In vivo light scattering for the detection of cancerous and precancerous lesions of the cervix

Judith R. Mourant,[1,*] Tamara M. Powers,[1] Thérese J. Bocklage,[2] Heather M. Greene,[3] Maxine H. Dorin,[4] Alan G. Waxman,[4] Meggan M. Zsemlye,[4] and Harriet O. Smith[5]

[1]Bioscience Division, Los Alamos National Laboratory, P.O. Box 1663, MS E535, Los Alamos, New Mexico 87545, USA

[2]Pathology Department, University of New Mexico Health Science Center, BMSB 335, 915 Camino de Salud, Albuquerque, New Mexico 87131, USA

[3]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University of New Mexico Health Sciences Center, 2211 Lomas Boulevard NE, Albuquerque, New Mexico 87131-5286, USA

[4]Department of Obstetrics and Gynecology, University of New Mexico Health Sciences Center, 2211 Lomas Boulevard NE, Albuquerque, New Mexico 87131-5286, USA

[5]Department of Obstetrics and Gynecology and Women's Health, Division of Gynecologic Oncology, Albert Einstein College of Medicine and Montefiore Medical Center, 1695 Eastchester Road, Suite 601, Bronx, New York 10461, USA

*Corresponding author: jmourant@lanl.gov

Received 30 June 2008; revised 12 November 2008; accepted 15 November 2008; posted 17 November 2008 (Doc. ID 97586); published 7 January 2009

A noninvasive optical diagnostic system for detection of cancerous and precancerous lesions of the cervix was evaluated *in vivo*. The optical system included a fiber-optic probe designed to measure polarized and unpolarized light transport properties of a small volume of tissue. An algorithm for diagnosing tissue based on the optical measurements was developed that used four optical properties, three of which were related to light scattering properties and the fourth of which was related to hemoglobin concentration. A sensitivity of ~77% and specificities in the mid 60% range were obtained for separating high grade squamous intraepithelial lesions and cancer from other pathologies and normal tissue. The use of different cross-validation methods in algorithm development is analyzed, and the relative difficulties of diagnosing certain pathologies are assessed. Furthermore, the robustness of the optical system for use by different doctors and to changes in fiber-optic probe are also assessed, and potential improvements in the optical system are discussed. © 2009 Optical Society of America

OCIS codes: 170.1610, 300.0300.

1. Introduction

Optical diagnostics have the potential to provide real-time diagnosis of tissue, and many optical diagnostic techniques are being developed. For example, fluorescence, light scattering, and a combination of the two continue to be investigated for their ability to accurately detect precancerous lesions of the cervix, and a summary of published clinical studies is given in Section 4. The motivation for developing optical methods for detection of cancerous and precancerous cervical lesions is that current methods have several shortcomings, including missed lesions [1,2] and loss of patients to follow up [3]. The diagnositic procedures currently in clinical practice are not suitable for "see and treat" methods that would allow treatment at the time of diagnosis [4,5].

In this study, light scattering spectroscopy alone is studied. Specifically, we have designed, built, and implemented a unique fiber-optic probe that determines both morphological and biochemical

0003-6935/09/100D26-10$15.00/0
© 2009 Optical Society of America

properties of tissue by measuring the transport of linearly polarized and unpolarized light through a small volume of tissue. The primary goal of the study was to determine the accuracy of the light scattering measurements for the detection of high grade squamous interepithelial lesions (HSILs), a precursor for cervical cancer. We have also investigated the effects of having different physicians using the instrument and of changes in optical probes during the course of the clinical study. The results of different resampling methods for determining the accuracy of our classification algorithms are also compared.

2. Methods

A. Instrumentation

The experimental measurement system and probe are illustrated in Fig. 1. The tungsten lamp box contains two tungsten lamps (Gilway, Model L1041) each with a UV filter (Hoya, Y-48, required by the FDA) and a shutter (Uniblitz VMM-D3). Light collected by each of the optical fibers is simultaneously dispersed using an Acton Spectra Pro 275 spectrograph with an attachment designed specifically for imaging optical fibers onto a CCD. A front-illuminated, TE-cooled CCD (Princeton Instruments) is used for light collection. To measure tissue, the probe is placed in gentle contact with the tissue and can be used to measure either polarized or unpolarized light transport. When the shutter for the lamp illuminating fiber DU is open, the light collected by fiber U is a measurement of unpolarized light transport. The center-to-center separation of fibers DU and U is 550 μm. The light source for polarized measurements is fiber DP, and the polarized collection fibers are 1, 3, and 4. There is a linear polarizer [3 M, HN 32% × 0.01 in. (0.03 cm)] over fibers DP, 1, and 4 that allows horizontally polarized light to pass. The linear polarizer over fiber 3 allows vertically polarized light to pass. All collection fibers are angled at 20° toward their respective delivery fiber in order to optimize sensitivity to epithelial tissue, and all optical fibers in the probe are 200 μm in diameter with a numerical aperture of ~0.37. Titanium dioxide in solid epoxy is used as the reference material. It is submerged in water, and the probe is placed in contact with this reference material for measurement. Nine parameters, which are described in detail below, are calculated from the spectra: total hemoglobin (Hb) concentration (multiplied by path length), a fraction of Hb that is oxygenated, a vessel "size" (divided by path length), exponent of the power law dependence of the unpolarized light scatter, amplitude of the unpolarized light scatter, slope of the unpolarized light scatter, the ratio of light collected by fibers 1 and 3, the ratio of light collected by fibers 1 and 4, and water concentration (multiplied by path length) [6].

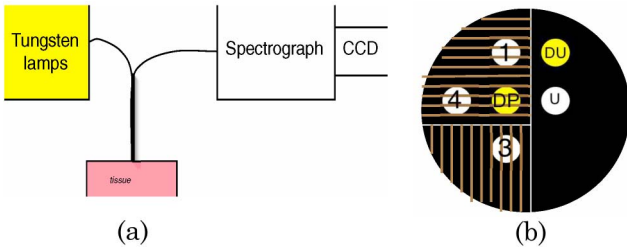

Fig. 1. (Color online) (a) Schematic of the fiber-optic system. The tungsten lamp box contains two lamps each with its own computer-controlled shutter. One lamp illuminates an optical fiber, DP, while the other lamp illuminates an optical fiber, DU. These optical fibers can transport the light to the tissue via an optical probe that is in gentle contact with the tissue. Light that is collected from the tissue by several optical fibers is simultaneously dispersed using a spectrograph, and spectra for each of the fibers are measured using a TE-cooled CCD. (b) Schematic of the distal end of the probe. There is a polarizer at top left, covering delivery fiber DP and collection fibers 1 and 4, that polarizes the light horizontally. A polarizer at bottom left, covering only fiber 3, polarizes the light vertically. There is no polarizer over delivery fiber DU and collection fiber U.

B. Spectral Data Fitting

To estimate Hb concentration and oxygenation, the portion of the unpolarized spectrum from 500 to 800 nm was fitted to Eq. (1):

$$I = (c_0\lambda^{-c_1})e^{-\mu_a lV}, \tag{1}$$

where

$$\mu_a l = C_{\mathrm{Hb}}\epsilon_{\mathrm{Hb}} + C_{\mathrm{HbO_2}}\epsilon_{\mathrm{HbO_2}}, \tag{2}$$

$$V = \frac{1 - \exp(-2\mu_a lR)}{2R\mu_a l}. \tag{3}$$

C_{Hb} and $C_{\mathrm{HbO_2}}$ are the concentrations of deoxygenated and oxygenated Hb, respectively, both multiplied by the path length. ϵ_{Hb} and $\epsilon_{\mathrm{HbO_2}}$ are the wavelength-dependent absorption coefficient of deoxygenated and oxygenated Hb, respectively, both taken from the literature. c_0 is the "amplitude" of the unpolarized light scatter, and c_1 is the exponent of the power law dependence of the unpolarized scattered light intensity, sometimes called scatter power. R is the "size" of the blood vessels divided by the path length the collected light traveled. Equation (1) is an equation for the absorption due to Hb in blood vessels [7]. The total Hb content (multiplied by optical path length) is given by $C_{\mathrm{Hb}} + C_{\mathrm{HbO_2}}$, while the fraction of Hb that is oxygenated is given by $C_{\mathrm{HbO_2}}/(C_{\mathrm{Hb}} + C_{\mathrm{HbO_2}})$. An example fit to Eq. (1) is given in Fig. 2. In some cases there was insufficient Hb absorption to determine vessel diameter, and in those cases V was set to 1. A more general expression that includes both Hb in blood vessels and Hb outside blood vessels has been described [7], however, for most of our data this expression resulted in overparameterization of the data.

The slopes of the unpolarized data from 690 to 790 nm are calculated by fitting a straight line to the data from 690 to 790 nm as shown in Fig. 2.

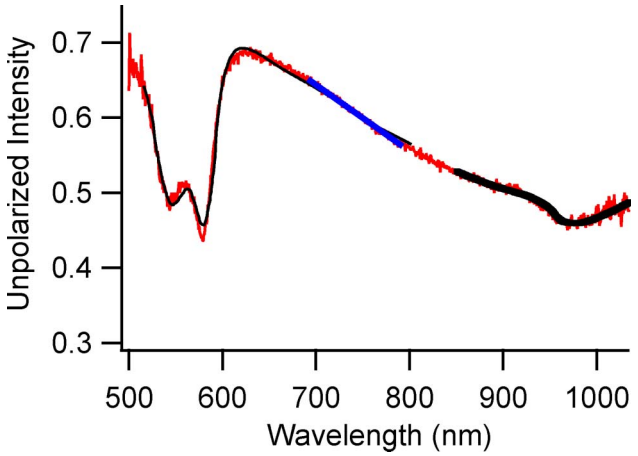

Fig. 2. (Color online) Representative unpolarized spectrum and the fits to that spectrum. The thick black line from 850 to 1035 nm is a fit to the water absorption. The thick solid line from 690 to 790 nm is a fit to the slope, and the thin black line is a fit to Hb absorption.

The slope was then divided by the area under the spectrum from 690 to 790 nm. Consequently, this slope value is proportional to 1/intensity as demonstrated in Eq. (4), where it is assumed that the data are a straight line from 690 to 790 nm and $\langle y \rangle$ is the average value of the intensity between 690 and 790 nm. The magnitude of the slope has been found to be greater for proliferating cells [8], and a greater slope magnitude indicates that the average size of structures scattering light is smaller [9,10]:

$$\text{slope} = \frac{\Delta x}{\Delta y}\frac{1}{\text{area}} \sim \frac{\Delta x}{\Delta y}\frac{1}{\langle y \rangle \Delta x} = \frac{1}{\langle y \rangle \Delta y}. \qquad (4)$$

The ratio of light intensity from fibers 1 and 3 and the ratio of light intensity from fibers 1 and 4 are calculated as a function of wavelength. At wavelengths past ~900, the polarizers do not polarize. Therefore this region can be used to normalize the data and correct for the different collection efficiencies of the optical paths, including fibers 1, 3, or 4. Specifically, I1/I3 and I1/I4 are normalized to 1 from 950 to 1000 nm. Example spectra are shown in Fig. 3 The physical interpretation of I1/I3 and I1/I4 has been determined in previous work. I1/I3 is greater for more strongly scattering tissue, and I1/I4 increases as the average size of scattering structures decreases [11]. Finally, water concentration was calculated in a manner analogous to that used for total Hb. An example of the fit is given in Fig. 2.

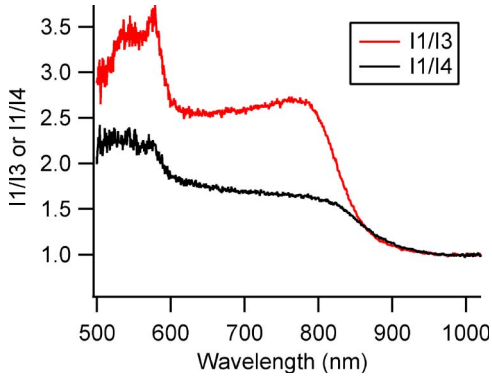

Fig. 3. (Color online) Representative I1/I3 and I1/I4 spectra.

C. Clinical and Histopathology

All tissue sites for which the colposcopist planned to take a biopsy as part of normal clinical procedure were measured with the spectroscopic system. Additionally, one or two normal sites were also measured but not biopsied. All sites were measured once with the spectroscopic system, and then the measurements were repeated. Subsequently, biopsies were obtained and each biopsy was placed in a separate container. After all spectroscopic measurements of a patient were completed, the probe was gently wiped off, and a reference measurement was made.

Each biopsy was characterized as normal, cervicitis, low-grade squamous intraepithelial lesion (LSIL), HSIL, or cancer by the study pathologist. The study pathologist also ranked the inflammation as none, a few clusters of inflammatory cells, or many inflammatory cells. Vascularity was paramaterized as normal or increased. The tissue site was determined by histopathology as ectocervix (squamous epithelium), endocervix (columnar epithelium), or squamous columnar junction (SCJ).

Data from 151 patients were acquired and analyzed. Data from several other patients could not be used primarily due to failure of our decade-old equipment. Human subjects review boards reviewed and approved this work at both the University of New Mexico and at Los Alamos National Laboratory. Consent was obtained from each patient by the study coordinator.

D. Correcting the Data for Small Differences in Probes

Data from 64 of the patients were acquired with the original fiber-optic probe dedicated to this study. When that probe broke irreparably, data from the rest of the patients were acquired with a replacement fiber-optic probe that was very similar but not perfectly identical. To determine if the change in probe had any effect on the measurements, data from the two probes were compared using Student's t-tests within each pathology classification. (The two instances of invasive cancer were not included in the comparison.) When statistically significant differences were found in multiple pathology categories, the measurements made with the second probe were multiplied by a correction ratio that was calculated as follows. For each pathology classification, the average for probe 1 divided by the average of probe 2 was calculated. The average of these ratios was the correction ratio. No significant differences were found after the data were corrected.

E. Identifying and Correcting for Differences between Doctors

Four doctors participated in this study and made spectroscopic measurements. Each patient was only

measured by only one doctor. The Student's t-test was used to determine if there were significant differences in the average values of the spectroscopic variables measured by the different doctors. The values of spectroscopic parameters measured by two doctors cannot be compared by simply using the mean and standard deviation of all measurements of a spectroscopic variable, because each doctor may have measured a different fraction of patients with a given pathology. Also, the colposcopically normal data were not used, because the tissue that one doctor used for the colposcopically normal measurements may be different from what another doctor chose.

Here is an example of how statistically significant differences were determined for I1/I3. The averages of all cervicitis measurements for each doctor were determined. Call these p_{i1}, where i goes from 1 to 4 for the four doctors. Then the averages of a second pathology were determined. Call these p_{i2}. Then the ratio of I1/I3 values for pathology 2 to cervicitis is calculated as

$$r_2 = \sum_{i=1}^{4} \frac{p_{i2}}{4p_{i1}}. \tag{5}$$

A new set of I1/I3 data for each doctor is then calculated as

$$D_{in} = \sum_{j=1}^{4} d_{ijn} r_j, \tag{6}$$

where d are the original data, j is over the four pathologies (HSIL and cancer were grouped together) and n is the subscript for an individual data point. (Note that the colposcopically normal data are not part of these data sets.)

The standard deviation and averages were calculated for these data sets in order to use the Student's t-test to determine whether there are significant differences between the new I1/I3 data sets for each doctor.

When differences in the data sets were found ($p < 0.05$), corrections were made to the original data. For I1/I3, slope, total Hb, and scatter power, the averages were similar for three doctors, while a fourth doctor had significantly different averages from two of the first three (in the case of total Hb it was only one of the first three). Therefore, the raw data for the fourth doctor was multiplied by a correction factor so that the data sets calculated by Eq. (6) had the same average for all four doctors. The "odd" doctor was not the same in every case. For total Hb and amplitude, doctors 1 and 2 had similar results, while doctors 3 and 4 had similar results. The results for doctors 3 and 4 were multiplied by a correction ratio, which was the average of doctors 3 and 4 divided by the average of doctors 1 and 2. The colposcopically normal data were multiplied by the same correction factor as the rest of the data.

F. Probability Distributions

Histograms of the number of sites with a given value of a spectroscopic variable were made for each diagnostic category for each measured variable. Histograms for slope, I1/I3 and I1/I4, were then fitted to Gaussians and normalized to yield probability distributions. These histograms provide a visual picture of the changes in spectroscopic values with tissue pathology and of the overlap between different categories.

G. Classification of the Measured Sites

We initially analyzed our data with the Mahalanobis distance metric, which is the analysis method used by Chang *et al.* [12] and Mirabal *et al.* [13] but found that significantly worse results were obtained for the testing sets than for the training sets, indicating that this method was overtraining. The following classification method was found to give more similar results between the training and the testing data sets. A vote is cast "by" each of three variables: slope, the average value of I1/I4 from 660 to 760 nm, slope, and the average value of I1/I3 from 660 to 760 nm. For each variable there is a cutoff value. If the measured value for a site is on one side of the cutoff then the vote is positive, i.e., for HSIL or cancer. If it is on the other side of the cutoff, it is for the negative category. Initial classification is then a two out of three vote. In addition, if total Hb is very high, an initially negative classification is changed into a positive classification if at least one of the three variables, I1/I4, slope, and I1/I3, had a positive vote. The cutoff values were optimized as follows. The data were normalized so that the data range for each variable was 0 to 10. The cutoff for total Hb was set at 3, and the cutoff for I1/I4 was fixed at 4.5. A wide range of I1/I3 cutoffs was then tested. For each I1/I3 cutoff a wide range of slope cutoffs was tried. The optimum cutoffs for I1/I3 and slope were defined as those providing the largest sum of sensitivity and specificity such that the sensitivity was greater than or equal to 80%. A sensitivity greater than 80% was required in order to limit the number of HSIL sites that were missed. Using fixed values for the cutoffs for I1/I4 and Hb reduced the variation in results of the training and testing sets. Furthermore, correlations between the different variables meant that disparate combinations of cutoff values would yield the same results. By holding I1/I4 constant, the optimization problem became much smaller with little change in the ultimate results. The value for the Hb cutoff was chosen such that the vast majority of measurements had total Hb less than the cutoff. (The distribution of Hb measurements is non-Gaussian.) The value for the I1/I4 cutoff was chosen to be a number with only two significant digits that was near to values commonly found in early optimization runs where I1/I4, I1/I3, and slope were all varied.

Fivefold cross validation was used as a validation method for the classification algorithms. The data were split into five subsets of approximately equal size with each subset containing approximately the same proportion of each pathology classification. Each of the five subsets was used once as a testing set, with the remaining data used for training in each case. Sensitivity and specificity were estimated by averaging the results for the five data sets. This validation method was chosen because resampling methods, such as *n*-fold cross validation, have been shown to be better at evaluating models than nonresampling methods. Furthermore, fivefold and tenfold cross validation have been recommended over leave-one-out (LOO) cross validation [14,15], because LOO cross validation can have large variance (e.g., the results for one trial of 25 patients may be very different from the results for a different trial of 25 patients) [14,16].

To examine the differences between fivefold cross validation and LOO cross validation, both methods were used to develop and evaluate classification algorithms that optimize the sum of specificity and sensitivity, while keeping sensitivity greater than 80% for training data.

3. Results

A. Spectra

Figure 2 shows a representative spectrum of collected unpolarized light and the fits to that spectrum. There are some small systematic errors in the Hb region which were fairly common. Examples of I1/I3 and I1/I4 are shown in Fig. 3. The distance light travels from the linearly polarized delivery fiber to the cross-polarized collection (I3) is longer than to the copolarized collection (I1), therefore the Hb bands show up as positive in the I1/I3 spectrum. Similarly, the path length of light traveling from the delivery fiber to I4 is longer than the path length from the delivery fiber to I1, [17] and the Hb bands are positive in the I1/I4 spectrum.

B. Dependence of Spectroscopic Measurements on Nondiagnostic Parameters

Two fiber-optic probes were used in this study. The mean values for I1/I4 and water differed for every pathology classification for the two probes used in this study, and therefore the data were corrected for these differences. In contrast, no significant differences were found between the probes for I1/I3 for any pathology category. For slope, a significant difference was found only for the category of "colposcopically normal" and was not corrected. The total Hb measurement was found to differ significantly for two categories and was corrected. Whatever physical differences in the probes caused water concentration measurements to differ also likely caused the Hb concentration measurement to vary. However, the distributions of Hb concentration in tissue are broader, and consequently differences between the probes are more difficult to detect.

This study was performed by four clinicians and a few systematic differences were found in the results for different doctors. The average values of I1/I3, slope, total Hb, scatter power, and amplitude were all significantly different between some of the doctors. Corrections to the data were made as described in Section 2.

We have also investigated how patient characteristics (e.g., patient age) affect the spectroscopic measurements. This work is described in a separate paper where dependencies on menstrual cycle and patient age are reported [18].

C. Pathologies, Epithelial Type, Inflammation, and Vascularity

The pathology of the measured sites is given in Table 1. A total of 362 sites were used in this analysis, half of which were biopsied sites and half of which were normal via colposcopic examination and not biopsied. The vast majority of biopsied sites were of the SCJ, which contains some combination of squamous, columnar, and metaplastic epithelium. 24 biopsies were confirmed to be of the ectocervix, which is usually squamous epithelium, and 11 biopsies were confirmed to be of the endocervix, which is usually columnar epithelium. On average, inflammation was increased for cervicitis and HSIL compared to the normal sites. Vascularity was more likely to be increased for cervicitis and HSIL than in the normal and LSIL biopsies.

D. Probability Distributions

Examination of the probability distributions for a given spectroscopic variable for each pathology provides insight into which pathology categories can be accurately diagnosed. Figure 4 shows the distributions of values of I1/I4 and slope obtained for the different diagnostic categories. The best separation is between the categories of colposcopically normal and HSIL. LSIL and HSIL have very similar distributions. The distribution for "no diagnostic abnormality" is as narrow as or narrower than the other distributions.

E. Diagnosing HSILs and Cancers

A goal of this work is to identify HSILs and cancers versus sites with other pathologies and normal tissue

Table 1. Pathology of the Measured Sites[a]

Pathology	Number of Sites
Colposcopy "normal"	181
Normal by histopathology	36
Cervicitis	44
LSIL	43
HSIL	56
Cancer	2
Total	362

[a]Colposcopy "normal" are nonbiopsied sites assumed to be normal by the colposcopist.

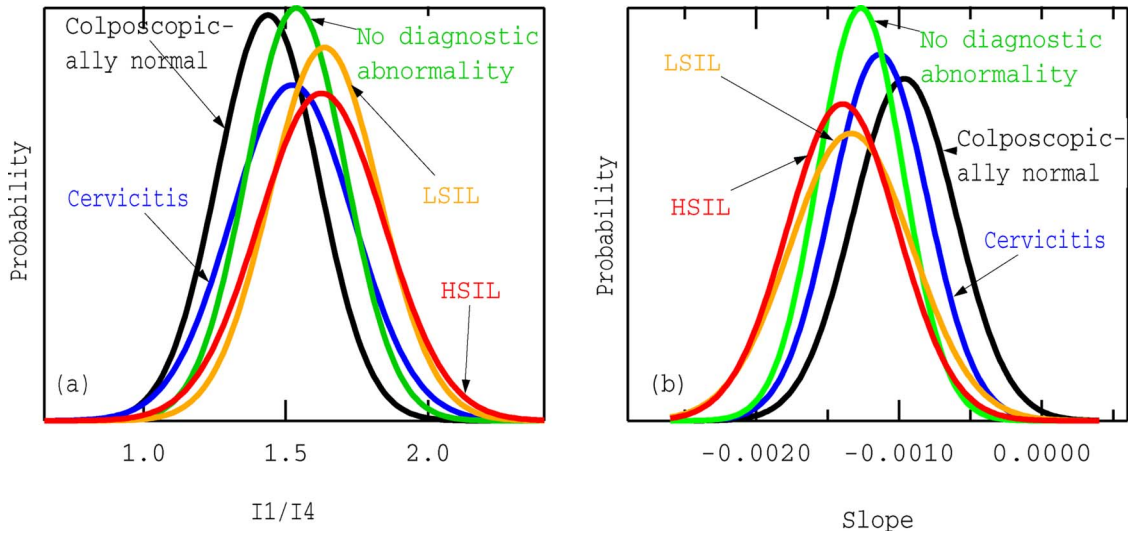

Fig. 4. (Color online) Probability distributions for the values of I1/I4 (a) and slope (b). The length of the x-axis is 1.4 times the range of measured values.

sites. There were several confounding factors. The measurements depended slightly on which doctor made the measurements and on which optical probe was used (Section 3.B). We also found that the some of the spectroscopic parameters depend on patient age and menopausal or menstral cycle status [18]. Corrections to the data for all these confounding effects were made before the ROC curves were calculated and the classification algorithms developed.

Reciever operating characteristic (ROC) curves for the diagnosis of HSIL and cancer versus the other pathologies are presented in Fig. 5. The areas under the ROC curves for slope, I1/I3, I1/I4, and total Hb are 0.69, 0.64, 0.70 and 0.64, respectively. ROC curves are not shown for water, oxyHb, vessel "size," and "amplitude" because the area under them was close to 0.5. The wavelength dependence of the unpolarized light scatter (scatter power) is very similar to the slope parameter. Since the area under the ROC curve for this parameter was 0.65,which is less than that for slope, this parameter was not used for classication. Because none of the areas under the ROC curves are near the perfect value of 1, a method of combining these metrics was desired. In the course of analyzing the data, several different methods were considered (e.g., classification by Mahalobonis distance). A voting method was chosen for simplicity and because of the similarity found between results for training and testing data sets. The inputs to the voting method are measured values for I1/I4, slope, I1/I3, and total Hb as described in Subsection 2.G. The results are shown in Table 2. The best results are obtained when the colposcopy normal sites are included, when the positive category is HSIL or cancer, and when the negative category is nondysplastic. The average results for the testing data sets are then a sensitivity of 77% and a specificity of 68%. When colposcopically normal sites are not included, the obtained sensitivity is 79% and the specificity is 47%.

F. Comparison of Validation Methods

LOO cross validation is a very common method for assessing the accuracy of a classification method. Table 3 compares results obtained with LOO cross validation and fivefold cross validation. In the top row, where the disease classification is HSIL and cancer and the nondisease classification is LSIL and nondysplastic, the sensitivities are the same, but the specificities are higher for LOO. In the bottom row, where the disease classification is HSIL and cancer and the nondisease classification is nondysplastic, the results are nearly identical.

4. Discussion

A. Fundamental Light Scattering and Comparison with *In Vitro* Work

Values of I1/I4 and slope have been previously shown in tissue phantom studies to correlate with the average size of the scattering structures [9–11]. I1/I4 and the slope magnitude (the slope is negative) are greater for HSIL than for nondysplastic tissue, indicating that the average size of the scattering centers decreases in HSIL. This change may be due to increased spatial fluctuations in DNA content in dysplastic nuclei, which have been shown to affect light scattering [19].

An increase in slope magnitude and greater values for I1/I3 were seen for HSIL sites in this study and for our tumorigenic model in our previous *in vitro* experiments. However, the changes in I1/I4 found in this clinical work differ from the changes in I1/I4 seen in our *in vitro* measurements comparing a tumorigenic and nontumorigenic model [20]. I1/I4 was smaller for the tumorigenic model than the nontumorigenic model. In contrast, I1/I4 was larger for *in vivo* precancerous and cancerous tissue. The reason for this difference is currently not known. However, there are several possibilities. Possibly, the

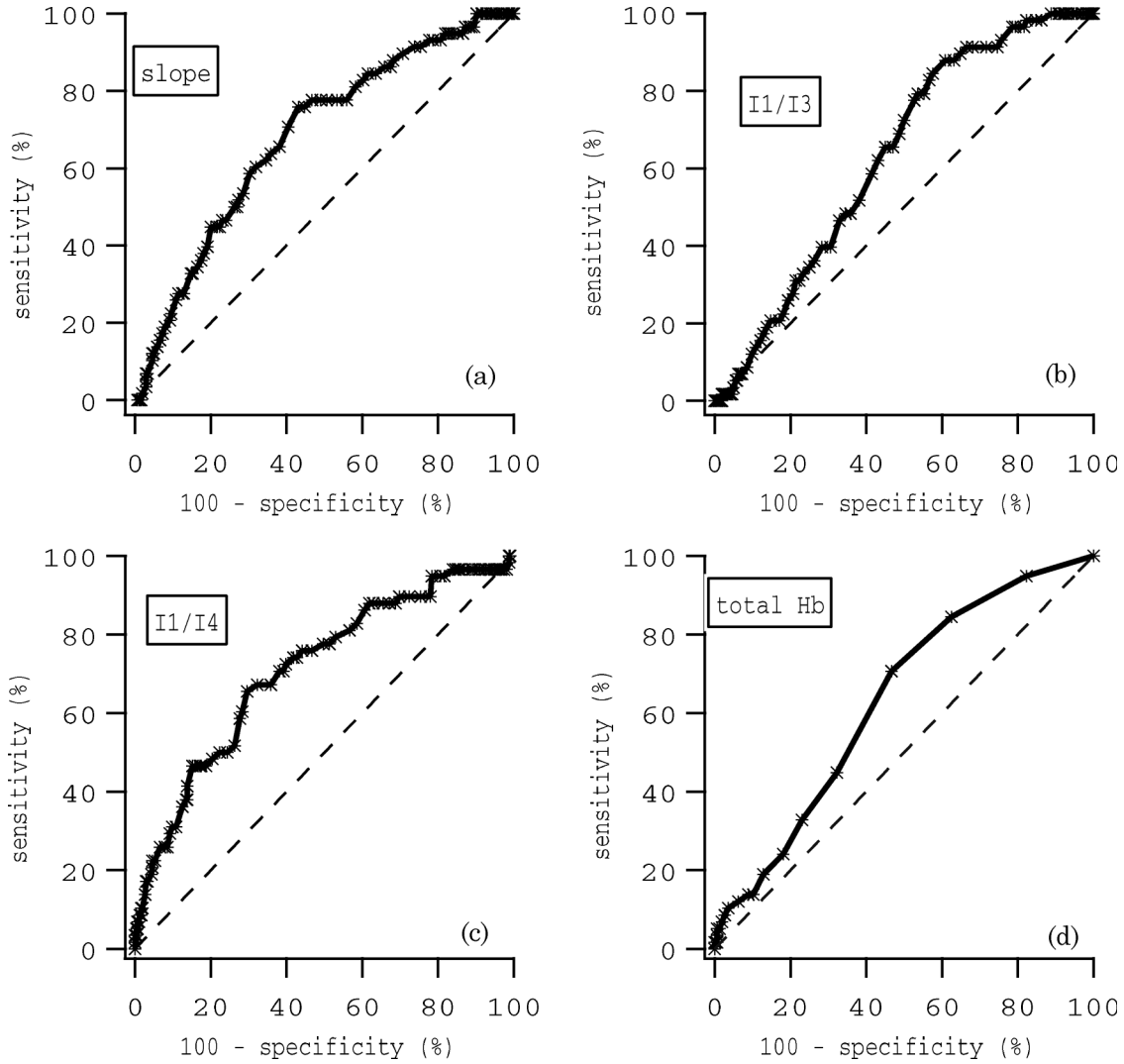

Fig. 5. (Color online) Receiver operating characteristic (ROC) curves. The dashed lines from (0,0) to (100,100) are to guide the eye.

morphological changes in the cells really are different between the *in vitro* fibroblast model and the *in vivo* cervical epithelial cells. Second, tissue is more complex than the cell models. *In vivo*, we measure not only cells, but also some of the underlying stroma. Both the thickness of the epithelial layer of cells and the properties of the stroma may change in precancerous tissue.

B. Comparison to Other *In Vivo* Studies

When comparing results of different studies, several details of the studies should be considered. Extremely important is the extent of inclusion of sites that are expected to be normal. We (Fig. 4 and Table 2) and others [21] have found that separating sites that appear normal by colposcopy from HSIL sites is easier than separating nonnormal appearing sites from HSIL.

Another consideration is the validation and resampling methods, particularly for small data sets. As noted in Subsection 2.G some resampling methods give more robust results. Closely related is the reported error in the presented sensitivity and specificity. Unfortunately, this is frequently not reported.

Table 2. Ability to Detect HSIL and Cancer[a]

	All Measured Sites				Colposcopically Abnormal Sites			
	Training Set		Testing Set		Training Set		Testing Set	
	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
HSIL & cancer vs. LSIL & nondysplastic	82.8	60.8	77.1 ± 4.5	62.2 ± 1.7	82.3	43.1	77.1 ± 4.5	43.8 ± 3.3
HSIL & cancer vs. nondysplastic	82.3	66.1	77.1 ± 4.5	67.8 ± 2.0	81.9	49.7	78.9 ± 4.4	47.3 ± 3.7

[a]The errors presented are standard errors of the mean. Sens., sensitivity; Spec., specificity.

Table 3. Comparison of the Ability to Detect HSIL and Cancer as Determined by Two Different Cross-Validation Methods[a]

	All Measured Sites				Colposcopically Abnormal Sites			
	LOO		Fivefold		LOO		Fivefold	
	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
HSIL & cancer vs. LSIL & nondysplastic	77.6	64.1	77.1 ± 4.5	62.2 ± 1.7	77.6	47.2	77.1 ± 4.5	43.8 ± 3.3
HSIL & cancer vs. nondysplastic	77.6	68.6	77.1 ± 4.5	67.8 ± 2.0	77.6	48.8	78.9 ± 4.4	47.3 ± 3.7

[a]The errors presented are standard errors of the mean. Sens., sensitivity; Spec., specificity; LOO, leave-one-out cross validation; fivefold, fivefold cross validation.

The number of patients in the study should also be considered. In our own work, we have found that better sensitivities and specificities are obtained when the sample size is smaller. The number of patients in this study, 151, is comparable to or significantly larger than that used in previously published studies.

Table 4 summarizes results from three distinct point measurement spectroscopy studies that used fluorescence, light scattering, or a combination of both. The first line is a study from 1996 that used fluorescence from three separate excitation wavelengths [22]. 59% of the samples were normal epithelium. The diagnostic algorithm was developed with a calibration model and tested on a testing data set. The results for the separate data set were slightly worse than for the calibration set, similar to our differences between average training and average testing set results. The results appear slightly better than those in this study, however, this result may be caused by the slightly higher percentage of expected normals. The second two lines in Table 4 are results from the same study. In one case fluorescence was used to perform the diagnosis, and in the second case reflectance was used. Our result of $77 \pm 4.5\%$ specificity is similar to these results, while our results of $68 \pm 2\%$ sensitivity are slightly lower than from this study that has a similar number of patients. Comparison of our results to the third study is difficult because pathology classifications are different and the study is quite a bit smaller [21]. Nonetheless, the bottom two lines of Table 3 demonstrate a very important point. The reported accuracy greatly decreases when expected normal sites are not included in the study.

Table 5 shows the results of three studies performed with imaging instruments. The first study (rows 1–4) was quite small and contained a large number of expected normal sites, 373 out of a total of 490 [23]. The second study (row 5) is the largest published study and used an excellent cross-validation method. The fraction of sites that was normal by colposcopy was probably smaller than our study (an exact number is not given) [24]. The final study (row 6) is again difficult to compare with ours, because the sensitivity and specificity are presented on a per patient basis and the validation method was unusual (see Table 5 caption) [25]. Overall, our results compare well to those in the peer-reviewed literature.

Results from a small portion of patients in this study have been previously reported [6]. Specifically, results from 29 patients were reported in a retrospective study with no validation of the classification method. In an analysis that excluded 3 sites, a sensitivity of 100% and a specificity of 80% were obtained. When the colposcopy normal sites were not included the specificity dropped to 55%. These results are somewhat better than those reported here for the much larger data set. The major reason for this change in accuracy is, most likely, that no validation was performed for the first study, while the larger study used fivefold cross validation. Furthermore, it seems plausible that better results are obtained with a small data set because classification parameters are optimized for the unique characteristics of that small data set. Because of the discrepancy in accuracy, we examined the significance of variables used for classification to determine if that had changed. For the small data set, slope and I1/I4 were found to have significantly different averages (i.e., $p < 0.05$) for the non-HSIL and the HSIL data regardless of whether the colposcopy normals were included [6]. Those results held for the larger data set. Similarly, I1/I3 was significant only when the colposcopy normals were included for both the small and large data sets.

An optical imaging system that performs both fluorescence and reflectance has been reported to

Table 4. *In Vivo*, Point Optical Studies of Cervical Pathologies[a]

Pathology stratification	Method	Validation	Use ENS	Sens. %	Spec. %	Patients	Sites	Ref.
HSIL vs. LSIL & nondysplastic	Fluorescence	Testing set	Yes	79 ± 2	78 ± 6	95	381	[22]
HSIL vs. nondysplastic	Fluorescence	LOO	Yes	83	80	161	324	[12]
HSIL vs. nondysplastic	Reflectance	LOO	Yes	72	81	161	324	[13]
SIL vs non-SIL	Fluor. & refl.	LOO	No	92	71	44	97	[21]
SIL vs non-SIL	Fluor. & refl.	LOO	Yes	92	90	44	97	[21]

[a]Sensitivity and specificity are on a per site basis. Sens., sensitivity; Spec, specificity; ENS, expected normal sites; fluor, fluorescence; refl, reflectance; LOO, leave-one-out cross validation.

Table 5. *In vivo*, Optical Imaging Studies of Cervical Pathologies[a]

Pathology stratification	Method	Validation	Use ENS	Sens. %	Spec. %	Patients	Sites	Ref.
HSIL vs. sq. normal[b]	Fluorescence	50/50 1000×	373/490	91	93	41	490	[23]
HSIL vs. metaplasia	Fluor.	50/50 1000×	373/490	90	87	41	490	[23]
HSIL vs. sq. normal[b]	Reflectance	50/50 1000×	373/490	82	67	41	490	[23]
HSIL vs. metaplasia	Refl.	50/50 1000×	373/490	77	76	41	490	[23]
HSIL vs. non-HSIL	Fluor. & refl.	70/30 100×	a little	~90	~50	271	1569	[24]
HSIL vs. non-HSIL[c]	Fluor. & refl.	Fivefold resampling[d]	Yes	95[b]	83[b]	111	NA	[25]

[a]Sens., sensitivity; Spec, specificity; ENS, expected normal sites, sq, squamous; fluor., fluorescence; refl, reflectance.
[b]Sensitivity and specificity are on a per patient basis. Squamous normal was determined by colposcopy not pathology.
[c]Diagnoses were determined by either pathology or colposcopy.
[d]Fivefold resampling was done for the ROC curves and yielded an area of 0.924 versus the original number of 0.947. The sensitivity and specificity are from the original ROC curve.

increase detection rates of HSIL and cancer [26]. In one published study, half of the patients went to a colposcopy-only arm and the others went to a colposcopy plus optical imaging arm. The percent of patients found to have HSIL or worse was significantly greater in the optical imaging plus colposcopy arm, 14.4% versus 11.4%. In a second study of 193 subjects, colposcopists completed their standard exam and then were instructed to take at least one biopsy in a region identified as high probability for HSIL or worse by the optical imaging system [27]. An additional 9 patients were identified as having HSIL or worse via this biopsy, above the 41 already identified by colposcopy. Importantly, approximately one more biopsy was taken per patient due to the use of the optical imaging system. The study did not demonstrate whether or not the increase in detection rates was simply due to the increase in the number of biopsies.

C. Clinical Utility

Information about how patient characteristics affect the spectroscopy data was used in this work to improve the quality of the data [18]. Most of this information is routinely acquired in a clinical exam (e.g., age), and the other information can easily be acquired. The values of the spectroscopic variables also depended slightly on the clinician making the measurement. In this work, we corrected for these effects. However, this correction will not generally be feasible. These differences are possibly due to differences in the time between when acetic acid is applied and measurements are made or more likely caused by how the doctors hold and use the fiber-optic probe. We are working on incorporating a pressure sensor into the probe so that all doctors will hold the probe against the tissue with the same pressure. Additionally, the manufacturing procedures are being modified in order to eliminate differences between optical probes.

Improvements in the accuracy of the optical system are clearly desired. To make these improvements it will be necessary to understand some fundamentals. For example, it is important to understand whether the widths of the probability distributions in Fig. 4 are instrumental, due to measurement technique, or biological in origin.

Alterations to the probe are planned, and implementation of the changes will be influenced by knowledge of the fundamental scattering processes that are the basis for this technique. For example, some alterations will provide additional light scattering information with very little complexity added to the probe. Other alterations will make the probe a more robust clinical tool. We have already built a prototype with an incorporated pressure sensor.

In order for this spectroscopy system to be used in the places where it is most needed (i.e., low income areas), the system must be made less expensive. There is tremendous potential for greatly decreasing the size and cost of the instrumentation used here, because light scattering is a strong and relatively easy to measure signal.

5. Conclusions

An elastic light scattering system that measures both polarized and unpolarized light transport in the cervical epithelium has a sensitivity of $77 \pm 5\%$ for detection of HSIL and a specificity of $44 \pm 3\%$ for colposcopically abnormal sites. An important result of this study is that much improved results are obtained if colposcopically normal sites are included in the analysis. Further conclusions are that spectroscopic measurements varied slightly depending on the doctor using the spectroscopic system and that similar results for sensitivity were obtained using either LOO cross validation or fivefold cross-validation, while specificity results were sometimes greater for LOO cross validation.

References

1. J. C. Gage, V. W. Hanson, K. Abbey, S. Dippery, S. Gardner, J. Kubota, M. Schiffman, D. Solomon, and J. Jeronimo, for the ASCUS LSIL Triage Study (ALTS) Group, "Number of cervical biopsies and sensitivity of colposcopy," Obstet. Gynecol. **108**, 264–272 (2006).
2. J. T. Cox, M. Schiffman, and D. Solomon, "Prospective follow-up suggests similar risk of subsequent cervical intraepithelial neoplasia grade 2 or 3 among women with cervical intraepithelial neoplasia grade 1 or negative colposcopy and directed biopsy," Am. J. Obstet. Gynecol. **188**, 1401–1412 (2003).

3. M. Spitzer, A. E. Cheryns, and V. L. Seltzer, "The use of large loop excision of the transformation zone in an inner-city population," Obstet. Gynecol. **82**, 731–735 (1993).
4. L. A. Dainty, J. C. Elkas, G. S. Rose, and C. M. Zahn, "Controversial topics in abnormal cervical cytology: "see and treat"," Clin. Obstet. Gynecol. **48**, 193–201 (2005).
5. R. Sankaranarayanan, R. Rajkumnar, P. O. Esmy, J. M. Fayette, S. Shanthakumary, L. Frappart, S. Thara, and J. Cherian, "Effectiveness, safety and acceptability of 'see and treat' with cryotherapy by nurses in a cervical screening study in India," Br. J. Cancer **96**, 738–743 (2007).
6. J. R. Mourant, T. J. Bocklage, T. M. Powers, H. M. Greene, K. L. Bullock, L. R. Marr-Lyon, M. H. Dorin, A. G. Waxman, M. M. Zsemlye, and H. O. Smith, "In vivo light scattering measurements for detection of precancerous conditions of the cervix," Gynecol. Oncol. **105**, 439–445 (2007).
7. R. L. P. van Veen, W. Verkryusse, and H. J. M. C. Sterenborg, "Diffuse-reflectance spectroscopy from 500 to 1060 nm by correction for inhomogeneously distributed absorbers," Opt. Lett. **27**, 246–248 (2002).
8. J. R. Mourant, M. Canpolat, C. Brocker, O. Esponda-Ramos, T. M. Johnson, A. Matanock, K. Stetter, and J. P. Freyer, "Light scattering from cells: the contribution of the nucleus and the effects of proliferative status," J. Biomed. Opt. **5**, 131–137 (2000).
9. J. R. Mourant, A. H. Hielscher, A. A. Eick, T. M. Johnson, and J. P. Freyer, "Evidence of intrinsic differences in the light scattering properties of tumorigenic and nontumorigenic cells," Cancer Cytopath. **84**, 366–374 (1998).
10. J. R. Mourant, J. P. Freyer, A. H. Hielscher, A. A. Eick, D. Shen, and T. M. Johnson, "Mechanisms of light scattering from biological cells relevant to noninvasive optical tissue diagnostics," Appl. Opt. **37**, 3586–3593 (1998).
11. J. R. Mourant, T. M. Johnson, and J. P. Freyer, "Characterizing mammalian cells and cell phantoms by polarized backscattering fiber-optic measurements," Appl. Opt. **40**, 5114–5123 (2001).
12. S. K. Chang, Y. N. Mirabal, E. N. Atkinson, D. Cox, A. Malpica, M. Follen, and R. Richards-Kortum, "Combined reflectance and fluorescence spectroscopy for *in vivo* detection of cervical pre-cancer." J. Biomed. Opt. **10**, 024031 (2005).
13. Y. N. Mirabal, S. K. Chang, E. N. Atkinson, A. Malpica, M. Follen, and R. Richards-Kortum, "Reflectance spectroscopy for *in vivo* detection of cervical precancer," J. Biomed. Opt. **7**, 587–594 (2002).
14. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)* (IJCAI, 1995), Vol. 2, 1137–1143.
15. L. Breiman and P. Spector, "Submodel selection and evaluation in regression—the X-random case," Int. Statist. Rev. **60**, 291–319 (1992).
16. B. Efron, "Estimating the error rate of a prediction rule: improvement on cross-validation," J. Am. Stat. Assoc. **78**, 316–331 (1983).
17. T. Johnson, J. Mourant, "Polarized wavelength-dependent measurements of turbid media," Opt. Express **4**, 200–216 (1999).
18. J. R. Mourant, T. J. Bocklage, T. M. Powers, H. M. Greene, M. H. Dorin, A. G. Waxman, M. M. Zsemlye, and H. O. Smith, "Detection of HSIL and cancers in cervical tissue by *in vivo* light scattering," J. Lower Genital Tract Disease, to be published.
19. R. Drezek, M. Guillard, 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).
20. J. Ramachandran, T. M. Powers, S. Carpenter, A. Garcia-Lopez, J. P. Freyer, and J. R. Mourant "Light scattering and microarchitectural differences between tumorigenic and nontumorigenic cell models of tissue," Opt. Express **15**, 4039–4053 (2007).
21. I. Georgakoudi, E. E. Sheets, M. G. Mulller, V. Backman, C. P. Crum, K. Badizadegan, R. R. Dasari, and M. S. Feld, "Trimodal spectroscopy for the detection and charaterization of cervical precancers *in vivo*," Am. J. Obstet. .Gynecol. **186**, 374–382 (2002).
22. 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. **64**, 720–735 (1996).
23. R. J. Nordstrom, L. Burke, J. M. Niloff, and J. F. Myrtle, "Identification of cervical neoplasia (CIN) using UV-excited fluorescence and diffuse-reflectance tissue spectroscopy," Lasers Surg. Med. **29**, 118–127 (2001).
24. W. K. Huh, R. M. Cestero, F. A. Garcia, M. A. Gold, R. S. Guido, K. McIntyre-Seltman, D. Harper, L. Burmke, S. T. Sum, R. F. Flewelling, and R. D. Avarez, "Optical detection of high-grade cervical intraepithelial neoplasia *in vivo*: results of 604 patient study," Am. J. Obstet. Gynecol. **190**, 1249–57 (2004).
25. D. G. Ferris, R. A. Lawhead, E. D. Dickman, N. Holtzapple, J. A. Miller, S. Grogan, S. Bambot, A. Agrawal, and M. A. Faupel, "Multimodal hyperspectral imaging for the noninvasive diagnosis of cervical neoplasia," J. Lower Gen. Tract Disease **5**, 65–72 (2001).
26. R. D. Alvarez and T. C. Wright, "Effective cervical neoplasia detection with a novel optical detection system: A randomized trial," Gynecol. Oncol.**104**, 281–289 (2007).
27. R. D. Alvarez and T. C. Wright, "Increased detection of high-grade cervical intraepithelial neoplasia utilizing an optical detection system as an adjunct to colposcopy," Gynecol. Oncol. **106**, 23–28 (2007).