Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

Detecting cervical cancer progression through extracted intrinsic fluorescence and principal component analysis

Seema Devi
Prasanta K. Panigrahi
Asima Pradhan

SPIE

Detecting cervical cancer progression through extracted intrinsic fluorescence and principal component analysis

Seema Devi,[a] Prasanta K. Panigrahi,[b] and Asima Pradhan[a,c,]
[a]Indian Institute of Technology, Department of Physics, Kanpur 208016, India
[b]Indian Institute of Science Education and Research (IISER), Kolkata 741246, India
[c]Indian Institute of Technology, Center for Lasers and Photonics, Kanpur 208016, India

Abstract. Intrinsic fluorescence spectra of the human normal, cervical intraepithelial neoplasia 1 (CIN1), CIN2, and cervical cancer tissue have been extracted by effectively combining the measured polarized fluorescence and polarized elastic scattering spectra. The efficacy of principal component analysis (PCA) to disentangle the collective behavior from smaller correlated clusters in a dimensionally reduced space in conjunction with the intrinsic fluorescence is examined. This combination unambiguously reveals the biochemical changes occurring with the progression of the disease. The differing activities of the dominant fluorophores, collagen, nicotinamide adenine dinucleotide, flavins, and porphyrin of different grades of precancers are clearly identified through a careful examination of the sectorial behavior of the dominant eigenvectors of PCA. To further classify the different grades, the Mahalanobis distance has been calculated using the scores of selected principal components. [DOI: 10.1117/1.JBO.19.12.127003]

Keywords: tissue; intrinsic fluorescence; precancers; principal component analysis.

Paper 140432RR received Jul. 5, 2014; accepted for publication Nov. 17, 2014; published online Dec. 11, 2014.

1 Introduction

Early detection of cancer can significantly improve chances of patient survival. Conventional techniques are often not sensitive enough to detect subtle changes that occur in the early stages of disease. Optical methods, however, are known to be sensitive to biochemical and morphological changes that occur during the progress of the disease. Biochemical changes occurring at a cellular level originate from biomolecules, e.g., nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FAD), and its derivatives and collagen.[1,2] Both NADH and FAD play a major role in the metabolic activities of the cell, whereas structural changes in the tissue are produced by changes of the collagen.[3,4] The ratio of NAD+ and NADH is known to change significantly in diseased tissue. NADH is bound to proteins in normal tissues as compared to the unbound one in tumors, exhibiting different behaviors as a result.[5–7] The collagen network, which forms the supportive backbone in the stroma, systematically breaks down with the progress of the disease.[8,9] These changes result in differing emission properties of the tissue in normal and diseased conditions and have been captured through fluorescence spectroscopy and imaging.[10,11] However, absorption by blood and scattering from cells and nuclei in tissue result in masking the weak fluorophore activities, making it difficult to observe these changes. Hence, isolating these environmental effects will undoubtedly help in identifying the differing optical activities, particularly in the early stage of cancer where these changes are expected to be small.

Various techniques have been used to correct for the attenuation due to scatterers and absorbers from the measured fluorescence of the tissue samples. Schemes developed with judicious combinations of measurements like an intensity ratio of the fluorescence to scattered light,[12] ratios of diffuse reflectance,[13,14] and ratios of autofluorescence[15,16] at different wavelengths are used to extract intrinsic fluorescence. However, their accuracy in correcting for the attenuation is limited since the wavelength dependence of optical parameters is not taken into account.[1] In this direction, several groups have developed transfer functions in terms of optical transport parameters of tissue[17] using Kubelka-Munk theory,[18,19] diffusion theory,[20,21] photon migration theory,[22–25] and Monte Carlo simulations.[26] Most of these techniques apply only to a limited wavelength range (500 to 800 nm), where the absorption coefficient (μ_a) is less dominant than the reduced scattering coefficient (μ_s'). The photon migration theory covers the entire bio-optically relevant wavelength region, which includes UV ($\mu_a \geq \mu_s'$), visible, and near-infrared ($\mu_a \ll \mu_s'$), however, prior knowledge of the optical parameters of the targeted samples are necessary.[23,25] Liu et al.[27] reported the development of a compact point-detection fluorescence spectroscopy system and an empirical ratiometric method to quantify the intrinsic fluorescence redox ratio and hemoglobin concentrations in *in vitro* orthotopic brain tumor rat model. We make use of a novel approach, developed earlier by our group, in effectively combining the measured polarized fluorescence with the polarized elastic scattering to remove the effects of absorption and diffuse scattering.[28,29] The main advantage of this correction technique is the recovery of both line shape and intensity information in a turbid medium with reasonable accuracy for the range of optical transport parameters,[28] where the ratio of μ_a/μ_s' lies between 0 and 1.8.

*Address all correspondence to: Asima Pradhan, E-mail: asima@iitk.ac.in

0091-3286/2014/$25.00

Decoupling the effects of scatterers and absorbers improves diagnostic efficacy since subtle biochemical changes that take place during the progress of tumors are visible in the intrinsic fluorescence. To further highlight these subtle changes, a principal component analysis (PCA) is applied which disentangles the collective behavior from smaller correlated clusters in a dimensionally reduced space. The use of PCA with suitable classification functions has been found useful in differentiating various classes of human tissues.[30–42] It is worth pointing out that Ramanujam et al. have developed a multivariate statistical algorithm for the screening and diagnosis of cervical precancers.[43,44]

To illustrate the general applicability of our approach, we focus here on cervical precancers. Studies on the detection of cervical tissue precancers have been pursued by several groups using different optical techniques and statistical methods, both *in vivo* and *in vitro*.[31,45–54] *In vivo* measurements by Nordstrom et al. have shown that UV fluorescence provides a better discrimination of precancers from normal squamous cervical tissue than diffuse reflectance.[45] A sensitivity of 86% and 87% specificity has been achieved by them in normal versus CIN1 and 91% sensitivity and 93% specificity for normal versus CIN2/3. Huh et al. have discriminated the high grades of CIN 2/3 with 90% sensitivity using *in vivo* data from both fluorescence and reflectance.[46] A binary classification of different sets of cervical tissue is reported by Kortum et al. They have reported a better discrimination with combined reflectance and fluorescence measurements (83% sensitivity and 80% specificity) compared to exclusively using reflectance spectroscopy in *in vivo* studies of squamous normal and high grades squamous intraepithelial lesions (HGSIL) of cervical tissue.[31,47] Isolating chronic inflammation from precancers is necessary for accurate detection. Rodero et al. have discriminated chronic inflammation and low grades of squamous intraepithelial lesions from HGSIL with 89% sensitivity and 100% specificity using fluorescence spectroscopy.[48] Other optical techniques like optical coherence tomography[49] and Raman spectroscopy[50,51] have also been utilized for cervical precancer detection. The absorption coefficient and reduced scattering coefficient obtained through *in vivo* diffuse reflectance spectroscopy measurements and Monte-Carlo simulations by Ramanujam et al. have shown differences among normal, CIN1, and CIN2 cervical tissue.[52] Fluorescence from body fluids have also displayed a potential for early detection.[53,54] Most of the groups mentioned above use a binary classification where all classes (CIN1, CIN2, CIN3, and normal) are not always discriminated with respect to one another. Here, we report differences in normal, CIN1, CIN2, and cancerous tissues with respect to one another, highlighting the extracted molecular information.

In this study, the merit of coupling experimentally extracted intrinsic fluorescence with PCA is clearly demonstrated in cervical precancers. Elimination of both diffuse scattering and absorption effects unambiguously reveals signatures of fluorophores in intrinsic fluorescence spectra as has been validated in our earlier work on tissue mimicking phantoms.[28] The combined approach highlights subtle biochemical changes in the activities of NADH, flavins, and collagen emission properties[55] for different grades of human cervical precancer. An autocorrelation and random-matrix approach[56–58] identifies the wavelength domain where these activities are most prominent. This not only helps identify the underlying fluorophores, but also vividly captures the most sensitive marker that indicates disease progression. Careful examination of the sectorial behavior of the dominant eigenvectors reveals the relative contributions of NADH, flavins, collagen, and porphyrin. To further classify, the Mahalanobis distance has been calculated using the scores of selected principal components.

2 Methods and Materials

2.1 *Data Acquisition*

A commercial spectrofluorometer (SPEX, Fluorolog 3, Model FL3-22), which uses a 450 W xenon lamp as the source, is used for acquisition of data from cervical tissues. The polarized fluorescence spectra of fresh cervical tissue specimens are recorded with excitation at 370 nm. All measurements are made at an angle of 22.5 deg with respect to the incident polarized light. The epithelial layer of the tissue is exposed to the incident light during the recording of the spectra. The scanning range of polarized fluorescence is 400 to 609 nm with an integration time of 0.1 s and a spectral resolution of 1 nm. In addition, corresponding polarized elastic scattering spectra are also recorded. Keeping the excitation polarizer vertical, copolarized (VV) and cross-polarized (VH) fluorescence and corresponding polarized elastic scattering are recorded with the emission polarizer in the vertical and horizontal positions, respectively.

2.2 *Protocol Used to Extract Intrinsic Fluorescence*

The intrinsic fluorescence is extracted by using the measured polarized fluorescence and polarized elastic scattering data, which remove the effects of diffuse scattering and absorption. The intrinsic fluorescence intensity ($I_{\text{int.fl.}}$) is given by

$$I_{\text{int.fl.}} = \frac{\{[I_{\text{VV}}(\lambda) - G(\lambda) \times I_{\text{VH}}(\lambda)]_{\text{fl}}\}}{[R]^{x}\{[I_{\text{VV}}(\lambda) - G(\lambda) \times I_{\text{VH}}(\lambda)]_{\text{scatt}}\}^{(1-x)}}. \quad (1)$$

Here, subscripts fl and scatt refer to the fluorescence and elastic scattering, respectively, with I_{VV} and I_{VH} representing the intensities of co- and cross-polarized channels. $G(=I_{\text{HV}}/I_{\text{HH}})$ corrects for the difference in instrumental response to vertically and horizontally polarized light.[55] Elastic scattering $(I_{\text{VV}} - G \times I_{\text{VH}})_{\text{scat}}$ at the excitation wavelength is denoted by R. On an average, fluorescence and elastic scattering contribute equally to the entire path traversed by the excitation photon inside the tissue.[24,29] The factor $x = 0.5$ takes into account this contribution. The validation of this technique has been previously reported on tissue mimicking phantoms.[28]

2.3 *Sample Collection and Handling*

This *in vitro* study involved fresh human cervical tissue provided by G.S.V.M. Medical College, Kanpur, within 2 h of excision. All samples are taken from hysterectomy cases and a written consent form is obtained from each patient by the doctors involved in this study. The protocol number for this work from the Institute Ethical Committee (IEC) of IIT Kanpur is IITK/IEC/2012-13/1/3. These samples are labeled as normal and abnormal on the basis of visual inspection. Sample dimensions are ~2 mm × 5 mm. The samples are rinsed with saline water to remove superficial blood before performing the experiment. Data acquisition time for each measurement is 3 min. After completing the signal acquisition, these tissue samples are sent back to the medical college for histopathology analysis

and to confirm classification as normal, CIN1, CIN2, and cancerous tissue.

2.4 *Data Analysis*

Polarized fluorescence spectral data, collected from 99 patients of varying ages and economic backgrounds, are categorized into different groups as normal (21 samples), CIN-1 (47 samples), CIN-2 (22 samples), and cancerous (9 samples) on the basis of histopathological reports. Intrinsic fluorescence is extracted for emission wavelengths from 400 to 609 nm using the earlier mentioned protocol. Subsequent to the extraction, a principal component analysis (PCA) is applied to the intrinsic fluorescence spectra for the dimensional reduction of the data without loss of its features. In addition to dimensional reduction, PCA identifies the spectrally prominent regions through the eigenvector components of the dominant eigenvalues. It also characterizes the noise using results from random matrix theory, making the spectral features unambiguous.[42,57] To identify the principal components (eigenvectors) of the spectra containing maximum variance of the complete data, an empirical correlation matrix C is constructed.[39–42]

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N} \delta I_i^T(k)\delta I_j(k) = \frac{A^T A}{N}. \tag{2}$$

Here, $\delta I_j(k) = A_{jk}$ is the mean subtracted intensity divided by the standard deviation, computed over the samples at each wavelength. Index i varies from 1 to 210, representing the wavelength range from 400 to 609 nm and k gives the number of samples. Subsequently, the principal components PC1, PC2, and PC3 corresponding to the first three highest eigenvalues of the correlation matrix are calculated, which capture ~99% variance of the original data. These eigenvalues lie well above the noise threshold.[36] Further, probability density functions are used to study the scores of the principal components for careful analysis of the distribution. In order to estimate the probability density function f of the scores x, a kernel-smoother density estimation is used:[40]

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{3}$$

where the entire data are divided into $n = 100$ equally spaced points x_i. Here h is the smoothing parameter, known as bandwidth, and K is the normal distribution given as

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-(x-x_i)^2/2\,h^2}. \tag{4}$$

Scores of the selected principal components of each tissue class are then divided into training and validation data sets. Mahalanobis distances of validation set data points (v) from the mean of training set data points (t_m) are calculated as given below:[31,45]

$$r = \sqrt{(v - t_m)' c_t^{-1} (v - t_m)}, \tag{5}$$

where c_t is the covariance matrix of training set data points. The different tissue types are classified on the basis of r. Ten samples each from normal and CIN2, 24 samples from CIN1, and 4 cancer samples are randomly selected for inclusion in the training data sets. The remaining 11 normal, 23 CIN1, 12 CIN2, and 5 cancer samples are used as the validation data sets. Mahalanobis distances are computed using training data sets for each grade, and these distances are then used to classify the validation sets of the corresponding grades. This process is repeated twice taking randomly selected samples in training and validation data sets, and this made no significant difference in classification. Finally, the above-computed Mahalanobis distance is used to classify the complete data set (which included both training and validation data sets).

3 Results and Discussion

Figures 1(a) and 1(b) display co- and cross-polarized components of the fluorescence and of the elastic scattering spectra of cervical tissue, respectively. Area normalized intensities of copolarized, cross-polarized, and intrinsic fluorescence spectra are shown in Fig. 1(c). The cross-polarized component travels a larger path inside the tissue and bears more absorption and diffuse-scattering effects as compared to the copolarized component. As a consequence, the absorption dips at 410 nm are enhanced in the cross-polarized components, as seen in Fig. 1. The removal of the dip in the intrinsic fluorescence spectrum is clearly visible in Fig. 1(c).

The technique, thus, successfully extracts intrinsic fluorescence by eliminating the effects of absorption and diffuse scattering. This can now be used to study the biochemical behavior of normal and different grades of cervical tissue samples with further statistical analysis. Figure 2(a) shows the area normalized mean intrinsic fluorescence spectra of normal, CIN-1, CIN-2, and cancerous cervical tissue samples. The regions in the wavelength ranges from 400 to 420 nm and 452 to 470 nm are attributed to the contribution from collagen and NADH, respectively.[3–9] These ranges are magnified in

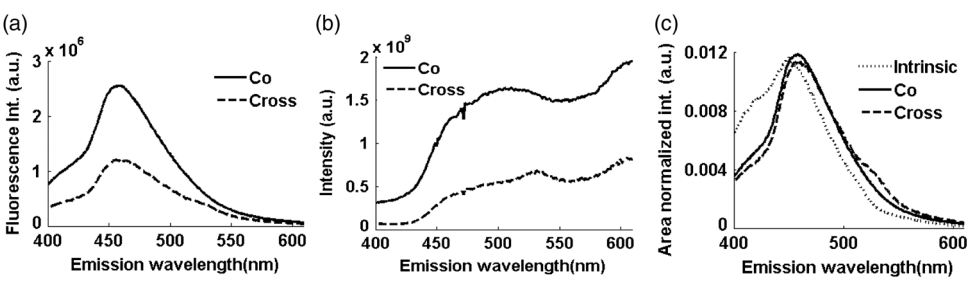

Fig. 1 Typical plots of co- and cross-polarized components of (a) fluorescence spectra, (b) elastic scattering spectra, and (c) typical area normalized spectra of copolarized, cross-polarized, and intrinsic fluorescence emission for human cervical tissue.

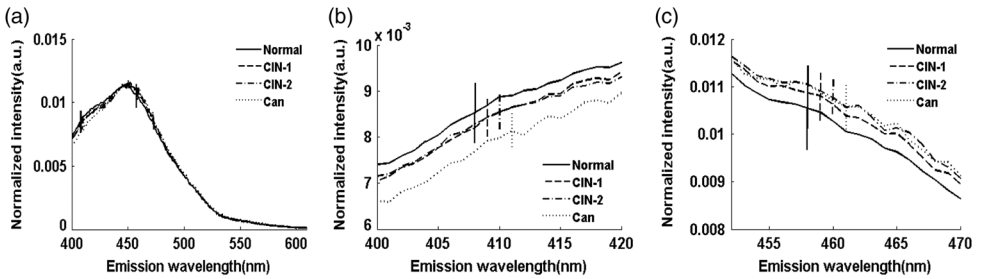

Fig. 2 Area normalized mean spectra of intrinsic fluorescence for normal and different grades of cervical tissues in the emission wavelength range of (a) 400 to 609 nm, (b) 400 to 420 nm, and (c) 452 to 470 nm. Figures 2(b) and 2(c) [magnified view of Fig. 2(a)] show the collagen and nicotinamide adenine dinucleotide behavior in the intrinsic fluorescence of normal and different grades of cervical tissues. Error bars in the figures show standard deviation.

Figs. 2(b) and 2(c) to highlight the reduction of collagen and increase in NADH contributions with development of the disease, respectively. The reduced collagen emission corresponds to the breaking of collagen matrix structures and the increased NADH emission is due to the enhanced metabolic activity of cells with the growth of abnormal cells.[5–9] It is pertinent to note at this point that even on extraction of the intrinsic fluorescence, the spectra do not display significant changes as the disease progresses.

Figures 3(a), 3(e), and 3(f) depict the correlation maps for intrinsic, co- and cross-polarized fluorescence components of the normal tissues. With 370-nm excitation, the spectra are scanned from 400 to 609 nm. The wavelength shift in the correlation map is calculated from 400 nm. Significantly high

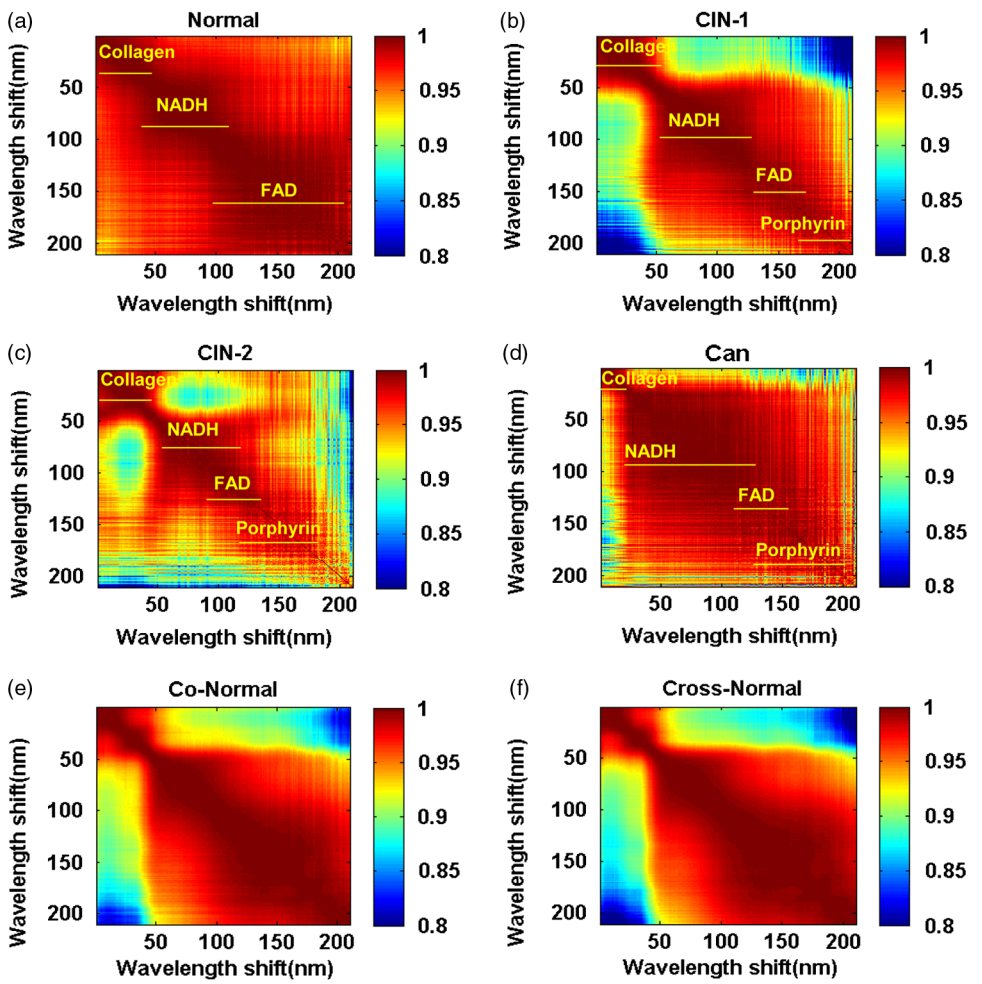

Fig. 3 Correlation maps of intrinsic fluorescence for (a) normal, (b) CIN-1, (c) CIN-2, and (d) cancerous cervical tissue. (e) and (f) show co- and cross-polarized fluorescence correlation matrices for normal cervical tissue. One observes well-demarcated domains in normal and different grades of precancer in the correlation maps.

Table 1 First, second, and third highest eigenvalues for normal and different grades of cervical tissues showing distinct differences between tissue types.

Eigenvalues	Normal	CIN-1	CIN-2	Can
First	206.0	199.2	198.1	203.7
Second	2.5	7.8	4.6	3.1
Third	0.9	1.6	3.5	2.0

correlation domains are seen in the correlation map of intrinsic fluorescence compared to the domains of co- and cross-polarized components. The high correlation occurs because the fluorescence is now free from the effects of absorption and diffuse scattering. The importance of extracting intrinsic fluorescence is clearly visible in these correlation maps where the domains of co- and cross-polarized components [Figs. 3(e) and 3(f)] are not as distinct as in the intrinsic case. The correlation maps of intrinsic fluorescence of CIN-1, CIN-2, and cancerous cervical tissues are displayed in Figs. 3(b), 3(c), and 3(d). In the correlation map of the normal tissue, three distinct domains are visible. These correspond to the fluorescence emission of collagen (400 to 420 nm), NADH (452 to 470 nm), and FAD (500 to 550 nm), as is evident from the wavelength regions they occupy.

The size of the first domain, which corresponds to the activity of collagen, decreases monotonically when going from normal to CIN-1, CIN-2, and cancer samples. The second domain corresponding to NADH activity is red-shifted in the case of CIN-1 and CIN-2 relative to the normal one and is most dominant in the correlation map of the cancer case. The contribution of the third domain, which represents the activity of FAD and its derivatives, reduces sequentially in the abnormal tissues relative to the normal ones. A small domain appears at approximately 580 nm in CIN-2 and cancer cases; this may be due to porphyrin contributions. There is evidence of porphyrin fluorescence in breast tumors from recent studies.[59–61] Thus, in spite of a large standard deviation as seen in Fig. 2, distinct differences are still observed in the corresponding correlation maps of intrinsic fluorescence. This demonstrates the strength of PCA in highlighting the molecular information of already corrected spectra.

We now proceed to select the significant eigenvalues and corresponding eigenvectors of the correlation matrices, which contain the maximum information of the original data. The first three highest eigenvalues chosen are listed in Table 1. Other eigenvalues have been found to be quite small and, hence, have not been considered. The eigenvectors (principal components) corresponding to these three highest eigenvalues are used to represent the complete data. This is evident in the spectra, reconstructed using eigenvectors corresponding to the first two and the first three highest eigenvalues and shown in Figs. 4(a) and 4(b), respectively. The reconstructed spectrum captures the original spectrum more accurately when the eigenvector corresponding to the third eigenvalue is included.

Probability densities of the PC1, PC2, and PC3 scores have been systematically analyzed, as shown in Figs. 5(a), 5(b), and 5(c), respectively. A kernel-smoother has been used to get the normal distributions of the scores of PC1, PC2, and PC3. It is observed that probability distribution curves of the scores of PC1, PC2, and PC3 are able to separate different tissue grades. The distribution curves of these groups have some overlap. However, the peak separation of the distribution curves of PC3 scores is more distinct as compared to the other two, bringing out the subtle differences more distinctly.

The principal components can be visualized as new coordinate axes along which the given data are optimally represented. Hence, any parametric relation between these can have empirical implications. Considerably different curvatures are observed in the three-dimensional scatter plots of PC1, PC2, and PC3 for different tissue types, as is clear in Fig. 6(a). Scatter plots for CIN-2 and cancerous tissue types are more dispersed compared to normal and CIN-1. The probability density function and scatter plots both show remarkable differences among different tissue types.

It is evident that the eigenvector entries and their distributions provide valuable insight into the relative contribution of different fluorophores to the emission spectra. The eigenvector entries of PC1, PC2, and PC3 corresponding to the first three largest eigenvalues are shown in Fig. 7 for normal and different grades of cancerous tissue. As is well known, the dominant eigenvalue corresponds to the collective motion, akin to the motion of the center of mass. It can be observed that PC1 bears the reflection

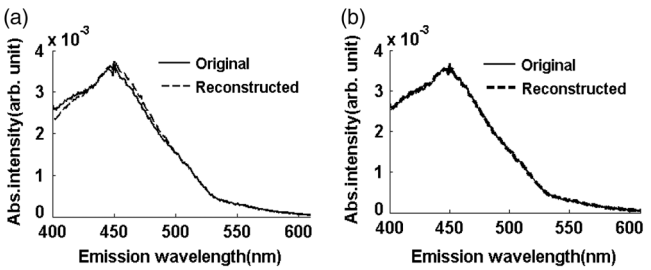

Fig. 4 Typical reconstructed spectrum using the eigenvectors corresponding to first (a) two and (b) three highest eigenvalues with the original spectrum of normal tissue. It is clear that, together, these three capture the maximum variance.

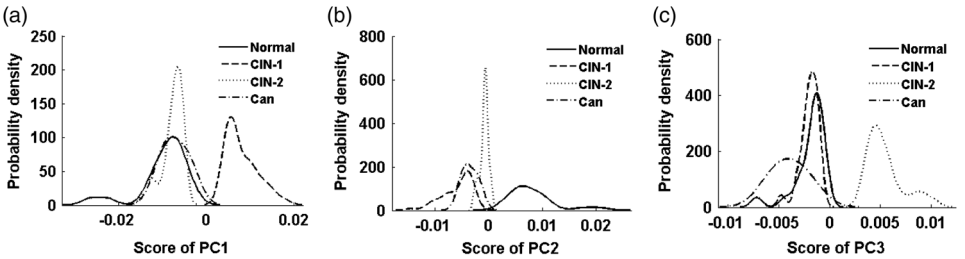

Fig. 5 Probability density for the scores of (a) PC1, (b) PC2, and (c) PC3 of intrinsic fluorescence intensity revealing significant differences between the tissue types.

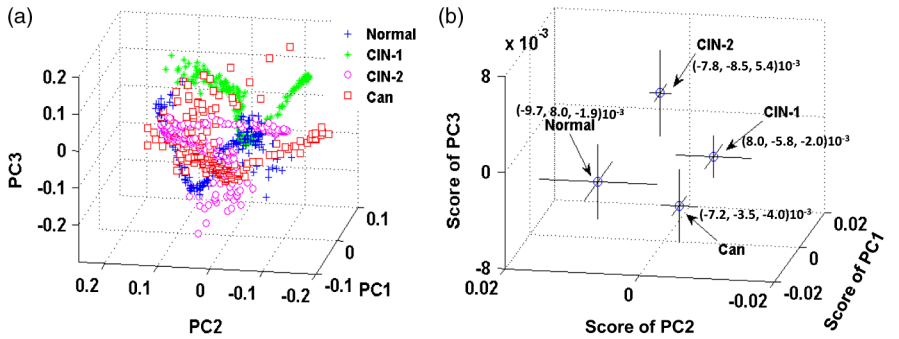

Fig. 6 (a) Three-dimensional scatter plots and (b) means of scores with error bars at 95% confidence intervals of PC1 versus PC2 versus PC3 for normal, CIN-1, CIN-2, and cancerous cervical tissue.

of the structures in the correlation matrices as it retains most of the information of the original data. These modes in a financial time series are known as market modes, representing the correlated behavior of the entire market.[56] It is clear from Fig. 7(a), which depicts the eigenvector entries of the dominant eigenvalue, that these are broadly distributed and show characteristic collective differences in the structure of the correlation matrices. The other smaller eigenvalues provide information about the subdominant correlated sectors. In the context of a financial time series, these refer to correlated sectors arising from the cluster of companies of similar kind. In the present context, these refer to different fluorophore activities and their interrelations. To systematically identify the correlated substructures, the eigenvalue entries are plotted with prior demarcation of different sectors corresponding to the activity of different fluorophores, namely collagen (400 to 420 nm), NADH (452 to 470 nm), FAD (~530 nm), and porphyrin (~580 nm).

The eigenvector entries corresponding to the second and third largest eigenvalues shown in Figs. 7(b) and 7(c) clearly segregate the domains of activity of different fluorophores.

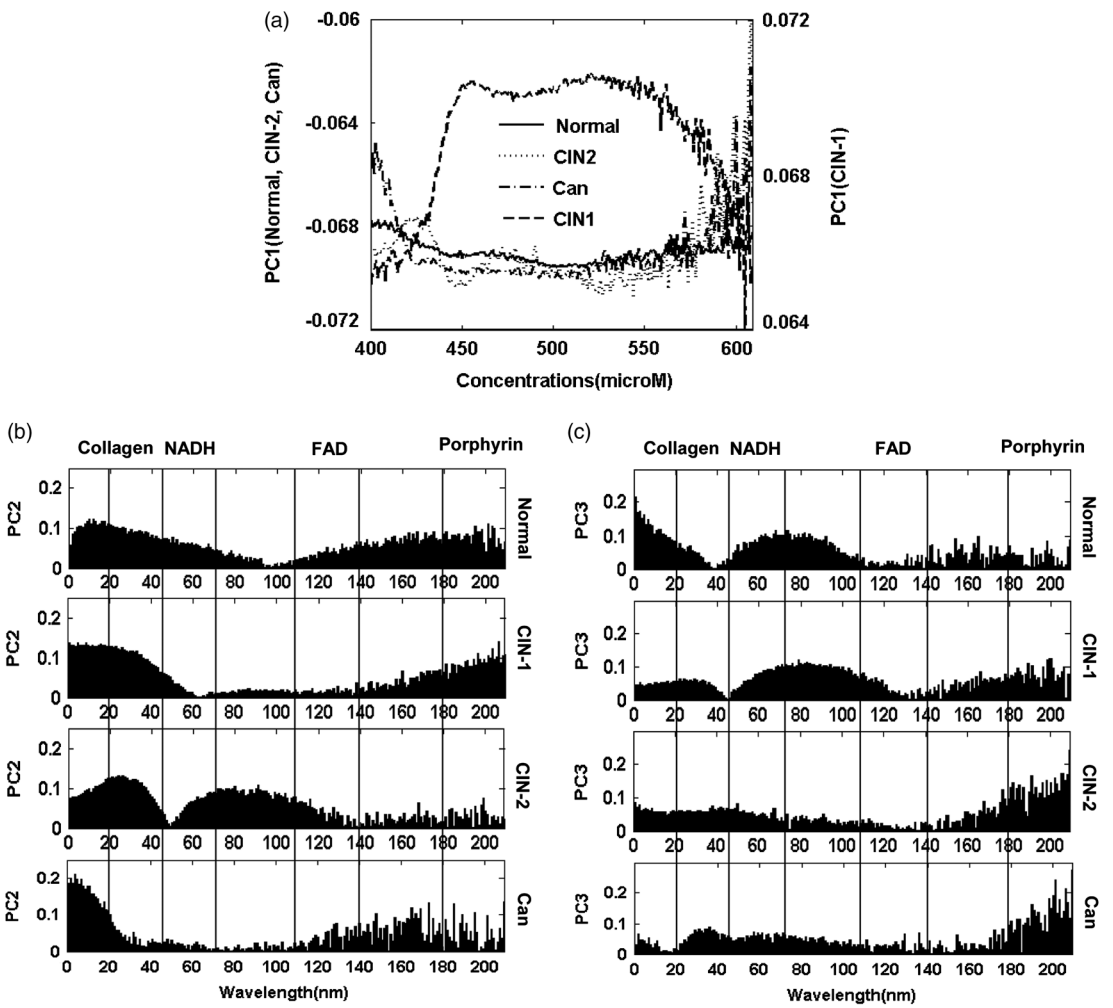

Fig. 7 Plots of the eigenvector entries corresponding to the (a) first, (b) second, and (c) third highest eigenvalues of the intrinsic fluorescence for normal, CIN1, CIN2, and cancer samples.

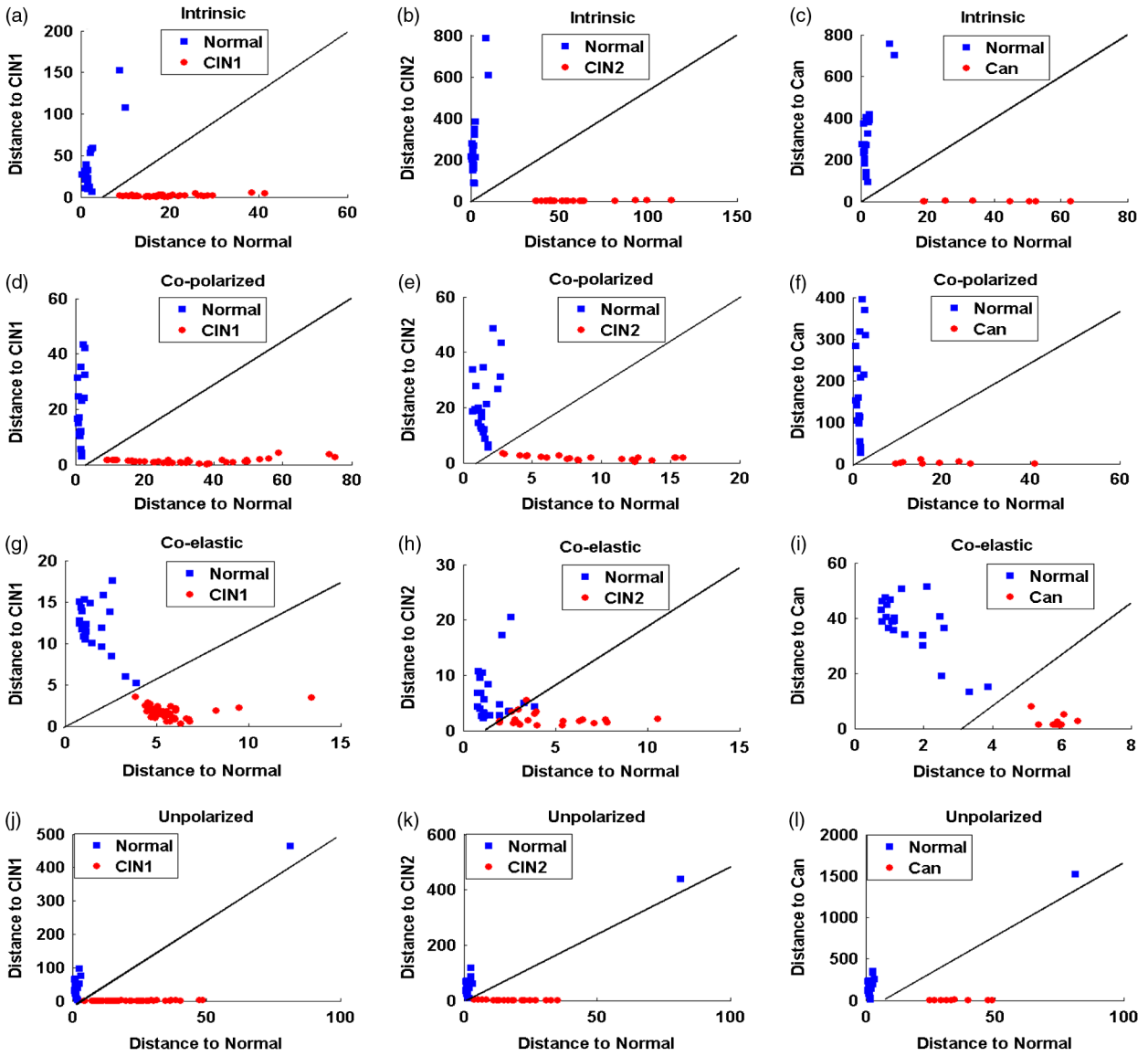

Fig. 8 Scatter plots of the Mahalanobis distances calculated for each sample from the training centroids of intrinsic [(a), (b), and (c)], copolarized [(d), (e), and (f)], copolarized elastic scattering [(g), (h), and (i)], and unpolarized [(j), (k), and (l)] fluorescence for tissue samples of different grades. Here, scores of PC1, PC2, and PC3 have been used for discrimination. Cross-polarized data of fluorescence and elastic scattering also show similar trends as copolarized data (not shown here).

The eigenvector entries (PC3) corresponding to the third largest eigenvalue show that collagen activity in the wavelength range from 400 to 420 nm is significantly reduced from normal to cancer. The NADH activity corresponding to the range from 452 to 470 nm shows a rapid increase in CIN-1 and then a gradual decrease as the disease progresses. The traces of its increased activity can then be tracked in the eigenvector entries of the second largest eigenvalue. In CIN-2, the levels of porphyrin activity corresponding to the range from 580 to 609 nm are increased considerably and remain high in cancer cases. Hence, the biochemical information is faithfully captured by the principal components and clearly demonstrates the changes as normal tissue progressively deteriorates to the higher grades.

Further, for a quantitative estimation of the spectral properties of tissue, the Mahalanobis distance has been utilized for discrimination.[31,45] Mahalanobis distances are calculated for co- and cross-polarized, unpolarized, intrinsic fluorescence, and elastic scattering data sets for comparison, using scores of PC1, PC2, and PC3 and their scatter plots as displayed in Fig. 8. Based on this analysis, the sensitivity and specificity of the intrinsic fluorescence are found to be 100% for classifying the normal from CIN1, CIN2, and cancerous tissue samples. The Mahalanobis distances of the copolarized fluorescence show very high scatter, while unpolarized data points are clumped together around the origin, although with 100% sensitivity and specificity.

The sensitivity and specificity for copolarized elastic scattering are (100%, 100%), (90%, 95%), and (100%, 100%), in differentiating normal from CIN1, CIN2, and cancerous samples [Figs. 8(g), 8(h), and 8(i)], respectively. The cross-polarized components of both fluorescence and elastic scattering follow similar trends as the copolarized data with 100% sensitivity and specificity. Among all the scatter plots displayed, the

intrinsic fluorescence scores display a clear separation with almost no dispersion. The fact that intrinsic fluorescence shows maximum separation indicates its usefulness in reliable discrimination.

4 Conclusion

In conclusion, the polarization-based technique presented here successfully extracts intrinsic fluorescence from the cervical tissue spectra after eliminating the absorption and diffuse scattering effects. In combination with PCA, the experimentally extracted intrinsic fluorescence clearly unmasks the contributions of collagen, NADH, FAD, and porphyrin during the progression of the disease. It is worth emphasizing the effective role of PCA, which separates the collective behavior through its eigenvectors corresponding to first three highest eigenvalues, thereby accurately demonstrating the relative change of the fluorophore activities in the different grades of precancer. The spectral features of these fluorophores manifest in different domains of the correlation matrices. The correlation maps of the intrinsic fluorescence reveal distinct differences among the normal and the different grades as compared to the subtle differences observed in their spectral profiles. Furthermore, the importance of extracting intrinsic fluorescence can be observed from the correlation map of intrinsic fluorescence, which gives three distinct domains that are not clearly visible in co- and cross-polarized components. The eigenvectors corresponding to the first three highest eigenvalues of the correlation matrices containing maximum information of the original data highlight the broad spectral structures and their differences in normal and different grades of different tissues. It can be observed that eigenvectors corresponding to the first three highest eigenvalues capture the relevant spectral features of the correlation matrices and clearly discriminate different tissue types. The sectorial behavior of the eigenvector entries of these dominant eigenvalues unambiguously highlights the differing activities of collagen, NADH, FAD, and porphyrin in different grades of precancer. In particular, the eigenvectors corresponding to the second and third eigenvalues clearly reveal the relative contribution of different flurophores to the correlation matrix. The contribution of the third eigenvalue may be emphasized, as it proved vital in remarkably pointing out the activities of the fluorophores with disease progression. In future, this study may be extended to other forms of cancer and to *in vivo* studies. Further quantification with Mahalanobis distance analysis has shown the reliability of using intrinsic fluorescence as a discriminating tool. Though the sensitivity and specificity of intrinsic, co-, and cross-polarized data turn out to be 100%, the dispersion in data of co- and cross-polarized states is high, while the intrinsic fluorescence displays maximum separation between normal and different grades, respectively. In conclusion, it is pertinent to note that *in vivo* results need not always follow those of excised tissue. The major concerns are the oxidation of major fluorophores and the weaker absorption by blood in *in vitro* conditions. These concerns may be answered only through *in vivo* studies. Nevertheless, validating the proposed technique does provide an insight into the possibility of using it in clinical environment. To take this study further, a polarization-based probe has been fabricated and is in the preliminary stage of testing.

Acknowledgments

The authors acknowledge Dr. Kiran Pandey and Dr. Asha Agarwal of GSVM Medical College, Kanpur, for providing well-characterized tissues for preliminary studies and Dr. Anita H. Gharekhan of C. U. Shah Science College, Ahmedabad, India, for helping in statistical analysis. Authors also acknowledge Mr. Swapnil Barmase, an integrated MS student of the Department of Chemical Sciences, IISER Kolkata, Mohanpur Campus Nadia, India, for his help in the sectorial representation of eigenvector entries.

References

1. R. S. Bradley and M. S. Thorniley, "A review of attenuation correction techniques for tissue fluorescence," *J. R. Soc. Interface* **3**, 1–13 (2006).
2. C. Macaulay et al., "Variation of fluorescence spectroscopy during the menstrual cycle," *Opt. Express* **10**, 493–504 (2002).
3. I. Georgakoudi et al., "NAD(P)H and collagen as in vivo quantitative fluorescent biomarkers of epithelial precancerous changes," *Cancer Res.* **62**(3), 682–687 (2002).
4. U. Utzinger et al., "Understanding the contributions of NADH and collagen to cervical tissue fluorescence spectra: modeling, measurements, and implications," *J. Biomed. Opt.* **6**, 385–396 (2001).
5. J. M. Salmon et al., "Microspectrofluorometric approach to the study of free/bound NAD(P)H ratio as metabolic indicator in various cell types," *Photochem. Photobiol.* **36**(5), 585–593 (1982).
6. T. Galeotti et al., "On the fluorescence of NAD(P)H in whole-cell preparation of tumors and normal tissues," *Eur. J. Biochem.* **17**, 485–496 (1970).
7. A. A. Heikal, "Intracellular coenzymes as natural biomarkers for metabolic activities and mitochondrial anomalies," *Biomark. Med.* **4**(2), 241–263 (2010).
8. S. Curran and G. I. Murray, "Matrix metalloproteinases in tumour invasion and metastasis," *J. Pathol.* **189**, 300–308 (1999).
9. C. Streuli, "Extracellular matrix remodeling and cellular differentiation," *Curr. Opin. Cell Biol.* **11**, 634–640 (1999).
10. A. J. Durkin, S. Jaikumar, and R. R. Kortum, "Optically dilute, absorbing, and turbid phantoms for fluorescence spectroscopy of homogeneous and inhomogeneous samples," *Appl. Spectrosc.* **47**, 2114–2121 (1993).
11. N. Das et al., "Probing multifractality in tissue refractive index: prospects for precancer detection," *Opt. Lett.* **38**, 211–213 (2013).
12. M. Canpolat and J. R. Mourant, "Optical measurement of photosensitizer concentration using a probe with a small source-detector fiber separation," *Proc. SPIE* **3911**, 10–18 (2000).
13. N. Subhash et al., "Oral cancer detection using diffuse reflectance spectral ratio R540/R575 of oxygenated hemoglobin bands," *J. Biomed. Opt.* **11**(1), 014018 (2006).
14. R. Mallia et al., "Oxygenated hemoglobin diffuse reflectance ratio for in vivo detection of oral pre-cancer," *J. Biomed. Opt.* **13**(4), 041306 (2008).
15. R. Mallia et al., "Laser-induced autofluorescence spectral ratio reference standard for early discrimination of oral cancer," *Cancer* **112**(7), 1503–1512 (2008).
16. R. J. Mallia et al., "Clinical grading of oral mucosa by curve-fitting of corrected autofluorescence using diffuse reflectance spectra," *Head Neck* **32**(6), 763–779 (2010).
17. C. M. Gardner, S. L. Jacques, and A. J. Welch, "Fluorescence spectroscopy of tissue: recovery of intrinsic fluorescence from measured fluorescence," *Appl. Opt.* **35**, 1780–1792 (1996).
18. A. J. Durkin et al., "Relation between fluorescence spectra of dilute and turbid samples," *Appl. Opt.* **33**, 414–423 (1994).
19. A. J. Durkin and R. R. Kortum, "Comparison of methods to determine chromophore concentrations from fluorescence spectra of turbid samples," *Lasers Surg. Med.* **19**, 75–89 (1996).
20. N. N. Zhadin and R. R. Alfano, "Correction of the internal absorption effect in fluorescence emission and excitation spectra from absorbing and highly scattering media: theory and experiment," *J. Biomed. Opt.* **3**, 171–186 (1998).
21. M. S. Nair et al., "Propagation of fluorescence in human breast tissues: a diffusion theory model," *Appl. Opt.* **41**, 4024–4035 (2002).
22. J. Wu, M. S. Feld, and R. P. Rava, "Analytical model for extracting intrinsic fluorescence in turbid media," *Appl. Opt.* **32**, 3585–3595 (1993a).

23. Q. Zhang et al., "Turbidity-free fluorescence spectroscopy of biological tissue," *Opt. Lett.* **25**, 1451–1453 (2000).
24. M. G. Müller et al., "Intrinsic fluorescence spectroscopy in turbid media: disentangling effects of scattering and absorption," *Appl. Opt.* **40**(25), 4633–4646 (2001).
25. Z. Volynskaya et al., "Diagnosing breast cancer using diffuse reflectance spectroscopy and intrinsic fluorescence spectroscopy," *J. Biomed. Opt.* **13**(2), 024012 (2008).
26. S. C. Kanick et al., "Extraction of intrinsic fluorescence from single fiber fluorescence measurements on a turbid medium," *Opt. Lett.* **37**(5), 948–950 (2012).
27. Q. Liu et al., "Compact point-detection fluorescence spectroscopy system for quantifying intrinsic fluorescence redox ratio in brain cancer diagnostics," *J. Biomed. Opt.* **16**(3), 037004 (2011).
28. N. Biswal et al., "Recovery of turbidity free fluorescence from measured fluorescence: an experimental approach," *Opt. Express* **11**(24), 3320–3331 (2003).
29. R. Singh et al., "Polarized fluorescence study in cervical tissue: change in auto-fluorescence through different excitation wavelengths," *Proc. SPIE* **7561**, 75610J (2010).
30. C. K. Brookner et al., "Cervical fluorescence of normal women," *Lasers Surg. Med.* **24**, 29–37 (1999).
31. Y. N. Mirabal et al., "Reflectance spectroscopy for in vivo detection of cervical precancer," *J. Biomed. Opt.* **7**(4), 587–594 (2002).
32. D. C. de Veld et al., "Autofluorescence characteristics of healthy oral mucosa at different anatomical sites," *Lasers Surg. Med.* **32**, 367–376 (2003).
33. S. K. Majumder et al., "Nonlinear pattern recognition for laser-induced fluorescence diagnosis of cancer," *Lasers Surg. Med.* **33**, 48–56 (2003).
34. G. S. Nayak et al., "Principal components analysis and artificial neural network analysis of oral tissue fluorescence spectra: classification of normal premalignant and malignant pathological conditions," *Biopolymers* **82**, 152–166 (2006).
35. S. M. Chidananda et al., "Optical diagnosis of cervical cancer by fluorescence spectroscopy technique," *Int. J. Cancer* **119**, 139–145 (2006).
36. M. Ringnér, "What is principal components analysis?," *Nat. Biotechnol.* **26**(3), 303–304 (2008).
37. J. L. Jayanthi et al., "Discriminant analysis of autofluorescence spectra for classification of oral lesions in vivo," *Lasers Surg. Med.* **41**, 345–352 (2009).
38. M. V. Chowdary et al., "Autofluorescence of breast tissues: evaluation of discriminating algorithms for diagnosis of normal, benign, and malignant conditions," *Photomed. Laser Surg.* **27**(2), 241–252 (2009).
39. X. Liu et al., "Principal components analysis of dynamic fluorescence diffuse optical tomography images," *Opt. Express* **18**(6), 6300–6314 (2010).
40. A. H. Gharekhan et al., "Distinguishing autofluorescence of normal, benign, and cancerous breast tissues through wavelet domain correlation studies," *J. Biomed. Opt.* **16**, 087003 (2011).
41. A. H. Gharekhan et al., "PCA based polarized fluorescence study for detecting human cervical dysplasia," *Proc. SPIE* **8580**, 85800N (2013).
42. A. H. Gharekhan et al., "Characterizing breast cancer tissues through the spectral correlation properties of polarized fluorescence," *J. Biomed. Opt.* **13**, 054063 (2008).
43. N. Ramanujam et al., "Development of a multivariate statistical algorithm to analyze human cervical tissue fluorescence spectra acquired in vivo," *Lasers Surg. Med.* **19**, 46–62 (1996).
44. N. Ramanujam et al., "Spectroscopic diagnosis of cervical intraepithelial neoplasia (CIN) in vivo using laser-induced fluorescence spectra at multiple excitation wavelengths," *Lasers Surg. Med.* **19**, 63–74 (1996).
45. R. J. Nordstrom et al., "Identification of cervical intraepithelial neoplasia (CIN) using UV-excited fluorescence and diffuse-reflectance tissue spectroscopy," *Lasers Surg. Med.* **29**, 118–127 (2001).
46. W. K. Huh et al., "Optical detection of high-grade cervical intraepithelial neoplasia in vivo: results of a 604-patient study," *Am. J. Obstet. Gynecol.* **190**, 1249–1257 (2004).
47. S. K. Chang et al., "Combined reflectance and fluorescence spectroscopy for in vivo detection of cervical pre-cancer," *J. Biomed. Opt.* **10**(2), 024031 (2005).
48. A. B. Rodero et al., "Fluorescence spectroscopy for diagnosis differentiation in uteri's cervix biopsies with cervical/vaginal atypical cytology," *J. Fluoresc.* **18**, 979–985 (2008).
49. A. F. Zuluaga et al., "Optical coherence tomography: a pilot study of a new imaging technique for noninvasive examination of cervical tissue," *Am. J. Obstet. Gynecol.* **193**, 83–88 (2005).
50. F. M. Lyng et al., "Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool," *Exp. Mol. Pathol.* **82**, 121–129 (2007).
51. S. Rubina et al., "Raman spectroscopic study on classification of cervical cell specimens," *Vib. Spectrosc.* **68**, 115–121 (2013).
52. N. Ramanujam et al., "Quantitative physiology of the precancerous pervix in vivo through optical spectroscopy," *Neoplasia* **11**(4), 325–332 (2009).
53. V. Masilamani et al., "Fluorescence spectra of blood and urine for cervical cancer detection," *J. Biomed. Opt.* **17**(9), 098001 (2012).
54. R. Rajasekaran et al., "Characterization and diagnosis of cancer by native fluorescence spectroscopy of human urine," *Photochem. Photobiol.* **89**, 483–491 (2013).
55. J. Lackowicz, *Principles of Fluorescence Spectroscopy,* chapter 5, pp. 11–150, Plenum Press, New York (1983).
56. V. Plerou et al., "Universal and nonuniversal properties of cross correlations in financial time series," *Phys. Rev. Lett.* **83**(7), 1471–1474 (1999).
57. A. M. Sengupta and P. P. Mitra, "Distributions of singular values for some random matrices," *Phys. Rev. E* **60**, 3389–3392 (1999).
58. A. H. Gharekhan et al., "Characterizing breast cancer tissues through the spectral correlation properties of polarized fluorescence," *J. Biomed. Opt.* **13**(5), 054063 (2008).
59. N. Agarwal et al., "Wavelet transform of breast tissue fluorescence spectra: a technique for diagnosis of tumors," *IEEE J. Sel. Topics Quantum Electron.* **9**(2), 154–161 (2003).
60. A. H. Gharekhan et al., "Characteristic spectral features of the polarized fluorescence of human breast cancer in the wavelet domain," *Appl. Spectrosc.* **66**(7), 820–827 (2012).
61. R. R. Alfano et al., "Stokes shift emission spectroscopy of human issues and key biomolecules," *IEEE J. Sel. Topics Quantum Electron.* **9**(2), 148–153 (2003).

Seema Devi is pursuing her PhD in the Department of Physics at the Indian Institute of Technology (IIT), Kanpur, India. Her areas of interest include intrinsic fluorescence spectroscopy, synchronous fluorescence spectroscopy, wavelet transform, principal component analysis, confocal imaging, and fluorescence life-time imaging for early detection of cancer.

Prasanta K. Panigrahi is currently a professor at the Indian Institute of Science Education and Research (IISER) Kolkata, India. He received his doctoral degree from the University of Rochester, New York, in 1988. His research interests lie in Bose-Einstein condensates, cold fermions, nonlinear dynamics, quantum computation and quantum information, wavelet transform, and field theory.

Asima Pradhan is a professor in the Department of Physics at the Indian Institute of Technology (IIT), Kanpur, India, and is also associated with the Center for Lasers and Photonics. She received her PhD from City University of New York in 1991. Her areas of interest include laser spectroscopy and biophotonics. Her expertise is in fluorescence spectroscopy, light scattering spectroscopy, time-resolved fluorescence, polarization-based fluorescence, and Mueller matrix imaging of biosamples.