Optimal Excitation Wavelengths for Discrimination of Cervical Neoplasia

Sung K. Chang, Michele Follen, Anais Malpica, Urs Utzinger, Gregg Staerkel, Dennis Cox, E. Neely Atkinson, Calum MacAulay, and Rebecca Richards-Kortum*

***Abstract*—Fluorescence spectroscopy has shown promise for the *in vivo*, real-time detection of cervical neoplasia. However, selection of excitation wavelength has in the past been based on *in vitro* studies and the availability of light sources. The goal of this study was to determine optimal excitation wavelengths for *in vivo* detection of cervical neoplasia. Fluorescence excitation-emission matrices (EEMs) were measured *in vivo* from 351 sites in 146 patients. Data were analyzed in pairs of diagnostic classes to determine which combination of excitation wavelengths yields classification algorithms with the greatest sensitivity and specificity. We find that 330–340-, 350–380-, and 400–450-nm excitation yield the best performance. The sensitivity and specificity for discrimination of squamous normal tissue and high-grade squamous intraepithelial lesion (HGSIL) were 71% and 77% on cross validation using three excitation wavelengths. These results are comparable with those found in earlier *in vivo* studies; however, in this study we find that the proportion of samples which are HGSIL influences performance. Furthermore stratification of samples within low-grade squamous intraepithelial lesion and HGSIL also appears to influence diagnostic performance. Future diagnostic studies should be carried out at these excitation wavelengths in larger groups so that data can be stratified by diagnostic subcategory, age and menopausal status. Similarly, large studies should be done in screening populations.**

***Index Terms*—Algorithm, cancer diagnosis, fluorescence spectroscopy.**

NOMENCLATURE

SN	Squamous normal tissue.
CN	Columnar normal tissue.
HPV	Human papilloma virus.
CIN 1	Grade-1 cervical intraepithelial neoplasia.
CIN 2	Grade-2 cervical intraepithelial neoplasia.
CIN 3	Grade-3 cervical intraepithelial neoplasia.
CIS	Carcinoma *in situ*.
SIL	Squamous intraepithelial lesion.
LGSIL	Low-grade squamous intraepithelial lesion.
HGSIL	High-grade squamous intraepithelial lesion.
EEM	Excitation emission matrix.
ESL	Eigenvalue significance level.

Manuscript received November 13, 2001; revised April 5, 2002. This work was supported in part by the U.S. National Cancer Institute under Grant PO1-CA82710. *Asterisk indicates corresponding author.*

S. K. Chang is with the Electrical and Computer Engineering Department, ENS 8, University of Texas at Austin, Austin, TX 78712 USA.

M. Follen is with the Biomedical Engineering Center, UT M.D. Anderson Cancer Center, Houston, TX 77030 USA.

A. Malpica and G. Staerkel are with the Department of Pathology, UT M.D. Anderson Cancer Center, Houston, TX 77030 USA.

U. Utzinger is with the Biomedical Engineering Department, University of Arizona, Tucson, AZ 85721 USA.

D. Cox is with the Department of Statistics, Rice University, Houston, TX 77030 USA.

E. N. Atkinson is with the Department of Biomathematics, UT M.D. Anderson Cancer Center, Houston, TX 77030 USA.

C. MacAulay is with the Department of Cancer Imaging, British Columbia Cancer Agency, Vancouver, BC, Canada.

*R. Richards-Kortum is with the Electrical and Computer Engineering Department, University of Texas at Austin, Austin, TX 78712 USA (e-mail: kortum@mail.utexas.edu)

Publisher Item Identifier 10.1109/TBME.2002.803597.

I. INTRODUCTION

PAPANICOLOAU smear, in which a small sample of cells collected from the cervical epithelium are diagnosed under the microscope by an expert, is at present the most comprehensive means of screening and detecting cervical cancer. Although the Papanicoloau smear has been effective in reducing the mortality due to cervical cancer [1], [2], it is highly dependent on the skill of the investigator. In fact, the mean sensitivity and specificity in screening using Papanicoloau smear are 73% and 63%, respectively [3].

An abnormal Papanicoloau smear is followed by colposcopy, where a mounted magnifying lens is used to view the cervix. The sensitivity of colposcopy is excellent (96%) but the specificity is poor (48%) [4]. The cervix is covered by two types of epithelial tissue; the ectocervix has a stratified squamous epithelium and the endocervix has a columnar epithelium. The junction between these two is known as the transformation zone. Cervical precancers usually originate on the squamous side of the transformation zone and can be recognized based on their characteristic colposcopic appearance. Typically, it is relatively easy to discern colposcopically the mature squamous epithelium of the ectocervix from the columnar epithelium of the endocervix. However, within the transformation zone the colposcopic features of the squamous metaplastic/neoplastic and columnar epithelium are sometimes not very distinct and contribute to the low specificity of colposcopy.

Fluorescence spectroscopy has been investigated as an effective and noninvasive method for screening and detecting cervical cancer. Fluorescence spectroscopy of the tissue is affected by various optical interactions. Changes in index of refraction in the tissue and scatterers such as the cell nuclei causes scattering of light. Hemoglobin molecules are significant light absorbers at certain wavelengths. Light is also absorbed by chromophores, which then emit fluorescent light. Biological chromophores such as NADH and flavins are closely related to cellular metabolism. Scattering, absorption and fluorescence properties convey significant morphologic, cytologic and histo-pathologic information of the tissue under investigation.

A number of clinical trials have shown that fluorescence spectroscopy has promise for *in vivo*, real-time detection of cervical neoplasia [5]–[10]. Typically in these trials, fluorescence emission spectra are measured at one to three excitation wavelengths and diagnostic algorithms are developed retrospectively based on features of these spectra. Ramanujam reports a sensitivity and specificity of 92% and 90%, respectively, using one excitation wavelength at 337 nm for detecting CIN 1 and above [5]. In a separate study, Ramanujam reports sensitivity and specificity of 82% and 68%, respectively, when three excitation wavelengths at 337, 380, and 460 nm were used to differentiate HPV and above from normal tissue [6]. Burke also reports a sensitivity and specificity of 93% and 94%, respectively, at 337-nm excitation for discriminating CIN against benign (including normal, inflammation and metaplasia) [8]. Based on the fluorescence spectroscopy algorithm developed in [6], LifeSpex Inc. has developed a system to image fluorescence from cervical epithelium at multiple excitation emission wavelength pairs. Over 100 patients were evaluated with this device; the initial data from the study show that the device discriminates precancerous cervical lesion from normal tissue with a sensitivity and specificity of 98% and 95.4% [9]. Recently, a similar device, which incorporates the ability to measure both reflectance and fluorescence was used to measure the colposcopically visible cervical epithelium [10]. 136 patients were measured in the colposcopy setting, of which 111 patients were included for analysis. An algorithm was derived to recognize cervices with grade-2 cervical intraepithelial neoplasia (CIN 2) or greater from CIN 1 and normal tissue. Encouraging sensitivities and specificities were reported (97% and 70% respectively). However in both studies [9], [10], algorithm results are reported from the same data set used to derive the algorithm; thus, estimates of sensitivity and specificity may be high due to over-training bias.

An important limitation of past studies is that the selection of excitation wavelength was based either on availability of a light source [8] or on the basis of small, *in vitro* studies surveying many different excitation wavelengths [11]. It is well known that the optical properties of epithelial tissue differ *in vitro*, implying that different excitation wavelengths may be optimal for *in vivo* studies [12], [13].

Recently, several groups have developed spectroscopic systems which enable measurement of fluorescence emission spectra at many excitation wavelengths *in vivo* [14]–[16]. These emission spectra can be assembled into an EEM, which contains the fluorescence intensity as a function of both excitation and emission wavelength. These systems provide a convenient way to characterize the autofluorescence properties of epithelial tissue over the entire UV-visible spectrum. While these research level systems enable clinical trials to determine the optimal excitation wavelengths for diagnostic purposes, they are not suited for office-based diagnosis. Cost-effective devices, using a smaller number of optimized excitation wavelengths will be required to allow the technology to enter wide scale clinical practice [17].

The goal of this study was to carry out *in vivo* measurements of fluorescence EEMs and analyze these data to determine the optimal excitation wavelengths for diagnosis of cervical neoplasia and to estimate the sensitivity and specificity at this combination of excitation wavelengths.

II. Methods

A. Materials

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. Eligible patients included those over the age of 18 who were not pregnant and who were referred to the Colposcopy Clinic at the UT M.D. Anderson Cancer Center with an abnormal Papanicolaou smear. After signing informed consent, all patients underwent a demographic interview, risk-factor questionnaire, complete history, and physical exam and pan-colposcopy of the vulva, vagina, and cervix. Initially, each patient underwent a urine pregnancy test, chlamydia and gonorrhea cultures and a Papanicolaou smear. Additionally, patients underwent Virapap testing (DiGene, Bethesda, MD) as well as HPV, DNA, and mRNA sampling. Each patient had blood drawn for follicle-stimulating hormone (FSH), estradiol, and progesterone levels. The last menstrual period and menstrual history were asked of each patient.

During colposcopy, two colposcopically normal sites and one colposcopically abnormal site were chosen by the physician (MF) or nurse colposcopist and fluorescence EEMs were measured from these three sites. It was noted whether these sites corresponded to squamous or columnar epithelium or the transformation zone.

Following fluorescence measurement, each site was biopsied and submitted for histopathologic diagnosis. Each Papanicolaou smear was read by the cyto-pathologist assigned to the case that day and was subsequently reviewed by the study cyto-pathologist (GS). Discrepant cases were reviewed a third time for consensus diagnosis by the study cytologist (GS). Each biopsy was read by the pathologist assigned to the case that day and was subsequently reviewed by the study histopathologist (AM). Again, discrepant cases were reviewed a third time for consensus diagnosis by the study histopathologist (AM). Standard diagnostic criteria were used [18] and consensus diagnostic categories included: normal SN, normal CN, HPV, CIN 1, CIN 2, and CIN 3. For initial analysis, HPV infection and CIN 1 were grouped together as LGSIL and CIN 2 and CIN 3 were grouped together as HGSIL.

B. Instrumentation

The spectroscopic system used to measure fluorescence EEMs has been described in detail previously [19], [20]. Briefly, the system measures fluorescence emission spectra at 16 excitation wavelengths, ranging from 330 nm to 480 nm in 10-nm increments with a spectral resolution of 5 nm. The system incorporates a fiber-optic probe, a Xenon arc lamp coupled to a monochromator to provide excitation light and a polychromator and thermo-electrically cooled charge-coupled device camera to record fluorescence intensity as a function of emission wavelength. The fiber-optic probe consists of 25 excitation fibers and 12 collection fibers, arranged randomly on a 2-mm-diameter quartz fiber at the tip.

C. Measurements

As a negative control, a background EEM was obtained with the probe immersed in a nonfluorescent bottle filled with distilled water at the beginning of each day. Then a fluorescence EEM was measured with the probe placed on the surface of a quartz cuvette containing a solution of Rhodamine 610 (Exciton, Dayton, OH) dissolved in ethylene glycol (2 mg/mL) at the beginning of each patient measurement.

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; in the visible an NIST traceable calibrated tungsten ribbon filament lamp was used and in the UV a deuterium lamp was used (550C and 45D, Optronic Laboratories Inc, Orlando, FL). Correction factors were derived from these spectra. Dark current subtracted EEMs from patients were then corrected for the nonuniform spectral response of the detection system. Variations in the intensity of the fluorescence excitation light source at different excitation wavelengths were corrected using measurements of the intensity at each excitation wavelength at the probe tip made using a calibrated photodiode (818-UV, Newport Research Corp.).

Before the probe was used it was disinfected with Metricide (Meterx Research Corp.) for 20 min. The probe was then rinsed with water and dried with sterile gauze. The disinfected probe was guided into the vagina and its tip positioned flush with the cervical epithelium. Then fluorescence EEMs were measured from the three cervical sites. Acetic acid, which enhances the fluorescence and the reflectance differences between normal and dysplastic tissue [21], was applied to the cervical epithelium prior to the placement of the probe. Measurement of each EEM required approximately 2 min.

D. Data Analysis

All spectra were reviewed by three investigators blinded to the pathologic results (SKC, UU, and RRK) prior to analysis. Spectra which indicated evidence of instrument error or probe slippage were discarded from further analysis.

Fluorescence data were analyzed to determine which excitation wavelengths contained the most diagnostically useful information and to estimate the performance of diagnostic algorithms based on this information. Initially, algorithms that discriminated between all pair-wise combinations of diagnostic categories were explored (Table I). In comparing all pairs of diagnostic classes, we can determine which categories differ spectroscopically and assess where these differences are greatest. This information can then be used to develop multi-step algorithms [6] to determine the tissue type of an unknown sample based on its fluorescence spectrum. For this purpose, we developed an algorithm based on multivariate discriminant techniques which selects a subset of spectra at excitation wavelengths that perform best from all possible combinations of emission spectra at all excitation wavelengths [19]. The algorithm, described in detail below, consists of the following major steps: 1) data preprocessing to reduce interpatient variations; 2) data reduction to reduce the dimensionality of the data set; 3) classification to classify the two given classes with maximum diagnostic performance and minimal likelihood of over-training in a training set; and 4) evaluation of these algorithms using the technique of cross validation.

TABLE I
PAIRS OF HISTOPATHOLOGIC CLASSES FOR WHICH DIAGNOSTIC ALGORITHMS WERE DEVELOPED

Class 1	Class 2
Squamous normal	Columnar normal
Squamous normal	LG-SIL
Columnar normal	LG-SIL
Squamous normal	HG-SIL
Columnar normal	HG-SIL

Fluorescence data from a single measurement site is represented as a matrix containing calibrated fluorescence intensity as a function of excitation and emission wavelength. Columns of this matrix correspond to emission spectra at a particular excitation wavelength; rows of this matrix correspond to excitation spectra at a particular emission wavelength. Each excitation spectrum contains 16 intensity measurements ranging from 330 to 480 nm in 10–nm increments; each emission spectrum contains between 50 and 130 intensity measurements ranging between 380 and 910 nm in 5-nm increments, depending on excitation wavelength. Finally, emission spectra were truncated at emission wavelength of 700 nm to eliminate the highly variable background due to room light present above 700 nm. Most multivariate data analysis techniques require vector input, so prior to analysis the column vectors containing the emission spectra at excitation wavelengths selected for evaluation were cropped to discard the noisy tails and then were concatenated into a single vector.

Our previous work illustrates that spectra of the cervix obtained *in vivo* show large patient-to-patient variations in intensity that can be greater than the differences between histopathologic categories [22]. Therefore, we utilized a preprocessing method to reduce the interpatient variations, while preserving intercategory differences: each emission spectrum in the concatenated vector was normalized to its respective maximum intensity.

Principal component analysis was performed on the entire dataset for dimensionality reduction. First, an input matrix was created for each excitation wavelength combination by placing the concatenated spectrum vector from each sample in rows. The eigenvectors of the corresponding co-variance matrix were then calculated, yielding the principal components. We used eigenvectors accounting for 65%, 75%, 85%, and 95% of the total variance to investigate the effect of ESL (eigenvector significance level) on the algorithm performance. The ESL represents the fraction of the total variance of the dataset accounted for by the linear combination of the first n eigenvectors. Principal component scores associated with these eigenvectors were calculated for each sample [23].

Classification functions were then formed to assign a sample to one of the two given classes. The classification was based on the Mahalanobis distance, which is a multivariate measure of the separation of a data point from the mean of a dataset in n-dimensional space [22]. The multivariate distance between the sample to be classified 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. From the available pool of eigenvectors at ESLs of 65%, 75%, 85%, and 95%, the single principal component score yielding the best initial performance was identified and then the principal component score that improved this performance most was selected. This process was repeated until performance was no longer improved by the addition of principal component scores, or all the available scores were selected.

The sensitivity and specificity of the algorithm at each excitation wavelength combination were then evaluated relative to diagnosis based on histopathology. Overall diagnostic performance was evaluated as the sum of the sensitivity and the specificity, thus minimizing the number of misclassifications. The risk of overtraining was assessed for each of the excitation wavelength combinations by comparing the training set performance with the performance of an algorithm developed from the same dataset after the diagnosis corresponding to each sample had been randomized. The number of samples in each class was preserved during diagnosis randomization. This provides a dataset with the same variance structure as the original dataset but where the diagnostic performance is not expected to exceed that of chance. Diagnostic algorithms were then developed based on the randomized diagnoses. Random diagnoses were assigned 50 times for each wavelength combination and the average of the sensitivities and the specificities of the 50 cases were calculated. Ideally for completely normally distributed data, the sum of the sensitivity and specificity should be one for randomized diagnosis at all levels of training significance. However, if overtraining occurs, this sum will be greater than one. Combinations of emission spectra from one up to four excitation wavelengths were considered. Limiting the device to four wavelengths allows for construction of a reasonably cost-effective clinical spectroscopy system [24]. To identify the optimal combination of excitation wavelengths, we evaluated all possible combinations of up to four wavelengths chosen from the 16 possible excitation wavelengths. This equated to 16 combinations of one, 120 combinations of two, 560 combinations of three, and 1820 combinations of four excitation wavelengths, for a total of 2516 combinations. The top-25 wavelength combinations were then ranked based in order of the increase in performance between the training set performance with the correct histopathologic diagnoses and the training set performance with random assignment of diagnosis. This method allows the top wavelength combinations to be ranked in order of their robustness, or lack of propensity to overtrain.

These estimates of algorithm performance are biased since they are based on the training set used to develop the algorithm. An unbiased performance estimate must be made to assess the true potential of each of the top-25 wavelength combinations. The effects of overtraining in performance estimation can be minimized by using separate training and validation sets, or by using the method of cross validation [25]. In the cross-validation method, a single data from the whole dataset is temporarily removed from the training dataset and the classification algorithm is developed using the remaining dataset as training set. The new classification algorithm is applied to the held out data. Each sample in the dataset was used as the test data in turn and the sensitivity and specificity were calculated by comparing the classification result of each sample with histopathologic diagnosis. Among the top-25 wavelength combinations, results from top-ten combinations based on the sum of the cross-validated sensitivity and cross-validated specificity are presented in this paper.

TABLE II
SUMMARY OF HISTOPATHOLOGIC SAMPLES USED FOR ANALYSIS. CIN 3+ REFERS TO CIN 3 AND CIS

Histopathologic category	Number of samples	Composition	
Squamous normal	233		
Columnar normal	23		
LG-SIL	64	HPV: 46	CIN 1:18
HG-SIL	31	CIN 2: 12	CIN 3+:19

III. RESULTS

A total of 373 EEMs from 147 patients were analyzed in this study. Of the 373 EEMs reviewed, 22 were identified as defective for analysis due to instrument error (ten sites) and probe movement (12 sites) and were discarded. Of the 351 remaining EEMs, 233 sites were normal squamous sites, 23 were columnar sites, 64 were LGSILs, and 31 were HGSILs. Of the 64 LGSIL sites, 46 were HPV, and 18 were CIN 1. Of the 31 HGSIL sites, 12 were CIN 2, 11 were CIN 3, and eight were CIS. Of the 233 squamous normal sites 107 sites had inflammation and/or metaplasia. 40 showed inflammation only and 14 showed metaplasia only. The rest 53 sites had both inflammation and metaplasia. Of the 23 columnar normal sites, 12 showed inflammation only and 11 showed both inflammation and metaplasia. None of the columnar sites had just metaplasia. The composition of data used for analysis is summarized in Table II.

Fig. 1 shows typical fluorescence EEMs from different sites in the same patient, including a normal squamous site, a normal columnar site and a site with HGSIL. The fluorescence EEMs are plotted as topographical maps, with excitation wavelength on the ordinate and emission wavelength on the abscissa. Contour lines connect points of equal fluorescence intensity. Several excitation-emission maxima are present; the peak at 350-nm excitation, 450-nm emission is consistent with emission of the co-factor NADH as well as collagen crosslinks. A less apparent shoulder at 370-nm excitation, 525-nm emission is consistent with emission of the co-factor FAD. The peak at 450-nm excitation, 525-nm emission is consistent with the co-factor FAD as well as structural protein fluorescence. In addition, fluorescence of endogenous porphyrins is present in the EEM of the HGSIL, with excitation maxima at 410 nm and emission maxima at 630 and 690 nm. Tissue vascularity can influence fluorescence spectra, when hemoglobin absorbs fluorescent light at 420, 540, and 580 nm, producing valleys in the EEMs parallel to the excitation and emission wavelength axes as seen in all three EEMs.

We illustrate the effect of increasing the number of excitation wavelengths on diagnostic performance initially for the discrimination between SN and HGSIL. Table III shows the cross-validated sensitivity and specificity for algorithms based on top-ten performing single-excitation wavelengths. An ESL of 75% was used; however, similar results were obtained at

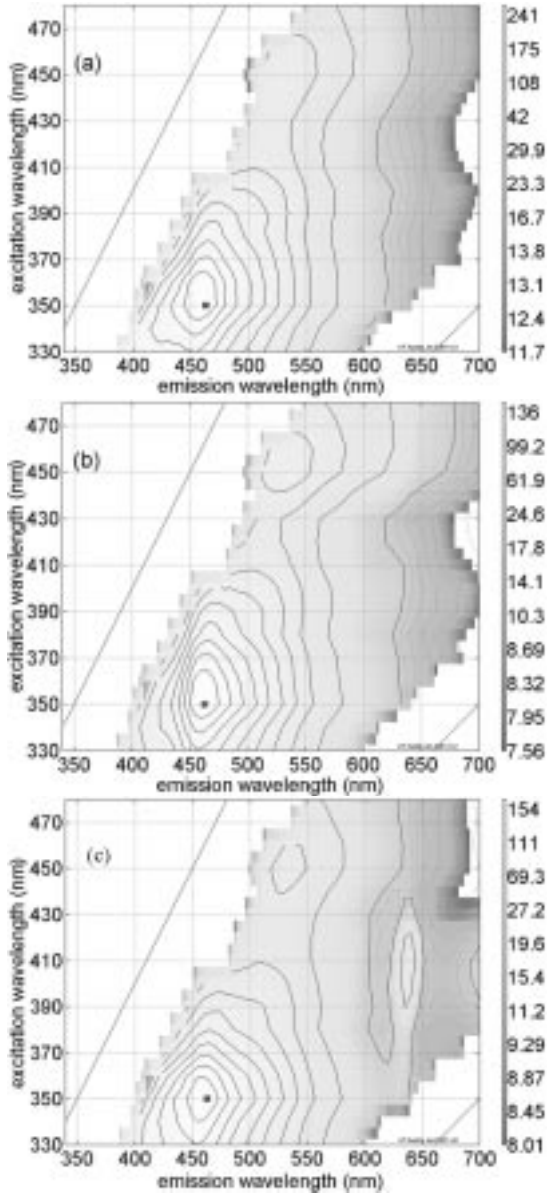

Fig. 1. Typical fluorescence EEM of (a) a normal squamous site, (b) a normal columnar site, and (c) HGSIL from the same patient. Excitation wavelength is shown on the ordinate and emission wavelength is shown on the abscissa. Contour lines connect points of equal fluorescence intensity.

TABLE III
CROSS-VALIDATED PERFORMANCE OF TOP-TEN PERFORMING SINGLE-EXCITATION WAVELENGTHS FOR DISCRIMINATION BETWEEN SN AND HGSIL AT AN ESL OF 75%. WAVELENGTHS ARE LISTED IN DESCENDING ORDER OF THE SUM OF THE CROSS-VALIDATED SENSITIVITY AND SPECIFICITY

Excitation Wavelength(nm)	Cross-validated Sensitivity	Cross-validated Specificity
390	0.74	0.74
350	0.71	0.73
400	0.68	0.76
440	0.71	0.72
340	0.71	0.71
380	0.68	0.68
330	0.65	0.71
430	0.65	0.70
370	0.65	0.70
420	0.55	0.78
MEAN	0.67	0.72

TABLE IV
TOP-TEN PERFORMING COMBINATIONS OF TWO EXCITATION WAVELENGTHS FOR DISCRIMINATION BETWEEN SN AND HGSIL AT AN ESL OF 75%. WAVELENGTH COMBINATIONS ARE LISTED IN DESCENDING ORDER OF THE SUM OF THE CROSS-VALIDATED SENSITIVITY AND SPECIFICITY

Excitation Wavelength 1 (nm)	Excitation Wavelength 2 (nm)	Cross-validated Sensitivity	Cross-validated Specificity
350	430	0.68	0.86
430	460	0.68	0.82
330	410	0.68	0.79
330	400	0.74	0.72
330	420	0.74	0.72
330	450	0.65	0.82
330	430	0.74	0.72
360	410	0.65	0.81
420	460	0.74	0.71
430	450	0.65	0.80
MEAN		0.70	0.78

TABLE V
TOP-TEN PERFORMING COMBINATIONS OF THREE EXCITATION WAVELENGTHS FOR DISCRIMINATION BETWEEN SN AND HGSIL AT AN ESL OF 75%. WAVELENGTH COMBINATIONS ARE LISTED IN DESCENDING ORDER OF THE SUM OF THE CROSS-VALIDATED SENSITIVITY AND SPECIFICITY

Excitation Wavelength 1 (nm)	Excitation Wavelength 2 (nm)	Excitation Wavelength 3 (nm)	Cross-validated Sensitivity	Cross-validated Specificity
420	430	460	0.71	0.79
330	460	470	0.65	0.85
330	420	470	0.65	0.85
340	380	420	0.68	0.79
330	340	420	0.68	0.79
340	420	470	0.74	0.72
350	430	440	0.74	0.72
350	440	470	0.74	0.72
340	420	460	0.74	0.72
330	400	470	0.74	0.71
MEAN			0.71	0.77

all ESLs. The excitation wavelengths are listed in descending order of the sum of the cross-validated sensitivity and specificity. Tables IV and V show the cross-validated sensitivity and specificity for algorithms based on the top-ten performing combinations of two and three wavelength combinations, respectively. Again, an ESL of 75% was used, but similar results were obtained at all ESLs. The excitation wavelength combinations in each table are listed in descending order of the sum of the cross-validated sensitivity and specificity. Table III shows that, while the four top excitation wavelengths have similar performance, sensitivity and specificity drop for the remaining single-excitation wavelengths. However, most of the top-ten combinations of two excitation wavelengths (Table IV) and all of the top-ten combinations of three excitation wavelengths (Table V) have similar performance. Interestingly, the four excitation wavelengths which give the best performance in Table III appear in many of the top-ten combinations identified in Tables IV and V.

Fig. 2 illustrates the effect of increasing the number of excitation wavelengths for algorithms that classify the important histopathological categories in the cervix. Fig. 2 shows the mean and the standard deviation of the cross-validated sensitivities

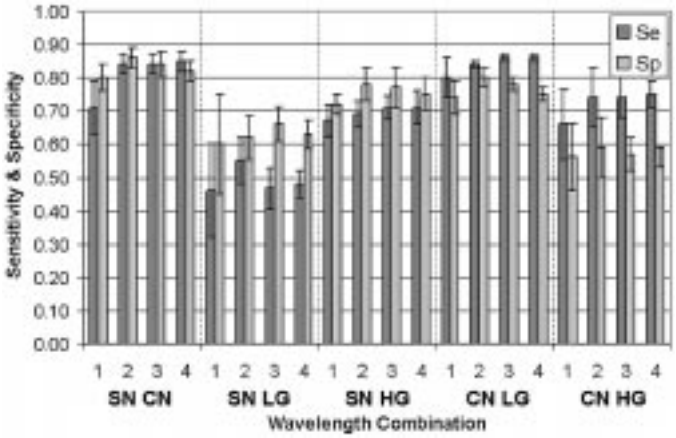

Fig. 2. Mean sensitivity and specificity for each single-excitation wavelength and the top-ten performing combinations of two, three, and four excitation wavelengths for distinguishing pairwise combinations of histopathologic categories at ESL of 75%. Dark gray and light gray plots are sensitivity and specificity, respectively. Error bars show plus and minus one standard deviation.

and specificities from the top-ten excitation wavelength combinations for each pair of classes tested at an ESL of 75%. In general, the performance of the algorithms developed for a pair of classes improves when the number of excitation wavelengths used in the combination is increased from one to two; however, adding a third or a fourth excitation wavelength does not generally improve diagnostic performance. The algorithm performs well for separating SN versus CN, SN versus HGSIL, and CN versus LGSIL. In contrast, lower performance is observed for separating SN versus LGSIL and CN versus HGSIL.

Fig. 3 illustrates the performance of ensemble classifiers based on the top-25 performing combinations of three and four excitation wavelengths at ESLs of 65%, 75%, 85%, and 95%. Fig. 3(a) shows the sensitivity (left) and specificity (right) for ensemble classifiers to separate SN versus CN. Performance of three types of ensembles are shown: in the first the sample was classified as CN if a majority of the 25 individual top performing excitation wavelength combinations indicated the sample was CN. Similarly, results from the classifiers which required 80% or 100% of the top performing 25 excitation wavelength combinations to classify the sample as CN are also shown. Again, as the number of excitation wavelengths is increased, little change in performance is seen. As the percentage of individual classifiers required to identify a sample as CN is increased from a majority to unanimous, little decrease is seen in the specificity, indicating that most excitation wavelength combinations yield the same classification. Finally, it is interesting to note that the sensitivity drops slightly at an ESL of 95% when all 25 individual classifiers are required to agree. This likely reflects an overtraining bias at this ESL.

Fig. 3(b) shows the performance of ensemble classifiers based on the top 25 performing combinations of three and four excitation wavelengths to separate SN from HGSILs. As the number of individual classifiers required to identify a sample as HGSIL increased from a majority to unanimous, sensitivity dropped while specificity remained fairly constant. Again, as the number of excitation wavelengths was increased from three to four, no significant increase in performance were observed. We further explored the sensitivity of detection for CIN 2 and CIN 3, the two subcategories which make up HGSIL, based onthe algorithm to discriminate HGSIL from SN. Fig. 3(c) shows the sensitivity of ensemble classifiers for CIN 2 (left) and CIN 3 (right). Sensitivity is much higher for CIN 3 than for CIN 2 for the three ensemble classifiers. This is consistent with the ability of the pathologist to identity CIN 3; interobserver agreement is much higher for diagnosis of CIN 3 than for CIN 2 [26].

Similarly, Fig. 3(d) shows the performance of ensemble classifiers which separate CN from LGSIL. In this case, neither sensitivity nor specificity varied substantially as the fraction of individual classifiers required to classify a sample as LGSIL was increased from 50% to 100%. Again, the diagnostic performance did not increase as the number of excitation wavelengths was increased from three to four. To explore whether performance varied for the two subcategories which make up LGSIL (HPV infection and CIN 1), we separately examined sensitivity for these two subcategories. Results are shown in Fig. 3(e). The sensitivity for discriminating CN and HPV (left) was higher than that for CIN 1 (right), indicating that CN spectra more closely resemble those of tissue with CIN 1 than with HPV infection. Although the plots are not shown, the ensemble classifiers for discriminating LGSIL from SN shows similar results where addition of an excitation wavelength to four wavelength combinationsdoes not increase performance. Investigation of the subcategories for LGSIL shows that difficulty of separating HPV from SN limits the performance.

In order to select the wavelengths most promising for discrimination between histopathologic classes of cervical tissue, we explored those excitation wavelengths which resulted in highest diagnostic performance. We created histograms indicating the frequency of occurrence of each excitation wavelength in the top-ten performing combinations of two excitation wavelengths at an ESL of 75% for all pairwise discriminations. Results are shown in Fig. 4. Fig. 4(a) shows the histogram for discrimination between SN and CN and indicates that 330–340 nm occurs most frequently and wavelengths of 410 and 420 nm also occur frequently. Fig. 4(b) shows the histogram for discrimination between SN and LGSIL. Again, 330–350-nm excitation occurs frequently, as well as 400–450-nm excitation. Fig. 4(c) and (d) shows similar wavelengths are useful for discriminating between SN and HGSIL as well as between CN and LGSIL. Fig. 4(e) shows that excitation wavelengths between 370–400 nm are most useful for discriminating CN and HGSIL.

IV. Discussion and Conclusion

In summary, fluorescence EEMs of SN, CN, LGSIL and HGSIL show characteristic differences that can be used to discriminate among histopathologic classes. In this study we measured fluorescence emission spectra at 16 different excitation wavelengths in an attempt to identify those excitation wavelength combinations which provide the most useful diagnostic information. We find that, to discriminate between pairs of histologic classifications, performance increases when data from two excitation wavelengths are combined. However, adding data from a third or a fourth excitation wavelength does not increase diagnostic performance. Best discrimination

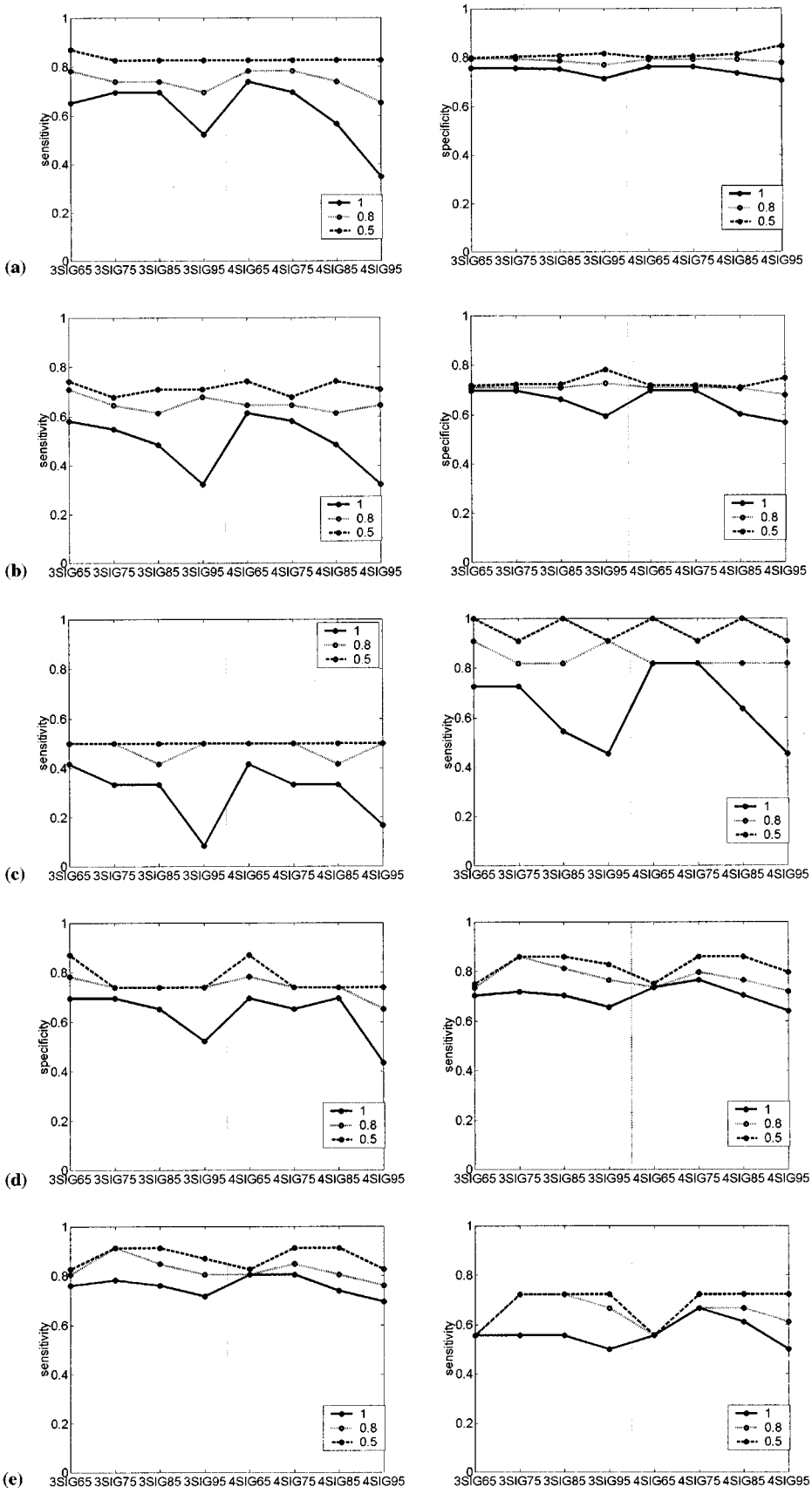

Fig. 3. Sensitivity and specificity for ensemble classifiers made up of the top 25 performing combinations of three and four excitation wavelengths at ESLs of 65%, 75%, 85%, and 95%. x-axis label shows the different combinations. For example, "3SIG65" denotes 3 wavelength combinations at 65% ESL. Results are shown for ensemble classifiers requiring agreement among 50% (dashed line), 80% (dotted line), and 100% (solid line) of the individual algorithms. Results are shown for algorithms which discriminate (a) SN against CN, (b) SN against HGSIL, (c) SN against CIN 2 and CIN 3, (d) CN against LGSIL, and (e) CN against HPV and CIN 1.

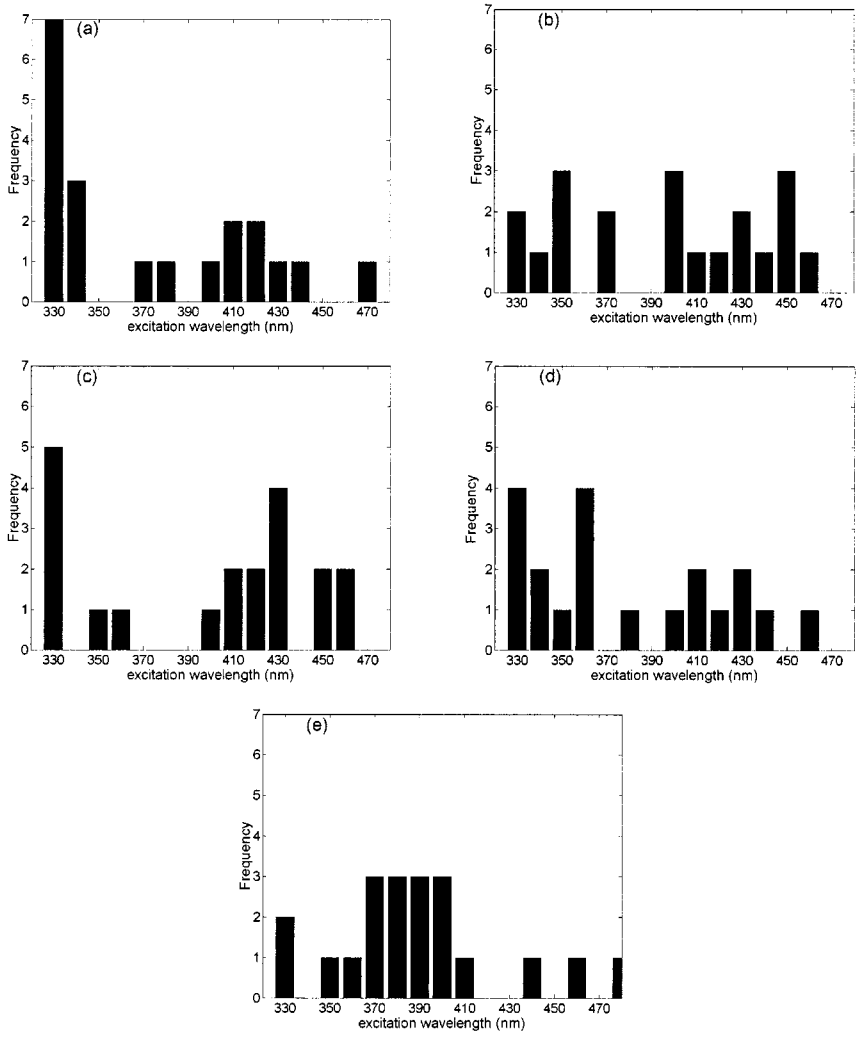

Fig. 4. Histogram indicating the frequency of occurrence of each excitation wavelength in the top ten performing combinations of two excitation wavelengths. ESL was 75%. Results are shown for pairwise discrimination between (a) SN and CN, (b) SN and LGSIL, (c) SN, and HGSIL, (d) CN and LGSIL, and (e) CN and HGSIL.

was found between spectra from SN and CN, between SN and HGSIL and between CN and LGSIL, with sensitivity and specificity averaging between 70% and 80%.

This pairwise analysis is useful to identify those excitation wavelengths that contribute to diagnostic information. However, in the clinical setting all of these tissue types are potentially encountered in each patient and for an algorithm to be clinically useful it must discriminate between all of them. The histograms in Fig. 4 show that approximately two excitation wavelengths appear optimal for pairwise classification. In Fig. 4(a), 330 and 420 nm appear frequently in classifying SN from CN. Subsequent figures in Fig. 4 show that 350 nm, 450 nm occur frequently in discriminating LGSIL from SN; 330 and 430 nm occur frequently in discriminating HGSIL from SN; 330 and 360 nm occur frequently in discriminating LGSIL from CN. Wavelengths between 370 and 400 nm appear frequently in discriminating HGSIL from CN. Based on these observations, three ranges of wavelengths appear important in order to develop a diagnostic algorithm that separates all of these tissue categories. These ranges include 330–340-, 350–380-, and 400–450-nm excitation.

In our previous work, using a system that could measure at only three excitation wavelengths, we examined fluorescence at 337-, 380- and 460-nm excitation [6]. This was based on an analysis of EEMs measured *in vitro* from tissue biopsies. Despite the limitations of working with tissue *in vitro*, these wavelength regions correspond reasonably well with those identified from this study.

While the performance results are encouraging for separating SN from CN, SN from HGSIL and CN from LGSIL, lower performance was observed for discriminating SN from LGSIL and CN from HGSIL (Fig. 2). Other strategies may be necessary to complement fluorescence spectroscopy in the separation of these tissue classes. Examples include other data reduction and normalization methods for fluorescence spectroscopy or the use of reflectance spectroscopy. We are exploring these alternatives.

We previously reported results from a study of 95 patients measured *in vivo* at 337-, 380-, and 460-nm excitation [6] and found a sensitivity and specificity of 82% and 68% for discriminating between normal tissues (squamous and columnar) and SILs. We identified that excitation wavelengths of 350, 370, and 420 nm are optimal for discriminating HGSIL against SN and

HGSIL against CN. Using the three wavelengths, an algorithm for classifying HGSIL against SN, CN, and LGSIL combined was developed. The results show a sensitivity and specificity of 84% and 66%, respectively, at an ESL of 85%. The results are comparable with those reported in [6]. However, there are some interesting differences in the composition of the patient histopathologic samples between the two studies. In our earlier study [6], the proportion of SILs which were diagnosed as HGSIL was 60%. In this study, the proportion of SILs diagnosed as HGSIL was only 33%. Proportions of disease within LGSIL and HGSIL also differed between the two studies. In our earlier study, 68% of the LGSILs were CIN 1, while in this study it was only 28%. In our earlier study, 54% of the HGSILs were CIN 3, while in this study, it was only 35%. Fig. 3(c) shows that there is greater separation between the SNs and CIN 3 lesions than between SNs and CIN 2 lesions. This suggests that the performance of fluorescence-based algorithms will be significantly influenced by the composition of the patient samples. These differences in fluorescence are consistent with the biological continuum which exists between normal tissue, LGSIL, and HGSIL. Further, they are consistent with studies which examine the intraobserver and interobserver agreement between pathologists for discrimination between various histopathologic categories, showing that there is greater agreement for CIN 3 lesions [26].

Future studies should use the excitation wavelengths identified in this study prospectively, stratifying results by age, menstrual status and histopathologic grade, since these variables appear to affect fluorescence characteristics of the cervix. Further, these combinations should be evaluated in the screening setting, with cohorts of patients that have a history of a normal Pap smear, similarly stratified by age and menstrual status. The results from the pair-wise analysis presented here and in the future work will provide the foundation for the development of a multi-step algorithm which will classify an unknown sample into various tissue types.

ACKNOWLEDGMENT

The authors would like to acknowledge the contributions of nurse colposcopists J. Sandella, A. Sbach, and K. Rabel.

REFERENCES

[1] L. G. Koss, "The Papanicoloau test for cervical cancer detection: A triumph and a tragedy," *JAMA*, vol. 261, pp. 737–43, 1989.

[2] R. J. Kurman, R. E. Herison, A. L. Herbst, K. L. Noller, and M. H. Schiffman, "Interim guidelines for management of abnormal cervical cytology," *JAMA*, vol. 271, pp. 1866–9, 1994.

[3] M. F. Mitchell, S. B. Cantor, C. Brookner, U. Utzinger, D. Schottenfeld, and R. Richards-Kortum, "Screening for squamous intraepithelial lesions with fluorescence spectroscopy," *Obstet. Gynecol.*, vol. 94, pp. 889–96, 1999.

[4] M. F. Mitchell, D. Schottenfeld, G. Tortolero-Luna, S. B. Cantor, and R. Richards-Kortum, "Colposcopy for the diagnosis of squamous intraepithelial lesions: A meta-analysis," *Obstet. Gynecol.*, vol. 91, pp. 626–631, 1998.

[5] N. Ramanujam, M. F. Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. C. Wright, N. Atkinson, and R. R. Richards-Kortum, "*In vivo* diagnosis of cervical intraepithelial neoplasia using 337-nm-excited laser-induced fluorescence," in *Proc. Nat. Acad. Sci. USA*, vol. 91, 1994, pp. 10 193–97.

[6] N. Ramanujam, M. F. Mitchell, A. Mahadevan-Jansen, S. L. Thomsen, G. Staerkel, 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.*, vol. 6, pp. 720–35, 1996.

[7] 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.*, vol. 19, pp. 63–74, 1996.

[8] L. Burke, M. Modell, J. Niloff, M. Kobelin, G. Abu-Jawdeh, and A. Zelenchuk, "Identification of squamous intraepithelial lesions: Fluorescence of cervical tissue during colposcopy," *J. Lower Genital Tract Disease*, vol. 3, pp. 159–62, 1999.

[9] S. Furlong, presented at the Industrial Roundtable, OSA Spring Topical Meeting on Biomedical Optical Spectroscopy and Diagnostics, Orlando, FL, 1998.

[10] D. G. Ferris, R. A. Lawhead, E. D. Dickman, N. Holtzapple, J. A. Miller, S. Grogan, S. Bambot, A. Agrawal, and M. L. Faupel, "Multimodal hyperspectral imaging for the noninvasive diagnosis of cervical neoplasia," *J. Lower Genital Tract Disease*, vol. 5, pp. 65–72, 2001.

[11] A. Mahadevan, M. F. Mitchell, E. Silva, S. Thomsen, and R. Richards-Kortum, "A study of the fluorescence properties of normal and neoplastic human cervical tissue," *Lasers Surg. Med.*, vol. 13, pp. 647–55, 1993.

[12] K. Shomacker, J. Frisoli, C. Compton, T. Flotte, J. M. Richter, N. Nishioka, and T. Deutsch, "UV LIF of colonic tissue: Basic biology and diagnostic potential," *Lasers Surg. Med.*, vol. 12, pp. 63–78, 1992.

[13] A. J. Welch, C. Gardner, R. Richards-Kortum, G. Criswell, E. Chan, J. Pfefer, and S. Warren, "Propagation of fluorescent light," *Lasers Surg. Med.*, vol. 21, pp. 166–78, 1997.

[14] 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. Spectroscopy.*, vol. 53, pp. 302–11, 1998.

[15] R. A. Zangaro, L. Silveira, R. Manoharan, G. Zonios, I. Itzkan, R. Dasari, J. Van Dam, and M. S. Feld, "Rapid multiexcitation fluorescence spectroscopy system for *in vivo* tissue diagnosis," *Appl. Opt.*, vol. 35, pp. 5211–9, 1997.

[16] J. A. Zuclich, T. Shimada, T. R. Loree, I. Bigio, K. Strobl, and Nie, "Rapid noninvasive optical characterization of the human lens," *Lasers Life Sci.*, vol. 6, pp. 39–53, 1994.

[17] S. B. Cantor, M. F. Mitchell, G. T. Luna, C. Bratka, D. Bodurka, and R. Richards-Kortum, "Cost-effectiveness analysis of diagnosis and management of cervical squamous intraepithelial lesion," *Obstet. Gynecol.*, vol. 91, pp. 270–77, 1998.

[18] R. J. Kurman, *Blaustein's Pathology of the Female Genital Tract*, 3rd ed, R. J. Kurman, Ed. New York: Springer-Verlag, 1987.

[19] D. L. Heintzelman, U. Utzinger, H. Fuchs, A. Zuluaga, K. Gossage, A. M. Gillenwater, R. Jacob, B. Kemp, and R. R. Richards-Kortum, "Optimal excitation wavelengths for *in vivo* detection of oral neoplasia using fluorescence spectroscopy," *Photochem. Photobiol.*, vol. 72, pp. 103–13, 2000.

[20] U. Utzinger, M. Brewer, E. Silva, D. Gershenson, R. C. Bast, and M. Follen *et al.*, "Reflectance spectroscopy for *in vivo* characterization of ovarian tissue," *Lasers Surg. Med.*, vol. 28, pp. 56–66, 2001.

[21] A. Agrawal, U. Utzinger, C. Brookner, C. Pitris, M. F. Mitchell, and R. Richards-Kortum, "Fluorescence spectroscopy of the cervix: Influence of acetic acid, cervical mucus and vaginal medications," *Lasers Surg. Med.*, vol. 25, pp. 237–49, 1999.

[22] N. Ramanujam, M. F. Mitchell, A. Mahadevan, S. Thomsen, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, "Development of a multivariate statistical algorithm to analyze human cervical tissue fluorescence spectra acquired *in vivo*," *Lasers Surg. Med.*, vol. 19, pp. 46–62, 1996.

[23] W. R. Dillon and M. Goldstein, *Multivariate Analysis: Methods and Applications*. New York: Wiley, 1984.

[24] E. V. Trujillo, D. R. Sandison, U. Utzinger, N. Ramanujam, M. F. Mitchell, and R. R. Richards-Kortum, "A method to determine tissue fluorescence efficiency *in vivo* and predict signal to noise ratio for spectrometers," *Appl. Spectroscopy*, vol. 52, pp. 943–51, 1998.

[25] P. A. Lachenbruch, *Discriminant Analysis*. New York: Hafner., 1975.

[26] M. F. Mitchell, W. Hittelman, W. K. Hong, R. Lotan, and D. Shottenfeld, "The natural history of CIN: An argument for intermediate endpoint biomarkers," *Cancer Epidemiol. Biomarkers Prev*, vol. 3, pp. 619–26, 1994.