

Comparison of autofluorescence, diffuse reflectance, and Raman spectroscopy for breast tissue discrimination

Shovan K. Majumder

Matthew D. Keller

Vanderbilt University
Department of Biomedical Engineering
Nashville, Tennessee 37235

Fouad I. Boulos

Vanderbilt University
Department of Pathology
Nashville, Tennessee 37235

Mark C. Kelley

Vanderbilt University
Department of Surgical Oncology
Nashville, Tennessee 37235

Anita Mahadevan-Jansen

Vanderbilt University
Department of Biomedical Engineering
Nashville, Tennessee 37235

Abstract. For a given diagnostic problem, important considerations are the relative performances of the various optical biopsy techniques. A comparative evaluation of fluorescence, diffuse reflectance, combined fluorescence and diffuse reflectance, and Raman spectroscopy in discriminating different histopathologic categories of human breast tissues is reported. Optical spectra were acquired *ex vivo* from a total of 74 breast tissue samples belonging to 4 distinct histopathologic categories: invasive ductal carcinoma (IDC), ductal carcinoma *in situ* (DCIS), fibroadenoma (FA), and normal breast tissue. A probability-based multivariate statistical algorithm capable of direct multiclass classification was developed to analyze the diagnostic content of the spectra measured from the same set of breast tissue sites with these different techniques. The algorithm uses the theory of nonlinear maximum representation and discrimination feature for feature extraction, and the theory of sparse multinomial logistic regression for classification. The results reveal that the performance of Raman spectroscopy is superior to that of all others in classifying the breast tissues into respective histopathologic categories. The best classification accuracy was observed to be ~99%, 94%, 98%, and 100% for IDC, DCIS, FA, and normal breast tissues, respectively, on the basis of leave-one-sample-out cross-validation, with an overall accuracy of ~99%.

© 2008 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.2975962]

Keywords: breast tumor; fluorescence; diffuse reflectance; Raman; multi-class diagnostic algorithm; posterior probability.

Paper 07414R received Oct. 3, 2007; revised manuscript received Mar. 17, 2008; accepted for publication Apr. 16, 2008; published online Sep. 22, 2008. This paper is a revision of a paper presented at the SPIE conference on Advanced Biomedical and Clinical Diagnostic Systems V, January 2007, San Jose, California. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 6430.

1 Introduction

Breast cancer is the most common form of malignant tumor found in women in the United States, with an estimated 240 000 new cases diagnosed in 2007.¹ The implementation of regular screening programs and increased use of x-ray mammography for early detection have had a significant impact on breast cancer mortality. Mammograms, however, still suffer from insufficient sensitivity and high rates of “false negatives,” which result in unnecessary follow-ups and biopsies that lead to patient trauma, time delay, and high medical costs.² In fact, 60% to 90% of the suspicious lesions detected by mammography are benign upon biopsy,³ which can range from the fine needle aspiration of single cells to the surgical removal of the entire suspicious mass. A technique that holds considerable promise to overcome these limitations is optical spectroscopy, thanks to its ability to provide biochemical and morphological information about a tissue in a near-real time, minimally or noninvasive manner.^{4–7} Another potential appli-

cation of optical spectroscopy that has currently drawn significant interest is therapeutic guidance,⁸ especially the evaluation of surgical margins in real time to guide tumor resection during breast conservative therapy (partial mastectomy).

Optical spectroscopic techniques that have been investigated for breast cancer detection to date include autofluorescence, diffuse reflectance, and Raman spectroscopies.^{4–9} Alfano and co-workers^{10–12} were the first to apply autofluorescence spectroscopy to the problem of identifying breast malignancy. A series of studies carried out by them^{10–12} showed that significant differences exist in the fluorescence signatures of malignant and nonmalignant human breast tissues. In particular, they showed that using discrimination indices based on ratio of intensities from emission spectra at 300-nm excitation and excitation spectra at 340-nm emission, they could discriminate malignant from fibrous breast tissues with 93% sensitivity and 95% specificity, but results were worse for discriminating normal fatty and malignant tissues. Gupta et al.¹³ and Majumder et al.¹⁴ measured emission spectra at 337- and 488-nm excitation wavelengths from normal,

Address all correspondence to Shovan Kumar Majumder, Department of Biomedical Engineering, Vanderbilt University, VU Station B, Box 351631, Nashville, TN 37235; E-mail: shovan.k.majumder@vanderbilt.edu

benign, and malignant breast tissue samples, and using the integrated emission intensity from the 337-nm excitation, they separated malignant tissues in a binary fashion from both benign and normal with sensitivity and specificity of up to 99.6%.^{13,14} More recently, Palmer et al.¹⁵ used the fluorescence emission spectra from multiple excitation wavelengths ranging from 300 to 460 nm to separate malignant from non-malignant samples. Using principal component analysis (PCA), followed by a Wilcoxon rank-sum test to identify significant components, and then entering those into a support vector machine (SVM) resulted in 70% sensitivity and 92% specificity for discriminating malignant from normal or benign tissues.¹⁵ Using a probe with three different delivery-to-collection fiber distances and similar data analysis procedures, the same group found that analyzing integrated fluorescence emission intensities from a single excitation wavelength at all three separations could provide results comparable to those of the previous study, but with a simpler experimental setup.¹⁶

A few groups have investigated the utility of UV-visible diffuse reflectance (or elastic scattering) spectroscopy, either alone or in conjunction with fluorescence, for breast cancer diagnosis. Of these, Bigio et al.⁸ used *in vivo* measurements in an attempt to both make a diagnosis and help guide resection, and using diagnostic algorithms based on artificial neural networks and hierarchical cluster analysis, they were able to distinguish malignant from normal breast tissue with sensitivities up to 69% and specificities up to 93%. Palmer et al.¹⁵ investigated the utility of combining diffuse reflectance measurements with fluorescence measurements, but they found that diagnostic performance for breast cancer was not significantly improved by doing so.

Perhaps the most widespread application of Raman spectroscopy in cancer research has been for breast cancer detection.^{17–24} Alfano et al.¹⁷ were the first to investigate the ability of Raman spectroscopy to distinguish normal from malignant breast tissue, using a Fourier transform Raman spectrometer at 1064-nm excitation. Later, Frank et al.^{18,19} and Redd et al.²⁰ employed Raman spectroscopy using visible and near IR excitation to study excised human breast tissues. More recently, Feld and colleagues^{21–24} have done extensive work on using Raman spectroscopy for breast cancer diagnosis. An early study²¹ similar to Redd et al.²⁰ showed comparable spectra, but multivariate statistical algorithms improved diagnosis. Over the past several years, this group has developed a system that classifies breast Raman spectra according to the modeled contributions of individual component spectra from materials such as fat, collagen, and DNA.²² They have used this system on tissue samples to discriminate invasive carcinoma from normal fatty, fibroadenoma, and fibrocystic change tissues with an overall accuracy of 86%,²³ and in a small pilot *in vivo* trial for guiding resection, normal, fibrocystic change, and malignant tissues were classified with an overall accuracy rate of 93%.²⁴

Although the studies described use autofluorescence, diffuse reflectance, combined autofluorescence and diffuse reflectance, or Raman spectroscopy for breast tissue discrimination, there is no published report of a direct, side-by-side comparison of the efficacy of these modalities for specific types of discrimination. The goal of this paper, then, is to report a comparative evaluation of the relative capabilities of fluorescence, diffuse reflectance, combined fluorescence and

diffuse reflectance, and Raman spectroscopy for discriminating the different histopathologic categories of human breast tissues. Fluorescence, diffuse reflectance, and Raman spectra were acquired *ex vivo* from human breast tissue samples belonging to four histopathologic categories: invasive ductal carcinoma (IDC), ductal carcinoma *in situ* (DCIS), fibroadenoma (FA), and normal. A probability-based multivariate statistical algorithm capable of direct multiclass classification²⁵ was developed to analyze the diagnostic content of these different sets of optical spectra measured sequentially from the same set of breast tissue sites. The results showed that although Raman spectroscopy allows for the most accurate tissue classification as may be useful for clinical diagnosis of the pathological state of a tumor, combined fluorescence and diffuse reflectance holds promise for use in clinical applications such as evaluating margin status during breast surgery in which imaging an entire tissue surface will be required for delineating normal from nonnormal breast tissue.

2 Materials and Methods

2.1 Breast Tissue Samples

The human breast tissue samples were obtained under a protocol approved by the Vanderbilt University Institutional Review Board. The freshly frozen samples were obtained either from the tissue bank at Vanderbilt Clinic or from the National Cancer Institute's (NCI) Cooperative Human Tissue Network. A total of 74 tissue samples were obtained for this study. Normal tissue samples were obtained from either reduction mammoplasty or uninvolved areas from radical mastectomy procedures, and tumor samples were partial sections of surgically removed breast lesions. All samples were stored at -80°C until spectroscopic study, at which point they were thawed to room temperature in buffered saline.

2.2 Instrumentation

In vitro fluorescence and diffuse reflectance spectra of breast tissue samples were measured using a portable spectroscopic system as illustrated in Fig. 1. A high-pressure nitrogen laser (Spectra Physics, Mountain View, CA) is used as the excitation source for fluorescence measurements, and a 150-W tungsten-halogen lamp (Ocean Optics, Dunedin, Florida) emitting broadband white light from 400 to 850 nm is used for diffuse reflectance measurements. Light delivery to and collection from the sample is achieved with a fiber-optic probe consisting of seven 300- μm core diameter fibers arranged in a six-around-one configuration (Romack, Williamsburg, Virginia). Two of the surrounding fibers deliver laser and white light consecutively to the tissue sample while the remaining fibers collect fluorescence and diffuse reflectance from the tissue sample. Diffuse reflectance and fluorescence emissions collected by the fiber-optic probe are serially dispersed and detected with a chip-based spectrometer (model number S-2000, Ocean Optics). For this study, the output power of the white light was ~ 0.6 mW at the tissue surface, and the nitrogen laser was operated at a 20-Hz repetition rate, 5-ns pulse width, and average pulse energy of 45 ± 5 μJ at the tissue surface. An integration time of 100 ms was used for each spectral measurement. The overall spectral resolution of the system was ~ 2 nm.

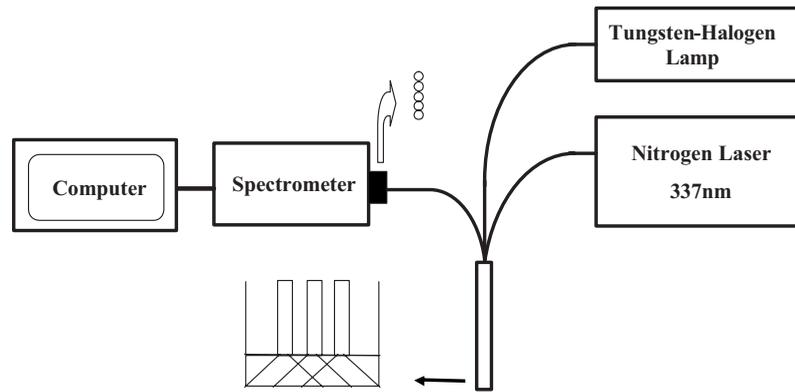


Fig. 1 Schematic of the spectroscopic system for fluorescence and diffuse reflectance measurements.

Raman spectra of the breast tissue samples were measured with a portable Raman spectroscopy system shown in Fig. 2. The system consists of a 785-nm diode laser (Process Instruments, Inc., Salt Lake City, Utah), a seven-around-one beam-steered fiber optic probe (Visionex Inc., Atlanta, Georgia), an imaging spectrograph (Kaiser Optical Systems, Inc., Ann Arbor, Michigan), and a back-illuminated, deep-depletion, thermoelectrically cooled charge-coupled device (CCD) (Princeton Instruments, Princeton, New Jersey), all controlled with a laptop computer. The beam-steered fiber optic probe delivers the 785-nm light, which is band pass filtered at the probe tip, onto the tissue and collects the Raman scattered light, which is then filtered with an inline notch filter within the probe itself. The light is then fed into the spectrograph where it is filtered again with a 785-nm holographic notch filter and dispersed onto the CCD, where the computer records the signal. For this study, the fiber optic probe delivered 80 mW onto the tissue and collected the scattered light for 1 to 3 s.

2.3 Experimental Methods

A standard protocol was followed for the spectral measurements and maintained for all the samples in this study. Prior to spectral acquisition, each sample was thawed to room tem-

perature in phosphate-buffered saline. For recording spectra, the tissue sample was kept on a sheet of aluminum foil, and the tips of the fiber-optic probes were placed normally in gentle contact with the target tissue. From each tissue sample, spectra were recorded from 2 to 6 sites depending on the size of the sample. From each individual site, autofluorescence, diffuse reflectance, and Raman spectra were measured sequentially. In all cases, the overhead fluorescent lights were turned off during the measurements. Following spectral acquisition from each sample, the investigated sites were inked to record their locations, and then fixed in formalin for standard hematoxylin and eosin staining and examination by an experienced pathologist (F.B.) blinded to the results of the optical spectra. The histopathology report of each site was considered to be the gold standard. All spectra were categorized according to their histological identities and grouped into IDC, DCIS, FA, or normal breast tissue.

2.4 Data Processing

After acquisition of autofluorescence and diffuse reflectance spectra, a set of reference spectra from a fluorescence and a reflectance standard were recorded to correct for intersample variability due to variations in laser-pulse energy and white light power. The fluorescence standard is a low-concentration Rhodamine 6G solution (2 mg/L) contained in a quartz cuvette, and the reflectance standard is a 20% reflectance plate (Labsphere, North Sutton, New Hampshire) placed in a black box. All *in vitro* raw fluorescence and diffuse reflectance spectra were processed to remove instrumentation-induced variations and to yield calibrated spectra, the details of which are described elsewhere.²⁶ The resultant fluorescence and diffuse reflectance spectra were further corrected for the nonuniform spectral response of the detection system. Fluorescence spectra were recorded from 370 to 650 nm, and diffuse reflectance spectra were recorded from 400 to 800 nm.

Prior to Raman spectral measurements, the wave number axis was calibrated with a neon-argon lamp, acetaminophen, and naphthalene standards. For each Raman spectrum measured, the signal from the CCD was binned along the vertical axis to create a single spectrum per measurement site. Prior to any signal processing, the spectrum was truncated to only include the region from about 990 to 1800 cm^{-1} to eliminate the large Raman peaks due to the silica present in the fiber-optic probe that obscure any tissue Raman peaks, as well as

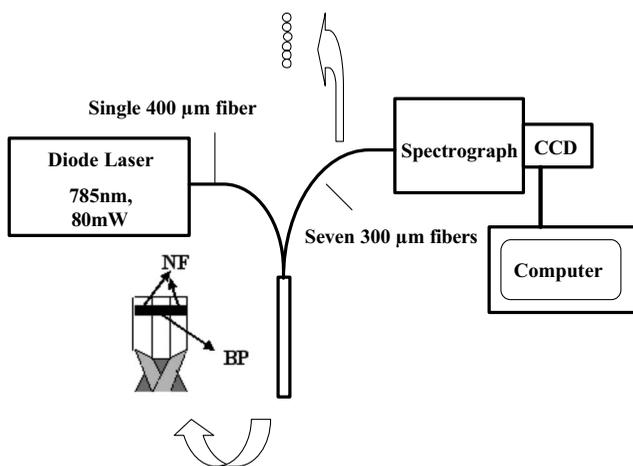


Fig. 2 Schematic of the experimental setup for Raman spectroscopy measurements (NF: notch filter at 785 nm; BP: bandpass filter at 785 nm).

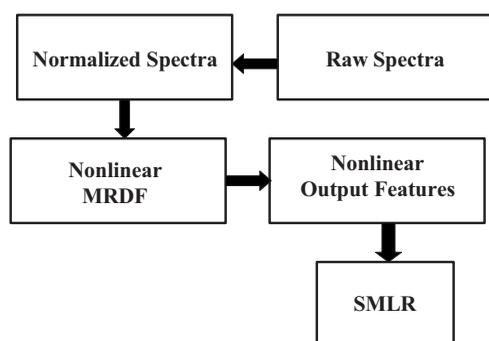


Fig. 3 Schematic overview of the diagnostic algorithm developed to analyze autofluorescence, diffuse reflectance, and Raman spectra of breast tissues.

the noise present at the very end of the spectral region. The spectrum was then binned along the wave number axis in 3.5-cm^{-1} intervals and filtered with a second-order Savitzky-Golay filter (window= 17.5 cm^{-1}) for noise smoothing. Fluorescence subtraction was accomplished using an automated, modified polynomial fitting method in which a fifth-order polynomial is fit to the fluorescence baseline.²⁷

Following data processing, a method of normalization was adopted to remove the absolute intensity information from the spectra that might be affected by many unavoidable experimental factors. In the case of fluorescence and diffuse reflectance, the spectrum from each site of a sample was normalized with respect to the integrated intensity (ΣI) from that site. In the case of Raman, each spectrum was normalized to its mean spectral intensity across all Raman bands.

The set of spectral data normalized in the manner just discussed were used for subsequent data analysis. For combined fluorescence and diffuse reflectance, the respective area-normalized spectra from each tissue site were concatenated end-to-end to form a single column vector. These column vectors were then further concatenated in rows to form the input data matrix (for all the tissue samples investigated).

2.5 Data Analysis

Figure 3 shows a flow chart of the diagnostic algorithm that was developed to analyze the breast tissue fluorescence, diffuse reflectance, and Raman spectra. The algorithm development was described previously²⁵ and consisted of two steps: (i) extraction of diagnostic features from the spectra using nonlinear maximum representation and discrimination feature (MRDF)²⁸ and (ii) development of a probabilistic scheme of classification based on sparse multinomial logistic regression (SMLR)²⁹ for classifying the nonlinear features into corresponding tissue categories. Each step is described in detail in the following.

Given a set of input data comprising samples from different classes with a given dimensionality, nonlinear MRDF²⁸ aims to find a set of nonlinear transformations of the input data that optimally discriminate between the different classes in a reduced dimensionality space. It invokes nonlinear transforms, in this case restricted order polynomial mappings of the input data,²⁸ in two successive stages. In the case of present spectral data, the aim of nonlinear MRDF is to compute $K < M$ ($< N$) nonlinear transformation vectors, Φ_K , from

N -dimensional (where N is the number of wavelengths over which spectra were recorded) spectra of breast tissue sites, such that the projections of the input data on Φ_K from the different tissue categories are statistically well separated from each other. In the first stage, the input spectral data \mathbf{x} (normalized intensities corresponding to wavelengths of the spectra) from each tissue type are raised to the power p' to produce the associated nonlinear input vectors $\mathbf{x}_{p'}$, which are then subject to a transform Φ'_M such that $\mathbf{y}'_M = \Phi'^T_M \mathbf{x}_{p'}$, are the first stage output features in the nonlinear feature space of reduced dimension $M \ll N$. In the second stage, the reduced M -dimensional output features \mathbf{y}'_M for each tissue type are further transformed nonlinearly to the power p to produce higher order features \mathbf{y}'_{Mp} , and a second transform Φ_K is computed so as to yield the final output features $\mathbf{y}_K = \Phi^T_K \mathbf{y}'_{Mp}$ in the nonlinear feature space of dimension K ($K \leq M$). Because the nonlinearities introduced in the two stages are different (p' in the first stage and p in the second stage), this is expected to produce more general nonlinear transforms on the input spectral data, leading to improved separation of the final nonlinear features for the tissue categories in the new feature space. Thus MRDF automatically finds a closed form solution for the best set of nonlinear transforms.

Classification with SMLR²⁹ is a probabilistic multiclass model based on the sparse Bayesian machine-learning framework of statistical pattern recognition. The central idea of SMLR is to separate a set of labeled input data into its constituent classes by predicting the posterior probabilities of their class membership. It computes the posterior probabilities using a multinomial logistic regression model and constructs a decision boundary that separates the data into its constituent classes based on the computed posterior probabilities following Bayes' rule. Classification of a given set of input data \mathbf{x} is based on the vector of posterior probability estimates yielded by the SMLR algorithm and a class is assigned to a data for which its posterior probability is the highest

$$\hat{y}(\mathbf{x}) = \arg \max P(y_i = 1 | \mathbf{x}).$$

An important task following development of the diagnostic algorithm was to evaluate its classification ability in an unbiased way through cross-validation. Because we had a limited number of spectra in each diagnostic category from a limited number of samples, the cross-validation of the algorithm was performed using leave-one-sample-out cross-validation. In this method, the training of the algorithm was performed using $N-1$ samples (where $N=74$ samples), and the test was carried out using the excluded sample.³⁰ This was repeated N times, each time excluding a different sample. Thus, training was achieved using, in a sense, all samples, and at the same time independence between the training and test sets was maintained.

2.5.1 Multiclass receiver operating characteristic analysis

To quantitatively compare the relative performance of the diagnostic algorithms developed for fluorescence, diffuse reflectance, combined fluorescence and diffuse reflectance, and Raman spectral data sets, a multiclass receiver-operating characteristic (ROC) analysis was carried out on the classifi-

Table 1 Histological distribution of the tissues.

Category	No. of Spectra	No. of Tissue Samples
Invasive ductal carcinoma (IDC)	86	25
Ductal carcinoma in-situ (DCIS)	18	6
Fibroadenoma (FA)	55	11
Normal (adipose, glandular)	134	32
Total	293	74

cation results yielded by the corresponding algorithms. The formulation developed by Hand and Till³¹ was followed for this purpose. The formulation extends the two-class ROC analysis in a straightforward way for multiclass case and computes a generalized metric indicative of overall performance measure of a given multiclass diagnostic algorithm. Given n number of classes, the Hand and Till measure (HTM) is the average of the pairwise area under the ROC curves between $n(n-1)/2$ pairs of classes:

$$HTM = \frac{2}{|n|(|n|-1)} \sum_{\{c_i, c_j\} \in C} AUC(c_i, c_j),$$

where AUC is the area under the two-class ROC curve involving classes c_i and c_j . The summation is calculated over all pairs of distinct classes, irrespective of order. Similar to the two-class case, the closer the HTM equals to 1, the more accurate the corresponding diagnostic algorithm is.

3 Results

In vitro fluorescence, diffuse reflectance, and Raman spectroscopic measurements were carried out on a total of 74 different tissue samples from 74 different patients. Optical spectra were acquired from 293 unique tissue sites on these samples. The details of the histopathological distribution of the tissue sites are summarized in Table 1.

Figures 4–6 show the average normalized fluorescence, diffuse reflectance, and Raman spectra for IDC (86), DCIS (18), FA (55), and normal breast tissues (134), with the error bars representing the spectral standard deviations. From the figures, it is evident that the variation in the measured spectral intensity is comparable for all the tissue types in all three sets of optical spectra. The percentage variation (σ/\bar{x}) in the spectral intensities from the different measurement sites was observed to lie in the range of $\sim 15\%$ to 35% over the respective number of tissue sites included in the four histopathological categories for all the three sets of spectra. Here, \bar{x} is the mean intensity value from different measurement sites of one category and σ is one standard deviation.

For comparison of spectral differences among the different tissue types, the average fluorescence, diffuse reflectance, and Raman spectra are plotted without error bars in Figs. 7–9. One can see that although fluorescence from FA tissue is visibly different than that from the rest of the tissue types throughout the spectral region, the differences among IDC,

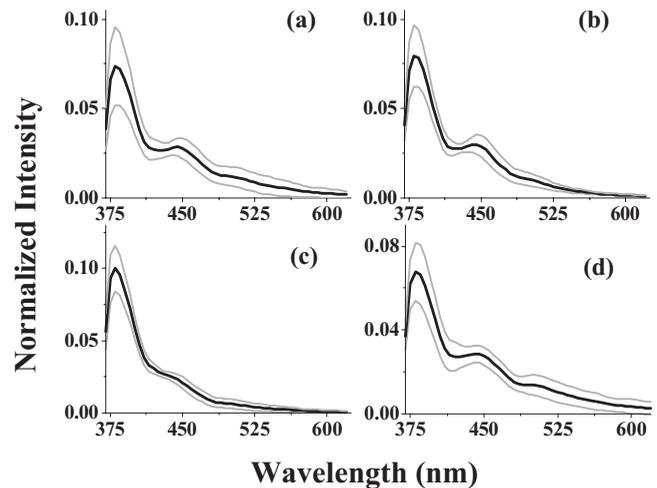


Fig. 4 Mean fluorescence spectra (thick solid lines) of the (a) invasive ductal carcinoma ($n=86$), (b) ductal carcinoma *in situ* ($n=18$), (c) fibroadenoma ($n=55$), and (d) normal ($n=134$) breast tissue samples. The gray lines above and below the thick black solid lines represent ± 1 standard deviation.

DCIS, and normal breast tissues are fairly subtle except in the 390-nm band region where the band intensities show prominent differences. Similarly, although the average reflectance spectra show visible differences among all the tissue types in general, almost no difference is seen between IDC and normal breast tissue in the 590 to 800-nm wavelength region. In contrast, the differences in the average Raman spectra appear to be somewhat more pronounced among all the tissue types across all the major Raman bands present over the entire wave number region.

Tables 2–5 show the diagnostic results in the form of confusion matrices displaying comparisons of the pathological diagnosis with that of the MRDF–SMLR–based spectroscopic diagnostic algorithms. In all instances, the classification re-

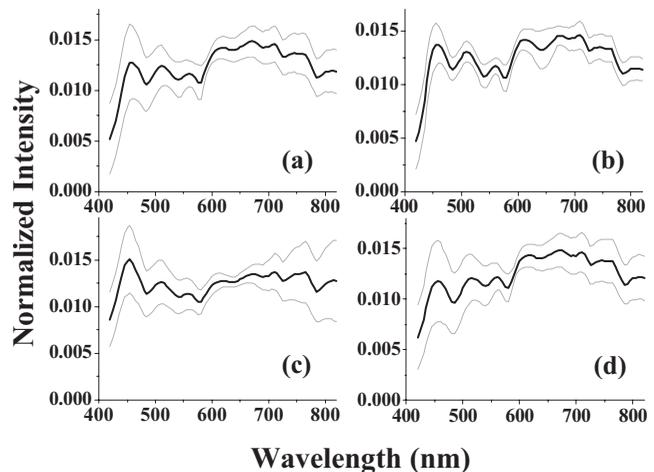


Fig. 5 Mean diffuse reflectance spectra (thick solid lines) of the (a) invasive ductal carcinoma ($n=86$), (b) ductal carcinoma *in situ* ($n=18$), (c) fibroadenoma ($n=55$), and (d) normal ($n=134$) breast tissue samples. The gray lines above and below the thick black solid lines represent ± 1 standard deviation.

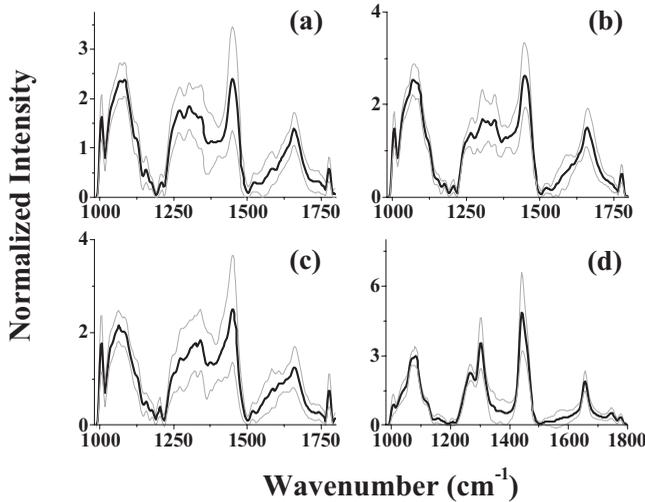


Fig. 6 Mean Raman spectra (thick solid lines) of the (a) invasive ductal carcinoma ($n=86$), (b) ductal carcinoma *in situ* ($n=18$), (c) fibroadenoma ($n=55$), and (d) normal ($n=134$) breast tissue samples. The gray lines above and below the thick black solid lines represent ± 1 standard deviation.

results were obtained based on leave-one-sample-out cross-validation of the entire data set. One can see that diffuse reflectance data alone achieved an overall classification accuracy of 72% (211 out of 293). Its best performance was in classifying FA tissue (85% accuracy), though it fared worse in classifying other tissue types, and errors were spread among the various classes. Fluorescence data alone provided the worst overall accuracy, correctly classifying only 209 measurement sites (71%). It proved most adept at classifying normal breast tissues, though accuracy in that diagnosis was still only 85%. When fluorescence and diffuse reflectance spectra were combined, the diagnostic accuracy improved quite a bit, classifying 247 out of 293 (84%) sites correctly. Normal tissues were correctly classified in 86% of cases, but DCIS and FA were classified correctly in 89% and 98% of the sites. The algorithm still struggled somewhat to classify IDC, but, in general, misclassifications appeared less random than with either modality alone. With Raman spectra as the input data, the algorithm discriminated all four classes very well, correctly

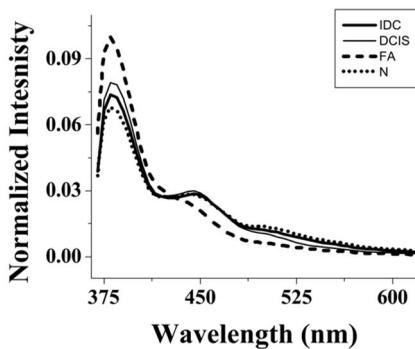


Fig. 7 Mean fluorescence spectra of the invasive ductal carcinoma, ductal carcinoma *in situ*, fibroadenoma, and normal breast tissue samples without error bars.

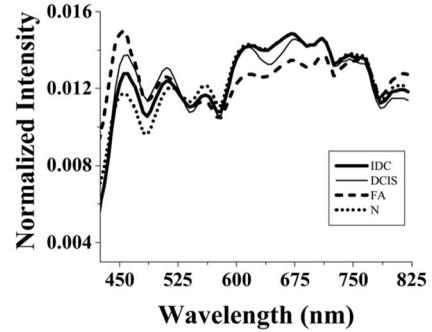


Fig. 8 Mean diffuse reflectance spectra of the invasive ductal carcinoma, ductal carcinoma *in situ*, fibroadenoma, and normal breast tissue samples without error bars.

classifying 290 out of 293 (99%) sites. Normal tissues, which include both fatty and glandular, were classified correctly for every site, and only one each IDC, DCIS, and FA sites were misclassified.

In addition to assigning class labels, all the diagnostic algorithms also yielded posterior probabilities of the measured tissue sites belonging to each breast tissue type. Figures 10(a) to 10(d) illustrate these posterior probabilities computed by the four different algorithms. The posterior probabilities are indicative of the certainty of classification, and they are plotted for all the different tissue sites included in each type of tissue. It is apparent from the figures that although more than 95% of the correctly classified tissue sites in each type have a posterior probability >0.80 with the algorithm based on Raman spectra of breast tissues, a significant fraction (up to $\sim 50\%$) of these sites classified correctly using the algorithm based on combined fluorescence and diffuse reflectance spectra are seen to have a posterior probability <0.80 . Similarly, although the accuracy obtained in correctly classifying FA and normal tissue sites is $>80\%$ with the algorithm based on either fluorescence or diffuse reflectance spectra alone, only $\sim 50\%$ to 60% of these tissue sites are seen to have a posterior probability >0.80 .

The multiclass ROC analyses of the classification results provided a quantitative evaluation of the overall performance of the diagnostic algorithms. Table 6 lists the *HTM* values

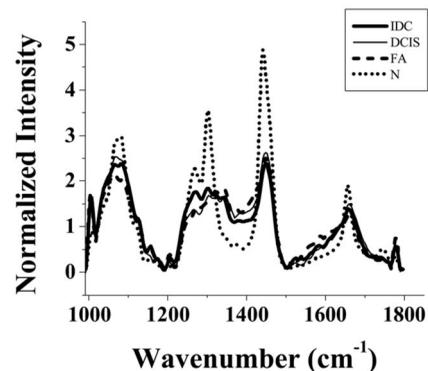


Fig. 9 Mean Raman spectra of the invasive ductal carcinoma, ductal carcinoma *in situ*, fibroadenoma, and normal breast tissue samples without error bars.

Table 2 Confusion matrix displaying classification of breast tissues using MRDF-SMLR-based algorithm with fluorescence spectra.

Pathology Diagnosis (no. of sites)	Fluorescence Diagnosis			
	IDC	DCIS	FA	Normal
IDC (86)	50	2	10	24
DCIS (18)	8	1	0	9
FA (55)	5	1	44	5
Normal (134)	11	3	6	114

The classification results represent leave-one-sample-out cross-validation. The bold type values appearing in the diagonal of each confusion matrix denote number of tissue sites correctly classified for each diagnostic category.

obtained for the algorithms based on fluorescence, diffuse reflectance, combined fluorescence and diffuse reflectance, and Raman spectra. The estimated *HTM* value of the algorithm based on fluorescence spectra alone is 0.83, and those for diffuse reflectance, combined fluorescence and diffuse reflectance, and Raman spectra based algorithms are 0.88, 0.95, and 0.99, respectively. It is important to mention here that the *HTM* value is a quantitative measure of the gross performance of an algorithm and the *HTM* for an ideal diagnostic algorithm will have a value of 1.

4 Discussion

Most of the studies reported to date on the application of optical spectroscopy for breast cancer detection⁸⁻²⁴ have used one of three spectroscopic techniques: fluorescence, diffuse reflectance, or Raman scattering. However, a comprehensive, side-by-side evaluation of the relative efficacies of these different methods has not been addressed in the literature. The goal of the present study is to evaluate and compare the relative capabilities of fluorescence, diffuse reflectance, combined fluorescence and diffuse reflectance, and Raman spectroscopy for simultaneously discriminating the different histopathologic categories of human breast tissues. Such an evaluation is important because it may help choose the optimal modality for a given diagnostic problem. In the present study, fluorescence, diffuse reflectance, and Raman spectra were acquired

Table 3 Confusion matrix displaying classification of breast tissues using MRDF-SMLR-based algorithm with diffuse reflectance spectra.

Pathology Diagnosis (no. of sites)	Diffuse Reflectance Diagnosis			
	IDC	DCIS	FA	Normal
IDC (86)	44	2	1	39
DCIS (18)	9	5	0	4
FA (55)	3	0	47	5
Normal (134)	13	2	4	115

The classification results represent leave-one-sample-out cross-validation.

Table 4 Confusion matrix displaying classification of breast tissues using MRDF-SMLR-based algorithm with combined fluorescence and diffuse reflectance spectra.

Pathology Diagnosis (no. of sites)	Combined Fluorescence and Diffuse Reflectance Diagnosis			
	IDC	DCIS	FA	Normal
IDC (86)	62	2	0	22
DCIS (18)	2	16	0	0
FA (55)	0	0	54	1
Normal (134)	17	2	0	115

The classification results represent leave-one-sample-out cross-validation.

ex vivo from human breast tissue samples belonging to four distinct histopathologic categories: IDC, DCIS, FA, and normal. A probability-based multivariate statistical algorithm capable of direct multiclass classification was developed to analyze the diagnostic content of these different sets of optical spectra measured sequentially from the same set of breast tissue sites.

The primary basis for optical detection using spectroscopic techniques is an array of biochemical changes that take place as tissue undergoes neoplastic transformations. For example, IDC, FA, and normal breast tissues are known to show variable amounts of collagen, and elastin, which is reflected in the fluorescence intensity of the 390-nm band characteristic of these connective tissue proteins.⁴⁻⁶ Similarly, the differences in the concentrations and oxidation states of coenzymes such as nicotinamide adenine dinucleotide (NADH) and flavin adenine dinucleotide (FADH₂) due to differences in metabolic activities in normal and neoplastic breast tissues contribute to the changes in the fluorescence intensity of the broad 460-nm band believed primarily to be due to these fluorophores.⁴⁻⁶ Some of the changes found in the fluorescence spectra of normal and abnormal breast tissues are also the result of changes in the wavelength-dependent absorption and scattering properties of tissues.⁴⁻⁶ However, these changes can be seen to be more prominent in the diffuse reflectance spectra of

Table 5 Confusion matrix displaying classification of breast tissues using MRDF-SMLR-based algorithm Raman spectra.

Pathology Diagnosis (no. of sites)	Raman Diagnosis			
	IDC	DCIS	FA	Normal
IDC (86)	85	0	0	1
DCIS (18)	0	17	0	1
FA (55)	0	1	54	0
Normal (134)	0	0	0	134

The classification results represent leave-one-sample-out cross-validation.

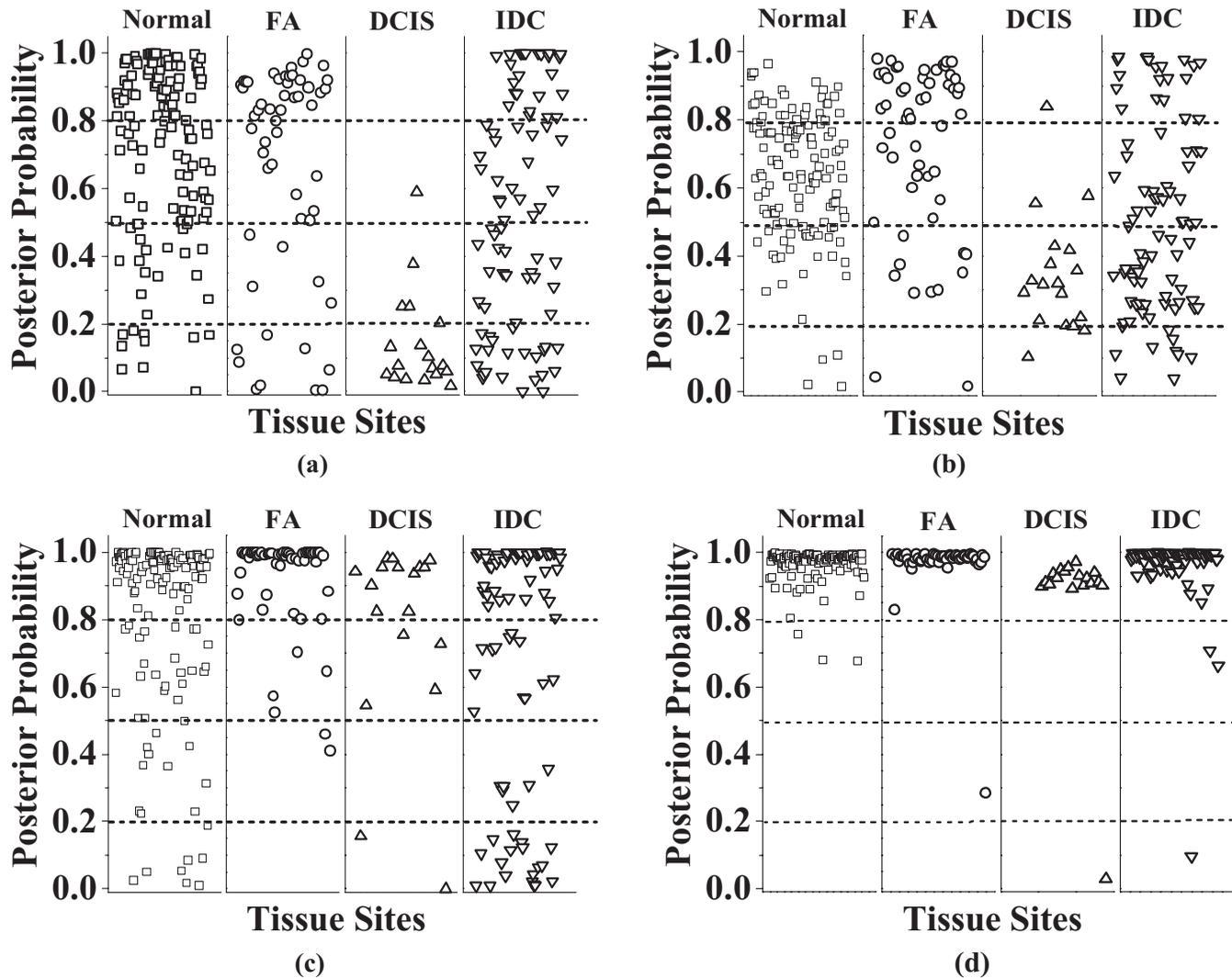


Fig. 10 Posterior probabilities of being classified as invasive ductal carcinoma (IDC), ductal carcinoma *in situ* (DCIS), fibroadenoma (FA), and normal for the (a) fluorescence, (b) diffuse-reflectance, (c) combined autofluorescence and diffuse reflectance, and (d) Raman spectra of the breast tissue sites.

the corresponding tissue types, as diffuse reflectance provides a direct measurement of tissue absorption as well as scattering.^{8,15} For example, the several dips found in the diffuse reflectance spectra of breast tissues represent the signatures of absorption by oxygenated and deoxygenated hemoglobin,^{8,15} which are major absorbers present in blood and have structured absorption bands spanning nearly the entire visible and near-infrared region.^{4-6,8} The differences in absorption properties between normal and abnormal breast tissues, known to be caused primarily by hemoglobin, are clearly seen in the measured diffuse reflectance spectra that show significant variation in spectral line shapes for the corresponding tissue types. On the other hand, Raman spectroscopy probes the vibrational energy levels of molecules, and specific peaks in the Raman spectrum correspond to particular chemical bonds or bond groups.⁷ Because of Raman's chemical specificity, it has the ability to discern the slight biochemical changes associated with neoplastic transformation.⁷ For example, the spectral variations between the different breast tissue pathologies observed at 1000 to 1150, 1170, 1200 to

1345, 1440, and 1650 cm^{-1} correspond to biochemical differences inherent in the different breast tissues, notably connective tissue proteins, and fatty acids.^{7,17} The normal breast tissue spectra are noticeably different as compared with all other categories and are dominated by Raman bands characteristic of fatty acids (1650, 1440, and 1300 cm^{-1}), whereas the intensities of these lipid-specific bands are much reduced in the Raman spectra of IDC, DCIS, and FA, implying a relative increase in protein content in these tissue types. In contrast, differences among IDC, DCIS, and FA are found to be subtler: the ratio of the bands between 1200 to 1300 cm^{-1} (tryptophan, phenylalanine, amide III: proteins) and 1440 cm^{-1} (CH_2 bending: proteins, lipids) varies with tissue types, and this variation is different between cancerous and noncancerous breast tissues. Similarly, the ratio of peaks at 1650 cm^{-1} (amide I: proteins, lipids) and the small peak near 1550 cm^{-1} (tryptophan) is different in the IDC and FA than in DCIS. Other variations include changes in the band patterns and intensities between 1000 and 1100 cm^{-1} (tyrosine, proline, phe-

Table 6 *HTM* values corresponding to the diagnostic algorithms based on autofluorescence, diffuse reflectance, combined autofluorescence and diffuse reflectance, and Raman spectra of breast tissues.

	Fluorescence	Diffuse Reflectance	Combined Autofluorescence and Diffuse Reflectance	Raman
<i>HTM</i>	0.83	0.88	0.95	0.99

nylalanine, proteins) evident as a function of pathology.

It is relevant to note here that although all the spectroscopic techniques are seen to lead to many observable spectral differences between different breast pathologies, it is more important to see the significance of these variations toward pathological classification. A gross comparison of the classification results (see Tables 2–5) yielded by the four diagnostic algorithms clearly indicates that even though fluorescence alone is the least capable in accurately discriminating among the four histopathologic categories of breast tissues based on their measured spectra at 337-nm excitation, the performance of diffuse reflectance alone appears to be a bit better. Although the results are improved to a large extent when fluorescence and diffuse reflectance spectra are combined, the performance of Raman spectroscopy alone is seen to be superior to that of all the other techniques employed. The large improvement in classification performance of the Raman-based algorithm likely originates from the greater diagnostic content of the tissue Raman spectra in comparison with that from either fluorescence or diffuse reflectance spectra. Although diffuse reflectance primarily probes the absorption and scattering properties of tissue,^{8,15,16} fluorescence, in principle, has the additional advantage of biochemical specificity that arises from the fact that tissue contains several intrinsic fluorophores that have their characteristic fluorescence emission.^{4–6} However, the significantly broad and overlapping emission profiles of these fluorophores make the appearance of the resulting tissue fluorescence (which is a superposition of the spectra of its constituent fluorophores modulated by tissue optical properties) spectra mostly flat and featureless,^{4–6} thus making it difficult, in practice, to fully exploit this advantage. In contrast, the Raman spectrum of a tissue consists of relatively narrower bands characteristic of specific molecular vibrations of a much larger number of Raman-active biochemicals present in tissue,⁷ thus allowing one to detect molecular information in a tissue in greater detail than with fluorescence or diffuse reflectance.

A critical evaluation of the diagnostic results listed in Tables 2–5 reveals some interesting points that are worth noting. One may see that both fluorescence and diffuse reflectance consistently misclassify 30% to 45% of IDC tissue sites as normal, and the situation does not improve much even with the combined approach, which still leads to a 25% misclassification rate. In contrast, the classification accuracy of Raman alone is 99% (85 out of 86 IDC tissue sites). The likely reason for this inferior classification performance of fluorescence and diffuse reflectance becomes apparent only when one critically examines the detailed physical appearance of all the tissue samples investigated. It was found that most of the malignant tissue sites that were misclassified as normal belonged to

those IDC samples that had thin layers of fat on their surfaces. It is known that both fluorescence and diffuse reflectance with UV and visible light excitation can probe only the superficial tissue layer,^{4–6} whereas Raman with near IR excitation can probe to a much greater depth inside tissue.⁷ On a similar note, one may find that a total of ~13% to 14% of the normal tissue sites are consistently misclassified as IDC, FA, or DCIS by either fluorescence or diffuse reflectance, and the results remain the same even with the combined approach with no apparent improvement in discrimination of normal tissue sites. Although the majority of the normal breast tissue samples investigated was predominantly fatty, a few of them were glandular or fibrous as well. It was noticed that the normal breast tissue sites that were consistently misclassified belonged to the samples that had more glandular or fibrous than fatty tissues. On the other hand, the DCIS tissue sites that show very poor classification accuracy with either fluorescence or diffuse reflectance alone are seen to have a much improved accuracy of 89% (16 out of 18) with the combined approach. However, in this case, the performance of Raman is only marginally better, with 17 out of 18 classified correctly (94% accuracy).

Recently, Palmer et al.¹⁵ demonstrated a comparison of diffuse reflectance and fluorescence spectroscopy for *ex vivo* characterization of different breast pathologies. They used PCA for reducing the dimensionality of the spectral data and linear SVM for classifying the resulting diagnostically relevant principal components. Even though their PCA-SVM algorithm was successful in discriminating malignant from nonmalignant breast tissues with a sensitivity and specificity of 70% and 92%, respectively, based on autofluorescence spectra alone, the performance of the algorithm was found to provide a sensitivity of only 30% and a specificity of 78% when only diffuse reflectance spectra of the same set of breast tissue samples were incorporated for discrimination analysis. Use of both fluorescence and diffuse reflectance spectra in combination, however, was observed to make the algorithm perform much better than that with diffuse reflectance alone, though the sensitivity and specificity values were the same as that obtained using autofluorescence alone. Although a direct comparison is not possible, mainly due to the differences in experimental and data analysis methods, these observations of Palmer et al. are seen to be grossly consistent with those of ours. For example, in our case, the maximum sensitivity and specificity achieved by the algorithm based on combined autofluorescence and diffuse reflectance in discriminating malignant from nonmalignant breast tissue sites were 72% and 89%, respectively, whereas the algorithm based on fluorescence spectra alone provided a poorer sensitivity and specificity of 58% and 77%, respectively. It is pertinent to note here that Palmer et al.¹⁵ used fluorescence spectra at nine excitation wavelengths (300 to 460 nm in 20-nm increments) as input to their diagnostic algorithm, whereas we used fluorescence spectra at only 337-nm excitation for our algorithm development. Use of multiple excitation wavelengths make it possible to probe most of the tissue fluorophores that have characteristic emission over the complete UV-visible wavelength region as compared with fewer fluorophores (primarily NADH and flavins) that can be probed with the 337-nm wavelength. The possibility of incorporating a much larger number of spectral features that can be of diagnostic rel-

evance in multiexcitation as compared with single-excitation fluorescence is most likely the reason for the superior classification performance of their multiexcitation fluorescence-based algorithm as compared with that of ours based on single excitation.

Although the earlier studies by several research groups^{17–20} demonstrated the applicability of Raman spectroscopy for breast cancer diagnosis, a series of systematic studies carried out recently by the Feld group^{21–24} has provided strong evidence that the dispersive Raman spectroscopic technique has enough promise to be used as a potential diagnostic tool for real-time assessment of various breast pathologies. The chemical and morphological models-based diagnostic algorithm developed by them²² shows excellent diagnostic accuracies in classifying *ex vivo* breast tissue samples according to their specific pathologic diagnoses, attaining 94% sensitivity and 96% specificity for distinguishing cancerous breast tissues from normal and benign tissues.²³ One may note here that although there is intrinsic difference between the MRDF-SMLR-based diagnostic algorithm²⁵ employed by us and the spectroscopic model-based algorithm²² used by them for the discrimination analysis of the breast tissue Raman spectra, the resulting discrimination results in both the cases are comparable with classification accuracy consistently being in the range of 90% to 100%. This is perhaps not unexpected given the wealth of information embedded in the Raman spectra of breast tissues that can serve as a basis for diagnosis.⁷

It is pertinent to mention here that in addition to investigating the capability of fluorescence, diffuse reflectance, combined fluorescence and diffuse reflectance, and Raman, a scheme was also investigated where all three sets of spectra were combined to develop a diagnostic algorithm. This was done to explore whether this kind of trimodal spectroscopy can result in any better classification outputs. This algorithm did not, however, lead to any improvement in the classification results that were already obtained using the Raman-based diagnostic algorithm alone. This perhaps shows that blindly adding together spectral data of all three spectroscopic techniques is not the way to address a given diagnostic problem for any better accuracy.

Another important point worth considering is that when it comes to distinguishing only normal from nonnormal tissues, both Raman and combined fluorescence and diffuse reflectance are seen to have the potential to provide such discriminations reasonably accurately. For example, the diagnostic results in Table 4 show that combined fluorescence and diffuse reflectance spectroscopy can discriminate tumor from normal breast tissues with sensitivity and specificity of 83% and 85%, respectively. If the objective is only to delineate tumor from normal breast tissues, as may be required for certain clinical procedures, the combined fluorescence and diffuse reflectance approach can serve as a method of choice. This technique can be modified as well, such as adding polarization optics to provide depth-dependent information³² to perhaps overcome the limitations of fluorescence and diffuse reflectance as discussed previously. A further incentive toward using this approach is that, given currently available instruments, combined fluorescence and diffuse reflectance is a significantly stronger candidate than Raman for imaging techniques. In many applications, such as evaluating margin status during breast conservative therapy, it is highly desirable to move to

an imaging-based system rather than point spectroscopy to gather more information from an entire tissue surface in a much quicker time. At this point, one can perhaps think of exploiting the advantages of all three spectroscopic techniques in one setting by combining imaging based on combined fluorescence and diffuse reflectance with Raman-based point spectroscopy if the goal is to cover a larger area while accurately distinguishing normal from nonnormal tissue.

Acknowledgments

The authors would like to thank Ms. Evelyn Okediji for help with preparing the tissue slides for pathology. The authors acknowledge the financial support of the NCI SPORE in Breast Cancer Pilot Project (5P50 CA098131-03).

References

1. American Cancer Society, *Cancer Facts and Figures 2006*. Available at <http://www.cancer.org/downloads/STT/CAFF2006PWSecured.pdf>. Accessed January 5, 2007.
2. G. Maranto, "Should women in their 40s have mammogram," *Sci. Am.* **275**, 113 (1996).
3. J. M. Johnson, R. R. Dalton, S. M. Wester, J. Landercasper, and P. J. Lambert, "Histological correlations of microcalcifications in breast biopsy specimens," *Arch. Surg. (Chicago)* **134**, 712–716 (1999).
4. G. A. Wagnieres, W. M. Star, and B. C. Wilson, "In vivo fluorescence spectroscopy and imaging for oncological applications," *Photochem. Photobiol.* **68**(5), 603–632 (1998).
5. E. Servick-Muraca and R. Richards Kortum, "Quantitative optical spectroscopy for tissue diagnosis," *Annu. Rev. Phys. Chem.* **47**, 556–606 (1996).
6. N. Ramanujam, "Fluorescence spectroscopy of neoplastic and non-neoplastic tissues," *Neoplasia* **2**(1–2), 89–117 (2000).
7. A. Mahadevan-Jansen, "Raman spectroscopy: from benchtop to bedside," in *Biomedical Photonics Handbook*, T. Vo-Dinh, Ed., CRC Press, Washington, D.C. (2003).
8. J. Bigio, S. G. Bown, G. Briggs, C. Kelley, S. Lakhani, D. Pickard, P. M. Ripley, I. G. Rose, and C. Saunders, "Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results," *J. Biomed. Opt.* **5**(2), 221–228 (2000).
9. S. Demos, A. J. Vogel, and A. H. Gandjbakhche, "Advances in optical spectroscopy and imaging of breast lesions," *J. Mammary Gland Biol. Neoplasia* **11**(2), 165–181 (2006).
10. G. C. Tang, A. Pradhan, and R. R. Alfano, "Spectroscopic differences between human cancer and normal lung and breast tissues," *Lasers Surg. Med.* **9**(3), 290–295 (1989).
11. Y. Yang, A. Katz, E. J. Celmer, M. Zurawska-Szczepaniak, and R. R. Alfano, "Fundamental differences of excitation spectrum between malignant and benign breast tissues," *Photochem. Photobiol.* **66**(4), 518–522 (1997).
12. Y. Yang, E. J. Celmer, M. Zurawska-Szczepaniak, and R. R. Alfano, "Excitation spectrum of malignant and benign breast tissues: a potential optical biopsy approach," *Lasers Life Sci.* **7**, 249–265 (1997).
13. P. K. Gupta, S. K. Majumder, and A. Uppal, "Breast cancer diagnosis using N2 laser excited autofluorescence spectroscopy," *Lasers Surg. Med.* **21**(5), 417–22 (1997).
14. S. K. Majumder, P. K. Gupta, B. Jain, and A. Uppal, "UV excited autofluorescence spectroscopy of human breast tissues for discriminating cancerous tissue from benign tumor and normal tissue," *Lasers Life Sci.* **8**, 249–264 (1998).
15. G. M. Palmer, C. Zhu, T. M. Breslin, F. Xu, K. W. Gilchrist, and N. Ramanujam, "Comparison of multiexcitation fluorescence and diffuse reflectance spectroscopy for the diagnosis of breast cancer (March 2003)," *IEEE Trans. Biomed. Eng.* **50**(11), 1233–1242 (2003).
16. C. Zhu, G. M. Palmer, T. M. Breslin, F. Xu, and N. Ramanujam, "Use of a multiseparation fiber optic probe for the optical diagnosis of breast cancer," *J. Biomed. Opt.* **10**(2), 024032 (2005).
17. R. R. Alfano, C. H. Liu, W. L. Sha, H. R. Zhu, D. L. Akins, J. Cleary, R. Prudente, and E. Celmer, "Human breast tissue studied by IR Fourier transform Raman spectroscopy," *Lasers Life Sci.* **4**, 23–28 (1991).
18. C. J. Frank, R. L. McCreery, and D. C. Redd, "Raman spectroscopy

- of normal and diseased human breast tissues," *Anal. Chem.* **67**(5), 777–783 (1995).
19. C. J. Frank, D. C. Redd, T. S. Gansler, and R. L. McCreery, "Characterization of human breast biopsy specimens with near-IR Raman spectroscopy," *Anal. Chem.* **66**(3), 319–326 (1994).
 20. D. C. B. Redd, Z. C. Feng, K. T. Yue, and T. S. Gansler, "Raman-spectroscopic characterization of human breast tissues-implications for breast-cancer diagnosis," *Appl. Spectrosc.* **47**(6), 787–791 (1993).
 21. R. Manoharan, K. Shafer, L. Perelman, J. Wu, K. Chen, G. Deinum, M. Fitzmaurice, J. Myles, J. Crowe, R. R. Dasari, and M. S. Feld, "Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging," *Photochem. Photobiol.* **67**(1), 15–22 (1998).
 22. K. E. Shafer-Peltier, A. S. Haka, M. Fitzmaurice, J. Crowe, J. Myles, R. R. Dasari, and M. S. Feld, "Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo," *J. Raman Spectrosc.* **33**(7), 552–563 (2002).
 23. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, and M. S. Feld, "Diagnosing breast cancer by using Raman spectroscopy," *Proc. Natl. Acad. Sci. U.S.A.* **102**(35), 12371–12376 (2005).
 24. S. Haka, Z. Volynskaya, J. A. Gardecki, J. Nazemi, J. Lyons, D. Hicks, M. Fitzmaurice, R. R. Dasari, J. P. Crowe, and M. S. Feld, "In vivo margin assessment during partial mastectomy breast surgery using raman spectroscopy," *Cancer Res.* **66**(6), 3317–3322 (2006).
 25. S. K. Majumder, S. C. Gebhart, M. D. Johnson, R. Thompson, W. C. Lin, and A. Mahadevan-Jansen, "A probability-based spectroscopic diagnostic algorithm for simultaneous discrimination of brain tumor and tumor margins from normal brain tissue," *Appl. Spectrosc.* **61**(5), 548–557, (2007).
 26. W. C. Lin, S. A. Toms, M. Johnson, E. D. Jansen, and A. Mahadevan-Jansen, "In vivo brain tumor demarcation using optical spectroscopy," *Photochem. Photobiol.* **73**(4), 396–402 (2001).
 27. C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl. Spectrosc.* **57**(11), 1363–1367 (2003).
 28. A. Talukder and D. Casasent, "General methodology for simultaneous representation and discrimination of multiple object classes," *Opt. Eng.* **37**(3), 904–913 (1998).
 29. B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(6), 957–968 (2005).
 30. A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(1), 4–37 (2000).
 31. D. J. Hand and R. J. Till, "A simple generalization of the area under the ROC curve for multiclass classification problems," *Mach. Learn.* **45**, 171–186 (2001).
 32. N. Ghosh, S. K. Majumder, and P. K. Gupta, "Depth-resolved fluorescence measurement in a layered turbid medium by polarized fluorescence spectroscopy," *Opt. Lett.* **30**(2), 162–164 (2005).