

# Skin lesion classification using oblique-incidence diffuse reflectance spectroscopic imaging

Mehrübe Mehrübeoğlu, Nasser Kehtarnavaz, Guillermo Marquez, Madeleine Duvic, and Lihong V. Wang

We discuss the use of a noninvasive *in vivo* optical technique, diffuse reflectance spectroscopic imaging with oblique incidence, to distinguish between benign and cancer-prone skin lesions. Various image features were examined to classify the images from lesions into benign and cancerous categories. Two groups of lesions were processed separately: Group 1 includes keratoses, warts versus carcinomas; and group 2 includes common nevi versus dysplastic nevi. A region search algorithm was developed to extract both one- and two-dimensional spectral information. A bootstrap-based Bayes classifier was used for classification. A computer-assisted tool was then devised to act as an electronic second opinion to the dermatologist. Our approach generated only one false-positive misclassification out of 23 cases collected for group 1 and two misclassifications out of 34 cases collected for group 2 under the worst estimation condition. © 2002 Optical Society of America

OCIS codes: 170.0170, 110.0110, 100.0100, 300.0300.

## 1. Introduction

Skin cancer is an increasing problem around the world. According to the annual Cancer Statistics Review, an estimated 42,000 new cases of melanoma of the skin in the United States were reported for 1998.<sup>1</sup> This number jumped to 47,700 estimated new cases in the year 2000.<sup>2</sup> Melanoma is expected to be the cause of 7700 deaths in 2000.<sup>2</sup> Nonmelanoma skin cancers account for ~40% of all the diagnosed cancers.<sup>3</sup> Among the nonmelanoma cancers, squamous cell carcinomas cause 1500 deaths in the United States each year.<sup>4</sup> Basal cell carcinomas are not usually fatal and grow slowly; however, they can penetrate deep into the bone and cause local destruction.<sup>5</sup> Basal cell carcinomas could cause morbidity if

untreated. Almost all skin cancers are curable, if detected early.

Currently, clinical dermatologists rely on visual inspection and experience to make an initial assessment of the skin lesion state. If further visual analysis does not provide a conclusive decision on the lesion state, such suspicious lesions are sent for biopsy analysis. Biopsy is unpleasant for the patient, slow in diagnostic results, and costly for the hospital on account of the wait time. Dermatologists would greatly benefit from a fast and noninvasive technique that could assist them in their clinical diagnostic decisions.

Several groups have worked on a number of noninvasive methods to detect abnormal cells.<sup>6-8</sup> Rönning and Riech worked on a skin cancer detection system based on computer vision that registered lesions on multiple patient visits and compared the successively collected lesion images with previous ones to determine changes in color, shape, size, and other features on the basis of asymmetry, border irregularity, color variation, and texture inside a mole.<sup>9</sup> Cooney *et al.* reported studies on the development of an optical system for detecting oral cancers with near-infrared spectroscopy.<sup>10</sup> In this system a Fourier-transform infrared spectrometer was used to obtain a single spectrum from each malignant, benign, and inflamed hamster cheek tissue. Feld's group reported research on polarized light-scattering spectroscopy for measuring epithelial cell structures

---

When this research was performed, M. Mehrübeoğlu (mehrubeoglum@esi.com), N. Kehtarnavaz, G. Marquez, and L. V. Wang were with A&M University, College Station, Texas 77842. M. Mehrübeoğlu, N. Kehtarnavaz, and L. V. Wang were with the Department of Electrical Engineering, and G. Marquez and L. V. Wang were with the Biomedical Engineering Program. M. Duvic was with the Department of Dermatology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030. M. M. Mehrübeoğlu is currently with ESI, Inc., 13900 NW Science Park Drive, Portland, Oregon 97229.

Received 20 October 2000; revised manuscript received 7 May 2001.

0003-6935/02/010182-11\$15.00/0

© 2002 Optical Society of America

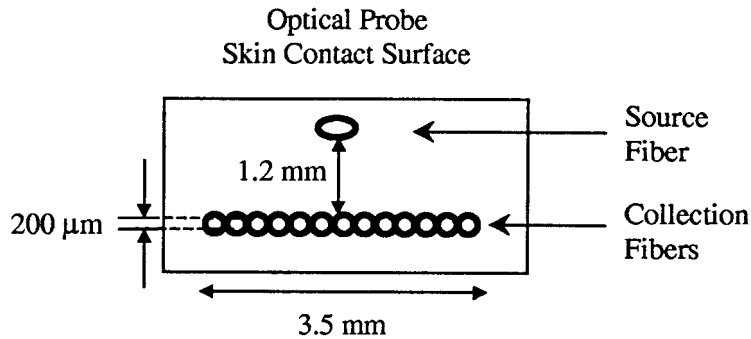


Fig. 1. Optical probe for light transmission. The source fiber transmits the white light onto the skin surface. The diffusely reflected light off the skin surface is collected by a series of 13 optical collection fibers.

quantitatively in an attempt to detect and diagnose precancerous changes in human tissues.<sup>11</sup> Wallace *et al.* discussed assessing pigmented skin lesions with spectrophotometry and feature selection techniques.<sup>12</sup> Some of their successful features included slopes, curvature, mean, and variance within regions of spectra.

In this paper we investigate diffuse reflectance spectroscopic imaging (DRSI), a fast and noninvasive technique, with oblique incidence,<sup>6</sup> to identify benign and cancerous or cancer-prone skin lesions for the purpose of reducing unwanted benign biopsies. DRSI with oblique incidence has an advantage over normal incidence by gathering information from superficial layers without penetrating deep tissue.<sup>6</sup> This is of particular importance, since skin cancer information is usually present in the top layers of the skin tissue, and deeper layers only add to the background noise in the signal.<sup>11</sup> In addition, oblique-incidence DRSI is sensitive to scattering and is able to separate absorption from scattering in homogeneous scattering media. Oblique-incidence DRSI is used in a multifiber array mode such that spatial information is also gathered along tissue surface. We have combined the DRSI technique with image processing and pattern recognition to provide the dermatologist with a computer-assisted tool to determine the state of unknown lesions in clinical settings.

In the following sections we describe the equipment and setup used in this study, images of skin lesions collected *in vivo*, feature extraction and selection, effective features, and classification. The information conveyed by the effective features is discussed in Section 7.

## 2. Equipment and Setup

The image acquisition for the DRSI system is composed of a light source, transmission (source and receiver) fibers, a spectrograph, a CCD camera, and a personal computer. The CCD camera has a  $512 \times 512$  pixel array of area  $9.7 \text{ mm} \times 9.7 \text{ mm}$ . A halogen lamp (Dolan-Jenner Industries, 100 W) is used as the broadband white-light source. The source light is delivered to the skin surface via a source fiber within a fiber-optic probe (Fig. 1). The fiber-optic probe is made from brass tubing and 200- $\mu\text{m}$ -diameter, low-OH optical fibers. The optical fibers consist of 1 source fiber and 13 receiver fibers. The source fiber is oriented at  $45^\circ$  with respect to the horizontal axis. The 13 vertical receiver fibers are arranged in a linear array spanning 3.5 mm. The receiver fibers collect the diffusely reflected light from the skin surface and deliver the collected light to the spectrograph (Oriel, Multispec 257; Fig. 2). The probe is positioned on the skin surface without pressure for image acquisition. The outputs from the receiver fibers are

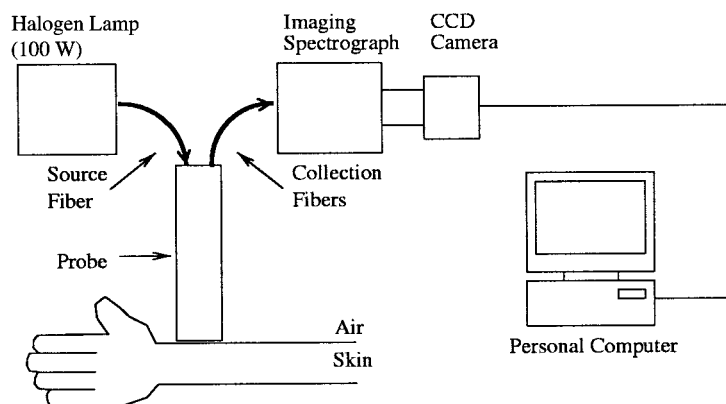


Fig. 2. DRSI system. The spectral information from the diffusely reflected light off the skin surface is transmitted to the CCD camera system via the probe and the imaging spectrograph.

placed at the object plane of the imaging spectrograph. The spectrograph disperses the one-dimensional (1-D) light from each fiber into the spectral components between 518 and 766 nm. The spectra from the 13 fibers are then projected in the form of a two-dimensional (2-D) image onto the CCD matrix (Princeton Instruments, Inc., Model 1530P), as shown in Fig. 2.

The vertical dimension of the CCD array represents the spatial distribution of the diffuse reflectance in terms of fiber locations. The horizontal dimension represents the spectral distribution for the diffusely reflected light collected by each receiver fiber. The personal computer controls the image-acquisition parameters and records the spectral images automatically.

### 3. Collected Images of *in vivo* Skin Lesions

Only the suspicious lesions, whose state the dermatologist could not determine by visual inspection, are included in the sample pool. The images are subdivided into two types of cancer groups, namely, carcinomas (basal and squamous cell carcinomas) (group 1) and melanomas (group 2). The dermatologist is able to separate the two cancerous subgroups, or suspicious lesions belonging to these subgroups, by color and other visual textural characteristics; however, she may not be able to separate the benign from the cancerous lesions within each subgroup.

The collected samples consist of benign (keratoses, warts) and cancerous (squamous and basal cell carcinoma) lesions from group 1 and of benign (common nevi) and intermediate or potentially cancerous (dysplastic nevi) lesions from group 2. These samples were obtained at the University of Texas MD Anderson Cancer Center in Houston, Texas. Among the tested lesions, biopsy reports have identified 15 lesions belonging to group 1 as benign and 8 lesions as cancerous for a total of 23 lesions. For group 2, 13 lesions are benign and 21 intermediate for a total of 34 lesions. The lesions are 2–8 mm in diameter.

Multiple (three to five) spectral images have been acquired from each suspicious lesion and the neighboring healthy skin. The images from the healthy skin are collected away from the lesion site to prevent signal interference from potential extended lesions beneath the skin. In group 1 a total of 222 image samples have been collected from 23 cases, 111 lesion and 111 healthy images (73 benign and 38 cancerous). In group 2, image samples from 34 cases have been acquired for a total of 340 images, 170 lesion and 170 healthy images (65 common and 105 dysplastic nevi). For each image the probe is removed and randomly repositioned on the lesion or healthy skin surface to obtain multiple and minimally dependent images from different locations or orientations on the skin surface.

### 4. Feature Extraction and Selection

For each image the individual spectra from the fibers are first corrected for the instrument function by a calibration factor. We obtain the calibration factor

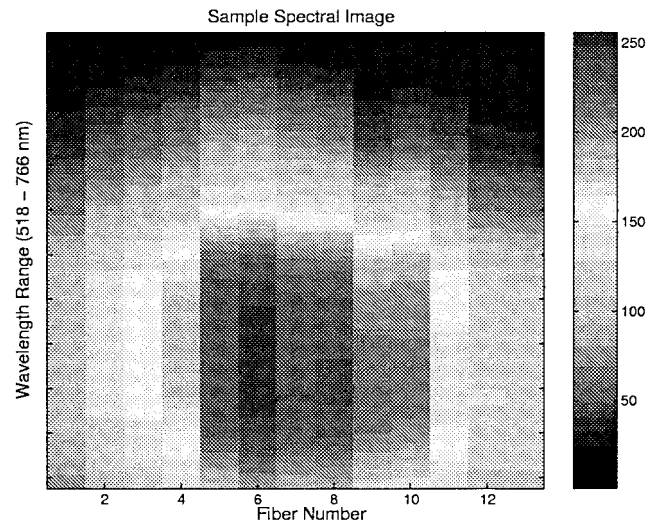


Fig. 3. Sample image. Horizontal axis, location of fibers. Vertical axis, wavelength at which the light was collected. Different shades of gray in the image represent the relative intensity of the received light. (Correct aspect ratio is not shown, for visual clarity.)

from the camera by turning off the delivery fiber light and acquiring a calibration image, which is necessary, since fiber material as well as mechanical mounting may cause variations in individual fiber response to the collected light. The individual spectra from each fiber are separately filtered with a 1-D five-point running-average filter and then intensity scaled. The scaling operation compresses the image intensities between 0 and 255 for an 8-bit-resolution (256 gray levels) image. The images from different lesion groups and classes look similar to the naked eye. A sample diffuse reflectance spectroscopic image is shown in Fig. 3.

During data collection the probe is placed on the apparent center of the lesion surface. The total probe contact area is 3.5 mm, which indicates that for lesions smaller than this size, the edge fibers collect diffusely reflected light from the surrounding apparently healthy tissue. For this reason information from the first and last two fibers (fibers 1, 2, 12, 13) are removed from the analysis of all the images, to allow a consistent analysis procedure for all the image samples.

Feature extraction involves first identifying features that discriminate among classes. In this study a comprehensive list of features have been tested, including 1-D signal features (spectral features from a single fiber and spatial features from multiple fibers) and 2-D texture features taking advantage of the combined frequency and spatial information. New features that capture the specific characteristics of the images acquired by the DRSI system have been introduced.

The classes to be separated must contain distinct within-class similarities and between-class differences. Features depicting these similarities and differences can then be extracted and used for image classification. Visual inspection of the lesion data

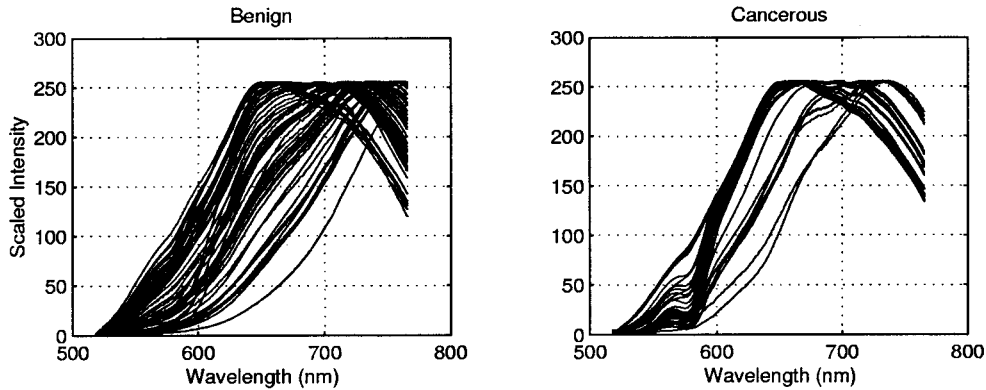


Fig. 4. Sample intensity-scaled single-fiber spectra from benign (left) and cancerous (right) skin lesions in group 1.

reveal only subregional local variations among the spectra belonging to the classes (see Fig. 4 for spectra from group 1 and Fig. 5 for group 2). Using information from specific spectral ranges instead of the entire spectrum allows the extraction of small differences between the classes that would otherwise be averaged out and missed when the entire spectrum is used. Using spectral ranges rather than the entire spectrum permits extraction of unique features effective at particular wavelength ranges.

To find regions and widths of regions that provide information for optimum class separation, a region search algorithm (RSA) has been developed to extract features at various spectral locations and region widths (1-D analysis). The results for each of the tested features are then plotted on separate scatter plots with each data point labeled according to the class to which the spectral subregion belongs. A statistical discriminant analysis is then performed to test the success of the extracted features in separating the classes, as further explained in Section 5.

Similarly, image block analysis (2-D) has been carried out with the same RSA. Features from subimages, obtained by means of combining information from selected fibers and wavelengths, have been tested with a moving window and varying window sizes along the length of the image to determine 2-D

regions that yielded information for maximum separability. The window *width* is kept at 9 pixels corresponding to the maximum number of nine analyzed fibers. The window *length*,  $L_i$ , is changed from 5 to 508 pixels as we search for the optimal window length and location for each tested feature. With this procedure, different optimum subregions are found for different features. A schematic diagram of the regions searched by the RSA is depicted in Fig. 6.

Fisher discriminant value and the area under receiver operating characteristic (ROC) curves are used as two measures of feature effectiveness. Fisher distance is a measure of class separability.<sup>13</sup> The mean and standard deviation of the feature values in each class are used to compute the Fisher distance. For a two-class problem, the Fisher distance is computed by

$$F_{bc} = |\mu_b - \mu_c| / (\sigma_b^2 + \sigma_c^2)^{1/2}, \quad (1)$$

where  $\mu_b$  is the mean of the feature values for the benign class,  $\mu_c$  is the mean of the feature values for the cancerous class,  $\sigma_b$  is the standard deviation of the feature values for the benign class, and  $\sigma_c$  is the standard deviation of the feature values for the cancerous class. (In group 2,  $\mu_c$  and  $\sigma_c$  correspond to the intermediate lesion parameters.)

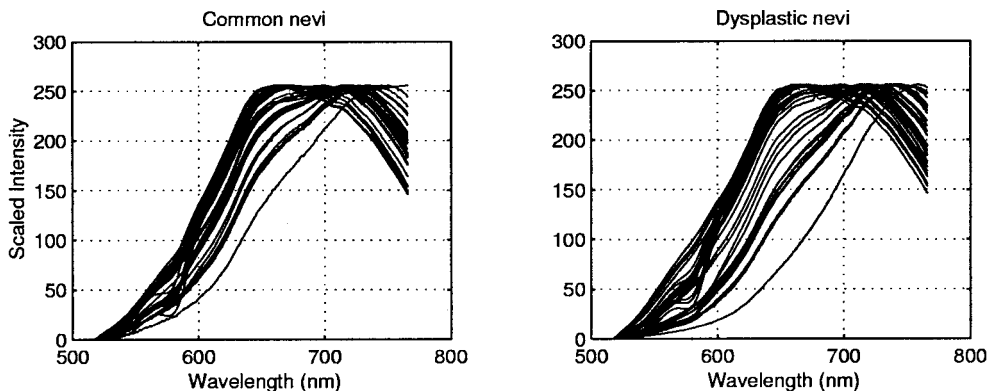


Fig. 5. Sample intensity-scaled single-fiber spectra from benign or common nevi (left), and intermediate or dysplastic nevi (right) lesions in group 2.



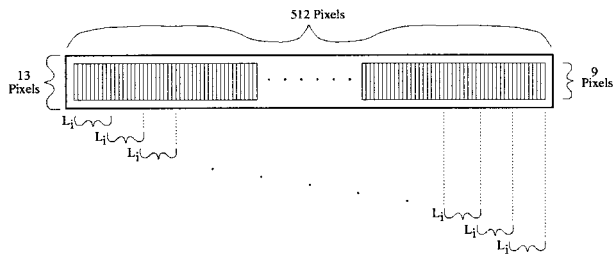


Fig. 6. Schematic diagram of the region scanned by the RSA. The width of the window corresponding to fiber number was kept at 9 pixels. The length,  $L_i$ , and location of the window were varied to find the optimum region and width of region for a given feature.

As can be seen from Eq. (1), the Fisher distance increases as the difference between the means of the feature values for the two classes increases or as the separability between the two classes increases. Similarly, as the standard deviation of the feature values in each class decreases, the Fisher distance increases. This indicates that the extracted features can be ranked with the Fisher distance, and features that generate relatively high Fisher distances can be considered effective.

In this paper we report the product of the Fisher distance and the area under the ROC curve as a measure of feature effectiveness for a more reliable feature-ranking scheme. ROC curves characterize the true positives and false positives for different threshold values on the scatter plots. The area under the ROC curve describes accuracy<sup>14</sup> and represents the efficiency or performance of the feature. For good separation, the area under the ROC curve must be maximized. The ROC curve is plotted as the true-positive fraction versus false-positive fraction at different threshold values. The ROC curves represent sensitivity versus  $(1 - \text{specificity})$ . Sensitivity is defined as the fraction of correctly identified (true) positive events [cancerous (group 1) or intermediate (group 2) lesions] over all the positive events. Specificity is given by the total correctly identified (true) negative events over all the negative events (benign lesions).<sup>15</sup> The features are arranged in descending order according to the product of the Fisher distance and the ROC area. The effective features are selected from the top of the list with the highest product values. The selected features are tested for correlation. The features contributing least to separation (features corresponding to the same outliers in scatter plots) are not selected.

## 5. Effective Features

The ideal features separate classes with 100% sensitivity and specificity. In reality, perfect separation is rarely achieved. The goal, then, is to identify different features that provide for separability of classes with minimum overlap for different data samples so that when multiple features are combined, classification can be performed successfully.

Table 1. List of Examined Features<sup>a</sup>

1-D and 2-D Features		
Statistical		
Mean	Kurtosis	Second moment
Variance	Energy	Third moment
Skewness	Entropy	Fourth moment
Co-occurrence matrix based		
Mean	Contrast	Cluster shade
Variance	Inertia	Cluster prominence
	(contrast)	
Energy	Inverse	Dissimilarity
	recursivity	
Entropy	Maximum	
	probability	
Wavelet based		
Mean	Kurtosis	Second moment
Variance	Energy	Third moment
Skewness	Entropy	Fourth moment
Fractal signature		
Principal component		
First	Fourth	Seventh
Second	Fifth	Eighth
Third	Sixth	Ninth

<sup>a</sup>Ref. 17.

Image texture features were studied to identify effective texture features in the spectroscopic skin images. Broadly speaking, image texture is defined as the visual effect produced by spatial distribution of tonal variations within small areas.<sup>16</sup> A wide collection of image texture features based on statistics, co-occurrence matrix, wavelet decomposition, fractals, and other geometric relations within the images were extracted and examined in order to obtain those features whose distributions were relatively nonoverlapping for the benign and cancerous cases (see Tables 1 and 2 for a list of examined features).

Tables 3 and 4 summarize the effective features both in 1-D and 2-D (texture related) found for groups 1 and 2, respectively. For group 1, wavelet-based energy, wavelet-based skewness, polar curvature, mean, skewness and volume-approximation features have been found effective, as listed in Table 3. Effective features found for group 2 include polar, statistical (variance, mean, skewness), subspectral dip, and wavelet-based entropy. The details of all the features used are beyond the scope of this paper. The interested reader is referred to Ref. 17 for details.

Table 2. List of Introduced Features<sup>a</sup>

Introduced Features	
Diffuse reflectance	Subspectral dip
Profile gradient	Volume approximation
Point-source shift	Polar coordinate

<sup>a</sup>Ref. 17.

**Table 3. Effective Features for Group 1 from Which the Classifier Features are Selected**

Feature	Distinguishing Spectral Region Pixels (Wavelength, nm)	FD <sup>a</sup>	ROCA <sup>b</sup>	FD * ROCA
Biorthogonal wavelet energy (H)	460–505 (741–763)	1.53	0.95	1.45
Daubechies wavelet skewness (H)	340–403 (682–713)	1.52	0.94	1.43
Skewness	115–142 (574–587)	1.53	0.90	1.38
Polar geometry (1-D) (area, fiber 7)	275–347 (651–686)	1.47	0.83	1.22
Mean	115–142 (574–587)	1.45	0.83	1.19
Volume approximation	480–498 (750, 759)	1.28	0.88	1.12

<sup>a</sup>FD, Fisher distance.

<sup>b</sup>ROCA, ROC curve area.

An overview of the features found effective is provided below.<sup>17</sup>

**A. Statistical Features**

Statistical features include mean (first moment), variance, skewness, kurtosis, energy, and other moments (second, third, and fourth). These statistical attributes are calculated for the entire image, and subimages (image subblocks) for 2-D analysis. The same attributes are computed for the entire spectrum from a particular fiber, or a subregion of a spectrum

for 1-D spectral analysis, as well as for spatial curves from multiple fibers at averaged wavelengths for 1-D spatial analysis. The RSA has been used for subimage and subspectral feature computations.

**B. Wavelet-Based Texture Features**

Multiresolution wavelet analysis is a recently developed scheme<sup>18</sup> that represents data at various scales with different time or spatial resolutions. Single-level wavelet transforms have been acquired for each image, generating four subimages, each half in reso-

**Table 4. Effective Features for Group 2 from Which the Classifier Features Are Selected**

Feature	Distinguishing Spectral Region Pixels (Wavelength, nm)	FD <sup>a</sup>	ROCA <sup>b</sup>	FD * ROCA
Polar statistical (third moment) (fiber 7)	215–251 (622–639) (622–639)	1.16	0.82	0.96
Variance (fiber 8)	235–307 (632–666)	1.10	0.83	0.91
Mean (fiber 6)	375–393 (699–708)	1.07	0.85	0.91
Skewness (fiber 6)	250–259 (639–643)	1.03	0.85	0.88
Subspectral dip (fiber 6)	160–178 (595–604)	1.01	0.82	0.83
Wavelet-based entropy	285–384 (656–704)	0.98	0.81	0.79
Polar statistical (mean) (fiber 5)	190–208 (610–619)	1.02	0.83	0.70

<sup>a</sup>FD, Fisher distance.

<sup>b</sup>ROCA, ROC curve area.

lution with respect to the height and width of the original image. These subimages comprise one coarse and three detail subimages. The horizontal (H), diagonal (D), and vertical (V) component subimages correspond to the high-frequency part, whereas the coarse component subimage (C) corresponds to the low-frequency part of the image. Statistical features are then computed from each subimage.

In this study various types of wavelets have been investigated, in particular, Daubechies and biorthogonal wavelets. For the data set examined, the mean and energy calculated from the biorthogonal and skewness from the Daubechies horizontal images have produced well-separated features for group 1. Only the wavelet-based energy feature has been found to be an effective feature for group 1 classification, since both the mean and energy features produced the same outlier samples (correlated feature set) in the scatter plots. The entropy from the Daubechies horizontal image has been found to be an effective feature for group 2.

### C. Introduced Features

#### 1. Subspectral Dip

As can be seen from Fig. 4, the benign and the cancerous classes (group 1) contain local differences in their single-fiber spectra. The region around pixels 87–142 corresponding to the wavelength range 560–587 nm, for example, displays visible differences for the benign and the cancerous classes (see Fig. 4). It was discovered that there is a prominent dip in the spectral information for cancerous lesion images that was absent or not so prominent in the benign lesions. This dip is characterized and used as an effective feature in separating the two classes.

We characterize the dip by first intensity-scaling the region of interest and then finding the best linear fit through scaled region. This method of line approximation is chosen because of its reduced sensitivity to edge and outlier points in the spectrum, which might otherwise adversely affect the subspectral area calculations. The calculated line is then moved by the mean of the scaled curve. We find the area between the best-fit line and the curve by taking the absolute difference between the two lines. The spectral region is characterized for each fiber separately. Scaling brings out the spectral changes in the tested region.

#### 2. Volume Approximation

We calculate this feature by finding the volume under subregional areas, using the same method to characterize the subspectral dip stated in Subsection 5.C.1. In this case, the information from multiple fibers is used to calculate the volume, incorporating the spatial information. The volume is calculated by summation of the area under multiple-fiber spectra. Only the conformational characteristics of the subregional area are necessary for this feature; therefore the subregional pixel values are scaled to 8-bit resolution before the volume is calculated.

### 3. Polar-Coordinate Geometry

Polar coordinates are used to transform the image information from the Cartesian coordinate system. Polar coordinates provide radial and angular information for each intensity point on the Cartesian system, from the center of origin. Polar coordinates are used to amplify the small angular changes in subregions of spectra, hence providing a measure of change within a given region in spectra between classes. Angular representation of a subregion also brings out spectral fluctuations.

### 4. Polar-Coordinate Statistics

The statistical attributes are applied to the polar curves. Some of the polar-coordinate-based statistical features have produced separable results in image subregions. The RSA is used to find the best spectral regions from individual fibers for a given statistical attribute.

## 6. Classification

As revealed by the scatter plots, each of these features generated false-negative or false-positive values. Using multiple features as the input to a classifier reduces the sensitivity to single-feature outliers and minimizes the likelihood of such false positives and negatives.

### A. Feature Conditioning

Correlation may exist among the multiple features chosen. We applied feature conditioning to the feature vectors before feeding them into the classifier. Also, feature conditioning further reduces dimensionality of the input feature vector. We achieve feature conditioning by first computing the within and the between scatter matrices,  $S_W$  and  $S_B$ , for different combinations of effective features of Tables 3 and 4. Then the separability measure,  $J$ , for the effective features is calculated per group as<sup>19</sup>

$$J = S_W^{-1} S_B. \quad (2)$$

Computing the eigenvalues and eigenvectors revealed that there is only one prominent eigenvalue when the effective features are linearly combined for both groups; therefore the conditioned feature,  $\mathbf{x}^*$ , is obtained from the most prominent eigenvalue,  $\lambda_1$ , and its corresponding eigenvector,  $v_1$ , as follows:

$$\mathbf{x}^* = \lambda_1^{-1/2} [v_1]^T (\mathbf{x} - \mu), \quad (3)$$

where  $v_1$  is a row vector,  $\mathbf{x}$  is the original input feature vector, and  $\mu$  is the overall mean vector.

The scatter plots for the conditioned feature used for classification are displayed in Figs. 7 and 8 for the two groups. The corresponding ROC curves computed from the Gaussian distributions are shown in Fig. 9. The classifier is described next.

### B. Bootstrap-Based Bayes Classifier

In the case of a limited number of samples and observations of these samples, the bootstrap method<sup>20</sup> is used to generate multiple observations,  $\mathbf{x}$ , of the

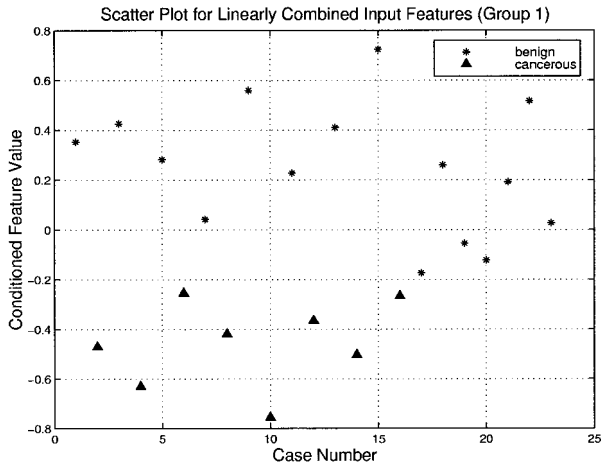


Fig. 7. Scatter plot for the conditioned classifier input feature for group 1.

samples from the original data set. For a parameter,  $\theta_k$ , that is to be estimated, first  $\mathbf{B}$  realizations of bootstrap samples are generated from the original observation by selection of random samples from the original data set with replacement. This means that in the newly generated sets some samples may appear more than once, whereas others may not appear at all.

In the case of Gaussian approximation for the estimated parameter,  $\theta_k$ , the mean vector,  $\mu_k$ , of the estimated parameter is calculated from the bootstrap samples as follows<sup>21</sup>:

$$\mu_k = \frac{1}{\mathbf{B}} \sum_{b=1}^{\mathbf{B}} \theta_k^b, \quad (4)$$

where  $\theta_k^b$  is the bootstrap estimation from the  $b$ th bootstrap-generated sample set. The mean vectors for the estimated parameters (mean and variance) are then used in the Gaussian likelihood function for classification.

The bootstrap estimation of feature parameters

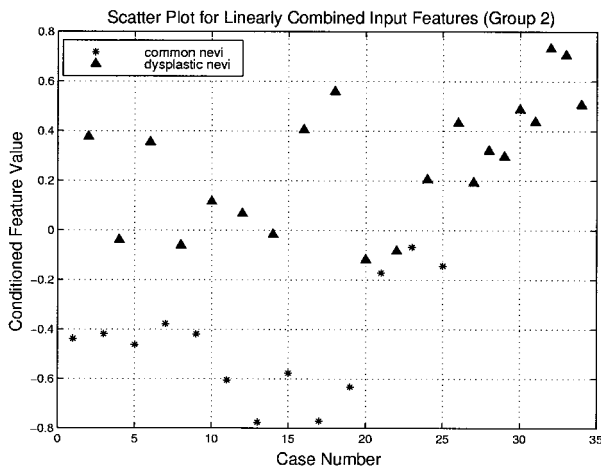


Fig. 8. Scatter plot for the conditioned classifier input feature for group 2.

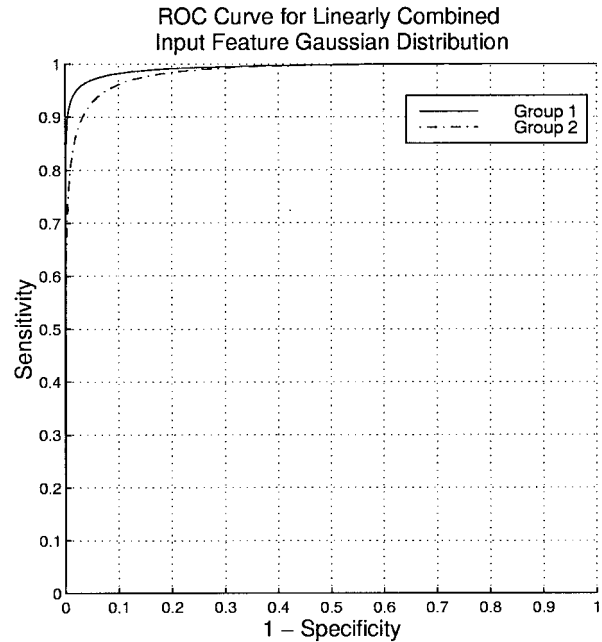


Fig. 9. ROC curve for the features displayed in Figs. 7 and 8.

has been applied to the effective features listed in Tables 3 and 4, respectively, after feature conditioning. We generated 1000 bootstrap samples for each class separately, using the conditioned features (single input feature). The results did not change when 10,000 bootstrap samples were generated. The mean vector for mean and covariance of the generated samples are computed by Eq. (4). For a single feature, the covariance matrix is simply the variance. The distributions for each feature can be obtained by use of the bootstrap estimation for mean and variance and substitution for these values in the Gaussian distribution equation.

The leave-one-out approach has been adopted for classification. With this approach, a feature belonging to a lesion is left out, and bootstrap samples are generated separately for each class with the remaining features. The unknown parameters (mean and variance) are computed for the Gaussian distributions. The classifier is tested with the left-out feature. The classification rate (hit or miss) is recorded. The procedure is repeated by means of reinserting the removed feature, taking out a different lesion feature, and regenerating new bootstrap samples, until all the lesion features have been removed and the classifier tested with all the left-out features. We note that the leave-one-out method provides a pessimistic estimate of the error probability, in that the error calculated by this method is the upper bound for classification error.<sup>18</sup> A more optimistic approach to estimating the classification rate (low classification error) would be to use the resubstitution method, in which all the lesions are used in generating bootstrap samples. The unknown parameters are estimated with the generated samples, and the classifier is tested with the original training



**Table 5. Performance of the Bootstrap-Based Classifier Designed for Lesions in Group 1<sup>a</sup>**

Number of Input Features	Added Features	Leave-One-Out Hit Rate (out of 23)	Resubstitution Hit Rate (out of 23)
1	Biorthogonal wavelet energy	19	21
2	Daubechies wavelet skewness	21	21
3	Skewness	21	21
4	Polar geometric	21	22
5	Mean	22	22
6	Volume approximation	22	22

<sup>a</sup>The classification rate versus the number of conditioned input features is shown.

features. The resubstitution method constitutes the lower bound for classification error.<sup>18</sup> Tables 5 and 6 show the bootstrap-based Bayes classifier performance with different number of input features for groups 1 and 2.

Using the bootstrap-based Bayes classifier with the leave-one-out method, we correctly identified all cancerous lesions (100% sensitivity) for group 1. There was only 1 false positive (1 benign lesion misclassified) in this group. Sensitivity is particularly important in clinical diagnoses so as not to miss any cancerous lesions. In group 2, 32 out of 34 cases were classified correctly, with 1 false positive and 1 false negative. With the resubstitution method, the classifier achieved 100% correct classification rate for group 2 (all benign and cancer-prone skin lesions correctly identified). The same hit rate (22 out of 23 lesions correctly identified) has been reached for group 1. The resubstitution method required fewer input features to generate the best outcome in comparison with the leave-one-out method, as intuitively expected. We note that including more features did not improve the classification outcome. This is primarily due to the reduced effectiveness of the added features.

Table 7 summarizes the procedure to classify unknown lesions as described in this paper.

### 7. Information Conveyed by Features

The nature of the acquired images and what they represent play a role in determining the extracted features and their effectiveness. The spectral nature of the images collected via multiple fibers suggests that there are particular regions in the images,

or spectra, that provide more information on the lesions than the entire image, since the analysis of the entire image would average out some of the small details present in spectral areas.

Correlating the effective features with the physiological parameters requires a close look at the spectral regions where these features are found to be effective, and the tissue response to light (physiological parameters) at these particular wavelengths. An effective feature reflects how much the corresponding spectral region (wavelengths) and spatial profile (fiber position) in the image differ between and within classes for a given lesion group. An effective feature with a large performance value (product of Fisher distance and ROC area, as adopted in this study) suggests that the compared spectral and spatial regions in the image display distinct variations between classes, and similarities within classes; therefore the effective features help in identifying regions of highest distinction between and lowest distinction within classes. The next task is to look at these spectral regions identified by the features within the image and investigate which physiological responses are known to occur in the found wavelength range for both benign and cancer-prone lesions.

One reason for the success of the subspectral image analysis can be explained by the absorption and scattering properties of the analyzed tissue. These optical properties are affected by the structure, composition, and homogeneity of the tissue at hand. The diffusely reflected spectral light exhibits variations in certain regions of the spectral curves depending on the hemoglobin content, nucleus size within

**Table 6. Performance of the Bootstrap-Based Classifier Designed for Lesions in Group 2<sup>a</sup>**

Number of Input Features	Added Features	Leave-One-Out Hit Rate (out of 34)	Resubstitution Hit Rate (out of 34)
1	Polar statistical (third moment)	27	29
2	Variance	27	29
3	Mean	28	31
4	Skewness	32	32
5	Subspectral dip	32	33
6	Daubechies wavelet entropy	32	34

<sup>a</sup>The classification rate versus the number of conditioned input features is shown.

**Table 7. Summary of Procedure to Classify Unknown Lesions**

Step	Procedure
1	During clinical examination, determine the lesion group (group 1 or group 2).
2	Acquire at least three images from the same unknown lesion and neighboring healthy skin. For each measurement, remove and reposition the probe on the lesion at random orientations.
3	Apply image calibration against camera and fiber response.
4	Extract effective features from the unknown images (use the spectral regions and corresponding effective features listed in Tables 3 and 4 for groups 1 and 2, respectively).
5	Apply feature conditioning to reduce multiple features to a single feature.
6	Classify the image with the one conditioned feature value and the devised classifier for the group.

the cell, the structure of other microstructures within the cell, as well as the homogeneity and compactness of these structures, which are all known changing parameters in cancerous tissues.<sup>12,22</sup> Hemoglobin, for example, has known absorption bands in the visible spectrum that peak at 540 and 576 nm (oxyhemoglobin) and at 555 nm (deoxyhemoglobin).<sup>22</sup> The different hemoglobin-carrying blood content among lesions could explain the spectral dips observed in the diffuse reflectance polarimetric images from carcinoma (cancerous) lesions that were generally absent in the keratosis and wart (benign) lesions. This could further explain the effective features, skewness and mean, in the spectral region 574–587 nm for group 1. Similar explanations could be given to spectral variations in other regions of the collected spectra, considering the dissimilarities in structural and molecular content among benign and cancerous lesions.

Researchers have shown that the direction and orientation at which the light is collected will affect the spectra of diffusely reflected light.<sup>23</sup> In fibrous or collagenous structures, for example, muscles, the orientation of the light-collecting probe with respect to the direction of the fibers within the tissue causes variations in the diffusely reflected spectra. Although the lesions are not fibrous in nature, the inhomogeneities within the lesions could explain the variance among the images and, hence, the features collected from the same lesions by means of repositioning and reorienting the light-collecting probe. This area needs further investigation.

In this paper, factors such as age, race, and gender of the patients from whom data were collected were not taken into consideration. Because of the limited number of lesion samples, all the images were grouped together in our analysis. The effects of age, race, and gender on research results in multiple fields are well documented. The melanin pigment attributed to skin color, for example, also has absorption characteristics in the studied visible spectral region and could cause further variations in the collected

images even within the same class (benign or cancerous). Ideally the above factors must be considered and data subgrouped accordingly.

## 8. Conclusions

Image-processing texture feature extraction provides a computer-based method for separating benign from precancerous skin lesions. Finding effective image features that describe each class most distinctly is one of the most challenging tasks in any image-classification problem. In this study various features have been extracted from benign (keratoses, warts) and cancerous (carcinomas) skin lesion images in group 1 in one and two dimensions. The same features have been applied on common and dysplastic nevi images of group 2. The analysis technique combines effective features (features that effectively separated the two lesion classes) with an automatic classification algorithm using a bootstrap-based Bayes classifier.

The designed leave-one-out bootstrap-based Bayes classifiers for group 1 and group 2 lesions have achieved classification rates of 95.7% and 94.1%, respectively, identifying 22 out of 23 and 32 out of 34 lesions correctly for the two groups. All cancerous lesions have been correctly identified for group 1 (sensitivity of 100%) with only 1 false positive (specificity of 93.3%). For group 2 there were 1 false positive and 1 false negative, reaching a sensitivity of 95.2% and specificity of 92.3% in the identification of common nevi and dysplastic nevi. The resubstitution method has produced more-optimistic results with 100% correct classification for group 2. The hit rate for group 1 has not improved with the resubstitution method; however, the same results have been obtained with a lower number of input features in comparison with the leave-one-out method.

The results are encouraging for the development of a noninvasive skin lesion classification system that dermatologists could use as an electronic second opinion in clinical settings. This research has identified effective features that could be used with the current images. More images need to be acquired and added to the current database to increase the confidence of the findings.

This research has been supported by the Texas A&M University Interdisciplinary Research Initiative and by a grant from the U.S. Department of Education. Institutional Review Board (IRB) approval was obtained for collecting skin images from patients at the University of Texas MD Anderson Cancer Center, Houston, Texas (IRB TAMU 2000-207). We thank the patients for their participation in this research.

## References

1. *Statistical Abstract of the United States 1998* (U.S. Census Bureau, Washington, D.C., 1998), Table 239, p. 152.
2. R. T. Greenlee, T. Murray, S. Bolden, and P. A. Wingo, "Cancer statistics, 2000," *CA Cancer J. Clin.* **50**, 7–33 (2000).
3. S. P. Stratton, R. T. Dorr, and D. S. Alberts, "The state-of-the-

- art in chemoprevention of skin cancer," *Eur. J. Cancer* **36**, 1292–1297 (2000).
4. R. W. Demetrius and H. W. Randle, "High-risk nonmelanoma skin cancers," *Dermatol. Surg.* **24**, 1272–1292 (1998).
  5. "The sun, UV, and you: a guide to sun wise behavior," Air and Radiation publication EPA 430-K-00–035 (U.S. Environmental Protection Agency, Washington, D.C., 1999).
  6. G. Marquez and L.-H. Wang, "White light oblique incidence reflectometer for measuring absorption and reduced scattering spectra of tissue-like turbid media," *Opt. Express* **1**, 454–460 (1997), <http://www.opticsexpress.org>.
  7. *Advances in Optical Imaging and Photon Migration*, C. Chang-Hasnain, ed., Vol. 15 of OSA Trends in Optics and Photonics Series (Optical Society of America, Washington, D.C., 1998).
  8. K. Tumer, N. Ramanujam, J. Ghosh, and R. Richards-Kortum, "Ensembles of radial basis function networks for spectroscopic detection of cervical precancer," *IEEE Trans. Biomed. Eng.* **45**, 953–961 (1998).
  9. J. Rönning and M. Riech, "Registration of nevi in successive skin images for early detection of melanoma," in *Proceedings of the Fourteenth International Conference on Pattern Recognition*, A. K. Jain, S. Venkatesh, and B. C. Lovell, eds. (IEEE Computer Society Press, Los Alamitos, Calif., 1998), Vol. 1, pp. 352–357.
  10. K. M. Cooney, K. W. Gossage, M. J. McShane, E. W. J. van der Breggen, M. Motamedi, and G. L. Coté, "Development of an optical system for the detection of oral cancer using near-infrared spectroscopy," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, H. K. Chang and Y. T. Zhang, eds. (Institute of Electrical and Electronics Engineers, Piscataway, N.J., 1998), Vol. 2, pp. 906–909.
  11. V. Backman, R. Gurjar, K. Badizadegan, I. Itzkan, R. R. Desari, L. T. Perelman, and M. S. Feld, "Polarized light scattering spectroscopy for quantitative measurement of epithelial cellular structures *in situ*," *IEEE J. Sel. Top. Quantum Electron.* **5**, 1019–1026 (1999).
  12. V. P. Wallace, D. C. Crawford, P. S. Mortimer, R. J. Ott, and J. C. Bamber, "Spectrophotometric assessment of pigmented skin lesions: methods and feature selection for evaluation of diagnostic performance," *Phys. Med. Biol.* **45**, 735–751 (2000).
  13. R. A. Vaughan, *Pattern Recognition and Image Processing in Physics: Proceedings of the Thirty-Seventh Scottish Universities Summer School in Physics* (Scottish Universities Summer School in Physics and Adam Hilger, Bristol, UK, 1991).
  14. D. R. Lovell, C. R. Dance, M. Niranjana, R. W. Prager, and K. J. Dalton, "Using upper bounds on attainable discrimination to select discrete valued features," in *Neural Networks for Signal Processing VI: Proceedings of the 1996 IEEE Signal Processing Society Workshop*, S. Usui, ed. (Institute of Electrical and Electronics Engineers, New York, 1996), pp. 233–242.
  15. J. M. DeLeo, "Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty," in *Proceedings of the Second International Symposium on Uncertainty Modeling and Analysis*, B. M. Ayyub, ed. (IEEE Computer Society Press, Los Alamitos, Calif., 1993), pp. 318–325.
  16. A. Baraldi and F. Parmiggiani, "An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters," *IEEE Trans. Geosci. Remote Sens.* **29**, 293–304 (1991).
  17. M. Mehrübeoğlu, "Diffuse reflectance imaging modalities for characterization of highly scattering turbid media," Ph.D. dissertation (Texas A&M University, College Station, Tex., 2000), pp. 38–142.
  18. S. Theodoridis and K. Koutroumbas, *Pattern Recognition* (Academic, San Diego, Calif., 1999).
  19. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. (Academic, Boston, 1990).
  20. B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Stat. Sci.* **1**, 54–75 (1986).
  21. P. M. Djurić, "Using the bootstrap to select models," in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (IEEE Computer Society Press, Los Alamitos, Calif., 1997), pp. 3729–3732.
  22. H. F. Bunn and B. G. Forget, *Hemoglobin: Molecular, Genetic and Clinical Aspects* (Saunders, Philadelphia, Pa., 1986).
  23. G. Marquez, L.-H. Wang, S.-P. Lin, J. A. Schwartz, and S. L. Thomsen, "Anisotropy in the absorption and scattering spectra of chicken breast tissue," *Appl. Opt.* **37**, 798–805 (1998).