

# Combined reflectance and fluorescence spectroscopy for *in vivo* detection of cervical pre-cancer

## Sung K. Chang

Massachusetts General Hospital  
Wellman Center for Photomedicine  
Boston, Massachusetts 02114

## Yvette N. Mirabal

University of Texas  
M.D. Anderson Cancer Center  
Biomedical Engineering Center  
Houston, Texas 77030

## Edward Neely Atkinson

University of Texas  
M.D. Anderson Cancer Center  
Department of Biomathematics  
Houston, Texas 77030

## Dennis Cox

Rice University  
Department of Statistics  
Houston, Texas 77030

## Anais Malpica

University of Texas  
M.D. Anderson Cancer Center  
Department of Pathology  
Houston, Texas 77030

## Michele Follen

University of Texas  
M.D. Anderson Cancer Center  
Biomedical Engineering Center  
Houston, Texas 77030

## Rebecca Richards-Kortum

University of Texas at Austin  
Department of Biomedical Engineering  
ENS Room 8  
Austin, Texas 78712  
E-mail: kortum@mail.utexas.edu

## 1 Introduction

Cervical cancer is the third most common cancer in women worldwide and the leading cause of cancer mortality for women in developing countries.<sup>1</sup> Early detection programs based on the Papanicolaou smear and colposcopy have helped reduce both the incidence and the mortality of cervical cancer.<sup>2</sup> However, the sensitivity and specificity of the Papanicolaou smear range from 11 to 99 and 14 to 97%, respectively.<sup>3</sup> Because of this limitation, colposcopy is performed following an abnormal Pap smear. Although colposcopy provides good sensitivity (>90%), its specificity is still poor (<50%), requiring a biopsy to confirm the diagnosis of cervical precancer.<sup>4</sup> The effectiveness of early detection can be improved by developing more sensitive methods for screening and diagnosis.

Address all correspondence to Rebecca Richards-Kortum, Univ. of Texas at Austin, ENS 610, Biomedical Engineering Program, Austin, TX 78712. Tel: 512-471-2104; Fax: 512-471-0616; E-mail: kortum@mail.utexas.edu

**Abstract.** Optical technologies, such as reflectance and fluorescence spectroscopy, have shown the potential to provide improved point-of-care detection methods for cervical neoplasia that are sensitive, specific, and cost-effective. Our specific goals are to analyze the diagnostic potential of reflectance and fluorescence spectra, alone and in combination, to discriminate normal and precancerous cervical tissue *in vivo* and to identify which classification features contain significant diagnostic information. Reflectance spectra are measured at four source-detector separations and fluorescence emission spectra are measured at 16 excitation wavelengths, from 324 sites in 161 patients. These 20 spectral features are permuted in all possible combinations of one, two, and three; and classification algorithms are developed to evaluate the diagnostic performance of each combination. Algorithms based on fluorescence spectra alone yield better diagnostic performance than those based on reflectance spectra alone. The combination of fluorescence and reflectance do not significantly improve diagnostic performance compared to fluorescence alone, except in the case of discriminating high-grade precancers from columnar normal tissue. In general, fluorescence emission spectra at 330- to 360-nm and 460- to 470-nm excitation provide the best diagnostic performance for separating all pairs of tissue categories. © 2005 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1899686]

Keywords: fluorescence; reflectance; tissues; biomedical optics; neoplasia; cervix; *in vivo* diagnosis.

Paper 04057 received Apr. 13, 2004; revised manuscript received Sep. 24, 2004; accepted for publication Sep. 30, 2004; published online Apr. 22, 2005.

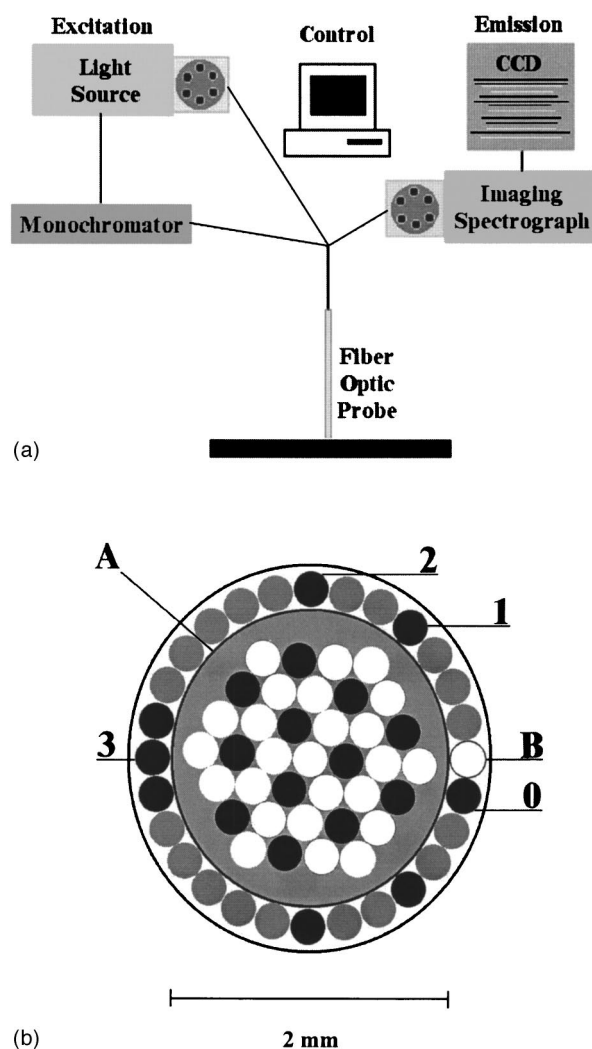
As an alternative, many groups have demonstrated that techniques based on quantitative optical spectroscopy show promise as a potential diagnostic tool for detecting epithelial neoplasia. In particular, diffuse reflectance and fluorescence spectroscopy have shown initial success for precancer detection in various organ sites including the bladder,<sup>5</sup> the uterine cervix,<sup>6-9</sup> the colon,<sup>10,11</sup> the esophagus,<sup>12,13</sup> and the breast.<sup>14</sup> Coppelson et al.<sup>9</sup> developed a fiber optic probe (polar probe) to measure cervical tissue reflectance at four wavelengths in the visible and near-IR (NIR) region. An empirical algorithm was developed using *in vivo* data from 77 volunteers; its diagnosis agreed with colposcopy and histology in 85 to 99% of measurements, depending on tissue type. Based on the increased penetration depth of light at larger separation between the source and detector fibers,<sup>15</sup> other studies have investigated spatially resolved measurements of diffusely reflected

light to enhance diagnostic performance.<sup>16,17</sup> Nordstrom et al.<sup>18</sup> carried out a clinical trial where 41 patients with abnormal colposcopy were measured *in vivo* using both diffuse reflectance and UV-excited fluorescence spectroscopy for detection of cervical precancer. A multivariate algorithm based on the Mahalanobis distance used reflectance spectra to discriminate normal squamous (SN) tissue and high-grade squamous intraepithelial lesions (HG-SIL) with a sensitivity and specificity of 82 and 67%, respectively, from reflectance measurements. A similar approach using fluorescence spectra resulted in a sensitivity and specificity of 91 and 93%, respectively.<sup>18</sup> Georgakoudi et al. used a combination of *in vivo* reflectance and fluorescence spectroscopy to calculate the intrinsic fluorescence from cervical tissue,<sup>8</sup> which resulted in a sensitivity and specificity of 62 and 92%, respectively.

The diagnostic capability of fluorescence and diffuse reflectance spectroscopy derives from the ability of these techniques to probe the metabolic and architectural changes at the cellular and molecular levels that accompany the development of neoplasia. For example, one of the important clinical markers for diagnosis of cervical precancer is hemoglobin concentration in tissue, which increases during dysplastic progression due to angiogenic developments.<sup>19</sup> Fluorescence and reflectance spectra collected from tissue effectively monitor the level of hemoglobin concentration by measuring the light absorption from the chromophore and thereby provide important diagnostic information.<sup>20</sup> On a similar note, reflectance spectroscopy is also sensitive to light scattering in tissue. Electromagnetic modeling predicts that the intensity of cellular light scattering increases with progression of cervical precancer due to changes in nuclear size and DNA content.<sup>21</sup> Collier used reflectance measurements from a confocal imaging system to quantify the level of nuclear scattering from the cervical epithelium, and demonstrated the diagnostic potential of cellular light scattering properties in separating normal and HG-SIL tissue.<sup>22</sup>

Among the various intrinsic fluorophores in tissue, fluorescence from cofactors NADH (reduced nicotinamide adenine dinucleotide) and FAD (flavin adenine dinucleotide) convey important information about the cellular metabolic state. Confocal fluorescence microscopy images of fresh tissue slices have revealed differences in cytoplasmic fluorescence patterns from normal and precancerous biopsy specimens,<sup>23</sup> possibly due to variations in the metabolic state of tissue during dysplastic progression.

In most cases, the diagnostic capabilities of fluorescence and reflectance spectroscopy have been investigated separately. Several recent small studies have suggested that the combinations of both techniques may yield improved diagnostic performance.<sup>8,18</sup> These studies have either been conducted at a single excitation wavelength for fluorescence<sup>18</sup> or a small number of excitation wavelengths.<sup>8</sup> Furthermore, the diagnostic performance of spatially resolved reflectance in combination with fluorescence measurements has not been investigated previously. In this paper, we explore the utility of combining spatially resolved reflectance spectra and fluorescence spectra measured at a wide range of excitation wavelengths for the detection of cervical precancer.



**Fig. 1** (a) System block diagram showing the light source assembly, a fiber optic probe for delivery and collection of light, and the spectrograph assembly, and (b) schematic diagram of the distal end of the probe: [A] fluorescence excitation (white circles) and collection (black circles) fiber bundle, [B] reflectance illumination fiber (white circle) and reflectance collection fibers at positions 0 to 3 (black circles labeled with 0 to 3 for each respective position).

## 2 Materials and Methods

### 2.1 Instrumentation

The spectroscopic system used to measure fluorescence and reflectance spectra has previously been described in detail.<sup>24</sup> Briefly, the system incorporates three main components: (1) a xenon arc lamp to provide broadband illumination for reflectance and fluorescence excitation light using bandpass filters; (2) a fiber optic probe that directs the light to tissue and collects diffusely reflected and fluorescent emission light; and (3) an optical assembly with an imaging spectrograph (Chromex 250 IS, Albuquerque, New Mexico) and a thermoelectrically cooled CCD camera (Spectrasource HPC-1, Westlake Village, California) to record the spectral data. Figure 1(a) illustrates the system.

The probe, whose distal end is illustrated in Fig. 1(b), utilizes a fiber optic bundle for fluorescence measurement in the

core surrounded by nine spatially separated optical fibers [200- $\mu\text{m}$ -diam fibers, numerical aperture (NA)=0.2] to measure reflectance. The fluorescence bundle consists of a random arrangement of 25 illumination and 12 collection fibers, with a 15-mm-long quartz mixing element (2 mm diameter) at the distal end of the bundle to diffuse the excitation and collection light at the measurement site. Fluorescence excitation wavelengths range from 330 to 480 nm in 10-nm increments and each emission spectrum is sampled at 5-nm intervals. Of the nine reflectance fibers, one excitation fiber provides broadband illumination and eight reflectance collection fibers are placed at four different source-detector separations (position 0, 250- $\mu\text{m}$  separation; position 1, 1.1-mm separation; position 2, 2.1-mm separation; position 3, 3.0-mm separation) to collect diffusely reflected light. The emission wavelength in each reflectance spectrum ranges between 355 and 655 nm in 2.5-nm intervals. A single spectroscopic measurement consists of fluorescence emission spectra from 16 different excitation wavelengths and four reflectance spectra measured in sequence in approximately 2 min.

## 2.2 Clinical Measurements

The study protocol was reviewed and approved by the Institutional Review Boards at the University of Texas M. D. Anderson Cancer Center and the University of Texas at Austin. Details of the clinical study are provided in Refs. 25 and 26. A health-care provider described the study to eligible patients who had been referred on the basis of an abnormal Papanicolaou smear; written consent was obtained from those agreeing to participate. Following colposcopic examination, but prior to biopsy, a fiber optic probe was advanced through the speculum and placed in contact with the cervix. Spectra were measured from up to four sites in each patient: one colposcopically normal cervical site covered with squamous epithelium, one or two colposcopically abnormal cervical sites, and if visible, one colposcopically normal cervical site covered with columnar epithelium. Following spectroscopic measurements, all sites were biopsied.

Within 2 h of each patient measurement, spectra from reflectance and fluorescence standards were measured. As a positive control for reflectance measurements, reflectance spectra were measured from a 1-cm-path length cuvette containing a suspension of 1.02- $\mu\text{m}$ -diam polystyrene microspheres (6.25% by volume). Fluorescence spectra measured from a solution of Rhodamine 610 (Exciton, Dayton, Ohio) dissolved in ethylene glycol (2 mg/ml) in a 1-cm-path length cuvette was used for positive control of fluorescence measurements. As a negative control, reflectance and fluorescence spectra were measured with the probe tip immersed in a large container of distilled water to record levels of various background signal.

Biopsies were fixed and submitted for permanent section. The 4- $\mu\text{m}$ -thick sections were stained with both hematoxylin and eosin (H&E) as well as Feulgen stains. Two pathologists who were blinded to the results of spectroscopy read each biopsy, with discrepant cases reviewed a third time for consensus diagnosis by the study histopathologist. Diagnostic classification categories included normal tissue, human papilloma virus infection (HPV), grade 1 cervical intraepithelial neoplasia (CIN 1), grade 2 cervical intraepithelial neoplasia (CIN 2), grade 3 cervical intraepithelial neoplasia (CIN 3),

and carcinoma *in situ* (CIS) using standard histopathologic criteria.<sup>2</sup> Normal tissues were divided into two categories based on colposcopic impression: normal squamous epithelium (SN) and normal columnar epithelium (CN). Tissues with acute/chronic inflammation or metaplasia were included in the corresponding SN or CN category. In accordance with the Bethesda system, HPV and CIN 1 were termed low grade squamous intraepithelial lesions (LGSILs), whereas CIN 2, CIN 3, and CIS were termed high grade squamous intraepithelial lesions (HGSILs). The diagnostic categories SN, CN, LGSIL, and HGSIL were used in this analysis.

## 2.3 Data Processing and Statistical Analysis

Three investigators (YM, DDC, RRK) blinded to the pathologic results reviewed all spectra. Spectra indicating evidence of user or instrument error, such as probe slippage, were discarded from further analysis. Reflectance spectra at each source-detector separation were normalized by the corresponding spectrum from the microsphere suspension to correct for the effects of the source spectrum, variations in the illumination intensity, and the wavelength-dependent response of the detection system. For each fluorescence measurement, variations in the source light were corrected with excitation illumination intensity measured at the probe tip using a calibrated photodiode (818-UV, Newport Research Corp.). To correct for the nonuniform spectral response of the detection system, the spectra of two calibrated sources were measured at the beginning of the study; a National Institute of Standards and Technology (NIST) traceable calibrated tungsten ribbon filament lamp in the visible range and a deuterium lamp (550C and 45D, Optronic Laboratories Inc, Orlando, Florida) in the UV range. System response correction factors for fluorescence emission spectra were derived from these calibration spectra.

Reflectance data from a single measurement site are represented as a matrix containing calibrated reflectance intensity as a function of source-detector separation and emission wavelength. Spectra from each of the four source-detector separation positions form column vectors containing 121 intensity measurements corresponding to emission wavelengths from 355 to 655 nm in 2.5-nm increments. Fluorescence data from a single measurement site are represented as an excitation-emission matrix (EEM), where the emission spectra at the various excitation wavelengths are concatenated into a 2-D matrix so that the calibrated fluorescence intensity is expressed as a function of excitation and emission wavelength. Columns of this matrix correspond to emission spectra at each excitation wavelength, containing between 50 to 130 intensity measurements ranging from 380 to 700 nm emission in 5-nm increments. The excitation wavelengths range from 330 to 480 nm in 10-nm increments.

The spectroscopic data were then analyzed to determine which reflectance source-detector separations and fluorescence excitation wavelengths, termed classification features, contained the most diagnostically useful information to separate a pair of diagnostic categories of the cervix. We developed classification algorithms to discriminate SN versus CN, SN versus LGSIL, SN versus HGSIL, CN versus LGSIL, and CN versus HGSIL from the following three datasets: combinations of four reflectance features, combinations of 16 fluorescence features, and combinations of 20 integrated features.

**Table 1** Number of different possible feature vector combinations used to evaluate diagnostic performance.

	1	2	3	4
Number of feature vectors in a combination	1	2	3	4
Number of combinations using only reflectance spectra	4	6	4	1
Number of combinations using only fluorescence emission spectra	16	120	560	Not used
Number of combinations using both reflectance and fluorescence emission spectra	20	190	1140	Not used

The diagnostic performance of each combination was evaluated with the classification algorithm. In the analysis using only the reflectance features, up to four spectra at different source-detector separations were considered as input to the classifiers, whereas in analyses of fluorescence alone and combination of fluorescence and reflectance, combinations of up to three reflectance spectra or fluorescence emission spectra were considered. Table 1 lists the number of different possible combinations of feature vectors considered in each analysis.

Classification algorithms were developed to separate data from the two diagnostic classes under analysis. The algorithm development was described previously,<sup>25</sup> and consists of data reduction using principal component analysis (PCA) followed by binary classification using Mahalanobis distance. Each step is described in detail in the following.

Prior to PCA, an input matrix was assembled with the specified feature vector combination from the two diagnostic classes. For each measurement, fluorescence and reflectance spectra from the combinational features were concatenated end-to-end as a single row vector. To reduce interpatient variation, each fluorescence spectrum was normalized by its maximum intensity prior to concatenation. These row vectors were concatenated in a column to form the input matrix.

Eigenvectors of the corresponding covariance matrix were then calculated to generate the principal components; those accounting for up to 65, 75, 85, and 95% of the total variance were retained for algorithm development. We denote the fraction of the total variance accounted for by the eigenvectors as the eigenvector significance level (ESL). Principal component scores of each measurement in the input matrix were calculated using the selected eigenvectors.

Classifiers based on the principal component scores were generated to perform binary classification into the two diagnostic classes under analysis. Classification is based on the Mahalanobis distance  $r_i^2$  which is a multivariate measure of the separation of a data point from the mean of a dataset in  $n$ -dimensional space:<sup>27</sup>

$$r_i^2 = (\mathbf{x} - \bar{\mathbf{x}}_i)' \cdot \mathbf{C}_x^{-1} \cdot (\mathbf{x} - \bar{\mathbf{x}}_i). \quad (1)$$

Here,  $\mathbf{x}$  is the vector containing principal component scores from a sample,  $\bar{\mathbf{x}}_i$  is the mean of the principal component scores from diagnostic class  $i$ , and  $\mathbf{C}_x$  is the covariance matrix. The multivariate distance between the sample to be clas-

**Table 2** Data set classified by histopathologic diagnosis.

Diagnostic Class	SN	CN	HPV	CIN 1	CIN 2	CIN 3/CIS	Total
Number of sites (161 patients)	227	18	52	9	3	15	324

sified and the means of the two possible classification groups was calculated; the sample was then assigned to the group that it was closest to in this multivariate space.

The performance of classification depends on the principal component scores included for analysis. For each eigenvector selected at an ESL, the corresponding set of principal component scores were applied to the Mahalanobis distance classification, and the set yielding the best initial performance was retained in the data matrix  $\mathbf{M}$  for analysis. Among the remaining eigenvectors, the set of principal component scores that improved this performance most when combined with  $\mathbf{M}$  was selected in sequence. This process was repeated until performance was no longer enhanced by the addition of principal components, or until all components were selected.

The diagnostic performance of the data matrix  $\mathbf{M}$  at each ESL was evaluated relative to the histopathologic diagnosis. The Mahalanobis classifier was trained and tested using all the samples in  $\mathbf{M}$ . In calculating the sensitivity and the specificity for each pair of diagnostic classes, diseased tissue was taken as the positive sample relative to either columnar or squamous normal tissue. However, when CN was discriminated against SN, columnar normal tissue was taken as the positive sample relative to squamous normal tissue.

A potential problem with this approach is that it may overestimate sensitivity and specificity due to overtraining. To minimize the effect of overtraining, we carried out each analysis once with the true diagnosis and 50 times when the diagnosis was randomly assigned. The total number of positive and negative samples was kept the same when generating the set of randomized diagnosis. We ranked each feature combination according to the difference in the sum of the sensitivity and specificity obtained with the true diagnosis with that of the average sensitivity and specificity from the randomized diagnosis. Since leave-one-out cross-validation provides a less biased estimate of algorithm performance,<sup>28</sup> the diagnostic performance of the top 25 ranking combinations was further evaluated using leave-one-out cross-validation.

### 3 Results

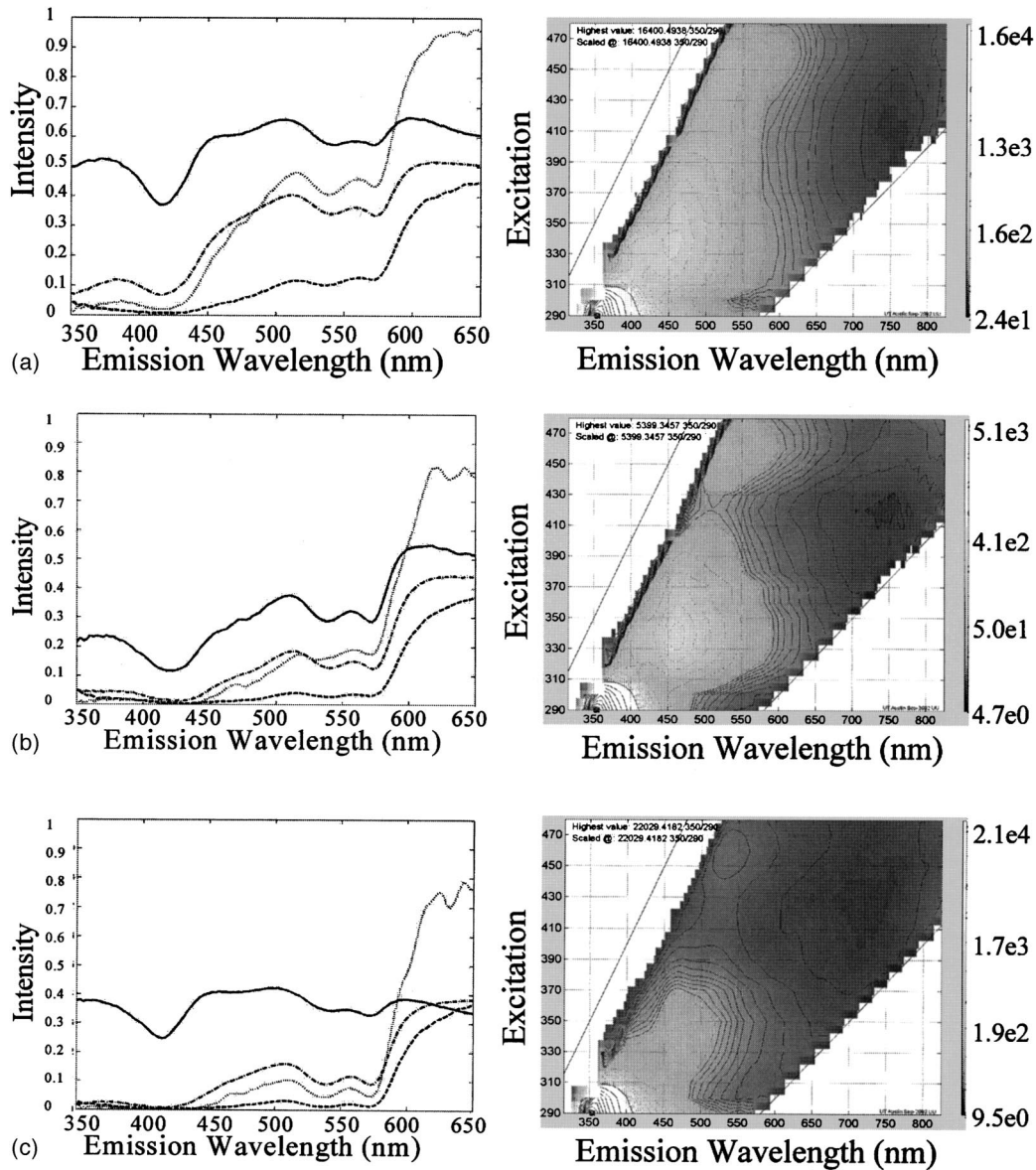
#### 3.1 Data Set

The data set consisted of a set of spectra from 324 sites in a group of 161 patients that were deemed adequate for both reflectance and fluorescence analysis by independent reviewers. Table 2 shows the diagnostic composition of the data set. Tissues with acute/chronic inflammation or metaplasia were included in the corresponding squamous or columnar normal category.

#### 3.2 Reflectance Spectra

Typical reflectance and fluorescence spectra from three measurement sites diagnosed as normal squamous [Fig. 2(a)], normal columnar [Fig. 2(b)], and CIS [Fig. 2(c)] are shown in





**Fig. 2** Typical *in vivo* reflectance spectra (left) and fluorescence EEMs (right) from cervical tissue: (a) normal squamous, (b) normal columnar, and (c) carcinoma *in situ*. In the left column, reflectance spectra at four different source-detector separations (position 0=—; position 1=— · —; position 2=— · — · —; position 3=·····), normalized by a standard microspheres solution, are shown. In the right column, fluorescence EEM data are shown.

Fig. 2. The reflectance spectra from each site at the four different source-detector separation positions are shown in the left column of Fig. 2. Positions 0, 1, 2, and 3 correspond to an increasingly greater source-detection separation. All reflectance spectra show valleys due to hemoglobin absorption at 420, 542, and 577 nm. In general, reflectance intensity decreases from SN tissue to abnormal tissue, with the most significant level of attenuation observed with HGSIL. Reflectance intensity from CN tissue is low compared to that from SN tissue.

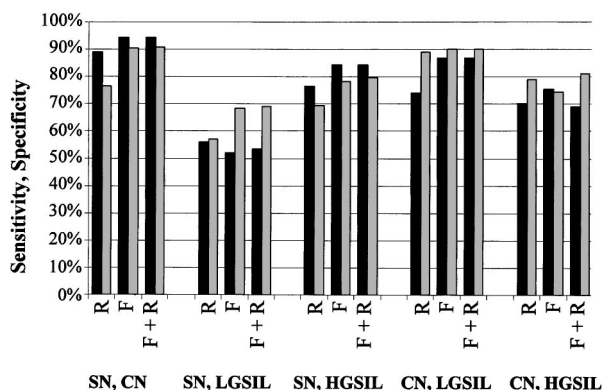
### 3.3 Fluorescence Spectra

The fluorescence EEMs measured at the identical sites are shown in the right column of Fig. 2. Fluorescence peaks from several fluorophores are evident. The peak at 350 nm

excitation/450 nm emission is due to cofactor NADH as well as collagen crosslinks, while the peak along 525-nm emission at both 370- and 450-nm excitation is due to cofactor FAD and collagen crosslinks. Fluorescence from endogenous porphyrin, if present, appears as a peak at 410 nm excitation/630 nm emission. Absorption due to hemoglobin causes valleys parallel to the excitation and emission wavelength axes along 420, 540, and 580 nm. As observed in the reflectance spectra, the hemoglobin absorption valley is generally more prominent in abnormal tissue compared to squamous normal tissue.

### 3.4 Statistical Analysis

Figure 3 shows the average cross-validated sensitivity and specificity of the five best-performing feature combinations among all possible combinations; results are shown from

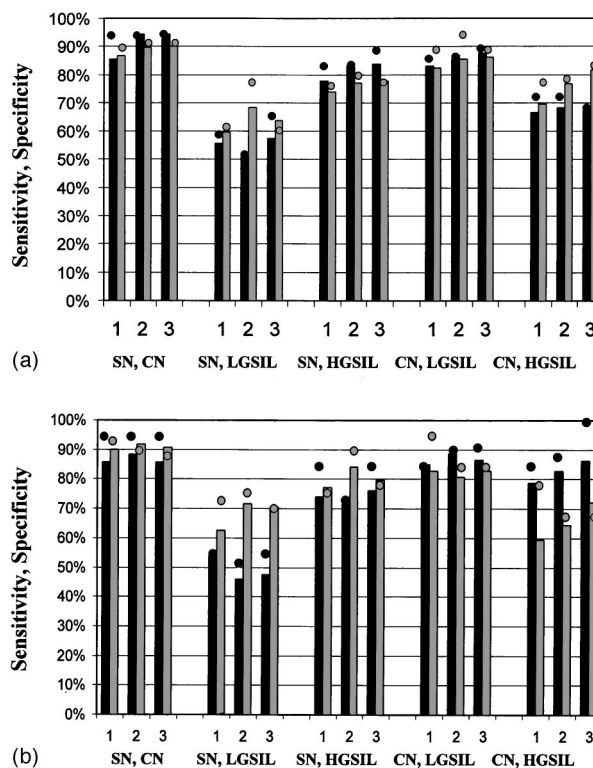


**Fig. 3** Average sensitivity (black) and specificity (gray) of the five best performing classification combinations in each pairwise analysis (at ESL of 65%). Results are shown for reflectance and fluorescence features combined (R+F) when selecting combinations of up to three features, fluorescence spectra alone (F) when selecting combinations of up to three excitation wavelengths, and reflectance spectra alone (R) when selecting combinations of up to four source-detector separations.

analyses using reflectance spectra alone, fluorescence spectra alone, and reflectance combined with fluorescence. In general, discrimination between SN and CN gave the best performance, followed by CN versus LGSIL, SN versus HGSIL, and CN versus HGSIL. Discrimination between SN and LGSIL resulted in the lowest sensitivity and specificity. For all pairs of diagnostic categories, the use of reflectance alone resulted in good diagnostic performance; however, better performance is achieved using fluorescence data alone. In the cases of SN versus LGSIL, SN versus HGSIL, and CN versus HGSIL, the addition of reflectance features to fluorescence features showed modest improvement in the average performance compared to the results using only the fluorescence spectra. For SN versus CN and CN versus LGSIL, the average performance from the combination of fluorescence and reflectance spectra was equal to those from fluorescence spectra alone.

The average cross-validated performances from the top 10 performing combinations of one, two, and three features among 20 reflectance and fluorescence features are shown in Fig. 4. The sensitivity and specificity of the single best performing combination in each analysis is indicated with black and gray dots, respectively. Figure 4(a) shows the average performance from eigenvectors selected at an ESL of 65%. The diagnostic performance is high when limited to the use of a single feature, and a small increase in performance is observed when a second feature is added. However, addition of a third feature does not result in increased performance in many cases. Again, the best performance is obtained when discriminating between SN and CN, reaching an average sensitivity of 94% and specificity of 90% with the use of two or three classification features. Increasing the ESL from 65 to 95% does not noticeably increase performance, as shown in Fig. 4(b).

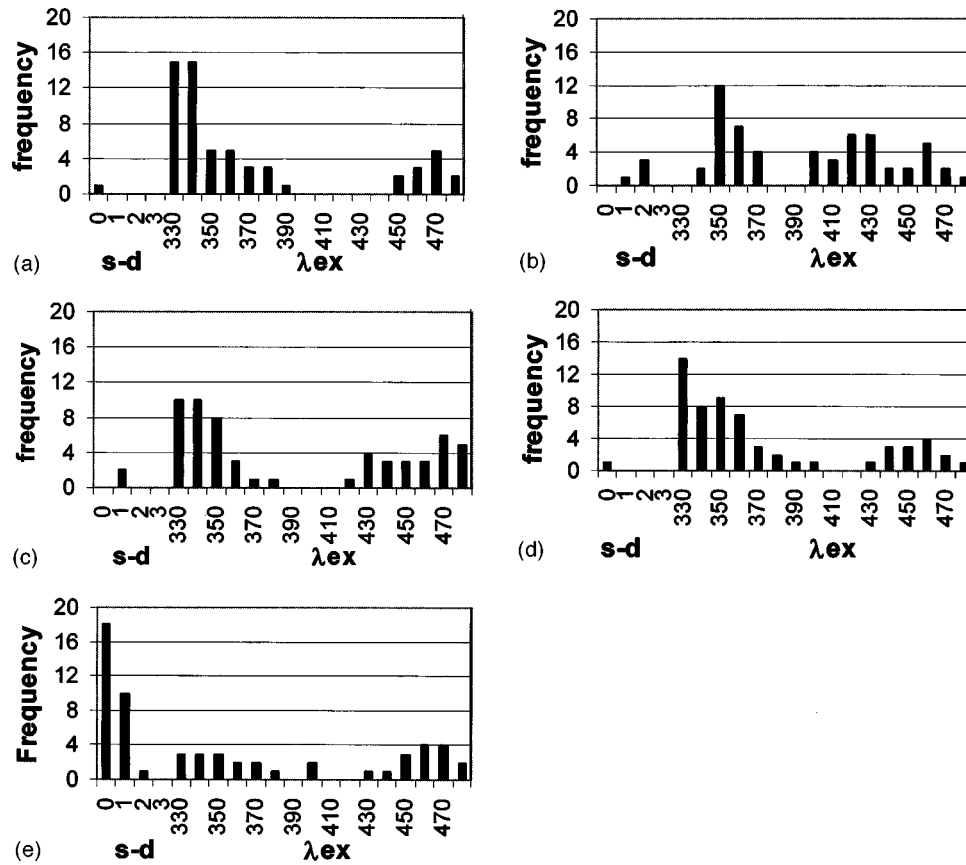
Figure 5 shows the frequency with which each classification feature appears among the 10 best-performing feature combinations selected in Fig. 4(a). Figure 5(a) shows that in discriminating SN and CN tissues, fluorescence excitation



**Fig. 4** Average sensitivity (black) and specificity (gray) of the top 10 performing combinations for each pairwise analysis when the 20 classification features from fluorescence and reflectance measurements are combined one, two, or three at a time at (a) an ESL of 65% and (b) an ESL of 95%. Black and gray dots indicate the sensitivity and specificity of the best performing combination, respectively.

wavelengths between 330 and 360 nm and 440 and 470 nm occur relatively frequently. Similarly, for discrimination of SN from HGSILs [Fig. 5(c)] and CN from LGSILs [Fig. 5(d)], fluorescence excitation wavelengths between 330 and 360 nm and 440 and 480 nm appear relatively frequently. Note that for separating CN tissue from HGSILs [Fig. 5(e)], reflectance source-detector separations 0 and 1 occurred more frequently than any fluorescence excitation wavelength or any other reflectance feature.

Table 3 compares the classification features identified as significant from the three different trials: one using only the fluorescence features, another using only the reflectance features and the other using integrated features from fluorescence and reflectance data. In each trial, significant features were identified for the following five different analyses: SN versus CN, SN versus LGSIL, SN versus HGSIL, CN versus LGSIL, and CN versus HGSIL. In the trial using reflectance features only, source-detector separation positions 0 and 1 appear significant in all five analyses. The trial using only fluorescence features shows that excitation wavelengths between 330 and 360 nm and those between 460 and 480 nm appear frequently in most analyses. Note that in four analyses (SN versus CN, SN versus LGSIL, SN versus HGSIL, and CN versus LGSIL), only the fluorescence excitation wavelengths are identified as significant features in the trial integrating both fluorescence and reflectance spectra. The selected wavelength ranges correspond well with those identified in the trial using fluores-



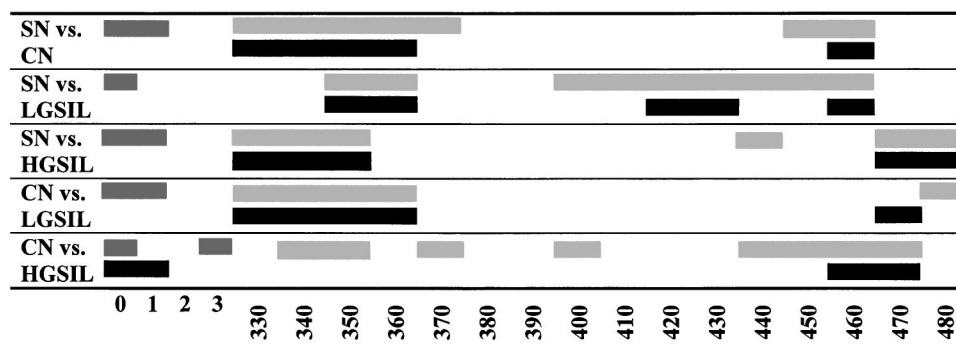
**Fig. 5** Histograms showing frequency of appearance of each classification feature among the top 10 performing combinations when considering up to three features at ESL of 65%. Results from the five pairwise analyses are shown: (a) SN versus CN, (b) SN versus LGSIL, (c) SN versus HGSIL, (d) CN versus LGSIL, and (e) CN versus HGSIL. ESL=65%. The four reflectance source detector separation positions and the 16 fluorescence excitation wavelengths are indicated by s-d and  $\lambda_{ex}$ , respectively.

cence features. However, when discriminating CN and HGSIL, reflectance positions 0 and 1 were selected in addition to the fluorescence features.

Based on the results in Table 3, we attempted to identify a set of three features that yields the best overall performance in all the five pairwise analyses. All combinations of the 20 classification features that appear in Table 3 were combined into sets of three, and the combinations that gave the best performance in each pairwise analysis were identified. The overall diagnostic performance of each combination in this analysis was calculated as the average sensitivity and specificity from the five pairwise classification algorithms. When the available features were limited to three, we found that fluorescence emission spectra at excitation wavelengths of 330, 430, and 470 nm resulted in optimal overall classification performance. The feature combination and the corresponding classification performance for each pairwise analysis from the limited-feature set are listed in Table 4. When the number of available features was increased to four, fluorescence excitation wavelengths 330, 360, 430, and 470 nm were selected as the set resulting in best overall performance (Table 5). Performance of the limited-feature sets is comparable to those when all possible combinations of classification features were tested [Fig. 4(a)]. Note that limiting the number of available features from three to four did not result in significant improvement in classification performance.

To investigate the diagnostic information inherent in the classification features that were selected as significant, we plotted the spectra of correctly classified and misclassified samples in each pairwise analysis of Table 5. In Fig. 6(a), the left plot shows the classification of samples from SN when discriminated against CN. The heavy black plot is the average of all the correctly classified SN samples, and the thin gray lines are the individual misclassified samples. The right plot corresponds to results of CN samples when discriminated against SN. The main discriminating factors between SN and CN samples in Fig. 6(a) appear to be the valley around 380-nm emission at 330-nm excitation and that around 570 nm at 470-nm excitation. Note that these valleys correspond to hemoglobin absorption peaks. Figure 6(b) shows similar plots for the case of SN versus LGSIL, where the plot on the left shows the samples from SN and that on the right shows the samples from LGSIL. We find that, on average, the peak of the correctly classified SN samples is toward the lower wavelengths compared to that from correctly classified LGSIL samples. Figure 6(c) shows equivalent plots for the case of SN versus HGSIL, with the plot for SN samples on the left and that for HGSIL samples on the right. In Fig. 6(d), the plot on the left shows samples from CN and that on the right shows those from LGSIL. In both pairwise analyses, the valley around 380-nm emission at 330-nm excitation appears to be an evident discriminating factor as well as the peak shift in

**Table 3** Classification features that appeared most frequently among the 10 best-performing combinations when taken up to three features for the following five analyses: SN versus CN, SN versus LGSIL, SN versus HGSIL, CN versus LGSIL, and CN versus HGSIL. Results from the following three trials are shown: one using only the reflectance features only (dark gray bars), another using only the fluorescence features only (light gray bars), and the other using both reflectance and fluorescence features (black bars).



either 430-nm excitation or the 360-nm excitation. The main discriminating factor between CN and HGSIL, as shown in Fig. 6(e), is the hemoglobin absorption valley around 580-nm emission, which is more prominent in CN (left figure).

#### 4 Discussion and Conclusions

In our study of the diagnostic potential of combined fluorescence and reflectance spectroscopy, we obtained cervical *in vivo* measurements at four distinct source-detector separation positions and for 16 fluorescence excitation wavelengths. Using Mahalanobis distance-based classification, we determined which classification combination contained the most diagnostically useful information. Results showed the sensitivity and specificity to be high when using a single classification feature at the lowest level of eigenvector significance considered. The addition of a second classification feature did increase the sensitivity and specificity; however, there was no noticeable increase in classification performance when data from higher ESLs are included. Furthermore, fluorescence excitation wavelengths between 330 and 360 nm and 460 and 470 nm and reflectance source-detector separations at positions 0 and 1 appear most frequently among the best performing classification feature combinations.

**Table 4** Diagnostic performance of each pairwise analysis when the available classification features are limited to three (fluorescence excitation wavelengths of 330, 430, and 470 nm). The three classification features were selected on the basis of best overall diagnostic performance in all five pairwise analyses at an ESL=65%.

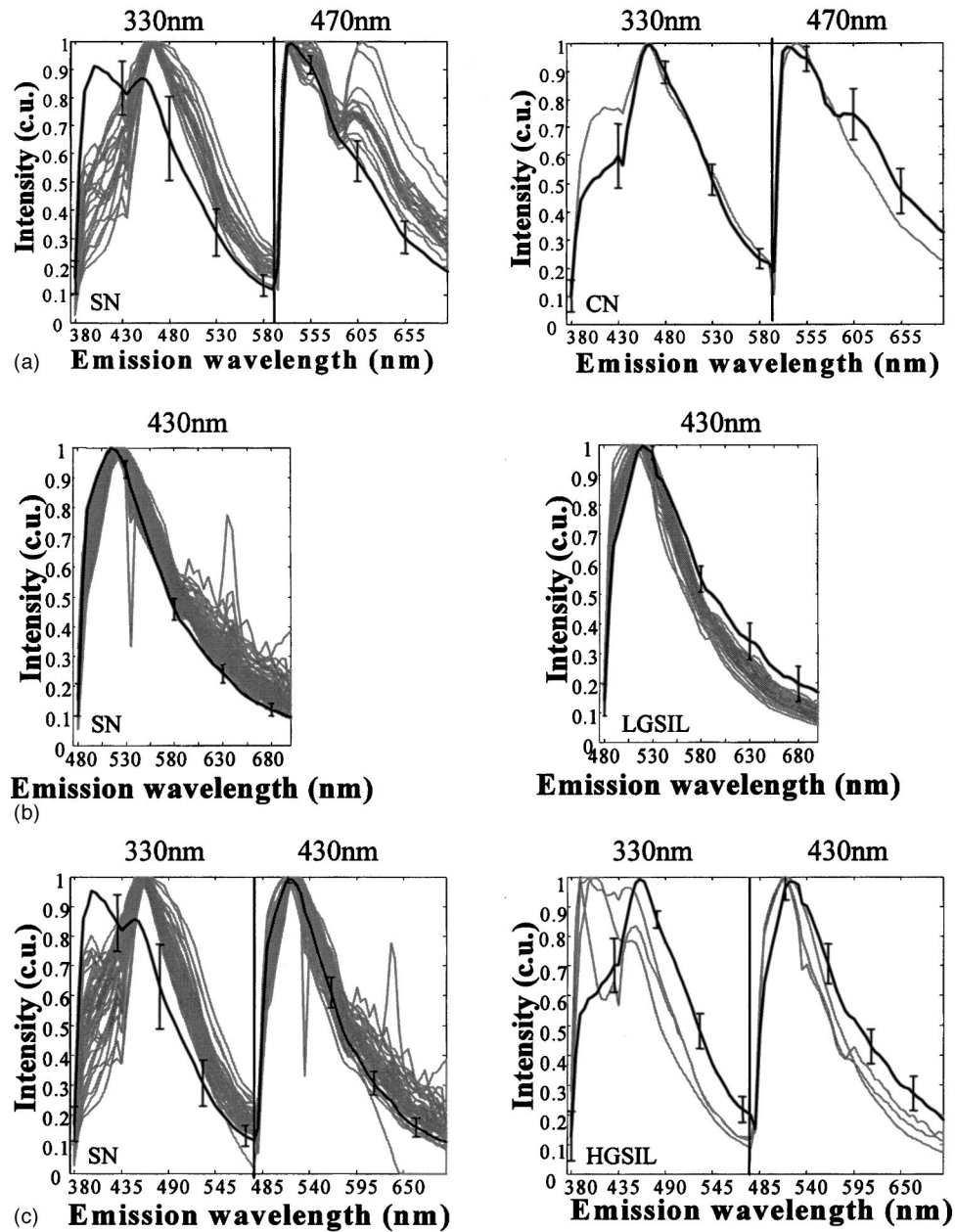
Diagnostic Pair	Sensitivity (%)	Specificity (%)	Feature Combination
SN versus CN	94	91	330 nm, 470 nm
SN versus LGSIL	55	63	430 nm
SN versus HGSIL	83	80	330 nm, 430 nm
CN versus LGSIL	90	83	330 nm
CN versus HGSIL	72	78	470 nm
Average	79	79	

In a previous study, *in vivo* fluorescence spectroscopy using 340, 380, and 460 nm excitation yielded a sensitivity of 79% and a specificity of 78% to discriminate HGSIL from all other cervical tissue types.<sup>7</sup> In a separate study, we were able to discriminate HGSIL from SN with a sensitivity and specificity of 71 and 79%, respectively, using fluorescence emission spectra from three excitation wavelengths.<sup>25</sup> However, we were able to discriminate HGSIL from CN with a very low sensitivity and specificity. We also performed a previous pairwise diagnostic-category comparison using reflectance spectroscopy alone, and found we could discriminate HGSIL from columnar normal tissue with a 72% sensitivity and an 83% specificity.<sup>26</sup> These initial findings would indicate the use of a strategy utilizing the respective strengths of fluorescence and reflectance spectroscopy. In this paper, we consider the additional diagnostic performance that can be obtained by combining fluorescence and reflectance spectra. We find fluorescence alone gives superior performance compared to reflectance alone and that the addition of reflectance spectra with fluorescence spectra provides a modest improvement in diagnostic performance using the empirical diagnostic algorithms considered here. In particular, reflectance spectroscopy provides good discrimination of CN and HGSIL tissues.

**Table 5** Diagnostic performance of each pairwise analysis when the available classification features are limited to four (fluorescence excitation wavelengths of 330, 360, 430, and 470 nm). The four classification features were selected on the basis of best overall diagnostic performance in all five pairwise analyses at an ESL=65%.

Diagnostic Pair	Sensitivity (%)	Specificity (%)	Combination
SN versus CN	94	91	330 nm, 470 nm
SN versus LGSIL	55	63	430 nm
SN versus HGSIL	83	80	330 nm, 430 nm
CN versus LGSIL	87	94	330 nm, 360 nm
CN versus HGSIL	72	78	470 nm
Averages	78	81	





**Fig. 6** Average spectra of correctly classified tissue measurements (heavy black line). Error bars represent one standard deviation. Individual spectra of the misclassified tissue measurements (thin gray lines) from each diagnostic class in the pairwise analysis when available features are limited to fluorescence excitation wavelengths of 330, 360, 430, and 470 nm. Results are shown for the following pairwise analyses: (a) SN versus CN, (b) SN versus LGSIL, (c) SN versus HGSIL, (d) CN versus LGSIL, and (e) CN versus HGSIL.

Several studies have investigated the diagnostic effectiveness of fluorescence and reflectance spectroscopy. Nordstrom et al.<sup>18</sup> investigated fluorescence and reflectance spectroscopy separately, and he reports that fluorescence spectroscopy yields higher classification performance in separating a pair of diagnostic classes except for the case of HGSIL versus metaplastic tissues, where reflectance spectroscopy performed better. In our study, metaplastic tissue was included in the SN category and yet we were able to achieve a high level of sensitivity and specificity when SN was classified from HGSIL using only fluorescence spectra. This could be attributed to the fact that a large number of fluorescence excitation

wavelengths were used in this study in contrast to the single fluorescence excitation wavelength (355 nm) used in Ref. 18. In fact, we have identified that fluorescence excitation wavelengths between 330 and 350 nm were significant in discriminating HGSIL from SN.

Our previous studies using fluorescence and reflectance spectroscopy individually indicate that stepwise diagnostic algorithms are required to determine the tissue type of an unknown sample based on its spectrum because of the large differences in optical properties of squamous and columnar cervical tissue.<sup>25,26</sup> The pairwise analysis presented here provides the foundation for this type of diagnostic algorithm. In a

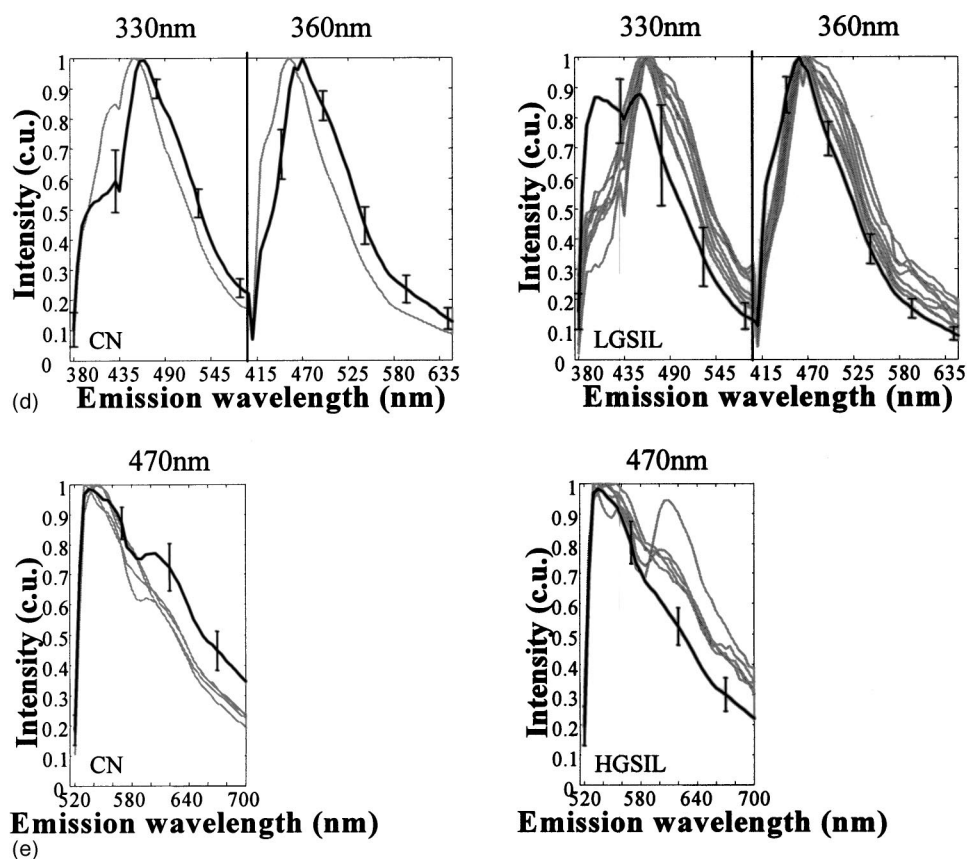


Fig. 6 (Continued.)

similar analysis, we examined fluorescence EEMs also for discrimination of all diagnostic categories. Hence, this information can be used toward the development of multistep classification algorithms to determine the tissue type of an unknown sample based on its reflectance and fluorescence spectra.

#### Acknowledgments

The authors gratefully acknowledge the contributions of the clinical research staff (Christina Amos, Joanne Baker, Glenda Dickerson, Kim Hagedorn, Patricia Trigo, and Christy Whitmore), nurse colposcopists (Judith Sandella, Alma Sbach, and Karen Rabel), and data managers Nan Earle and Trey Kell. Financial support from the National Cancer Institute (PO1-CA82710) is gratefully acknowledged.

#### References

- P. Pisani, D. M. Parkin, F. Bray, and J. Ferlay, "Estimate of the worldwide mortality from 25 cancers in 1990," *Int. J. Cancer* **83**, 18–29 (1999).
- T. C. Wright, R. J. Kurman, and A. Ferenczy, "Cervical intraepithelial neoplasia," in *Pathology of the Female Genital Tract*, A. Blaustein, Ed., Springer-Verlag, New York (1994).
- M. T. Fahey, L. Irwig, and P. Macaskill, "Meta-analysis of pap test accuracy," *Am. J. Epidemiol.* **141**(7), 680–689 (1995).
- M. F. Mitchell, "Accuracy of colposcopy," *Clin. Consult. Obstetrics Gynecol.* **6**(1), 70–73 (1994).
- I. J. Bigio, T. R. Loree, J. Mourant et al., "Spectroscopic diagnosis of bladder cancer with elastic light scattering," *Lasers Surg. Med.* **17**(4), 350–357 (1995).
- W. S. Glassman, C. H. Liu, G. C. Tang, S. Lubicz, and R. R. Alfano, "Ultraviolet excited fluorescence spectra from non-malignant and malignant tissues of the gynecological tract," *Lasers Life Sci.* **5**, 49–58 (1992).
- N. Ramanujam, M. Follen Mitchell, A. Mahadevan-Jansen, S. L. Thomsen, G. Staerckel, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, "Cervical precancer detection using a multivariate statistical algorithm based on laser induced fluorescence spectra at multiple excitation wavelengths," *Photochem. Photobiol.* **6**, 720–735 (1996).
- I. Georgakoudi, E. E. Sheets, M. G. 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*," *Am. J. Obstet. Gynecol.* **186**(3), 374–382 (2002).
- M. Coppleston, B. L. Reid, V. Skladnev, and J. C. Dalrymple, "An electronic approach to the detection of precancer and cancer of the uterine cervix: a preliminary evaluation of polar probe," *Int. J. Gynecol. Cancer* **4**, 79–93 (1994).
- R. M. Cothren, M. V. Sivak, J. Van Dam, R. E. Petras, M. Fitzmaurice, J. M. Crawford, J. Wu, J. F. Brennan, R. P. Rava, R. Manoharan, and M. S. Feld, "Detection of dysplasia at colonoscopy using laser-induced fluorescence: a blinded study," *Gastrointest Endosc.* **44**, 168–176 (1996).
- 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).
- R. R. Alfano, G. C. Tang, A. Pradham, W. Lam, D. S. J. Choy, and E. Opher, "Fluorescence spectra from cancerous and normal human breast and lung tissues," *IEEE J. Quantum Electron.* **QE23**(10), 1806–1811 (1987).
- T. Vo-Dinh, M. Panjehpour, B. F. Overholt, C. Farris, F. P. Buckley, and R. Sneed, "In vivo cancer diagnosis of the esophagus using differential normalized fluorescence (DNF) indices," *Lasers Surg. Med.* **16**, 41–47 (1995).

14. R. R. Alfano, B. B. Das, J. Cleary, R. Prudente, and E. J. Celmer, "Light sheds light on cancer—distinguishing malignant tumors from benign tissues and tumors," *Bull. N. Y. Acad. Med.* **67**, 143–150 (1991).
15. J. R. Mourant, J. Boyer, A. H. Hielscher, and I. J. Bigio, "Influence of the scattering phase function on light transport measurement in turbid media performed with small source detector separations," *Opt. Lett.* **21**(7), 546 (1996).
16. S. P. Lin, L. Wang, S. L. Jacques, and F. K. Tittel, "Measure of tissue optical properties by the use of oblique-incidence optical fiber reflectometry," *Appl. Opt.* **36**(1), 136–143 (1997).
17. M. G. Nichols, E. L. Hull, and T. Foster, "Design and testing of a white-light, steady-state diffuse reflectance spectrometer for determination of optical properties of highly scattering systems," *Appl. Opt.* **36**, 93 (1997).
18. R. J. Nordstrom, L. Burke, J. M. Niloff, and J. F. Myrtle, "Identification of cervical intraepithelial neoplasia (CIN) using UV-excited fluorescence and diffuse-reflectance tissue spectroscopy," *Lasers Surg. Med.* **29**(2), 118–127 (2001).
19. A. Dellas, H. Moch, E. Schulthesis, G. Feichter, A. C. Almendral, F. Gudat, and J. Torhorst, "Angiogenesis in cervical neoplasia: microvessel quantitation in precancerous lesions and invasive carcinomas with clinicopathological correlations," *Gynecol. Oncol.* **67**, 27–33 (1997).
20. C. H. Liu, G. C. Tang, A. Pradhan, W. L. Sha, and R. R. Alfano, "Effects of self-absorption by hemoglobins on the fluorescence spectra from normal and cancerous tissues," *Lasers Life Sci.* **3**, 167–176 (1990).
21. R. Drezek, M. Guillaud, T. Collier, I. Boiko, A. Malpica, C. MacAulay, M. Follen, and R. Richards-Kortum, "Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, DNA content, and chromatin texture," *J. Biomed. Opt.* **8**, 7–16 (2003).
22. T. Collier, D. Arifler, A. Malpica, M. Follen, and R. Richards-Kortum, "Determination of epithelial tissue scattering coefficient using confocal microscopy," *IEEE J. Sel. Top. Quantum Electron.* **9**, 307–313 (2003).
23. I. Pavlova, K. Sokolov, R. Drezek, A. Malpica, M. Follen, and R. Richards-Kortum, "Microanatomical and biochemical origins of normal and precancerous cervical autofluorescence using laser-scanning fluorescence confocal microscopy," *Photochem. Photobiol.* **77**, 550–555 (2003).
24. A. Zuluaga, U. Utzinger, A. Durkin, H. Fuchs, A. Gillenwater, R. Jacob, B. Kemp, J. Fan, and R. Richards-Kortum, "Fluorescence excitation emission matrices of human tissue: a system for *in vivo* measurement and data analysis," *Appl. Spectrosc.* **53**, 302–311 (1999).
25. S. K. Chang, M. Follen, A. Malpica, U. Utzinger, G. Staerckel, D. Cox, E. N. Atkinson, C. MacAulay, and R. Richards-Kortum, "Optimal excitation wavelengths for detection of cervical neoplasia," *IEEE Trans. Biomed. Eng.* **49**, 1102–1111 (2002).
26. Y. 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).
27. W. R. Dillon and M. Goldstein, *Multivariate Analysis: Methods and Applications*, Wiley, New York (1984).
28. U. Utzinger, V. Trujillo, E. N. Atkinson, M. F. Mitchell, S. B. Cantor, and R. Richards-Kortum, "Performance estimation of diagnostic tests for cervical precancer based on fluorescence spectroscopy: effects of tissue type, sample size, population and signal to noise ratio," *IEEE Trans. Biomed. Eng.* **46**, 1293–1303 (1999).