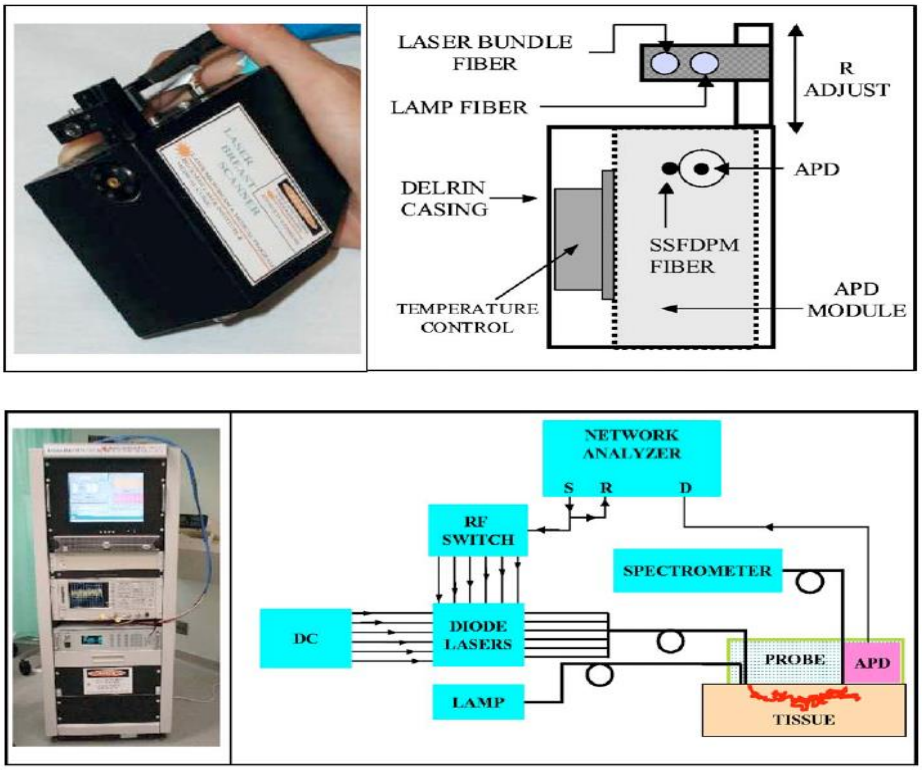


Figure 4: The Diffuse Optical Spectroscopy System [12].

Also, an important application of diffuse optical spectroscopy is monitoring the response of patients to cancer therapies. One of the cancer therapy methods for women in advanced stages of breast cancer is pre-surgical chemotherapy, which can shrink tumors to a smaller size, so that they can be operated and surgically removed. In order to maximize the therapeutic effectiveness, the above proposed non-invasive optical method can be used to examine and assess the physiological properties of the tumor. Figure 4 shows the Diffuse Optical Spectroscopy System.

2.2.2 Canine Mammary Tumor Detection

In a paper by Gurfinkel [13], fluorescent dyes have been used to discern canine adenocarcinoma from normal tissue. Two fluorescent dyes: indocyanine green (ICG) and carotene-conjugated 2-devinyl-2-(1-hexyloxyethyl) pyropheophorbide (HPPH-car) were administered intravenously in a canine patient. For illumination of the tissue, a 660 nm laser diode beam was used. The

fluorescence generated by the dyes upon propagating to the tissue surface, is captured by a CCD camera. The uptake and release rates of the dye were studied and they varied considerably in the diseased and normal tissue. A pharmacokinetic model was used to monitor the delivery of the dyes. The need to administer a fluorescent dye renders this procedure mildly invasive, whereas the experiment presented in this thesis was completely non-invasive. Figure 5 shows the instrumentation for the multipixel imaging device.

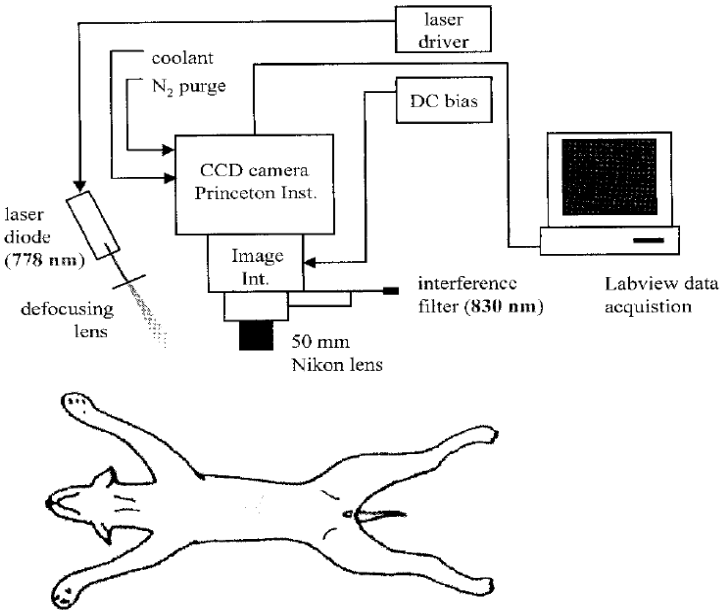


Figure 5: Instrumentation for the multipixel imaging device [13].

2.3 Other Diseases Diagnosed By Infrared Hyperspectral Imaging

Literature shows that there are various diseases that have been diagnosed non-invasively by near-infrared hyperspectral imaging. In this section, detection of prostate, gastric, tongue and skin cancer, lung emphysema, and intestinal ischemia using NIR hyperpsectral imaging is discussed.

2.3.1 Prostate Cancer Detection

A paper by Akbari *et al.* [4] describes a non-invasive detection method of prostate cancer cell in mice using hyperspectral image processing and machine learning techniques. Using least square support vector machines, the authors classify the hyperspectral images of the cancerous and normal tissue. They also indicate the differences of reflectance properties in cancer and normal tissue using spatially resolved images. The experiment was carried out on diseased mice (*in-vivo*) as well as on pathological slides (*in-vitro*). Figure 7 shows the detection of prostate cancer tissue using the classification method with least square support vector machine. Figure 6 shows the *in-vivo* spectral signature of benign and malignant lesions.

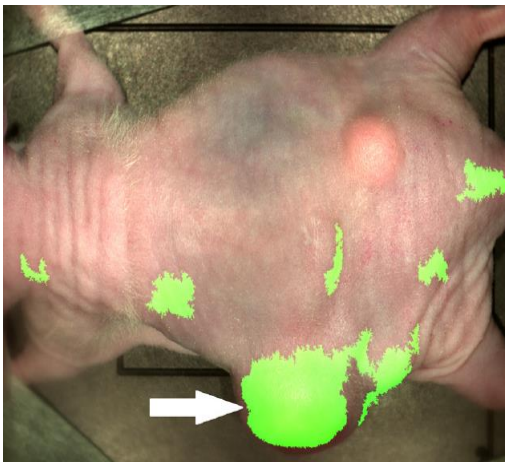


Figure 7: Detection of prostate cancer tissue (green area) using the classification method with least square support vector machine [4].

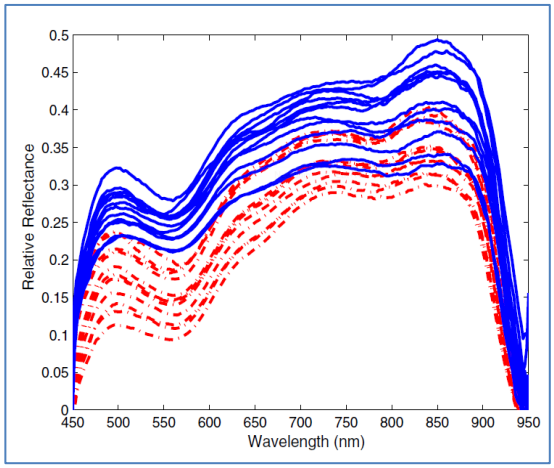


Figure 6: *In-vivo* spectral signature of benign and malignant lesion. The dashed red lines represent the malignant lesion and the continuous line represents the benign regions [4].

To remove the spectral non-uniformity of the illumination device, the authors used a normalization technique. The technique minimized the influence of dark current. To solve the classification problem, the authors employed a least square support vector machine in which a linear equation is solved (instead of quadratic programming) in a high dimensional space or kernel space. In any least square method, the sum of the squared error is minimized.

The authors evaluate the classification technique using two parameters: sensitivity and specificity. Sensitivity measures the percentage of correctly identified tumor tissue and specificity measures the percentage of healthy tissue correctly identified. The results show the specificity of the classification technique is 92.8% (with a margin of 2.0%) and sensitivity of 96.9 % (with a margin of 1.3%) for 11 mice [4].

The method described above is superior to conventional methods like tissue biopsy because it measures spectra of tissue point by point. It also spectrally and spatially verifies the spectral variations of different types of tissue. Therefore, it is a continuous non-intervening evaluation technique for suspicious cancer tissues and can be used as a virtual biopsy [4].

2.3.2 Tongue Cancer Detection

Liu [5] describes tongue cancer detection using hyperspectral imaging. The reflectance spectra dataset of 34 tongue tumors were collected using an acousto-optic tunable filter (AOTF) and a spectral adapter. For medical applications, most hyperspectral systems use tunable filters rather than spatial scanning, because they do not need mechanically moving parts and are faster. AOTF is a solid state tunable filter and is based on the principle of acoustic diffraction of light in an anisotropic medium. They are advantageous over other spectrometers, because they are not sensitive to mechanical shock or vibration. AOTF technology also provides a wider tuning range, more field of view, and can be easily programmed. For illumination purposes, the authors have used 500W halogen lamps. Noise is present in the tumor spectral data because of the saliva on the tongue and its instinctive squirming. For denoising, the authors used median filters. To remove the spectral non-uniformity, a normalization procedure has been used using standard reflectance boards.

Sparse representation (SR) has been used as the classification algorithm to detect the cancerous tongue tumors. Sparse signal representation has been proven to be an extremely powerful

computer vision tool for representing high dimensional signals [5]. This is because some classes of signals have naturally occurring sparse properties. The authors have first shown the difference in the spectral properties of the tumor and normal tissue. The SR classifier was used to develop an algorithm to detect the cancerous pixels from the non-cancerous pixels. The authors achieved a recognition rate of 96.5%.

To prove that the SR algorithm is computationally superior to traditional hyperspectral algorithms like support vector machine (SVM) or relevance vector machine (RVM), the authors have compared the performance of the SR algorithm to that of the SVM and RVM. The SR method worked better than the other two even when the tumors were more than 3mm deep and was covered with mucosa. The authors have also shown that the classification time for their method is less than SVM and RVM. The false negative rate (FNR) and the false positive rate (FPR) were calculated for each hyperspectral image. The SR method showed a FNR and FPR of 6.3% and 8.7% respectively [5].

Thus this method shows promise in the detection of tongue cancer using hyperspectral imaging and sparse representation. This computer aided tool can help the doctors decide non-invasively as to whether a patient has tongue cancer.

2.3.3 Gastric Cancer Detection

A paper by Akbari [4] describes gastric cancer detection using infrared hyperspectral imaging. Gastric tumors were imaged from 10 human subjects and the results were correlated with the histopathology results. A total of 101 images were taken from 10 patients. Each sample was imaged 10 times, to ensure repeatability of the system. The NIR spectral range of 1000-2500 nm was used for this project. The hyperspectral imaging device was a tunable optical band pass filter, and the spectral resolution of the camera was 6.29 nm. The light source for illumination consisted of two 150W halogen lamps, with diffuse reflectors. In this experimental setup, the camera was

attached to a frame and the frame was fixed (Figure 8). The frame was controlled by a linear actuator. This made sure that the distance between the lens and the sample is constant. This is a very important factor that should be kept in mind while designing our canine experiment.

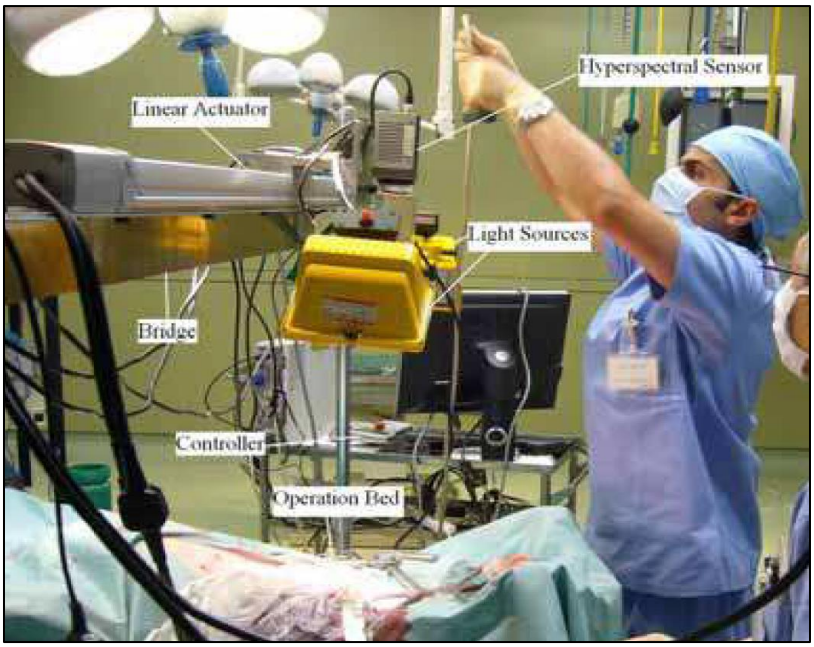


Figure 8: The hyperspectral imaging setup used by the authors for the gastric cancer experiment [4].

In this paper, the authors have normalized the data using standard reflectance boards. The noise was removed by median filtering. The differences in the spectral characteristics of the tumor and the normal tissue have been shown. The standard deviation of the spectral signature of the tumors is higher compared to the normal tissue. Four different classification methods were used to detect the cancerous tumors. The four classification methods were a support vector machine (SVM) based method, a standard deviation (SD) method, an integral method and the normalized difference cancer index (NDCI) method. Out of these four algorithms, NDCI gave the best results. Using this method, the authors obtained a sensitivity and specificity of 93% and 91% respectively [4].

The above proposed method shows potential in the non-invasive detection of gastric cancer. This method could detect tumors less than 0.5 mm in size [4]. Also, determining the tumor bed during surgery is also challenging for the doctors. Such imaging technology can be used to determine the tumor bed during surgery and also to ensure that the entire tumor has been removed after tumor resection. This it can help in the segmentation of tumors and remove ambiguity between the malignant and benign regions.

2.3.4 Skin Cancer Detection

Skin cancer is a common form of cancer, with more than a million people affected annually by this form of cancer in the United States. Dermatologists identify skin melanomas, but a biopsy is required for definitive evaluation. A biopsy is not feasible for a patient with a large number of melanomas. Therefore, a non-invasive method to detect skin cancer in place of the traditional biopsy is in great demand.

Balas [10] describes a method of skin cancer detection using Raman spectroscopy. In this paper, the authors develop a clinical research prototype that can obtain Raman spectra from skin *in vivo* within 1 second. The imaging instrument consists of a diode laser (785 nm), a spectrograph, a CCD detector and Raman optical probe. The probe was designed in such a way that the collection of Raman signals is maximum. In order to process the raw spectral data, a polynomial fitting algorithm was used to remove the background fluorescence. Data from 274 skin lesion were analyzed which included cancers, precancerous tumors and benign lesions. Using Raman spectroscopy, the authors obtained sensitivity and specificity of 90% and 75% respectively.

2.3.5 Intestinal Ischemia and Lung Emphysema Detection

Hyperspectral imaging has also been used to detect diseases such as intestinal ischemia and lung emphysema [8] [9]. Paper by Akbari [8] describes the detection of intestinal ischemia using hyperspectral imaging both in the visible and the invisible wavelength range. Intestinal Ischemia

is caused by the insufficient flow of oxygenated blood to the intestine, which may result in bloody diarrhea, gangrene and other infection. The spectral signatures of the intestinal ischemic region of a pig's intestine were extracted using two hyperspectral imagers (400-1000 nm and 900-1700 nm). Two filters were designed as the classification algorithm and the results were compared with that of the traditional support vector machine (SVM) method. One of the filter classification methods could detect ischemic tissue from normal intestinal tissue and other organs with a FPR (False Positive Rate) and FNR (False Negative Rate) of 1% and 1% respectively. The SVM method gave a FPR and FNR of 5% and 6% respectively.

A paper by Jong-Ha Lee [9] describes the characterization of emphysema from healthy tissue using hyperspectral imaging. The NIR spectral range of 650nm to 1100nm was used. Peak absorption intensity was observed at four wavelengths (760, 805, 915, and 970 nm), which is characteristic of the four tissue chromophores. Spectral signatures from the normal pig's lung phantom and the smoker's lung phantom were extracted and compared. The reflectance intensity of the smoker's lung tissue was considerably higher than the normal lung tissue over the entire wavelength range (650-1100nm). This information could be useful for detecting lung emphysema from healthy lung tissue.

CHAPTER 3

DIFFERENTIATION OF MALIGNANT AND BENIGN TUMORS

3.1 Introduction

In this chapter, we will be discussing the various algorithms that will be applied on the canine spectral data set in order to differentiate the malignant and benign spectra. Also, normalization and preprocessing techniques will also be described.

3.2 Tissue Optical Indices

Some tissue optical indices (TOI) have been established that correlate with the pathology of the breast tissue [11].To differentiate malignant and benign breast lesions, tissue optical indices have previously been defined as follows:

$$TOI = \frac{[H_2O][HbT]}{[Lipid][StO_2]} \tag{1}$$

Higher total hemoglobin content [HbT] indicates higher tissue blood volume and angiogenesis; higher water content [H_2O] suggests abnormal accumulation of interstitial fluid in the tissue; decreased tissue oxygen saturation, which is the ratio of oxygenated hemoglobin to total hemoglobin [StO_2], indicated tissue hypoxia (a pathological condition in which tissue is deprived of adequate oxygen supply) and was caused by metabolically active tumor cells; decreased lipid content [Lipid] suggests that the parenchymal adipose have been displaced. These changes can be grouped together to enhance contrast through the formation of the TOI, where elevated TOI values suggest high metabolic activity and malignancy [12].

In this project, we do not have the concentration information of the chromophores. So we will reframe equation 1 in terms of reflectance intensity. Here we will be establishing a relationship between the concentration of the tissue components and their corresponding reflectance intensity values.

The dependence of the chromophore concentration, *c*, on the absorbance of light is,

$$A = \varepsilon[c]l \tag{2}$$

where A is the absorption of NIR light, ε is the molar extinction coefficient (mol /litre/cm), [c] is the concentration of chromophore (mol/litre), and l is the photon path length (cm). Here the molar extinction coefficient is compound and wavelength dependent. So we are assuming it to be a constant for a particular chromophore and a particular wavelength. Also, the path length l is dependent on the scattering properties of the tissue and is not known *a priori*. So for this calculation, we assume it to be constant.Thus, we can say that, absorbance, A, is directly proportional to the concentration of the chromophore.

$$A \propto [c] \tag{3}$$

Absorbance, A, is related to reflectance intensity [arbitrary unit] as follows:

$$A = \log_{10} \frac{1}{R} \tag{4}$$

Now let us say there are two chromophores 1 and 2. The concentrations of the two chormophores are $[c_1]$ and $[c_2]$. The absorption and the reflectance intensity values are A_1, A_2 and R_1, R_2 respectively. The total concentration of the chromophore is the sum of the individual concentrations. Let the total concentration of the chromophore be c_t and the total absorbance and the total reflectance intensity be A_t and R_t respectively.

$$[c_t] = [c_1] + [c_2] \tag{5}$$

Since, $A \propto [c]$, we can write that

$$[A_t] = [A_1] + [A_2] \tag{6}$$

$$\log_{10} \frac{1}{R_t} = \log_{10} \frac{1}{R_1} + \log_{10} \frac{1}{R_2} \tag{7}$$

$$\log_{10} \frac{1}{R_t} = \log_{10} \frac{1}{R_1 R_2} \tag{8}$$

$$\frac{1}{R_t} = \frac{1}{R_1 R_2} \tag{9}$$

From the definition of Tissue Optical Indices from above we have,

$$TOI = \frac{A_{H_2O} \cdot A_{HbT}}{A_{lipid} \cdot A_{StO_2}} \tag{10}$$

$$TOI = \frac{\log_{10}(1/R_{H_2O}) \log_{10}(1/(R_{H_2O}\, R_{Hb}))}{\log_{10}(1/R_{lipid})(\log_{10}(1/R_{HbO}))/\log_{10}(1/(R_{HbO} R_{HbO}))} \tag{11}$$

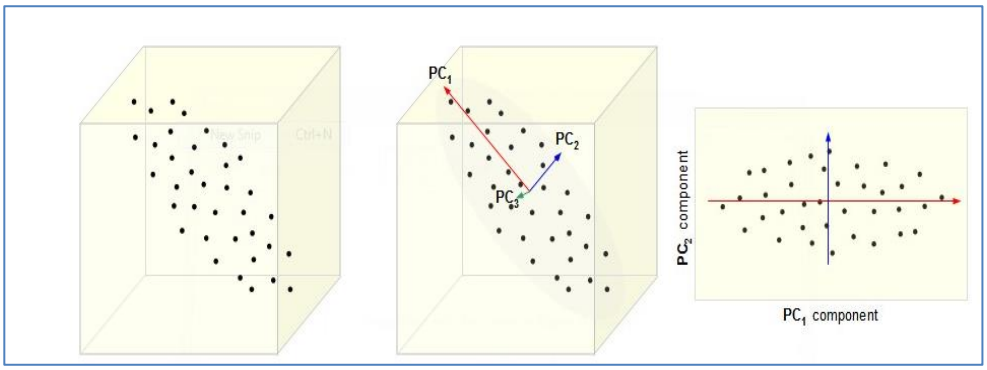
Equation 11 was used to calculate the Tissue Optical Index value, which will be further discussed in Chapter 4.

3.3 Principal Component Analysis

Principal Component Analysis (PCA) is a well-known multivariate statistical technique, that converts a larger number of observed variables into a smaller number of variable (called principal components), that will take in account most of the variance in the observed variables. Number of

original variables is more than or equal to the number of principal components. In this technique, the first principal component has the highest variance; the second principal component has the second highest variance and so on. But each principal component should be orthogonal to the one before that.

PCA helps us to identify pattern in data and emphasize on the similarities and differences of the dataset. When the data had a large number of variables, patterns in the dataset are difficult to be recognized. So PCA is a useful tool for analyzing large dimensional data (Figure 9). In this project, we will use PCA on the raw canine data, and also the normalized data to see if we get separate clusters of the benign, malignant and normal Region of Interests (ROI). The filter in our imaging system captures an image at an increment of 10 nm. So we have 46 images from 650 nm to 1100 nm at an interval of 10 nm. So our data has 46 dimensions. Thus it is difficult to visualize and classify the spectral data. Applying PCA would give a principal scores plot, which would be two dimensional. Clusters in the Principal Scores plot would indicate spectral similarity and would enable us to classify the tumors as malignant or benign. We will be using the PCA plot tool in the Unscrambler 10.1 for this purpose.



(a) **(b)** **(c)**

Figure 9: An illustration of PCA. a) A data set given as 3-dimensional points. b) The three orthogonal Principal Components (PCs) for the data, ordered by variance. c) The projection of the data set into the first two PCs, discarding the third one. [18].

3.4 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a classification method that provides a linear transformation of n-dimensional feature vectors into an m-dimensional space ($m < n$). LDA is a supervised classification method, as the categories to which objects are to be classified is known before the model is created. The objective of LDA is to determine the best fit parameters for classification of samples by developing a model. The model can then be used to classify unknown samples.

LDA is a classification method that is based on Bayes' formula. From Bayes' rule one develops a classification model assuming the probability distribution within all groups is known, and that the prior probabilities for groups are given, and sum to 100% over all groups. It is based on the normal distribution assumption and the assumption that the covariance matrices of the two (or more) groups are identical. This means that the variability within each group has the same structure. The only difference between groups is that they have different centers. LDA considers both within-group variance and between-group variance. The estimated covariance matrix for

LDA is obtained by pooling covariance matrices across groups. When the variability of each group does not have the same structure (unequal covariance matrix), the shape of the curve separating groups is not linear, and therefore quadratic discriminant analysis will provide a better classification model [19].

LDA is used for classifying objects (samples, people, foods, etc.) into groups based on features that can be used to describe the objects. This could include developing classifications models for a library of products, good vs. bad quality product, or healthy vs. cancerous cells. A typical example related to classifying objects or, more generally, recognizing patterns is not a simple task for automated procedures, particularly when the objects are of biological interest. For example, identifying species, predicting species distributions or finding gene expression patterns that

predict the risk of developing a particular type of tumor are generally difficult tasks. Data can be different analytical techniques related to chromatographic hyphenated techniques, like liquid chromatography with diode array detection (LC-DAD), where a set of UV-Vis spectra are used for classification. Data from any type of measurement, including spectroscopic data, imaging data, or generic data such as a table of physical properties of samples, can be used for classification, if those measurements have features which describe the objects. But for an LDA to be a well-posed problem, the number of objects in the calibration set should be larger than the number of variables

In order to overcome the constraint of requiring more objects than features, we may use PCA-LDA which reduces the data dimensionality using PCA prior to running LDA. The number of components would still need to be less than the number of objects in each class. PCA-LDA makes use of a common projection space for all the classes.

To explain the concept of LDA from a more mathematical point of view, let us assume that we have a N dimensional sample $\{x_1, x_2, x_3, ..., x_N\}$, N_1 of which belongs to class C_1 and N_2 of which belongs to class C_2. . LDA seeks to reduce dimensionality while preserving as much class discriminatory information as possible. We obtain a scalar *y* by projecting the samples *x* onto a line

$$y = w^T x \tag{12}$$

Of all the possible lines we would like to select the one that maximizes the separability of the scalars (Figure 10).

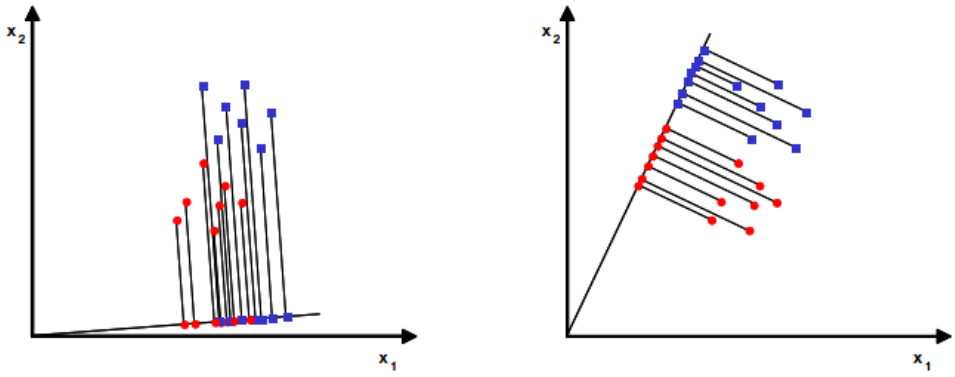


Figure 10: In LDA, the line that maximizes the separability of scalars is selected.

In this thesis, we intend to use PCA-LDA model on the canine dataset. Since LDA is a supervised classifier, it requires that the input vectors are independent. So as input, we use the principal component scores. To build the model, we need to supply the information if the tumor is cancerous or not (this information we would get from histopathology results). Then we would use *k- fold* cross validation to test the PCA-LDA model.

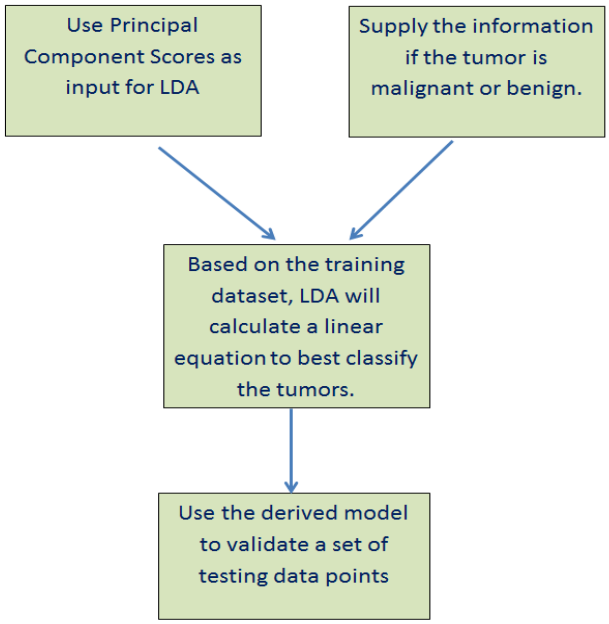


Figure 11: Flowchart showing PCA-LDA used for validation of model. Here 44-fold cross validation is used.

3.5 Normalization and Preprocessing

This section describes the normalization and preprocessing techniques that will be investigated. In the first section, data normalization using standard reflectance boards will be described. This is the standard normalization procedure followed by most research groups working on near infrared hyperspectral imaging. However, these standards were not used during our canine experiment. There are a number of normalization techniques which are available in 'The Unscrambler 10.1'.These techniques are also widely used for normalizing NIR data. Taking first and second derivative of the spectral data is a way of preprocessing the data to correct the baseline effects. Derivative methods and their applications are also discussed in this section.

3.5.1 Data Normalization by the Radiometric Reflection Calibration Process

The captured data should be normalized to treat the spectral non-uniformity of the illumination device. The raw data changes due to several factors like illumination conditions, temperature and non-uniform contour of the subject. To minimize these varying conditions, spectral data has to be normalized. The radiometric reflectance normalization process involves a pixel-to-pixel normalization of the hyperspectral image data to percentage reflectance. This is the most common approach for radiometric calibration and has been widely used [3] [4] [5]. White reference and dark current are two data that should be captured for normalization. White reference is the spectrum acquired by the hyperspectral sensor by placing the white reference board in the field of view. For this purpose, Spectralon diffuse reflectance standard SRS-99 for an approximately 99% reflectance (Labsphere Inc., Sutton, NH) should be used. The dark current is measured by taking a dark image in the absence of light. For this purpose, Spectralon diffuse reflectance standard SRS-02 for an approximately 2% reflectance (Labsphere Inc., Sutton, NH) should be used [3]. White reflectance is used to show the maximum reflectance in each wavelength. Dark current is used to address the defects by calculating the peaks in the dark current spectrum with temperature. Then the raw data is corrected using the following equation:

$$R(\lambda) = \frac{I_{raw}(\lambda) - I_{dark}(\lambda)}{I_{white}(\lambda) - I_{dark}(\lambda)} \quad (13)$$

Where $R(\lambda)$calculated reflectance value for each wavelength is, $I_{raw}(\lambda)$ is the reflectance intensity of the raw data, and $I_{dark}(\lambda)$ and $I_{white}(\lambda)$ are the reflectance intensity of the 2% reflectance standard board and the 99% reflectance standard board, respectively. Figure 12 shows the Spectralon reflectance standards SRS-99 and SRS-02**.**

Figure 12: The Spectralon reflectance standards SRS-99 and SRS-02.

3.5.2 Other Normalization Methods

Normalization is a family of transformations that are computed sample-wise. Its purpose is to scale samples in order to get all data on approximately the same scale. The following normalization methods are also used in normalizing NIR spectral data [14]:

Area Normalization

This transformation normalizes a spectrum by calculating the area under the curve for the observation. The area under the curve becomes the same for all samples. It attempts to correct the transmission spectra for indeterminate path length when there is no way of measuring it. Figure 13 shows the area normalized spectra of the normal, benign and malignant ROIs of all canine patients.

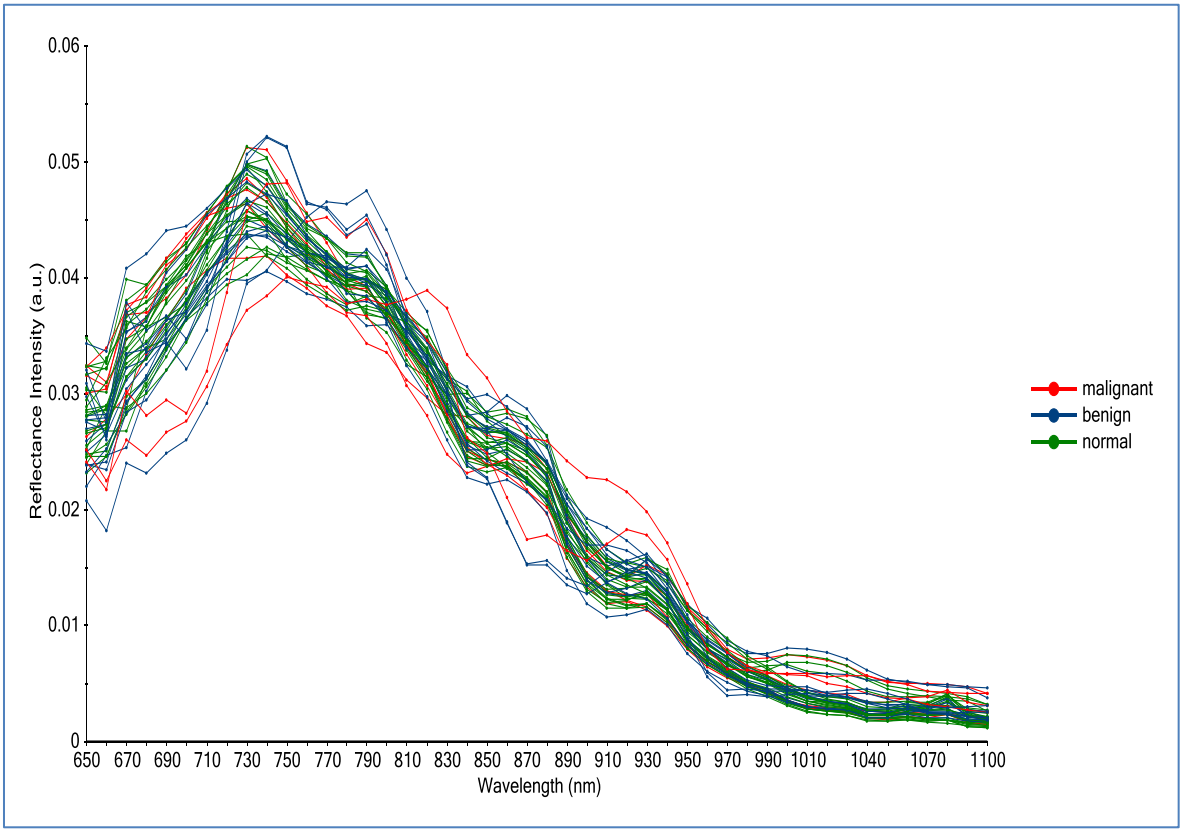


Figure 13: The smoothed area normalized spectra of normal, benign, and malignant ROIs of all canine patients.

Unit Vector Normalization

This transformation normalizes sample-wise data to unit vectors. It is useful for pattern recognition applications. The normalized samples have a length (norm) of 1. Figure 14 shows the smoothed unit vector normalized spectra of the normal, benign and malignant ROIs of all canine patients.

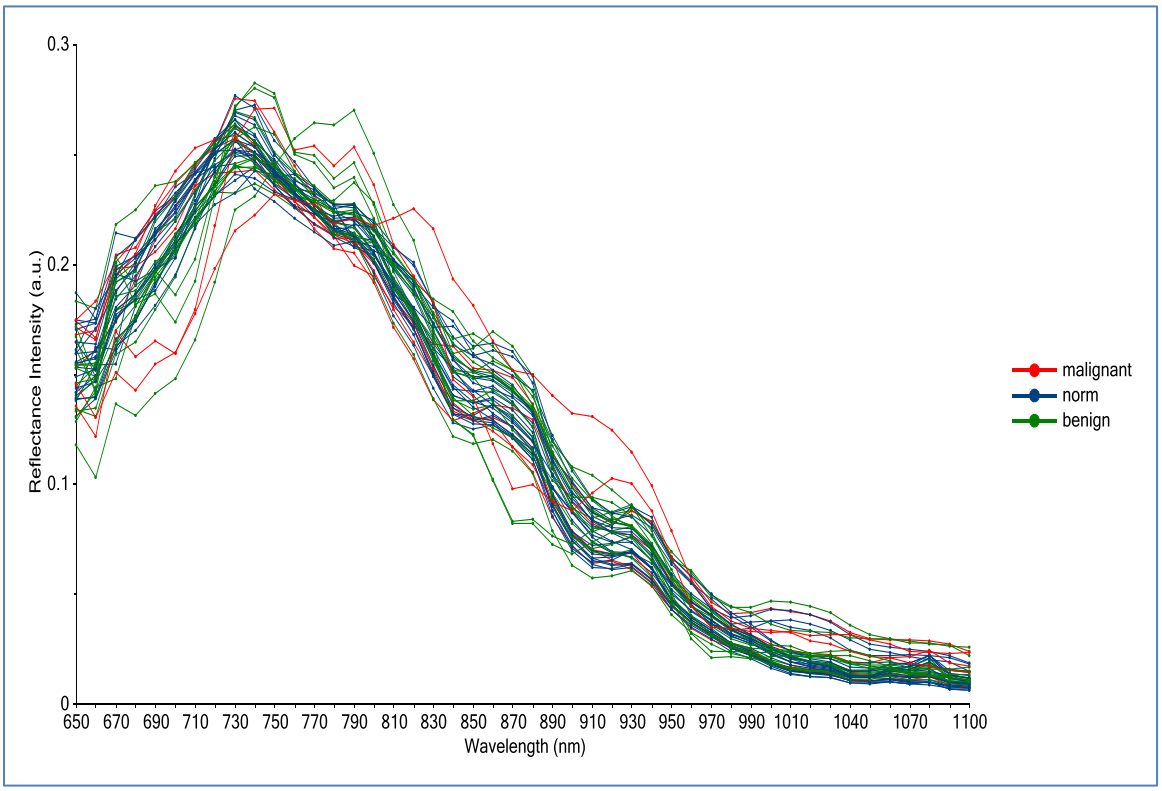


Figure 14: The smoothed unit vector normalized spectra of normal, benign, and malignant spectra of all canine patients.

Mean Normalization

This is the most classical case of normalization. It consists in dividing each row (each observation) of a data matrix by its average. This transformation is used in chromatography to express the results in the same units for all samples, no matter which volume was used for each of

them. Figure 15 shows the mean normalized spectra of the normal, benign and malignant ROIs of all canine patients.

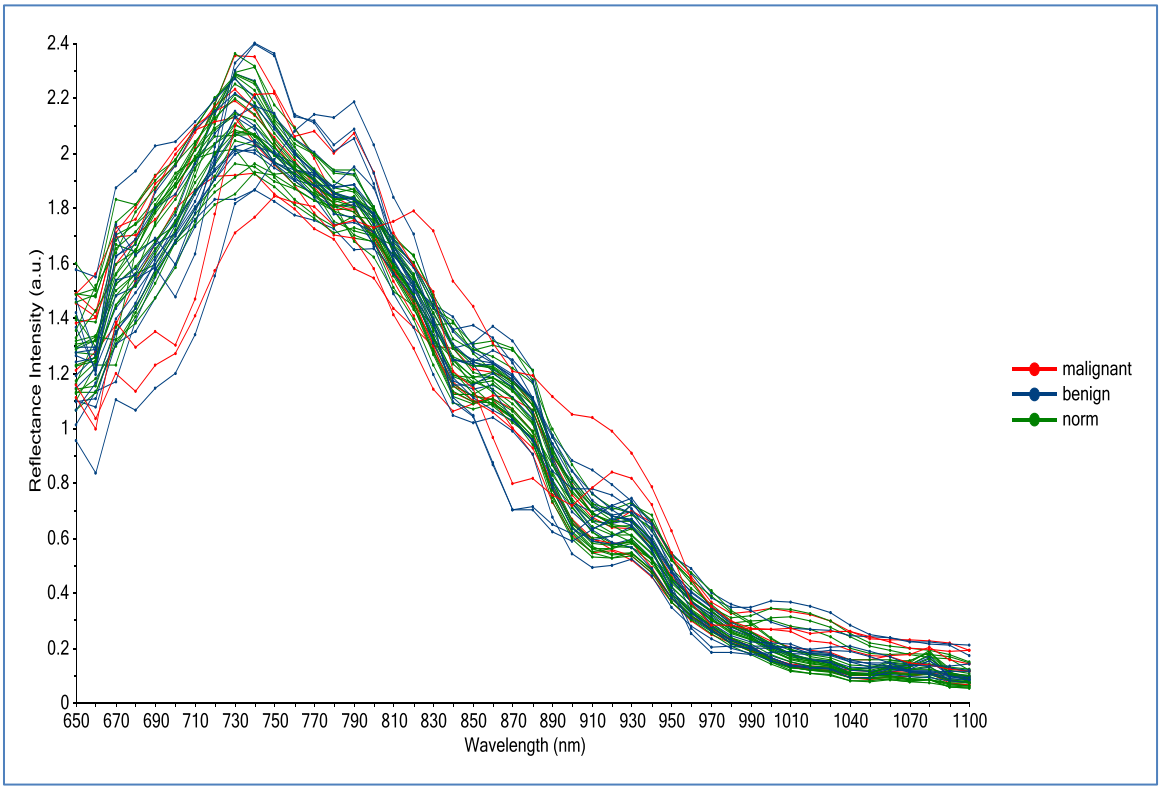


Figure 15: The smoothed mean normalized spectra of normal, benign, and malignant spectra of all canine patients.

Maximum Normalization

This is an alternative to classical normalization which divides each row by its maximum absolute value. If all values are positive the maximum value becomes +1. If all values are negative the minimum value becomes -1. If the sign of the values changes over the curve either the maximum value becomes +1 or the minimum value becomes -1. Figure 16 shows the maximum normalized spectra of the normal, benign and malignant ROIs of all canine patients.

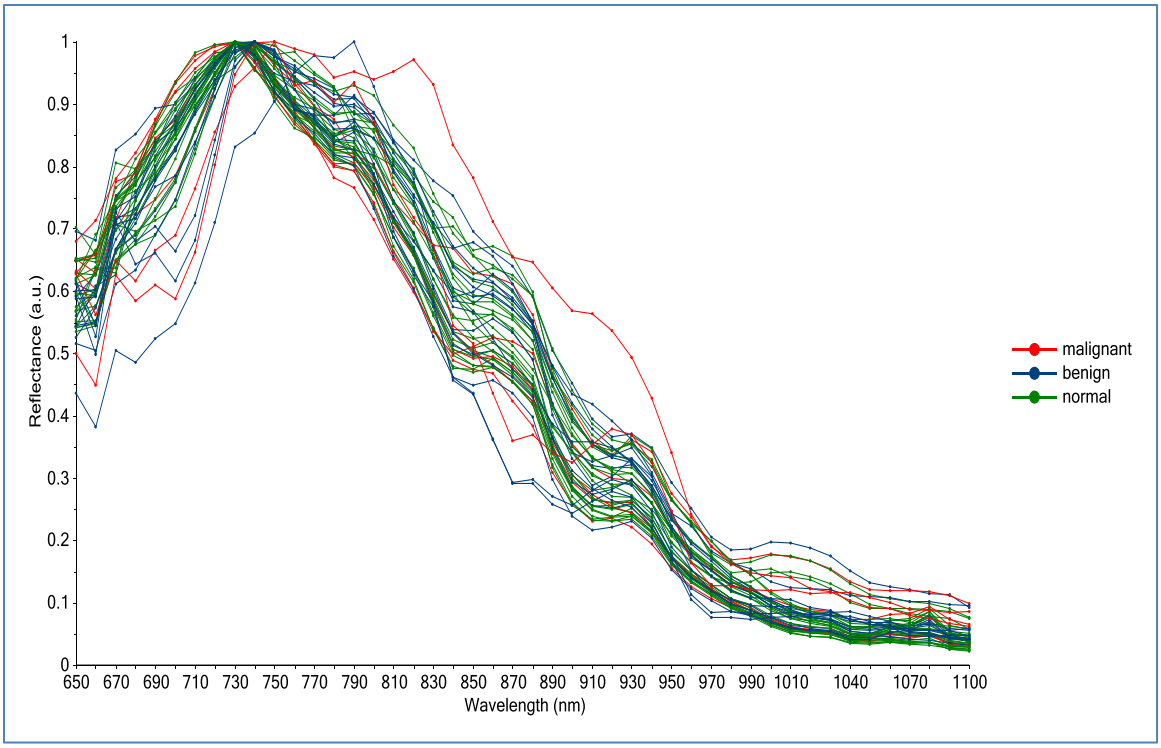


Figure 16: The smoothed maximum normalized spectra of normal, benign, and malignant spectra of all canine patients.

Range Normalization

Here each row is divided by its range, i.e. "max value – min value".

In Chapter 4, we will investigate which normalization method works the best for the canine dataset.

3.5.3 Smoothing and Differentiation

Smoothing of the reflectance spectra is done to minimize random noise and improve the signal to noise ratio. Removal of noise should be done without unduly degrading the underlying information in the signal. Savitzky-Golay smoothing filter determines the smoothed value at each point, on a series of values which are equally spaced, by performing a local polynomial regression.

Savitzky-Golay smoothing process uses the method of least squares. A set of points are to be fitted to some curves, for example, the curve

$$a_3x^3 + a_2x^2 + a_1x + a_0 = y \tag{14}$$

The a's are to be selected such that when each abscissa point is substituted into this equation, the square of the differences between the computed numbers and the observed numbers is a minimum for the total of the observations used in determining the coefficients.

This main advantage of the Savitzky-Golay method has a good balance between smoothing the spectra and preserving the shape of the spectra

Derivative method is a preprocessing tool and it is applied to the reflectance spectra. Derivatives are applied to correct the baseline effects in spectra for the purpose of removing nonchemical effects and creating robust calibration models. Derivatives may also aid in resolving overlapped bands.

First derivative

The first derivative of a spectrum is simply a measure of the slope of the spectral curve at every point. The slope of the curve is not affected by baseline offsets in the spectrum, and thus the first derivative is a very effective method for removing baseline offsets. However, peaks in raw spectra usually become zero-crossing points in first derivative spectra, which can be difficult to interpret.

Second derivative

The second derivative is a measure of the change in the slope of the curve. In addition to ignoring the offset, it is not affected by any linear "tilt" that may exist in the data, and is therefore a very effective method for removing both the baseline offset and slope from a spectrum. The second derivative can help resolve nearby peaks and sharpen spectral features. Peaks in raw spectra change sign and turn to negative peaks with lobes on either side in the second derivative. Also,

we can identify the wavelengths characteristic of deoxy-hemoglobin, oxy-hemoglobin, lipid and water respectively by applying second derivative method on the reflectance spectra of the canine spectral dataset. The negative peaks obtained in the second derivative curve would give the wavelengths corresponding to the chromophores.

CHAPTER 4

EXPERIMENTS AND RESULTS

In this chapter, the hyperspectral imaging system used in this project will be described. Experiments to characterize the hyperspectral system such as depth of penetration, repeatability of the hyperspectral system will be described and the experimental results will be discussed. Also, the method of acquiring hyperspectral images of the canine mammary tumors from the canine patients will be described. Method of analysis of those images and application of the TOI model and the PCA-LDA model will be detailed.

4.1 Hyperspectral Imaging System Description

To capture the hyperspectral imaging data, a portable hyperspectral tunable imaging system was used. It consists of digital imager (Qimaging Inc., Surrey, British Columbia), a Liquid Crystal Tunable Filter (LCTF Cambridge Research & Instrumentation Inc., Massachusetts), and LCTF controller. The digital imager was a 1.4 megapixel, mono-cooled CCD (charged coupled device) camera; with a cell size of 6.45 μm x 6.45μm and 12 bit output. The filter is placed in front of the digital imager and has a transmission wavelength range between 650 to 1100nm with a 10nm increment. Care should be taken that the light falls as uniformly as possible on the subject.

The LCTF controller synchronizes between digital imager and LCTF. It also varies the programmed sequential bands of filter. The tuning speed of the filter is between 50 ms to 150 ms. During acquisition of the images, dual 500W quartz tungsten halogen lamps have been used for illumination. An Apple Macbook Pro laptop computer is used for the spectral image acquisition purposes. The image is visualized and the spectral signature is plotted using ENVI software (Ver. 4.5, ITT Visual information solutions, Boulder, CO) and 'The Unscrambler' version 10.1 (CAMO Software AS, Oslo, Norway) . Figure 17 shows the proposed HSI system setup.

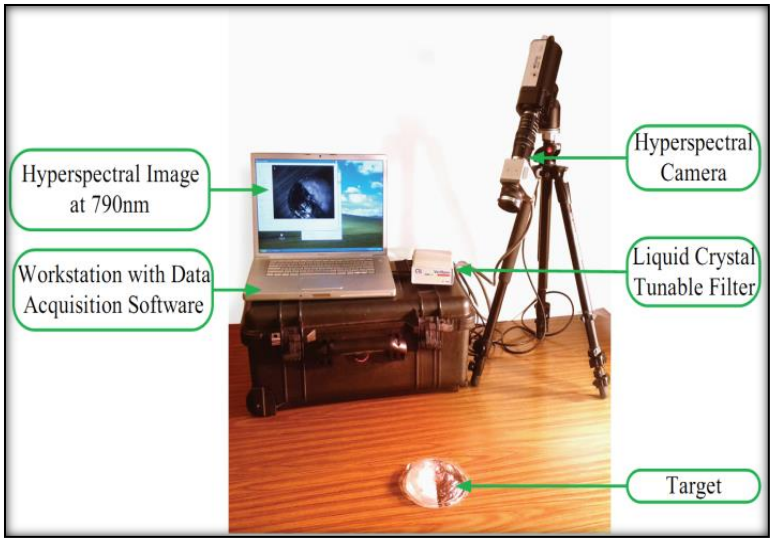


Figure 17: Schematic view of the hyperspectral imaging acquisition system. The LCTF and the computer are connected by a firewire (1394) cable. The camera and the controller are connected by a USB cable. To connect the LCTF controller and the camera, a parallel port cable has been used.

4.2 Experiments to Characterize the Hyperspectral Imaging System

In this section, experiments for the purpose of characterizing the hyperspectral imaging system will be conducted. First, the repeatability of the hyperspectral system was tested. Then, an experiment was conducted to quantify the depth of penetration of NIR light into the chicken breast tissue. Third, an experiment was carried out to test whether the lens to sample distance has any effect on the reflectance spectra of neoprene rubber.

4.2.1 Testing the Repeatability of the Hyperspectral Imaging System

Objective: To investigate the repeatability of the hyperspectal imaging system

Materials Required:

- The hyperspectral imaging system which consists of 1.4 megapixel, 12 bit digital imager (Qimaging Inc., Surrey, British Columbia), liquid crystal tunable filter

(LCTF, Cambridge Research and Instrumentation Inc., Woburn, Massachusetts), and LCTF controller.

- Dual 500W white quartz tungsten halogen lamps for illumination
- A laptop computer (Apple Macbook Pro, Cupertino, CA)
- A firewire IEEE 1394 cable to connect the computer to the LCTF controller
- A parallel port cable to connect the LCTF controller to the camera
- A USB cable to connect the camera to the computer
- Neoprene rubber sheet
- Temperature and humidity meter

Procedure:

Repeatability of a hyperspectral imaging system should be assessed. Hyperspectral images captured over consecutive days of the same target should be identical or nearly identical. However, there might be differences in the spectral properties of the same target taken over consecutive days. These variations are due to: ambient temperature and humidity, lighting conditions and inherent instrumentation error.

The hyperspectral image of a neoprene rubber sheet was captured for 5 consecutive days. The ambient temperature and the humidity were recorded. The distance between the lens and the sample was kept constant each time. External conditions such as lighting were kept as similar as possible. The reflectance spectra of the neoprene rubber captured over the 5 days were compared.

The following was the temperature and humidity measured over the 5 days:

Table 1: Weather, Temperature and Humidity Measurement

Day	Temperature	Humidity	Weather conditions
1	22.9 ˚C	28%	Cloudy
2	22.7 ˚C	29%	Cloudy
3	23.6 ˚C	26%	Sunny
4	23.8 ˚C	25%	Sunny
5	24.1˚C	25%	Sunny

Results and Discussion:

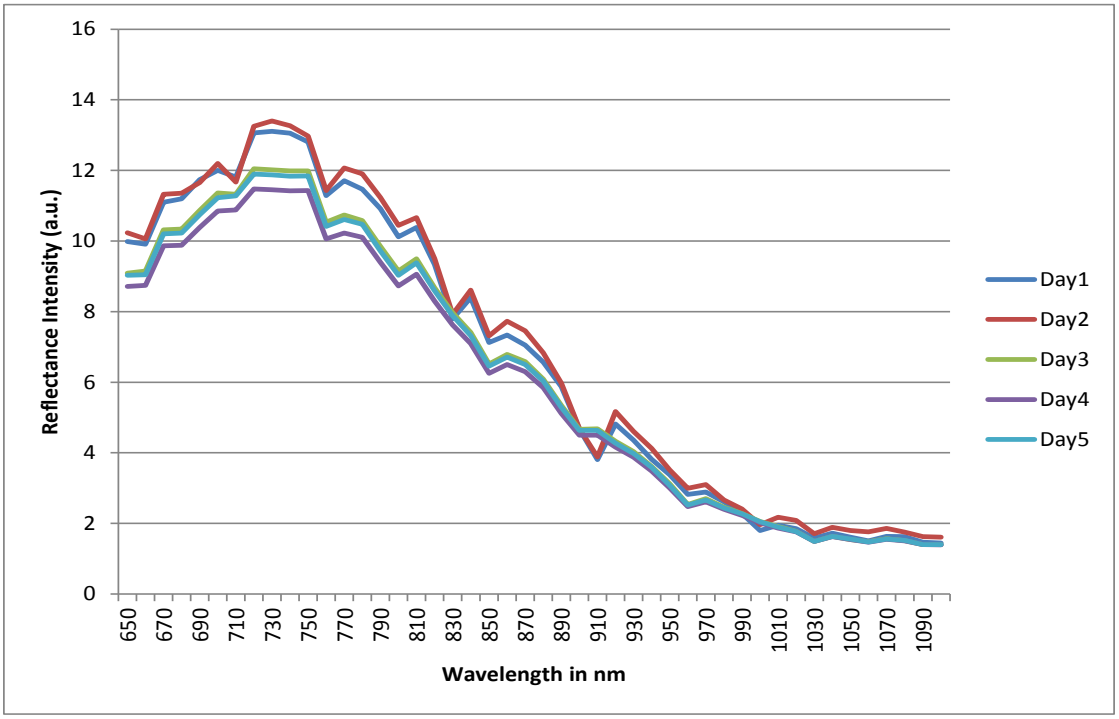


Figure 18: Graph showing the raw spectral data (wavelength vs. reflectance intensity) measurement of neoprene rubber sheet over five days.

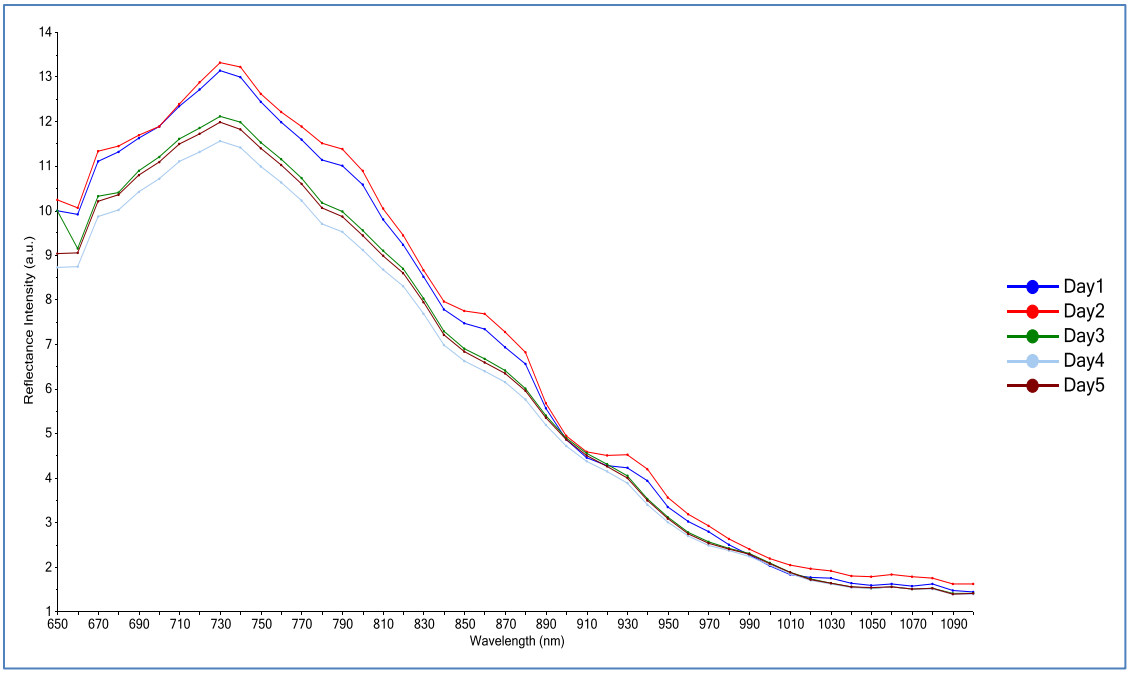


Figure 19: Graph showing the smoothed spectral data (wavelength vs. reflectance intensity) measurement of neoprene rubber sheet over five days.

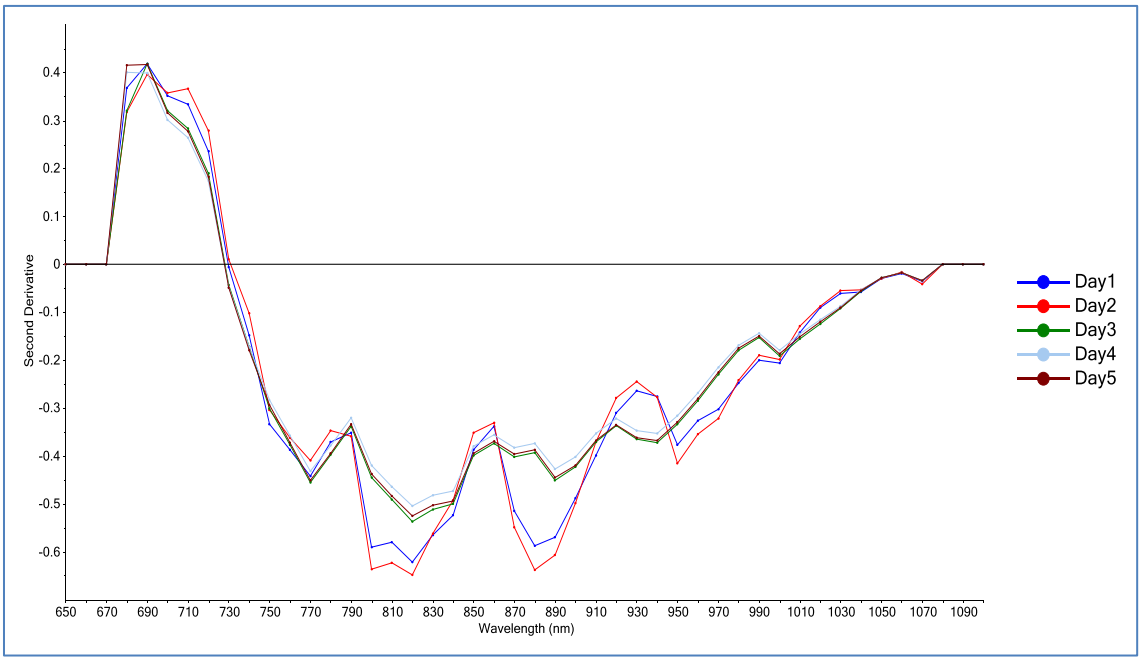


Figure 20: Second derivative reflectance spectra of smoothed spectral data of neoprene rubber.

For the repeatability experiment, the data was smoothed using Savitzky-Golay filter (Figure 19) and then the second derivative reflectance spectra was plotted (Figure 20). From the second derivative spectra, we can conclude that Day 3, Day 4, Day 4 spectra are repeatable, but Day 1 and Day 2 spectra are different from Day 3, Day 4, Day 5 spectra. In the Day 1 and Day 2 spectra there are negative peaks at 875 nm and 950 nm. Most probably, that is because of the fact that the humidity was more in the first two days. Water gets absorbed at in the wavelength range 900-1000nm [11].

4.2.2 Depth of Penetration of NIR Light into Chicken Breast Tissue

Objective: To quantify the depth of penetration of near-infrared light (650-1100 nm) into chicken breast sample using hyperspectral imaging system.

Materials and equipment required:

- A hyperspectral imaging system which consists of 1.4 megapixel, 12 bit digital imager (Qimaging Inc., Surrey, British Columbia), liquid crystal tunable filter (LCTF, Cambridge Research and Instrumentation Inc., Woburn, Massachusetts), and LCTF controller.
- Dual 500W white quartz tungsten halogen lamps for illumination.
- A laptop computer (Apple Macbook Pro, Cupertino, CA).
- A firewire IEEE 1394 cable to connect the computer to the LCTF controller.
- A parallel port cable to connect the LCTF controller to the camera.
- A USB cable to connect the camera to the computer.
- Chicken Breast.
- Neoprene Rubber Sheet.

Procedure:

In this experiment, chicken breast tissue was cut into sections of various thicknesses. The neoprene rubber sheet was kept underneath the chicken slice, and then investigated at what minimum width of the chicken slice we were getting the spectral effect of neoprene rubber. Prior to this, the hyperspectral images of the chicken slice and the neoprene rubber were captured separately.

While capturing the image, the distance between the lens and the subject should be constant. And care was taken that the light falls as uniformly as possible on the subject. The chicken breast was cut into pieces of the following width: 1.23 mm, 3 mm, 5 mm, 7.27 mm, 10.7 mm, 40mm and 60 mm respectively. The neoprene rubber sheet was kept below the chicken slices and hyperspectral images were captured. Hyperspectral image of only the neoprene sheet and the chicken breast (60 mm width) was also captured. Then, we compared the reflectance spectra of neoprene rubber, chicken breast and different widths of chicken breast over the neoprene rubber. The logic behind this test is that, once the chicken breast is thick enough, the neoprene would not influence the spectra anymore, and the shape and of the resulting spectra would remain fairly constant.

Results and Discussion:

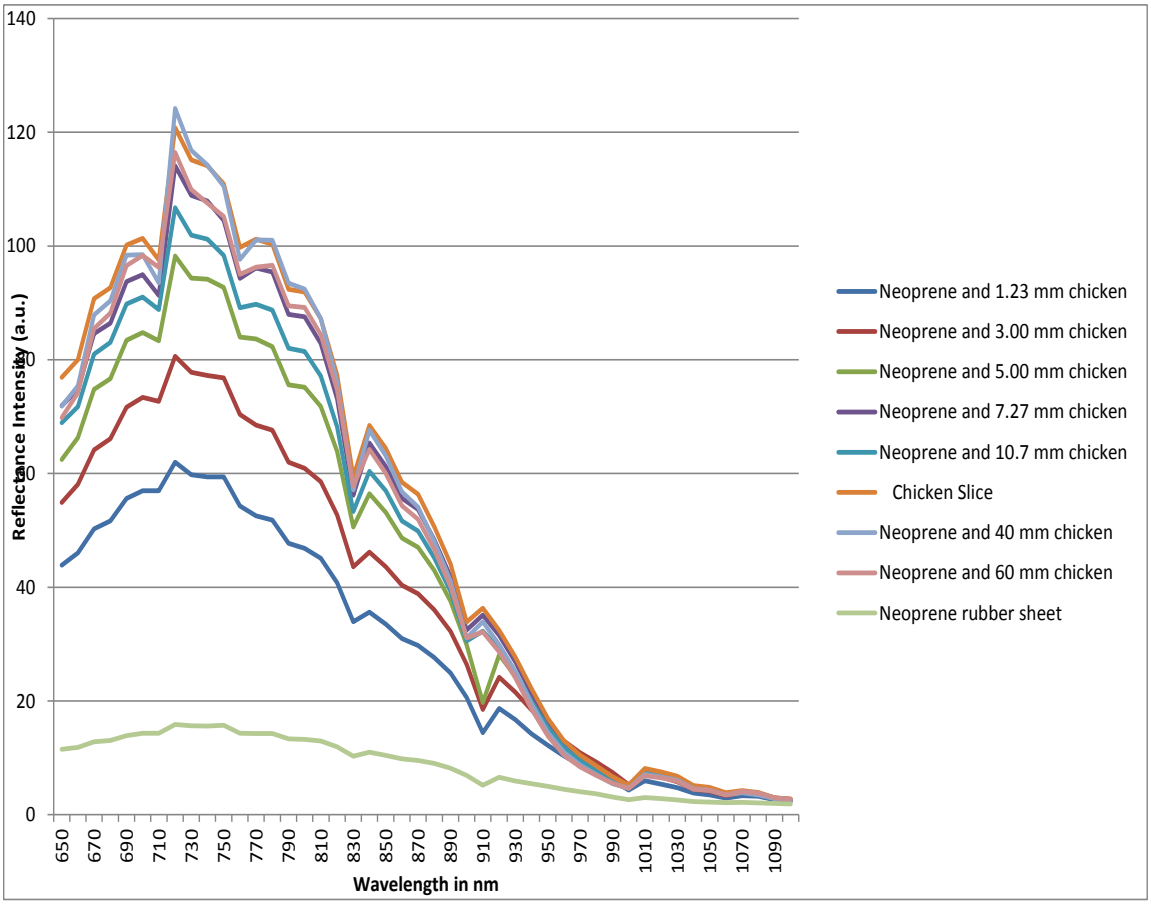


Figure 21: Graph showing the reflectance intensity vs. the wavelength for neoprene rubber sheet, chicken slice and varying width of chicken slices on the neoprene sheet.

The depth of penetration of NIR light into chicken breast can be quantified to be between 3 and 5 mm. From figure 21, it is seen that the neoprene rubber sheet, which was black in color, has very low reflectance intensity compared to the chicken breast. However, the slices of chicken of different widths are placed on the neoprene sheet, it is observed that as the slice of the chicken width increases, the reflectance intensity also increases. From 1.23 mm to 3 mm there is a sharp increase of reflectance intensity of approximately 30 a.u. Also, there is a sharp increase in reflectance intensity from 3mm to 5mm. However, after 5 mm, the increase is not drastic. The shape of the curve also remains fairly uniform after 5 mm. So we can estimate the depth of penetration of NIR light into chicken breast to be between 3mm to 5mm.

4.2.3 Experiment to Demonstrate Effect on Camera to Sample Distance on the Reflectance Spectra of Chicken Tissue

Objective:

To demonstrate the effect of the camera to sample distance on the reflectance spectra of chicken tissue.

Materials Required:

- The hyperspectral imaging system which consists of 1.4 megapixel, 12 bit digital imager (Qimaging Inc., Surrey, British Columbia), liquid crystal tunable filter (LCTF, Cambridge Research and Instrumentation Inc., Woburn, Massachusetts), and LCTF controller.
- Dual 500W white quartz tungsten halogen lamps for illumination
- A laptop computer (Apple Macbook Pro, Cupertino, CA)
- A Firewire IEEE 1394 cable to connect the computer to the LCTF controller
- A parallel port cable to connect the LCTF controller to the camera
- A USB cable to connect the camera to the computer
- Chicken Breast

Procedure:

The distance between the camera and the sample might affect the overall spectra of the sample. So, in this experiment, the distance between the sample and the camera was variedneach time and the hyperspectral image of the same target was captured. All other external conditions such as lighting, position of the camera were kept same for each time. Chicken breast was used as the sample. The distances between the camera and the sample were varied as follows: 23 cm, 26 cm, 35 cm, 39 cm, and 47 cm.

From Figure 22, it is seen that the reflectance spectra remain fairly constant even if the distance between the camera and the sample are varied each time. So changing the camera to sample distance does not have much effect on the reflectance spectra.

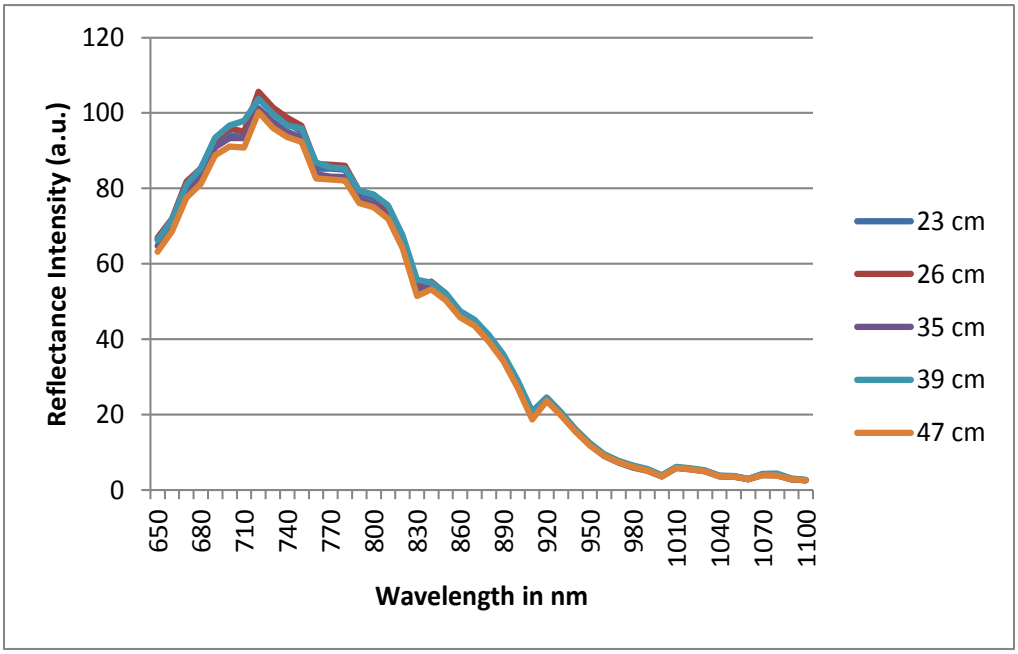


Figure 22: The reflectance spectra of the chicken sample with the distance between the camera and the sample varying each time.

4.3 Canine Mammary Tumor Experiment

In this section, the process of acquiring data from the canine patients is described. The reflectance spectra were extracted from the hyperspectral images. The spectra was smoothed and normalized. Four wavelengths were identified characteristic of four tissue chromophores, from the second derivative absorption spectra. Then the 'Tissue Optical Index' model and the PCA-LDA model were applied on the normalized and smoothed reflectance spectra.

4.3.1 Data Acquisition from Hospital

The data was acquired in collaboration with the University of Pennsylvania Veterinary Hospital. These canine patients had multiple mammary tumors in their abdomen. The animal experiments were approved by the University of Pennsylvania IACUC Protocol #803829. In this study, the hyperspectral imaging experiment was performed on 9 dogs. Prior to image acquisition, the tumors were marked with a black marker, so that it would be easier to recognize the tumor during image analysis. During image acquisition, the dogs were held by the veterinary doctor. Each dog had to undergo biopsy to detect the malignancy of the tumors. The histopathology results were compared with our imaging analysis results. Figure 23 shows a canine patient with multiple mammary tumors. Figure 24 shows the hyperspectral image of the canine patient at 760 nm wavelength and figure 25 shows the digital image of the canine patient.

Figure 23: The canine patient with multiple mammary tumors held by the doctor as images are being captured by the hyperspectral system.

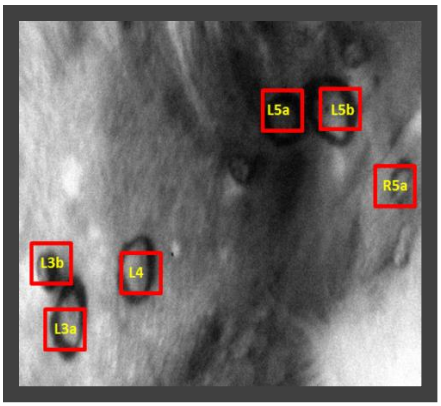


Figure 25: The hyperspectral image of canine mammary tumors at 760nm wavelength

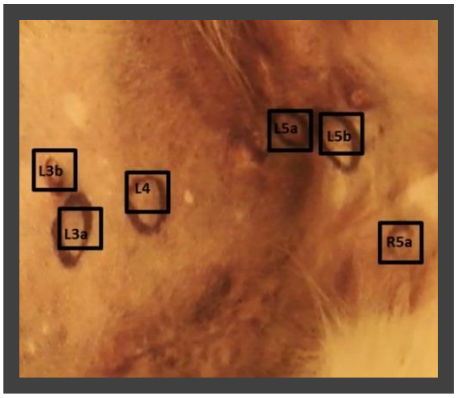


Figure 24: The digital image of canine mammary tumors.

4.3.2 Extracting the Reflectance Spectra of the Tumors

In this section, the hyperspectral image was visualized and the spectral signature of the tumor and normal tissue regions were plotted using ENVI software (Ver. 4.5, ITT Visual information solutions, Boulder, CO). The Region of Interest was chosen as the size of the entire tumor. There were 7 malignant tumors, 15 benign tumors and 22 normal region ROIs.

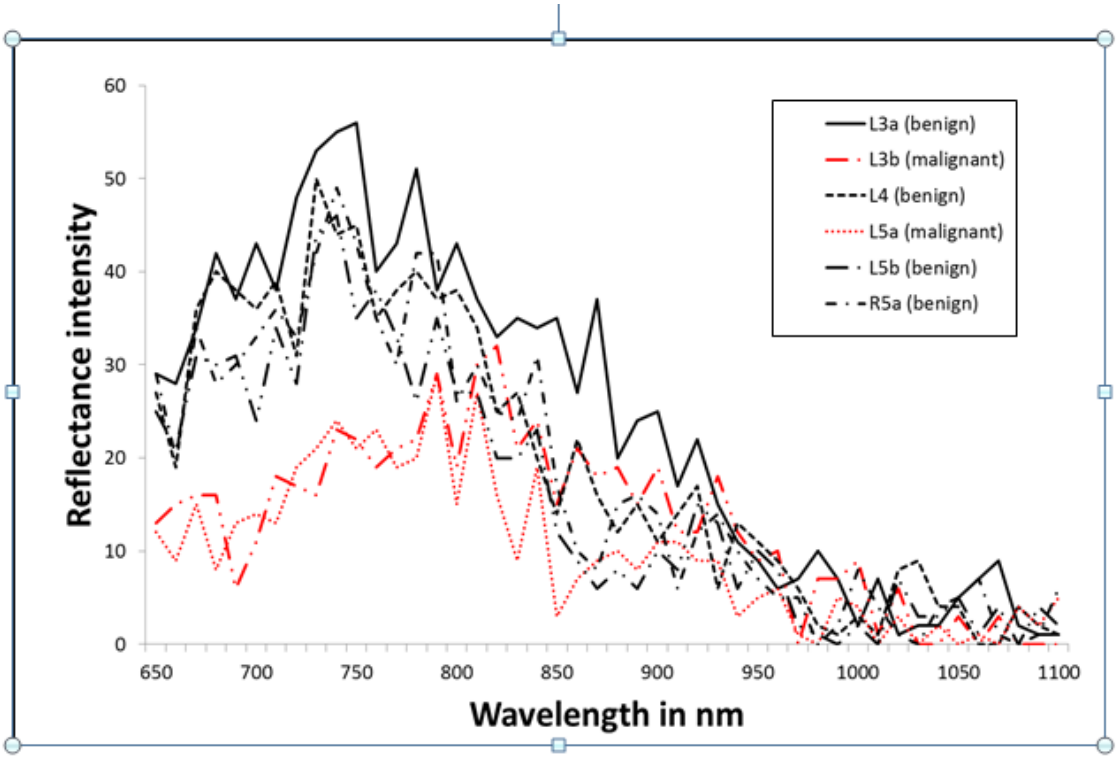


Figure 26: The reflectance spectra of one of the canine patients. The raw reflectance spectra have a low signal-to-noise ratio. To mitigate the effect of noise, we smoothen the signal.

4.3.3 Smoothing and Normalization

The raw reflectance spectra appeared to have random noise. To minimize the noise, the reflectance spectra were smoothed using Savitzky Golay Smoothing Process (second order polynomial and 70 nm window). The smoothed reflectance spectra were normalized using range normalization technique. For normalization and smoothing, 'The Unscrambler' version 10.1 (CAMO Software AS, Oslo, Norway) was used. Figure 27 shows the smoothed reflectance tumor spectra of all canine patients. Figure 28 shows the normalized reflectance tumor spectra of all canine patients.

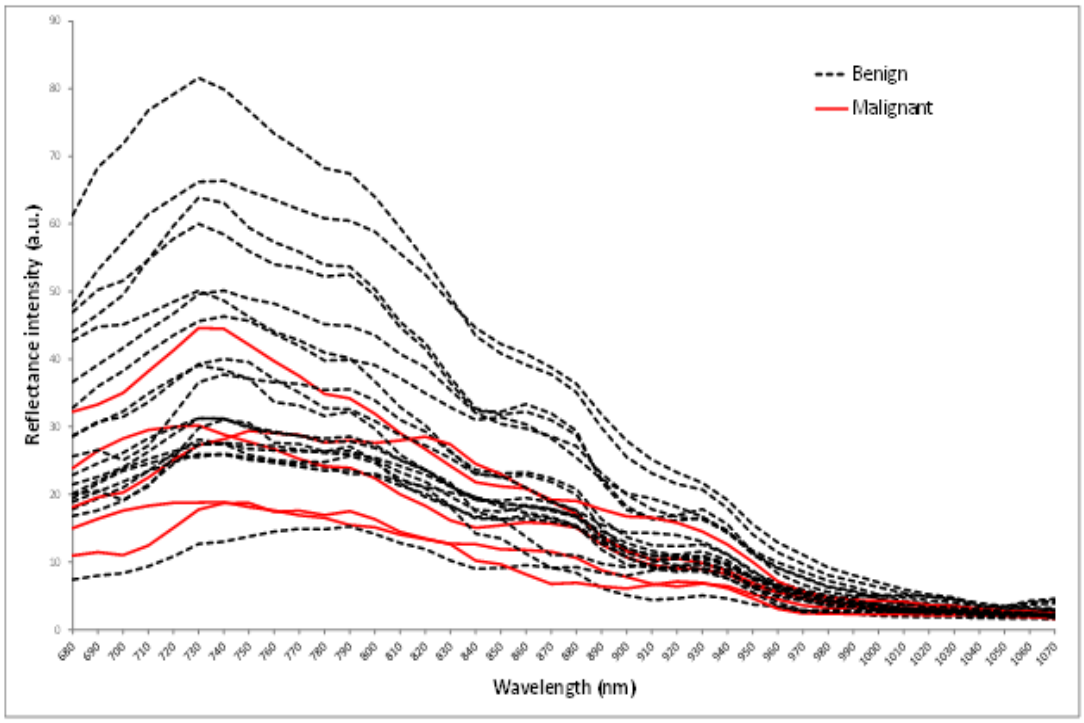


Figure 27: The smoothed reflectance spectra of the all canine patient tumors.

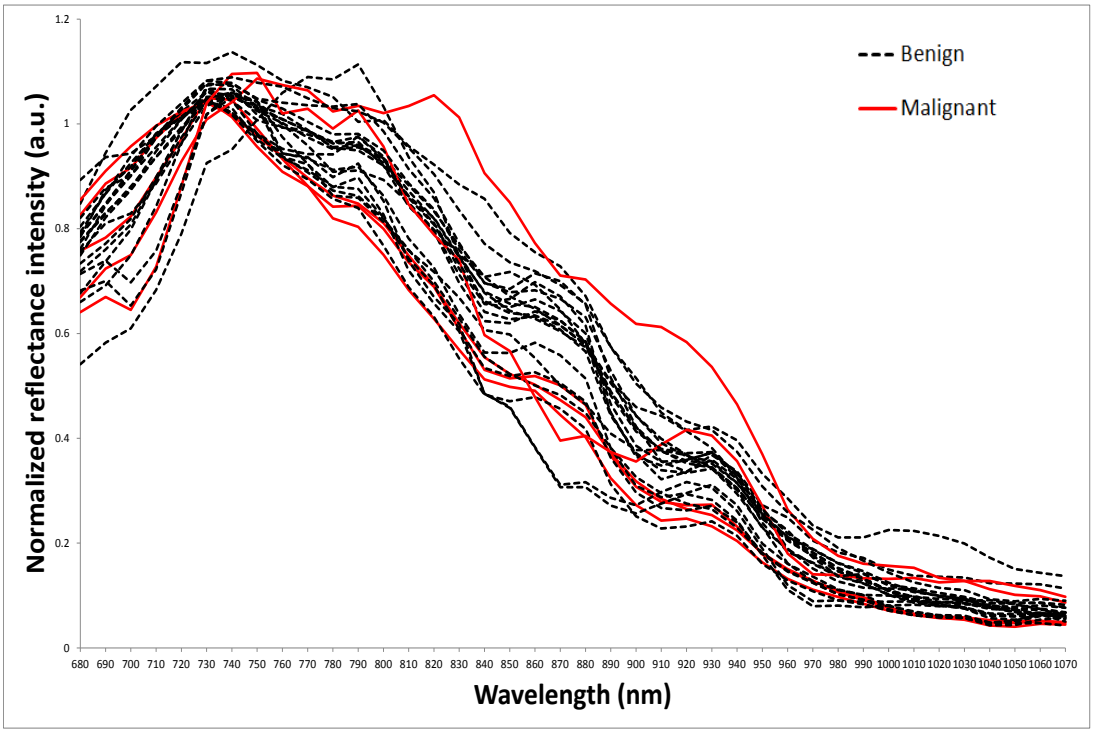


Figure 28: The range normalized reflectance spectra of the all canine patient tumors.

4.3.4 The Reflectance Spectra of Canine Patients

In this section, the smoothed reflectance spectra of malignant, benign, and normal ROIs of the 9 canine patients are given.

Canine Patient #1

Name of canine patient: Maxine

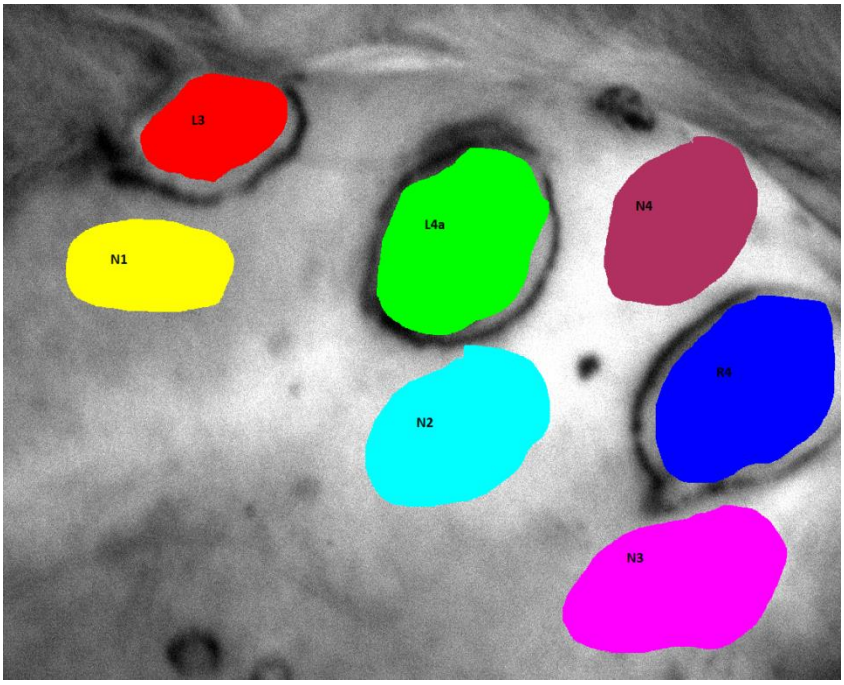


Figure 29: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Maxine.

Table 2: Clinical Results of Maxine

Tumor Label	Pathology Results
R4	Malignant
L3	Malignant
L4a	Benign

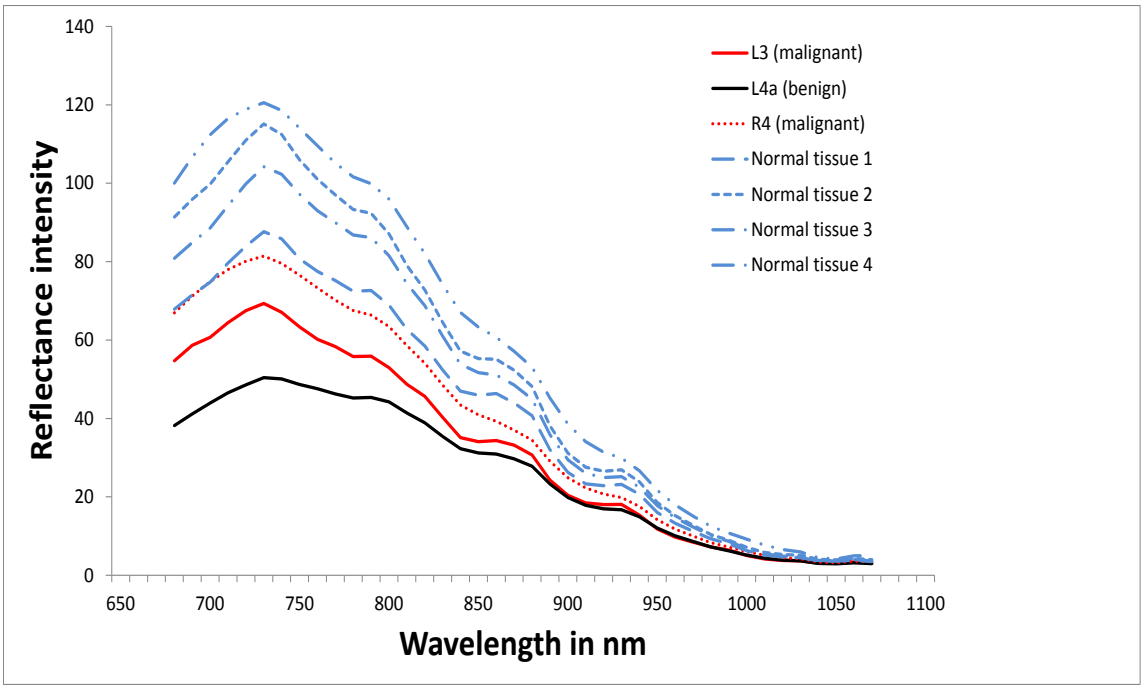


Figure 30: The spectral plot of the malignant, benign and normal tissue of canine patient Maxine.

Canine Patient #2

Name of Canine Patient: Chelsey

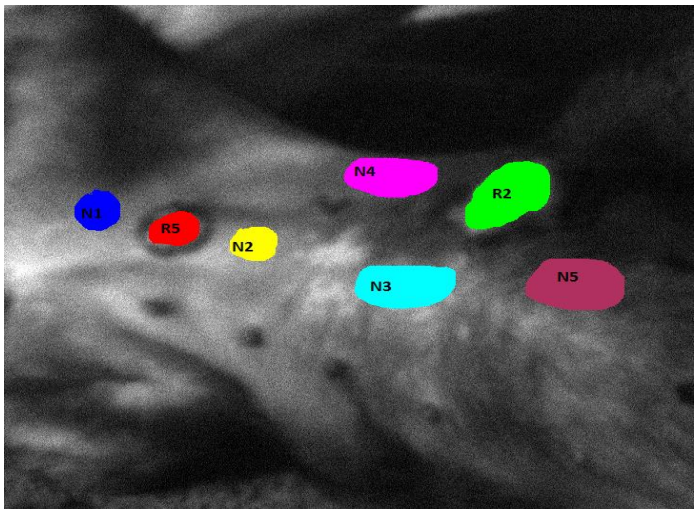


Figure 31: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Chelsey

Table 3: Clinical result of Chelsey

Tumor Label	Pathology Results
R2	Benign mixed tumor
R5	Benign mixed and lobular hyperplasia

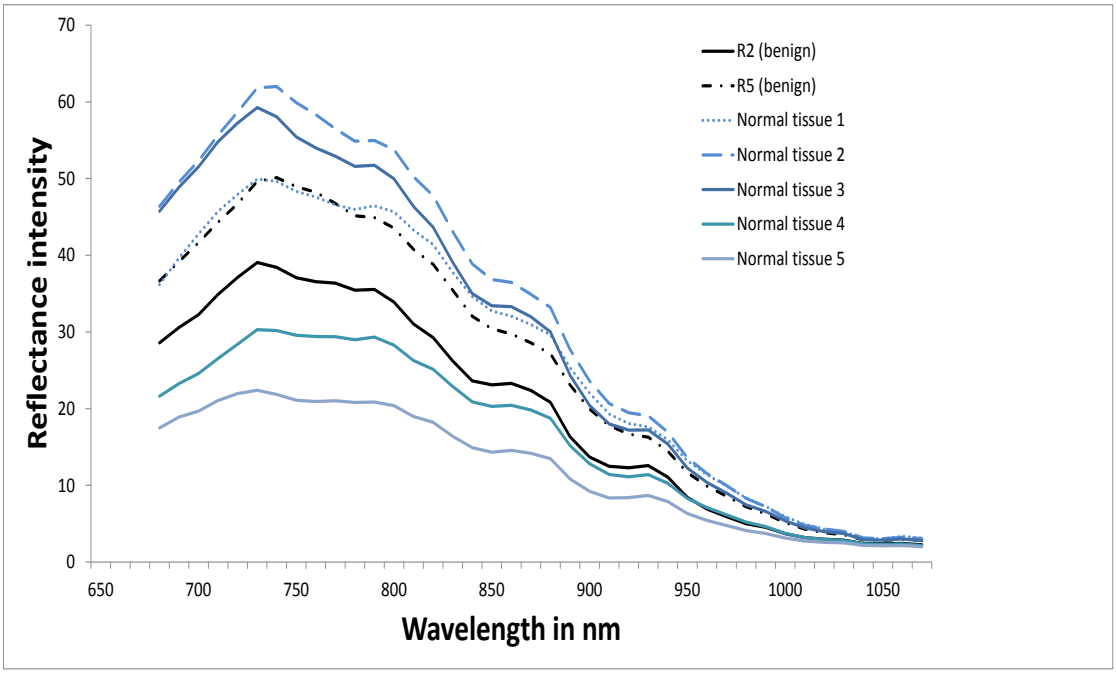


Figure 32: The spectral plot of the benign and normal tissue of canine patient Chelsey.

Canine Patient #3

Name of Canine Patient: Hanna

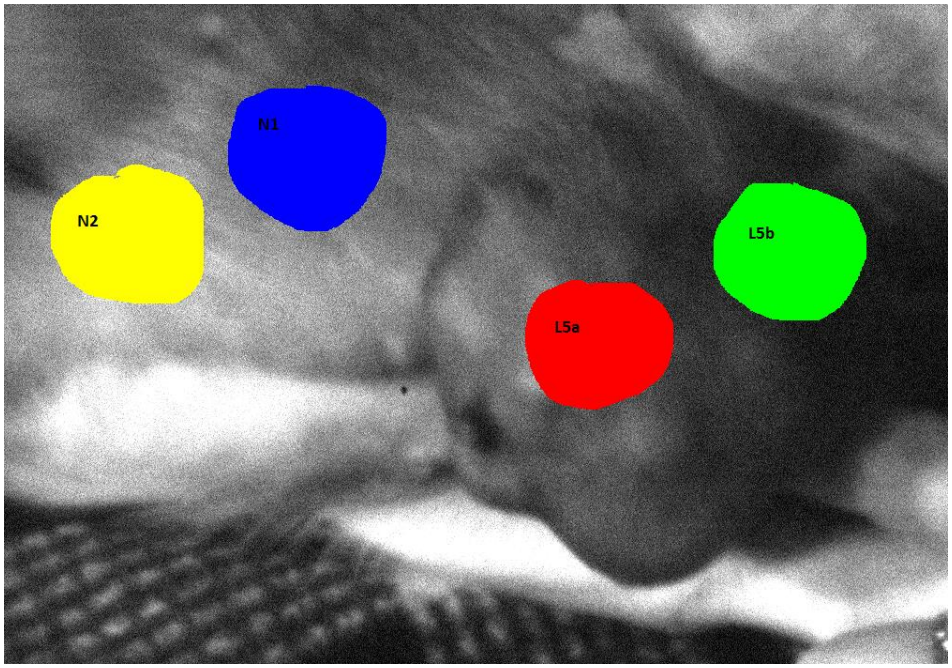


Figure 33: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Hanna.

Table 4: Clinical result of Hanna

Name of Tumor	Type of tumor
L5a	Benign mixed tumor
L5b	Benign mixed tumor

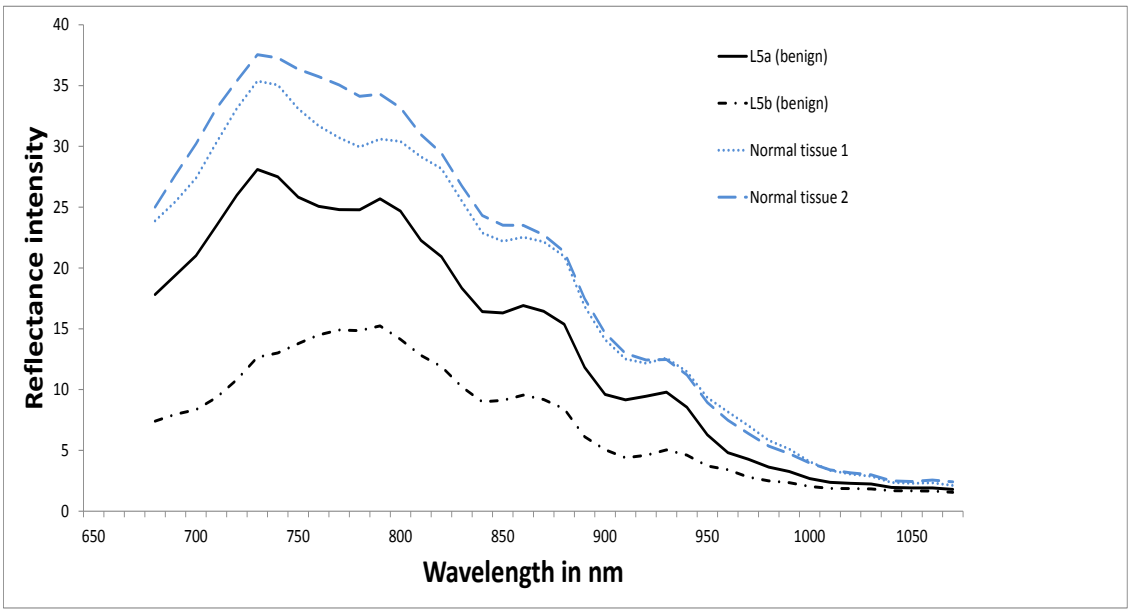


Figure 34: The spectral plot of the benign and normal tissue of canine patient Hanna.

Canine Patient #4

Name of Canine Patient: Lady

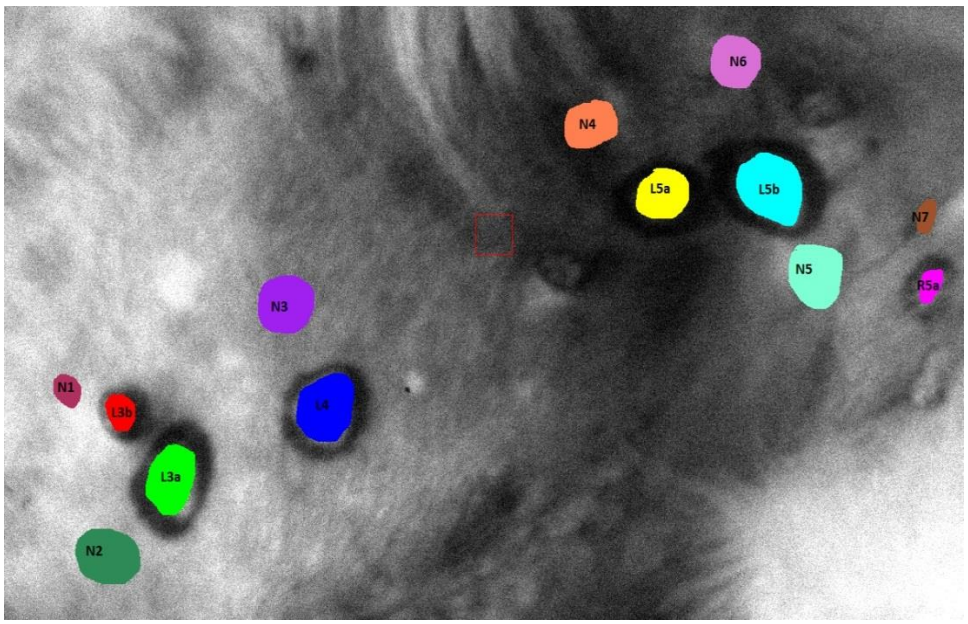


Figure 35: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Lady.

Table 5: Clinical result of Lady

Name of Tumor	Type of tumor
L3a	Adenoma (benign)
L3b	Epitheliosis **(malignant)**
L4	Papillary Adenoma(benign)
L5a	Intraductal Papillary Carcinoma**(malignant)**
L5b	Intraductal papillary Adenoma(benign)
R5a	Complex Adenoma(benign)

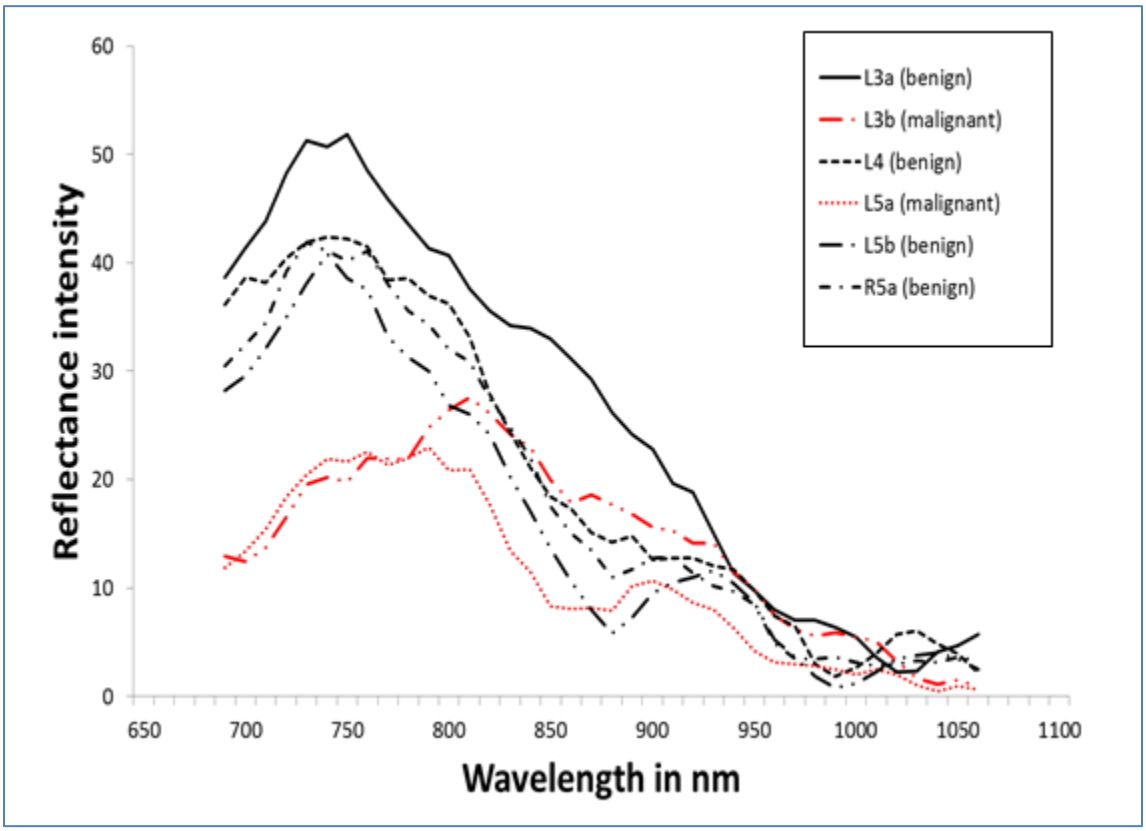


Figure 36: The spectral plot of the benign and malignant lesion of canine patient Lady.

Canine Patient #5

Name of Canine Patient: Suzanna

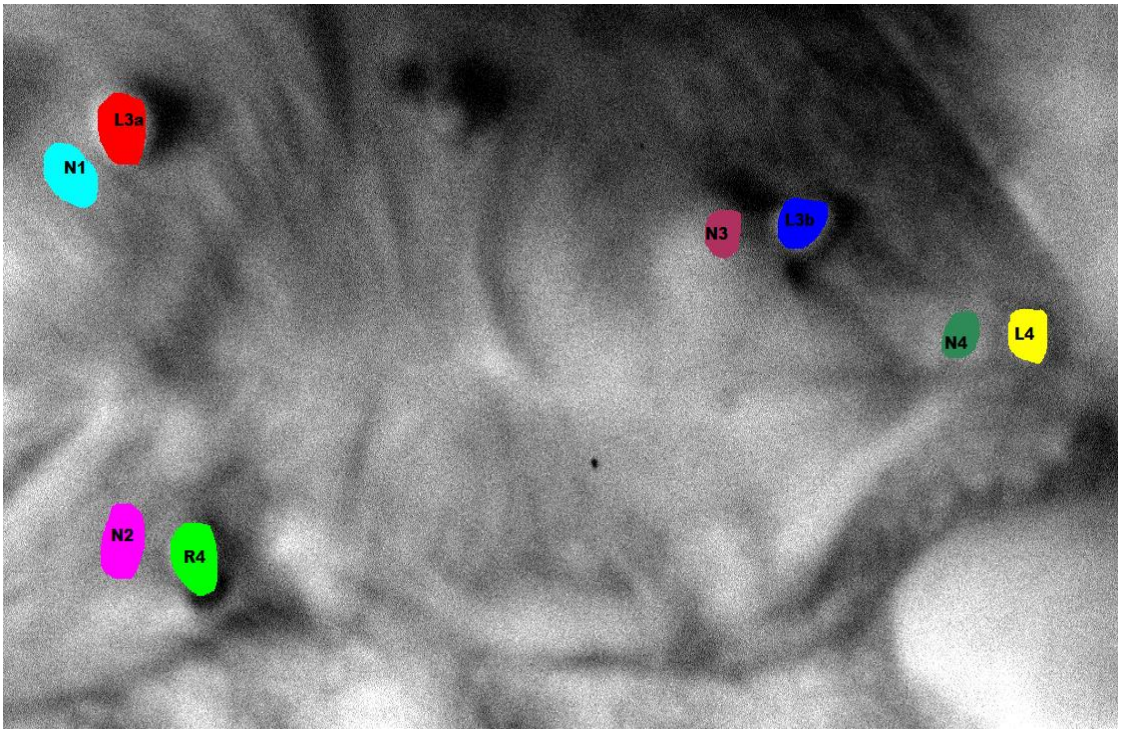


Figure 37: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Suzanna.

Table 6: Clinical result of Suzanna

Name of Tumor	Type of tumor
L3a	Benign mixed tumor
L3b	Benign mixed tumor
L4	Complex Adenoma (benign)
R4	Benign mixed tumor

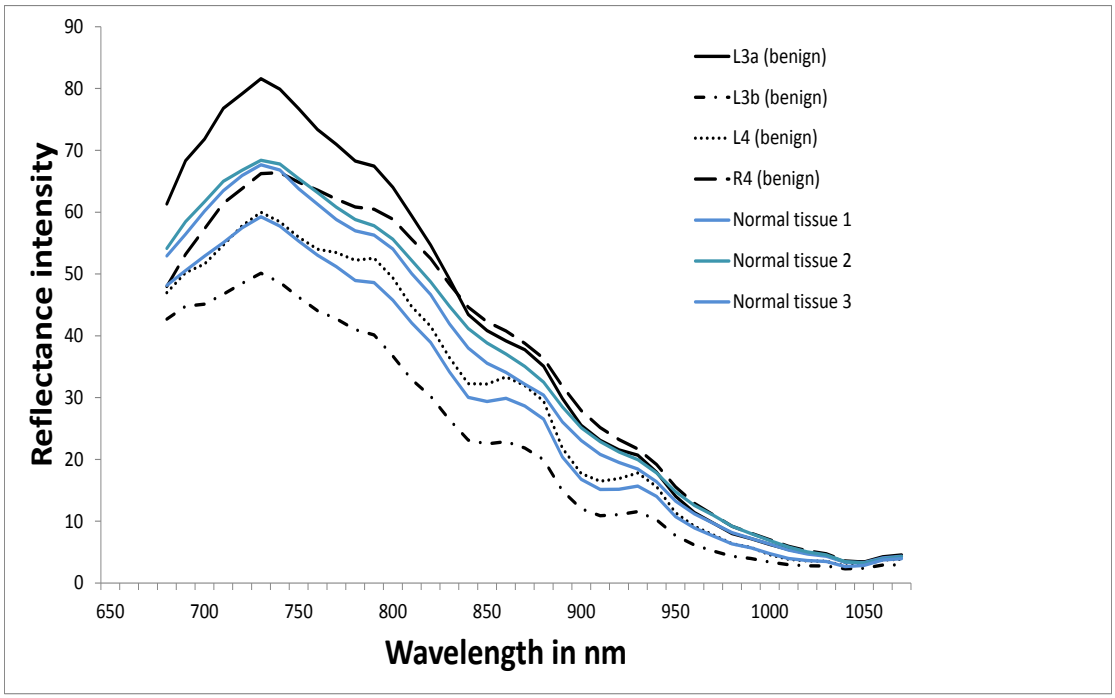


Figure 38: The spectral plot of the benign lesion and normal tissue of canine patient Suzanna.

Canine Patient #6

Name of Canine Patient: Queenie

Table 7: Clinical result of Queenie

Tumor Label	Pathology Results
R4	Carcinoma arising in a benign mixed tumor (malignant)

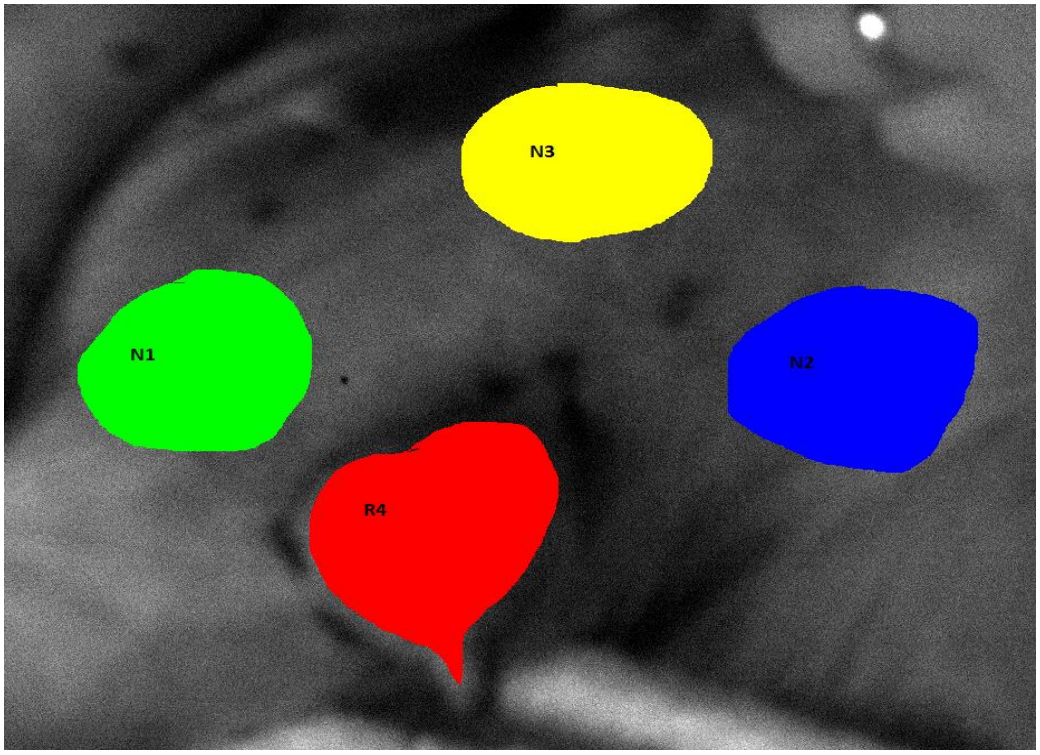


Fig 39: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Queenie.

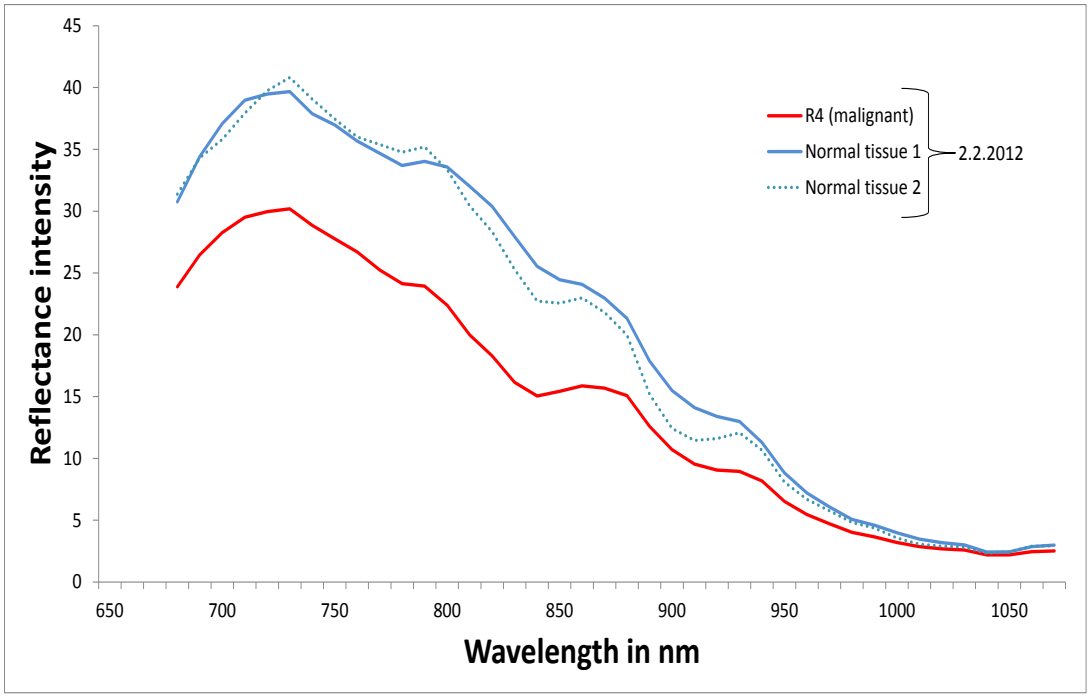


Fig 40: The spectral plot of the malignant lesion and normal tissue of canine patient Queenie.

Canine Patient #7

Name of Canine Patient: Tinkerbell

Table 8: Clinical result of TinkerBell

Tumor Label	Pathology Results
L1	Adenocarcinoma (malignant)

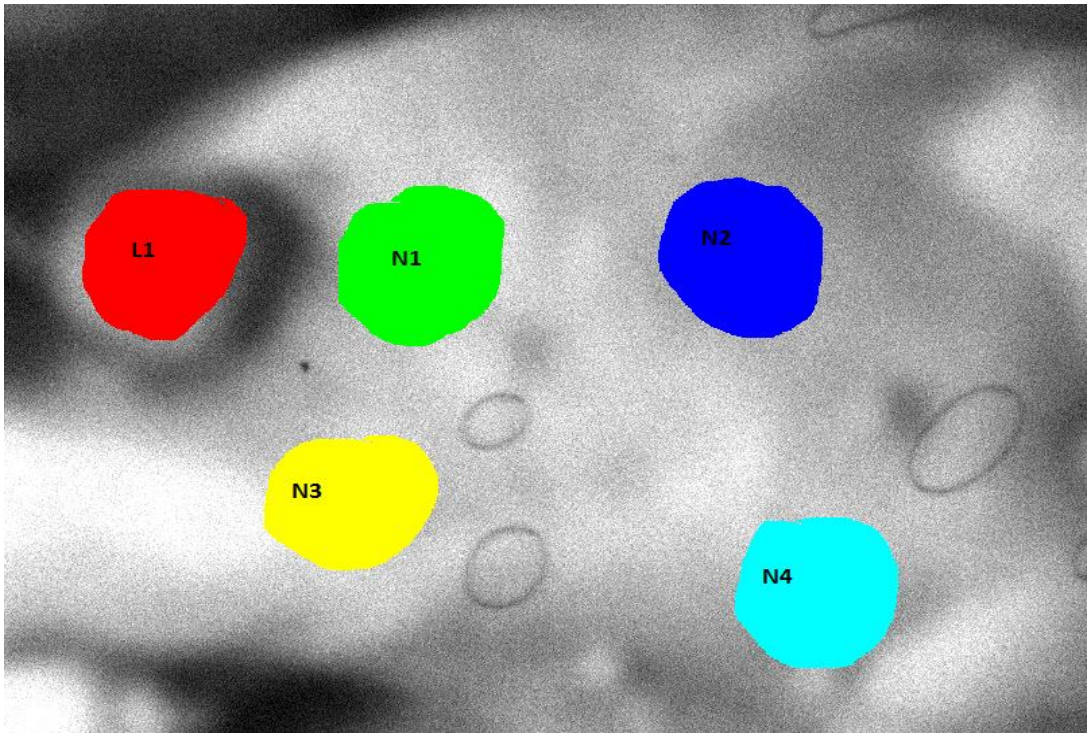


Fig 41: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Tinkerbell.

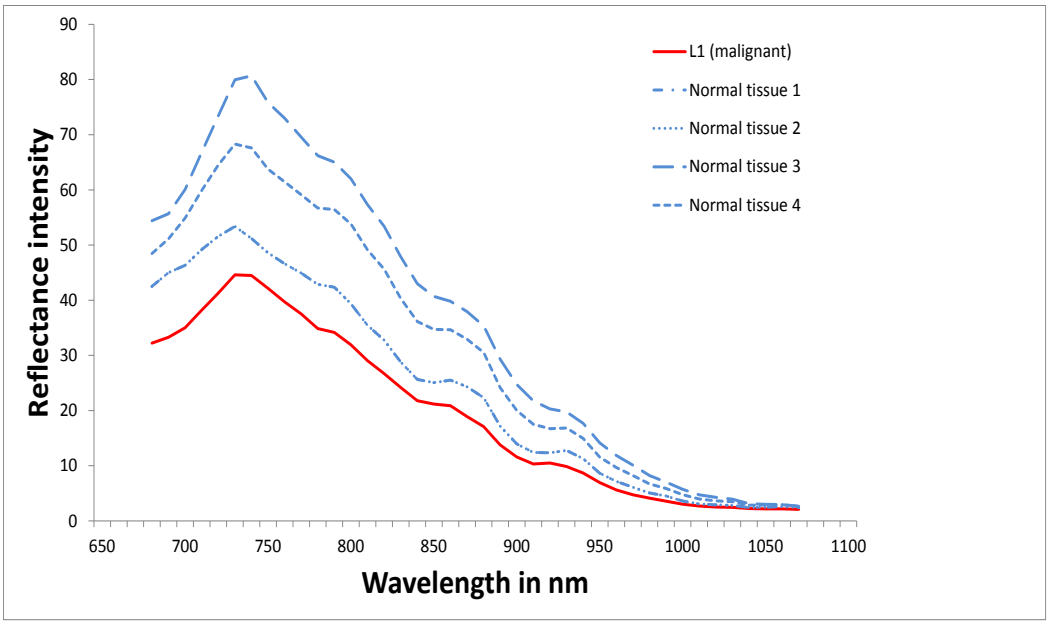


Fig 42: The spectral plot of the malignant and normal tissue of Tinkerbell

Canine Patient #8

Name of Canine Patient: Gretchen

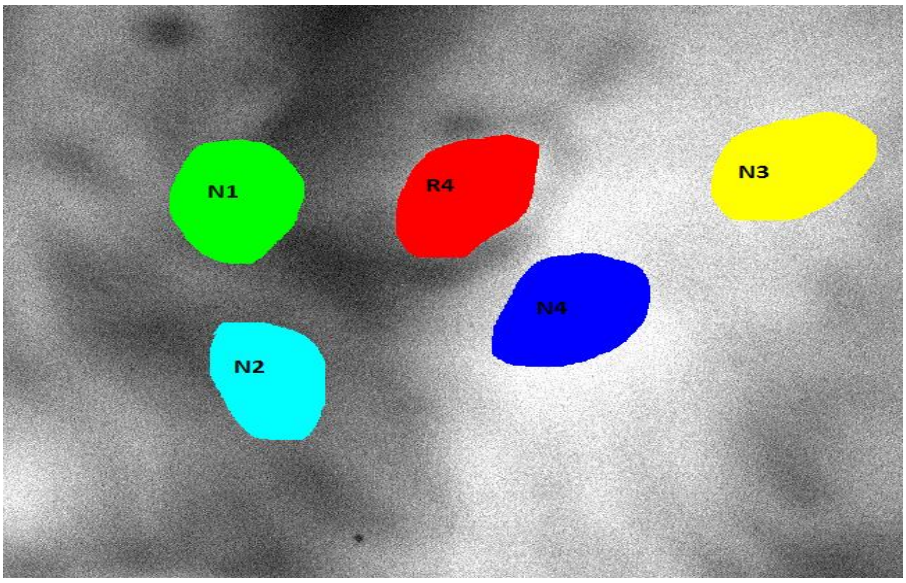


Fig 43: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Gretchen.

Table 9: Clinical result of Gretchen

Tumor Label	Pathology Results
R4	Benign

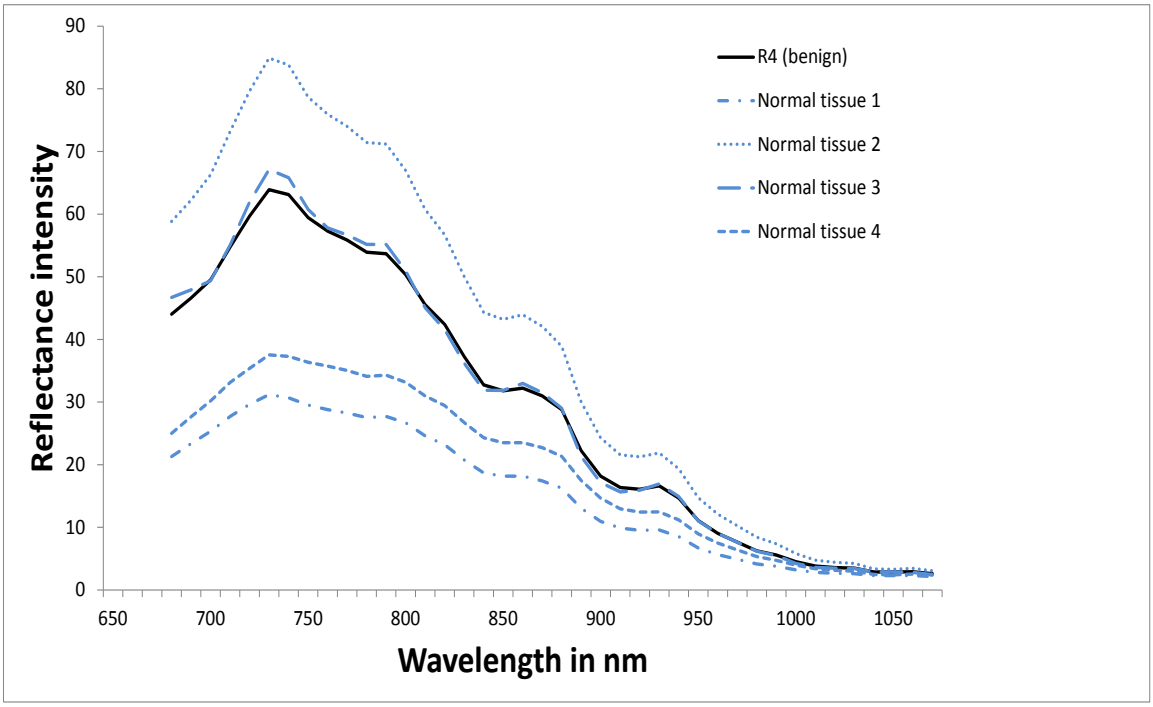


Figure 44: The spectral plot of the benign lesion and normal tissue of canine patient Gretchen

Canine Patient #9

Name of Canine Patient: Kiki

Table 10: Clinical result of Kiki

Tumor Label	Pathology Results
R1	Benign
L3	Malignant

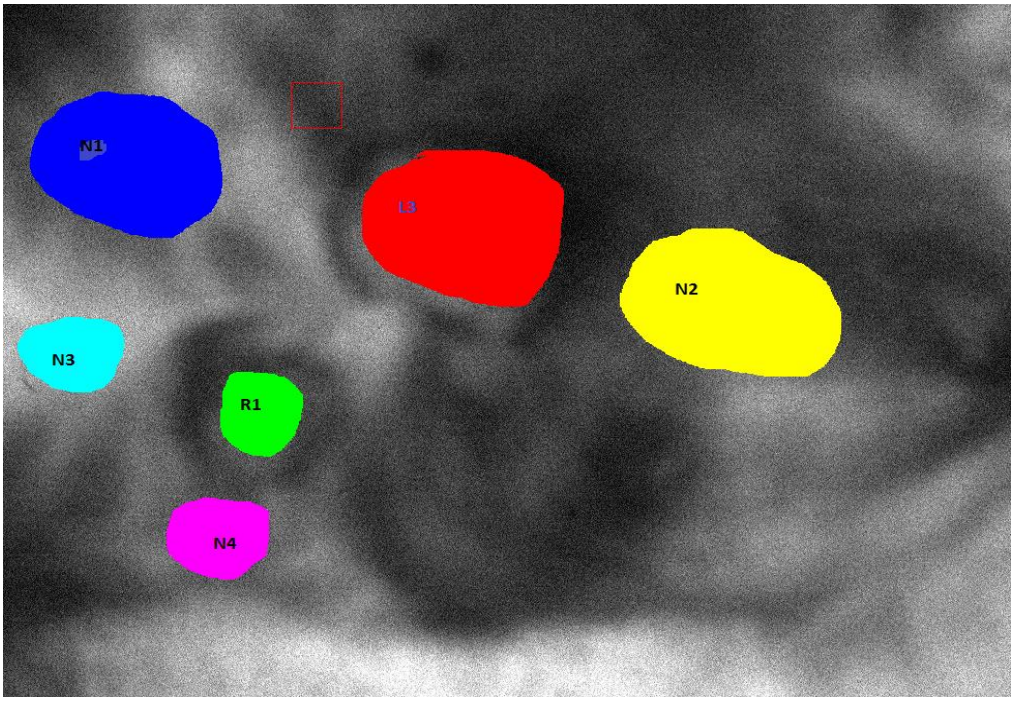


Figure 45: The ROI of the tumor tissue and the normal tissue of hyperspectral image of canine patient Kiki.

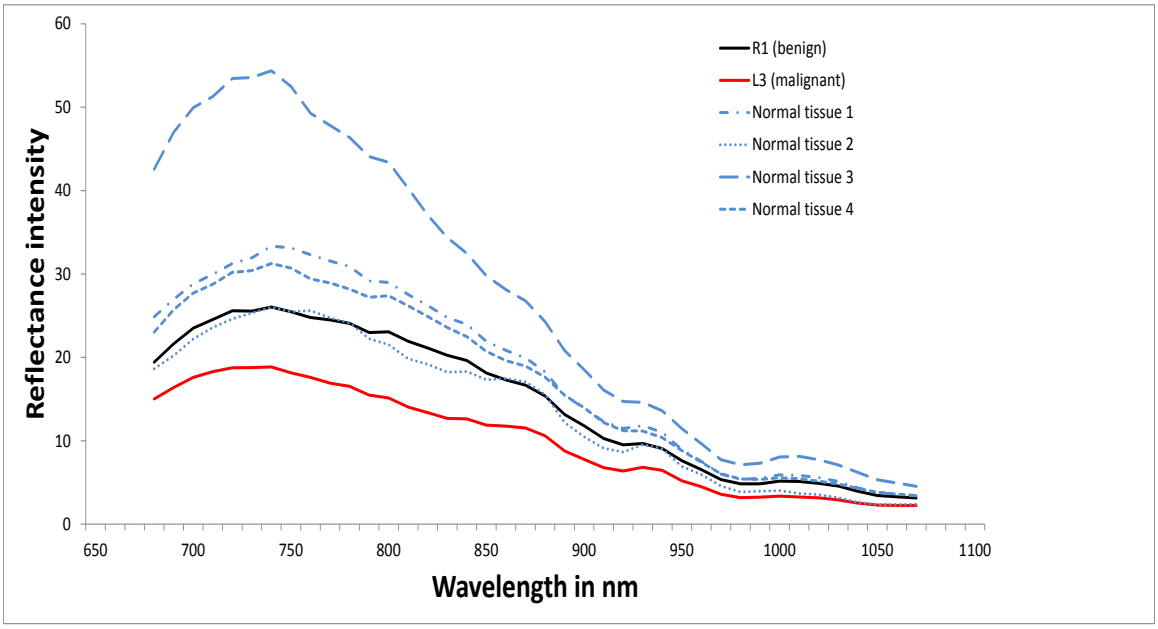


Figure 46: The spectral plot of the malignant and normal tissue of canine patient Kiki

The spectra of malignant tumor tissue were clearly different from those of benign and normal tissue (Figure 36, 40, 42, 46). Within a single patient the reflectance intensities from 650 to 110 nm of cancerous tissue were relatively low compared to the benign and normal tissue. Previous studies have shown that cancerous tissues have a higher hemoglobin and water content, and lower lipid concentration with respect to benign tissues. The decreased reflectance intensity observed for malignant tumors is likely due to the increased microvasculature and therefore higher blood content of malignant tissue relative to benign tissue.

Table 11: The reflectance intensity values all ROI at 700, 840, 900 and 970 nm.

Tumor Label	Type of Tumor	Reflectance intensity at 700 nm	Reflectance intensity at 840 nm	Reflectance intensity at 900 nm	Reflectance intensity at 970 nm
Kiki R1	Benign	49.94	32.46	18.54	7.73
Kiki L3	Malignant	35.01	21.80	11.59	4.73
Kiki N1	Normal	57.76	35.18	19.62	8.13
Kiki N2	Normal	46.32	25.62	13.95	6.09
Kiki N3	Normal	60.09	43.01	24.71	10.11
Tinkerbell N1	Normal	60.69	35.14	20.37	8.43
Tinkerbell L1	Malignant	43.94	32.29	19.81	8.65
Tinkerbell N2	Normal	74.90	43.41	24.88	10.03
Tinkerbell N3	Normal	74.76	46.94	26.25	11.24
Maxine L3	Malignant	99.76	57.16	31.15	12.76
Maxine L4a	Benign	88.60	53.82	29.44	12.34
Maxine R4	Malignant	28.26	15.05	10.71	4.73
Maxine N1	Normal	37.08	25.53	15.49	6.08
Maxine N2	Normal	35.81	22.73	12.42	5.76
Maxine N3	Normal	32.25	23.61	13.68	5.92
Queenie R4	Malignant	41.66	32.05	19.95	8.59
Queenie N1	Normal	42.78	34.63	22.02	9.91

Queenie N2	Normal	52.31	38.88	23.59	9.98
Chelsea R2	Benign	51.53	34.99	20.41	9.00
Chelsea R5	Benign	49.51	32.73	18.17	7.65
Chelesa N1	Normal	25.34	18.69	10.96	4.91
Chelsea N2	Normal	66.30	44.30	24.27	10.26
Chelsea N3	Normal	49.37	31.90	17.16	7.62
Gretchen R4	Benign	21.01	16.41	9.60	4.28
Gretchen N1	Normal	8.34	8.99	5.06	2.80
Gretchen N2	Normal	27.39	22.88	14.11	7.04
Gretchen N3	Normal	43.38	31.93	19.11	8.90
Hanna L5a	Benign	71.81	43.46	25.51	9.70
Hanna L5b	Benign	45.11	23.11	12.00	5.18
Hanna N1	Normal	51.61	32.27	17.70	7.78
Hanna N2	Normal	57.32	44.62	27.90	11.05
Suzzana L3a	Benign	60.18	38.01	23.06	9.67
Suzzana L3b	Benign	61.69	41.17	25.11	11.00
SuzzanaL4a	Benign	52.90	30.04	16.79	7.61
Suzzana R4	Benign	38.26	31.00	20.14	6.63
Suzzana N1	Normal	20.28	24.51	16.73	5.63
Suzzana N2	Normal	31.47	23.02	14.33	4.81
Suzzana N3	Normal	11.02	10.19	6.07	2.40

4.3.5 Finding the wavelength characteristic of four chromophores

Chromophore specific wavelengths were selected by applying second derivative reflectance spectra. Peaks at 700, 840, 900 and 970 nm were observed in the second derivative reflectance spectra, these peaks were attributed to deoxy-hemoglobin, oxy-hemoglobin, lipid and water respectively. These four wavelengths were chosen because based on literature, the absorption of oxy-hemoglobin and de-oxyhemoglobin dominates at wavelength ranging from 600-850 nm and

the absorption of lipid and water is more at higher wavelengths (900-1000 nm) [4]. The second derivative spectra were plotted using 'The Unscrambler' version 10. Figure 47 shows the second derivative reflectance spectra of the reflectance data of all canine patients.

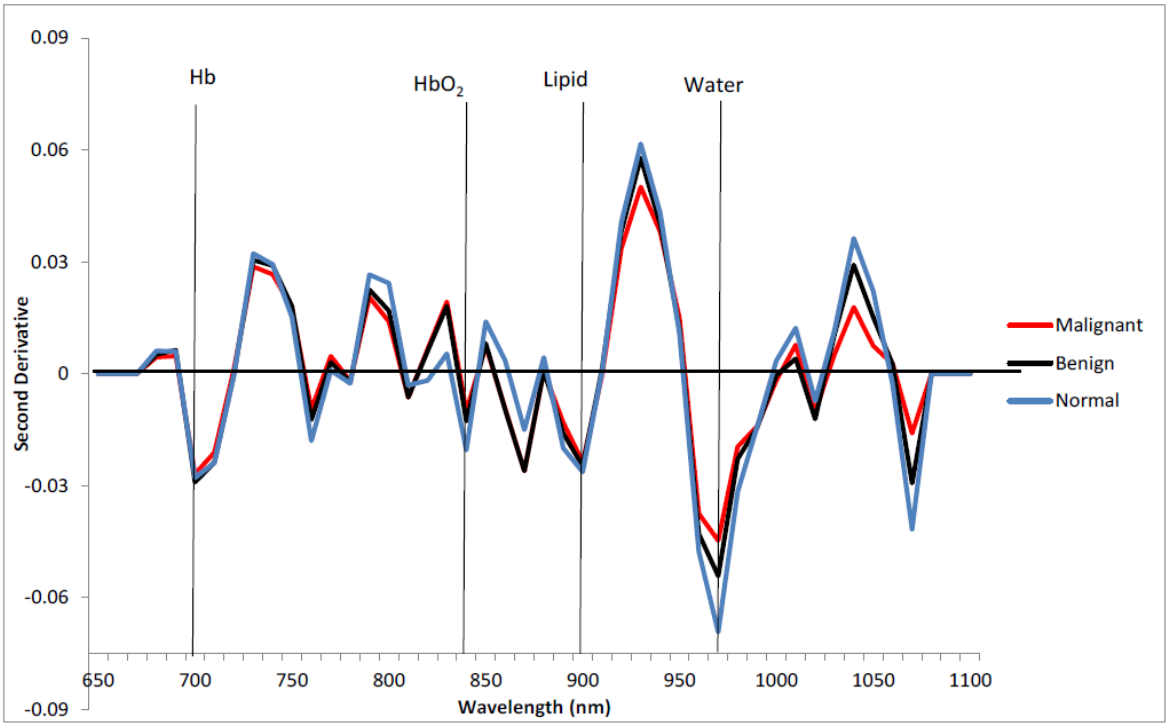


Figure 47: The representative second derivative absorption spectra of the reflectance data of all canine patients. Four negative peaks were observed at the wavelengths of 700, 840, 900 and 970 nm.

4.3.6 Data analysis using TOI method and PCA-LDA method

4.3.6.1 Tissue Optical Indices Method

<u>Computation of TOI using whole tumor as ROI</u>

While applying the TOI using the whole tumor as the ROI, the ROI (region of interest) of the malignant and benign tumors are taken as the size of the entire tumor (Figure 48). The ROI of the normal tissue is taken adjacent to the tumor tissue, in approximately the same pixel size as the

tumor tissue. The total reflectance intensity within the ROI is averaged to compute the reflectance intensity. The spectral information is then obtained using 'ROI Tools' in ENVI 4.5. The spectral information gives the plot of the reflectance intensity vs. the wavelength. These spectral files were then saved as ASCII files.

To analyze the results and smoothen the data, the ASCII spectral files were opened in Excel. This procedure is as follows: Open text file in Excel →select delimited→click next→select Tab→then click Finish. This will give experimental details (e.g. file name, date), band numbers, average reflectance intensity and standard deviation and reflectance intensities for the ROI. The average reflectance intensities for all tumors and normal tissue for all patients were copied into a single Excel file. The Excel file was then imported into 'The Unscrambler' version 10.1 (CAMO Software AS, Oslo, Norway) and Savitzky-Golay smoothing process was used (second order polynomial and 70 nm window) to smoothen the data. The Unscrambler software is a statistical program primarily designed for multivariate analyses. Using the Unscrambler software, data may be preprocessed (e.g. Savitzky-Golay smoothing) and analyzed using methods such as principal components analysis (PCA), multivariate regression and partial least squares. The smoothed data was copied and put back into an excel file so that it could be easily plotted and examined using any computer. Then equation 11 was used to calculate the values of the TOIs.

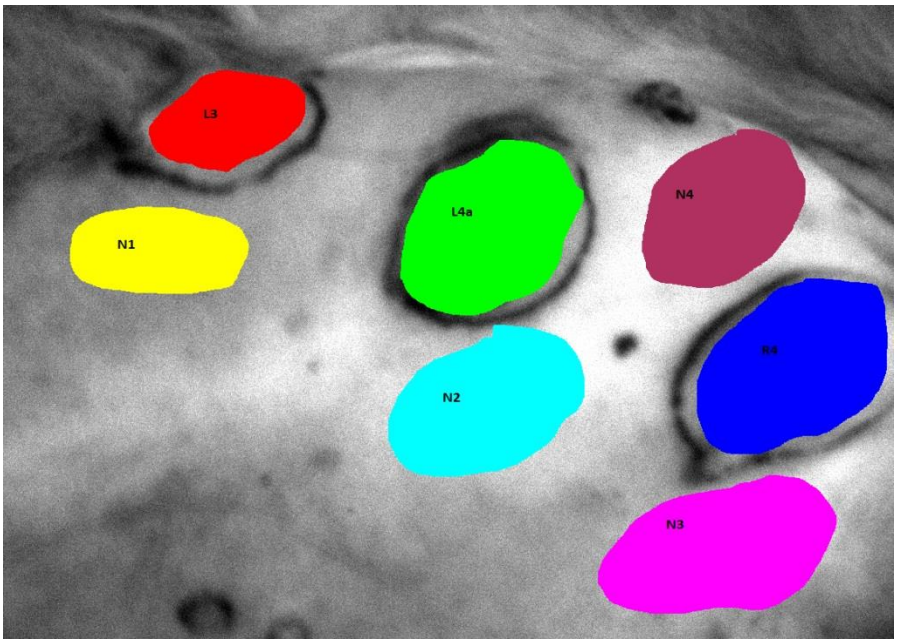


Figure 48: The whole tumor ROIs of one of the canine patients

Results:

A 'Tissue Optical Index' was developed as described in Section 3.1. The Tissue Optical Index values were calculated using equation 11. Malignant tumors had higher TOI values than benign tumors (Figure 49). Using a TOI threshold of 2.00 units (threshold was obtained by 44-fold cross validation), 6 out of 7 malignant tumors, 13 out of 15 benign tumors, and all of 22 normal tissue ROIs were correctly identified. So the sensitivity and specificity of the proposed method were 86% and 95% respectively.

Table 12: Tissue Optical Indices value of the benign, normal and malignant ROIs as calculated from the reflectance spectra. For ease of visualization, malignant, benign, and normal ROIs are shaded in red, green and blue respectively

Name of Patient	Label and Type of tumor	TOI
Kiki	R1 (benign)	1.07
	L3 (malignant)	2.56
	N1 (normal)	1.12
	N2 (normal)	1.34
	N3 (normal)	0.99

Tinkerbell	L1 (malignant)	2.52
	N1 (normal)	1.42
	N2 (normal)	1.22
	N3 (normal)	1.39
Queenie	R4 (malignant)	6.14
	N1 (normal)	1.52
	N2 (normal)	1.59
Lady	L3a (benign)	1.22
	L3b (malignant)	1.09
	L4 (benign)	1.45
	L5a (benign)	0.96
	L5b(malignant)	2.35
	R5a(benign)	1.78
Chelsey	R2 (benign)	1.81
	R5 (benign)	1.59
	N1 (normal)	1.29
	N2(normal)	1.72
	N3(normal)	1.44
Gretchen	R4 (benign)	1.96
	N1 (normal)	1.72
	N2(normal)	1.41
	N3(normal)	1.77
Suzanna	L3a (benign)	3.29
	L3b (benign)	4.74
	L4 (benign)	1.56
	R4 (benign)	1.90
	N1 (normal)	1.11
	N2(normal)	1.87
	N3(normal)	1.22
Maxine	L3 (malignant)	3.42
	L4a (benign)	1.86
	R4 (malignant)	4.26
	N1(normal)	1.66
	N2(normal)	1.98
	N3(normal)	1.78
Name of Patient	Label and Type of tumor	TOI
Hanna	L5a (benign)	1.72
	L5b(benign)	1.25
	N1(normal)	1.26
	N2(normal)	1.69

.

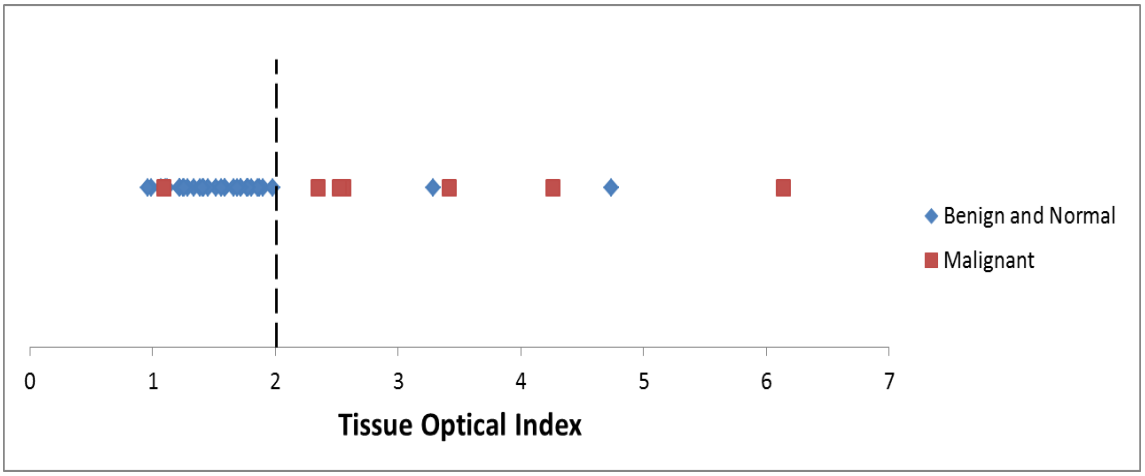


Figure 49: Graph showing that the malignant tumors generally have a higher TOI than benign tumors. TOI threshold of 2 units was used.

Computation of TOI by taking multiple ROIs in a tumor

The number of ROIs obtained in the previous method is 44, out of which only 7 are malignant. So using the whole number as the ROI, we might not obtain enough ROIs to prove that the results are statistically significant. So in this method, we took 4 ROIs in each tumor, and also would take more number of normal ROIs (Figure 50). This would increase the overall number of ROIs, resulting in more number of test points.

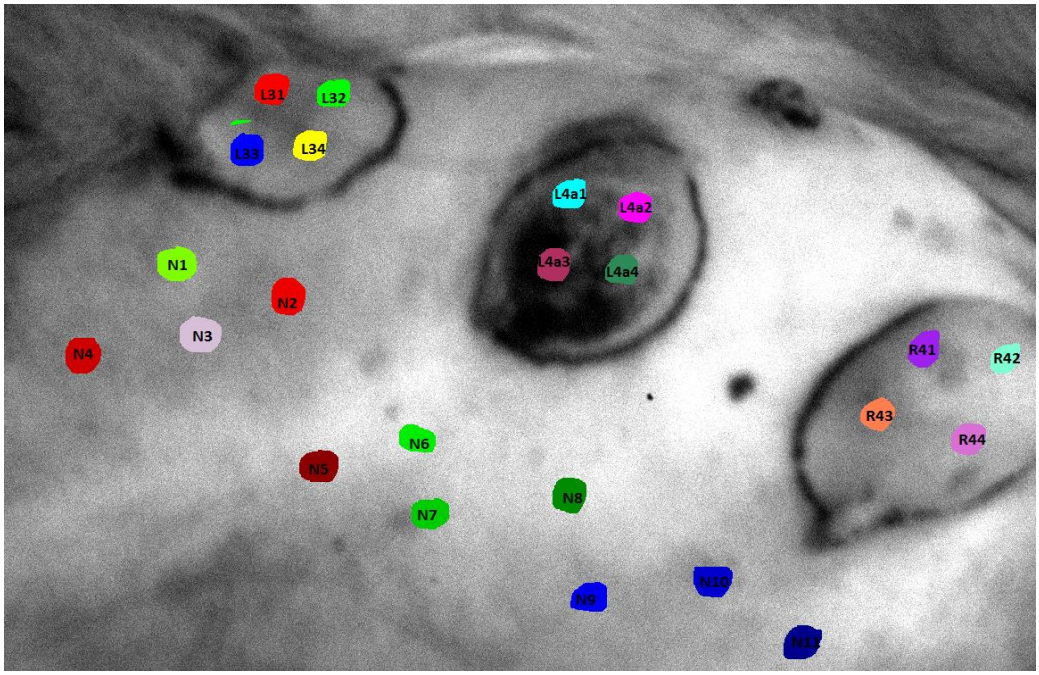


Figure 50: Multiple ROIs (4 in each tumor) taken to increase the overall number of ROIs.

Results:

Equation 11 was applied on the multiple ROI spectra. Again taking 2.00 as the threshold, the sensitivity and specificity in this case was 47% and 56% respectively. So applying Tissue Optical Index on multiple ROI data does not work as well as applied TOI model on whole tumor ROIs.

4.3.6.2 PCA-LDA model and cross validation

In this model, we applied Principal Component Analysis and Linear Discriminant Analysis on the smoothed and normalized reflectance canine spectral dataset. Then we used a k-fold validation method to validate the model.

Validation of a model means testing the performance of the model according to an a priori given set of test result specifications. For example, in prediction model validation, validation testing is concerned with its prediction ability on a new data set, which has not been used in the development of the model.

In this case we have in all 44 Region of interests. So we do not have enough data points to construct separate training set and testing set. Because of that, we would use a k-fold cross validation. In k-fold cross-validation, we make as many sub-models as there are objects, each time leaving out just one of the objects and only using this for testing. In this case, there are 44 ROIs, each sub-model will thus be made on 43 samples, and the remaining sample would be used for cross validation. In this case we would have a 44-fold cross validation.

For cross validating the TOI model, we applied Principal Component Analysis on the smoothed and normalized reflectance spectra. Then Linear Discriminant Analysis was applied on the PC data. Since LDA is a supervised classifier, we supplied the information if the tumor is malignant or not. This information we obtained from the TOI model. We applied a 44 fold cross validation method on our dataset. The details of the procedure of cross validation using 'Unscrambler 10.1' are described in the Appendix A.

Results:

Table 13: Results of a 44-fold cross validation using PCA-LDA.

Name of Patient	Label and Type of tumor	Correctly predicted using Cross validation
Kiki	R1 (benign)	Yes
	L3 (malignant)	Yes
	N1 (normal)	Yes
	N2 (normal)	Yes
	N3 (normal)	Yes
Tinkerbell	L1 (malignant)	Yes
	N1 (normal)	No
	N2 (normal)	Yes
	N3 (normal)	Yes
Queenie	R4 (malignant)	Yes
	N1 (normal)	No
	N2 (normal)	No
Lady	L3a (benign)	Yes
	L3b (malignant)	No
	L4 (benign)	Yes

	L5a (benign)	Yes
	L5b(malignant)	Yes
	R5a(benign)	Yes
Chelsey	R2 (benign)	Yes
	R5 (benign)	Yes
	N1 (normal)	Yes
	N2(normal)	Yes
	N3(normal)	Yes
Gretchen	R4 (benign)	Yes
	N1 (normal)	Yes
	N2(normal)	Yes
	N3(normal)	Yes
Suzanna	L3a (benign)	No
	L3b (benign)	No
	L4 (benign)	Yes
	R4 (benign)	Yes
	N1 (normal)	Yes
	N2(normal)	Yes
	N3(normal)	Yes
Maxine	L3 (malignant)	Yes
	L4a (benign)	Yes
	R4 (malignant)	Yes
	N1(normal)	Yes
	N2(normal)	Yes
	N3(normal)	Yes
Name of Patient	Label and Type of tumor	
Hanna	L5a (benign)	Yes
	L5b(benign)	Yes
	N1(normal)	Yes
	N2(normal)	Yes

The sensitivity and specificity is 86% and 86 % respectively.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Discussion

In this study, we have used a hyperspectral imaging system for canine mammary tumor characterization. Within a single canine patient, we were able to effectively differentiate the malignant and benign spectra based on the reflectance intensities. Spectra of tumors had lower reflectance intensities than benign spectra over the NIR spectral range of 650-1100 nm. Malignant tumors exhibit increased metabolic activity and thus the tissue concentration of oxy-hemoglobin and deoxy-hemoglobin is greater in malignant tumors, which results in lower reflectance intensities values relative to benign tissue, especially in the 650-850 nm range.

There are several issues to be resolved before routine hyperspectral imaging of the mammary tumors of canine patients is possible. The lighting of the sample should be as uniform as possible. In this study 500W white tungsten halogen lights were used (Figure 51). However, the temperature of these lights was very high and it was therefore uncomfortable for the patient to be under such lighting for the 4 to 5 minutes required to collect an image. In future experiments, we will investigate alternative lighting for example, the use of fiber optic cabling (Figure 52). By using fiber optic cabling to deliver light to the sample we should observe a reduction of the sample surface and have a more flexible experimental setup.

Figure 51: The 500W quartz tungsten halogen lamp used in the canine experiment.

Figure 52: The 300W Sunbeam Xenon Lightsource to which a fiber optic cable can be attached.

The TOI method was effective inspite of the non-uniform skin color of the dogs. The use of a ratio of the reflectance intesnity of certain wavelengths to calculate the TOI mitigates any effect skin color may have.

Normalization of the data is also an important issue. Using the reflectance standards to normalize the data is the most common approach. The images of the reflectance standards should be taken under similar conditions and at the same time as the sample is imaged. However, during our data collection procedure, we did not use the standards. Therefore, we will investigate the best way to normalize the data using one of the software normalization methods. But in future experiments, we should definitely use the reflectance standards to normalize the data.

Taking the Region of Interest (ROI) as the whole tumor worked much better than taking multiple ROIs in the same tumor, as seen from the sensitivity and specificity results. This would mean that the reflectance spectra does not vary uniformly in the tumor. One of the resons of this might be non-uniform lighting. In future experiemnts, care should be taken to ensure that the lighting is as uniform on the sample as possible.

This evolving imaging modality may have many potential applications in the medical field. Early diagnosis and treatment of cancer leads to a better prognosis and a much greater chance of full recovery. HSI can be used not only for diagnosis, but also to determine the tumor margin during surgery.

Table 14: Sensitivity and specificity of TOI and PCA-LDA method applied for the canine spectral dataset compared to the results of other algorithms used for different types of cancer.

Type of cancer	Sensitivity and specificity	Method Used	Sample Size	Is the test set independent?	Invasive?
Prostate	93% and 97%	Support vector Machine	251 ROIs from 11 mice tumors	Yes	No
Gastric	93% and 91%	Normalized Difference Cancer Index	101 ROIs from 10 tumors	Yes	Yes
Tongue	93% and 91%	Sparse Representation	34 tumors	Yes	No
Canine Mammary Cancer	86% and 95% (TOI) 86% and 86% (PCA-LDA)	Tissue Optical Index PCA-LDA	44 ROIs from 22 tumors	No (cross validation used)	No

Preliminary results with 22 canine mammary tumors showed that the sensitivity and specificity of the TOI method was 86 % and 95% respectively. The sensitivity and specificity of the TOI method was 86% and 86% respectively. This is comparable with the results obtained by researchers using the same technology, but for other types of cancer. However, their test set was independent. In case of prostate cancer, the authors used prostate tumors in mice. Mice is a

simpler model to test on than dogs. Also in case of gastric cancer, surgery was performed and then the stomach sample was imaged. So the sample did not have skin. Considering all these factors, we can say that the sensitivity and specificity obtained for identifying malignant and benign canine mammary tumors respectively, in this thesis , is comparable to that of results in literature [3][4][5].

5.2 Conclusion

We used a hyperspectral imaging system to characterize malignant and benign canine mammary tumors. A 'Tissue Optical Index' was developed to classify canine cancer. Preliminary results with 22 canine mammary tumors showed that the sensitivity and specificity of the proposed method was 85.7% and 94.6% respectively. This method worked when we took the whole tumor as the Region of Interest (ROI). But when multiple ROIs were taken in a tumor, the method did not work as well. The sensitivity and specificity of the method was 47% and 56% respectively. We also applied PCA-LDA algorithm on the canine spectral dataset. After cross validation of the method the sensitivity and specificity came out to be 85.71% and 86.48 %. The hyperpsectral system that we used was found to be fairly repeatable. Also the depth of penetration of NIR light over chicken breast tissue was quantified to be between 3 to 5mm. We found out that varying the sample to camera distance does not affect the reflectance spectra using neoprene rubber as the sample. HSI offers a potential non-invasive tool that allows surgeons to inspect and assess a large area of tissue without having to take any tissue samples for pathology examinations. An advantage of this technique is its capability to spatially constrain the spectral variations of different tissue types. This study shows promise in the optical diagnosis of canine mammary cancer.

5.3 Future Work

The current study can be used for early detection of the mammary cancerous tumor in canine population barring a few limitations faced with the representative quality of the dataset. Further work needs to be done to collect a much more comprehensive dataset spread across the canine breeds and distributed demographic to be able to generalize an application for the predictions put forward by the current study. Secondly, since the dataset being used currently is collected primarily for surface tumors, the reflectance value recorded by the hyperspectral imaging system is directly used as the reflectance value of the tumor without adjusting for any aberrations resulting from skin depth. These aberrations can possibly induce an error in the device recorded reflectance value in skin-deep tumor detection. We can use the image of the resected tumor cells after biopsy for an *in-vitro* spectral imaging study, measure the depth of the overlaying skin and the compare the difference in the reflectance value to study the role of skin depth and adjust for any possible aberrations.

Standard reflectance should be used to normalize the raw data in future experiments. Also, in future we should design an algorithm that would give us the result in real-time. That would help surgeons during operations to identify the tumor margin. Also, the following tasks should be accomplished in future:

1. For the depth of penetration experiment, a literature search is to be conducted on materials that would have a sharp peak between 650 – 1100 nm, and use that material instead of black neoprene rubber sheet.
2. For depth of penetration experiment, use ‘no neoprene’ as a control mechanism. This means that when imaging the neoprene with the chicken sample, an image should be taken without the neoprene as well.
3. Find out if there are any other chromophores that can be identified in the range 650 – 1100 nm. (example calcification chromophore).

4. Investigate if we could differentiate between normal tissue regions and benign tumors.
5. Machine Learning and Pattern Recognition algorithms such as Random Forest should be investigated to see if it would work better than just using TOI.

REFERENCES

[1] The Merck Veterinary Manual, "Mammary Tumors: Introduction", 2006 < http://www.merckmanuals.com/vet/reproductive_system/mammary_tumors/overview_of_mammary_tumors.html?qt=&sc=&alt=> (July 2011) http://www.merckmanuals.com/vet/reproductive_system/mammary_tumors/overview_of_mammary_tumors.html?qt=&sc=&alt=

[2] Merlo, D. F., Rossi, L., Pellegrino, C., Ceppi, M., Cardellino, U., Cappuro, C., Ratto, A., Sambucco, P.L., Sestito, V., Tanara, G., and Bocchini, V., "Cancer Incidence in Pet Dogs: Findings of the Animal Tumor Registry of Genoa, Italy", Journal of Veterinary Internal Medicine, 22(4), 976-984 (2008).

[3] Akbari, H., Halig V. L., Schuster, D. M., Osunkoya, A., Master, V., Nieh, P. T., Chen, G. Z., Fei, B., "Hyperspectral imaging and quantitative analysis for prostate cancer detection", Journal of Biomedical Optics, 17(7), 076005 (2012).

[4] Akbari, H., Kuniaki U., Kosugi, Y., Kojima, K., Tanaka, N., "Cancer detection using infrared hyperspectral imaging", Cancer Science, 102(4), 852-857 (2011).

[5] Liu, Z., Wang, H., Li, Q., "Tongue Tumor Detection in Medical Hyperspectral Images", Sensors, 12(1), 162-174 (2012).

[6] Medina, J. M., Pereira, L. M., Correia, H.T., Nascimento, S. M. C., "Hyperspectral optical imaging of human iris in vivo: characteristics of reflectance spectra", Journal of Biomedical Optics, 16(7), 076001 (2011).

[7] Panasyuk, S. V., Yang, S., Faller, D. V., Ngo, D., Lew, R. A., Freeman, J. E., Rogers, A. E., "Medical hyperspectral imaging to facilitate residual tumor identification during surgery", Cancer Biology and Therapy, 6(3), 439-46 (2007).

[8] Akbari, H., Kosugi, Y., Kojima, K., Tanaka, N., "Detection and Analysis of The Ischemia Using Visible and Invisible Hyperspectral Imaging", IEEE Transactions of Biomedical Engineering, 57(8), 2011-2017, (2010).

[9] Lee, J., Won, C. H., "Characterization of Lung Tissues using Liquid-Crystal Tunable Filter and hyperspectral Imaging System," Proc. IEEE EMBC 09, 1416-1419 (2009).

[10] Balas, C., Themelis, G., Papadakis, A., Vasgiouraki, E., Argyros, A., Koumantakis, E., Tosca, A., Helidonis, E., "A Novel Hyper-Spectral Imaging System : Application on in-vivo Detection and Grading of Cervical Precancers and of Pigmented Skin Lesions", In Proc. of "Computer Vision Beyond the Visible Spectrum" CVBVS'01 Workshop, Hawaii, USA, (2001).

[11] Fantini, S., and Sassaroli, A., "Near-Infrared Optical Mammography for Breast Cancer Detection with Intrinsic Contrast", Annals of Biomedical Engineering, 40(2), 398-407 (2011).

[12] Shah, N., Cerussi, A. E., Jakubowski, D., Hsiang, D., Butler, J., and Tromberg, B. J., "The role of diffuse optical spectroscopy in the clinical management of breast cancer" , Disease Markers 19, 95–105 (2003).

[13] Gurfinkel, M., Thompson, A. B., Ralston, W., Troy, T. L., Moore, A. L., Moore, T. A., Gust, J. D., Tatman, D., Reynolds, J. S., Muggenburg, B., Nikula, K., Pandey, R., Mayer, R. H., Hawrysz, D. J. and Sevick-Muraca, E. M., "Pharmacokinetics of ICG and HPPH-car for the Detection of Normal and Tumor Tissue Using Fluorescence, Near-Infrared Reflectance Imaging: A Case Study", Photochemistry and Photobiology, 72(1), 94-102 (2000).

[14] Rinnan, A., Berg, F., and Engelsen, S., "Review of the most common pre-processing techniques for near-infrared spectra", Trends in Analytical Chemistry, 28(10), 1201-1222 (2009).

[15] A.F. H. Goetz, "Three decades of hyperspectral remote sensing of the earth: a personal view," Remote Sens. Environ. 113 (Suppl. 1), S5-S16, 2009

[16] Aurora Health Care, Fine Needle Biopsy, http://www.aurorahealthcare.org/yourhealth/healthgate/getcontent.asp?URLhealthgate=39734.html.

[17] David Bannon,"Hyperspectral Imaging: Cubes and Slices," Nature Photonics, 3, 627-629 (2009).

[18] Lydia E. Kavraki, "Dimensions Reduction Method for Molecular Motion", < http://cnx.org/content/m11461/latest/>

[19] The Unscrambler 10.1 Manual < http://www.camo.com/rt/Products/Unscrambler/unscrambler.html>

APPENDIX A

Steps to cross validate the data using LDA and PCA in Unscrambler 10.1

1. First we have to normalize the data. To do this

Task>Transform>Normalize

2. Select Maximum normalization . We have to make sure the right column range is selected, i.e. 650-1100.

3. Now we need to work out how many PCs we are going to use in LDA model. To complete PCA go Tasks>Analzye>Principal Components Analysis

4. The following is the display of PCA results. We have to use this to select the number of PCs to use.

5. To calculate LDA model Tasks>transform>Linear Discriminant Analysis

6. We have to make sure the right matrix is selected for descriptors and classification. Select the columns (i.e. the wavlengths) to be used to base the model on and the column with the class identify in it.

7. Now for the cross validation aspect, we have to exclude one sample. To do this click on define and then put in spectrum number in rows of Keep Out. In the example below spectrum 1 should be excluded.

8. In the options tab of LDA dialogue box make sure the right number of components is selected.

9. Once we have calculated the LDA model we can use it to predict the class of the sample we left out.

10. Select the LDA model just calculated. Make sure the right matrix and columns are selected.

11. We have to exclude all the samples used to calculate the model so click on define and in the keep out rows now enter all the samples used in the model. In this example spectrum 1 from excluded from the model calculation so keep out rows should read 2-44. If for example spectrum 5 was excluded then the keep out rows for the prediction step would be 1-4, 6-44.

12. Once we have run the prediction you will get the following results. Record what group the spectrum was assigned to.

13. Need to go through this 44 times, excluding one sample at a time. By the end we should have predicted classes for each of the spectra in the data set. From this list calculate correctly sensitivity, specificity etc.