Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results

Irving J. Bigio*
Los Alamos National Laboratory
MS-E535, Los Alamos, New Mexico 87545

Stephen G. Bown
Gavin Briggs
National Medical Laser Centre
University College London
London W1P 7LD, United Kingdom

Christine Kelley
Los Alamos National Laboratory
MS-E535, Los Alamos, New Mexico 87545

Sunil Lakhani
Department of Histopathology
The Royal Free and University College Medical School
London W1P 7LD, United Kingdom

David Pickard
National Medical Laser Centre
University College London
London W1P 7LD, United Kingdom

Paul M. Ripley
Ian G. Rose
Los Alamos National Laboratory
MS-E535, Los Alamos, New Mexico 87545

Christobel Saunders
Department of Surgery
The Royal Free and University College Medical School
London W1P 7LD, United Kingdom

Abstract. We report on the first stages of a clinical study designed to test elastic-scattering spectroscopy, mediated by fiberoptic probes, for three specific clinical applications in breast-tissue diagnosis: (1) a transdermal-needle (interstitial) measurement for instant diagnosis with minimal invasiveness similar to fine-needle aspiration but with sensitivity to a larger tissue volume, (2) a hand-held diagnostic probe for use in assessing tumor/resection margins during open surgery, and (3) use of the same probe for real-time assessment of the ''sentinel'' node during surgery to determine the presence or absence of tumor (metastatic). Preliminary results from *in vivo* measurements on 31 women are encouraging. Optical spectra were measured on 72 histology sites in breast tissue, and 54 histology sites in sentinel nodes. Two different artificial intelligence methods of spectral classification were studied. Artificial neural networks yielded sensitivities of 69% and 58%, and specificities of 85% and 93%, for breast tissue and sentinel nodes, respectively. Hierarchical cluster analysis yielded sensitivities of 67% and 91%, and specificities of 79% and 77%, for breast tissue and sentinel nodes, respectively. These values are expected to improve as the data sets continue to grow and more sophisticated data preprocessing is employed. The study will enroll up to 400 patients over the next two years. © *2000 Society of Photo-Optical Instrumentation Engineers.* [S1083-3668(00)00302-6]

1 Introduction and Background

Various types of optical spectroscopy have been investigated as methods of ''optical biopsy,'' with a majority of research efforts focusing on UV-induced fluorescence or Raman spectroscopies. (For a review of the topic, see, for example, Ref. 1 or references contained therein.) Recently, elastic-scattering spectroscopy (ESS) (sometimes called ''diffuse reflectance spectroscopy'') has been studied as a method for minimally invasive optical diagnosis of tissue pathologies, with emphasis on distinguishing dysplasia and cancer from normal tissue or other benign conditions[2–4] or for distinguishing different normal tissue types.[5] ESS, when performed using an appropriate optical geometry,[6,7] is sensitive to the sizes and structures of the subcellular components that change upon transformation to premalignant or malignant conditions (e.g., the nucleus, mitochondria, etc).[8] In work reported to date, clinical application of the ESS technique has been fundamentally noninvasive, with access to the tissue mediated by endoscopes[2–4] or by means of direct topical access as would be appropriate for diagnosis on the cervix.[9,10]

Within solid organs, such as the breast, the least invasive approach requires access through a needle. In screening, most early breast cancers are detected when an abnormality is seen on mammography, although some patients may detect a lump themselves (palpation). In either case, before any treatment is initiated, the diagnosis must be confirmed by fine-needle aspiration cytology (FNA) or biopsy (which is often done as an open procedure). Approximately 50 000 diagnostic lumpectomies are performed annually in the U.S. Of those, only about 12 000 turn out to be malignant when histology is performed on the excised tissue by a pathologist.[11] If it had been known in advance that the remaining 38 000 lesions were benign, the potentially disfiguring surgery could have been avoided, along with the trauma and cost, as many benign lesions resolve spontaneously in time, without intervention. A core biopsy, through a large-gauge needle, is a frequent alternative to surgical resection or lumpectomy for biopsy. Although less invasive than core biopsy, FNA is only infrequently used in the U.S. because false-negative rates with FNA are often in the range of 12%–15%,[12,13] and its reliability is under critical review. This is due to the heterogeneity of breast lesions and the relatively small number of cells accessed by FNA. Despite having the same level of invasiveness as FNA, the ESS

*Address all correspondence to Irving J. Bigio. Tel: 505-667-7748; Fax: 505-665-4637; E-mail: ijb@lanl.gov

method poses the potential advantage of immediate diagnosis, and there is the possibility of improved sensitivity when compared with FNA, due to the larger tissue volume that is sensed by the ESS method. If the optical biopsy technique based on ESS proves sufficiently reliable, it gives an immediate result, which minimizes the time a patient must wait for a diagnosis and might make it realistic to treat a small cancer at the same clinic visit. This will reduce patient anxiety and reduce costs for the large number of patients who have suspicious areas seen on mammography. Image-guided optical biopsy either with magnetic resonance, ultrasound, or conventional mammography could prove particularly valuable for patients with more than one suspicious area.

However, two other applications of ESS diagnosis to breast cancer may provide even greater patient benefits than the transdermal needle measurement. For open surgery procedures, the first of these is the object of our effort to develop a probe for the surgeon to use during breast-conserving surgery (wide local excision or partial mastectomy) for determining the status of the resection margins (the tissue surface, or ''tumor bed,'' that is left after the suspect tissue has been excised) in real time. In current surgical practice, especially when the tumor limits are not clearly visible, the surgeon (and anesthetized patient) are required to wait for pathology results on a frozen section of the excised tissue to determine if tumor-free margins of excision have been established. Immediate frozen-section pathology is not always available, and, in real practice, with delayed pathology positive margins are found in as many as 20%–55% of all breast-conserving surgeries,[14] requiring a second surgical procedure.

The other potentially important surgical application being investigated in this study is real-time assessment of a ''sentinel'' lymph node in the axilla. For years, there has been controversy about how the axilla should be treated in breast cancer patients. Recent research has shown that if the main node in the axilla draining a tumor area, called the sentinel node, is removed and does *not* show cancer, then the chances of any other nodes in the axilla showing cancer are approaching 1%.[15] Thus, if the sentinel node does not show cancer, the rest of the axillary nodes can be left in place, but if it does show cancer, then a full surgical axillary-node clearance must be done. The sentinel node can be identified by injecting a radioactive marker and then scanning the axilla about 24 h later (during surgical preparation) or by using a dye such as methylene blue, which can be easily detected in tissue during surgery. The ESS method may be able to provide immediate assessment of the sentinel node during surgery.

Our diagnostic study is coupled, utilizing the same patient cohort, with a study of minimally invasive optical treatment. In the case of focal lesions, treatment can be accomplished using interstitial laser photocoagulation with laser energy from a diode laser, again mediated by fiber optics through a transdermal needle, to nonsurgically treat both adenocarcinomas and fibroadenomas. Results for that part of the study will be reported in separate publication(s).

2 Materials and Methods

The clinical instrumentation based upon elastic-scattering spectroscopy is essentially similar to that described in publications on earlier clinical studies.[2,3] The system (see Figure 1)
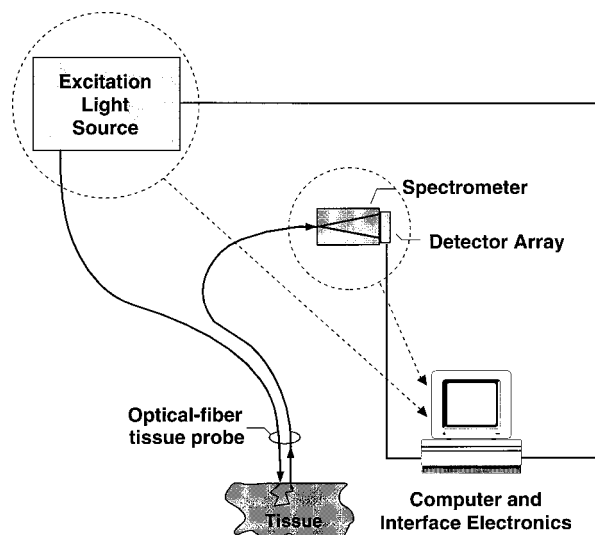

Fig. 1 Schematic diagram of the principal components of the ESS diagnostic system. All components (except for the fiber probe) are located inside a small, portable chassis.

consists of a pulsed xenon-arc lamp (EG&G) for the light source, a PC-compatible spectrometer, (a modified version of a spectrometer manufactured by Ocean Optics, Dunedin, FL), which employs a linear charge-coupled device (CCD) array for detection, an optical-fiber based probe, and a laptop computer for system control and data display. The wavelength range of the system is from 300 to 900 nm, but the range used for these studies is 330–750 nm, which covers the near-UV-visible part of the electromagnetic spectrum. As depicted in Figure 2, the probe is designed to be used in (gentle) optical contact with the tissue and incorporates two optical fibers, which are selected for their broadband light transmission over the spectral range used in the study. The output of the arc lamp is coupled to the illumination fiber, with a core diameter of 400 μm, which transmits the light to the tissue target site. A second, adjacent and parallel fiber within the probe, with a core diameter of 200 μm, collects a small fraction of the scattered light from the tissue. The collected light is then guided to the spectrometer where an optical spectrum is generated for further processing. For all of the measurements in this study, the probe design fixes the center-to-center separation of the collection and delivery fibers at 350 μm, at the tip of the probe. For this probe geometry, the volume of tissue
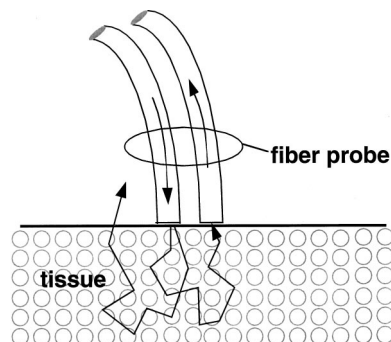

Fig. 2 Depiction of the optical geometry for the fiberoptic probe.

visited by the collected photons occupies a zone approximately 500 μm long, 300 μm wide and 300 μm deep. This has been determined from computational simulations using our Monte Carlo code, which incorporates Mie theory for the details of the scattering events.[6]

In clinical use the tip of the fiber probe is momentarily placed in contact with the suspect tissue, and the measurement is activated, at the keyboard or with a foot pedal. The system automatically takes a background (ambient+ ''dark'') measurement without firing the lamp, followed immediately (within 100 ms) by an ESS measurement with the pulsed lamp being triggered, and then subtracts the background spectrum from the ESS spectrum. The entire measurement process, i.e., activating the spectrometer, triggering the arc lamp, reading the detector array with an analog/digital (A/D) converter, etc., is computer controlled by a laptop PC, and can be activated with a foot pedal. This allows both accurate and reproducible measurements within the clinical setting. Furthermore, it also provides the clinician with the advantages of rapid data acquisition and a graphical display for inspection. Typical data acquisition and display time is less than 1 s for each site measurement. More specific details about the basic concepts of ESS, and discussion about the optical system and the design philosophy of the optical fiber probe, can be found in previous publications.[6,16,17]

Prior to any clinical measurements the ESS system and probe are calibrated with a reflectance standard. The reflectance standard (Spectralon™, Labspere, Inc., North Sutton, NH) has a spectrally flat, diffuse reflectance >98%, over the entire wavelength range of the system. The purpose of referencing the system to a known standard is to allow the normalization of spectral data against the overall system response. This technique effectively minimizes any variations in the system response due to variations in the spectral transmission among different probes, thermal effects, coupling efficiency of the fiber probe, drifts in detector/spectrometer response, etc. The reference standard and probes are sterilized by autoclave, and the housing of the reference standard protects it from any airborne contaminants during surgery.

Two different optical-probe configurations have been fabricated for use in the clinical program. Both probes incorporate the same ''standardized'' optical design, which specifies the diameters of the illumination and collection fibers, as well as the distance between the centers of the fibers, as specified above. It should be noted that details of the probe's optical design, including the fiber sizes and separation, have a significant effect upon the characteristic spectra obtained for a particular tissue. Therefore, these parameters are standardized for each clinical study in order to optimize the sensitivity of the system to tissue variations and also to prevent spectral variations due to instrumental artifacts. The main difference between the two probes designed for this study is to be found in the mechanical housing of the probe, allowing each to be optimized for the different specific clinical procedures. For interstitial (transdermal) measurements the probes have a small-diameter ($\leq$1 mm) stainless-steel outer sheath, which houses the optical fibers. The outer diameter of these probes was carefully chosen to be compatible with the inner bore of current core-biopsy needles in use at the clinics, so that a probe could safely and easily be passed down the needle and presented to the tumor tissue under investigation. The second probe design incorporates a larger stainless-steel sheath, $\sim$5 mm in diameter, to house the optical fibers, with no change to the fiber dimensions or separation. This creates an ergonomically convenient, pen-like design, which is utilized, hand held, for optical measurements during open surgery, when both resected tumor margins and lymph nodes are addressed.

3 *In Vivo* Measurement Procedures

All clinical measurements reported here have been performed at the Middlesex Hospital, University College London Hospitals, UK. Histological examination of all tissue samples has been performed by the same pathologist (S.L.). Before any *in vivo* measurements were made, *ex vivo* tissue samples were obtained from resected breast specimens from 15 patients and were measured with the ESS system. These initial findings were used to optimize the operational settings of the system, and also helped to establish protocols for marking the optical biopsy sites for histological examination and correlation. All *in vivo* measurements are performed under a protocol approved by the Ethical Review Board of the Middlesex Hospital. Prior to surgery every patient within the study is informed about the research program, and informed consent for data to be taken is provided by the patient. As mentioned in the previous section, three separate areas are currently under investigation, and each warrants discussion due to the techniques employed in order to obtain the ESS spectra.

For interstitial measurements with the ESS system, a core-biopsy (Tru-cut™) needle is guided into the palpable lesion, sometimes with the aid of ultrasound or magnetic-resonance imaging. The appropriate ESS optical probe is then inserted through the needle and gently placed in direct contact with the tissue at the needle's distal end. After the optical measurement has been made, the fiber probe is withdrawn, and the inner component of the biopsy tool is reinserted and a core biopsy is then taken from the same site, without moving the needle. This produces a small worm-shaped biopsy sample, one end of which has been measured optically. The other end of the core sample (which is opposite to the end of the sample that was interrogated by the ESS probe) is marked with India ink so that proper orientation is preserved and identification of the end examined with the ESS probe can be maintained and correlated with histology.

Following the core-biopsy procedures, some of the patients immediately undergo surgery in order to have the tumor resected. In these and other surgical cases (not preceded by core biopsy), and depending upon the exact nature of the surgical procedure, measurements are taken from one or more locations of the tumor bed, during and immediately after resection. Each optical measurement is followed by a small surgical biopsy of the same site, and the biopsy samples are coded to correlate with the ESS probe measurements.

With a subset of the surgical patients, assessment of the sentinel node is performed in order to determine the presence of metastatic disease. (In a number of patients more than one node was assessed.) During the surgical procedure the sentinel node is located with the aid of radioactive tracer and/or blue dye (both injected previously into the tissue space near the tumor). The node is resected and cut in half, and ESS measurements are made on one or more sites on the cut surface of the node. Once again, the ESS measurement sites are

Table 1 Histopathology of optically measured tissue sites.

	Number of patients	Number of biopsies	Number of cancer sites	Number of normal sites
Breast tissue	24	72	13	59
Sentinel nodes	21	54	12	42

indicated to assist the pathologist in providing the corresponding histopathology information about the specific locations measured.

4 Methods of Classification

The data reported here were taken on the first 31 patients of the study (which will eventually enroll up to 400 patients over the next two years), with breast-tissue measurements (either as an interstitial measurement or during surgery) being made on 24 of them and lymph-node measurements (during surgery) on 21 of them. Thus, both breast-tissue measurements and sentinel-node measurements were made on many, but not all, of the patients. The numbers of patients and measurements are summarized in Table 1. The last two columns list the histopathology designations for the sites measured optically, i.e., the number of true positives and true negatives. In many cases two or three optical measurements were made for a given biopsy site to test repeatability of the spectra. However, it is not appropriate to treat multiple spectra from the same biopsy site as independent measurements, as this could result in incorrectly high *or* low sensitivity and/or specificity. Consequently, in such cases the multiple spectra (which were very similar) were averaged together and treated as a single spectrum. Thus, one spectrum is correlated with each biopsy site. An effort was always made to mark precisely each site of optical measurement, so that corresponding histopathology could be properly correlated.

Figures 3 and 4 show examples of tissue spectra, malignant and nonmalignant, for breast tissue and sentinel nodes, respectively. The spectra in Figure 3 are examples that are representative of some but not all of the data. It would be misleading to assume otherwise, given the heterogeneity of breast tissue (glandular, adipose, connective, fibrous, ductal structures, etc.) and the resulting large spectral variations. Consequently, it would be risky to base diagnosis directly on a model for any specific/common signature causes. For example, Figure 3 does not show a difference in hemoglobin saturation between normal breast tissue and ductal carcinoma *in situ* (DCIS), although there is evidence of more perfusion in the DCIS, which is expected. (On the other hand, Figure 4 does indicate some desaturation for the metastatic sentinel node, and the spectra are more representative of other node spectra, since nodes are less heterogeneous.) Reduced saturation *can* indicate cancer indirectly because necrotic regions are more likely to be cancer than benign, but DCIS is an early, pre-invasive stage, and does not generally invoke necrosis. Thus, reduced saturation can be indicative of cancer (although other conditions can also, more rarely, produce necrosis), but good saturation is *not* a reliable indicator of benign conditions. Artificial-intelligence pattern-recognition (AI-PR) methods are well suited for spectral classification when model-based analysis is made difficult by the large number of necessary input parameters resulting from heterogeneity. Nonetheless, there certainly are some spectral indicators that are understood, as discussed below, and we have ensured that the input parameters (derived from the spectra) for our AI analyses contain that information.

Two different AI-PR methods of spectral classification have been employed to assess the degree of correlation between pathology and spectral pattern differences: artificial neural networks and hierarchical cluster analysis. Artificial neural networks (ANNs) were selected for study because of the expectation by our group and other researchers[4] that ANNs would prove to be a generally useful method of tissue spectral classification. ANNs (and other AI-PR methods) are
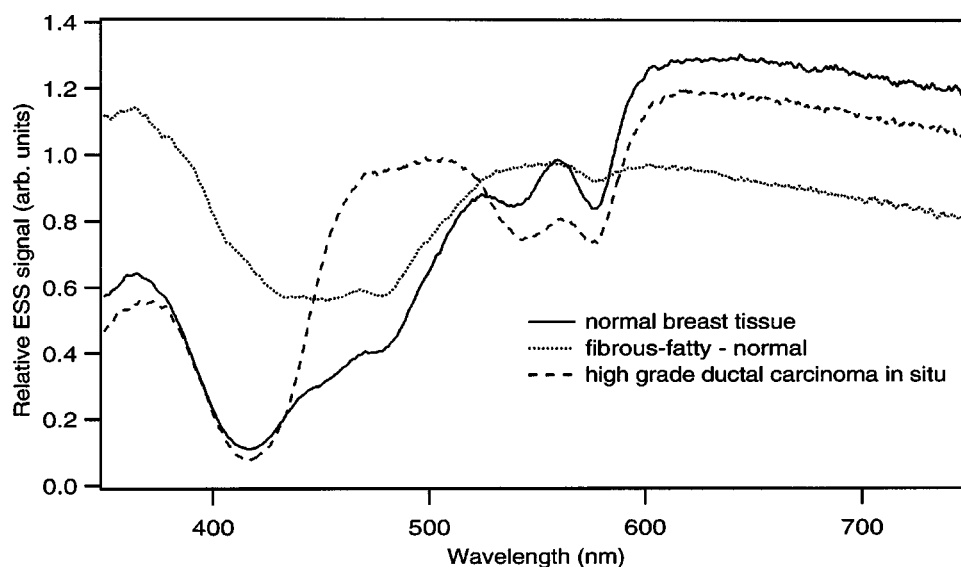

Fig. 3 Examples of ESS spectra for normal and malignant breast-tissue conditions.
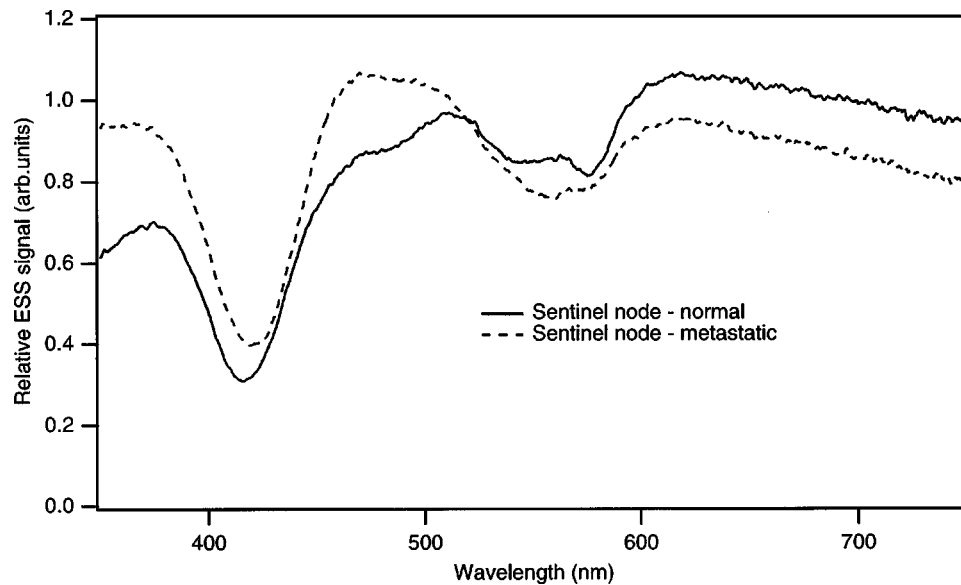
Fig. 4 Examples of ESS spectra for a reactive node, without tumor, and for node containing metastatic tumor.

known to be well suited for classification in systems where model-based classification is difficult.[18] As stated above, such is the case with ESS spectra of breast tissue because of its remarkable heterogeneity of tissue types. In light of this expectation, there was interest in our group to evaluate the relative merits of ANNs in ESS spectral analysis. Hierarchical cluster analysis (HCA) was also selected for study because it can serve as an alternative to the many approaches to classification that provide unbounded class regions (including ANNs, as well as linear discriminant analysis, regression analysis, etc.). The problem of unbounded class regions, and the benefits of using cluster analysis to avoid this problem, are discussed in detail by Osbourn et al.[19] In general, the problem with unbounded class regions is that a new sample will be classified in one of the available classes of a previously trained system, even if the input parameters of this new sample differ markedly from the other samples in the class to which it is assigned. This is not a problem when new samples are expected to be reasonably similar to previous samples. With optical biopsies, however, a number of variables can significantly alter the measured spectra. In addition to tissue heterogeneity, and even with relatively homogeneous tissue types, there can be experimental factors such as the presence of fluid (especially blood) trapped under the probe tip, or simply as a consequence of improper probe contact with tissue. It would thus be beneficial for a classification technique to be able to recognize samples that may have been adversely affected by such variables (i.e., ''outliers'' samples that differ markedly from previous samples). Having bounded class regions, such as those produced through HCA, provides this ability to reject outliers. A more detailed description and comparison of these artificial intelligence methods for classification of tissue spectra for several different organ areas is the subject of a future publication currently in preparation.

Although both ANN and HCA methods are intrinsically statistical-based classification methods, some preprocessing or weighting of input parameters can be implemented to ''assist'' classification when some things about the underlying tissue optical properties are known. (Such preprocessing can, in effect, combine benefits of both statistical-based analysis and model-based analysis.) In its simplest form this means that the input parameters for training the AI methods, which are derived from the raw spectra, should be structured so as to include the spectral information known to have diagnostic relevance. For example, the presence of adipose (fatty) tissue in the measured volume is much more frequently associated with normal or benign breast conditions (for a given biopsy site) than with adenocarcinoma. Especially in postmenopausal women, and in mixed tissues, a developing adenocarcinoma will often ''push'' the fat out of the tissue volume where growth is occurring. Since adipose tissue has the characteristic absorption feature of beta carotene, the spectral bands containing that feature can be accounted for. (The absorption feature of beta carotene (450–500 nm) can be seen in both the fibro-fatty and normal spectra in Figure 3.) Another example is the mean redox state of the hemoglobin in the tissue. Hemoglobin desaturation is generally indicative of necrotic/malignant lesions, although the obverse is not true, and hypervascularization more frequently accompanies malignant tissue than normal or benign conditions. In short, it is valuable to include enough detail of the beta carotene and hemoglobin bands among the AI-PR input parameters. Finally, broad (large-spectral-range) slope changes are expected for malignant conditions in glandular tissue due to enlarged and denser nuclei, and other scattering differences are expected among tissue types.

For the data analyses presented here, some model-based preprocessing was implemented to ensure sensitivity to known factors, but no weighting of any specific parameters was implemented at this stage. This and other methods of data preprocessing can improve the classification accuracy, and will be treated more extensively in a future publication.

5 Data Processing and Analysis Results

For all analyses reported here the spectra from breast tissues and sentinel nodes were analyzed separately since they are

fundamentally different classes of tissue. All breast-tissue spectra were preprocessed by first normalizing each spectrum to the same total integral over the spectral range of 330–750 nm. Sentinel-node spectra were normalized for the range 330–590 nm. (See discussion below.) Thus, only spectral shapes were compared. These normalized spectra comprise information on both the spectral dependence of scattering parameters and absorption bands, but not total scattering efficiencies. Such data treatment eliminates concerns about differences in optical coupling and/or transmission among the several fiber probes used, although any potentially useful information in the total scattering efficiency is, admittedly, also lost. In the case of sentinel nodes, many exhibited strong absorption due to the presence of methylene blue dye (Patent Blue™), which had been used to assist the surgeon in locating the node during surgery. Therefore, sentinel-node spectra were also treated to subtract the feature of that absorption band, and were analyzed for the range of 330–590 nm.

In presenting the statistical results, sensitivity and specificity are defined in the standard way

$$\text{Sensitivity} \equiv \text{SE} = \text{TP}/(\text{TP}+\text{FN}),$$

$$\text{Specificity} \equiv \text{SP} = \text{TN}/(\text{TN}+\text{FP}),$$

where TP, FP, TN, and FN represent the numbers of true positives, false positives, true negatives, and false negatives, respectively, as determined by the corresponding histopathology.

5.1 ANN Classification

For ANN classification the normalized spectra were divided into 21 wavelength bands of 20 nm width (330–750 nm), with some of the intervals being smaller for sentinel-node spectra because of the reduced total spectral range (330–590 nm), and an average spectral intensity for each interval was calculated. For classification of breast-tissue spectra we also included slopes of the intervals as input parameters, and although more sophisticated ANNs can handle a large number of input nodes, ours was limited to about 20. Therefore, for breast-tissue spectra, principal component analysis was performed on the interval intensities to establish a smaller number of input parameters.[20] The resulting smaller number (4–6) of principal components based on spectral intensities were combined with a number of slopes for various spectral intervals, for a total of 12–16 input parameters.

With ANN classification a variety of network designs and transfer functions between neuron layers are possible.[21] The analyses presented here used a relatively simple flat, three-layer network, with input and output layers and one hidden layer of neural nodes, or ''neurodes.'' As mentioned above, the input layer had 12–16 input parameters (neurodes) and the output layer always had two parameters, for malignant and nonmalignant conditions. The choice for the number of hidden-layer neurodes is somewhat arbitrary, and a typical starting point is to set it at half the total of input and output parameters, although sometimes a somewhat larger number is chosen if training is too slow. The transfer function (between layers) chosen for these analyses is a unipolar sigmoid (from 0 to 1), which is often used for parameters that can have smoothly varying values.[21] Network training was accomplished by ''error back propagation,'' wherein the correct diagnoses (from histopathology) are presented at the output layer, and the network is run (transfer functions are varied) to minimize the error between the classified output and the known parameter. Eighty percent of the data samples where selected randomly for training and the remaining 20% were used for testing. This was repeated five times with disjoint testing sets, each with a different 20% of the samples, so that all of the data samples were used for testing, but none were used for both training and testing in the same run. Statistics were derived from the average of all five runs. Given the modest size of the data set, better statistics might have been achieved by the ''leave-one-out'' method where only one data sample is used for testing (and $n-1$ are used for training) and repeating the process n times, but this would have taken prohibitively long computer time.

5.2 HCA Classification

For HCA classification of breast-tissue spectra, the spectra were again divided into 20 nm intervals, and a preliminary cluster-template analysis was performed, as described below, in order to pre-select a subset of interval intensities to be used for input parameters. Similarly, preliminary template analysis was also performed with interval slopes in order to pre-select a reduced number of slopes as input parameters. A combination of these pre-selected intensities and slopes was then used as the set of input parameters for the full HCA analysis. For sentinel-node spectra only the interval intensities were used as input parameters, without any pre-selection.

With the hierarchical cluster analysis method the ''training'' consists of trying all possible different combinations of the input parameters, using multidimensional pattern recognition schemes to find clusters with no *a priori* input of actual classifications, and then comparing the resulting clusters with the known classifications of their members. Different criteria are possible for determining the multidimensional Euclidian ''distance'' between a sample point and a given cluster: nearest neighbor, farthest neighbor, or average distance.[20] For the analyses presented we chose the average-distance method. The combination of input parameters that yields the clusters of samples that best agree with each other, by having the same histology classifications, is the one that is chosen. For a given number of input parameters HCA generally requires more data than ANNs for statistical stability of the clusters (outputs). However, training can be faster than for ANNs. Once the input parameters for the best clustering are established, a new data sample can be tested for those parameters, to see which (if any) cluster it falls into. With our modest-sized data sets, and with the multiple input parameters required, the clusters found are sensitive to the specific choice of training set. Therefore, instead of an 80/20 split of the data, we applied the leave-one-out method, as described above, and the statistics are determined from the sum of all of the leave-one-out tests.

The results of data analysis are summarized in Tables 2 and 3. It can be seen that with the HCA method a few samples appear as unclassifiable, rather than as false positives or false negatives. It is especially interesting that, upon examination of the individual unclassified samples, about half of these outliers corresponded to samples that were classified incorrectly by ANN. The resulting sensitivities and specificities were

Table 2 Classifications for breast-tissue spectra.

	ANN	HCA
Number of specimens showing cancer histologically	13	13
Number of cancers optically diagnosed as cancer (true positives)	9	8
Number of cancers optically diagnosed as normal (false negatives)	4	4
Number of cancers with an indeterminate optical diagnosis (HCA only)		1
Sensitivity (%)	69%	67%
Number of specimens found normal histologically	59	59
Number of normals optically diagnosed as normal (true negatives)	50	44
Number of normals optically diagnosed as cancer (false positives)	9	12
Number of normals with an indeterminate optical diagnosis (HCA only)		3
Specificity (%)	85%	79%
% Classified (HCA only)		91.5%

Table 3 Classifications for sentinel-node spectra.

	ANN	HCA
Number of specimens showing cancer histologically	12	12
Number of cancers optically diagnosed as cancer (true positives)	7	10
Number of cancers optically diagnosed as normal (false negatives)	5	1
Number of cancers with an indeterminate optical diagnosis (HCA only)		1
Sensitivity (%)	58%	91%
Number of specimens found normal histologically	42	42
Number of normals optically diagnosed as normal (true negatives)	39	26
Number of normals optically diagnosed as cancer (false positives)	3	8
Number of normals with an indeterminate optical diagnosis (HCA only)		8
Specificity (%)	93%	76.5%
% Classified (HCA only)		83.3%

comparable for breast tissue, but for sentinel nodes HCA yielded significantly better sensitivity (91%) than ANN (58%), albeit at the cost of somewhat reduced specificity. It is our expectation that with adequately large data sets HCA with properly chosen input parameters will generally yield similar or better sensitivity and specificity, as compared with ANNs, but with HCA having the advantage of identifying outliers as difficult to classify, rather than blindly forcing a classification. We believe this is a more realistic approach for clinical application of any new diagnostic method, and is comparable to other clinical measurement situations: for example, an invalid EEG measurement resulting when an electrode has poor contact. As expected with data sets for which there were significantly more negatives than positives but comparable accuracy, the negative predictive values (NPV) (87%–96%) were significantly higher than the positive predictive values (PPV) (40%–75%), where $NPV \equiv TN/(TN+FN)$, and $PPV \equiv TP/(TP+FP)$.

6 Discussion and Problems Encountered

It is not the purpose of this paper to present a detailed discussion of the structures and functions of different artificial-intelligence, pattern-recognition schemes for classification of tissue spectra, but rather to present the results of their implementation, with minimal refinement, for these preliminary data of our breast-cancer study. Although our ANN was limited to about 20 input parameters, with some ANN codes it is possible to have as many as 250 input parameters, which would have permitted us to present every spectral resolution element, for each spectrum, to a separate input node. This would have permitted capturing all spectral features. Also, a variety of other improvements and refinements are possible, including variations in the design of the ANN, and more sophisticated data preprocessing and choice of input parameters for HCA. A careful discussion of the relative merits of AI methods for tissue-spectra classification will require a full separate publication, which (as mentioned above) is in preparation. Also, with HCA robust clustering requires that the number of samples be much greater than the number of input parameters. For the sizes of our data sets this condition was marginally met, and we expect significant improvement in the statistics as data sets grow. Nonetheless, we believe these preliminary results are compelling enough to merit presentation to the scientific community.

In actual clinical implementation of the ESS instrumentation, it became evident that some attention must be given to avoiding excessive amounts of blood on the tip of the fiber probe, as this can obscure information about the amount of tissue perfusion, or even block much of the scattering spectral information (due to the strong absorption by hemoglobin). Thus, it was sometimes necessary to sponge, or rinse with saline, the surgical resection surface to be measured, and/or to wipe the tip of the probe. Also, for sentinel nodes the accuracy in co-registration between optical and histology sites left some uncertainty when only a small region of the node exhib-

ited metastasis. Improvements are planned for the co-registration between optical measurement and histology for sentinel nodes.

Another problem encountered on a few occasions was that the ambient lighting of the surgical suite (with the high-intensity directed-beam lighting typically found in theater) would occasionally cause too much scatter through tissue into the collection fiber of the probe, using up much of the detector's dynamic range, and making it difficult for the system to perform an accurate background subtraction. This was remedied by either having the surgeon momentarily shadow the strong light from the measurement site, or by a redesign of the probe fixture to incorporate a small shadow mask. (Of course, the strong lights could be temporarily aimed away from the surgery or reduced in intensity while the ESS measurements are being made, but we wished to minimize any inconvenience to the surgical team.)

Finally, in the development of any new diagnostic technique, the accuracy and robustness of the ''gold'' standard, histopathology, are always a concern. In our case all pathology reports were provided by a specialist breast pathologist (S.L.) of the University College London Hospitals. (For the purposes of these analyses, ductal carcinoma *in situ* (DCIS) was always classified as cancer, since the treatment consequences are the same regardless of whether the DCIS already shows signs of invasiveness or not.) Careful attention has been paid to adequacy of tissue samples, correlation with optical measurement sites, and consistency of pathology reporting terminology, although further improvements are planned. As the program enlarges, and incorporates other medical centers, agreement among pathologists will be addressed, and slides will be read by multiple pathologists. Methods will be assessed to resolve conflicting reports. Nonetheless, the reliability of histopathology being less than 100%, this type of correlation is fundamentally limited in accuracy, and determination of the ultimate capability of the new method requires long-term outcome studies.

7 Conclusions

We have described a research program designed to test the value of elastic-scattering spectroscopy as a real-time diagnostic tool and as a diagnostic aid to surgical/therapeutic procedures for breast cancer. We have presented, as preliminary data, the results from the first 31 patients of a larger program. The modest sizes of the data sets notwithstanding, the results of spectral classification by two different methods of ''artificial intelligence'' pattern recognition show promise for good agreement with pathology. This allows us to be hopeful that as the data sets grow, we will be able to successfully test the predictive capabilities of already trained spectral classification schemes.

Acknowledgments

References

1. I. J. Bigio and J. R. Mourant, ''Ultraviolet and visible spectroscopies for tissue diagnosis...,'' *Phys. Med. Biol.* **42**, 803–814 (1997).

2. J. R. Mourant, I. J. Bigio, J. Boyer, T. Johnson, J. Lacey, A. G. Bohorfoush, and M. Mellow, ''Elastic scattering spectroscopy as a diagnostic tool for differentiating pathologies in the GI tract...'' *J. Biomed. Opt.* **1**, 192–199 (1996).

3. J. R. Mourant, I. J. Bigio, J. Boyer, R. L. Conn, T. Johnson, and T. Shimada, ''Spectroscopic diagnosis of bladder cancer with elastic light scattering,'' *Lasers Surg. Med.* **17**, 350–357 (1995).

4. Z. Ge, K. T. Schomacker, and N. S. Nishioka, ''Identification of colonic dysplasia and neoplasia by diffuse reflectance spectroscopy and pattern recognition techniques,'' *Appl. Spectrosc.* **52**, 833–839 (1998).

5. M. Johns, C. Giller, and H. Liu, ''Computational and in vivo investigation of optical reflectance from human brain to assist neurosurgery,'' *J. Biomed. Opt.* **3**, 437–445 (1998).

6. J. R. Mourant, J. Boyer, A. Hielscher, and I. J. Bigio, ''Influence of the scattering phase function on light transport measurements in turbid media performed with small source-detector separations,'' *Opt. Lett.* **21**, 546–548 (1996).

7. L. T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Can Dam, J. M. Crawford, and M. S. Feld, ''Observation of periodic fine structure in reflectance from biological tissue: A new technique for measuring nuclear size distribution,'' *Phys. Rev. Lett.* **80**, 627–630 (1998).

8. J. R. Mourant, A. H. Heilscher, A. A. Eick, T. M. Johnson, and J. P. Freyer, ''Evidence of intrinsic differences in the light scattering properties of tumorigenic and nontumorigenic cells,'' *Cancer (N.Y.)* **84**, 366–374 (1998).

9. A. F. Zuluaga, U. Utzinger, A. Durkin, H. Fuchs, A. Gillenwater, R. Jacob, B. Kemp, J. Fan, and R. Richards-Kortum, ''Fluorescence excitation emission matrices of human tissue: A system for in vivo measurement and method of data analysis,'' *Appl. Spectrosc.* **53**, 302–311 (1999).

10. I. J. Bigio, T. M. Johnson, J. R. Mourant, B. Tromberg, T. Tadir, M. Fehr, H. Nilsson, and V. C. Darrow, ''Determination of the cervical transformation zone using elastic-scattering spectroscopy,'' *Proc. SPIE* **2679**, 85–91 (1996).

11. *Cancer Facts and Figures—1995*, The American Cancer Society (Washington, DC, 1995).

12. S. L. Willis and I. Ramzy, ''Analysis of false results in a series of 835 fine-needle aspirates of breast lesions,'' *Acta Cytol.* **39**, 858–864 (1995).

13. P. Purasiri, M. Abdalla, S. D. Heys, A. K. Ahsee, M. E. McKean, F. J. Gilbert, G. Needham, H. E. Deans, and O. Eremin, ''A novel diagnostic index for use in the breast clinic,'' *J. R. Coll. Surg. Edinb* **41**, 30–34 (1996).

14. J. Walls, F. Knox, A. D. Baildam, D. L. Asbury, R. E. Mansel, and N. J. Bundred, ''Can preoperative factors predict for residual malignancy after breast biopsy for invasive cancer?,'' *Ann. R. Coll. Surg. Engl* **77**, 248–251 (1995).

15. H. Snider, K. Dowlatshahi, M. Fan, W. M. Bridger, G. Rayudu, and D. Oleske, ''Sentinel node biopsy in the staging of breast cancer,'' *Am. J. Surg.* **176**, 305–310 (1998).

16. J. Boyer, J. R. Mourant, and I. J. Bigio, ''Theoretical and experimental demonstrations of elastic scattering spectroscopy as a diagnostic for tissue pathologies,'' in *Optical Imaging and Photon Migration*, R. R. Alfano, Ed. (Optical Society of America, Washington, DC, 1994).

17. J. R. Mourant, T. Fuselier, J. Boyer, T. Johnson, and I. J. Bigio, ''Predictions and measurements of scattering and absorption over broad wavelength ranges in tissue phantoms,'' *Appl. Opt.* **36**, 949–957 (1997).

18. D. A. Cirovoc, ''Feed-forward artificial neural networks: Applications to spectroscopy,'' *TrAC, Trends Anal. Chem.* **16**, 148–155 (1997).

19. G. Osbourn, J. Bartholomew, A. Bouchard, and R. Martinez, ''Automated pattern recognition based on the visual empirical region of influence (VERI) method: A user's guide,'' web publication: *http://www.sandia.gov/1100/1155Web/1155home.htm*.

20. R. Gnanadesikan, *Methods for Statistical Analysis of Multivariate Observations*, 2nd ed., Wiley, New York (1997).

21. C. G. Loony, *Pattern Recognition Using Neural Networks*, Oxford University Press (1997).