IMPROVING THE ACCURACY OF DIAGNOSTIC BREAST ULTRASOUND

Michael P. André, Michael Galperin, Linda K. Olson, Katherine Richman, Susan Payrovi, Peter Phan*

1. INTRODUCTION

The National Cancer Institute estimates that approximately 700,000 women undergo breast biopsies (surgical or needle) in the U.S. each year. Approximately 80% of tumors biopsied are benign, 20% are malignant. Surgical biopsies--the most common--cost between $2,500 and $5,000 while needle biopsies cost from $750 to $1,000. Patients experience both physical and emotional effects when undergoing biopsy procedures and internal scarring may be problematic since it complicates interpretation of future mammograms. Until fairly recently, ultrasound in the U.S. has been used only to distinguish cystic from solid breast masses and to guide needle biopsies. A number of positive studies in Europe, Asia and the U.S. indicate that high-quality ultrasound can provide radiologists with a high degree of confidence in differentiating many benign from malignant or suspicious lesions detected by mammography.[1] Results suggest that ultrasound could help reduce the number of biopsies of benign masses by 40% with a cost savings of as much as $1 billion per year in the U.S.

Work to improve the accuracy of diagnostic breast ultrasound has led to the development of a well-defined system for scoring the level of suspicion (LOS) based on parameters describing the ultrasound appearance of breast lesions. An extensive reporting lexicon for breast ultrasound is also being established by the American College of Radiology following the Breast Imaging Reporting And Data System (BIRADS) approach.[2] While acceptance and utilization of LOS and the new lexicon is increasing, it is difficult to teach the method and many radiologists feel uncomfortable with the number of benign and malignant masses that overlap in appearance. In addition, studies indicate that there is high variability between radiologists both in the analysis of the ultrasound appearance of a mass and in the final assessment. It is suggested that this variability can be reduced with a more structured reporting system and peer review.[3] The specific guidelines for differentiation of breast lesions are shown in Table 1 while the LOS score

* Department of Radiology 114, University of California, SD Veterans Healthcare System, San Diego, California 92093.

is assigned based on the number of benign and malignant criteria found as shown in Table 2.

A number of promising efforts to improve the specificity of breast lesion classification using ultrasound may be grouped in two categories: 1) analysis of features in the display (image processing),[4,5] and 2) analysis of the ultrasound signal properties (tissue characterization).[6-8] Much of this work confirms that it is difficult to precisely classify masses because there is overlap in the acoustic properties of many solid benign and malignant lesions. Computer-aided diagnosis with artificial neural networks (ANN), a form of regression analysis, attempts to aid the radiologist in locating suspicious regions that might otherwise be missed.[9] We chose to concentrate our efforts on the complementary problem of improving confidence in benign findings.

Table 1

Criteria Associated with Benign Lesions	Criteria Associated with Malignant Lesions
Spherical/ovoid/lobulated	Irregular shape
Linear margin	Poorly defined margin
Homogeneous texture	Central shadowing
Isoechoic/anechoic	Distorted architecture
Edge shadow	Calcifications
Parallel to the skin	Skin thickening
Distal enhancement	
Dilated duct/mobile	

Table 2

LOS	Diagnosis	Number of Criteria
5	Malignant	5 malignant criteria
4	Probably malignant	3-4 malignant criteria
3	Indeterminate criteria	1-2 malignant criteria
2	Probably benign	0 malignant criteria
1	Benign	0 malignant criteria & all benign criteria

Image texture was shown by a number of investigators to have ability to identify malignant versus benign breast lesions.[5] Image texture is a parameter with numerous definitions, many not suitable for ultrasound, commonly used for general segmentation and classification.

2. MATERIALS AND METHODS

The core technology of the software to be adapted for this project was originally developed for satellite image processing and analysis for ocean monitoring (Almen Laboratories, Inc., Escondido, CA). The software system contains a very large number of filtering, shading and image sharpening tools that may be cascaded into a series of operations by the user for any specific application and stored as a macro. Segmentation is accomplished through several regimes including multi-level pixel thresholding. Figures 1 and 2 illustrate the approach for US images of two different types of breast masses.

A large number of parameters of each image and each segmented mass are stored with a reference file containing the values of the measured features. The features listed in Table 3 were developed to correspond to the image criteria of Table 1. We utilize several texture features including 1) those measured from the pixel histogram of a defined region of interest, considered first order statistics, 2) second order parameters that involve spatial distribution and relationships such as Markovian and features of co-occurrence matrices, 3) probability distributions including measures of the angular second moment, sum and difference entropy, sum and difference variance, correlation, contrast, etc., 4) and run-length measures including fractal dimension. The two-dimensional Fourier transform and power spectrum may also be used to estimate spatial properties of the texture. At this early phase of development we include a large number of parameters in order to determine which have a promising degree of association with lesion characterization, particularly those lesions with a lower LOS score.

These parameters represent an N-dimensional vector **P** that may be used to calculate the "Relative Similarity," **R**, of one lesion to another. Appropriate weighting factors, ω, may be applied to these results to enhance the classification; they are established by regression analysis of cases with known findings. A new case with an "unknown" finding is compared directly to the database of stored images and a measure of **R** is computed for different benign and malignant lesions. Similarity is calculated for a particular lesion $\mathbf{P}_{it}$ (the index of this "template" object) compared to the other lesions, $\mathbf{P}_k$ (k=1,…L) where L is the number of objects.

$$R = \left(\sum_{k=1}^{L} \left(p_k^t - P_k^t \right)^s \cdot \omega_k \right)^{1/s}$$

where ω is the statistical weight resultant from multi-factorial regression analysis. Figure 3 illustrates this method for a complex cyst compared to a collection of different breast masses. The term relative similarity means that the detected lesion is compared to the database of previously analyzed patients and the cases most "similar" to this suspicious mass are automatically retrieved and displayed. Only the portion of the stored image that is most similar to the unknown is initially retrieved as a thumbnail, so performance is nearly instantaneous.

Instead of an attempt to find an absolute measure of likelihood, we measure a "relative similarity" within defined boundaries--the classification space. Such an approach is known to alleviate the problems of "under-training" and "over-training" with classification schemes such as ANN techniques. The advantage of this approach is that we have an established highly specific, rule-base scheme to describe a breast mass on the basis of its features, namely the LOS system. Unlike many ANN's, which are employed with explicit rules are difficult to define, we do not need to optimize the number of features empirically since all combinations will be evaluated directly.[10] Importantly, we are able to update the feature set to optimize classification after each new case and as the "expert" database grows.

As discussed above, the weight assigned to a given parameter during this comparison process may be manually set by the user or preferably set using a statistical method, especially useful when there is a structured set of rules for object characteristics. These data can be analyzed to determine how strongly different parameters of the parameter set values correlate with the presence or absence of the specific trait. The weight used for a given parameter in the comparison process may thus be derived from the values of the

parameter vectors associated with the detected objects in the image database. In using this method a system is represented as a totality of factors. The mathematical simulation tools are correlation, regression, and multi-factor analyses, where the coefficients of pair-wise and multiple correlation are computed and both a linear and non-linear regression may be obtained. The data for a specific model experiment are represented as a matrix whose columns stand for factors describing the system and the rows for the experiments (values of these factors).

The factor Y, for which the regression is obtained, is referred to as the system response. (Responses are integral indicators but theoretically, any factor can be a response. All the factors describing the system can be successively analyzed. In breast cancer Y could be a biopsy result, lesion class or any other clinical indicator that is impacted by analyzed factors). The regression and covariance help to "redistribute" the multiple determination coefficient among the factors; in other words the "impact" of every factor to response variations is determined. The specific impact indicator of the factor is the fraction to which a response depending on a totality of factors in the model changes due to this factor. This specific impact indicator may then be used as the appropriate weight to assign to that factor (i.e., parameter set associated with the objects). The impact of a specific factor is described by a specific impact indicator, which is computed by the following algorithm: $\gamma_j = \alpha * [b_j * c_{0j}]$, j=1,2,...,k where γ is the specific impact indicator of the j-th factor; k is the number of factors studied simultaneously; b_j is the j-th multiple regression coefficient; c_{0j} – covariance coefficient and α - is the fraction of multiple determination related to the impact of the factor.

Table 3. Default image parameters for lesion classification

Image Criteria	Sample of Associated Parameters
Spherical/ovoid vs. irregular shape	Formfactor Equivalent circular diameter/Form factor Perimeter/Area Perimeter/Equivalent circular diameter Aspect ratio
Linear margin vs. poorly defined margin	Edge gradient
Homogeneous texture vs. internal echoes Isoechoic/anechoic vs. echoic Calcifications	Homogeneity (multiple texture parameters) Relief Contrast Optical density Integrated density Scatterer density, scatterer size 2^{nd}, 3^{rd}, 4^{th} moments of inertia
Edge shadowing vs. Central shadowing Distal enhancement	Density measures of a Distal ROI defined by X- and Y-Ferret coordinates
Parallel to skin vs. irregular	X-Ferret/Y-Ferret Aspect ratio Relative angle

3. RESULTS OF FEASIBILITY STUDY

A study was conducted to examine feasibility of this approach and to estimate the degree of inter-observer variability in our institution. Results are encouraging and were used to estimate the minimum size of the patient population needed in a follow-up prospective clinical trial. Diagnostic breast ultrasound (US) image files for 112 women were retrieved chronologically (not randomly) from the image library of one of our

hospitals under institutional review board approval. These files were sampled from patients who underwent core biopsy, fine needle aspiration or surgical biopsy in the year 2000. This group ranged in age from 41-67 and presented 97 benign (including solid and cystic) and 15 malignant breast masses. In this preliminary study, some of the images were acquired from different ultrasound systems so instrumentation variables are included. An additional group of ten malignancies and ten simple cysts for whom follow up was negative were also acquired from examinations in 1999.

A preliminary segmentation algorithm was devised and applied to all images. In every image the mass was successfully identified. Two sub-specialty breast imaging radiologists independently scored the US images for LOS and were asked to decide whether to biopsy the mass or not. A third radiologist has not yet completed the review. Relative Similarity was computed from the parameters of Table 3, a very simple linear mapping to LOS score was devised following Table 2 and no weighting factors were applied. LOS of 1, 2 or 3 was considered a benign finding (or follow up), while 4 or 5 were taken as a recommendation for biopsy. Table 4 shows results for the radiologists that are similar to those reported elsewhere. Simple cysts are almost always correctly identified and there are very few circumstances in which a histologically malignant mass is scored benign on ultrasound (low false negatives, high sensitivity). Our computer-aided imaging system (CAIS) showed higher specificity and positive predictive value (PPV), lower sensitivity by the traditional measures, but a significantly higher area under the Receiver Operator Characteristic (ROC) curve, A_Z (Figure 5).[11] Simple linear correlation to the biopsy findings by the radiologists was 0.41 and 0.34, while the CAIS system showed a higher correlation, 0.69. These results are surprising given that we have not yet attempted optimization of the processing system.

Table 4. Results of feasibility study with 112 patients.

Observer	Sensitivity	Specificity	PPV	NPV	Corr. Coeff.
Radiologist A	87	76	36	97	0.41
Radiologist B	93	66	30	99	0.34
CAIS	80	98	80	97	0.69

The radiologists independently scored the US images for LOS using the entire image set available for each patient including the mammograms by completing a checklist of criteria derived from Table 1. The radiologists then assembled as a group and reach a consensus on the borders and extent of the lesions for each patient. Using the workstation they viewed the digitized sonograms in the experimental software and manually trace (segment) the boundaries of the lesions for at least two US views for each patient.

In every one of the 112 cases, automated segmentation was considered "successful" in that the mass in question was always correctly contoured and analyzed. Examples of the segmentation for several different types of masses are shown in Figures 1-3.

4. DISCUSSION

Although these results are incomplete and subject to several flaws, they suggest that the CAIS approach may be successful in its goal of aiding the reduction of biopsies on benign masses. We require a high specificity (high TN, low FP) while maintaining a high Negative Predictive Value (NPV). Furthermore, the estimated variances suggest that we

will detect a change in radiologists' performance at the 0.05 confidence level with a study population of 250 subjects.[27]

We do not claim novelty of the well-known algorithms we use to measure features such as homogeneity, shape, location, etc. (except possibly Relief and Homogeneity through the 4th moments of inertia). Rather, the novelty comes from the application, especially the technique to store, process, retrieve and compare images on the basis of information content. This is a timely approach since there is a recognized effort to promote the ACR LOS method. Also, ultrasound is the sub-specialty that has adopted most widely PACS workstations for interpretation. CAIS, if successful, is readily adapted to routine use in a diagnostic workstation and through HL-7 messaging could incorporate clinical findings. In fact, the method is readily adapted to a number of other diagnostic imaging applications as well. The system we describe is essentially an expert electronic teaching file operating in background and available to the radiologist as desired. In addition, the system incorporates the clinical experience of the radiologist and builds automatically his electronic "long-term memory."

5. REFERENCES

1. A.T. Stavros, D. Thickman, C.L. Rapp, M.A. Dennis, S.H. Parker, and G. A. Sisney, "Solid Breast Nodules: Use of Sonography to Distinguish between Benign and Malignant Lesions," Radiology 196, pp. 123-134, 1995.
2. American College of Radiology (ACR). Breast Imaging Reporting and Data System (BI-RADS™). Third Edition. Reston, VA: American College of Radiology; 1998.
3. J.A. Baker, P.J. Kornguth, M.S. Soo, et al., "Sonography of solid breast lesions: Observer variability of lesion description and assessment," Amer J Roentgenol 172, 1621-1625, 1999.
4. V. Goldberg, A. Manduca, D.L. Ewert, J.J. Gisvold, J.F. Greenleaf, "Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence," Med. Phys. 19, pp. 1475-1481, 1992.
5. B.S. Garra, B.H. Krasner, S.C. Horii, et al., "Improving the distinction between benign and malignant breast lesions: The value of sonographic texture analysis," Ultrasonic Imag 15, pp. 267-285, 1993.
6. J. Bamber, "Ultrasound propagation properties of the breast," In: Ultrasonic Examination of the Breast, J. Wiley and Sons, Chichester, 1983.
7. MP André, HS Janée, GP Otto, PJ Martin, BA Spivey, DA Palmer: "High-speed data acquisition in a diffraction tomography system employing large-scale toroidal arrays." Intl J Imaging Systems Technol 8(1), pp. 137-147, 1997.
8. Lefebvre F, Meunier M, Thibault F, et al., "Computerized ultrasound B-scan characterization of breast nodules," Ultrasound in Med & Biol, 26(9):1421-1428, 2000.
9. M. Kallergi, G.M. Carney and J. Gaviria, "Evaluating the performance of detection algorithms in digital mammography," Med. Phys 26, pp. 267-275, 1999.
10. M.A. Kupinski and M. L. Giger, "Feature selection with limited dataset," Med. Phys. 26(10), pp. 2176-2182, 1999.
11. C.E. Metz, "ROC methodology in radiographic imaging," Invest. Radiol., 21, pp. 720-733, 1986.

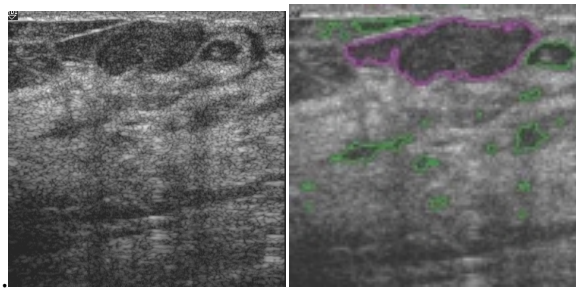

Figure 1. US image (left) of a superficial benign fibroadenoma Automated segmentation (right). This mass has features with an intermediate LOS score. Non-important components of the image are not selected by practitioner and are not used for further quantification and classification

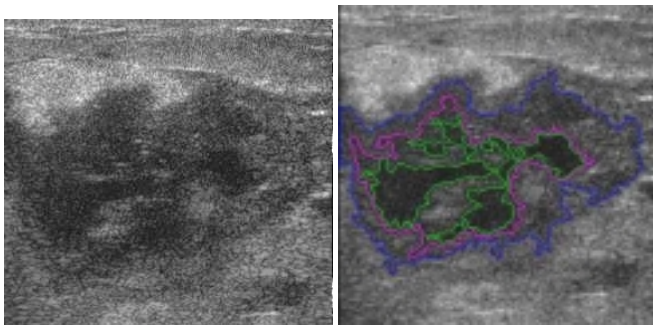

Figure 2. US image (left) of a biopsy proven carcinoma. Automated segmentation detected three-layer structure of the lesion (contours). The lesion parameters were used by for classification and calculation of relative similarity to other digital template with cases with known findings. Practitioner selected to use the intermediate layer of the lesion (middle contour) for further classification and comparison with other cases.

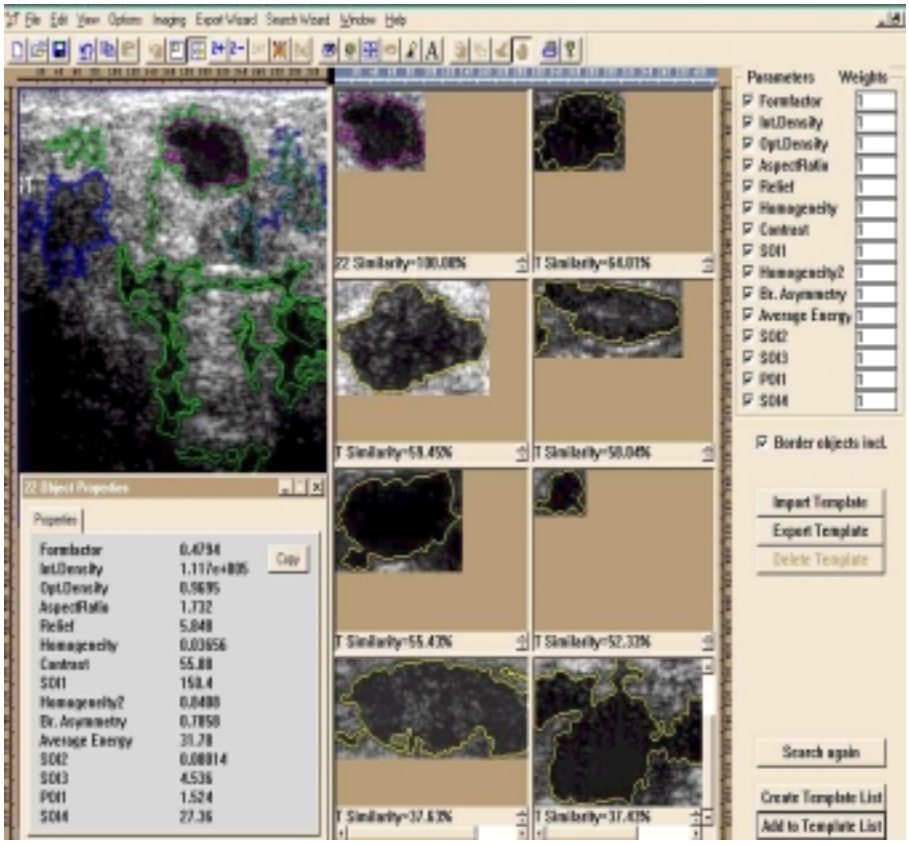

Figure 3. Complex cyst (benign) is compared to other images in the template database. The "unknown" mass in the upper left portion of the screen is a dark, relatively echo-free consistent with fluid-filled cyst but with irregular indistinct margins more consistent with a solid mass that might have a higher suspicion for cancer. Software automatically locates the mass contour. Measurements are made of the mass and its "relative similarity" is compared to a digital template database with known findings. Cases most “similar” to this suspicious mass are automatically retrieved and displayed in the thumbnail images on the right listed in rank order of this value (light contours in the left half of the screenshot). In this case, all of the “similar” masses were proven to be benign.

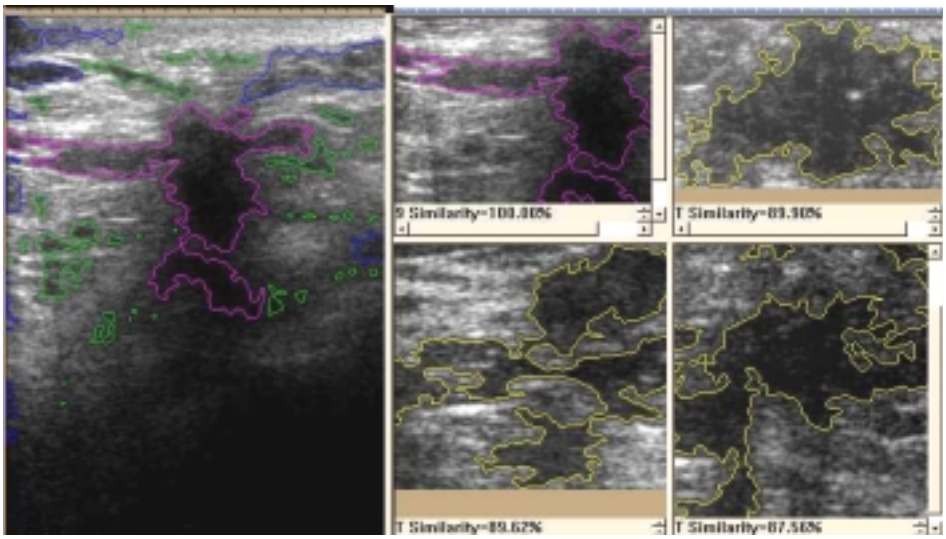

Figure 4. An unknown lesion on the left (purple contour) is automatically segmented and compared to a full digital database of lesion templates with know findings (130 cases). The system identified all "closest" cases (yellow contours on the right) to be malignant and calculated score of suspicion at the highest level of 5. It is important to bring attention to the fact that the system does not try to "match" (like in ANN approaches) or to diagnose the lesion. The system finds the closest relatively similar lesions in the digital database in accordance with the set of parameters and lesion contour selected by practitioner. That is why there are additional perks and uses of the developed methodology such as retrieval by the medical content of an image, as well as training and education for the practitioners, because the system will enable them to evaluate their judgment in comparison with biopsy proven cases.

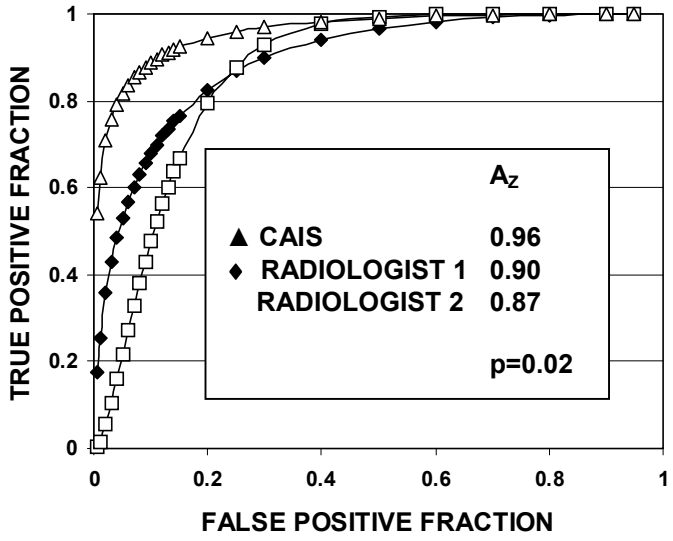

Figure 5. ROC for breast US LOS (112 patients).