

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

**Chih-wen Kan**

**Andy Y. Lee**

The University of Texas  
Department of Biomedical Engineering  
1 University Station  
Austin, Texas 78712-1084

**Linda T. Nieman**

**Konstantin Sokolov**

The University of Texas  
Department of Biomedical Engineering  
1 University Station  
Austin, Texas 78712-1084  
and

The University of Texas  
M. D. Anderson Cancer Center  
Department of Imaging Physics  
1515 Holcombe Boulevard  
Houston, Texas 77030

**Mia K. Markey**

The University of Texas  
Department of Biomedical Engineering  
1 University Station  
Austin, Texas 78712-1084

**Abstract.** We present an approach to adaptively adjust the spectral window sizes for optical spectra feature extraction. Previous studies extracted features from spectral windows of a fixed width. In our algorithm, piecewise linear regression is used to adaptively adjust the window sizes to find the maximum window size with reasonable linear fit with the spectrum. 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 data set of diffuse reflectance spectra of oral mucosa lesions. Eight features were extracted from each window. We performed classifications using linear discriminant analysis with cross-validation. Using windowing techniques results in better classification performance than not using windowing. The area under the receiver-operating-characteristics curve for windowing techniques was greater than a nonwindowing technique for both normal versus mild dysplasia (MD) plus severe high-grade dysplasia or carcinoma (SD) (MD+SD) and benign versus MD+SD. Although adaptive and fixed-size windowing perform similarly, adaptive windowing utilizes significantly fewer windows than fixed-size windows (number of windows per spectrum: 8 versus 16). Because 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. © 2010 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3481143]

**Keywords:** spectroscopy signal processing; adaptive spectral window size; feature extraction; data compression; optical spectroscopy; computer-aided diagnosis; ROC analysis.

Paper 10203PR received Apr. 18, 2010; revised manuscript received Jun. 26, 2010; accepted for publication Jul. 1, 2010; published online Aug. 20, 2010.

## 1 Introduction

Cancer is a major health problem throughout the world.<sup>1</sup> Early diagnosis of precancerous neoplasia has been shown to reduce mortality dramatically. Thus, there is a pressing need for accurate and low-cost screening and diagnostic techniques to identify curable precancerous sites. Epithelial precancers are characterized by a variety of architectural and morphological features, including increased nuclear size, increased nuclear-cytoplasmic ratio, hyperchromasia, and pleomorphism. Currently, the architectural and morphological changes related to carcinogenesis are assessed by biopsy, which is invasive. Moreover, biopsy is not able to monitor real-time dynamic changes associated with the disease, which are important for assessing treatment response. In contrast, several studies have shown that optical technologies hold great promise for noninvasive and real-time assessment of precancers.<sup>2-14</sup> This study concerns diffuse reflectance spectroscopy for epithelial cancer diagnosis. Reflectance spectroscopy captures optical scatter-

ing and absorption of epithelial tissue, which are altered when precancerous abnormalities exist. Although optical spectroscopy has been demonstrated to capture properties that are, on average, different between normal and cancerous tissues, variation in the spectral response between patients makes accurate diagnoses difficult to achieve. Therefore, clinicians will need decision support tools to help them predict cancer status from optical spectroscopy with acceptable sensitivity and specificity.

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.<sup>15</sup> 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

---

Address all correspondence to: Mia K. Markey, The University of Texas Department of Biomedical Engineering, 1 University Station, Austin, Texas 78712-1084. Tel: 512-471-1711; Fax: 512-471-0616. E-mail: mia.markey@mail.utexas.edu

to prevent overtraining. PCA is a linear transformation technique for reducing the dimensionality of data. Spectral feature extraction describes spectral signatures without prior knowledge of the tissue's physical nature, while model-based extraction requires prior knowledge of the tissue's properties. Most previous studies processed spectroscopy signals by employing spectral intensity signals directly as extracted features, and performed PCA for feature reduction (e.g., Refs. 5, 10, and 16–20), but some use model-based features (e.g., Refs. 13 and 21–28). The second category of feature extraction methods is model-based feature extraction. Model-based methods analyze the wavelength or angle-dependence optical signals to determine physical properties of cells and tissues.<sup>12,14,29</sup> These properties are generally believed to relate to disease status; hence, they can be indications of disease. The current trend leans more toward model-based feature extraction and hybrid feature extraction. Hybrid feature extraction, the fourth category of feature extraction methods, is a mixture of statistical features and model-based features.<sup>29,30</sup> These approaches utilize model-based preprocessing of spectra, based on underlying biology to extract input parameters, and analyze these parameters with statistical methods.

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 maximum 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<sup>7,31,32</sup> 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 a spectral region for feature extraction makes a difference in the performance of the extracted features.<sup>7,31,32</sup>

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.<sup>5</sup> 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.<sup>5</sup> Mourant et al.'s study on spectroscopic diagnosis of bladder cancer with elastic light scattering<sup>31</sup> 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

**Table 1** Eight spectral features were employed in this study. These local features were extracted in each spectral window.

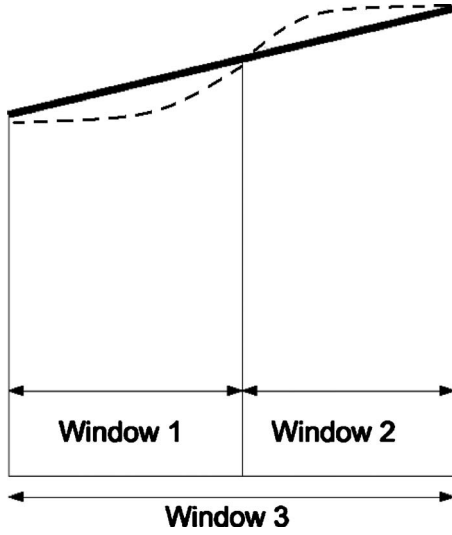
Feature		
No.	Name	Proposed by
1	Slope from linear regression	Bigio et al., <sup>5</sup> Mourant et al. <sup>31</sup>
2	Intercept from linear regression	Mueller et al. <sup>12</sup>
3	Minimum intensity	New
4	Average intensity	Bigio et al., <sup>5</sup> Mourant et al. <sup>31</sup>
5	Median intensity	New
6	Maximum intensity	New
7	Standard deviation of the intensities	Kamath and Mahato <sup>19</sup>
8	Signal energy of the intensities	Kamath and Mahato <sup>19</sup>

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 feature 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-size windowing; however, adaptive windows require significantly less windows than fixed-size 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 Sec. 2.

## 2 Theoretical Derivation

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

Table 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



**Fig. 1** Illustration of an adaptive window and two fixed windows. The dashed line is the real spectrum, and the straight line is the fitted line. Windows 1 and 2 are fixed-sized windows while window 3 is the adaptive window.

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 Fig. 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

$$\sigma_1 = \sqrt{\frac{\sum_{i=1}^{N/2} [I(i) - \bar{I}_1]^2}{N/2}}, \quad (1)$$

$$\sigma_2 = \sqrt{\frac{\sum_{i=N/2+1}^N [I(i) - \bar{I}_2]^2}{N/2}}, \quad (2)$$

where  $I(i)$  is the intensity at the  $i$ 'th wavelength point and  $N$  is the number of points in the window. The mean intensity is given by

$$\bar{I}_1 = \frac{\sum_{i=1}^{N/2} I(i)}{N/2}, \quad \text{and} \quad \bar{I}_2 = \frac{\sum_{i=N/2+1}^N I(i)}{N/2}. \quad (3)$$

In window 3, the standard deviation can be written as

$$\sigma_3 = \sqrt{\frac{\sum_{i=1}^N [I(i) - \bar{I}_3]^2}{N}}, \quad (4)$$

$$\sigma_3^2 = \frac{1}{N} \left[ \sum_{i=1}^{N/2} [I(i) - \bar{I}_3]^2 + \sum_{i=N/2+1}^N [I(i) - \bar{I}_3]^2 \right], \quad (5)$$

$$\sigma_3^2 = \frac{1}{N} \left[ \sum_{i=1}^{N/2} [I^2(i) - 2I(i)\bar{I}_3 + \bar{I}_3^2] + \sum_{i=N/2+1}^N [I^2(i) - 2I(i)\bar{I}_3 + \bar{I}_3^2] \right], \quad (6)$$

$$\sigma_3^2 = \frac{1}{N} \left[ \sum_{i=1}^{N/2} I^2(i) + \sum_{i=N/2+1}^N I^2(i) \right] - 2\bar{I}_3\bar{I}_3 + \bar{I}_3^2, \quad (7)$$

$$\sigma_3^2 = \frac{1}{N} \left[ \sum_{i=1}^{N/2} I^2(i) + \sum_{i=N/2+1}^N I^2(i) \right] - \bar{I}_3^2. \quad (8)$$

We rewrite Eq. (1) as

$$\sigma_1^2 = \frac{1}{N/2} \left[ \sum_{i=1}^{N/2} I^2(i) - 2\bar{I}_1 \sum_{i=1}^{N/2} I(i) + \frac{N}{2} \bar{I}_1^2 \right], \quad (9)$$

$$\sigma_1^2 = \frac{\sum_{i=1}^{N/2} I^2(i)}{N/2} - \bar{I}_1^2. \quad (10)$$

Thus, equating Eq. (10) ,

$$\frac{\sum_{i=1}^{N/2} I^2(i)}{N} = \frac{\sigma_1^2 + \bar{I}_1^2}{2}. \quad (11)$$

Substituting Eq. (8) with Eq. (11), we get

$$\sigma_3^2 = \frac{\sigma_1^2 + \bar{I}_1^2}{2} + \frac{\sigma_2^2 + \bar{I}_2^2}{2} - \bar{I}_3^2. \quad (12)$$

Because  $\sigma_1 = \sigma_2$ , we get

$$\sigma_3^2 = \sigma_1^2 + \frac{1}{4}(\bar{I}_1 + \bar{I}_2)^2. \quad (13)$$

Therefore, from Eq. (12), we can conclude that feature 7 has a linear relationship between adaptive windows and fixed-size windows.

Feature 8, signal energy, is defined as

$$SE = \sum_{i=1}^N I^2(i), \quad (14)$$

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  $i$ 'th point in the spectral window.

This can be rewritten as

$$\sum_{i=1}^N I^2(i) = \sum_{i=1}^{N/2} I^2(i) + \sum_{i=N/2+1}^N I^2(i), \quad (15)$$

$$SE_3 = SE_1 + SE_2. \quad (16)$$

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 Materials and Methods

#### 3.1 Overview

We compare three windowing techniques: (i) No windowing is performed (i.e., features are extracted from all wavelengths 400–725 nm), (ii) fixed window size of 20 nm (window size adopted from Ref. 5), and (iii) adaptive spectral window sizes determined by the novel algorithm presented in this paper.

#### 3.2 Data Collection

This analysis used a diffuse reflectance spectroscopy data set obtained in a previous investigation of oblique polarization reflectance spectroscopy (OPRS) of oral mucosa lesions.<sup>30</sup> In this study, we used only the diffuse reflectance subset of the OPRS data set because 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.

The data set was acquired at The University of Texas M. D. Anderson Cancer Center (UT MDACC) in Houston, Texas, as previously described.<sup>30</sup> Briefly, 27 patients over the age of 18 years were referred to the Head and Neck Clinic at UT MDACC with oral mucosa lesions suspicious for dysplasia or carcinoma.<sup>30</sup> Spectroscopic measurements were typically performed on one or two visually abnormal sites and one visually normal site. Biopsies were taken of all measured tissue sites. We measured a total of 57 sites, of which 22 were visually and histopathologically normal (Normal), 13 sites were visually abnormal but histopathologically normal (Benign), 12 were visually abnormal sites that proved to be mild dysplasia (MD) on histopathology, and 10 were visually abnormal sites that proved to be severe high-grade dysplasia or carcinoma (SD) on histopathology. The resulting data set consisted of 618 diffuse reflectance spectroscopy data points within the wavelength range of 400–725 nm for 57 measurements. The resolution of the spectra is one data point every 0.5 nm.

We used MATLAB R14 (The MathWorks, Natick, Massachusetts) and its statistics toolbox for data analysis in this study.

#### 3.3 Preprocessing

The spectra were normalized to remove interpatient variation. As suggested by our previous study,<sup>30</sup> 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.

#### 3.4 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.

#### 3.5 Feature Extraction

Within every spectral window, eight spectral features are extracted to describe the signal (Table 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,<sup>12,19</sup> breast,<sup>5</sup> and bladder,<sup>31</sup> with diffuse reflectance spectroscopy,<sup>5,12,19,31</sup> and fluorescence spectroscopy.<sup>12</sup>

Features 1, 2, 4, 7, and 8 are adopted from literature.<sup>5,12,19,31</sup> 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.

#### 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 to which of the two diagnostic categories each spectrum belongs. We consider only two of the possible diagnostic tasks in order to simplify the analyses, as follows:

1. Normal versus MD+SD
2. Benign versus MD+SD

Task 1 is very commonly studied in the literature because it is assumed to be, pathologically, the most distinct comparison.<sup>30</sup> Task 2 is arguably the most important task clinically, because 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 Sec. 4.1, (i) no windowing, (ii) fixed-size windowing, and (iii) 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 is used to combine the eight features. A leave-one-out cross-validation strategy is employed, and the area under the receiver-operating-characteristic (ROC) curve (AUC) 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 the adaptive and fixed-size windowing methods, the wavelength space is divided into fixed-size or adaptive size windows; thus, each window results in one AUC value. In comparison, in the no-windowing method, only one set of features is extracted; thus, there is only one AUC value for the entire wavelength space.

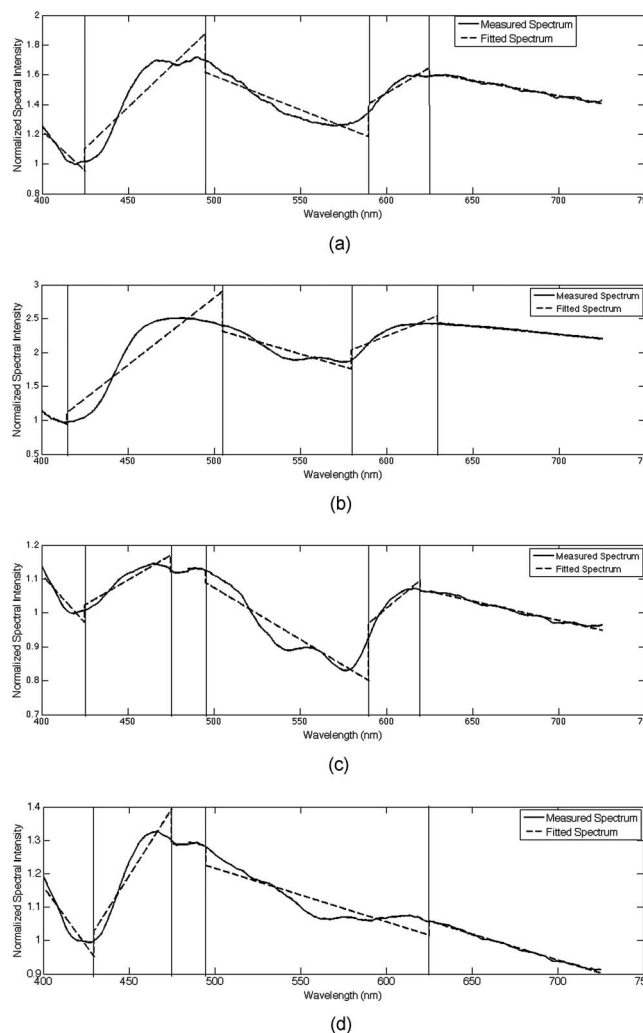
### 3.7 Classification and ROC Analysis

The classification task in this study is carried out using LDA, which is a linear classifier that searches for a hypersurface in the feature space that has an orientation to effectively discriminate between the two classes of data. The choice of this classifier is based on both its high computational efficiency and the previous success of LDA with this type of data set. In our previous study, we used LDA exhaustively to search for the most discriminatory features and identified that LDA can achieve high AUC on this task.<sup>30</sup> The purpose of this study is to compare the different windowing techniques in feature extraction; thus, the same classifier is used throughout this paper.

ROC analysis is used to quantify performance for two-class classification tasks.<sup>33</sup> Sensitivity and specificity indicate the ability of the diagnostic method to distinguish between two groups (e.g., healthy and diseased). By varying the threshold, an ROC curve of sensitivity versus (1-specificity) is generated. However, the AUC is often used as a metric to quantitatively summarize the performance of a clinical decision support system. Therefore, we use AUC as the metric to evaluate performance. The higher the AUC metric is, the better the performance is. An AUC value of 1 represents perfect discrimination, while an AUC value of 0.5 represents performance expected by chance alone.

### 3.8 Denoising

Noisy data can interfere with classifier training. We used two procedures for denoising: (i) Thresholding on classifier outputs and (ii) outlier removal. These two procedures are described as follows.



**Fig. 2** Sample diffuse reflectance spectra for (a) normal (b) benign (c) MD, and (d) SD patients. The solid curve indicates the measured diffuse reflectance spectrum, while the dashed curve indicates the fitted spectrum based on our piecewise linear regression model.

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 being 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, then we consider the predictive LDA model to be too noisy to make a proper prediction and we remove this LDA and its corresponding AUC from 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 is based on the Mahalanobis distance (MDist) measure.<sup>34</sup> The MDist is a multivariate measure (in square units) of the separation of an unknown data set from a known

set (with mean  $\mu$  and covariance matrix  $S$ ) in space. The MDist of a data set, 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  $\chi^2$  distribution with the degrees of freedom being the number of the variables.<sup>34</sup> We consider a spectrum to be an outlier if its MDist is larger than the critical point (significance level  $\alpha = 0.001$ ) of the  $\chi^2$  distribution with the degrees of freedom being the number of variables participating in the MDist.

The denoising of the data is different from the preprocessing step described in Sec. 4.3. The denoising 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.

### 3.9 Statistical Comparison of AUC Values

A bootstrapping technique was used to estimate the significance of the observed difference in the AUC between LDA models.<sup>35</sup>  $P$  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 versus 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.

### 3.10 Studying 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. Because 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) Large-range variations of initialization points: three initialization points were tested: 400, 562, and 725 nm. These points represent the smallest, middle, and largest wavelengths in our data. Because 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 the right. For 725 nm, we start from the right 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 the adaptive windowing technique.

## 4 Results

### 4.1 Piecewise Linear Regression Models

Sample results of our piecewise linear regression model are shown in Fig. 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, Fig. 2(a) has a window that starts at 425 nm and ends at 500 nm—a window size of 175 nm, which is about nine 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 large or small windows. The intensity features (features 4–7) may change for larger windows, but they are linearly proportional as shown in previous theoretical derivations (Sec. 2) 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 windows 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.

### 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 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. Because there are multiple windows involved in windowing techniques (fixed-size, adaptive windows), multiple AUCs are reported. In Table 2, the maximum and median AUCs of each windowing method are listed. The complete list of AUCs (not only maximum and median) are shown in Table 3. The classifiers based on features extracted

**Table 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 versus 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 versus 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 versus MD+SD			Benign versus MD+SD		
	Adaptive window	Fixed-size window (20 nm)	No windowing	Adaptive window	Fixed-size window (20 nm)	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	

from adaptive 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 versus MD+SD and benign versus 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 based on features extracted from the entire spectrum (no windowing, AUC=0.61) for normal versus 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 versus MD+SD and also for benign versus MD+SD with adaptive (max AUC 0.79 versus AUC 0.68,  $p=0.04$ ) and fixed windows (max AUC 0.83 versus 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 versus MD+SD and benign versus MD+SD, and these AUCs are significantly higher ( $p < 0.05$ ) in all of four comparisons.

Because 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 versus MD+SD and benign versus MD+SD.

On the other hand, the median AUCs for both normal versus MD+SD and benign versus 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 4). 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 versus MD+SD and benign versus 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 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 Sec. 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.

### 4.3 Effect of Different Initialization Points

As described in Sec. 3.10, there are three experiments conducted. First, the adaptive windowing technique was applied with different initialization points: 400, 405, 407, 410, 415, and 420 nm (Fig. 3). In Fig. 3, we selected one spectrum of category SD to visualize the window definitions of different initialization points. It is visually apparent that the adaptive window definitions are only slightly shifted from each other.

**Table 3** Supplementary data. A list of all the AUCs generated using LDA classifiers trained using all eight features from each spectral window. The maximum and median values of these AUCs are shown in Table 2. The number of AUCs is higher for adaptive windows because they have more unique (nonrepeated) features. Repeated features are commonly seen in fixed windows; they are removed here. The no-windowing technique takes the entire spectrum to do feature extraction and classification; therefore, only one AUC is calculated.

Normal versus MD+SD			Benign versus MD+SD		
Adaptive	Fixed	No windowing	Adaptive	Fixed	No windowing
0.42	0.55	0.61	0.68	0.69	0.68
0.68	0.50		0.65	0.78	
0.60	0.57		0.69	0.72	
0.60	0.55		0.65	0.80	
0.77	0.46		0.74	0.68	
0.60	0.73		0.74	0.69	
0.51	0.49		0.71	0.83	
0.64	0.44		0.64	0.68	
0.64	0.52		0.77	0.77	
0.59	0.49		0.77	0.74	
0.71	0.53		0.69	0.73	
0.73	0.49		0.68	0.72	
0.72	0.51		0.59	0.63	
0.74	0.57		0.50	0.67	
0.66	0.46		0.64	0.71	
0.72			0.69	0.73	
0.69			0.71	0.61	
0.72			0.61		
0.69			0.66		
0.74			0.69		
0.62			0.70		
0.72			0.74		
0.66			0.67		
0.74			0.71		
0.65			0.71		
0.60			0.70		
0.63			0.73		
0.66			0.70		
0.65			0.75		
0.49			0.68		

**Table 3** (Continued.)

Normal versus MD+SD			Benign versus MD+SD		
Adaptive	Fixed	No windowing	Adaptive	Fixed	No windowing
0.56			0.76		
0.52			0.70		
0.53			0.70		
0.56			0.74		
0.56			0.70		
0.59			0.79		
0.60			0.77		
0.60			0.72		
0.60			0.75		
0.60			0.75		
0.61			0.72		
0.59			0.75		
			0.64		
			0.64		
			0.67		
			0.70		
			0.67		
			0.67		
			0.67		
			0.70		
			0.70		
			0.70		
			0.70		
			0.64		
			0.62		
			0.67		
			0.67		



**Table 4** Numbers of unique features extracted and of windows per spectrum are presented. It shows that adaptive windows require fewer windows on average, but produce most unique features. The no-windowing technique has 44 and 35 “total number of windows” because there are 44 and 35 spectra in each case, and each spectrum has one window.

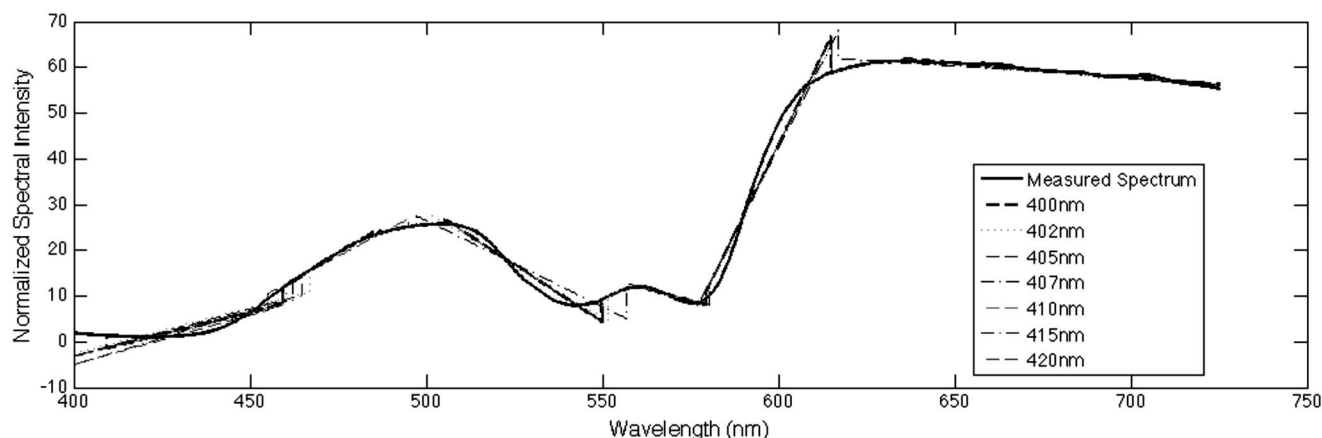
	Normal versus MD+SD			Benign versus MD+SD		
	Adaptive window	Fixed-size window (20 nm)	No windowing	Adaptive window	Fixed-size window (20 nm)	No windowing
Number of unique features extracted	60	17	1	60	14	1
Average number of windows per spectrum	8.5	16	1	7.71	16	1
Total number of windows	374	704	44	270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

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, because 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.

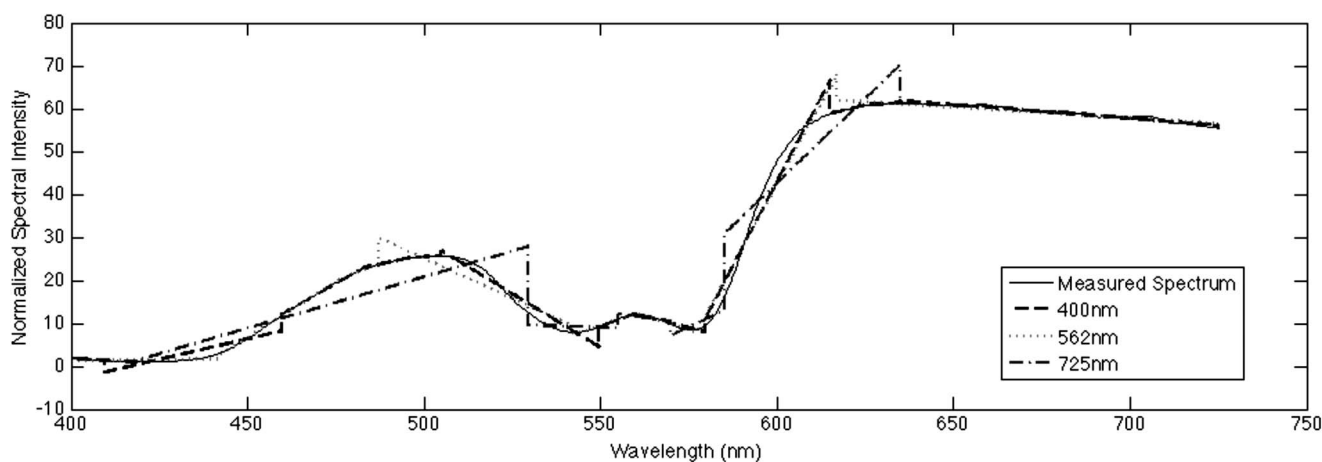
In the second experiment, we investigated a larger range of the initialization points. Because the ranges are larger, we modified the algorithm such that we do not discard the spectral data to the left of the initialization point (as we did in the previous experiment). The modified version of the algorithm “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 wavelength). The results are shown in Fig. 4. Similar to Fig. 3, in Fig. 4, 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; whereas 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 5) based on the choice of initialization point.



**Fig. 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 data set. The patient belongs to SD pathology group. The solid black curve is the measured spectrum, and the dashed straight lines are the windows. This example shows that the 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).



**Fig. 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 data set. The patient belongs to SD pathology group. 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) from the previous experiment (Figure 3); Windows “grow” to either the 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. This example shows that the window definitions are visually similar for 400 and 562 nm, but the ones with 725 nm are visually distinguishable.

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 5 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

**Table 5** AUCs calculated from different initialization points. It is shown that the AUCs of adaptive windows differ when the initialization point is varied, whereas the AUCs from fixed windows remain the same. Some of these AUC changes are statistically significant (0.82 compared to 0.65). This shows that adaptive windowing has a dependency on where the window definition starts, which gives the adaptive windows flexibility to achieve higher accuracy to predict the disease (AUC=0.82). However, this also shows that fixed windows have better robustness because adaptive windows depend on their initialization points.

	Maximum/median AUC for normal versus MD+SD	
	Adaptive windows	Fixed windows
400	0.73/0.62	0.73/0.51
405	0.78/0.66	0.73/0.51
407	0.82/0.66	0.73/0.51
410	0.73/0.58	0.73/0.51
415	0.65/0.56	0.73/0.51
420	0.70/0.62	0.73/0.51
562	0.70/0.59	0.73/0.51
725	0.78/0.56	0.73/0.51

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 and 725 nm may seem very different, their AUCs are not that different. The AUCs for the 725 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.

### 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 versus MD+SD and benign versus 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 data sets<sup>7,18</sup> that chose fixed-size windowing techniques. Bigio et al.<sup>5</sup> divided the measured spectrum into fixed-size wavelength bands of 20 nm from 330 to 750 nm, followed by feature extraction of average intensity within the wavelength band, and then PCA to reduce dimensionality for the input of artificial neural network. In addition, Johnson et al.<sup>18</sup> also used PCA to select only the wavelength regions with large variabil-

ity. These studies divide spectra into wavelength bands under the assumption that some wavelength 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-size windows and adaptive-size windows, we conclude that adaptive windowing more efficiently captures diagnostically relevant spectral features without a statistically significant decrease in classifier performance.

Our adaptive window technique defines spectral feature extraction regions using linear regression (Fig. 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 (Sec. 2) and experimentally (Sec. 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 (Fig. 1), where the signal in two fixed-size windows (windows 1 and 2) and one adaptive-size window (window 3) share the same regression line (solid line). In this case, we demonstrated that the adaptive-size 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 the features we chose have a linear relationship, proving that the adaptive-size window can completely represent the fixed-size windows. Adaptive windows can be viewed as linear “combinations” of smaller and potentially redundant fixed-size windows.

We further showed a reduction of redundancy in diagnostic information in Table 4. 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 data set. While the dimensionality of the feature space was reduced, there were no significant AUC differences observed between adaptive windows and fixed-size windows. Therefore, adaptive windowing is capable of teasing out unwanted redundancies that are inherent in fixed-size 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 shift. 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 points (562 and

725 nm) show 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 leads 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 data set. 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.

## 6 Conclusion

In recent years, there has been a debate on which diagnostic algorithm to use for bio-optical cancer-detection modalities.<sup>29</sup> Various analysis methods have been used, including model-based analysis, statistics-based analysis, and hybrid analyses. This paper focuses on providing a new aspect for statistics-based analysis. First, we verified the hidden assumption by Bigio et al.<sup>5</sup> and Johnson et al.<sup>18</sup> 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. Because 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.

### Acknowledgments

Financial support from the Whitaker Foundation and the National Institute of Biomedical Imaging and Bioengineering Grant No. EB003540 is gratefully acknowledged. The authors also thank Bryan Jiang and Nhi Pham for their help in developing the recursive algorithm for piecewise linear regression, and Arjun Ramachandran for his technical help.

### References

1. A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun, "Cancer statistics, 2009," *Ca-Cancer J. Clin.* **59**, 225–229 (2009).
2. K. Sokolov, M. Follen, and R. Richards-Kortum, "Optical spectroscopy for detection of neoplasia," *Curr. Opin. Chem. Biol.* **6**, 651–658 (2002).
3. R. Richards-Kortum and E. Sevick-Muraca, "Quantitative optical spectroscopy for tissue diagnosis," *Annu. Rev. Phys. Chem.* **47**, 555–606 (1996).
4. 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," *Appl. Opt.* **44**, 4291–4305 (2005).
5. 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," *J. Biomed. Opt.* **5**, 221–228 (2000).
6. 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," *Gastrointest. Endosc.* **63**, 257–261 (2006).
7. I. J. Bigio and S. G. Bown, "Spectroscopic sensing of cancer and cancer therapy: current status of translational research," *Cancer Biol. Therapy* **3**, 259–267 (2004).
8. G. A. Wagnieres, W. M. Star, and B. C. Wilson, "In vivo fluorescence spectroscopy and imaging for oncological applications," *Photochem. Photobiol.* **68**, 603–632 (1998).
9. V. Backman, M. B. Wallace, L. T. Perelman, J. T. Arendt, R. Gurjar, M. G. Müller, Q. Zhang, G. Zonios, E. Kline, T. McGillican, S. Shapshay, T. Valdez, K. Badizadegan, J. M. Crawford, M. Fitzmaurice, S. Kabani, H. S. Levin, M. Seiler, R. R. Dasari, I. Itzkan, J. Van Dam, and M. S. Feld, "Detection of preinvasive cancer cells," *Nature* **406**, 35–36 (2000).
10. I. Georgakoudi, B. C. Jacobson, J. Van Dam, V. Backman, M. B. Wallace, M. G. Müller, Q. Zhang, K. Badizadegan, D. Sun, and G. A. Thomas, "Fluorescence, reflectance, and light-scattering spectroscopy for evaluating dysplasia in patients with Barrett's esophagus," *Gastroenterology* **120**, 1620–1629 (2001).
11. M. Hunter, V. Backman, G. Popescu, M. Kalashnikov, C. W. Boone, A. Wax, V. Gopal, K. Badizadegan, G. D. Stoner, and M. S. Feld, "Tissue self-affinity and polarized light scattering in the Born approximation: a new model for precancer detection," *Phys. Rev. Lett.* **97**, 138102 (2006).
12. M. G. Müller, T. A. Valdez, I. Georgakoudi, V. Backman, C. Fuentes, S. Kabani, N. Laver, Z. Wang, C. W. Boone, and R. R. Dasari, "Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma," *Cancer* **97**, 1681–1692 (2003).
13. L. T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J. M. Crawford, and M. S. Feld, "Observation of periodic fine structure in reflectance from biological tissue: a new technique for measuring nuclear size distribution," *Phys. Rev. Lett.* **80**, 627–630 (1998).
14. 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**, 6628–6637 (1999).
15. C. W. Kan, L. T. Nieman, K. Sokolov, and M. K. Markey, "AI in clinical decision support: applications in optical spectroscopy for cancer detection and diagnosis," in *Advanced Computational Intelligence Paradigms in Healthcare 3*, M. Sordo, S. Vaidya, and L. C. Jain, Eds., pp. 27–49, Springer, New York (2008).
16. Y. N. Mirabal, S. K. Chang, E. N. Atkinson, A. Malpica, M. Follen, and R. Richards-Kortum, "Reflectance spectroscopy for in vivo detection of cervical precancer," *J. Biomed. Opt.* **7**, 587–594 (2002).
17. M. P. L. Bard, A. Amelink, M. Skurichina, M. den Bakker, S. A. Burgers, J. P. van Meerbeeck, R. P. W. Duin, J. G. J. V. Aerts, H. C. Hoogsteden, and H. J. C. M. Sterenberg, "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**, 41–47 (2005).
18. 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," *J. Biomed. Opt.* **9**, 1122–1128 (2004).
19. S. D. Kamath and K. K. Mahato, "Optical pathology using oral tissue fluorescence spectra: classification by principal component analysis and k-means nearest neighbor analysis," *J. Biomed. Opt.* **12**, 014028 (2007).
20. N. Ramanujam, M. F. Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, "Spectroscopic diagnosis of cervical intraepithelial neoplasia (CIN) in vivo using laser-induced fluorescence spectra at multiple excitation wavelengths," *Lasers Surg. Med.* **19**, 63–74 (1996).
21. 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," *J. Biomed. Opt.* **12**, 014030 (2007).
22. X. Wang and L. V. Wang, "Propagation of polarized light in birefringent turbid media: a Monte Carlo study," *J. Biomed. Opt.* **7**, 279–290 (2002).
23. X. Wang, L. V. Wang, C.-W. Sun, and C.-C. Yang, "Polarized light propagation through scattering media: time-resolved Monte Carlo simulations and experiments," *J. Biomed. Opt.* **8**, 608–617 (2003).
24. 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," *J. Biomed. Opt.* **12**, 014029 (2007).
25. B. C. Wilson and G. Adam, "A Monte Carlo model for the absorption and flux distributions of light in tissue," *Med. Phys.* **10**, 824–830 (1983).
26. C. Zhu, G. M. Palmer, T. M. Breslin, J. Harter, and N. Ramanujam, "Diagnosis of breast cancer using fluorescence and diffuse reflectance spectroscopy: a Monte-Carlo-model-based approach," *J. Biomed. Opt.* **13**, 034015 (2008).
27. T. M. Johnson and J. R. Mourant, "Polarized wavelength-dependent measurements of turbid media," *Opt. Express* **4**, 200–216 (1999).
28. 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," *J. Biomed. Opt.* **12**, 014021 (2007).
29. I. J. Bigio, "Clinical applications of elastic-scattering spectroscopy beyond proof-of-principle: what really matters," *Biomedical Application of Light Scattering, Proc. SPIE* **6446**, 64460J (2007).
30. L. T. Nieman, C.-W. Kan, A. Gillenwater, M. K. Markey, and K. Sokolov, "Probing local tissue changes in the oral cavity for early detection of cancer using oblique polarized reflectance spectroscopy: a pilot clinical trial," *J. Biomed. Opt.* **13**, 024011 (2008).
31. J. R. Mourant, I. J. Bigio, J. Boyer, R. L. Conn, T. Johnson, and T. Shimada, "Spectroscopic diagnosis of bladder cancer with elastic light scattering," *Lasers Surg. Med.* **17**, 350–357 (1995).
32. 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**, 205–215 (2004).
33. J. A. Hanley, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology* **143**, 29–36 (1982).
34. T. Raykov and G. A. Marcoulides, *An Introduction to Applied Multivariate Analysis*, Routledge/Psychpress, New York (2008).
35. B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, New York (1993).