Reflectance spectroscopy for in vivo detection of cervical precancer

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

Sung K. Chang  
University of Texas at Austin  
Department of Biomedical Engineering  
Austin, Texas 78712

Edward Neely Atkinson  
University of Texas M.D. Anderson Cancer Center  
Department of Biomedical Engineering Center  
Houston, Texas 77030

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

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

Rebecca Richards-Kortum  
University of Texas at Austin  
Department of Biomedical Engineering  
Austin, Texas 78712

Abstract. Optical technologies, in particular fluorescence spectroscopy, have shown the potential to provide improved detection methods for cervical neoplasia that are sensitive and cost effective through accurate, objective, instantaneous point-of-care diagnostic tools. The specific goals of this study were to analyze reflectance spectra of normal and neoplastic cervical tissue in vivo and to evaluate the data for use in diagnostic algorithm development. Spectroscopic measurements were obtained at four distinct source-detector separations from 324 sites in 161 patients. As the source-detector separation increases, greater tissue depth is probed. The average spectra of each diagnostic class differed at all source-detector separations, with the greatest differences occurring at the smallest source-detector separations. Algorithms, based on principal-component analysis and Mahalanobis distance classification, were developed and evaluated for all combinations of source-detector separations relative to the gold standard of colposcopically directed biopsy. The diagnostic combination of squamous normal versus high-grade squamous intraepithelial lesions gave good discrimination with a sensitivity of 72% and a specificity of 81%; discrimination of columnar normal versus high-grade squamous intraepithelial lesions was good, with sensitivity of 72% and specificity of 83%. Thus, reflectance spectroscopy appears promising for in vivo detection of cervical precancer. Strategies that combine fluorescence and reflectance spectroscopy may enhance the discrimination capabilities. © 2002 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1502675]

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

1 Introduction

Cervical cancer is the third most common cancer in women worldwide and the leading cause of cancer mortality for women in developing countries. When precancerous lesions are detected early they are easily treatable by ablation or excision. At more advanced stages, cervical cancer often requires hysterectomy, chemotherapy, radiation therapy, or combined chemoradiation therapy. In the developed world we have good screening and detection programs, but these are cumbersome and costly. Kurman et al. estimated that $6 billion are spent on the diagnostic evaluation and treatment of 2.5 million atypias and low-grade squamous intraepithelial lesions (LGSILs) alone. In the developing world, resources for screening are not available and, as a result, many young women die of a preventable disease.

Currently, screening for cervical cancer and its precursors is accomplished using the Papanicolaou (Pap) smear, which is unable to achieve a concurrently high sensitivity and high specificity; values for sensitivity and specificity range from 11% to 99% and from 14% to 97%, respectively. Results are typically available in two weeks; an abnormal Pap smear is followed by a colposcopy. In colposcopy a mounted magnifying lens is used to view the cervix after a 3%–6% acetic acid solution (vinaigre) is applied to increase the contrast between the normal cervix and areas of precancer, which turn white after application of acetic acid. Colposcopy provides excellent sensitivity (>90%), but poor specificity (<50%) for localizing cervical neoplasia, even in the hands of experienced practitioners. Because the specificity is poor, a directed biopsy is required to confirm the diagnosis, with results typically available one week later, thereby requiring a return visit for the patient if treatment is indicated. Currently, women often wait up to eight weeks to be treated as part of the standard of care in the diagnosis and treatment of the abnormal Pap smear.

Many groups have demonstrated that techniques based on quantitative optical spectroscopy have the potential to fulfill the need for improved screening and diagnosis of neoplasia. Optical measurement of tissue provides quantitative information that can be analyzed instantaneously and produce an objective diagnosis even in the hands of a nonexpert operator. Devices to make these measurements have
become inexpensive, robust, and portable because of advances in computing, fiber optics, and semiconductor technology.

Approaches based on reflectance, fluorescence, and Raman spectroscopies have shown potential for improved detection of epithelial neoplasia. In particular, investigations using reflectance spectroscopy indicate initial success in the diagnosis of neoplasia in the bladder, uterine cervix, and gastrointestinal tract. In reflectance spectroscopy, light emitted from tissue has undergone a combination of elastic scattering and absorption, providing information about tissue absorbers and scatterers, which are believed to change with neoplastic conversion.

The development of neoplasia is accompanied by local architectural changes at the cellular and subcellular levels, including (1) changes in the nuclear-to-cytoplasmic ratio of epithelial cells; (2) neovascularization; and (3) changes in stromal properties. These changes are hypothesized to affect the elastic scattering properties of tissue as follows. (1) Electromagnetic modeling predicts that the intensity of cellular light scattering increases with the progression of cervical precancer because of changes in nuclear size and DNA content. Mourant et al. measured changes in scattering in tumorigenic and nontumorigenic cells which confirmed increased scattering from tumorigenic cells from 500 to 790 nm. (2) Dallas et al. found that neoplastic lesions showed increased angiogenesis compared with adjacent normal cervical epithelium—a process that significantly progresses with the severity of the precancer. (3) Tromberg presented a pilot study that indicated that differences in the tissue absorption and scattering coefficients in the near-infrared (NIR) spectral regions can be used to discriminate between normal and precancerous cervical epithelia.

Other studies have shown that spatially resolved measurements of diffusely reflected light could be used to estimate tissue optical properties in vivo. It is well known that the depth of tissue penetration changes as the separation between the source and detector fibers is altered. In this paper, we explore the utility of spatially resolved reflectance spectra for the detection of cervical precancer in a preliminary analysis of data from a large diagnostic trial of in vivo fluorescence and reflectance spectroscopy with biopsy.

2 Materials and Methods

2.1 Instrumentation

The spectroscopic system used to measure reflectance spectra has previously been described in detail elsewhere. Briefly, the system consists of three main components: (1) a light source assembly that provides broadband excitation using a xenon arc lamp; (2) a fiber-optic probe that directs excitation light to tissue and collects diffusely reflected light; and (3) an optical assembly with a polychromator. Figure 1 illustrates the system. The probe, illustrated in Figure 2, utilizes nine optical fibers placed in direct contact with the tissue [200 μm diameter, numerical aperture (NA) = 0.2]. One fiber provides broadband illumination. Eight collection fibers at four different source–detector separations (position 0: 250 mm separation, position 1: 1.1 mm separation, position 2: 2.1 mm separation, position 3: 3.0 mm separation) collect diffusely reflected light.
2.2 Clinical Measurements
The study protocol was reviewed and approved by the Institutional Review Boards at the University of Texas M. D. Anderson Cancer Center and the University of Texas at Austin. Women who were 18 years of age or older, were not pregnant, and had been referred on the basis of an abnormal Pap smear were eligible. A health-care provider described the study to eligible patients; written consent was obtained from those agreeing to participate. Patients participating in the study received no incentive. Study participation included a Pap smear and colposcopic examination of the vulva, vagina, and cervix. The cervical smear was obtained using an Ayre’s spatula and a Cytobrush. Smears were routinely processed for clinical evaluation and submitted for Feulgen staining for quantitative evaluation. Human papilloma virus (HPV) testing was performed using hybrid capture and polymerase chain reaction (PCR) for negative specimens. In addition, for premenopausal women who were not on oral contraceptive pills (OCPs), the phase of menstrual cycle at the time of spectroscopic measurement was determined by serum estradiol, progesterone, follicle stimulating hormone (FSH), and luteinizing hormone (LH) levels.

Following colposcopic examination, but prior to biopsy, a fiber-optic probe was advanced through the speculum and placed in gentle contact with the cervix. Spectroscopic measurements were obtained from up to two colposcopically abnormal cervical sites, one colposcopically normal cervical site covered with squamous epithelium and; if visible, one colposcopically normal cervical site covered with columnar epithelium. Following spectroscopic measurements, all sites interrogated with the fiber-optic probe were biopsied.

Within 2 h of each patient measurement, standard spectra were measured. As a positive control, reflectance spectra were measured from a 1 cm pathlength cuvette containing a suspension of 1.02 μm diam polystyrene microspheres (6.25 vol %) to mimic the optical properties of tissue. As a negative control, reflectance spectra were measured with the probe tip immersed in a large container of distilled water. With the negative control, very low levels of reflectance are expected due to the small index mismatch between the optical fibers and the water. Low signal levels indicated acceptable levels of stray light, dark current, or other sources of background signal.

Biopsies were fixed and submitted for permanent section. Four-micron thick sections were stained with hematoxylin and eosin and then Feulgen stained. All routinely stained cytology and pathology specimens were submitted for diagnosis by experienced cytologists and pathologists, who were blinded to the results of the spectroscopy. Two cytologists read each Pap smear, and discrepant cases were reviewed a third time for consensus diagnosis by the study cytopathologist. Also, two pathologists read each biopsy, with discrepant cases reviewed a third time for consensus diagnosis by the study histopathologist. Diagnostic classification categories included normal tissues, HPV infection, grade 1 cervical intraepithelial neoplasia (CIN 1), grade 2 cervical intraepithelial neoplasia (CIN 2), grade 3 cervical intraepithelial neoplasia (CIN 3), and carcinoma in situ (CIS) using standard histopathologic criteria. Normal tissue was divided into two categories based on colposcopic impression: normal squamous (SN) epithelium and normal columnar epithelium (CN). LGSILs include HPV and CIN 1 and HGSILs include CIN 2, CIN 3, and CIS.

2.3 Data Processing and Statistical Analysis
Three investigators blinded to the pathologic results reviewed all spectra. Spectra indicating evidence of user or instrument error, such as probe slippage, were discarded from further analysis. To remove the effects of the source spectrum, variations in the illumination intensity, and the wavelength-dependent response of the detection system, we divided tissue spectra by the corresponding spectra measured from the microsphere standard. This normalization was performed at each source–detector separation. While this procedure results in spectra that describe the transport of light in tissue relative to that in microsphere suspension, the transport properties of microspheres are well known and can easily be described using Monte Carlo based models.

Reflectance data from a single measurement site are represented as a matrix containing calibrated reflectance intensity as a function of source–detector separation position and emission wavelength. Spectra for each of the four positions were column vectors containing 121 intensity measurements corresponding to emission wavelengths from 355 to 655 nm in 2.5 nm increments.

Reflectance spectra were then analyzed to determine which source–detector separations contained the most diagnostically useful information to separate each of the different types of tissue found in the cervix. An algorithm to separate each pairwise combination of diagnostic categories was developed and evaluated relative to the gold standard of colposcopically directed biopsy. 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 multistep classification algorithms to determine the tissue type of an unknown sample based on its reflectance spectrum. Algorithm development consisted of the following steps: (1) selection of the source–detector separations to analyze; (2) data reduction using principal component analysis (PCA); (3) feature selection and classification using Mahalanobis distance with cross validation. Each step is described in detail below.

To identify the optimal combination of source–detector separations, we evaluated all possible combinations when taken one, two, three, or four at a time. There were a total of 15 combinations considered—four of one, six of two, four of three, and one of four.

Prior to PCA, vectors containing the reflectance spectra for the source–detector separations of interest were concatenated into a single vector. A data matrix was assembled from these vectors, including all measurements from the two diagnostic categories being compared. Eigenvectors of the corresponding covariance matrix were then calculated; those accounting for up to 65%, 75%, 85%, and 95% of the total variance were retained for algorithm development. We denote the fraction of the total variance accounted for as the eigenvalue significance level (ESL). For each eigenvector, we calculated a principal-component score for each sample in the data matrix.

Classification functions were then formed to assign a sample to one of the two given classes. Classification was based on the Mahalanobis distance, $r^2$, the multivariate measure of the separation of a data point $x_i$ from the mean ($x_i$)
of a dataset in \( n \)-dimensional space [Eq. (1)]. This distance is given as

\[
r^2 = (x - x_m)^\top C_x^{-1} (x - x_m),
\]

where \( C_x \) is the covariance matrix. The multivariate distance between the measurement to be classified and the means of the two histopathologic categories were calculated, and the sample was assigned to the group it was closest to in multivariate space.

The performance of classification depends on the principal-component scores included for analysis. From the available pool of eigenvectors the single principal-component score yielding the best initial performance was identified, then the score that improved this performance most was selected. This process was repeated until performance was no longer enhanced by the addition of principal components, or until all components were selected. The sensitivity and specificity of diagnostic algorithms for each combination of source–

Table 1 Reflectance data set by histopathologic category (colposcopically directed biopsy gold standard).

<table>
<thead>
<tr>
<th>Diagnostic class</th>
<th>SN</th>
<th>CN</th>
<th>HPV</th>
<th>CIN 1</th>
<th>CIN 2</th>
<th>CIN 3/CIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sites</td>
<td>227</td>
<td>18</td>
<td>52</td>
<td>9</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

Fig. 3 Average reflectance spectra by tissue diagnostic classification for different source–detector separations: (a) position 0; (b) position 1; (c) position 2; (d) position 3. Diagnostic classifications: SN = ··· ···; CN = ··· ···; HPV = ··· ···; CIN 1 = ···; CIN 2 = ··· ···; CIN 3/CIS = ···. Error bars indicate ± one standard deviation. Tissue spectra were normalized by standard spectra from a suspension of 6.25 vol % polystyrene microspheres (1.02 \( \mu \text{m} \) diameter).
3 Results

3.1 Data Set

The data set consisted of spectra from 324 sites in 161 patients that were deemed adequate. Table 1 indicates the number of measurements within each diagnostic category. Tissue with acute or chronic inflammation or metaplasia was included in the corresponding squamous or columnar normal category.

3.2 Reflectance Spectra

Average reflectance spectra for each diagnostic category for the four different source–detector separation positions are shown in Figure 3. Each position (0, 1, 2, and 3) corresponds to an increasingly greater source–detection separation, which probes increasingly greater tissue depth. All spectra show valleys due to hemoglobin absorption at 420, 542, and 577 nm. As the source–detector separation increases, the relative level of elastic scattering in the longer wavelengths increases due to the increased penetration depth of light. Average spectra of each diagnostic class differ for all source–detector separations. At all separations, the mean reflectance of squamous normal tissue is most intense; there is gradual decrease in mean reflectance intensity from HPV to CIN 1 to CIN 2 to CIN 3, and the mean reflectance of columnar normal tissue is least intense. The greatest separation between average spectra of different diagnostic classes occurs at position 0.

3.3 Statistical Analysis

Figure 4 shows the sensitivity and specificity for the best performing combination of one, two, three, and four source–detector separations at an ESL of 65%. Results are shown separately for each pairwise combination of diagnostic categories. The diagnostic performance is high when using only a single separation, and does not noticeably increase when data from additional separations are included. The best performance is obtained when discriminating between SN and CN, with sensitivity of 89% and specificity of 77%, and CN and LGSIL, with sensitivity of 75% and specificity of 89%. It is most difficult to discriminate between SN and LGSIL. Increasing the ESL from 65% to 95% does not result in an increase in performance (Figure 5).

Table 2 gives the sensitivities and specificities for the combinations of source–detector separations that yielded the best performance. Many combinations of positions gave equally good results for discrimination between SN and CN and between CN and LGSIL. Positions 0 and 1 appear in all of the optimal combinations for discriminating between tissue types, except for SN vs HGSIL.

4 Discussion

In our study of the diagnostic potential of reflectance spectroscopy, we obtained cervical in vivo measurements at four distinct source–detector separation positions. Using Mahalanobis distance classification, we determined which source–detector separations contained the most diagnostically useful
information. The results showed the sensitivity and specificity to be high when using a single source–detector separation at the lowest level of eigenvector significance considered, and do not noticeably increase when data from additional separations or ESLs are included. Furthermore, the smaller source–detector separations of positions 0 and 1 appear more frequently than the greater separations of positions 2 and 3 in the positional combinations that performed best. Positions 0 and 1 seem effective for all pairwise discrimination except for SN vs HGSIL.

Our results show that HGSIL can be discriminated from squamous and columnar normal tissue with high sensitivity (72% and 81%, respectively) and high specificity (89% and 89%, respectively). These results indicate slightly lower sensitivity and improved specificity for discrimination of SN vs HGSIL found in Ref. 43 where sensitivity of 82% and specificity of 67% were obtained using analysis of a single reflectance spectrum. Furthermore, we found that LGSIL could be separated from columnar normal tissue with similar high sensitivity (75%) and specificity (89%); however, discrimination between LGSIL and squamous normal tissue was more difficult using reflectance spectroscopy in this study. In all cases, the specificity associated with reflectance spectroscopy was significantly higher than that previously reported for colposcopy (48%).

The sensitivities and specificities reported for reflectance spectroscopy here are based on leave-one-out cross validation. Ideally, separate training and validation sets should be used to provide unbiased estimates of algorithm performance; however, previous work using fluorescence spectroscopy has shown that cross validation provides a good estimate of the performance with a separate validation set.

### Table 2: Source–detector separation positional combinations that give the best overall sensitivity and specificity values for each histopathologic category pairwise combination, with the eigenvector significance levels specified. SN = squamous normal, CN = columnar normal, LGSIL = low-grade squamous intraepithelial lesion, HGSIL = high-grade squamous intraepithelial lesion.

<table>
<thead>
<tr>
<th>Diagnostic pairwise combination</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>ESL</th>
<th>No. of source–detector positions in combination</th>
<th>Corresponding source–detector positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN vs CN</td>
<td>89</td>
<td>77</td>
<td>0.65</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SN vs CN</td>
<td>89</td>
<td>77</td>
<td>0.65</td>
<td>2</td>
<td>0, 1</td>
</tr>
<tr>
<td>SN vs CN</td>
<td>89</td>
<td>77</td>
<td>0.65</td>
<td>3</td>
<td>0, 1, 2</td>
</tr>
<tr>
<td>SN vs CN</td>
<td>89</td>
<td>77</td>
<td>0.75</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SN vs CN</td>
<td>89</td>
<td>77</td>
<td>0.75</td>
<td>2</td>
<td>0, 1</td>
</tr>
<tr>
<td>SN vs CN</td>
<td>89</td>
<td>77</td>
<td>0.75</td>
<td>3</td>
<td>0, 1, 2</td>
</tr>
<tr>
<td>SN vs LGSIL</td>
<td>72</td>
<td>56</td>
<td>0.95</td>
<td>4</td>
<td>0, 1, 2, 3</td>
</tr>
<tr>
<td>SN vs HGSIL</td>
<td>72</td>
<td>81</td>
<td>0.95</td>
<td>2</td>
<td>2, 3</td>
</tr>
<tr>
<td>CN vs LGSIL</td>
<td>75</td>
<td>89</td>
<td>0.65</td>
<td>2</td>
<td>0, 1</td>
</tr>
<tr>
<td>CN vs LGSIL</td>
<td>75</td>
<td>89</td>
<td>0.65</td>
<td>3</td>
<td>0, 1, 2 and 0, 1, 3</td>
</tr>
<tr>
<td>CN vs LGSIL</td>
<td>75</td>
<td>89</td>
<td>0.75</td>
<td>2</td>
<td>0, 1</td>
</tr>
<tr>
<td>CN vs LGSIL</td>
<td>75</td>
<td>89</td>
<td>0.75</td>
<td>3</td>
<td>0, 1, 2 and 0, 1, 3</td>
</tr>
<tr>
<td>CN vs LGSIL</td>
<td>75</td>
<td>89</td>
<td>0.85</td>
<td>2</td>
<td>0, 1</td>
</tr>
<tr>
<td>CN vs LGSIL</td>
<td>75</td>
<td>89</td>
<td>0.85</td>
<td>3</td>
<td>0, 1, 2 and 0, 1, 3</td>
</tr>
<tr>
<td>CN vs HGSIL</td>
<td>72</td>
<td>83</td>
<td>0.65</td>
<td>3</td>
<td>0, 1</td>
</tr>
<tr>
<td>CN vs HGSIL</td>
<td>72</td>
<td>83</td>
<td>0.75</td>
<td>3</td>
<td>1, 2, 3</td>
</tr>
</tbody>
</table>
Given the single-cell epithelial thickness of columnar tissue, it is not surprising that its mean reflectance intensity is lower than the multilayered squamous tissues at the various normal and dysplastic states. The single-layer epithelium allows more light to reach and be absorbed by hemoglobin in stromal tissue. All other diagnostic classes are multilayered squamous epithelium and, as such, show a stronger mean reflectance intensity, which decreases with progression of disease. This is consistent with the results of Dellas et al., who found increased angiogenesis with increased severity of precancer.40 Using a Monte Carlo model of light transport in normal cervical epithelium, we estimate that, at 420 nm, the average penetration depth of photons detected at position 0 is 450 μm and at position 3 is 550 μm. Similarly, at 500 nm, the average penetration depth of photons detected at position 0 is 680 μm and at position 3 is 1180 μm. These simulations assumed an epithelial thickness of 350 μm and values of μs.epithelium = 105 cm⁻¹, μa.epithelium = 3 cm⁻¹, μs.stroma = 280 cm⁻¹, μa.stroma = 40 cm⁻¹ at 420 nm and μs.epithelium = 82 cm⁻¹, μa.epithelium = 2 cm⁻¹, μs.stroma = 230 cm⁻¹, and μa.stroma = 4 cm⁻¹ at 500 nm. In these simulations, the anisotropy factor (g) of epithelium and stroma were 0.95 and 0.88, respectively. The NA of the optical fiber was 0.22. The refractive indices of the epithelium and the stroma were assumed to be 1.4. The code was verified with the results of Welch et al.32 for the two-layer model and with those of Mourant et al.,48 for detector modeling. By keeping track of the photon position in tissue after each scattering event, the maximum depth the photon reaches can be calculated.

5 Conclusions

Previously, we examined the use of fluorescence spectroscopy alone to discriminate cervical precancer using 340, 380, and 460 nm excitation. We found sensitivity of 79% and specificity of 78% to discriminate HSIL from all other tissue types.31 This performance is comparable to that found for the pairwise diagnostic-category comparisons performed using reflectance spectroscopy alone in this study. Our previous studies using fluorescence spectroscopy indicate that stepwise diagnostic algorithms are required in order to determine the tissue type of an unknown sample based on its fluorescence spectrum because of the large differences in optical properties of squamous and columnar cervical tissue.20 A parallel strategy for reflectance spectroscopy and the pairwise analysis presented here provide the foundation for this analysis. In a similar analysis, we have examined fluorescence excitation emission matrices (EEMs) also for discrimination of all diagnostic categories.36 It is interesting to note that reflectance spectroscopy gives good performance for discriminating categories where fluorescence alone performs less well. Analogous results were noted in Ref. 43. Thus, fluorescence and reflectance spectroscopy appear to be complementary diagnostic technologies, and future studies should pursue combined analysis.

Acknowledgments

The authors gratefully acknowledge the contributions of Dizem Arifler (Monte Carlo modeling); the clinical research staff (Christy Whitmore, Joanne Baker, Kim Hagedorn, Christina Amos, Glenda Dickerson, and Patricia Trigo) and nurse colposcopists (Judith Sandella, Alma Sbach, and Karen Rabbel); the data managers Nan Earle and Trey Kell; and editor/assistant Stacy Crain. Financial support from the National Cancer Institute (PO1-CA82710) is gratefully acknowledged.

References

1. http://www.nccn-online.org


594 Journal of Biomedical Optics • October 2002 • Vol. 7 No. 4