The Dissertation Committee for Chih-wen Kan
certifies that this is the approved version of the following dissertation:

Optical Reflectance Spectroscopy for Cancer Diagnosis: Analysis and Modeling

Committee:

Mia K. Markey, Supervisor

Konstantin V. Sokolov, Supervisor

Alan C. Bovik

Andrew K. Dunn

Linda T. Nieman

Optical Reflectance Spectroscopy for Cancer Diagnosis: Analysis and Modeling

by

Chih-wen Kan, B.S.; M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2010

Acknowledgments

I am deeply appreciative for the direction and support of my advisors, Dr. Mia K. Markey and Dr. Konstantin "Kostia" Sokolov. I am extremely fortunate to have had the chance to learn from Mia, not only in the professional realm but also personally. I could not have finished my degree without her persistent inspiration and motivation. During my time as a graduate student, I was given the enviable freedom to explore multiple areas of research. For that, I am ever thankful. I am also grateful for the support that Kostia provided. His comments were always sharp, concise, and insightful.

I thank all the members in my dissertation committee: Dr. Al Bovik, Dr. Andrew Dunn, and Dr. Linda Nieman for their valuable suggestions on my project. My work included collaboration with Linda, from whom I gleaned much of my knowledge in optics. I also would like to thank Dr. Daniel Côté, who I collaborated with on the Monte Carlo project. His generous sharing of his Pol-MC code and his technical guidance led me past many difficulties, and the time I spent in his lab was simply wonderful.

I want to extend my thanks to all labmates at the Biomedical Informatics Laboratory and the Biomedical Optics and Nanodiagnostics Lab. Shalini, who was both my role model and dear friend, was a large influence in my life. I also wish to give my appreciation to Kort, who committed numerous

hours on the Monte Carlo project; working with him was not only enjoyable, but enlightening. I had a wonderful time working with Jin, Mehul, Gautam, Juhun, Clement, Shuang, Justina, Tim, Ryna, Pratixa, Bryan, Harry, Andy, Alfredo, and James, all brilliant people and great friends.

I desire to thank my friends, Stacy and Ya-lan, who in essence were my family in Austin; volleyball friends, Chris, Mei-i, Chung-hsiang, Ferdinand, Patrick, Kuei, Kim, Ashwin, Maggie, for all the laughter and good times. I would also like to thank Mike for his support and always keeping me motivated in my studies.

Last, but the most important, I thank my dear father, mother, and little brother, Jerry, for their love and support. I am proud of myself because they are proud of me.

Optical Reflectance Spectroscopy for Cancer Diagnosis: Analysis and Modeling

Publication No. _____

Chih-wen Kan, Ph.D.
The University of Texas at Austin, 2010

Supervisors: Mia K. Markey
Konstantin V. Sokolov

This dissertation focuses on the development of algorithms for analyzing and modeling of the signals from optical spectroscopy. This dissertation is motivated by the detection of oral cancer, but some of the methods developed can be generalized to epithelial cancers of other sites.

Two main topics are covered in this dissertation: Analysis and Modeling. For analysis, the focus is on developing algorithms to make diagnostic predictions. The analysis methods are empirically tested using an oral cancer dataset. Statistical analyses show that polarized reflectance spectroscopy has the potential to aid screening and diagnosis of oral cancer. Also, a novel adaptive windowing technique is developed to extract spectral features with fewer windows that retain the diagnostic information. For modeling, a Monte Carlo model simulating light-tissue interactions is presented to aid in the design of diagnostic instrumentation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Optical approaches have been studied for the detection and diagnosis of epithelial cancer. Due to the biochemical and structural changes that occur in cancerous cells, malignant, benign, and normal tissues have different optical properties.

In this dissertation, the two main focuses are the analysis and modeling of polarized reflectance spectroscopy. This chapter reviews the background of applying computation technology to optical spectroscopy for cancer detection and diagnosis. This chapter is an updated version of my publication "AI in Clinical Decision Support: Applications in Optical Spectroscopy for Cancer Detection and Diagnosis" in Advanced Computational Intelligence Paradigms in Healthcare (Editors: M. Sordo, *et al.*), 2008 [2].

1.1 Optical spectroscopy in cancer detection

Cancer is the second leading cause of death in the United States, exceeded only by heart disease [3]. Early detection and treatment of cancer is essential to improve the survival rate. Among diagnostic modalities, optical methods stand out since they employ non-ionizing radiation, are non-invasive,

and the equipment is moderate in cost.

Optical instrumentation provides information at biochemical, structural, and physiological levels for clinical decision making. Over the past few years, the term "optical biopsy" has been widely used in biomedical optics. "Optical biopsy" is commonly used to describe the idea that a non-invasive optical device could augment biopsy, the removal of tissue for analysis. Alternatively, an optical exam could be used as a screening step to reduce the number of benign biopsies. In other words, optical biopsy would be performed first and only those sites that are positive for disease according to the optical method would be subjected to biopsy.

A review of research on the interaction of light with tissue for disease detection and diagnosis was presented by Richards-Kortum and Sevick-Muraca [4]. As early as 1965, studies revealed the potential to use quantitative fluorescence spectroscopy for discriminating normal and malignant tissues [5]. In the past few decades, rapid developments of small light sources, detectors, and fiber optic probes have provided opportunities to quantitatively measure light-tissue interactions. Progress using fluorescence, reflectance, and Raman spectroscopy for cancer detection was reviewed in several recent articles [1, 6–20]. Whereas previous reviews have emphasized the underlying biophysical processes and the design of instrumentation to measure them, this chapter is focused on approaches to analyzing optical spectra in support of clinical decision making.

What is optical spectroscopy?

A typical optical fiber based spectroscopic experiment is shown in Figure 1.1. Conceptually it is very simple: an optical fiber delivers light to the tissue region of interest and a second optical fiber collects the remitted photons. The tissue can be illuminated with a single wavelength or a range of wavelengths, such as broad-band visible light. Photons interact with tissue through scattering or absorption. In the visible wavelength range, elastic scattering is often the dominant form of optical-tissue interaction used for cancer detection [21]. Figure 1.2 shows an example of a diagnostic system set-up. The light generated by the excitation source is coupled into a fiber-optic probe for delivery to the tissue. The collected photons are spectrally dispersed by a spectrometer and the spectrum, which is a plot of intensity as a function of wavelength, is measured by a camera or other light detector. The information is read out and analyzed using the computer and interface electronics.

Since cancer cells and normal cells have different morphologies, their optical properties are similarly different (Figure 1.3). Changes in architectural and morphological features as a result of cancer progression, including increased nuclear/cytoplasmic ratio, hyperchromasia, and pleomorphism, affect the nature of the scattering events when light interacts with the tissue. Therefore, these changes complicate the interpretation of spectra as they relate to tissue disease status. For example, cancer cells are well documented to have increased nuclear size, and decreased cell differentiation [22] due to the

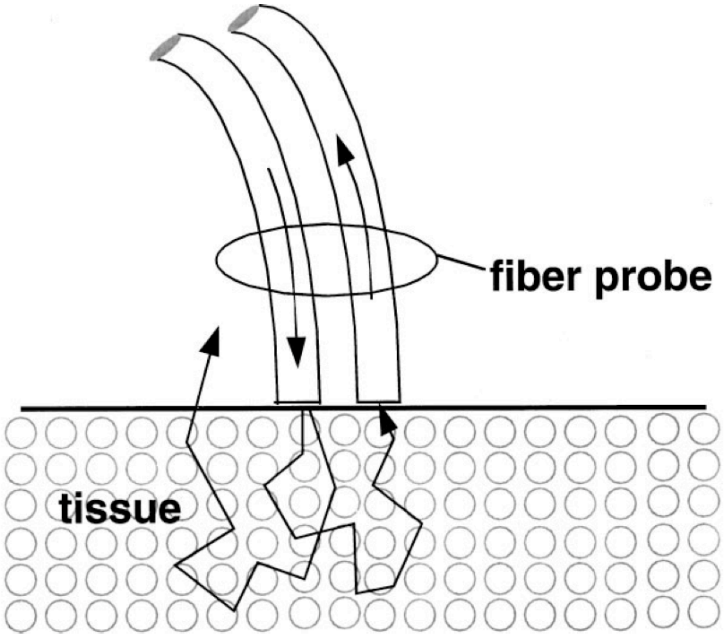


Figure 1.1: Optical geometry of the fiber-optic probe. The light is delivered into the tissue from an optical fiber probe and interacts with tissue. The remitted light is then collected by another optical fiber for analysis. This figure is reproduced with permission from [1] (©2007 SPIE)

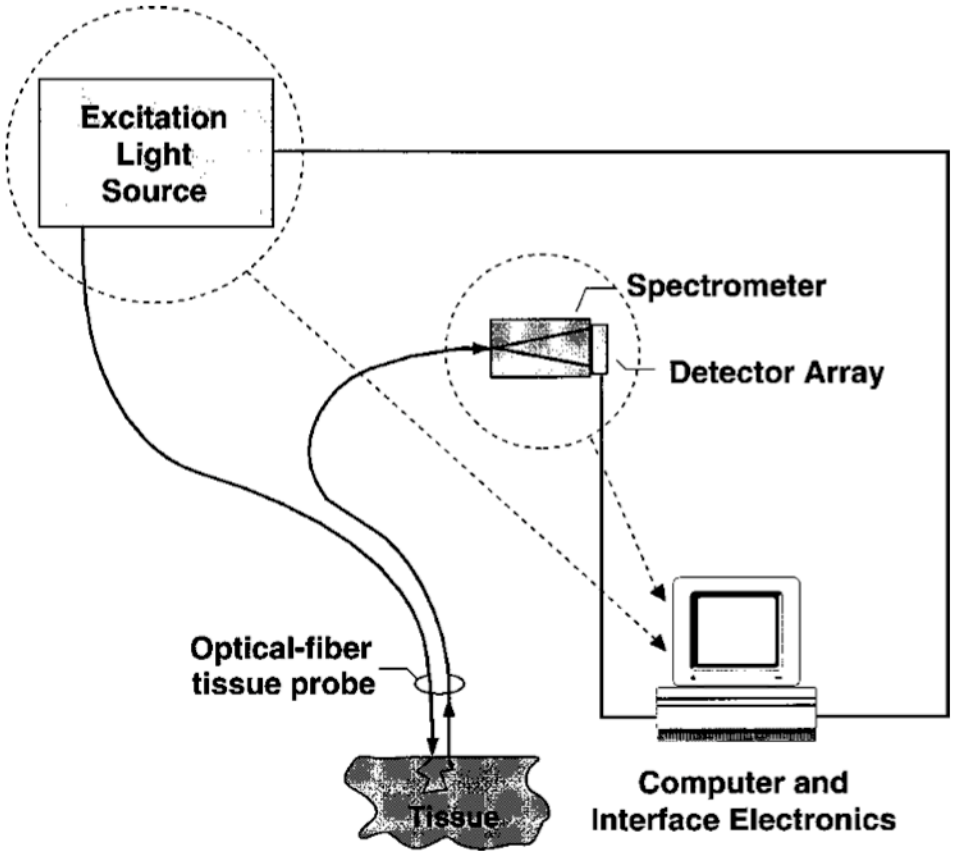


Figure 1.2: A diagram of a diagnostic system setup. The excitation light source generates photons to impinge on tissue with an optical fiber probe, and the light that leaves the tissue is collected by another fiber and is spectrally dispersed by a spectrometer. The computer and interface electronics control the light source and the spectrometer. This figure is reproduced with permission from [1] (©2007 SPIE)

abnormal duplication of cancerous cells. Larger nuclei result in more backward scattering events, which leads to the collection of more light at the optical fiber [23]. Therefore, the light intensities can be used, to some extent, as features for diagnostic decisions.

The spectral distribution may also be altered by the concentration of hemoglobin. Initially, non-necrotic cancerous tissues contain a higher concentration of hemoglobin due to the increased consumption of more nutrients than normal cells. In Figure 1.4, the extinction coefficients of oxy-hemoglobin and deoxy-hemoglobin are shown. The extinction coefficient is an important optical property as it determines how much light can be absorbed. Clearly, tissues containing different oxy- and deoxy-hemoglobin concentrations will have different spectra.

1.2 AI methods used for spectral classification

Artificial intelligence (AI) is a key technology for developing clinical decision support systems. AI uses computational techniques to achieve goals that require intelligence. AI methods are used to extract features and to select the features that are most indicative of cancer status in order to classify spectral patterns. In the following sections, we review the general steps taken to develop a clinical decision support system in the context of analyzing optical spectroscopy data. First, preprocessing techniques are applied to the optical spectra to reduce the influences of noise and artifacts and to normalize the spectra of different samples. Second, features are extracted from the spectra

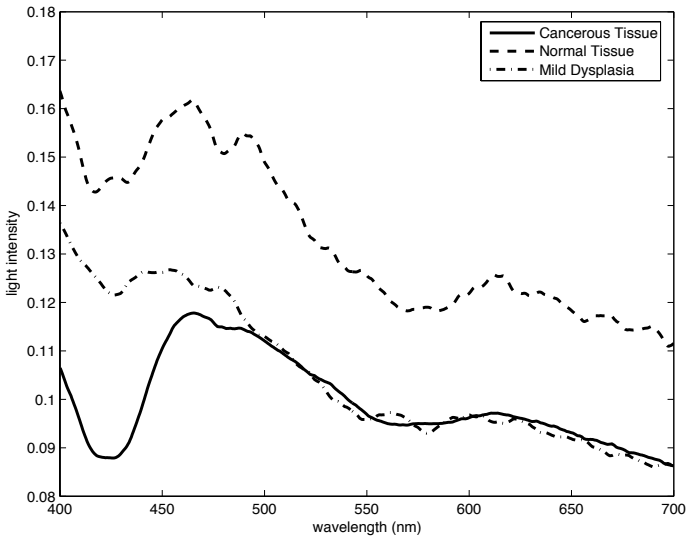


Figure 1.3: Example diffuse reflectance spectra for normal and malignant sites in oral tissue. The solid curve indicates cancerous oral tissue, while the dash-dotted curve indicates precancerous tissue, and the dashed curve represents normal tissue.

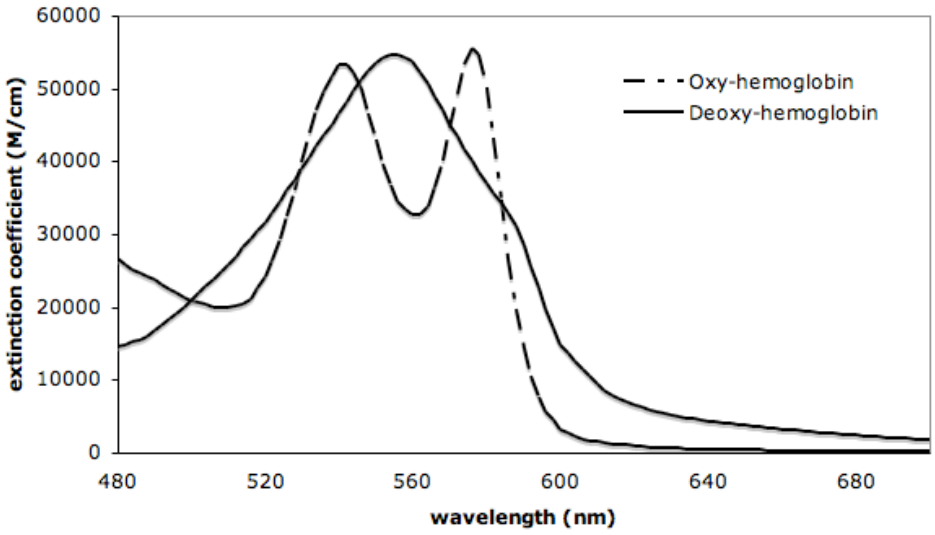


Figure 1.4: Absorption extinction coefficient of oxy-hemoglobin (dashed curve) and deoxy-hemoglobin (solid curve). The absorption extinction coefficient determines how much light can be absorbed by the object. These two spectra indicate that oxy-hemoglobin absorbs more light at 580 nm and deoxy-hemoglobin absorbs more at 560 nm.

to summarize the key information content. Then, feature selection is performed to reduce the number of redundant and irrelevant features. After that, classifiers are trained to distinguish between different histopathology groups. The choice of evaluation metric and testing paradigm is critical to accurately evaluate the performance of a clinical decision support system.

1.2.1 Preprocessing

Preprocessing techniques are needed since real-world data are often noisy, missing, or inconsistent [24]. Biomedical signals are especially notorious in this regard. Hence, it is desirable to have preprocessing techniques to improve the quality of data and the efficiency of data analysis. There are two broad types of preprocessing techniques. *Data cleaning* refers to the removal of noise and imputation of missing data. *Data normalization* transforms the data such that they are standardized. Data normalization may significantly improve the accuracy and efficiency of data analysis [24].

There are several factors that influence optical spectra. Some of these factors are critical for making clinical decisions, such as the structure of the cells and the hemoglobin concentrations, while other factors are not related to disease status. For example, fluctuation of the light source and light detector errors both cause slight changes to the measured optical spectra.

Normalization is a preprocessing step that standardizes the spectral data and enhances spectral features that distinguish between different histopathologies. The simplest approach assumes that the error is uniform in each

measurement. In other words, it assumes the same intensity of error at all wavelengths. In this case, the most common normalization method is to divide the spectrum by the baseline intensity of a specific wavelength. The choice of baseline wavelength is often at the extreme ends of a spectrum - either at a short or long wavelength [25], i.e., wavelengths that are outside of the range that usually interact with tissue. Another strategy is to normalize each spectrum by its highest intensity [26]. Normalizing by the peak intensity standardizes the spectra and makes the highest point the same in every spectrum. Unlike normalizing by an extreme wavelength, this method does not make any assumptions of the wavelength ranges that usually interact with tissue. Instead, it chooses the peak intensity to be where the most scattering or absorbing events occur. In a similar manner, one can normalize each spectrum by its lowest intensity, the wavelength at which the least scattering or absorbing events occur. Other researchers have normalized the spectra such that each spectrum has the same area under the plot [26]. Since the area under a spectrum can be viewed as the total energy received by the light collecting probe, this method assumes that the total energy received remains constant and uses this property to reduce the variability between different measurements [26].

Normalization is also key for effective visualization of spectra by clinicians. Noise and artifacts in the spectra may influence the ability of humans to make clinical decisions. Therefore, it is desirable to visually enhance the spectral signatures. Normalization methods used to enhance visualization include smoothing and sharpening of optical spectra and spectral enhancing near a

specific wavelength. This is an interesting topic for future work since there are few previous studies addressing normalization for visualization purposes. Moreover, although the normalization goals are different between visualization by humans and analysis by computers, the normalization methods may sometimes be the same.

Other preprocessing methods are used to reduce the noise due to instrumentation, patient, or equipment operator influences. To eliminate the effect that is specific to the patient, a spectrum may be normalized by the spectrum recorded from the contralateral position in the same patient, for example, the breast on the opposite side of the patient [26]. However, a contralateral position measurement is not always available for asymmetric organs or organs that have disease throughout.

Instrumentation effects are composed of various spectral responses from different optical components. To reduce the impact of these interferences, different noise models have been developed [27, 28]. Fawzy *et al.* [27] designed an experiment that measured the true tissue response and the signal reflected from a standard disc for reference. Two linear models were developed to account for the instrument's spectral response.

$$I_{m1}(\lambda) = a_1 I(\lambda) + b_1 I(\lambda) R_{tm}(\lambda) \tag{1.1}$$

where $I(\lambda)$ is the instrument's spectral response and $R_{tm}(\lambda)$ is the true tissue diffuse reflectance to be derived. a_1 and b_1 are the weights of the instrumen-

tation spectral response and that of the true tissue reflected response, respectively. The signal measured from tissue $I_{m1}(\lambda)$ is divided by the reflectance signal measured from a reflectance standard disc to account for instrument spectral response. The reflectance signal measured from the standard disc is

$$I_{m2}(\lambda) = a_2 I(\lambda) + b_2 I(\lambda) R_s \tag{1.2}$$

where $I(\lambda)$ is the instrument's spectral response, and R_s is the reflectivity of the standard disc, which is approximately constant across the whole visible wavelength range, and a_2 and b_2 are the weights of the instrumentation spectral response and the spectral response reflected from the standard disc, respectively.

Dividing Eq. 1.1 and Eq. 1.2 and rearranging the equation:

$$R_m(\lambda) = \frac{I_{m1}(\lambda)}{I_{m2}(\lambda)} = a_0 + b_0 R_{tm}(\lambda) \tag{1.3}$$

where $R_m(\lambda)$ is the apparent reflectance spectra measured, $R_{tm}(\lambda)$ is the true tissue diffuse reflectance to be derived, and a_0 and b_0 are the additive offset and multiplicative factor, respectively, which depend on the measurement conditions during each *in vivo* measurement. In Eq. 1.3, a_0 and b_0 can be measured by controlled calibration experiments, allowing the true diffuse reflectance, R_{tm}, to be determined. A study conducted by Marín *et al.* pointed out the importance of establishing calibration standards in fluorescence spectroscopy [29]. A consensus calibration standard enables meaningful comparison of data from multiple devices and unambiguous interpretations of experiments.

In the study of Fawzy *et al.*, the weights a_0 and b_0 in their noise model can be measured. In some other studies, the weights in the noise models are adjusted such that the effect of instrumentation noise is minimized. To achieve this goal, it is necessary to define a merit function that quantifies the effect of instrumentation noise. Ntziachristos *et al.* (2002) defined a merit function that was a summation of the tumor absorption coefficients and the absorption coefficients of background noise [28]. In other words, the merit function was the difference between the noise estimated by the model and the measured signals. This merit function was minimized by a χ^2 minimization technique. Since the noise components vary by different instruments, operators, and locations of the instrument [29], different noise models are needed for different situations. There exists no single noise model that fits all the cases.

1.2.2 Feature Extraction

Spectral feature extraction and selection are critical in designing clinical decision support systems. There are four major categories of feature extraction methods in spectral signal processing: principal component analysis (PCA), model-based feature extraction, spectral feature extraction, and hybrid feature extraction [2]. Quantitative features are extracted from optical spectra to describe different spectral patterns, while feature selection aids in the identification of those optically derived features that are diagnostically relevant and the elimination of redundant features that are strongly related to selected features.

1.2.2.1 Principal component analysis (PCA)

PCA is commonly used for feature extraction and reducing the dimensionality of data [1, 16, 25, 30–33]. PCA is a linear transformation of the data to a new coordinate system such that the data projection on the first principal component has the greatest variance. One can use PCA to reduce the dimensionality of the data by keeping the lower-order principal components and ignoring the higher-order ones. However, PCA transforms the original features into new features without direct physical or biological interpretation. Thus, PCA is not ideal for use in designing a clinical decision support system, in which it is desirable to make decisions based on features that can easily be directly related to the biophysics of the disease process. That being said, several studies have demonstrated the potential of PCA in analyzing optical spectra. Bard *et al.* applied PCA on reflectance spectra of neoplastic lesions of the bronchial tree [25]. In their study, they retained the first 10 principal components which captured 99.99% of the total variation of data. Several other studies also retained a pre-determined number of principal components [1, 16, 31, 32]. Other studies specified the percentage of the total variances. Mirabal *et al.* retained the principal components that account for 65%, 75%, 85%, and 95% of the total variance as their features [30]. Setting the number of PCs up front has the advantage of fixing the number of inputs to the classifiers being considered. However, the second strategy of selecting the number of PCs that account for a specified percentage of variance controls the amount of information contained in the features.

1.2.2.2 Spectral Features

In non-model based feature extraction, calculations are made based on the statistical properties of the spectrum. In optical spectroscopy, the spectral intensities themselves are informative since the intensities are related to how many photons are scattered or absorbed. The means of the spectra are commonly used to summarize the amount of light scattered from the tissue [1]. Furthermore, it is possible to divide each spectrum into several spectral bands and calculate the average spectral intensity of each band as a feature [1]. In addition to the first moment of the spectrum, higher order moments, such as the standard deviation, can also be used as features.

Mourant *et al.* (1995) conducted a study on spectroscopic diagnosis of bladder cancer with elastic light scattering [34]. They discovered that the spectral slopes over the wavelength range $330 - 370\ nm$ have positive values for nonmalignant tissues, and negative values for malignant ones. Bigio *et al.* also used spectral features from a spectral range ($330 - 590\ nm$) divided into 21 wavelength bands of 20 nm width. Both the average intensities and the slopes of each interval were calculated as spectral features [1].

In feature extraction, one must understand the biophysical properties that underlie optical phenomena. Also, it is well accepted that some wavelengths may be more discriminatory than others. This is because light of different wavelengths behaves differently when interacting with tissues. As the physical interpretation of light-tissue interactions is difficult, feature extraction may help us understand the underlying physics. The advantage of

feature extraction is to explicitly encode spectral information in a biophysically meaningful manner.

1.2.2.3 Model-based Features

Model-based features are extracted by building quantitative models of the tissue and inversely calculating the parameters in the model. These parameters contain important tissue information which may be indicative of the cancer status. For example, the sizes of scatterers in the tissue are typically larger in cancerous tissues; thus, the sizes of scatterers can be used as a feature indicative of cancer status. In these quantitative models, absorption coefficients and scattering coefficients are commonly inversely-determined from optical spectra. However, several other parameters such as tissue blood volume fractions or hemoglobin concentration can also be obtained from these models [16].

Obtaining parameters from quantitative models requires prior knowledge of the tissue and the light used in the system. That is, one must select the model that best approximates the properties of the tissue and the light. For example, the diffusion equation is valid for cases with low to moderate tissue absorption relative to scattering. A rule of thumb is that if

$$\mu_a << 3(1-g)\mu_s \tag{1.4}$$

then diffusion equation should be appropriate, where μ_a is the absorption coefficient $[1/m]$, μ_s is the scattering coefficient $[1/m]$, and g is the anisotropy

factor. Therefore, the diffusion equation is suitable for red light and near-infrared light systems where scattering dominates the light-tissue interaction [35].

Four models are commonly used to numerically and analytically extract the absorption coefficients from diffuse reflectance spectroscopy - radiative transfer theory (RTT), Monte Carlo (MC) modeling, empirical methods, and Mie theory. While the first three methods view light as a flow of photons traveling through a medium, the last method treats light as an electromagnetic wave. These modeling techniques are described below.

Radiative transfer theory (RTT) has been developed largely without reference to electromagnetic theory [41, 116]. It is based on the transfer of energy through a turbid medium. RTT has a stationary form of

$$\hat{s} \cdot \nabla I(\vec{r}, \hat{s}) + \mu_t(\vec{r}) I(\vec{r}, \hat{s}) = \mu_s(\vec{r}) \int_4 \pi p(\hat{s}, \hat{s}') I(\vec{r}, \hat{s}') d\vec{\omega}' + S(\vec{r}, \hat{s}) \qquad (1.5)$$

where $I(\vec{r}, \hat{s})$ is the radiance $[W/(m^2 \cdot sr)]$ along the $\hat{s}$ direction per unit solid angle and per unit area at location $\vec{r}$, μ_t is the total extinction coefficient $[1/m]$, μ_s is the scattering coefficient $[1/m]$, p is the scattering phase function $[1/sr]$, and S is the source term that corresponds to the power generated at $\vec{r}$ in the $\hat{s}$ direction $[W/(m^3 \cdot sr)]$. As there is no general solution available to this equation, an analytical approximation is used to model diffuse scattering. A common analytical approximation of RTT is called the diffusion equation.

Monte Carlo modeling provides a flexible approach to studying light propagation in biological tissues [36]. This method views light as particles and

traces three-dimensional random walks of photons in a medium. The two key parameters in Monte Carlo modeling are the scattering angle and the mean free path for a photon-tissue interaction. The mean free path is determined by the probability that a photon is scattered or absorbed after a given step size. This probability is determined by local optical properties: μ_s, the scattering coefficient, and μ_a, the absorption coefficient. Similarly, the scattering angle is determined by the anisotropy factor g. Monte Carlo modeling generates uniformly distributed random numbers and transforms them to follow the distributions of the mean free paths and scattering coefficients. These random attributes are then used to simulate and record the paths of photons. When modeling stratified tissues, multi-layer models are often needed. Each layer is described by its thickness, refractive index n, absorption coefficient μ_a, scattering coefficient μ_s, and anisotropy factor g [37, 38]. Internal reflection or refraction at the medium boundaries can also be simulated.

Another approach to inverse modeling is to use an empirical method. Empirical methods do not make explicit assumptions about the interaction of light and tissue. Instead, the idea is to use a classifier to "learn" the light-tissue interaction from experimental data. For example, Pfefer *et al.* [39] developed an empirical method for the extraction of absorption and scattering coefficients from diffuse reflectance spectra. They trained a neural network with phantoms and then used the network to extract optical parameters from another set of phantoms.

Mie theory is a complete analytical solution of Maxwell's equations for

the scattering of electromagnetic radiation by spherical particles [40–42]. It assumes a homogeneous, isotropic, and optically linear material irradiated by an infinitely extending plane wave. Mie theory assumes that the scatterers are spheres of arbitrary size and have a homogeneous index of refraction. Mie theory is often combined with the other three models for calculating the size of scatterers after obtaining the absorption and scattering coefficients of the tissue [19].

The choice of model depends on the system studied, although it is not uncommon for multiple quantitative models to be employed to elucidate the biophysics of the optical-tissue interaction. The main drawback of all these models, however, is that they either require a priori knowledge of optical parameters or they require simplifying assumptions, such as sphericity of scatterers, which may not be physically realistic. Despite their assumptions, these quantitative models provide valuable insight into the alterations of optical spectra with changing tissue state.

1.2.2.4 Hybrid features

The current trend leans more toward model-based feature extraction and hybrid feature extraction. Hybrid feature extraction, the fourth category of feature extraction method, is a mixture of statistical features and model-based features [12, 43]. These approaches utilize model-based preprocessing of spectra based on underlying biology to extract input parameters, and analyze these parameters with statistical methods.

1.2.3 Feature Selection

A large number of features, obtained from quantitative modeling or the measured spectra directly, can be potentially related to health status. Feature selection can aid identification of those optically derived features that are diagnostically relevant and those that are strongly related to each other and thus are redundant. Minimizing the number of features is important to reduce computation complexity, processing time, and to prevent overtraining of classifiers.

Since reflectance spectroscopy measures the combination of the elastic scattering from different organelles, it is not surprising that some wavelengths indicate cancer progression better than others. It is possible that one or several wavelengths contain information on the pathology status. Therefore, it is necessary to select subsets of these features to reduce the redundancy and improve the performance.

There are three main approaches to feature selection: *filters*, *wrappers*, and *embedded methods* [44]. Many feature selection algorithms use filters to select variables by giving ranks to individual features [44]. For example, Marín *et al.* (2005) selected their features using a combination of Principal Component Analysis (PCA) and a two-sample t test. They selected principal component scores (PCS) identified as statistically significant using a two-sample t test for independent samples, with equality of variance between the two groups (positive and negative for dysplasia) based upon an F test [45].

However, it is known that a feature which provides little information by itself can be valuable when combined with others [44]. Therefore, subsets of features can have better predictive power than would be expected by ranking of variables according to their individual predictive power. Thus, wrapper methods have been developed to select feature subsets rather than individual features. Wrapper methods use the prediction performance of a given classifier to assess the relative usefulness of subsets of features. If the number of features is not too large, an exhaustive search can also be considered. An example of a wrapper method is stepwise LDA (Linear Discriminant Analysis). It performs a greedy search of the possible feature combinations [46].

Embedded methods perform feature selection as part of the training process. Thus, they may be more efficient in the training process. For example, decision trees such as CART (Classification and Regression Trees) have a built-in mechanism to perform variable selection [47]. A study conducted by Atkinson *et al.* [48] used CART to analyze fluorescence spectra of suspected cervical intraepithelial neoplasia (CIN) at colposcopy.

1.2.4 Classification

There are two major types of machine learning algorithms: unsupervised learning and supervised learning. In unsupervised learning or clustering, the algorithm identifies clusters in the feature values based on criteria defining the desired properties of groups. Unsupervised learning techniques are useful for assessing the discriminatory power of features. The visualization of the

clustering distributions of features provides a qualitative evaluation of their potential for distinguishing between pre-defined categories, such as healthy vs. diseased [32]. In developing clinical decision support systems, supervised learning is more commonly used than unsupervised learning since the task is typically one of prediction, e.g., to predict disease status. Supervised learning refers to an algorithm that uses a training set of items for which target or truth labels are provided to learn a mapping from feature values to target values. In the case of cancer diagnosis, histopathological assessment of a biopsy sample is typically taken as the gold standard for establishing the truth state for classifier training and evaluation. In other words, the task of a supervised learning algorithm or classifier is to use the features provided by the feature extractor and selected in a feature selection step to predict the assignment of the sample to a diagnostic category [49]. Most clinical diagnostic decisions cannot be made based on a single feature; thus, classifiers play an important role in that they determine a function for combining two or more features to make predictions. The most commonly used classifiers for cancer detection/diagnosis from optical spectroscopy are LDA [50] and Artificial Neural Networks (ANN) [51, 52].

The difficulty of a classification task depends on the variability of the feature values within a class relative to the feature variability between classes. For example, suppose the size of a lesion is a feature that we can use to discriminate between benign and malignant lesions. All else being equal, it will be more difficult to distinguish between benign lesions with an average size of

1 mm and malignant lesions with an average size of 2 mm, as compared to the case in which the benign lesion mean size is 1 mm and the malignant lesion mean size is 10 mm. Likewise, for a fixed difference in the class means, the greater the variability within each class, the more challenging the classification task. Therefore, the underlying probability model of the categories determines the difficulty of the classification problem.

Some classifiers are more complex than others in the sense of the range of models that they can describe [49]. For example, a *support vector machine (SVM)* is capable of distinguishing samples by forming a non-linear function of the features, while a linear discriminant analysis (LDA) model is only able to solve linearly separable tasks. However, this does not mean that complex classifiers are always "better" than the simple ones. Complex classifiers suffer from generalization issues, i.e., overtraining. It is very easy to "tune" a complex classifier to the particular training samples, rather than to the real underlying characteristics of the diagnostic categories. A large training set will alleviate generalization issues, but assembling an extensive heterogeneous data set can be challenging. Thus, a simpler classifier may be used to avoid over-training given a limited amount of data. Another motivation for using linear classifiers is that they are more computationally efficient.

The choice of classifier depends on how much prior knowledge we have about the classification task. For example, if we have prior knowledge that the classification problem is linear, it will be most efficient to use a linear classifier rather than a non-linear one. However, in the case where we have

little or no prior knowledge about the problem, there is no simple answer as to how to choose the best classifier. The No Free Lunch theorem [53] states that if algorithm A outperforms algorithm B on some cost functions, there must exist exactly as many other functions where B outperforms A. In other words, if one algorithm seems to outperform another in a particular situation, it is because the algorithm is a better fit for that particular problem, not that that algorithm is generally "better" than the other one [49]. Thus, the selection of an algorithm for a practical classification task is an empirical choice because there is typically little prior knowledge of the underlying probability model.

1.2.5 Evaluation

To evaluate the performance of a clinical decision support system, we compare the diagnostic decisions suggested by our system to a gold standard, e.g., biopsy outcome. A performance measure, such as the accuracy, is computed to quantitatively summarize the efficacy of the system.

In two-class classification problems, Receiver Operating Characteristic (ROC) analysis is widely used for analyzing the classifier performance [54]. Sensitivity and specificity indicate the ability of the diagnostic method to distinguish between two groups, e.g., healthy and disease. By varying the threshold, a ROC curve of sensitivity versus (1-specificity) is generated. The area under the ROC curve (AUC) is often used as a metric to quantitatively summarize the performance of a clinical decision support system.

In contrast, for a multi-class classification task, there is not a widely ac-

cepted performance metric [55, 56]. Multiple research groups have been working on developing ROC-type analyses for multi-class problems. Several approaches have been proposed, so we briefly review only the most common and refer the reader to other resources for a broader summary [55, 56]. Hand and Till use the average AUCs of the binary one-versus-one comparisons in multi-class problems [57]. Mossman (1999) developed a three-way ROC method [58] that uses the correct classification rates as two separate decision thresholds are varied to form a 3-dimensional plot. The volume under the surface (VUS) of this plot serves as the performance metric. Moreover, Edwards *et al.* creates a ROC hypersurface, which is a two-dimensional plane in a six-dimensional space, and calculates the hypervolume under the hypersurface as a measure of performance [59]. We emphasize that no single metric has been widely adapted.

Generally speaking, three independent sample sets are desired to design and evaluate a classifier [60]. A training set is used for training a classifier. A validation set is used during or after classifier training, in order to adjust the classifier to prevent over-training. A test set is used for evaluating a classifier to report its performance. Ideally, these three sets should be randomly selected from the relevant population. However, there are practical issues to be considered. In particular, since the number of samples available is often limited, it may not be possible to construct truly independent sets, which causes bias. Therefore, sampling techniques such as *cross-validation* [61] and *bootstrap sampling* [62–64] are often necessary in experimental design.

Cross-validation refers to the partitioning of data into non-overlapping subsets of equal size such that the analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis [61]. The training/testing process is repeated until each partition has been used in turn as the testing partition. *Leave-one-out cross-validation* is a special case of cross-validation where one sample is assigned to the testing set and the rest are in the training set.

Bootstrap sampling is another technique that can be used to estimate the performance of a classifier. Bootstrap sampling creates a new set by sampling with replacement from the original set [62–64]. Typically, each bootstrap replicate has the same number of observations as the original sample. The process is repeated to create hundreds or thousands of bootstrap replicates. Then, a performance metric is computed from all of the bootstrap replicates and the average performance metric is taken as the estimate of the system performance.

The choice of appropriate evaluation methods is critical. Since clinical studies are often restricted to a small sample size, the analyses can suffer from biased data and little variation between samples [65]. Appropriate evaluation methods enable reliable estimation of system performance, which is essential for designing accurate and reliable clinical decision support systems.

1.3 Current challenges

It is difficult for a human to visually assess the optical spectra and make a diagnosis despite of the capability of optical spectroscopy for capturing cancer characteristics. The influence of cancer cell morphology on its optical properties is complicated, hence a clinical decision support system is necessary. This chapter shows that AI provides support for clinicians to make diagnostic decisions based on optical spectroscopy. Moreover, AI can also help scientists understand the optical properties of different tissue pathologies. For example, model-based feature extraction enables determining optical parameters of the measured tissue. This example shows how AI methods are used in data mining, which extracts useful information from large databases.

To develop a clinical decision support system for detection of epithelial cancer, it is important to understand the underlying physics. How the optical signals differ to capture cancer characteristics is the key to the correct diagnosis. We first analyze the signals from a pilot clinical trial dataset, then look deeper into modeling the light propagation through tissue. This dissertation covers two main topics: Analysis and Modeling. In Chapter 2, a dataset from a pilot clinical trial is presented. The analysis methods are empirically tested using an oral cancer dataset. Statistical analyses show that polarized reflectance spectroscopy has the potential to aid screening and diagnosis of oral cancer. In Chapter 3, a novel adaptive windowing technique is developed to extract spectral features with fewer windows that retain the diagnostic information. In Chapter 4, a Monte Carlo model simulating light-tissue

interactions is presented to aid in the design of diagnostic instrumentation.

Chapter 2

Probing local tissue changes in the oral cavity for early detection of cancer using oblique polarized reflectance spectroscopy: a pilot clinical trial

This chapter is modified from "Probing local tissue changes in the oral cavity for early detection of cancer using oblique polarized reflectance spectroscopy: a pilot clinical trial" published in Journal of Biomedical Optics, 2008 [43]. The instrumentation design was developed by Dr. Linda T. Nieman, and I contributed the statistical analyses development.

In this chapter, a feature set is extracted for early oral cancer detection using oblique polarized reflectance spectroscopy (OPRS). This feature set, consisting of the nuclear size and ten spectral features, is analyzed independently and in combination to assess discrimination of diagnostic class. The promising results suggest that OPRS has the potential to aid screening and diagnosis of oral precancer and cancer.

2.1 Oral cancer facts

Current oral cancer screening methods are limited by the variety of tissue architecture and by the similarity of appearance of benign inflammatory conditions to premalignant and malignant lesions. Further complications arise in high-risk patients, who often have carcinogenic exposure that covers the entire mucosal lining. After successful treatment of a cancerous tumor, secondary tumors can develop in adjacent areas over time. Monitoring these high-risk patients requires multiple biopsies taken routinely over many years. Clearly, noninvasive optical modalities to detect the early stages of oral cancer have the potential to reduce patient pain, morbidity, and mortality. To this end, significant effort have been directed toward exploring optical imaging and optical spectroscopy techniques to aid screening and diagnosis of the early stages of oral cancer [18, 26, 66–76]. A comprehensive overview of the new emerging methods for the detection and treatment of oral carcinoma has been given in references [77] and [78].

2.2 Overview for optical imaging and spectroscopy

As stated in previous section (1.1), optical imaging modalities such as confocal microscopy, optical coherence tomography (OCT), and nonlinear optical microscopy have shown great potential for oral lesion discrimination. While these nascent approaches alone or in combination with exogeneous dyes, vital stains, or nanoparticles are highly promising, they require some degree of equipment complexity and operator training. Autofluorescence imaging and

spectroscopy has been shown to give improved lesion contrast, which has been attributed in part to porphyrin fluorescence. However, it has been argued that porphyrin fluorescence is not a good diagnostic indicator, as it is synthesized by bacteria not only on ulcerating tumors, but also on the dorsums of normal tongues and on gingival plaques [69]. The presence of porphyrin fluorescence can therefore obscure detection of lesion grade using fluorescence techniques. Spectroscopic approaches that combine tissue autofluorescence and scattering showed promising sensitivity and specificity in pilot clinical trials, however, it required stratification of tissue sites according to level of keratinization before statistical analysis [18].

2.3 Specific aims

In this chapter, we present the results of the first pilot clinical trial that assesses the optical technique of oblique polarized reflectance spectroscopy (OPRS) to discriminate oral precancers and cancers from normal or benign tissue. OPRS is a noninvasive optical modality that employs polarized light illumination and polarization sensitive detection. This method is very simple and robust and, thus, provides an attractive approach for earlier cancer detection in the oral cavity including screening in a high risk population. OPRS is based on the following concept: the electric field orientation, or polarization, of the incident light remains unchanged after interaction with the optically dilute epithelial layer. In contrast, photons that propagate deeper to the optically dense stroma are remitted from tissue with their polarization state

randomized. In OPRS two scattered signals are collected: one with polarization parallel and the other with polarization perpendicular relative to the illumination polarization. The small epithelial signal can therefore be isolated by subtracting the perpendicular polarization intensity from the parallel polarization intensity [79–83]. Combining polarization sensitive detection with an oblique endoscopic collection geometry further resolves scattering signals from the upper epithelial layer and from the lower stromal layer [84]. This additional depth dependent optical information carried by polarized light has the potential to enhance discrimination of the varying grades of dysplasia and carcinoma from visually indistinguishable benign lesions.

The two scattering signals collected by OPRS can be used separately or in combination to yield diagnostically relevant parameters. In this pilot clinical study we extracted ten spectroscopic features and nuclear size of epithelial cells with the goal to identify key parameters for detection and monitoring of precancerous lesions in the oral cavity. A thorough statistical analysis of the spectral parameters and extracted nuclear size was performed using Linear Discriminant Analysis (LDA) (Section 1.2.4) and evaluated using Receiver Operating Characteristic (ROC) analysis (Section 1.2.5). OPRS was found to be sensitive to four clinically relevant histological groups: normal, benign, mild dysplasia, and severe dysplasia (defined as tissue requiring surgical excision for treatment). We demonstrated that the features that provide the best discrimination differ according to diagnostic category. This result emphasizes that a combination of features is needed to efficiently tackle the multi-task

problem of cancer detection and diagnoses.

2.4 Materials and Methods

2.4.1 Clinical Measurement

A pilot clinical study was conducted with informed consent on 27 patients over the age of 18 that were referred to the Head and Neck clinic at The University of Texas M. D. Anderson Cancer Center (MDACC) with oral mucosa lesions suspicious for dysplasia or carcinoma. A medical doctor performed a standard oral cavity examination, followed by spectroscopic measurements which were typically performed on 1 - 2 visually abnormal sites and 1 visually normal site. In some cases, more than one measurement was performed on the same tissue site without moving the probe to assess the repeatability of measurement, which was found to be good. Nuclear extraction was performed on 15 randomly selected patient sites that had more than one measurement. The average nuclear size difference between the measured spectra and the mean spectrum was 0.45 microns. All measurements from the same tissue site were averaged to give a single spectrum. A calibration spectrum was acquired before each patient evaluation using a diffuse reflectance substrate standard (Labsphere, Inc.). Data from three patients were removed from the analysis because of improper handling of the endoscope or malfunctioning of the clinical device. Biopsies were taken of all measured tissue sites. The biopsied tissue was sectioned into 4 micron transverse slices and mounted onto microscope slides. The slides were stained with a hematoxylin and eosin (H&E) stain for

standard histological analysis. Paired normal and abnormal slides were reviewed by a trained pathologist at MDACC. Detailed descriptions were made of each slide indicating the extent of dysplasia, inflammation, keratinization, and hyperplasia.

2.4.2 Instrumentation

Although detailed extensively in references [84] and [85], we briefly describe the instrument used in this clinical study for convenience. The illumination source was a broadband white light Xe pulsed lamp with ca. 4 μs pulse widths and wavelength range of 400-700 nm. Light was delivered to the tissue site of interest though a single optical fiber with a core diameter of 200 μm and 0.22 NA. The power delivered to the tissue was approximately 100 μW, well below the acceptable Threshold Limit Value given by the American Conference of Governmental Industrial Hygienists. Two identical optical fibers were placed on either side of the illumination fiber for collection of the remitted light. Two pieces of polarizing film with an extinction transmittance of 0.002% were adhered to the distal end of the fiber optic probe. These polarizers set the orientation of the illumination and collection polarization states and were oriented orthogonal to each other. The two collection fibers had polarizing film with transmission axes either parallel or orthogonal to the illumination polarization state. A protective fused silica window of defined thickness was placed over the fibers and polarizing film. The illumination fiber was oriented normal to the silica window surface and the collection fibers were oriented at

ca. 37 degrees with respect to the illumination fiber. With this geometry, the collection fiber's acceptance cones cross each other and with the illumination beam in the superficial tissue layer as shown in Figure 2.1. The collection efficiency of a single obliquely oriented beveled fiber peaks at the maximum overlap of the illumination beam with the collection acceptance cone. The probe used in this study has a fused silica window that acts as a spacer such that the maximum overlap is at the tissue surface. This overlap drops to 50% at a depth of 300-400 microns, the typical thickness of oral epithelium. Hence, superficial traveling photons approximately 0.07 mm^3 below the tissue surface are collected with greater efficiency than deeper traveling photons. The collection fibers deliver the remitted light to a grating spectrograph coupled to a gated intensified photodiode array detector. Operating in gated mode permits all measurement to be conducted under room light illumination so that OPRS measurements could blend seamlessly into a clinical examination or preoperative surgical procedure.

2.4.3 Data Analysis

2.4.3.1 Preprocessing

The collected parallel and perpendicular spectra were dark subtracted and then divided by the sum of the scattered light collected through both collection channels from a diffuse reflectance standard (Labsphere, Inc.) to correct for the wavelength-dependent response of the detection system and the spectral profile of the source. The parallel and perpendicular signals were

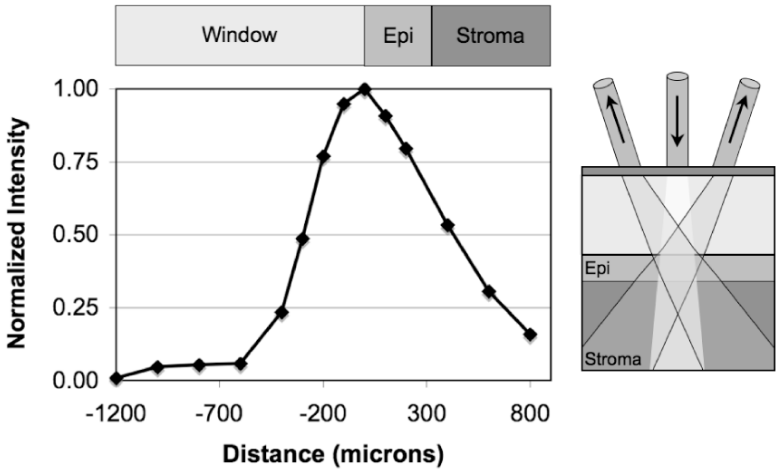


Figure 2.1: Overlap of illumination and collection acceptance cones as a function of distance from the distal end of the illumination and collection fibers. Distance zero corresponds to the epithelium surface. Measurements were made in water using a reflectance standard (Labsphere, Inc.)

studied alone and in the following combinations: the ratio of parallel to perpendicular, parallel minus perpendicular, and the sum of the parallel and perpendicular signals. The sum of the parallel and perpendicular spectra is equivalent to the diffuse reflectance spectrum and their difference is defined as the depolarization ratio. Typically, the majority of the diffuse reflectance is compromised of photons that have undergone many tissue interactions and therefore have probed more deeply, while the depolarization ratio is viewed as a measure of superficial scattering. Spectra were down-sampled using an averaging window with a spectral width of 5 nm to reduce data size and computation time.

The spectra were normalized to remove inter-patient variation. We re-

viewed various normalization methods in Section 1.2.1. Here, three approaches were tested: (1) no normalization, (2) division of entire spectra by the intensity value at 420 nm, and (3) additive DC offset applied to the entire spectra such that value at 420 nm was equal to the mean for all spectra within the same spectral type (e.g., parallel or ratio of parallel and perpendicular signals, etc.). This normalization was intended to preserve the relative intensity scale between different spectral types. Normalization method (2) gave the highest area under the ROC curve for all classification tasks considered and hence was used throughout this study.

2.4.3.2 Determination of the most discriminatory wavelength

The most discriminatory wavelengths were determined using the area under the ROC curve (method reviewed in Section 1.2.5). The area under the non-parametric curve (AUC) was computed using the trapezoid rule. On the occasion when ties existed, the best wavelength was chosen manually such that wavelength variations were minimized.

2.4.3.3 Features

We utilized a hybrid feature set in this study: one model-based feature and ten spectral features were extracted (Hybrid feature method is reviewed in Section 1.2.2). Two features were extracted from each spectrum of the five spectral types (parallel, perpendicular, diffuse, depolarization ratio, parallel / perpendicular), one being the mean intensity across the entire spectrum and

the other being the intensity at the most discriminatory wavelength, yielding 10 spectral features per measurement site. In addition, nuclear size was extracted from the depolarization ratio spectrum using a Mie theory based algorithm described in references [83] and [86]. Figure 2.2 shows the mean nuclear size per diagnostic category extracted from OPRS measurements and from direct measurements of the biopsied tissue histology slides. Hence, a total of 11 features were extracted from each measurement site for use in eight binary classification tasks: (1) Normal vs. SD, (2) Normal vs. MD, (3) Normal vs. MD and SD combined, (4) MD vs. SD, (5) Benign vs. SD, (6) Benign vs. MD, (7) Benign vs. MD and SD combined, and (8) Benign vs. Normal.

2.4.3.4 Selection of the most discriminatory features

In many classification tasks, a combination of features yields better discrimination than can be achieved with any single feature. Since there are 11 features in each of the two-class classification problems, $2^{11}-1 = 2047$ different combinations of features are possible, for example: feature #1 alone, features #1 and #2 combined, features #1 and #3 combined, features #2 and #3 combined, features #1, #2, and #3 combined, etc. We exhaustively searched through these 2047 combinations to identify the feature or combination of features that best discriminates between two diagnostic classes using LDA [49]. LDA was chosen because it works well with small datasets and it preserves the physical origins of features (refer to Section 1.2.4. Leave-one-out cross validation was employed to train and test all LDA models.

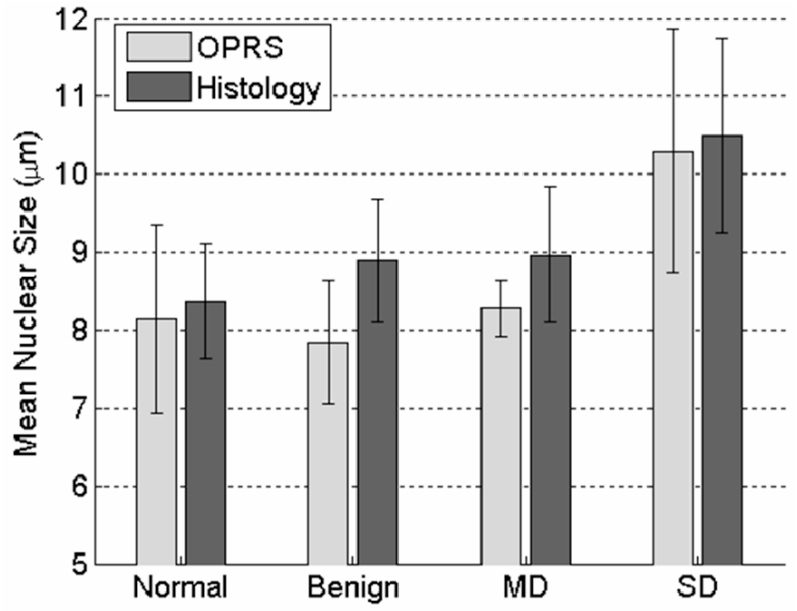


Figure 2.2: Bar graph of mean nuclear size per diagnostic category. Light gray bars indicate nuclear size extracted using OPRS measurements. Dark gray bars indicate nuclear size obtained directly from stained and sliced biopsied tissue mounted on microscope slides. Standard deviation of measurements shown as whiskers.

The performance of a feature combination was evaluated using the area (AUC) under the non-parametric ROC curve generated from the LDA decision variable. All possible LDA models were compared. The best LDA model was defined as that which had the highest area under the ROC curve. LDA calculations were carried out using the classify function in MATLAB R7 Statistics Toolbox (The MathWorks, Natick, MA). For each binary classification task, several of the 2047 LDA models can have statistically equivalent discrimination. A bootstrapping technique was used to estimate the mean difference in the AUC between LDA models and the two-sided p value of that difference[64]. P values below the conventional threshold of 0.05 were regarded as statistically significant. The top N models with AUC's statistically indistinguishable from the maximum AUC observed were considered to be comparable models.

We hypothesized that dominant features would appear with higher frequency, while irrelevant features would appear randomly. Consequently, the fractional occurrence of individual features within the top N LDA models within each diagnostic category was counted. Those features that appeared with a frequency of 0.5 or greater are considered to be of diagnostic importance.

2.4.3.5 Precautions taken to avoid overtraining

Overtraining is a concern any time one develops a classification model with a small dataset. To reduce this risk, we used leave-one-out cross validation to generate the LDA models. As a further check, a permutation test was

applied in which the pathology definition of each patient measurement was randomly shuffled while the prevalence of disease was kept constant [87]. The shuffling was repeated 1000 times for each binary classification task, so that the mean and standard error of the area under ROC curves could be calculated and compared with the observed area under ROC curves.

2.5 Results

2.5.1 Sample Distribution

Table 2.1 summarizes the distribution of sites from the oral cavity that were measured, then subsequently biopsied. A total of 57 sites in 24 patients were measured and analyzed in this study. Table 2.1 is divided into two categories based on the examining physician's visual impression at the time of biopsy: (1) Normal, (2) Abnormal. All visually normal sites were confirmed to be normal by histological analysis. The clinically appearing abnormals were further subdivided into three categories according to their histological diagnosis: benign (B), mild dysplasia (MD), severe dysplasia (SD). We defined SD as tissue that requires surgical excision; those include moderate to severe dysplasia and carcinoma. In this study all but one SD sites were classified as carcinoma (the other was classified as moderate dysplasia). Figure 2.3 shows images taken of representative histopathology slides from this study illustrating the morphology of the different diagnostic categories. All binary combinations of the four diagnostic classes were used to test classification of OPRS data.

Table 2.1: Distribution of anatomical sites within the oral mucosa measured with OPRS. The physical appearance was noted prior to OPRS measurement as either visually normal or abnormal. Biopsies were taken of all 57 measured sites. Histopathology of abnormal biopsies is categorized as benign, mild dysplasia, or severe dysplasia. *Severe dysplasia is defined as tissue that requires surgery for treatment; it includes tissue that has the histopathological diagnosis of moderate to severe dysplasia or carcinoma.

	Clinical appearance		Diagnosis for Abnormals		
Location	*Normal*	*Abnormal*	*Benign (B)*	*MD*	*SD*[a]
Tongue	7	22	8	7	7
Buccal	13	7	3	3	1
Floor of mouth	1	2	0	1	1
Gingiva	1	2	1	0	1
Soft Palate	0	2	1	1	0
Total measured	22	35	13	12	10

2.5.2 Polarized reflectance spectra

Figure 2.4 shows the measured spectra before normalization. The diagnostic category is indicated by the color of the curves: solid green - normal, dashed blue - benign, dash-dot red - mild dysplasia, dotted black - severe dysplasia. The first row shows all measured spectra from the parallel and perpendicular collection channels. The second and third rows show the averaged spectra per diagnostic class for the parallel, perpendicular, diffuse reflectance, and the depolarization ratio. Analysis of the mean spectra showed a good, albeit qualitative, separation of all diagnostic classes. A qualitative look at the

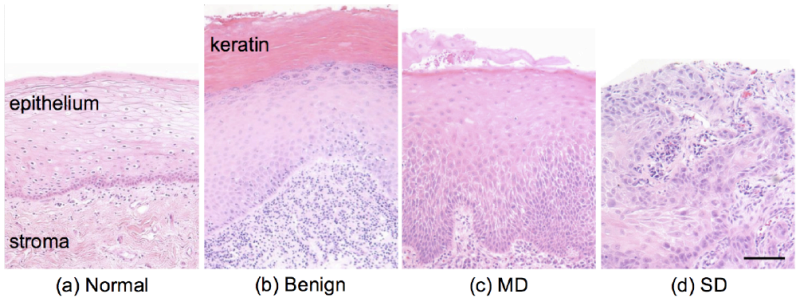


Figure 2.3: Representative images of biopsied tissue diagnosed as (a) normal, (b) benign, (c) mild dysplasia, or (d) severe dysplasia. Tissue was stained with hematoxylin and eosin (H&E) for standard histopathological analysis. Scale bar is 100 microns.

mean spectra reveals differences in the total intensity, which is modulated by hemoglobin absorption. Within a narrow wavelength band, the mean spectral differences can be quite large. For example, the mean perpendicular spectra have the largest separation between diagnostic classes for wavelengths shorter than 450 nm. Similarly, the mean depolarization ratio spectra have the largest separation between normal and SD in the red or long wavelength region. As a whole, the mean spectra hint at the possibility of using select wavelength regions for improved diagnostic discrimination, thus prompting an analysis of the optimal wavelength for maximum separation of two diagnostic classes for each of the spectral types.

2.5.3 Diagnostically relevant features

Table 2.2 shows the best wavelengths determined for the eight two-outcome diagnostic classification tasks and five spectral types. All spectra used in this analysis were normalized to one at 420 nm as described in sec-

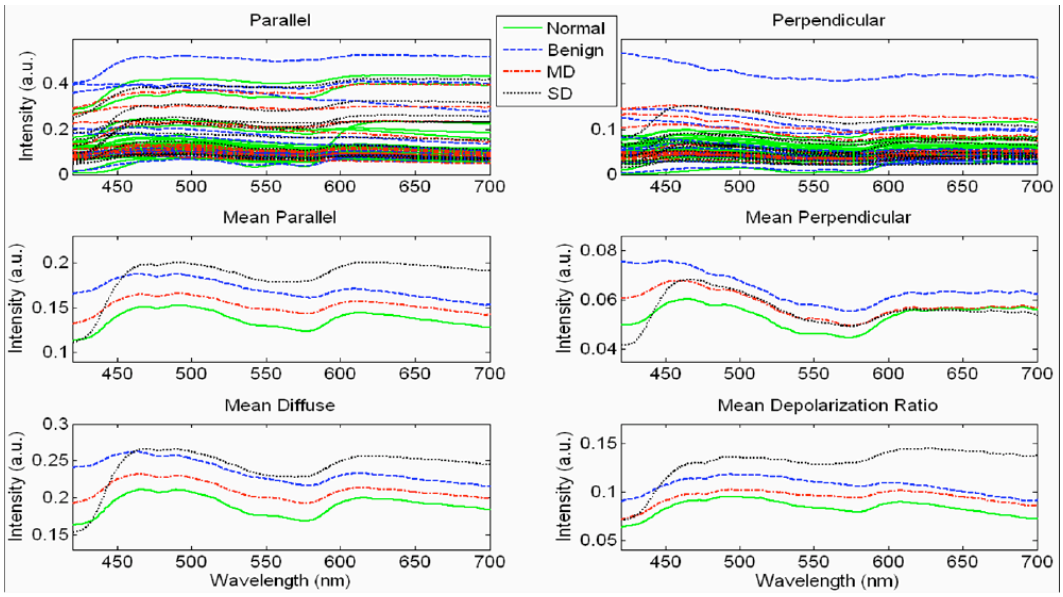


Figure 2.4: Measured spectra. The first row shows all the spectra collected from the two collection channels. The second and third rows show the mean spectra according to diagnostic class for the parallel, perpendicular, diffuse reflectance, and depolarization ratio. Spectra from normal tissue are shown as solid green curves, benign tissue shown as dashed blue curves, mild dysplasia as red dash-dotted curves, and severe dysplasia as black dotted curves.

tion 2.4.3.1. Table 2.3 lists the AUC's for individual features and the best combination of features, as determined by exhaustive LDA search, for each classification task considered. An AUC equal to one corresponds to perfect classification while an AUC of 0.5 corresponds to chance performance. Comparison of the individual feature AUC and the LDA AUC per classification task shows that the LDA model typically performs better. It is important to note that the performance of individual features varies quite drastically depending on the classification task. These results demonstrate that a combination of features is needed to efficiently tackle the multi-task classification problem involved in cancer detection and diagnosis.

Table 2.2: The wavelengths, given in nanometers, that provided the maximum separation between two diagnostic classes.

Spectral Type	*Normal from SD*	*Normal from MD*	*Normal from MD and SD*	*MD from SD*	*Benign from SD*	*Benign from MD*	*Benign from MD and SD*	*Benign from Normal*
Parallel	581	465	655	581	539	597	497	465
Perpendicular	454	576	454	454	454	602	491	486
Diffuse	539	465	718	586	539	497	497	497
Depolarization ratio	602	465	628	454	539	497	581	465
Parallel/perpendicular	633	565	612	633	644	512	644	512

Discrimination of normal from SD has a relatively high LDA AUC. This is consistent with our expectations since normal and SD are two extremes of the continuum of histopathological status. Likewise, histopathologically similar tissue such as normal and MD is more difficult to discriminate. The

Table 2.3: Area under the ROC curve for individual features and the best LDA combination of features. Eight binary classification tasks, indicated by the column headings, were considered. Features identified by x nm correspond to the intensity at the most discriminatory wavelength. These results demonstrate that a combination of features is needed to efficiently tackle the multi-task classification problem involved in cancer detection and diagnosis.

	Area under ROC							
	Normal from			*MD from*	*Benign from*			
Individual Features	*SD*	*MD*	*MD and SD*	*SD*	*SD*	*MD*	*MD and SD*	*Normal*
Nuclear size	0.79	0.69	0.74	0.65	0.75	0.62	0.68	0.56
Mean parallel	0.81	0.50	0.64	0.82	0.88	0.65	0.76	0.69
Mean perpendicular	0.69	0.60	0.53	0.77	0.82	0.73	0.77	0.81
Mean diffuse	0.79	0.53	0.62	0.82	0.86	0.69	0.77	0.74
Mean depolarization ratio	0.84	0.53	0.67	0.82	0.90	0.60	0.74	0.57
Mean par/per	0.73	0.66	0.69	0.69	0.70	0.55	0.62	0.63
Parallel, *x* nm	0.84	0.60	0.68	0.87	0.90	0.68	0.78	0.77
Perpendicular, *x* nm	0.78	0.65	0.57	0.86	0.89	0.74	0.78	0.83
Diffuse, *x* nm	0.82	0.60	0.64	0.86	0.90	0.69	0.78	0.78
Depolarization ratio, *x* nm	0.85	0.61	0.71	0.86	0.91	0.64	0.76	0.72
Par/per, *x* nm	0.77	0.67	0.71	0.71	0.71	0.58	0.62	0.70
Best LDA combination	0.89	0.72	0.74	0.87	0.91	0.76	0.78	0.84

similarity of normal and MD tissue can be seen in Figure 2.3, where the areas of MD (indicated by increased nuclear density) encompass a small section at the basal layer. It is not uncommon for MD to be focally located amidst normal tissue. In an OPRS measurement, the optical signal from MD is weighted

by the surrounding normal tissue. Similarly, the AUC for the classification task of normal from MD and SD combined yield a similar AUC to that of normal from MD. In comparison, clear diagnostic potential is shown for the classification task of MD vs. SD. One of the most clinically challenging classification task requires the ability to distinguish between dysplastic tissue and benign tissue, which has the outward appearance of dysplasia or carcinoma but is histologically normal. Therefore, it is encouraging that some features in our study show clear separation between benign and SD sites. Similar to what is observed when normal tissue is used as the reference, there is less discriminatory power for the classification tasks differentiating benign from MD or benign from MD and SD combined.

Due to the limited amount of data in this pilot study, there is not enough statistical power to distinguish between small differences in ROC area. For example, in the classification task of normal from SD, 36 LDA feature sets with the highest ROC areas were found to be statistically indistinguishable. Rather than attempting to identify a single best model, which is impossible to do in a pilot study, we identified the features that most frequently appear in the set of statistically indistinguishable best performing models. Table 2.4 shows the frequency with which each feature occurs in the top N feature sets with statistically similar ROC areas for each classification task. Dashed lines delineate features that appear with a frequency of 0.5 or greater. We suggest that these features warrant the greatest attention in future large clinical trials.

Table 2.4: For each classification task, the frequency of appearance of individual features within the best performing statistically comparable LDA feature sets is shown. Features are sorted such that those that occur with a frequency greater than 0.5 appear above the dashed line. These features are considered to be the most diagnostically relevant.

Normal from SD		*Normal from MD*		*Normal from MD and SD*		*MD from SD*	
Depol ratio, *x* nm	0.69	Nuclear size	0.64	Nuclear size	0.77	Nuclear size	0.31
Mean depol ratio	0.56	Perpendicular, *x* nm	0.32	Depol ratio, *x* nm	0.41	Mean diffuse	0.31
Nuclear size	0.53	Par/per, *x* nm	0.26	Mean par/per	0.39	Mean perpendicular	0.31
Parallel, *x* nm	0.53	Mean depol ratio	0.25	Par/per, *x* nm	0.38	Perpendicular, *x* nm	0.31
Mean diffuse	0.44	Parallel, *x* nm	0.22	Mean depol ratio	0.36	Mean par/per	0.23
Mean perpendicular	0.44	Mean perpendicular	0.19	Mean perpendicular	0.35	Par/per, *x* nm	0.23
Par/per, *x* nm	0.36	Depol ratio, *x* nm	0.19	Mean parallel	0.34	Parallel, *x* nm	0.23
Perpendicular, *x* nm	0.31	Diffuse, *x* nm	0.19	Mean diffuse	0.31	Depol ratio,*x* nm	0.15
Mean parallel	0.25	Mean parallel	0.18	Parallel, *x* nm	0.29	Diffuse, *x* nm	0.15
Diffuse, *x* nm	0.25	Mean par/per	0.17	Diffuse, *x* nm	0.29	Mean parallel	0.08
Mean par/per	0.17	Mean diffuse	0.16	Perpendicular, *x* nm	0.23	Mean depol ratio	0.00
Benign from SD		*Benign from MD*		*Benign from MD and SD*		*Benign from Normal*	
Mean perpendicular	0.54	Par/per, *x* nm	0.90	Mean perpendicular	0.53	Par/per, *x* nm	1.00
Mean parallel	0.50	Parallel, *x* nm	0.87	Perpendicular, *x* nm	0.40	Perpendicular, *x* nm	0.98
Parallel, *x* nm	0.47	Diffuse, *x* nm	0.87	Diffuse, *x* nm	0.27	Parallel, *x* nm	0.79
Mean diffuse	0.46	Depol ratio, *x* nm	0.75	Parallel, *x* nm	0.20	Mean perpendicular	0.55
Diffuse, *x* nm	0.46	Mean diffuse	0.66	Mean depol ratio	0.13	Mean depol ratio	0.36
Perpendicular, *x* nm	0.46	Mean parallel	0.57	Depol ratio, *x* nm	0.13	Mean diffuse	0.36
Nuclear size	0.37	Mean depol ratio	0.56	Mean parallel	0.07	Mean parallel	0.31
Depol ratio, *x* nm	0.37	Mean par/per	0.49	Mean diffuse	0.07	Diffuse, *x* nm	0.29
Mean depol ratio	0.34	Mean perpendicular	0.38	Nuclear size	0.00	Mean par/per	0.26
Mean par/per	0.13	Perpendicular, *x* nm	0.34	Mean par/per	0.00	Depol ratio, *x* nm	0.26
Par/per, *x* nm	0.11	Nuclear size	0.32	Par/per, *x* nm	0.00	Nuclear size	0.19

2.5.4 Check for overtraining

With a small data set, over training is always a concern. As noted in the methods section, leave-one-out cross validation was used for classifier training to reduce this risk. As an additional check, a permutation test was conducted where the diagnostic state of the measured site was randomly assigned for a given classifier task while preserving the number of patients within each class. The results are shown in Figure 2.5. The mean ± standard deviation of the AUCs for a randomly shuffled task is shown as a gray cross-hair with error bars, while the real AUC of the top LDA model is shown as a filled black circle with a single bar indicating the extent of AUC values within the top N comparable LDA models. The randomly assigned permutations have AUC's that are clustered around 0.5 or chance performance, whereas the real AUC's are well above the error bars. The p values of the permutation tests showed significant difference between the real AUC and the randomly shuffled AUCs. For example, the permutation test for classifying normal and SD has a zero p value. These results demonstrate that the LDA model is capturing meaningful differences between the diagnostic classes as opposed to merely magnifying chance differences in the feature values.

2.6 Discussion

Given the variety of oral mucosa and the resulting spectral diversity that can confound classification, it is noteworthy that statistical significance was obtained across diagnostic classes without the need to stratify the data

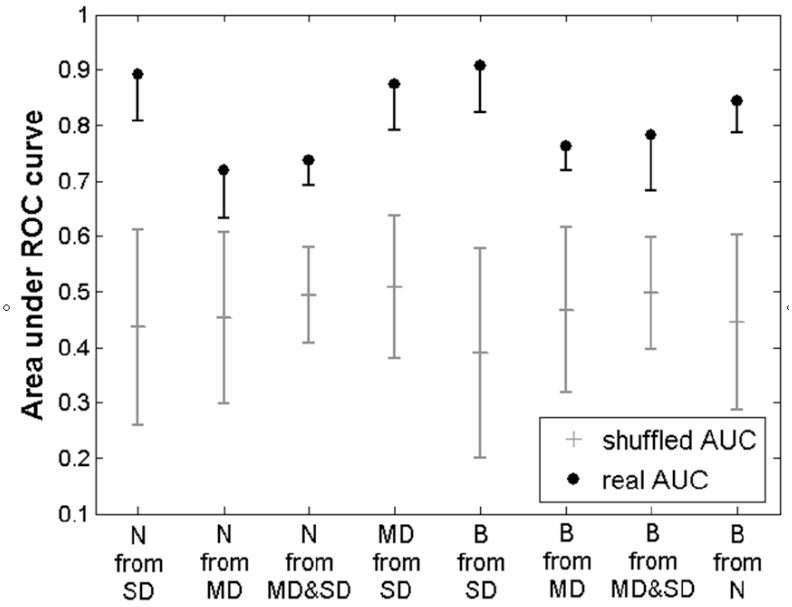


Figure 2.5: Permutation test to check for overtraining. The diagnosis assignment was shuffled for each classification task while preserving the number of patients within each class. The mean and standard deviation of the area under the ROC curve, AUC, obtained using the shuffled data sets is shown as gray cross hairs. The AUC for the true LDA feature sets is shown in black. The highest LDA AUC is shown as a filled black circle, while a lower black whisker indicates the lowest statistically equivalent LDA AUC. The true LDA AUC's are well above the permuted AUC's, which are grouped around chance performance, demonstrating that the data is not being overtrained.

according to tissue location (i.e., buccal, tongue, etc.) or tissue keratinization. The detailed interpretation of these findings as they relate to tissue morphology is somewhat difficult, although a qualitative understanding can be obtained, thus guiding future studies and probe designs.

In Figure 2.3 the progression of normal tissue to premalignancy then to malignancy can be viewed in terms of the physical alteration of normal tissue from a homogeneous two-layer structure to a very heterogeneous single layer tissue. It has been shown that changes in spectral profiles that accompany progression to carcinoma can be related to changes in both the epithelium and the stroma such as increased microvascularization and scattering alterations [6, 88].

The frequency of appearance of features in Table 2.4 points toward their diagnostic importance. The features that occur most consistently are nuclear size, the intensity ratio of parallel to perpendicular channels, and the mean perpendicular signal. Nuclear size appears with the highest frequency in classification tasks involving normal tissue while the mean perpendicular feature and the ratio of parallel to perpendicular feature dominate cases that involve benign tissue. Clearly nuclear size is a measure of the morphological changes that occur in shallow tissue. The regular appearance of nuclear size is consistent with histopathology where epithelial cell nuclei are well documented to enlarge with the progression of cancer [22]. In Figure 2.2 both the extracted and measured nuclear size shows an increase with disease progression.

The ratio of the parallel to the perpendicular signal, on the other hand,

can be interpreted as the ratio of shallow to deep tissue changes. This can be seen if the parallel / perpendicular signal is rewritten as

$$\frac{I_{par} - I_{per}}{I_{per}} + 1 \tag{2.1}$$

where $(I_{par} - I_{per})$ represents photons that have undergone few scattering events, while Iper represents photons that have had many scattering interactions. The mean perpendicular feature is, therefore, a measure of the interaction of photons in tissue below the epithelium where increases in capillary density will manifest as hemoglobin modulation of the perpendicular scattering spectrum.

Recent work in other organ sites indicates that changes in blood content and oxygenation that occur below the epithelium can be related to tumor development and, potentially, to premalignant lesion formation. Siegel, *et al.* reported increased blood supply in subepithelial mucosa before the development of dysplasia in adenomatous human colon biopsies and rat colons treated with a carcinogen [89]. Zonios *et al.* found an increase in hemoglobin concentration in adenomatous colon polyps, but not hemoglobin oxygenation [20]. In contrast, Bard *et al.* found that endobroncho tumors were characterized by lower blood oxygenation [25]. In another study, Fawzy *et al.* demonstrated that malignant lung lesions had differences in blood volume fraction and oxygen saturation when compared to normal or benign lesions [27]. In their investigation, the blood volume fraction was significantly higher in malignant lesions than benign lesions.

These findings can be extended to the oral mucosa where 85% of all precancerous lesions have the clinical appearance of white patches or leukoplakia [90]. A review of the prognosis of oral premalignant lesions in several countries including the USA, India, Hungary, Netherlands, and Norway by J. Rebeil has shown that the rate of malignant transformation of leukoplakia can range from less than 1% to 18%, where the highest transformation rate was found in the USA [91, 92] Less frequently encountered is a red patch or erthyroplakia, which is nearly always associated with dysplasia or carcinoma at the time of identification [93]. Both erythroplakia and their mixtures with leukoplakia are at a higher risk for malignancy. The clinical description of erythroplakia as a red patch indicates an increase in blood perfusion. The implication that capillary density is correlated with precancer progression in the oral cavity is a subject of interest for future work.

Adding to the interpretation of the mean perpendicular feature is the fact that the polarization change that gives rise to the perpendicular signal can be viewed as a diffusion process where depolarization increases with increasing optical-tissue interaction [80]. Consequently, areas of increased scattering in superficial regions of tissue such as keratin or dysplasia will also contribute to the perpendicular signal. The oblique collection geometry of our probe makes it more sensitive to superficial tissue changes as the collection efficiency is greatest in the first 300-400 microns of tissue. This sensitivity enhances the effects of increased superficial scattering. We hypothesize that it is this additional contribution from changes in the local tissue scattering, resulting

from changes in the local tissue morphology, that causes the perpendicular and the parallel / perpendicular features to appear with such high frequency in the top LDA models for classification tasks involving benign tissue. It is well known that increased architectural and morphological heterogeneity of oral cavity lesions is correlated with transformation to invasive cancer [94–96]. Work in the breast, cervix, and bronchus has also indicated that the local variation of tissue can potentially yield diagnostically relevant information [25, 97–99]. Further study is needed of these high frequency features to fully assess their physical meaning and their impact on diagnostic classification.

Another important outcome of the statistical analysis is the ability of OPRS to separate benign lesions from all other lesion types despite having the same or similar outward appearance. Many optical techniques are capable of discerning normal from malignant oral mucosa with a high degree of sensitivity and specificity, but discrimination of benign lesions from precancer and cancer is more elusive. A few groups have reported discrimination of the differing grades of abnormal human oral cavity tissue in vivo using imaging techniques. Wang *et al.* were able to separate benign from dysplastic and cancerous buccal mucosa autofluorescence (sensitivity of 81%, specificity of 96%) using a partial least squares artificial neural network analysis [100]. Onizawa was able to separate benign from cancerous oral cavity tissue with a sensitivity and specificity of 91% and 84%, and also benign from dysplasia plus cancer with a sensitivity of 94% and a specificity of 96% using UV flash photography [66]. Kulapaditharom *et al.* reported a sensitivity and specificity of

100% and 73%, respectively, for separation of benign from dysplastic plus malignant tissue [101]. Although these imaging studies showed good sensitivity and specificity they had certain limitations. The results by Wang *et al.* were limited by the isolation of their study to a single oral cavity location (buccal mucosa) and the similar history of carcinogenic exposure (areca quid chewing and smoking). The results achieved by Onizawa and Kulapaditharom could be attributed to increased porphyrin fluorescence with dysplasia. However, Betz, *et al.* indicated that porphyrin fluorescence is not a good indicator of disease as it was only present in one-third of tumors they studied and was also present on normal tongue and gingival plaques [69].

Overall spectroscopic studies have shown less optimistic results. Muller *et al.* found a sensitivity and specificity of 64% and 90%, respectively, for separating dysplasia from cancer using trimodal spectroscopy [18]. De Veld *et al.* recently concluded that neither diffuse reflectance spectroscopy nor tissue autofluorescence (alone or in combination) could distinguish benign tissue from dysplastic and cancerous tissue based on a clinical trial of 134 abnormal lesions[102]. The results of our pilot clinical trial are more consistent with the cited imaging work. We have also found that the OPRS is capable of distinguishing benign tissue from normal, precancerous, and cancerous tissue with good sensitivity and specificity. Although it is more informative to report AUC's rather than a single point on the ROC curve, Table 2.5 lists the sensitivity and specificity of OPRS for all diagnostic classification tasks for easy comparison to work by others. In each case, the ROC threshold was chosen

such that it gave both high sensitivity and high specificity. We suggest that multiple diagnostically relevant features that can be extracted from a single OPRS measurement are the key to its diagnostic performance.

Separation of normal tissue from SD yielded a sensitivity of 90% and a specificity of 86%. Discrimination of identical looking benign lesions from severe dysplasia was also achieved with high sensitivity (100%) and specificity (85%). Evaluation of all dysplasia (including carcinoma) against normal or benign tissue yielded sensitivity and specificities of 73% and 64%, and 86% and 61% respectively. Lesser diagnostic grades such as mild dysplasia (MD) and benign lesions do not require surgical resection, but must be monitored for possible transformation to malignancy. OPRS was able to discriminant MD from SD with a sensitivity of 80% and specificity of 83%. These compelling findings emphasize the need for a large prospective study to assess OPRS as an adjunct to clinical examination for the detection and monitoring of precancerous and cancerous tissue.

2.7 Conclusion

The oral cavity has a variety of tissue architectures in addition to a whole host of benign conditions such as leukoplakia, erythroplakia, and lichen planus that can mask precancer and cancer. Using a multi-pronged approach to discriminate the earliest stages of precancer could solve this problem. An ideal pared down system would be simple, low-cost, robust, and noninvasive. We believe that OPRS has the potential to fulfill these requirements. Our

Table 2.5: Sensitivity and specificity of OPRS for all binary combinations of four distinct histological groups: normal, benign, mild dysplasia, and severe dysplasia.

	Normal from			*MD from*
	MD	*MD and SD*	*SD*	*SD*
Sensitivity (%)	75	73	90	80
Specificity (%)	73	64	86	83
	Benign from			*Benign from*
	MD	*MD & SD*	*SD*	*Normal*
Sensitivity (%)	92	86	100	85
Specificity (%)	69	61	85	73

results from a preliminary pilot clinical trial have demonstrated the ability of OPRS to discriminate, with high sensitivity and specificity, normal tissue from high grade dysplasia and cancer (SD). Further, OPRS can discriminate visually identical lesions such as benign from SD and benign from premalignant and malignant lesions. These promising results suggest that OPRS has the potential to augment current clinical practice for diagnosis and monitoring of oral premalignancies and malignancies.

This study has also pointed out the importance of wavelength dependencies. Some wavelengths may carry more diagnostic information than other wavelengths as we did feature selection based on the most discriminatory wavelength (Section 2.4.3.4). In the next chapter (Chapter 3, a novel method for

spectral feature extraction will be presented to address spectral information in optical spectroscopy analysis.

Chapter 3

Adaptive spectral window sizes for extraction of diagnostic features from optical spectra

In Chapter 2, the analysis for a oral cancer dataset is presented as a pilot clinical trial. This chapter is modified from a manuscript accepted by Journal of Biomedical Optics, "Adaptive spectral window sizes for extraction of diagnostic features from optical spectra".

Here we present an approach to adaptively adjust the spectral window sizes for optical spectra feature extraction. This adaptive windowing technique ensures the signal linearity in defined windows; hence, adaptive windowing technique retains more diagnostic information while using fewer windows. This method was tested on the same dataset presented in Chapter 2 with the use of only the diffuse reflectance spectra. Eight features were extracted from each window. While adaptive and fixed size windowing perform similarly, adaptive windowing utilizes significantly fewer windows than fixed sized windows (Number of windows per spectrum: 8 vs. 16). Since adaptive windows retain most diagnostic information while reducing the number of windows needed for feature extraction, our results suggest that it isolates unique diagnostic features in optical spectra.

3.1 Background and motivation: spectral feature extraction

As reviewed previously in Section 1.2.2, spectral feature extraction and selection are critical in designing clinical decision support systems. There are four major categories of feature extraction methods in spectral signal processing: principal component analysis (PCA), model-based feature extraction, spectral feature extraction, and hybrid feature extraction [2]. Quantitative features are extracted from optical spectra to describe different spectral patterns, while feature selection aids in the identification of those optically derived features that are diagnostically relevant and the elimination of redundant features that are strongly related to selected features. Minimizing the number of features is important to reduce computation complexity, processing time, and to prevent overtraining.

The method we utilize is a variation on statistical feature extraction, which is a more heuristic approach to find the correlation between statistics from spectroscopy and known pathology status. In prior studies, spectral features of optical spectroscopy were extracted from the entire spectrum; however, that may not be optimal. Some groups investigated dividing the entire spectrum into smaller spectral regions for feature extraction. When extracting features from a spectral region, how large should the spectral region be? Take the maximal intensity feature, for example; should we use global maximum of the entire spectrum? Or is a local maximum more meaningful? It is widely accepted that some wavelengths may be more discriminatory than

others [11, 34, 68] because light of different wavelengths behaves differently when interacting with tissues. Features extracted from spectral regions help us understand the underlying morphology and optical properties of the tissue. Undoubtedly, the choice of spectral region for feature extraction makes a difference in the performance of the extracted features [11, 34, 68].

A study by Bigio *et al.* employed spectral features extracted by dividing the spectra into several spectral windows of a fixed width of 20 nm [1]. In addition to the average intensities of spectral windows, slopes of these spectral windows were also extracted as features because it was discovered that broad (large-spectral-range) slope changes were observed for malignant conditions due to enlarged and denser nuclei [1]. Mourant's study on spectroscopic diagnosis of bladder cancer with elastic light scattering [34] used a similar method of dividing the spectrum into smaller spectral bands (20 nm) and searching for the most discriminatory one. However, the choice of spectral window size was arbitrary in both of these prior studies.

In this study, we present an approach for adaptively adjusting the spectral window sizes for feature extraction from optical spectra. This approach uses simple linear regression to make a piecewise model of the measured optical spectra. Features such as average intensity and the slope and intercept of the piecewise linear regression were investigated. By adaptively adjusting the spectral window sizes, the trends in the data are captured more succinctly than when a small fixed window size is used. In other words, this method reduces the feature redundancies that exist when fixed-size windows are used for fea-

ture extraction. The results of this study show that windowing techniques have better diagnostic performance than no windowing. Also, adaptive windowing have similar diagnostic power to fixed-sized windowing; however, adaptive windows require significantly less windows than fixed-sized windows. This shows that adaptive windowing technique preserves the information needed for diagnosis with fewer windows used. The main contribution of this adaptive window approach is in statistical feature extraction. In addition, when choosing what features to use, adaptive windowing is most appropriate when the statistical features have linear relationships across windows, as theoretically illustrated in Section 3.2.

3.2 Theoretical derivation

Within every spectral window, eight spectral features are extracted to describe the signal (Table 3.1). In this section, we derive how adaptive windows can reduce the redundancies from fixed sized windows. In particular, we use an example where the adaptive window is exactly twice the size of a fixed size window (see Figure 3.1); a proof under general conditions is not applicable since adaptive windows are variable across spectra and can start and end anywhere. In Figure 3.1 we can see an illustration of adaptive and fixed-sized windows. The dashed line is the real spectrum, and solid line is the fitted line. Windows 1 and 2 are fixed-sized windows while window 3 is the adaptive window. For theoretical explanations, we assume that the dashed line (measured signal) can be represented using the fitted line (solid line).

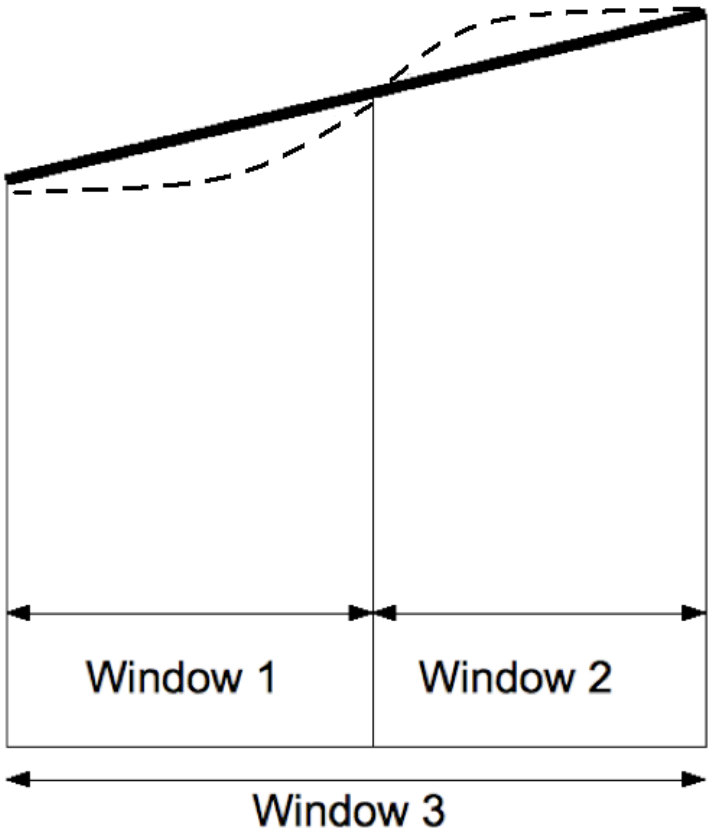


Figure 3.1: Illustration of an adaptive window and two fixed windows. Dashed line is the real spectrum, and straight line is the fitted line. Windows 1 and 2 are fixed-sized windows while window 3 is the adaptive window.

Table 3.1: Eight spectral features were employed in this study. These local features were extracted in each spectral window.

Feature Names	Proposed by
1. slope from linear regression	Bigio *et al.* [1] and Mourant *et al.* [34]
2. intercept from linear regression	Mueller *et al.* [18]
3. minimum intensity	new
4. average intensity	Bigio *et al.* [1] and Mourant *et al.* [34]
5. median intensity	new
6. maximum intensity	new
7. standard deviation of the intensities	Kamath *et al.* [32]
8. signal energy of the intensities	Kamath *et al.* [32]

Table 3.1 lists the features that we explored. Column 2 references other researchers that have used the given feature for diagnostic purposes. The eight features are defined as follows: Features #1 and #2 are the slope and intercept extracted from performing linear regression within each spectral window, which remain the same between smaller fixed window sizes and larger windows for a linear signal within the window. Features #3-#6 are, respectively, minimum, average, medium, and maximum intensities. These are intensities within the window and can vary with different window sizes. However, the variation is linearly proportional. For example, in Figure 3.1, feature #3, the minimal intensity, extracted from windows 1 and 2 differs only by an offset value. Therefore, there is a linear relationship in feature #3 extracted from fixed-size windows (windows 1 and 2,) and adaptive windows (window

3). Likewise, the values of features #4-#6 for fixed-size windows (windows 1 and 2,) and adaptive windows (window 3) are linearly proportional. Feature #7 is the standard deviation of the intensities in the window. The standard deviation in windows 1 and 2 can be written as

$$\sigma_1 = \sqrt{\frac{\sum_{i=1}^{N/2}(I(i) - \bar{I}_1)^2}{N/2}} \tag{3.1}$$

$$\sigma_2 = \sqrt{\frac{\sum_{i=N/2+1}^{N}(I(i) - \bar{I}_2)^2}{N/2}} \tag{3.2}$$

where $I(i)$ is the intensity at the i-th wavelength point, N is the number of points in the window. The mean intensity is given by $\bar{I}_1 = \frac{\sum_{i=1}^{N/2} I(i)}{N/2}$, and $\bar{I}_2 = \frac{\sum_{i=N/2+1}^{N} I(i)}{N/2}$

In window 3, the standard deviation can be written as

$$\sigma_3 = \sqrt{\frac{\sum_{i=1}^{N}(I(i) - \bar{I}_3)^2}{N}} \tag{3.3}$$

$$\sigma_3^2 = \frac{1}{N}\left[\sum_{i=1}^{N/2}(I(i) - \bar{I}_3)^2 + \sum_{i=N/2+1}^{N}(I(i) - \bar{I}_3)^2\right] \tag{3.4}$$

$$\sigma_3^2 = \frac{1}{N}\left[\sum_{i=1}^{N/2}\left(I^2(i) - 2I(i)\bar{I}_3 + \bar{I}_3{}^2\right) + \sum_{i=N/2+1}^{N}\left(I^2(i) - 2I(i)\bar{I}_3 + \bar{I}_3{}^2\right)\right] \tag{3.5}$$

$$\sigma_3^2 = \frac{1}{N}\left[\sum_{i=1}^{N/2} I^2(i) + \sum_{i=N/2+1}^{N} I^2(i)\right] - 2\bar{I}_3\bar{I}_3 + {\bar{I}_3}^2 \tag{3.6}$$

$$\sigma_3^2 = \frac{1}{N}\left[\sum_{i=1}^{N/2} I^2(i) + \sum_{i=N/2+1}^{N} I^2(i)\right] - {\bar{I}_3}^2 \tag{3.7}$$

We rewrite Equation (3.1) as

$$\sigma_1^2 = \frac{1}{N/2}\left[\sum_{i=1}^{N/2} I^2(i) - 2\bar{I}_1 \sum_{i=1}^{N/2} I(i) + \frac{N}{2}{\bar{I}_1}^2\right] \tag{3.8}$$

$$\sigma_1^2 = \frac{\sum_{i=1}^{N/2} I^2(i)}{N/2} - {\bar{I}_1}^2 \tag{3.9}$$

Thus, equating Equation (3.9),

$$\frac{\sum_{i=1}^{N/2} I^2(i)}{N} = \frac{\sigma_1^2 + {\bar{I}_1}^2}{2} \tag{3.10}$$

Substitute Equation (3.7) with Equation (3.10), we get

$$\sigma_3^2 = \frac{\sigma_1^2 + {\bar{I}_1}^2}{2} + \frac{\sigma_2^2 + {\bar{I}_2}^2}{2} - {\bar{I}_3}^2 \tag{3.11}$$

Since $\sigma_1 = \sigma_2$, we get

$$\sigma_3^2 = \sigma_1^2 + \frac{1}{4}(\bar{I}_1 - \bar{I}_2)^2 \tag{3.12}$$

Therefore, from Equation (3.11), we can conclude that feature #7 has a linear relationship between adaptive windows and fixed windows.

Feature #8, signal energy, is defined as:

$$SE = \sum_{i=1}^{i=N} I^2(i) \tag{3.13}$$

where N is the number of spectral points inside the spectral window (i.e., the number of wavelengths) and $I(i)$ is the intensity at the ith point in the spectral window. It can be rewritten as

$$\sum_{i=1}^{i=N} I^2(i) = \sum_{i=1}^{i=N/2} I^2(i) + \sum_{i=N/2+1}^{i=N} I^2(i) \tag{3.14}$$

$$SE_3 = SE_1 + SE_2 \tag{3.15}$$

From the derivation above, we conclude that features extracted from larger windows identified by the adaptive windowing technique are linearly related to features extracted from smaller, fixed-size windows. Thus, the adaptive windowing technique should enable the use of a smaller number of windows, with little loss of diagnostic information.

3.3 Materials and Methods

3.3.1 Overview

We compare three windowing techniques: (1) No windowing is performed, i.e., features are extracted from all wavelengths 400 nm to 725 nm, (2) Fixed window size of 20 nm (window size adopted from literature [1]), (3) Adaptive spectral window sizes determined by the novel algorithm presented in this paper.

3.3.2 Data Collection

This analysis used a diffuse reflectance spectroscopy dataset that was described in detail in Section 2.4.1 [43]. In this chapter, we used only the diffuse reflectance subset of the OPRS data set since diffuse reflectance spectroscopy is more commonly used than polarized reflectance spectroscopy. Moreover, the purpose of this study is to demonstrate the effectiveness of a new windowing method that can be used with a broad range of spectral data; thus, the selection of a particular spectroscopy technique as an exemplar is arbitrary.

3.3.3 Preprocessing

The spectra were normalized to remove inter-patient variation (refer to Section 1.2.1). As described by our previous chapter (Chapter 2 [43]), the spectra were dark subtracted, then divided by the diffuse reflectance from a white standard (Labsphere, SRS-99) to correct for the spectral response of the system and spectral profile of the source. Then each spectrum was normalized by dividing each intensity value by the intensity at 420 nm. No downsampling was performed in this study because detailed data were needed for piecewise linear regression.

3.3.4 Piecewise Linear Regression for Spectral Feature Extraction

We developed an algorithm to adaptively adjust the spectral window size for feature extraction from optical spectra. The spectral window sizes are maximized given defined acceptable linear fits on the spectrum.

Our method first sets an initial spectral window size of 5 nm. The choice of this initial spectral window size is based on the spectral resolution (5 nm) of our measured spectra. The spectral window size is iteratively increased by 5 nm and simple linear regression is performed within the spectral window. A stopping criterion of R^2 of 0.8 is applied to ensure the goodness of each fit. The spectral window size, and stopping criterion can be adjusted if other spectroscopy data, such as fluorescence spectra, are used. This algorithm ensures each piece of the linear model achieves the largest possible window size to fit a linear regression line with an R^2 value of 0.8 or greater. Once the R^2 value falls below 0.8, the iteration ends, and the starting position of the next window is set at the ending position of the current window. The new window size is re-initialized to 5 nm and the calculation for R^2 is repeated.

3.3.5 Feature extraction

Within every spectral window, eight spectral features are extracted to describe the signal (Table 3.1). Six features have been used previously by others to capture diagnostic information to detect cancer. These features were used to detect cancer in various organ sites - oral [18, 32], breast [1], and bladder [34], with diffuse reflectance spectroscopy [1, 18, 32, 34] and fluorescence spectroscopy [18].

Features #1, #2 #4, #7, and #8 are adopted from literature [1, 18, 32, 34]. In these previous studies, the features were extracted and fed to classifiers or clustering methods as inputs. In our study, we try to leverage this by

extracting similar features to investigate the effect of adaptive windowing.

We contribute novel features #3, #5, and #6, which are the minimum, median, and maximum intensities, respectively, within each spectral window. These extreme points (maximum, minimum) provide additional information not represented by other summary features such as the slope and intercept.

Note that the features used in this chapter differ from those extracted in Chapter 2. This is because we are investigating regional wavelength bands here, but not just extracting one general feature for the entire spectrum.

3.3.6 Performance evaluation

In this study, and generally in spectroscopy data processing, performance is evaluated based on individual windows. In other words, performances of wavelength bands are evaluated separately. This step is necessary for two reasons. First, the number of features is very large if all of the wavelengths are involved in the analysis, especially with fixed size windows. Second, this method can provide important insights for instrument development. Analyzing different wavelength bands enables measurements only at certain wavelengths making it inherently suitable for filter based imaging instrumentation design.

The wavelength-based performance analysis uses all eight features extracted from a window to predict which of the two diagnostic categories each spectrum belongs. We consider only two of the possible diagnostic tasks in order to simplify the analyses: 1. Normal vs. MD+SD 2. Benign vs. MD+SD

Task 1 is very commonly studied in the literature because it is assumed

to be pathologically the most distinct comparison [43]. Task 2 is arguably the most important task clinically, since distinguishing disease cases from visually abnormal but pathologically benign cases is the key challenge faced by the physician. In our analyses, three windowing techniques are evaluated, as described earlier in Section 4.1, (1) no windowing, (2) fixed size windowing, and (3) adaptive windowing. For each window throughout the spectrum, the eight features listed in Table 1 are extracted. A two-class Linear Discriminant Analysis (LDA) classifier (refer to Section 1.2.4) is used to combine the eight features. Leave-one-out cross-validation strategy is employed and the area under the Receiver Operating Characteristic (ROC) curve (AUC) (refer to Section 1.2.5) is used as an evaluation metric for the diagnostic power of each wavelength. The features are extracted from each window, but the AUCs are evaluated per wavelength because the window definitions vary across spectra.

One AUC value is reported for each window. Therefore, in windowing methods (2) and (3), the wavelength space is divided into fixed or adaptive sized windows, so each window results in one AUC value. In comparison, in method (1) wherein no windowing technique is used, only one set of features is extracted, so there is only one AUC value for the entire wavelength space.

3.3.7 De-noising

Noisy data can interfere with classifier training. We used two procedures for de-noising: (1) Thresholding on classifier outputs and (2) Outlier removal. These two procedures are described as follows.

Thresholding on classifier outputs uses the classifier output values to identify situations where the classifier has not been properly trained. Specifically, if all of the normalized classifier outputs fall in the range 0.4 to 0.6 (0 being negative cases and 1 positive cases), the classifier has not been effectively trained to distinguish between the target groups. Likewise, if the mean classifier output for the positive cases is smaller than the mean classifier output for negative cases, then the classifier has not been effectively trained to distinguish between the two classes. If either of these two criteria is met, we consider the predictive LDA model to be too noisy to make a proper prediction, and we remove this LDA and its corresponding AUC in our analysis.

Unlike thresholding of classifier outputs, outlier removal is performed to identify a particular spectrum as distinct from all other spectra. Outlier removal is performed in the feature space and it is based on the Mahalanobis distance (MDist) measure [110]. The MDist is a multivariate measure (in square units) of the separation of an unknown dataset from a known set (with mean μ and covariance matrix S) in space. The Mahalanobis distance of a dataset when applied to itself can be used to find outliers. It has been shown that for a large sample of multivariate normal data, the MDist follows approximately a chi-square distribution with the degrees of freedom being the number of the variables [110]. We consider a spectrum to be an outlier if its MDist is larger than the critical point (significance level α=0.001) of the chi-square distribution with the degrees of freedom being the number of variables participating in the MDist. The de-noising of the data is different from the

preprocessing step described in section 4.3. The de-noising of the data is necessary because it removes unwanted information from the feature space, while the data preprocessing rescales the data from the spectroscopic space.

3.3.8 Statistical comparison of AUC values

A bootstrapping technique was used to estimate the significance of the observed difference in the AUC between LDA models [64]. P values below the conventional threshold of 0.05 were regarded as statistically significant. In some comparisons in this study, a few cases were removed by the outlier identification algorithm for one windowing method but not the other. For example, case #19 is removed in adaptive windowing in Normal vs. MD+SD, but not removed in the fixed windows analysis. In order to compare the two windowing techniques, we remove all of the cases that are defined as outliers in either of the two methods under study.

3.3.9 Studying the effects of different initialization points

In principle, the initialization point of the first spectral window in piecewise linear regression may be a factor that affects performance. In other words, different starting points may cause the entire spectrum to be adaptively divided into different windows. To investigate this possibility, we experimented with multiple initialization points to determine if different initialization points result in different piecewise linear regression models. Three different experiments were conducted:

(1) Small range variations of initialization points: Several initialization points were used: 400, 405, 407, 410, 415, and 420 nm. Since these changes are small, we start the windows from these initialization points, and discard the data points before the initialization points. We then visually assess these regression models and the window definitions.

(2) Larger range variations of initialization points: three initialization points were tested: 400, 562, and 725 nm. These points represent the smallest, the middle, and the largest wavelengths in our data. Since these variations are too large to discard any data as in our first experiment, we modified the window growing direction. For 400 nm, we start from the left and grow to the right. For 562 nm, we start at the middle of the spectra and grow both to the left and to the right. For 725 nm, we start from the right end and grow solely to the left. We then visually assess these regression models and the window definitions.

(3) For all those window definitions obtained from steps (1) and (2), we extract features from these windows with different initialization points and evaluate the AUC to see whether the initialization point is an important factor for adaptive windowing technique.

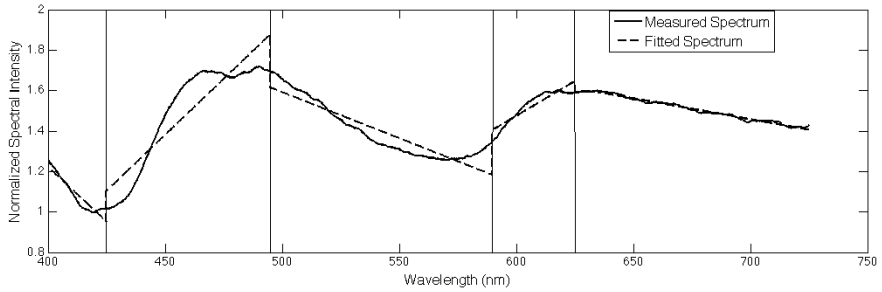
3.4 Results

3.4.1 Piecewise linear regression models

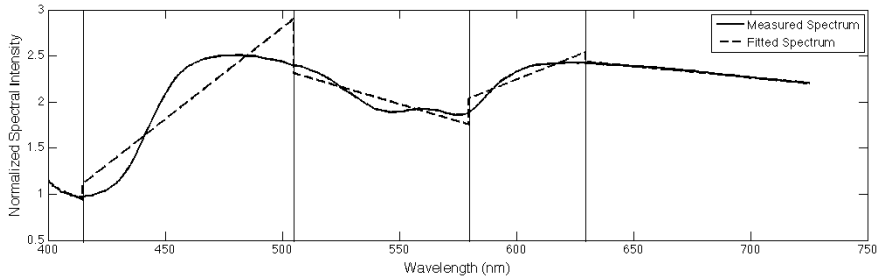
Sample results of our piecewise linear regression model are shown in Figure 3.2. These four examples show diffuse reflectance spectra measured on

four sites with different histopathology statuses (solid curves). These measured spectra are fitted by our piecewise linear regression models to iteratively search for maximum window sizes with acceptable goodness of fit (dashed curves). These fitted piecewise linear models define different sizes of spectral windows for feature extraction. These fitted spectra are not intended to replace measured signals; on the contrary, features are extracted from the measured spectra.

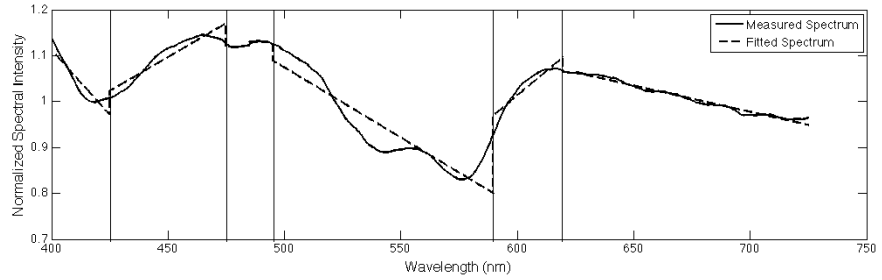
A qualitative look at these piecewise linear regression models reveals that the regression models capture most of the variability within each spectrum; thus, most of the spectral information is retained while the number of windows needed in feature extraction is reduced relative to using a smaller fixed window size. For example, Figure 3.2(a) has a window that starts at 425 nm and ends at 500 nm, a window size of 175 nm, which is about 9 times the size of the 20 nm fixed windows. This larger spectral window captures the relevant features necessary for classification that are also contained in several smaller windows, but using one large window has less redundancy. The features slope (Features #2) and intercept (Feature #3) are the same with big or small windows. The intensities features (Features #4 - #7) may change for bigger windows, but they are linearly proportional as shown in previous theoretical derivations (Section 3) such that the classifier will compensate for it. Therefore, this piecewise linear regression method decreases the redundancies in feature extraction relative to a fixed window size. An important finding of this study is that the adaptive windowing technique uses fewer win-



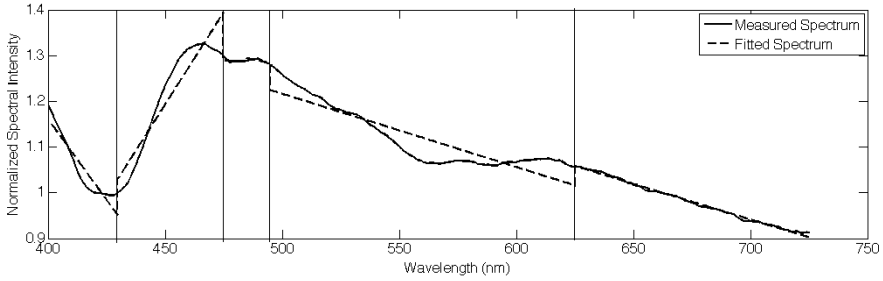
(a)



(b)



(c)



(d)

Figure 3.2: Sample diffuse reflectance spectra for (a) Normal (b) Benign (c) MD (d) SD patients. The solid curve indicates the measured diffuse reflectance spectrum, while the dashed curve indicate indicates the fitted spectrum based on our piecewise linear regression model.

dows to cover regions that behave similarly. In other words, several fixed size windows may cover a wide range of the spectrum that could instead be represented equally effectively in a single window. Adaptive windowing, on the other hand, preserves diagnostic information while decreasing the redundancy.

3.4.2 Feature extraction

After defining the adaptive windows, eight features are extracted from each window. In this section, we evaluate the three windowing techniques (no windowing, fixed size windowing, and adaptive windowing) by looking at the AUC performances, and the feature redundancies of each windowing technique.

In Table 3.2, we compare the predictive power of the three windowing techniques. For each window, all eight features were used to train a LDA classifier via leave-one-out cross validation to obtain the AUC. Since there are multiple windows involved in windowing techniques (fixed size, adaptive windows), multiple AUCs are reported. In Table 3.2, the maximum and the median AUCs of each windowing method are listed. The classifiers based on features extracted from adaptive windows or fixed-size windows outperform the classifier based on features extracted from the entire spectrum (no windowing). This same trend is apparent for both the Normal vs. MD+SD and Benign vs. MD+SD classification tasks. Specifically, the maximum AUC of classifiers trained on features extracted from adaptive windows (max AUC = 0.73) is statistically significantly ($p = 0.04$) larger than that of the classifier

Table 3.2: Statistics from three windowing techniques. The maximum and median AUCs for LDA classifiers trained using all eight features from each spectral window. Pairwise comparison of the windowing techniques found that the maximum AUC for both the adaptive and fixed window technique perform significantly better ($p = 0.04$) than the no windowing method for the classification task Normal vs. MD+SD. Both the adaptive and fixed window techniques also showed improved performance ($p = 0.04$ and 0.01) over the no windowing technique for the classification task Benign vs. MD+SD. For both classification tasks, the median AUC did not show statistically better performance over the no windowing technique. The results demonstrate the value of windowing, as adaptive/fixed windowing higher maximum AUCs than no windowing. It also suggests that the adaptive windowing technique yields classifiers as effective as the fixed windowing technique.

	Normal vs. MD+SD				Benign vs. MD+SD		
	Adaptive window	Fixed-size window (20nm)	No windowing		Adaptive window	Fixed-size window (20nm)	No windowing
Maximum AUC	0.7259	0.7273	0.6114		0.7879	0.8333	0.6818
Median AUC	0.6245	0.5136			0.6996	0.7235	

based on feature extracted from the entire spectrum (no windowing, AUC = 0.61) for Normal vs. MD+SD. Likewise, the maximum AUC of the classifiers trained based on features extracted from fixed windows (max AUC = 0.73) is statistically significantly larger ($p = 0.04$) than that of the classifier based on features extracted from the entire spectrum (no windowing, max AUC = 0.61) for Normal vs. MD+SD and also for Benign vs. MD+SD with adaptive (max AUC 0.79 vs. AUC 0.68, $p = 0.04$) and fixed windows (max AUC 0.83 vs. AUC 0.68, $p = 0.01$). From these statistical analyses, we found that the maximum AUCs by windowing techniques are all higher than no windowing

techniques, in both Normal vs. MD+SD and Benign vs. MD+SD. And these AUCs are significantly higher ($p < 0.05$) in all of 4 comparisons.

Since the maximum AUCs represent the best diagnostic power of the spectrum across all wavelengths, we showed that windowing techniques provide better diagnostic accuracy than no windowing. Moreover, the classifiers based on adaptive windows are as good as those based on fixed windows in both Normal vs. MD+SD and Benign vs. MD+SD. On the other hand, the median AUCs for both Normal vs. MD+SD and Benign vs. MD+SD did not show a statistically significant improvement over the no windowing technique. The results tell us that the median AUCs of windowing and those of no windowing are similar. This is equivalent to comparing a group of mediocre performers to the average of all performers. Therefore, comparison of median performance does not predict which window has the most diagnostic power.

In addition to evaluating the windowing techniques based on the resulting classifier efficacy, we also investigated the efficiency of the classifiers. Specifically, we examined the number of unique features extracted, the average number of windows used per spectrum, and the total number of windows used (Table 3.3). The unique features are calculated by simply removing repeated features extracted in our program. These repeated features are commonly seen in fixed-size windows. For both the Normal vs. MD+SD and Benign vs. MD+SD diagnostic tasks, the adaptive windowing technique requires fewer windows (8 windows instead of 16) but produces more unique features (60 unique features instead of 17) relative to the fixed windowing method. This

Table 3.3: This table shows numbers of unique features extracted, and the numbers of windows per spectrum. It shows that adaptive windows require fewer windows on average, but produces the most unique features. The reason that "no windowing technique" has 44 and 35 total number of windows instead of 1 is because there are 44 and 35 spectra in each case, and each spectrum has one window.

	Normal vs. MD+SD			Benign vs. MD+SD		
	Adaptive window	Fixed-size window (20nm)	No windowing	Adaptive window	Fixed-size window (20nm)	No windowing
Number of unique features extracted	60	17	1	60	14	1
Average number of windows per spectrum	8.5	16	1	7.71	16	1
Total number of windows	374	704	44	270	560	35

comparison demonstrates that adaptive windowing is able to maximize the information obtained in one window, and consequently reduce the number of windows needed to maintain the diagnostic power. In other words, adaptive windowing avoids the use of redundant windows that are employed by the fixed-size windowing method. Likewise, while reducing the number of data points used for feature extraction, the number of unique features remains high in adaptive windows. The adaptive windowing technique is able to retain the variability of data while reducing the data dimensionality in feature space.

This observation agrees with our theoretical assessment in Section 3.2

that adaptive windowing preserves information by using a larger adaptive window to cover the region where fixed size windows behave similarly and extract similar and redundant features.

3.4.3 Effect of different initialization points

As described in Section 3.3.9, there are three experiments conducted.

First, the adaptive windowing technique was applied with different initialization points: 400, 405, 407, 410, 415, and 420 nm (Figure 3.3). In Figure 3.3, we selected one spectrum of category SD to visualize the window definitions of different initialization points. It is visually apparent that the adaptive windows definitions are only slightly shifted from each other. For a range of 20 nm changes in initialization points, the window definitions are only shifted within 10 nm. Compared to the wavelength span of 325 nm (400-725) of the spectra, the shift is very small. Hence, from our first experiment, since the window definitions are similar, the features extracted from windows identified with different initialization points should be similar. We further investigated this in the third experiment. In the second experiment, we investigated a larger range of the initialization points. Since the ranges are larger, we modified the algorithm such that we do not discard the spectral data to the left of the initialization point (as we did in the previous experiment). The modified version of the algorithm grows windows to the left or right side of the initialization point. Three initialization points were tested: 400 (the smallest wavelength), 562 (the median of the wavelengths), and 725 (the largest

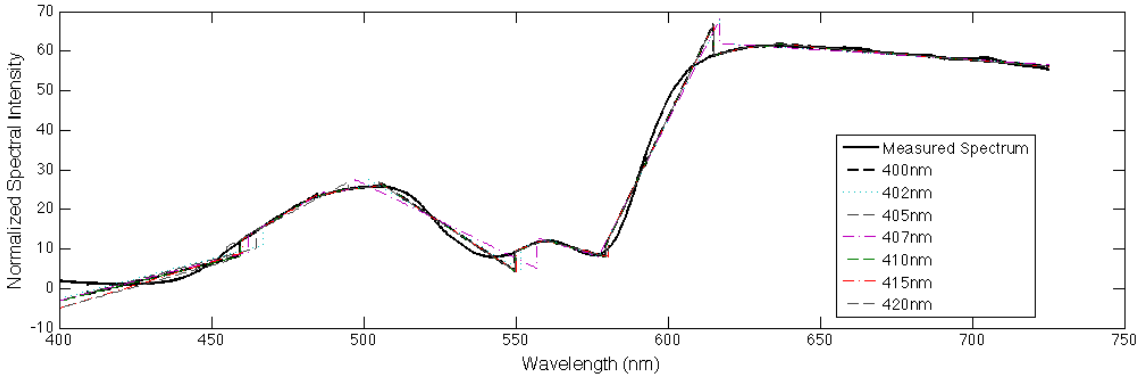


Figure 3.3: Adaptive window technique applied on a spectrum with variable initialization points within a small range. This is one example out of the 57 spectra in the dataset. The patient belongs to pathology group SD. The solid black curve is the measured spectrum, and the dashed straight lines are the windows. It shows that in this example, variable initialization points do not affect the window definitions very much, i.e. the changes are within 10 nm, which is 1.4% of the entire spectrum.

wavelength). The results are shown in Figure 3.4. Similar to Figure 3.3, In Figure 3.3, the same spectrum is used to visualize the window definitions of different initialization points. By visually assessing the window definitions, we can see that the initialization point of 400 and 562 nm have more similar window definitions; while the ones with 725 nm have larger variations from those of 400 and 562 nm.

In the third experiment, we tested the performance of these different initialization points. We calculated the AUCs for the classifiers based on features extracted from the adaptive windows defined by these different initialization points. All eight features are used to train an LDA classifier using leave-one-out cross validation. The maximum AUC ranges from 0.65 to 0.82 (Table 4)

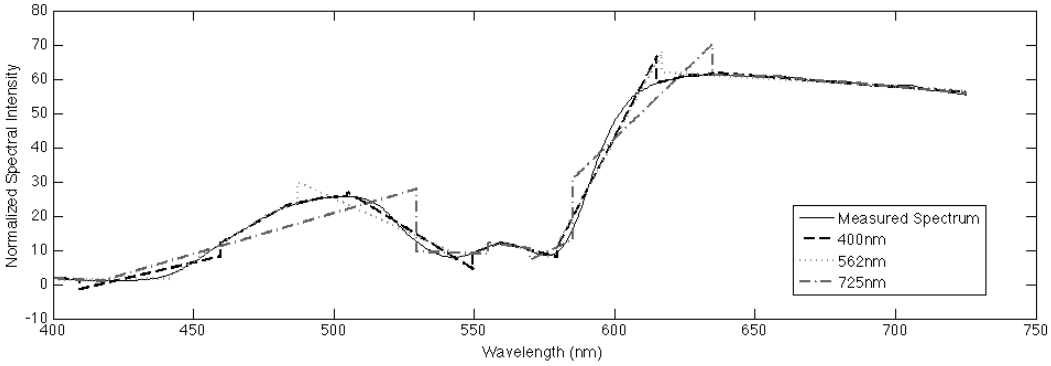


Figure 3.4: Adaptive window technique applied on a spectrum with variable initialization points within a large range. This is one example out of the 57 spectra in the dataset. The patient belongs to pathology group SD. The solid black curve is the measured spectrum, and the dashed straight lines are the windows. In this experiment, the initialization points differ in a larger range than the previous experiment (shown in Figure 3.3). Windows "grow" to either left or right side of the initialization point they grow to the right with 400 nm initialization, to both left and right with 562 nm, to left side with 725 nm. In this example, it shows that the window definitions are visually similar for 400 and 562 nm, but the ones with 725 nm are visually distinguishable.

based on the choice of initialization point. Similarly, the median AUC ranges from 0.56 to 0.66 based on the choice of the initialization point. Conversely, the AUCs for fixed windows do not change at all with different initialization points. These results in Table 4 give us two interesting perspectives: first, the adaptive windowing technique has flexibility to achieve higher AUC (0.82) to outperform fixed windowing (0.73). However, this also raises the concern that the adaptive windowing has variability where it is not guaranteed to have the optimum performance and consistency in terms of the initialization points. In other words, the initialization point needs to be optimized to produce the best performance. Second, while the window definitions of initialization points 400 nm and 725 nm may seem very different, their AUCs are not that different. The AUCs for the 725 nm and 405 nm initialization points are the same (AUC = 0.78). Therefore, optimizing the AUC from adaptive window does not require running through all the possible wavelengths. Instead, running a number of initialization points in a small range and looking for the best AUC may be sufficient; in our example, we considered 6 points in 20 nm. This 20 nm range is adjustable, and needs to correspond with the starting window size of the adaptive window-defining algorithm.

3.5 Discussion

We investigated the impact on classifier performance of using three different windowing techniques: adaptive windowing or fixed size windowing, compared to using no windowing. The results in Table 2 show that significant

differences in the AUCs ($p < 0.05$) were observed using classifiers trained on features extracted when windowing techniques were employed and classifiers trained on features extracted from the entire spectrum (i.e., no windowing) for two diagnostic tasks (Normal vs. MD+SD and Benign vs. MD+SD). In other words, from the observations in maximum AUCs, either windowing technique is better than no windowing in terms of providing accurate diagnostic information. This result agrees with previous studies on different optical spectroscopy datasets [11, 31] that chose fixed-size windowing techniques. Bigio *et al.* [1] divided the measured spectrum into fixed-size wavelength bands of 20 nm from 330 - 750 nm, followed by feature extraction of average intensity within the wavelength band, and then PCA to reduce dimensionality for input to an Artificial Neural Network (ANN). In addition, Johnson *et al.* [31] also used PCA to select only the wavelength regions with large variability. These studies divide spectra into wavelength bands under the assumption that some wavelengths bands have more diagnostic power. The statistically significant AUCs observed in our study verify this hypothesis.

We also found that significant differences were not observed between adaptive and fixed-size windowing, suggesting that these two windowing techniques perform equally well in the two diagnostic tasks explored. This underscores the main focus of this study: that adaptive windows capture piecewise linear information in a more adaptive and flexible fashion. From the comparison of classifiers based on fixed-sized windows and adaptive-sized windows, we conclude that adaptive windowing more efficiently captures diagnostically rel-

evant spectral features without a statistically significant decrease in classifier performance.

Our adaptive window technique defines spectral feature extraction regions using linear regression (Figure 3.1). We utilize features that capture the linearity of the spectrum. Therefore, it is intuitive that this adaptive windowing technique captures regional changes more efficiently than fixed size windowing. To further support this intuition, we also showed both theoretically (Section 3.2) and experimentally (Section 3.4) that the adaptive windows have less redundancy than fixed windows. The purpose of our theoretical derivations is to prove that features extracted from two fixed-size windows can be replaced by features extracted from a single adaptive window. To achieve this, we examined a common situation (Figure 3.1) where the signal in two fixed-size windows (window 1, window 2) and one adaptive-sized window (window 3) share the same regression line (the solid line). In this case, we demonstrated that the adaptive-sized window has all the information that the fixed-size windows can provide, while reducing the data size by half. In our derivation, we successfully showed that all of the features we chose have a linear relationship, proving that the adaptive-sized window can completely represent the fixed-size windows. Adaptive windows can be viewed as linear combinations of smaller and potentially redundant fixed sized windows. We further showed a decrease in the redundancy of diagnostic information in Table 3.3. The average and total number of windows used is lower for adaptive windows than fixed size windows. Furthermore, adaptive windows provide more unique features extracted

from the overall dataset. While the dimensionality of the feature space was reduced, there were no significant AUC differences observed between adaptive windows and fixed sized windows. Therefore, adaptive windowing is capable of teasing out unwanted redundancies that are inherent in fixed-sized windowing methods.

To investigate the uniqueness of the adaptive windows, we applied different initialization points to adaptive windowing. In an effort to minimize computation time, we restricted optimization of the initialization point to the range 400 - 420 nm. We also investigated two initialization points within a larger range (562 nm, 725 nm). The results of visual assessment showed that with a variation of 20 nm in initialization points, the resulting adaptive windows only have a 10 nm range shifts. The highest AUC (0.82) was found for an initialization point of 407 nm. An exhaustive search over the entire wavelength range may provide better AUC. However, while larger variations in the initialization point (562 nm and 725 nm) shows detectable differences of window definitions, their AUCs do not show significant improvement for this data set. These results demonstrate that the choice of initialization point is important and can affect AUC. AUCs for fixed windows do not change with different initialization points. Depending on the initialization point, the adaptive window AUCs can have better performance than both the no window technique and the fixed window method. Consequently, the adaptive window method has flexibility to achieve higher accuracy to predict the disease.

Decreasing redundancies in features can be very beneficial. In practical

instrumentation design, memory can be an important concern when trying to implement rapid diagnostic analysis of patient spectra. Adaptive windowing has fewer features extracted, thus, less computational memory is required. In addition, a large number of features with limited number of subjects often lead to overtraining of classifiers. Reducing the dimensionality of feature space helps alleviate overtraining. In previous studies we surveyed, overtraining concern is dealt with by applying PCA to reduce the dimensionality. However, PCA removes all physical property information, limiting understanding of underlying biophysical processes during disease progression. Therefore, the adaptive windowing technique is preferred over PCA for reducing the dimensionality. A further benefit of adaptive windowing is that it permits the isolation of unique diagnostic features without a priori knowledge of tissue properties. This unbiased perspective can be used as a complement to physical models of light tissue interaction, aiding the elucidation of the biophysical processes underlying disease development.

In this study, we tested an adaptive windowing algorithm on a diffuse reflectance spectroscopy dataset. We note that our adaptive method may be suitable for other spectroscopy signals if the spectra have a smooth pattern similar to that found in diffuse reflectance spectroscopy or fluorescence spectroscopy. In other words, when the tendency of spectrum is smooth (not having too many peaks in a short wavelength range), the adaptive windowing technique significantly reduces the number of windows used relative to a fixed size windowing approach. When the signal has high variation in one specific

wavelength band, such as in a Raman spectrum, the number of adaptive windows becomes larger than that of fixed windows. However, in that situation, adaptive windows might be beneficial for segmentation of peaks.

3.6 Conclusion

In recent years, there has been a debate on which diagnostic algorithm to use for bio-optical cancer detection modalities [12]. Various analysis methods have been used, including model-based analysis, statistics-based analysis, and hybrid analyses (Section 1.2.2). This chapter focuses on providing a new aspect for statistics-based analysis. First, we verified the hidden assumption by Bigio *et al.* [11] and Johnson *et al.* [31] that wavelength bands (defined by fixed-size windows) need to be separated in analyses because they have different diagnostic power. Second, we proposed a new adaptive windowing technique that avoids the feature redundancies from fixed-size windows feature extraction. Since adaptive windows retain most diagnostic information while reducing the number of windows needed for feature extraction, our results suggest that it is useful for data compression in optical spectra feature extraction.

As described in Chapters 2 and 3, I have developed algorithms for analyzing optical spectra for making clinical predictions for diagnosis. One thing remained a question: are these optical signals "good enough" for us to make predictions? are they probing the tissue regions where cancer and precancer appear? In the following chapter, we have put in effort to investigate

the light propagation in tissue to answer these questions.

Chapter 4

"Virtual probe design": Monte Carlo simulation in the design of diagnostic instrumentation

In the previous chapters, we focused on analyzing the optical spectral signals to make diagnostic predictions. In this chapter, a Monte Carlo simulation is build to model light propagation in multiple layered tissue. This simulation tool (Pol-MC) was originally developed by Dr. Daniel Côté [122], with several extended implementations described in this chapter. Pol-MC uses Mie scattering instead of Henyey-Greenstein as its phase function, which is closer to reality. This simulation method provides researcher a novel aspect of utilizing simulation as "virtual design tool" to guide the design of instrumentation. It offers desirable information that are unavailable in experiments, such as the maximum signal-to-noise ratio.

4.1 Background: modeling light-tissue interaction

Polarized light sources, and polarization-sensitive detectors, have been used to provide better separation between shallow and deep signals collected from tissue. This is highly advantageous, as cancer generally originates near

the surface layers of organs, and therefore, statistics with more diagnostic relevance will be obtained from these enhanced superficial signals. For example, polarized incident light and polarization-enabled collection fibers have been widely used to probe the structure of superficial tissue. This superficial signal enhancement is due largely to the fact that a significant difference between singly and multiply scattered light is that the latter does not preserve the incident polarization. It has been shown both experimentally and in simulation that polarized light is useful in studying tissue, and that tissue properties can be extracted through changes in the light's polarization state [20, 80, 83, 85, 111–115].

Radiative transfer theory (RTT) and the associated radiative transfer equation (RTE) [41, 116], is a general theory dealing with the transport of energy through physical media, such as scattering media. The RTT has been the most successful technique applied to the modelling of light propagation in tissue. However, due to its inherent complexity, various approximations have been proposed to simplify the RTE. These approximations concern constraints on the distribution of the sizes of the scatters, constraints on whether scattering or absorption dominates the optical properties, and constraints on whether polarization is or is not included in the model [117, 118]. For example, the diffusion approximation is a low order analytical solution to the RTE, which most commonly assumes multiple scattering events in the tissue and neglects any polarization of the propagating light. Such approximations however, are not always appropriate; and therefore, a numerical solution to the RTE

is desirable. The Monte Carlo technique provides a descriptive and rigorous numerical solution to the RTE. In general, photon Monte Carlo methods are a type of stochastic integration technique, which solve the RTE through the use of a 3D random walk of energy packets, or "photons", through the tissue volume. Although these methods provide accurate solutions, the drawback of Monte Carlo methods is that an extremely large number of photons must be simulated in order to produce statistically reliable results. Monte Carlo Multi Layer (MCML), implemented by Wang *et al.* [119], is one of the most commonly known implementations of Monte Carlo. Although MCML does not consider polarization, it is considered the standard of the field because it has been tested and validated extensively. More recent Monte Carlo implementations do incorporate polarization [112, 120–126], although many of these implementations are limited in other respects. For example, most polarized Monte Carol implementations offer only very limited geometric capability; for example, they do not allow modelling of multiple layers, or of complex objects or of more physically realistic scattering descriptions [112, 115, 120–122]. For these reasons, only a very few contemporary Monte Carlo implementations are practically useful for instrumentation design and validation.

When light is used for diagnosis, it is important to ensure that it travels through target areas where cancer and pre-cancer are located. Liu *et al.* [121] investigated the depth selectivity of polarization gating for tissue characterization, and concluded that polarization-gated signal is generated primarily by photons emerging from the surface of the medium within a few mean free path

lengths from the point of incidence. This study, however, was done considering only one layer of tissue. Guo *et al.* [126] studied the pathlength distribution of polarized light in turbid media, and concluded that the average pathlength of polarization-preserving photons was two to three times smaller than the average pathlength of all collected photons (i.e., where collected photons include all polarizations).

Several fundamental questions must be addressed in order to evaluate the effectiveness of the use of polarization for diagnostic purposes. What are the specific signals collected by parallel and perpendicular polarizers? Is the penetration depth of the light significantly different for multiple layer models? What is the relationship between penetration depth and pathlength? Can signals be discriminated through the use of singly scattered and multiply scattered light? And most importantly, is the instrument actually effectively collecting the signal from the target layer, which is the precancer layer? Answers to these questions will be critical to guide the development of new diagnostic instrumentation, and in addition to validate existing instrumentation designs. Particularly important are considerations of how probe geometry affects the collection of light. For example, in a previous study by our group, a probe was developed to improve the depth resolution by using beveled collection fibers [127], and it would be beneficial to be able to use simulation to validate the signal-to-noise ratio achievable through the use of this type of angled-fiber design in comparison to simpler probe designs using only a flat collection fibers.

In this study, we use a polarization-sensitive Monte Carlo simulation

(Pol-MC) to aid in the design of instrumentation for the early detection of epithelial cancer. In contrast to other Monte Carlo implementations, which primarily use the Henyey-Greenstein phase function, the Pol-MC numerical code accurately calculates scattering phase functions based on comprehensive Mie theory calculations, and fully retains polarization information at both scattering and interface transitions. In addition to these advanced features, we have extended the Pol-MC capability to include full multilayer capability, as well as the capability to use scattering mixture-models, which enable a physically realistic treatment for complex scattering media such as biological tissue.

By analysis of simulation results, we compared the use of angled and flat detector geometries in a hypothetical diagnostic instrument. Further, by taking advantage of the stochastic nature of the Monte Carlo technique itself, we are able to evaluate information not normally accessible through any physical measurement; for example, we are able to evaluate the maximum possible signal-to-background ratio achievable by the instrument through the sub-classification of scattered photon paths, as represented in the simulation, and this information is also applicable to assist in the development of algorithms for the analysis of the instrument's measurements. These promising results provide a proof-of-concept for the application of this type of stochastic computational technique to diagnostic instrument design.

4.2 Methods

4.2.1 Polarization sensitive Monte Carlo simulations (Pol-MC)

The numerical simulation of polarization sensitive Monte Carlo is performed using a program, Pol-MC, developed by Dr. Daniel Côté [122]. Several implementation extensions, including multiple layers, geometry checks, detector implementations, and analysis tools, are developed by the authors of this study.

4.2.1.1 Geometry

The simulation geometry to Pol-MC is specified by a user-supplied input file, in XML format, which may be quite general, and defines the relative positions and optical properties of a collection of geometric objects, such as boxes, infinite layers, and cylinders, at specified locations, along with a number of possible photon source types. In addition to geometric attributes, several optical attributes characterize each object: such as the size of scatters, refractive index for the scatter and its surrounding medium, and absorption coefficient μ_a and scattering coefficient μ_s. Boundaries of geometric objects are represented through the use of primitive surface elements, which then collectively specify the complete geometry. For example, a box has six surface elements, an infinite layer has two, and significantly more complicated "generic" objects are possible which can be represented through the import of any suitable surface mesh. In addition to its location, each surface element retains a set of local basis vectors consisting of two unit vectors ($\mathbf{a}$ and $\mathbf{b}$) in the plane of the sur-

face element and a unit normal vector($\hat{n}$), pointing towards the exterior of the object for which the surface element forms part of the boundary. The check for geometric consistency using these surface basis vectors is a critical part of our multilayer implementation, which is further discussed in Section 4.2.2.2.

4.2.1.2 Photon propagation

Once the simulation geometry and the illumination sources are fully specified, multiple photons propagating through the tissue geometry are simulated. At each scattering event, interface transition, and or collection or termination event, the program accumulates statistics on current properties of the associated photon. The "photon" structure itself holds the current position, propagation direction, and Stoke's parameters of the form [42]:

$$S = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} E_{\parallel}E_{\parallel}^{*} + E_{\perp}E_{\perp}^{*} \\ E_{\parallel}E_{\parallel}^{*} - E_{\perp}E_{\perp}^{*} \\ E_{\parallel}E_{\perp}^{*} + E_{\parallel}^{*}E_{\perp} \\ i(E_{\parallel}E_{\perp}^{*} - E_{\parallel}^{*}E_{\perp}) \end{pmatrix} \tag{4.1}$$

where $E_{\parallel}$ and $E_{\perp}$ are the complex electric field components. I represents the intensity of the beam, Q and U represent the linear polarization, and V represents the circular polarization. The specific definitions of $E_{\parallel}$ and $E_{\perp}$ are with respect to an arbitrarily selected *plane of reference* including the direction of propagation. The sense is chosen so that $E_{\perp} \times E_{\parallel}$ is in the direction of propagation. Whenever a scattering event takes place, this plane of reference is *defined* so that it is the plane containing both the incident and the scattered beams[41]. This plane of reference is no longer arbitrary thereafter.

Each photon interaction event occurs statistically, with a distance d determined from the probability distribution $exp(-\mu_t d)$, where the extinction coefficient μ_t is the sum of the scattering and absorption coefficients $\mu_s + \mu_a$. The photon is then moved by this distance d, while appropriately treating any intervening interfaces [119].

4.2.1.3 Mie scattering

As described previously, a key difference between an advanced contemporary Monte Carlo implementation such as Pol-MC and a classical Monte Carlo implementation such as MCML is in the treatment of polarization. When a scattering event takes place, both the polarization state and the propagation direction of the photon change. Higher order Mie theory calculations are used to model scattering in Pol-MC; thus, both the polarization state and the propagation direction are updated at each scattering event. As Mie theory in general describes quite complex scattering phase functions, which have no analytic inverse, stochastic sampling algorithms must be applied to generate distribution function samples corresponding to these phase functions. The Mueller matrix of the scattering event is obtained through the use of the SCATMECH library developed by Germer, which is available to the public [128]. Using the Mueller matrix of the scatterer and the current polarization state of the photon, a scattering plane and a scattering angle are statistically sampled. The probability density distribution $P(\theta, \phi)$ of scattering events in a small solid angle around θ and ϕ is $P(\theta, \phi) \equiv \mathbf{I} \cdot \mathbf{S'sin}\theta$. This probability

density distribution itself is sampled by the algorithm presented by Jaillon [123].

4.2.2 Contributions to Pol-MC

4.2.2.1 Multiple layers implementation

Côté's [122] original implementation of Pol-MC did not allow the convenient simulation of complex multilayer structures. Prior to the extensions implemented by the authors, all the simulation objects were required to be placed in the interior of the "World" object, which was defined to be vacuum medium; further, object surfaces could not be coincident, that is, no touching layers were possible. In order to correctly treat this latter case, when objects in the simulated space are touching, it is necessary to specially treat this case of coincident surfaces. In our extension, we handle this situation by automatically detecting surface elements that are coincident, then explicitly initializing the surface element structure attributes (`ObjectOutside`) to point the appropriate adjacent object. Figure 4.1 shows a graphic illustration of how this "touching" of layers is accomplished. Both layer objects have two surface elements: one on top and one on the bottom. When the two surface elements are detected to be at the same location, their `ObjectOuside` are modified so that they correctly point to the adjacent layer.

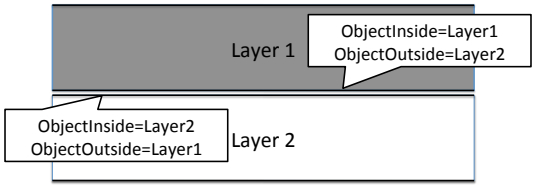


Figure 4.1: Illustration of how we handle layers "touching" each other.

4.2.2.2 Geometry checks

Any computer simulation involves calculation at a finite numerical precision. By implication, simulations containing an extremely large number of geometric computations must correctly deal with surface locations in a manner that is not subject to errors inherent in floating-point computation, such as numerical round-off error. To this end, the manner in which interface locations are tested, and the manner in which a photon makes an interface transition are quite carefully implemented in Pol-MC, and we also added additional refinements to this mechanism. We have incorporated two *safety distances* for geometry checks: PHOTON_SAFETY_DISTANCE is used for securely moving the photon from one side of an interface to another, OBJECT_SAFETY_DISTANCE = 0.5·PHOTON_SAFETY_DISTANCE is used for checking if a certain location is within the interior of an object, and this modification is necessary in order to take correct consideration of surfaces that touch each other, which might otherwise be treated as a geometry initialization error. For example, in Fig. 4.1, the surface elements that touch each other occupy the same location in space, which is *not* due to geometry setup errors, but is at the explicit request of the end-user.

4.2.2.3 Detector implementation in Pol-MC simulation

In order to realistically simulate physical instruments, it is necessary to correctly model various detection geometries; for example, detectors in the simulation must behave analogously to collection optical fibers used in experiments. The detector objects used in Pol-MC are fundamentally "box" objects with special settings that allow the collection of photons with no further propagation; that is, a detector is the end-point for any detected simulation photon. Although at first this may seem simple, it should be noted that implicit to every surface element participating in the Pol-MC simulation geometry is the capability of accumulating full angular and polarization statistics on every photon passing through the surface, and this capability is shared by the surface elements defining the detector. Although many real-world detectors have much more complicated geometries, such as the circular surfaces of optical fibers, this simple detector implementation allows for computational simplicity. We use both flat and tilted detectors in the present simulation. In order to tilt a detector, i.e., to place a detector that isn't perpendicular to the xy-plane, the user explicitly specifies Euler rotation angles (α, β, γ) as part of the detector specification in the input file. The specific angled detector used by our simulation is tilted at 45° from the z-axis.

4.2.2.4 Mixture model

One of the most significant extensions we made to the Pol-MC program is that of a scattering mixture-model. This model allows the arbitrary mixing

of any number of component scatterer types, including even other mixture-model scatterers. For example, to accurately model a distribution of nuclei sizes in scattering tissue, the scatterer mixture would be adjusted to match the size distribution function governing the scatterers. Similarly, a simpler case with perhaps more immediate physical relevance to the present simulation involves the mixture of a small amount of Rayleigh-type scattering associated with the tissue matrix, with the Mie-regime scattering from the cell nuclei in the various model layers. This type of scattering model may be especially important for modeling precancers, since histologically the precancerous basal layer begins with a mixture of scatterers associated with different cell types. Further, the properties of each type of cell may be governed by a statistical distribution. For example, one cell type may scatter analogously to a size distribution of spheres (with optical properties equivalent to normal cell nuclei), and another cell type has a distribution representing abnormal nuclei with appropriate sizes and other properties. The mixture model implemented here is a piece-wise uniform distribution over a user defined number of scatterers, where the probability of the participation is specified through its associated scattering coefficient μ_s. It will immediately be seen that any probability distribution can be modeled to any desired degree of accuracy using this type of mixture, simply by including enough components. Since $\mu_s = \sum_{i=1}^{N} \mu_{si}$, where i is the index of scatterer ranging from 1 to N, and μ_{si} is the scattering coefficient of the scatterer i, we can scale the scatterers with the random variable:

$$Pr(X = k) = \frac{\mu_{sk}}{\sum_{i=1}^{N} \mu_{si}} \tag{4.2}$$

where X is the uniform random variable representing the probability that the scatterer is of type k. Although our implementation of this extension to Pol-MC is complete, we do not use this model in the simulations shown below, and for this reason further discussion is beyond the scope of this study.

4.2.2.5 Parallel computing

The main drawback of Monte Carlo methods is that they are extremely computationally intensive. However, it is natural to parallelize photon Monte Carlo since each photon is independent from every other. Thus, photon Monte Carlo can be run on as many separate processes occupying separate cores and separate nodes in a cluster computer, the results of which are subsequently combined. To implement this high degree of scalability, it is important to make the properly initialize the random number generator; otherwise, running many jobs concurrently may cause re-use of the same random number seed, resulting in wasteful duplications. Naive initialization of the random seed is often performed from the system time stamp or from the system process ID number, either of which can be duplicated over a large number of processes running on different nodes. In order to ensure that distinct seeds are used, regardless of the number of independent processes, we modified Pol-MC so that the seed initialization step used the system random device "/dev/urandom"; this choice of initialization effectively reduces the probability that different runs can accidentally use the same random number seed to zero.

4.2.3 Analysis for the simulation outputs

For this type of simulation, two types of analysis are pertinent. First, one may evaluate the experiments with which simulated variables correspond to physical measurements. We refer to this type of analysis as "outside-the-box". Second, one can investigate variables that it is impossible to physically measure with current technology. This "inside-the-box" output is possible because complete information about the history of each simulated photon path, including all scattering and interface transitions that the photon has experienced, is available. We will show that this "inside the box" information, when correlated with the model geometry, provides important conclusions with respect to instrument feasibility and can be used to guide the design of instruments and data analysis algorithms.

4.2.3.1 "Outside-the-box" analyses

For "outside-the-box" analyses, our simulation is configured to deliver variables analogous to those that would be measured by a given instrument. The simulation can provide not only the light intensity collected by the fiber, but also the entire set of Stoke's parameters, which provide information relevant to polarization sensitive detection. Physical property information such as reflectance or transmittance can also be calculated. In order to obtain full spectral information, the incident wavelength is varied, and the simulation is repeated until the full wavelength range is obtained. Although "outside-the-box" simulations are often less expensive than actually building the in-

strumentation and conducting experiments, this is not always the case, and the specifics must be judged carefully. The relative costs of running Monte Carlo simulations, which can often be very computationally intensive must be considered.

4.2.3.2 "Inside-the-box" analyses

A unique advantage of using Monte Carlo simulation is that it produces a complete list of photon traveling paths, i.e., the position, weight, and polarization state of every photon at every step at every move through the simulation geometry. This list of photon traveling record is not easily accessible experimentally, if at all. It should be noted that with respect to effectively using this type of information, it is perhaps even more important than it normally is that the model parameters accurately represent what is being simulated. For example, with respect to simulation of biological tissue specifically, although much information is available, almost as much still remains to be discovered [38].

The signal-to-background ratio (SBR) of an instrument is an important indicator of whether a given design is feasible. Here we give an example of how comprehensive photon path information can be used to estimate the maximum possible s/b achievable by an instrument. With respect to this analysis it must be stressed that whether or not an instrument can actually achieve this s/b depends on the specifics of the physical measurement; that is, whether or not the signal can be physically discriminated from the noise. Here we define

the signal as photons detected from a specific layer of tissue via scattering processes. Noise is defined as all detected photons that are not signal. In our analysis of the photon paths recorded, the collected photon packets are classified into the following categories:

I Photons that passed above the basal/precancer layer;

II Photons that scattered *once* from the basal/precancer layer;

III Photons that scattered *more than once* with at least once from the basal/precancer layer; Note that III and II are similar, but II is defined more strictly, and is included in III. Both III and II are possibly considered "signal".

IV Photons that passed below the basal/precancer layer.

The signal-to-background ratio (SBR) is calculated as

$$SBR_k = \frac{\Phi_k}{I - \Phi_k} \tag{4.3}$$

where I is the total intensity, k is the index of the photon category, and Φ_k is the surface fluence at the detector surface for all photons from category k, which is defined as:

$$\Phi_k = \sum_{i=1}^{N} w_i \text{ [Watt]} \tag{4.4}$$

where w_i is the weight of photon i that hits the surface of the detector.

By definition, fluence is the integral of the radiance over all directions [38]:

$$\phi(\mathbf{r}) = \int_{\mathbf{4}\pi} \mathbf{L}(\mathbf{r}, \hat{\mathbf{s}}) \mathbf{d}\omega \tag{4.5}$$

Following this definition, the volume fluence from our simulation can be calculated as:

$$\Phi(x, y, z) = \sum_{i=1}^{N} w_i(x, y, z) \text{ [Watt]} \tag{4.6}$$

where $w_i(x, y, z)$ is the weight of the photon (with index i) at position (x, y, z). Note that volume fluence is different from surface fluence defined in Eq. 4.4 by their spatial definitions, and with respect to volume fluence specifically, for reasons of computational expedience, most Monte Carlo simulations accumulate this parameter only at photon *scattering* events, and this preferentially scales the sum in 4.6 by a factor of μ_t. Keeping this latter in mind, this volume fluence gives us a spatially-resolved volume to show where scattering occurs most often. Combining Eq. 4.6 and the aforementioned photon categories, we can visualize and compare the different categories of photons detected and see where they were scattered from.

4.3 Results

4.3.1 Simulation parameters

In our simulation, the geometry is set up as illustrated in Fig. 4.2. A four-layer model is simulated: (1) epithelial layer, (2) thin basal layer or precancer layer, (3) basement membrane, implemented as a thin transparent layer, (4) stroma, implemented as a thick layer of scattering material. We also have a collimated light source with photons entering the tissue perpendicular to the surface. Two detectors are placed above the epithelium. The angle detector has a 45° angle with the epithelium, while the flat detector is aligned

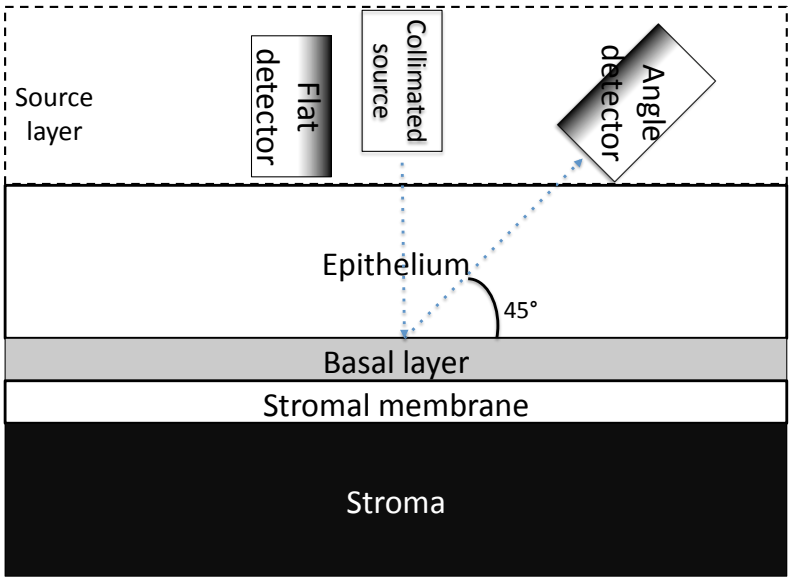


Figure 4.2: Illustration of the geometry in our simulation. Photons enter the tissue from a collimated source. The tissue consists of four layers: epithelium, basal/precancer layer, basement membrane, and stroma. These four layers are infinite layers and they touch each other with no gaps in between. Two detectors are placed above the epithelium with very small gaps. The angled detector has a 45° angle with the epithelium, while the flat detector is aligned with the surface of the epithelium. Note that a source layer is placed on top of the epithelium layer both to provide some degree of index-matching and for computational purposes.

Table 4.1: Tissue layer thickness used in this study. Four sets of simulation parameters are presented here.

thickness and radius (μm)	Model 1	Model 2	Model 3	Model 4
Source layer	10000	10000	10000	10000
Epithelium	250	250	230	190
Basal/precancer layer	20	20	40	80
Basement membrane	30	30	30	30
Stroma	1000	1000	1000	1000
Basal layer scatterer radius	6 (normal)	8 (abnormal)	8 (abnormal)	8 (abnormal)

with the surface of the epithelium.

In the simulation, the optical properties are defined in the input XML files. The wavelength used is 633 nm and all simulation results shown (excepting the wavelength-series) are simulated with 1.28×10^9 photons. Tab. 4.4 and Tab. 4.1 show some of the most important optical properties selected in this study. These parameters are chosen carefully to match the values reported in literature, on the conservative side [38]. The properties that we have used for the stromal membrane are selected to be consistent with those of other biological membranes. Although these properties are plausible, we believe that the specific properties of this membrane may be unknown.

4.3.2 Scattering event distribution

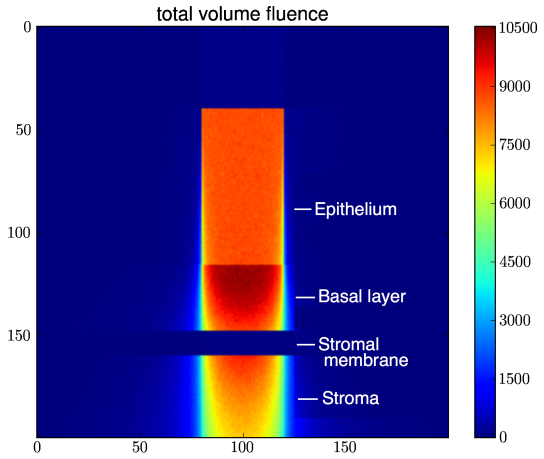
We generate the total fluence plots (as defined in Eq. 4.6 to check the scattering distribution spatially. The total scaled fluence for Model 4 is shown in Fig. 4.3(a). In the figure, we can interpret the intensity as the density of scattering events in each location. The basal/precancer layer has a large

number of scattering events both because it is closer to the source, and because its scattering coefficient (μ_s) is 120 cm^{-1}, is slightly higher than that of the epithelium which is 100 cm^{-1}. The dark gap between the basal/precancer layer and the stroma corresponds to the stromal membrane; with μ_s of zero, the scaled fluence for this region is also zero. Finally, although we would expect the stroma to have high scaled fluence because it is highly scattering ($\mu_s = 150$), it does not show high intensity in the figure; this is mostly because there are less photons that enter the stroma.

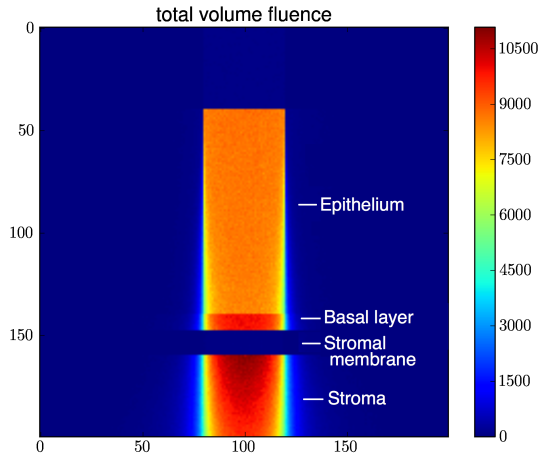
With respect to the scattering intensity in the stroma, this idea is borne out by comparison between the two models; if we compare Model 4 (Fig. 4.3(a)) with Model 2 (Fig. 4.3(b)), with all the parameters being the same except the basal/precancer layer thickness, we find the scaled fluence in the stroma much higher in Model 2 than in Model 4. This shows that because of the thicker basal/precancer layer, and also due to the more complicated scattering geometry, fewer photons are likely to penetrate through the basal/precancer layer and enter the stroma in Model 4.

4.3.3 Studying detector angles

In order to study the specifics of detector geometries, we classified the photon traveling paths and plotted the fluence plots only for photons that are collected by a given detector. This method allows us to specifically evaluate the photons detected, which means it effectively allows us to analyze where scattering events have happened, after the fact of detection, for de-



(a)



(b)

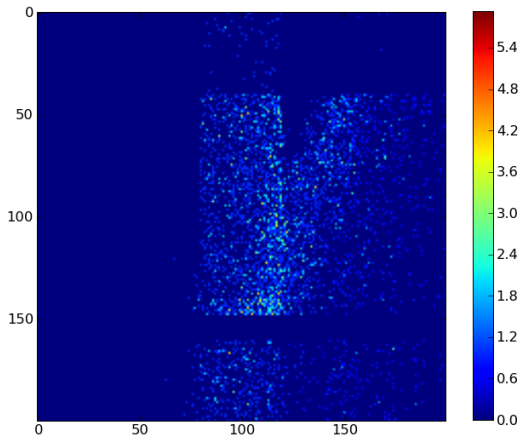
Figure 4.3: Total fluence for Model 4 (Fig. 4.3(a)) and Model 2 (Fig. 4.3(b)). Photons enter the tissue from the top. These volume fluence plots represent the total spatial scattering density. The basal/precancer layers in both models have high fluence and show a clear boundary between the epithelium and basal/precancer layers. Stromal membranes do not scatter, so they have a dark color in these figures. Stroma layers are also highly scattering, but the intensity also depends on how many photons actually enter the layers. Since Model 4 has a thick basal/precancer layer, fewer photons enter the stroma and less scattering is indicated.

tected photons. Fig. 4.4(a) and Fig. 4.4(b) are two examples of fluence plots from detectors, one from the angled detector and one from the flat detector. We see very clearly from both images that the photons form two beams: one going into the tissue, and one scattered near the bottom of epithelium and in the basal/precancer layer. We can also tell from these figures that the fluence is always most significant where the two beams overlap. This means that since the flat detector is placed right next to the source, its distance from the incident beam is smaller, allowing more overlap between the beams. Since the difference between these two fluence plots is largely caused by distances between the detectors and the light source, it suggests that the rotated angle of the detector is not as important a factor as the distance from the source. We find this conclusion to be particularly interesting, as it is counter-intuitive to analysis based strictly on ray-tracing considerations.

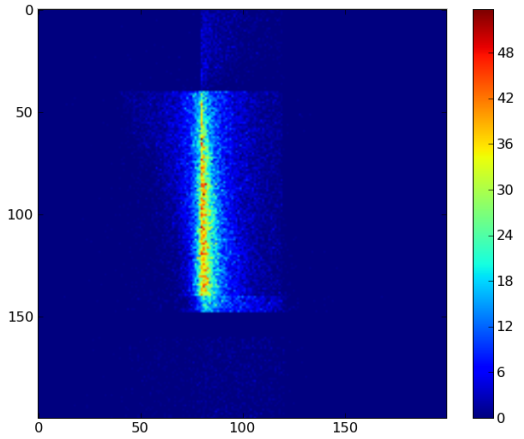
4.3.4 Detectors with collimation

Collimated light is light traveling in a nearly parallel direction. For a source in a Monte Carlo simulation, which can be seen as a type of ray-optics computation, a collimated source is often perfectly collimated, that is, it is comprised of exactly parallel rays, with no divergence at all. This means that any spread in the beam diameter with propagation is associated with either scattering events, or with interface transitions.

In simulation, we can study the effects of collimation by adjusting the acceptance angle range for photons incident on the detector. In a physical



(a)



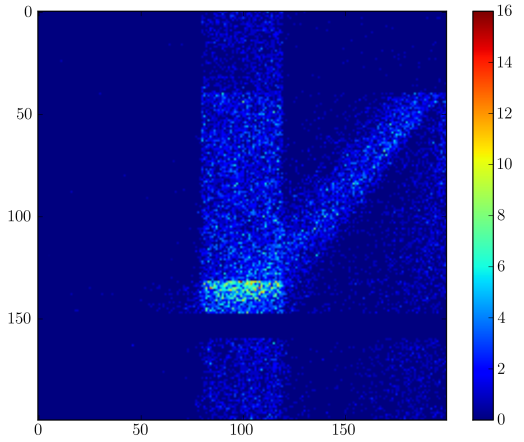
(b)

Figure 4.4: Fluence for Model 1 with an angle-detector(Fig. 4.4(a)) and flat-detector (Fig. 4.4(b)). This image shows the fluence associated with all photons collected by a detector. The photons form two beams: one going into the tissue, and one scattered from the bottom of epithelium and in the basal/precancer layer. The fluence is always most significant when the two beams overlap. This plot is for uncollimated detectors, which means all photons arriving at the detector are collected, regardless of their incident angle at the detector surface.

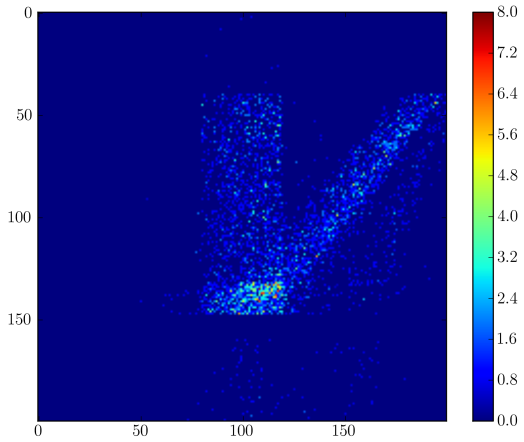
measurement, this selection is automatic due to the *numerical aperture (NA)* of the physical detector. In Fig. 4.5(a) and Fig. 4.5(b), we show the fluence plots of photons that scatter from the basal/precancer layer, detected by an angle detector. In Fig. 4.5(a), we can clearly see many photons that have scattering events right below the detector but not on the straight line path. These scattered photon are eliminated in Fig. 4.5(b) with a collimated detector. These results further confirm our speculation in Section 4.3.3 that if detector collimation is not considered the detectors would perform similarly regardless of rotation angles. However, when the detectors are collimated, which here specifically means when we consider only photons with acceptance angles with minimum acceptance cosine = 0.95, fewer photons are collected by the detector, and further, photons that carry information not relevant to the target layer are eliminated by the collimated detectors. This acceptance cosine of 0.95 corresponds to NA = 0.35, which is within the range of optical fibers.

4.3.5 SBR analysis

The SBR can be calculated by classifying the photon traveling paths in order to determine if the signals collected by detectors were actually from the region of interest. Here we present the SBR from an angle detector, with the signal defined as photons scattering from the basal/precancer layer. From Tab. 4.2 and Tab 4.3, we can see that SBR for collimated detectors are significantly higher than SBR for uncollimated detectors as much as 12–25X higher.



(a)



(b)

Figure 4.5: Fluence for Model 1 with an uncollimated angle-detector(Fig. 4.4(a)) and collimated angle-detector (Fig. 4.4(b)). These images show the fluence caused by all the photons detected by a detector. The collimated detector accepts photons that have acceptance cosine values between 0.95 and 1, whereas the uncollimated detector accepts all photons that hit the detector, regardless of acceptance cosine. These figures show that collimated detectors eliminate many photons that scatter from the stroma and other undesirable locations, and that they enhance the collection of photons scattered from the basal/precancer layer. This acceptance cosine of 0.95 corresponds to NA = 0.35, which is within the range of optical fibers.

Table 4.2: SBR calculation for an angle detector: collimated (minAcceptanceCosine: 0.95). All four models are used here.

	Epithelium	Basal	Stroma	SBR
Model 1	8986	33158	1746	3.09
Model 2	636	7007	1192	3.83
Model 3	558	13459	1156	7.85
Model 4	385	23491	1003	16.92

This result agrees with the results in Section 4.3.4 that collimated detectors remove unwanted noise associated with excessive scattering from outside the region of interest.

Also, in Tab. 4.2 and Tab. 4.3, we see that the thickness of the target layer (in this case, basal/precancer layer) dominates SBR. Model 4 has the thickest basal/precancer layer of 80 μm, while that of Model 1 and 2 is 20, and Model 3 is 40. An additional factor for SBR may be related to the scatterer size. The difference between Models 1 and 2 is the scatterer size of their basal/precancer layers. Model 2 has an abnormal scatterer size of 8 μm, while Model 1 is 6. However, due to the combined interactions of increasing basal/precancer layer thickness, and increasing scatterer size, which are present in these simulations, the isolated effect of either factor alone cannot be determined without further simulations.

4.3.6 Wavelength series analysis

By varying the wavelength in the simulation, and running multiple simulation runs, a full spectrum can be calculated. This corresponds to what

Table 4.3: SBR calculation for an angle detector: uncollimated (minAcceptanceCosine: 0.0). All four models are used here.

	Epithelium	Basal	Stroma	SBR
Model 1	163053	51622	43415	0.25
Model 2	111217	21343	25490	0.16
Model 3	110727	42719	23006	0.32
Model 4	108527	87431	19167	0.68

we refer to as a "physical simulation" in the preceding discussion: that is, it simulates what could actually be measured by an instrument. The simulated spectra are shown in Fig. 4.3.6 with both flat and angle detectors of Model 1 and Model 3. The wavelength range is 500–700 nm, and 10 spectral points are simulated in this range, which results in a resolution of 20 nm. Although there are "ripples" in these spectra, which may be caused by Mie scattering effects; we find it unlikely that these spectra are representative of physical spectra from biological tissue. It seems likely that this type of ripple effect would usually be are "washed out" through the combined action of different sizes of scatterers and multiple scattering effects. For this reason, our implementation of the scattering mixture model may be particularly relevant to future simulation work.

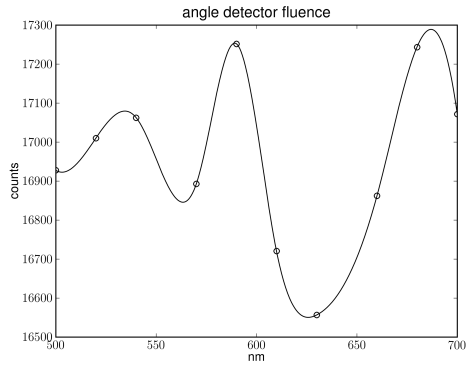
We notice that in the angle-detector spectra, the photon counts are significantly lower than those for flat-detectors. This is predominantly due to the fact that the angle-detectors are further away from the source. The low number of photons used to generate these spectra make us skeptical of the correctness of these plots. However, it may not be practical, at least in

this study, to simply increase the number of photons used in the simulation until the spectra are fully resolved. Note that these simulations are extremely intensive in terms of computational time. Each wavelength takes 10^9 photons, which means having 10 points in a spectrum takes 10×10^9 photons, which is very time consuming. At present, it appears that using the Monte Carlo technique to run full spectral simulations with any but the simplest source and detector configurations (i.e. even simpler than those used here) is essentially unfeasible.

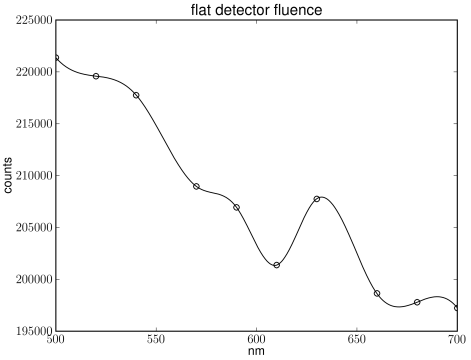
4.4 Discussion and conclusion

Tissue optics has demonstrated its potential in recent years, and various experimental instrument designs have been developed to non-invasively measure signals specific to precancer diagnosis. However, whether the instrument is actually specifically probing the desired signal or not has remained an open question. Researchers indirectly measure instrument parameters such as depth sensitivity, but there is no standard metric to use in instrument design. This fact is particularly evident when considering complex photon path behavior in scattering media, where it is not at all evident that considerations based on simple ray-optics estimates, or analogous measurement, are applicable.

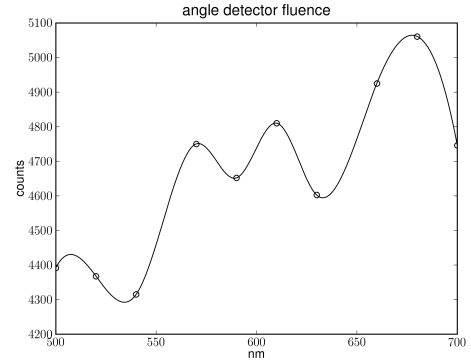
Pol-MC is a tool that provides additional, highly valuable information relating to the history of the photon traveling paths, including polarization. For example, by classifying photon paths into various categorize it is possible to understand where the scattering events have the highest density, and the


angle detector fluence
counts
nm
17300
17200
17100
17000
16900
16800
16700
16600
16500
500
550
600
650
700

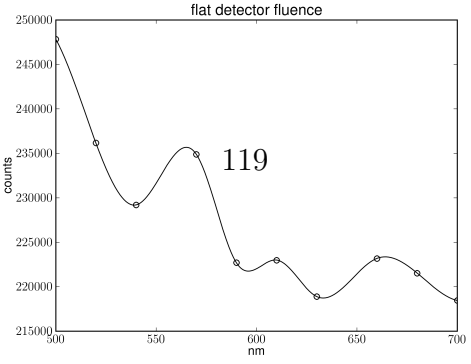
(a)


flat detector fluence
counts
nm
225000
220000
215000
210000
205000
200000
195000
500
550
600
650
700

(b)


angle detector fluence
counts
nm
5100
5000
4900
4800
4700
4600
4500
4400
4300
4200
500
550
600
650
700

(c)


flat detector fluence
counts
nm
250000
245000
240000
235000
230000
225000
220000
215000
500
550
600
650
700

(d)

associated concept of the SBR allows us to evaluate whether a given source / detector configuration is collecting from the region of interest. All of this information has great potential to offer investigators a way to understand the full degree of interaction of light with biological tissue, and in addition it offers a standard tool to evaluate multiple objectives in instrument design. Furthermore, as most of the widely-used Monte Carlo simulations use the unpolarized Henyey-Greenstein phase function as the scattering angle sampling mechanism, Pol-MC offers a much more realistic model for scattering, including higher order Mie scattering, with full polarization information.

Our study of comparing angle- and flat-detectors shows that for uncollimated detectors, the rotation angle of the detector has little effect on the collection efficiency. However, by adjusting the acceptance angle for photon incidence, a collimated detector can be simulated. Specifically, when the minimum acceptance cosine at the detector has a value of 0.95, the angled-detector is able to eliminate a large fraction of the photons that are scattered from locations outside the region of interest. The results from SBR confirmed this: the collimated detector has a SBR that is much higher than that of the uncollimated detector. This observation is in correspondence to physical intuition, as angled detectors aim at the region overlapping with source illumination [127]; however, there are fine details of the behavior that are counter-intuitive to expectations based solely on a ray-optics analysis.

In our wavelength series analysis, we were able to simulate a series of wavelengths in the range 500–700 nm. However, comparing the spectra from

flat- and angle-detectors, it seems very likely that the spectra from angle-detectors are exhibiting variations which have much more to do with the small photon number contributing to the spectra than to actual spectral features which may be present. Since collimated angle-detectors provide very good SBR, this type of detector does eliminates many unwanted photons, and this effect in terms of noise reduction is a reliable result. Nonetheless, this elimination of undesired photons has significant drawbacks in terms of the computational resources actually required to effectively simulate these detectors, in order to achieve the same level of signal strength as uncollimated or flat detectors. From Tab. 4.3, we can find that approximately half of the photons of uncollimated detectors are eliminated when we collimate the detector. This implies that we may need twice as many photons at the detector, which means more than 10 times in the simulation, for angle-detectors to achieve the same signal strength to get a reasonable spectrum. As the wavelength series analysis is already very computationally expensive, this raises a potential concern for the feasibility of Monte Carlo technique in general to provide spectra with good quality, even for this simple simulation geometry.

There are two random factors that we want to consider in the simulation. One is the variance of scatterers, and the other is inter-patient variance. We have briefly discussed the mixture model we contributed to Pol-MC (Section 4.2.2.4). This mixture model is not investigated in this chapter because it is not in the scope of the preliminary part of this study that we have discussed. However, the mixture model brings in another random variable to model the

variance of scatterers. With the use of a model, it will be possible to simulate more than one type of scatterer mixed in an object, which is closer to reality. It will immediately be seen that issues relating to inter-patient variance are significantly more complicated than a simple physical technique such as a mixture model. However, if we extend the stochastic formalism of which the Monte Carlo technique is an example to its full conclusion, it is possible that even a factor such as inter-patient variance could be included in the simulation as simply another random variable. In this case, the simulation would generate a "population" of measurements, which would further facilitate the development of diagnostic algorithms.

The simulation tool we present can be extended to consider various modalities. Considering its ability to provide information such as polarization states of the photon, it can be used to study effects caused by changes of optical properties. Our analyses provide a platform for comparing optical property changes: fluence plots and SBR can be used as metrics to study changes of optical parameters. This "virtual probe design" tool may guide future clinical applications for detection of epithelial cancer.

Table 4.4: Tissue properties used for this study. *Note that some of the parameters vary between different runs, the layer thickness are specified in Tab. 4.1. **HG: Henyey-Greenstein.

	thickness (μm)	scattering	g	index(media)	index(scatterer)	μ_a	μ_s	scatterer radius(μm)
Source layer	10000	HG**	0.5	1.335	—	0.1	1	—
Epithelium	varies	Jaillon	—	1.369	1.41	1	100	6
Basal/precancer layer	varies	Jaillon	—	1.369	1.41	1	120	varies
Basement membrane	30	HG**	0.5	1.41	—	1	0	—
Stroma	1000	Jaillon	—	1.369	1.41	1	150	6

Chapter 5

Conclusions

Optical spectroscopy is a technology that has high clinical potential for the detection of epithelial cancers. This dissertation extends previous instrumentation design studies, and focuses on the signals measured by these instruments. We answered some important questions in the field; for example, how these signals can be interpreted and processed in a clinically relevant manner, and how simulation can help improve the instrumentation design.

In Chapter 2 , we report the results of a pilot clinical trial carried out to detect early precancer and cancer in an oral cavity using optical polarized reflectance spectroscopy (OPRS). The epithelial cell nuclear sizes and ten spectral features were extracted for classification of their diagnostic categories. The ten features consist of two features extracted from each of the five spectral types: average intensity, and spectral intensity from a single, most discriminatory wavelength. Our results suggest two findings: First, a combination of features is needed for better performance in diagnostic classifications. Second, the diagnostic category "Benign" can be distinguished from severe dysplasia, which is one of the most clinically challenging classification tasks. These promising results demonstrated the ability of OPRS to augment

current clinical practice for diagnosis and monitoring of oral premalignancies and malignancies.

A novel algorithm is developed to extract features from optical spectroscopy in a more efficient manner (Chapter 3). We present an approach to adaptively adjust the spectral window sizes for optical spectra feature extraction. This adaptive windowing technique ensures the signal linearity in defined windows; hence, the adaptive windowing technique retains more diagnostic information while using fewer windows. This method was tested on a dataset of diffuse reflectance spectra of oral mucosa lesions. While adaptive and fixed size windowing enable similar classification accuracy, adaptive windowing utilizes significantly fewer windows than fixed sized windows. Since adaptive windows retain most diagnostic information while reducing the number of windows needed for feature extraction, our results suggest that it is useful for data compression in optical spectra feature extraction.

A simulation is developed to aid in the design of diagnostic instrumentation (Chapter 4). This study includes Mie scattering and polarization in Monte Carlo simulation, and it investigates the signal-to-noise ratio and fluence plots of a flat- and an angled-detector. This simulation is a novel application utilizing computational techniques to gain information from "inside-the-box" that is unavailable in the real world. The results show that this Monte Carlo simulation can be a useful tool to study where the collected photons are scattered from, and the signal-to-noise ratio provides a metric to evaluate the accuracy of the instrumentation. This study lays the groundwork for "virtual probe

design, " which has the potential to become the standard tool to guide design of optical instrumentation.

Future research could build upon this dissertation in several directions. For example, the Monte Carlo program could be modified to accommodate changes to tissue geometry, so it will be able to model arbitrary geometries. Because of the current limitations of Pol-MC, it is difficult to make it flexible. A better design of the code structure is desired to model various tissue geometries extensively. It will also be interesting to investigate the relationship between spectral features and Monte Carlo for virtual instrumentation design. If designed correctly, this could lead to optimization of instrumentation such that it collects spectra featuring specific spectral patterns, which makes spectral signal processing and pattern recognition easier. Furthermore, the Monte Carlo method can be studied extensively across a wide range of tissue properties with a mixture model. The implementation of mixture model is introduced in this dissertation, however, since it is extremely expensive in computational resources, the mixture model is not investigated thoroughly. To achieve that, we need either more computing resources, or we need to entirely parallalize Pol-MC with more powerful computing devices such as GPU (Graphics processing unit).

Bibliography

[1] I.J. Bigio, S.G. Bown, G. Briggs, C. Kelley, S. Lakhani, D. Pickard, P.M. Ripley, I.G. Rose, and C. Saunders. Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results. *Journal of Biomedical Optics*, 5:221, 2000.

[2] Chih-Wen Kan, Linda T. Nieman, Konstantin Sokolov, and Mia K. Markey. *AI in Clinical Decision Support: Applications in Optical Spectroscopy for Cancer Detection and Diagnosis*, chapter 2, pages 27–50. Springer, 2008.

[3] A. Jemal, R. Siegel, J. Xu, and E. Ward. Cancer statistics, 2010. *CA: A Cancer Journal for Clinicians*, 2010.

[4] R. Richards-Kortum and E. Sevick-Muraca. Quantitative optical spectroscopy for tissue diagnosis. *Annual Review of Physical Chemistry*, 47(1):555–606, 1996.

[5] RM Lycette and RB Leslie. Fluorescence of malignant tissue. *Lancet*, 40:436, 1965.

[6] I.J. Bigio and J.R. Mourant. Ultraviolet and visible spectroscopies for tissue diagnostics: fluorescence spectroscopy and elastic-scattering spectroscopy. *Phys. Med. Biol*, 42(5):803–814, 1997.

[7] LP Choo-Smith, HGM Edwards, HP Endtz, JM Kros, F. Heule, H. Barr, JS Robinson Jr, HA Bruining, and GJ Puppels. Medical applications of raman spectroscopy: from proof of principle to clinical implementation. *Biopolymers (Biospectroscopy)*, 67(1):1–9, 2002.

[8] N. Ramanujam. Fluorescence spectroscopy of neoplastic and non-neoplastic tissues. *Neoplasia*, 2(1/2):89–117, 2000.

[9] K. Sokolov, M. Follen, and R. Richards-Kortum. Optical spectroscopy for detection of neoplasia. *Curr. Opin. Chem. Biol*, 6(5):651–658, 2002.

[10] G.A. Wagnieres, W.M. Star, and B.C. Wilson. In vivo fluorescence spectroscopy and imaging for oncological applications. *Photochem. Photobiol*, 68(5):603–632, 1998.

[11] IJ Bigio and SG Bown. Spectroscopic sensing of cancer and cancer therapy: current status of translational research. *Cancer Biol Ther*, 3(3):259–67, 2004.

[12] Irving J. Bigio. Clinical applications of elastic-scattering spectroscopy beyond proof-of-principle: what really matters. *Biomedical Applications of Light Scattering*, 6446(1):64460J, 2007.

[13] D. Arifler, R.A. Schwarz, S.K. Chang, and R. Richards-Kortum. Reflectance spectroscopy for diagnosis of epithelial precancer: model-based

analysis of fiber-optic probe designs to resolve spectral information from epithelium and stroma. *Applied Optics*, 44(20):4291–4305, 2005.

[14] A. Dhar, K.S. Johnson, M.R. Novelli, S.G. Bown, I.J. Bigio, L.B. Lovat, and S.L. Bloom. Elastic scattering spectroscopy for the diagnosis of colonic lesions: initial results of a novel optical biopsy technique. *Gastrointestinal Endoscopy*, 63(2):257–261, 2006.

[15] V. Backman, MB Wallace, LT Perelman, JT Arendt, R. Gurjar, MG Muller, Q. Zhang, G. Zonios, E. Kline, T. McGillican, et al. Detection of preinvasive cancer cells. *Nature*, 406(6791):35–36, 2000.

[16] I. Georgakoudi, B.C. Jacobson, J. Van Dam, V. Backman, M.B. Wallace, M.G. Muller, Q. Zhang, K. Badizadegan, D. Sun, G.A. Thomas, et al. Fluorescence, reflectance, and light-scattering spectroscopy for evaluating dysplasia in patients with barrett's esophagus. *Gastroenterology*, 120(7):1620–1629, 2001.

[17] Martin Hunter, Vadim Backman, Gabriel Popescu, Maxim Kalashnikov, Charles W. Boone, Adam Wax, Venkatesh Gopal, Kamran Badizadegan, Gary D. Stoner, and Michael S. Feld. Tissue self-affinity and polarized light scattering in the born approximation: A new model for precancer detection. *Physical Review Letters*, 97(13):138102, 2006.

[18] M.G. Mueller, T.A. Valdez, I. Georgakoudi, V. Backman, C. Fuentes, S. Kabani, N. Laver, Z. Wang, C.W. Boone, and R.R. Dasari. Spectro-

scopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma. *Cancer*, 97(7):1681–1692, 2003.

[19] LT Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, et al. Observation of periodic fine structure in reflectance from biological tissue: A new technique for measuring nuclear size distribution. *Physical Review Letters*, 80(3):627–630, 1998.

[20] G. Zonios, L.T. Perelman, V. Backman, R. Manoharan, M. Fitzmaurice, J. Van Dam, and M.S. Feld. Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo. *Appl. Opt*, 38(31):6628–6637, 1999.

[21] X.H. Hu and J.Q. Lu. Optical detection of cancers. *Encyclopedia of Biomeaterials and Biomedical Engineering*, pages 1–10, 2005.

[22] RS Cotran, V. Kumar, SL Robbins, et al. Pathological basis of disease. *Philadelphia: WB Saunders*, pages 231–240, 1994.

[23] D. Arifler, M. Guillaud, A. Carraro, A. Malpica, M. Follen, and R.R. Richards-Kortum. Light scattering from normal and dysplastic cervical cells at different epithelial depths: finite-difference time-domain modeling with a perfectly matched layer boundary condition. *Journal of Biomedical Optics*, 8:484, 2003.

[24] J. Han and M. Kamber. *Data Mining: Concepts and Techniques.* Morgan Kaufmann, 2000.

[25] M.P.L. Bard, A. Amelink, M. Skurichina, M. den Bakker, S.A. Burgers, J.P. van Meerbeeck, RP Duin, J. Aerts, H.C. Hoogsteden, and H. Sterenborg. Improving the specificity of fluorescence bronchoscopy for the analysis of neoplastic lesions of the bronchial tree by combination with optical spectroscopy: preliminary communication. *Lung Cancer*, 47(1):41–47, 2005.

[26] D.C.G. de Veld, M. Skurichina, M.J.H. Witjes, R.P.W. Duin, H.J.C.M. Sterenborg, and J.L.N. Roodenburg. Clinical study for classification of benign, dysplastic, and malignant oral lesions using autofluorescence spectroscopy. *Journal of Biomedical Optics*, 9(5):940–950, 2004.

[27] Y.S. Fawzy, M. Petek, M. Tercelj, and H. Zeng. In vivo assessment and evaluation of lung tissue morphologic and physiological changes from non-contact endoscopic reflectance spectroscopy for improving lung cancer detection. *Journal of Biomedical Optics*, 11:044003, 2006.

[28] V. Ntziachristos, AG Yodh, M.D. Schnall, and B. Chance. Mri-guided diffuse optical spectroscopy of malignant and benign breast lesions. *Neoplasia*, 4(4):347–354, 2002.

[29] N.M. Marín, N. MacKinnon, C. MacAulay, S.K. Chang, E.N. Atkinson, D. Cox, D. Serachitopol, B. Pikkula, M. Follen, and R. Richards-Kortum. Calibration standards for multicenter clinical trials of fluorescence spectroscopy for in vivo diagnosis. *Journal of Biomedical Optics*, 11:014010, 2006.

[30] Yvette N. Mirabal, Sung K. Chang, Edward Neely Atkinson, Anais Malpica, Michele Follen, and Rebecca Richards-Kortum. Reflectance spectroscopy for in vivo detection of cervical precancer. *Journal of Biomedical Optics*, 7(4):587–594, 2002.

[31] K.S. Johnson, D.W. Chicken, D.C.O. Pickard, A.C. Lee, G. Briggs, M. Falzon, I.J. Bigio, M.R. Keshtgar, and S.G. Bown. Elastic scattering spectroscopy for intraoperative determination of sentinel lymph node status in the breast. *Journal of Biomedical Optics*, 9:1122, 2004.

[32] S.D. Kamath and KK Mahato. Optical pathology using oral tissue fluorescence spectra: classification by principal component analysis and k-means nearest neighbor analysis. *Journal of Biomedical Optics*, 12:014028, 2007.

[33] N. Ramanujam, M.F. Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. Wright, N. Atkinson, and R.R. Kortum. Spectroscopic diagnosis of cervical intraepithelial neoplasia(cin) in vivo using laser-induced fluorescence spectra at multiple excitation wavlengths. *Lasers in Surgery and Medicine*, 19(1):63–74, 1996.

[34] JR Mourant, IJ Bigio, J. Boyer, RL Conn, T. Johnson, and T. Shimada. Spectroscopic diagnosis of bladder cancer with elastic light scattering. *Lasers Surg Med*, 17(4):350–7, 1995.

[35] G.M. Palmer and N. Ramanujam. Monte carlo-based inverse model for

calculating tissue optical properties. part i: Theory and validation on synthetic phantoms. *Applied Optics*, 45(5):1062–1071, 2006.

[36] R.Y. Rubinstein. *Simulation and the Monte Carlo Method*. John Wiley & Sons, Inc. New York, NY, USA, 1981.

[37] BC Wilson and G. Adam. A monte carlo model for the absorption and flux distributions of light in tissue. *Medical Physics*, 10:824, 1983.

[38] A.J. Welch, M.J. Gemert, et al. *Optical-Thermal Response of Laser-Irradiated Tissue*. Springer, 1995.

[39] T.J. Pfefer, L.S. Matchette, C.L. Bennett, J.A. Gall, J.N. Wilke, A.J. Durkin, and M.N. Ediger. Reflectance-based determination of optical properties in highly attenuating tissue. *Journal of Biomedical Optics*, 8:206, 2003.

[40] A. Ungut, G. Grehan, and G. Gouesbet. Comparisons between geometrical optics and lorenz-mie theory. *Applied Optics*, 20(17):2911–2918, 1981.

[41] HC van de Hulst. *Light Scattering by Small Particles*. 1957.

[42] C.F. Bohren and D.R. Huffman. *Absorption and scattering of light by small particles*. Wiley New York, 1983.

[43] Linda T. Nieman, Chih-Wen Kan, Ann Gillenwater, Mia K. Markey, and Konstantin Sokolov. Probing local tissue changes in the oral cav-

ity for early detection of cancer using oblique polarized reflectance spectroscopy: a pilot clinical trial. *Journal of Biomedical Optics*, 13(2):024011, 2008.

[44] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(7-8):1157–1182, 2003.

[45] N.M. Marín, A. Milbourne, H. Rhodes, T. Ehlen, D. Miller, L. Benedet, R. Richards-Kortum, and M. Follen. Diffuse reflectance patterns in cervical spectroscopy. *Gynecologic Oncology*, 99(3 Supplement 1):S116–S120, 2005.

[46] H.P. Chan, B. Sahiner, K.L. Lam, N. Petrick, M.A. Helvie, M.M. Goodsitt, and D.D. Adler. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. *Medical Physics*, 25:2007, 1998.

[47] L. Breiman. *Classification and Regerssion Trees*. Chapman & Hall/CRC, 1998.

[48] EN Atkinson, MF Mitchell, N. Ramanujam, and R. Richards-Kortum. Statistical techniques for diagnosing cin using fluorescence spectroscopy: Svd and cart. *J Cell Biochem Suppl*, 23:125–30, 1995.

[49] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience, 2000.

[50] Z. Ge, KT Schomacker, and NS Nishioka. Identification of colonic dysplasia and neoplasia by diffuse reflectance spectroscopy and pattern recognition techniques. *Applied Spectroscopy*, 52(6):833–839, 1998.

[51] DA Cirovic. Feed-forward artificial neural networks: applications to spectroscopy. *TrAC Trends in Analytical Chemistry*, 16(3):148–155, 1997.

[52] GS Nayak, S. Kamath, K.M. Pai, A. Sarakar, S. Ray, J. Kurien, L. D'almeida, BR Krishnand, C. Santhosh, VB Kartha, and KK Mahato. Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra: Classification of normal premalignant and malignant pathological conditions. *Biopolymers*, 82(2):152–166, 2006.

[53] D.H. Wolpert and W.G. Macready. No free lunch theorems for search. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997.

[54] JA Hanley and BJ McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143:29–36, 1982.

[55] A.C. Patel and M.K. Markey. Comparison of three-class classification performance metrics: a case study in breast cancer cad. *Medical imaging*, pages 581–589, 2005.

[56] Mehul P. Sampat, Amit C. Patel, Yuhling Wang, Shalini Gupta, Alan C. Bovik Chih-Wen Kan, and Mia K. Markey. Indices for multi-class

classification performance assessment - an empirical comparison. *IEEE Transactions on Information Technology in Biomedicine (under review)*, 2008.

[57] D.J. Hand and R.J. Till. A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine Learning*, 45(2):171–186, 2001.

[58] D. Mossman. Three-way rocs. *Medical Decision Making*, 19(1):78, 1999.

[59] *Hypervolume under the ROC hypersurface of a near-guessing ideal observer in a three-class classification task*, volume 5372, 2004.

[60] T. Mitchell, B. Buchanan, G. DeJong, T. Dietterich, P. Rosenbloom, and A. Waibel. Machine learning. *Annual Review of Computer Science*, 4(1):417–433, 1990.

[61] *A study of cross-validation and bootstrap for accuracy estimation and model selection*, volume 2. Morgan Kaufmann, 1995.

[62] B. Efron and G. Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983.

[63] B. Efron and R. Tibshirani. Statistical data analysis in the computer age. *Science*, 253(5018):390, 1991.

[64] B. Efron and R.J. Tibshirani. *An introduction to the bootstrap*. Chapman & Hall New York, 1993.

[65] M. Bland. *An introduction to medical statistics*. Oxford University Press New York, 2000.

[66] K. Onizawa, H. Saginoya, Y. Furuya, H. Yoshida, and H. Fukuda. Usefulness of fluorescence photography for diagnosis of oral cancer. *International journal of oral and maxillofacial surgery*, 28(3):206–210, 1999.

[67] B. Kulapaditharom and V. Boonkitticharoen. Laser-induced fluorescence imaging in localization of head and neck cancers. *The Annals of otology, rhinology & laryngology*, 107(3):241–246, 1998.

[68] E. Svistun, R. Alizadeh-Naderi, A. El-Naggar, R. Jacob, A. Gillenwater, and R. Richards-Kortum. Vision enhancement system for detection of oral cavity neoplasia based on autofluorescence. *Head & Neck*, 26(3):205–215, 2004.

[69] CS Betz, M. Mehlmann, K. Rick, H. Stepp, G. Grevers, R. Baumgartner, and A. Leunig. Autofluorescence imaging and spectroscopy of normal and malignant mucosa in patients with head and neck cancer. *Lasers in surgery and medicine*, 25(4):323–334, 1999.

[70] A. L. Clark, A. Gillenwater, R. Alizadeh-Naderi, A. K. El-Naggar, and R. Richards-Kortum. Detection and diagnosis of oral neoplasia with an

optical coherence microscope. *Journal of Biomedical Optics*, 9(6):1271–1280, 2004.

[71] Petra Wilder-Smith, Tatiana Krasieva, Woong-Gyu Jung, Jun Zhang, Zhongping Chen, Katherine Osann, and Bruce Tromberg. Noninvasive imaging of oral premalignancy and malignancy. *Journal of Biomedical Optics*, 10(5):051601, 2005.

[72] P.M. Lane, T. Gilhuly, P. Whitehead, H. Zeng, C.F. Poh, S. Ng, P.M. Williams, L. Zhang, M.P. Rosin, and C.E. MacAulay. Simple device for the direct visualization of oral-cavity tissue fluorescence. *Journal of Biomedical Optics*, 11:024006, 2006.

[73] Ann Gillenwater, Rhonda Jacob, Ravi Ganeshappa, Bonnie Kemp, Adel K. El-Naggar, J. Lynn Palmer, Gary Clayman, Michele Follen Mitchell, and Rebecca Richards-Kortum. Noninvasive diagnosis of oral neoplasia based on fluorescence spectroscopy and native tissue autofluorescence. *Arch Otolaryngol Head Neck Surg*, 124(11):1251–1258, 1998.

[74] S.P. Schantz, V. Kolli, H.E. Savage, G. Yu, J.P. Shah, D.E. Harris, A. Katz, R.R. Alfano, and A.G. Huvos. In vivo native cellular fluorescence and histological characteristics of head and neck cancer. *Clinical Cancer Research*, 4(5):1177, 1998.

[75] J.K. Dhingra, D.F. Perrault Jr, K. McMillan, E.E. Rebeiz, S. Kabani, R. Manoharan, I. Itzkan, M.S. Feld, and S.M. Shapshay. Early diagnosis

of upper aerodigestive tract cancer by autofluorescence. *Archives of Otolaryngology- Head and Neck Surgery*, 122(11):1181, 1996.

[76] D.L. Heintzelman, U. Utzinger, H. Fuchs, A. Zuluaga, K. Gossage, A.M. Gillenwater, R. Jacob, B. Kemp, and R.R. Richards-Kortum. Optimal excitation wavelengths for in vivo detection of oral neoplasia using fluorescence spectroscopy. *Photochemistry and Photobiology*, 72(1):103–113, 2000.

[77] A. Gillenwater, V. Papadimitrakopoulou, and R. Richards-Kortum. Oral premalignancy: new methods of detection and treatment. *Current oncology reports*, 8(2):146–154, 2006.

[78] DCG De Veld, MJH Witjes, H. Sterenborg, and JLN Roodenburg. The status of in vivo autofluorescence spectroscopy and imaging for oral oncology. *Oral oncology*, 41(2):117–131, 2005.

[79] SG Demos and RR Alfano. Temporal gating in highly scattering media by the degree of optical polarization. *Optics letters*, 21(2):161–163, 1996.

[80] S.L. Jacques, J.R. Roman, and K. Lee. Imaging superficial tissues with polarized light. *Lasers in Surgery and Medicine*, 26(2):119–129, 2000.

[81] R. Anderson. Polarized light examination and photography of the skin. *Archives of dermatology*, 127(7):1000, 1991.

[82] V. Backman, R. Gurjar, K. Badizadegan, I. Itzkan, R.R. Dasari, L.T. Perelman, and M.S. Feld. Polarized light scattering spectroscopy for quantitative measurement of epithelial cellular structures in situ. *IEEE Journal of Selected Topics in Quantum Electronics*, 5(4):1019, 1999.

[83] K. Sokolov, R. Drezek, K. Gossage, and R. Richards-Kortum. Reflectance spectroscopy with polarized light: is it sensitive to cellular and nuclear morphology. *Optics Express*, 5(13):302–317, 1999.

[84] L. Nieman, A. Myakov, J. Aaron, and K. Sokolov. Optical sectioning using a fiber probe with an angled illumination-collection geometry: Evaluation in engineered tissue phantoms. *Applied Optics*, 43(6):1308–1319, 2004.

[85] K. Sokolov, LT Nieman, A. Myakov, and A. Gillenwater. Polarized reflectance spectroscopy for pre-cancer detection. *Technol Cancer Res Treat*, 3(1):1–14, 2004.

[86] A. Myakov, L. Nieman, L. Wicky, U. Utzinger, R. Richards-Kortum, and K. Sokolov. Fiber optic probe for polarized reflectance spectroscopy in vivo: Design and performance. *Journal of Biomedical Optics*, 7:388, 2002.

[87] N. Ancona, R. Maglietta, A. Piepoli, A. D'Addabbo, R. Cotugno, M. Savino, S. Liuni, M. Carella, G. Pesole, and F. Perri. On the statistical assessment of classifiers using dna microarray data. *BMC Bioinformatics*, 7(1):387, 2006.

[88] R. Drezek, M. Guillaud, T. Collier, I. Boiko, A. Malpica, C. Macaulay, M. Follen, and R.R. Richards-Kortum. Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, dna content, and chromatin texture. *Journal of biomedical optics*, 8:7, 2003.

[89] M.P. Siegel, Y.L. Kim, H.K. Roy, R.K. Wali, and V. Backman. Assessment of blood supply in superficial tissue by polarization-gated elastic light-scattering spectroscopy. *Applied optics*, 45(2):335–342, 2006.

[90] JJ Sciubba. Oral leukoplakia. *Critical reviews in oral biology and medicine: an official publication of the American Association of Oral Biologists*, 6(2):147, 1995.

[91] S. Silverman Jr, M. Gorsky, F. Lozada, et al. Oral leukoplakia and malignant transformation. a follow-up study of 257 patients. *Cancer*, 53(3):563–568, 1984.

[92] J. Reibel. Prognosisof oral pre-malignant lesions: Significanceof clinical, histopathological, and molecular biological characteristics. *Critical Reviews in Oral Biology & Medicine*, 14(1):47, 2003.

[93] J.J. Sciubba. Oral cancer and its detectionhistory-taking and the diagnostic phase of management. *The Journal of the American Dental Association*, 132(suppl_ 1):12–18, 2001.

[94] P. Holmstrup, P. Vedtofte, J. Reibel, and K. Stoltze. Long-term treatment outcome of oral premalignant lesions. *Oral oncology*, 42(5):461–474, 2006.

[95] L. Zhang, M. Williams, C.F. Poh, D. Laronde, J.B. Epstein, S. Durham, H. Nakamura, K. Berean, A. Hovan, N.D. Le, et al. Toluidine blue staining identifies high-risk primary oral premalignant lesions with poor outcome. *Cancer research*, 65(17):8017, 2005.

[96] S.J.R. Silverman and P.B. Sugerman. Oral premalignancies and squamous cell carcinoma. *Clinics in dermatology*, 18(5):563–568, 2000.

[97] A. Amelink, H.J.C.M. Sterenborg, M.P.L. Bard, and S.A. Burgers. In vivo measurement of the local optical properties of tissue by use of differential path-length spectroscopy. *Optics Letters*, 29(10):1087–1089, 2004.

[98] R.L.P. Veen, A. Amelink, M. Menke-Pluymers, C. Pol, and H.J.C.M. Sterenborg. Optical biopsy of breast tissue using differential path-length spectroscopy. *Physics in Medicine and Biology*, 50:2573, 2005.

[99] I. Georgakoudi, E.E. Sheets, MG Muller, V. Backman, C.P. Crum, K. Badizadegan, R.R. Dasari, and M.S. Feld. Trimodal spectroscopy for the detection and characterization of cervical precancers in vivo. *American journal of obstetrics and gynecology*, 186(3):374–382, 2002.

[100] C.Y. Wang, T. Tsai, H.M. Chen, C.T. Chen, and C.P. Chiang. Pls-ann based classification model for oral submucous fibrosis and oral carcinogenesis. *Lasers in surgery and medicine*, 32(4):318–326, 2003.

[101] B. Kulapaditharom and V. Boonkitticharoen. Performance characteristics of fluorescence endoscope in detection of head and neck cancers. *The Annals of otology, rhinology & laryngology*, 110(1):45–52, 2001.

[102] D.C.G. de Veld, M. Skurichina, M.J.H. Witjes, R.P.W. Duin, H.J.C.M. Sterenborg, and J.L.N. Roodenburg. Autofluorescence and diffuse reflectance spectroscopy for oral oncology. *Lasers in Surgery and Medicine*, 36(5):356–364, 2005.

[103] I.S. Seo, J.S. You, C.K. Hayakawa, and V. Venugopalan. Perturbation and differential monte carlo methods for measurement of optical properties in a layered epithelial tissue model. *Journal of Biomedical Optics*, 12:014030, 2007.

[104] Xueding Wang and Lihong V. Wang. Propagation of polarized light in birefringent turbid media: A monte carlo study. *Journal of Biomedical Optics*, 7(3):279–290, 2002.

[105] Xueding Wang, Lihong V. Wang, Chia-Wei Sun, and Chih-Chung Yang. Polarized light propagation through scattering media: time-resolved monte carlo simulations and experiments. *Journal of Biomedical Optics*, 8(4):608–617, 2003.

[106] M.F.G. Wood, X. Guo, and I.A. Vitkin. Polarized light propagation in multiply scattering media exhibiting both linear birefringence and optical activity: Monte carlo model and experimental methodology. *Journal of Biomedical Optics*, 12:014029, 2007.

[107] Changfang Zhu, Gregory M. Palmer, Tara M. Breslin, Josephine Harter, and Nirmala Ramanujam. Diagnosis of breast cancer using fluorescence and diffuse reflectance spectroscopy: a monte-carlo-model-based approach. *Journal of Biomedical Optics*, 13(3):034015, 2008.

[108] T. Johnson and J. Mourant. Polarized wavelength-dependent measurements of turbid media. *Optics Express*, 4(6):200–216, 1999.

[109] P. Taroni, D. Comelli, A. Pifferi, A. Torricelli, and R. Cubeddu. Absorption of collagen: effects on the estimate of breast composition and related diagnostic implications. *Journal of Biomedical Optics*, 12:014021, 2007.

[110] T. Raykov and G.A. Marcoulides. *An introduction to applied multivariate analysis*. Routledge/Psychpress, 2008.

[111] M.R. Antonelli, A. Pierangelo, T. Novikova, P. Validire, A. Benali, B. Gayet, and A. De Martino. Mueller matrix imaging of human colon tissue for cancer diagnostics: how monte carlo modeling can help in the interpretation of experimental data. *Optics Express*, 18(10):10200–10208, 2010.

[112] Xueding Wang, Lihong V. Wang, Chia-Wei Sun, and Chih-Chung Yang. Polarized light propagation through scattering media: time-resolved monte carlo simulations and experiments. *Journal of Biomedical Optics*, 8(4):608–617, 2003.

[113] J.R. Mourant, A.H. Hielscher, A.A. Eick, T.M. Johnson, and J.P. Freyer. Evidence of intrinsic differences in the light scattering properties of tumorigenic and nontumorigenic cells. *Cancer*, 84(6):366–374, 1998.

[114] A.H. Hielscher, J.R. Mourant, and I.J. Bigio. Influence of particle size and concentration on the diffuse backscattering of polarized light from tissue phantoms and biological cell suspensions. *Applied optics*, 36(1):125–135, 1997.

[115] X. Wang, G. Yao, and L.V. Wang. Monte carlo model and single-scattering approximation of the propagation of polarized light in turbid media containing glucose. *Applied optics*, 41(4):792–801, 2002.

[116] S.S. Chandrasekhar. *Radiative Transfer*. Courier Dover Publications, 1960.

[117] AJ Welch, G. Yoon, and MJ van Gemert. Practical models for light distribution in laser-irradiated tissue. *Lasers Surg Med*, 6(6):488–93, 1987.

[118] M.S. Patterson, B.C. Wilson, and D.R. Wyman. The propagation of optical radiation in tissue i. models of radiation transport and their

application. *Lasers in Medical Science*, 6(2):155–168, 1991.

[119] L. Wang, S.L. Jacques, and L. Zheng. Mcmlmonte carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedicine*, 47(2):131–146, 1995.

[120] G. Yao and L. Wang. Propagation of polarized light in turbid media: simulated animation sequences. *Optics Express*, 7(5):198–203, 2000.

[121] Yang Liu, Young Kim, Xu Li, and Vadim Backman. Investigation of depth selectivity of polarization gating for tissue characterization. *Opt. Express*, 13(2):601–611, 2005.

[122] D. Côté and I. Vitkin. Robust concentration determination of optically active molecules in turbid media with validated three-dimensional polarization sensitive monte carlo calculations. *Optics Express*, 13(1):148–163, 2005.

[123] Franck Jaillon and Hervé Saint-Jalmes. Description and time reduction of a monte carlo code to simulate propagation of polarized light through scattering media. *Appl. Opt.*, 42(16):3290–3296, 2003.

[124] Sebastian Bartel and Andreas H. Hielscher. Monte carlo simulations of the diffuse backscattering mueller matrix for highly scattering media. *Appl. Opt.*, 39(10):1580–1588, 2000.

[125] B. Kaplan, G. Ledanois, and B. Villon. Mueller matrix of dense polystyrene latex sphere suspensions: Measurements and monte carlo simulation. *Appl. Opt*, 40(16):2769–2777, 2001.

[126] X. Guo, M.F.G. Wood, and A. Vitkin. Monte carlo study of path-length distribution of polarized light in turbid media. *Optics Express*, 15(3):1348–1360, 2007.

[127] L.T. Nieman, M. Jakovljevic, and K. Sokolov. Compact beveled fiber optic probe design for enhanced depth discrimination in epithelial tissues. *Opt. Express*, 17:2780–2796, 2009.

[128] TA Germer. Scatmech: polarized light scattering c++ class library.

Vita

Chih-wen Kan was born in Taichung, Taiwan August 31, 1982. She received Bachelor of Science and Master of Science degree in Electrical Engineering from National Tsing Hua University in 2004 and 2005, respectively. She joined the Department of Biomedical Engineering in The University of Texas at Austin in 2005.

Permanent address: 88 Donghai Rd., Taichung, Taiwan

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.