

# UNCLASSIFIED

AD NUMBER
ADB277494
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to U.S. Gov't. agencies only; Proprietary Info.; Jul 2001. Other requests shall be referred to U.S. Army Medical Research and Materiel Command, 504 Scott St., Ft. Detrick, MD 21702-5012.
AUTHORITY
USAMRMC ltr, 28 Aug 2002

THIS PAGE IS UNCLASSIFIED

AD\_\_\_\_\_

Award Number: DAMD17-98-1-8194

TITLE: Computerized Analysis of MR and Ultrasound Images  
of Breast Lesions

PRINCIPAL INVESTIGATOR: Maryellen Giger, Ph.D.

CONTRACTING ORGANIZATION: The University of Chicago  
Chicago, Illinois 60637

REPORT DATE: July 2001

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Distribution authorized to U.S. Government agencies only (proprietary information, Jul 01). Other requests for this document shall be referred to U.S. Army Medical Research and Materiel Command, 504 Scott Street, Fort Detrick, Maryland 21702-5012.

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20020329 188

## NOTICE

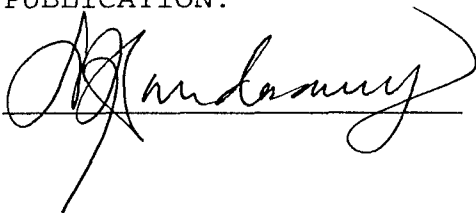
USING GOVERNMENT DRAWINGS, SPECIFICATIONS, OR OTHER DATA INCLUDED IN THIS DOCUMENT FOR ANY PURPOSE OTHER THAN GOVERNMENT PROCUREMENT DOES NOT IN ANY WAY OBLIGATE THE U.S. GOVERNMENT. THE FACT THAT THE GOVERNMENT FORMULATED OR SUPPLIED THE DRAWINGS, SPECIFICATIONS, OR OTHER DATA DOES NOT LICENSE THE HOLDER OR ANY OTHER PERSON OR CORPORATION; OR CONVEY ANY RIGHTS OR PERMISSION TO MANUFACTURE, USE, OR SELL ANY PATENTED INVENTION THAT MAY RELATE TO THEM.

### LIMITED RIGHTS LEGEND

Award Number: DAMD17-98-1-8194  
Organization: The University of Chicago

Those portions of the technical data contained in this report marked as limited rights data shall not, without the written permission of the above contractor, be (a) released or disclosed outside the government, (b) used by the Government for manufacture or, in the case of computer software documentation, for preparing the same or similar computer software, or (c) used by a party other than the Government, except that the Government may release or disclose technical data to persons outside the Government, or permit the use of technical data by such persons, if (i) such release, disclosure, or use is necessary for emergency repair or overhaul or (ii) is a release or disclosure of technical data (other than detailed manufacturing or process data) to, or use of such data by, a foreign government that is in the interest of the Government and is required for evaluational or informational purposes, provided in either case that such release, disclosure or use is made subject to a prohibition that the person to whom the data is released or disclosed may not further use, release or disclose such data, and the contractor or subcontractor or subcontractor asserting the restriction is notified of such release, disclosure or use. This legend, together with the indications of the portions of this data which are subject to such limitations, shall be included on any reproduction hereof which includes any part of the portions subject to such limitations.

THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.

  
\_\_\_\_\_

12/17/97

  
\_\_\_\_\_

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE July 2001	3. REPORT TYPE AND DATES COVERED Final (01 Jul 98 - 30 Jun 01)		
4. TITLE AND SUBTITLE Computerized Analysis of MR and Ultrasound Images of Breast Lesions		5. FUNDING NUMBERS DAMD17-98-1-8194		
6. AUTHOR(S) Maryellen Giger, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Chicago Chicago, Illinois 60637  E-Mail: m-giger@uchicago.edu		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		10. SPONSORING / MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Distribution authorized to U.S. Government agencies only (proprietary information, Jul 01). Other requests for this document shall be referred to U.S. Army Medical Research and Materiel Command, 504 Scott Street, Fort Detrick, Maryland 21702-5012.			12b. DISTRIBUTION CODE	
<p>Although general rules for the differentiation between benign and malignant mammographically identified breast lesions exist, considerable misclassification of lesions occurs with the current methods. The main goal of the proposed research is to develop noninvasive, computerized methods for analyzing ultrasound and MR (magnetic resonance) images of breast lesions to aid radiologists in their workup of suspect lesions.</p> <p>We currently have retrospectively collected over 400 ultrasound cases of mass lesions, all that had gone on to either biopsy or cyst aspiration. We extracted and calculated features related to lesion margin, shape, homogeneity (texture) and the nature of the posterior acoustic attenuation pattern. Linear discriminant analysis round-robin runs yielded <math>A_z</math> values of 0.94 and 0.87 in the task of distinguishing between benign and malignant lesions in the entire database and the equivocal database, respectively. The "equivocal" database contained lesions that had been proven to be benign or malignant by either cyst aspiration or biopsy.</p> <p>Dynamic MR data was analyzed with the computerized classification method, which includes temporal features of normalized speed and inhomogeneity of uptake, and spatial features of margin descriptors such as circularity and irregularity. Results indicate that classification based on temporal and spatial features combined can yield a positive predictive value of 94%, and has the potential to reduce the number of unnecessary biopsies by approximately 92%. We also have developed a new method for automatically extracting the lesion from the 3D image set of the breast. ROC analysis yielded <math>A_z</math> values of 0.90 when the manual segmentation was used in the classification and 0.93 when automatic segmentation was included.</p>				
14. SUBJECT TERMS Breast Cancer			15. NUMBER OF PAGES 80	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

## Table of Contents

Cover.....	
SF 298.....	2
Table of Contents.....	3
Introduction.....	4
Body.....	5
Key Research Accomplishments.....	10
Reportable Outcomes.....	10
Conclusions.....	10
References.....	11
Appendices.....	13

# **Computerized Analysis of MR and Ultrasound Images of Breast Lesions**

**Maryellen L. Giger, Ph.D.**

## **INTRODUCTION**

The main hypothesis to be tested is that, computerized analysis of breast ultrasound and MR images should yield new methods for distinguishing between malignant and benign lesions and thus, reduce the number of unnecessary biopsies. In addition, even higher performance is expected when a combination of features from mammographic, MR, and ultrasound images is used as an aid to radiologists in the task of distinguishing between malignant and benign lesions. The main goal of the proposed research is to develop noninvasive, computerized methods for analyzing ultrasound and MR (magnetic resonance) images of breast lesions to aid radiologists in their workup of suspect lesions. The specific objectives of the research to be addressed are: 1. Create a database of ultrasound and MR images including malignant lesions, benign solid masses, and complex cysts; 2. Develop noninvasive, computerized methods for characterizing the lesions to yield an output related to the probability of malignancy; and 3. Evaluate the efficacies of the new image analysis methods in the task of distinguishing between malignant and benign lesions. It is expected that the results from this research will aid radiologists in determining the likelihood of malignancy and in reducing the number of benign cases sent to biopsy. Computerized image analysis techniques that can objectively and reliably classify lesions based upon reported sonographic and/or MR characteristics of benign and malignant masses, especially if combined with their mammographic features, could significantly improve the specificity of breast imaging and the evaluation of breast masses. The proposed work is novel in that computer-aided diagnosis techniques applied to gray-scale sonographic images has not yet been reported. In addition, computerized analysis of MR images of the breast has mainly been limited to only temporal analysis using contrast media.

## **BODY**

Breast cancer is a leading cause of death in women, causing an estimated 46,000 deaths per year (1). Mammography is the most effective method for the early detection of breast cancer, and it has been shown that periodic screening of asymptomatic women does reduce mortality (2-4). Many breast cancers are detected and referred for surgical biopsy on the basis of a radiographically detected mass lesion or cluster of microcalcifications. Although general rules for the differentiation between benign and malignant mammographically identified breast lesions exist (5, 6), considerable misclassification of lesions occurs with the current methods. On average, less than 30% of masses referred for surgical breast biopsy are actually malignant (7).

Breast sonography is used as an important adjunct to diagnostic mammography and is typically performed to evaluate palpable and mammographically identified masses in order to determine their cystic vs. solid natures. The accuracy of ultrasound has been reported to be 96-100% in the diagnosis of simple benign cysts (8). Masses so characterized do not require further evaluation; however, 75% of masses prove to be indeterminate or solid on sonography and are candidates for further intervention (9). With the advent of modern high-frequency transducers that have improved spatial and contrast resolution, a number of sonographic features have emerged as potential indicators of malignancy, while other features are typical for benign masses (10,11). Benign features include hyperechogenicity, ellipsoid shape, mild lobulation, and a thin, echogenic pseudocapsule. Malignant features include spiculation, angular margins, marked hypoechogenicity, posterior acoustic shadowing, and a depth:width ratio greater than 0.8. Recently, Stavros, et al., used these and other features to characterize masses

as benign, indeterminate, and malignant (12). Their classification scheme had a sensitivity of 98.4% and a negative predictive value of 99.5%. However, the sonographic evaluation described by these investigators is much more extensive and complex than is usually performed at most breast imaging centers.

Breast MR imaging as an adjunct to mammography and sonography reveals breast cancer with a higher sensitivity than do mammography and sonography only (13). However, using all three methods in the human interpretation process yielded a lower specificity. It also has been shown that temporal analysis from dynamic MR correlates with intensity of fibrosis in fibroadenomas (14). Some computerized analyses of spatial features are being performed. Adams et al. achieved a separation between malignant and benign lesions using a statistical analysis, however, their database consisted of only 16 cases (15).

Computerized image analysis techniques that can objectively and reliably classify lesions based upon reported sonographic and/or MR characteristics of benign and malignant masses, especially if combined with their mammographic features, could significantly improve the specificity of breast imaging and the evaluation of breast masses. Computer-aided techniques have been applied to the color Doppler evaluation of breast masses with promising results (16). However, color Doppler imaging is a technique which focuses only upon the vascularity of lesions. Since not all sonographically visible cancers have demonstrable neovascularity, this technique is inherently somewhat limited. On the other hand, computer-aided diagnosis techniques applied to gray-scale sonographic images has not yet been reported. In addition, computerized analysis of MR images of the breast has mainly been limited to only temporal analysis using contrast media.

Comprehensive summaries of investigations in the field of mammography CAD have been published by the co-P.I. (17, 18). In the 1960s and 70s, several investigators attempted to analyze mammographic abnormalities with computers. These previous studies demonstrated the potential capability of using a computer in the detection of mammographic abnormalities. Gale et al. (19) and Getty et al. (20) are both developing computer-based classifiers, which take as input diagnostically-relevant features obtained from radiologists' readings of breast images. Getty et al. found that with the aid of the classifier, community radiologists performed as well as unaided expert mammographers in making benign-malignant decisions. Swett et al. (21) are developing an expert system to provide visual and cognitive feedback to the radiologist using a critiquing approach combined with an expert system. At the University of Chicago, we have shown that the computerized analysis of mass lesions (22) and clustered microcalcifications (23) on digitized mammograms yields performances similar to an expert mammographer and significantly better than average radiologists in the task of distinguishing between malignant and benign lesions.

The proposed work is novel in that computer-aided diagnosis techniques have not yet been applied to gray-scale breast ultrasound and/or MR images. In addition, future research involving the use computers to merge features from mammographic, MR, and ultrasound images, as an aid to radiologists, has not yet been investigated.

The main goal of the proposed research is to develop noninvasive, computerized methods for analyzing ultrasound and MR (magnetic resonance) images of breast lesions to aid radiologists in their workup of suspect lesions. The specific objectives of the research to be addressed are: 1. Create a database of ultrasound and MR images including malignant lesions, benign solid masses, and complex cysts; 2. Develop noninvasive, computerized methods for characterizing the lesions to yield an output related to the probability of malignancy; and 3. Evaluate the efficacies of the new image analysis methods in the task of distinguishing between malignant and benign lesions. It is expected that the results from this research will aid

radiologists in determining the likelihood of malignancy and in reducing the number of benign cases sent to biopsy.

## **1. Establishment of a database of ultrasound and MR images**

### **Methods**

Approximately 500 sonographically demonstrated lesions will be collected which will include aspirated complex cysts, and biopsied solid benign and malignant masses. The database of these collected cases will include the MR, sonographic, and mammographic images as well as the lesions' ultimate dispositions and diagnoses. (Note that funding already exists for the computerized analysis of the mammographic lesions). Based upon our current case load, we estimate that approximately 30% of the lesions will be complex cysts which required aspiration to prove their cystic nature, 40% will be benign solid masses, and 30% will be cancers. Palpable and mammographically identified masses are evaluated sonographically by representative images in orthogonal planes, obtaining measurements in these same planes, and most masses are also evaluated with color Doppler imaging. Although the preliminary studies on ultrasound images involved the digitization of ultrasound films, the ultrasound images in this new database will be obtained directly from an ATL ultrasound machine, which produces digital image data. In addition, approximately 50 cases of MR images of the breast will be collected with a T1-weighted sequence, using coronal slices. After injection of GD contrast, 4 to 6 scans of both breasts will be obtained at 90 sec. time intervals. Biopsy results will be used to determine truth regarding malignancy.

### **Results to Date**

We currently have retrospectively collected over 400 ultrasound cases of mass lesions, all that had gone on to either biopsy or cyst aspiration. The images are obtained from University of Chicago and Northwestern University. The images are transferred in digital format from the ATL unit. The digital images within the ATL unit are obtained by screen capture. For each case we have at least two views of the lesion. We are currently collecting the corresponding mammograms for the study. We have digitized over 100 cases with 2 to 7 films per case. We expect to finish the database by the end of summer 2001. Approximately 150 cases will not be digitized due to the case either having a non-mammographically seen lesion or a lesion which caused a call back for ultrasound but did not correspond to the lesion interpreted on the ultrasound.

We currently have retrospectively collected 35 coronal MR cases from University of Muenster, 362 sagittal MR cases of the breast from University of Pennsylvania, and 90 cases from the University of Berlin (which follow a protocol similar to University of Muenster). These are all volume datasets. Of the 362 sagittal cases, 253 are focal (192 malignant, 51 benign, 10 normal), 74 are diffuse lesions (48 malignant and 25 benign), 10 are ductal (9 malignant and 1 benign), and 25 showed no enhancement (3 malignant, 19 benign, 3 normal).

## **2. Development of computerized method for the classification of lesions**

The computerized method will include the image analysis of the texture within the lesion, the analysis of the margin of the lesion, and a comparison of the lesion with its surrounding tissue. Computerized analysis of the texture pattern in the lesion will be based on various texture analysis methods we have been investigating in our laboratory including Fourier spectra analysis and artificial neural networks. We note that it is extremely important to understand the relationship between the mathematical texture measures and the physical nature of the breast parenchyma.

The computerized analysis of the margin characteristics (edge definition) will involve feature extraction using radial edge-gradient analysis. We have done similar analysis on radiographic masses in determining their margin characteristics (spiculated and ill defined) (22).



Two promising measures are the FWHM and the average radial gradient which correspond to the degree of spiculation and how ill-defined is the margin, respectively. From the computer-extracted margin, we will also determine the shape and irregularity of each lesion.

Specifically for the ultrasound images, comparison of the "density" and the texture patterns of the lesion with neighboring regions, including those deep to the lesion, will be performed in order to quantify its echogenicity and the amount of any posterior acoustic shadowing or enhancement. This will be performed by comparing feature values "below" the lesion to those obtained along side and below the lesion.

Temporal features will be determined from analyzing the MR image data over time. The contrast medium uptake curve will be analyzed at various spatial locations within and around the suspect lesion. Temporal operators include the maximum uptake, mean gradient of uptake, and rms variation. Both two-dimensional and three-dimensional features will be calculated, e.g., irregularity and margin gradient characteristics. In addition, the spatial features will be investigated as a function of time.

We plan to use artificial neural networks along with other measures of the mass in question to obtain an estimate of the likelihood of malignancy. We will investigate merging the ultrasound image features and MR features with those from mammographic images of the same lesion. We already have funding support for the investigation involving radiographic imaging of masses.

The various features will serve as input data and will be supplied to the input units of the artificial neural network. Prior to input to the ANN, the features will be normalized between 0 and 1. The output data from the neural network are then obtained through successive nonlinear calculations in the hidden and output layers. The calculation at each unit in a layer includes a weighted summation of all entry numbers, an addition of a certain offset number, and a conversion into a number ranging from 0 to 1 using a sigmoid-shape function such as a logistic function. The neural network will be trained by a back-propagation algorithm using pairs of training input data and desired output data. The desired output data will be initially 1 if features of a malignant lesion are input and 0 otherwise. Once trained, the neural network will accept features of a lesion and will output a value that will be related to a likelihood of malignancy. Feature selection will be performed by analyzing the average and standard deviation of the various features for both malignant and benign lesions.  $A_z$  values will be calculated for each feature as well as for the output of the ANNs. In addition, genetic algorithms, which we have used, in a pilot study, for optimizing feature selection for the task of distinguishing true-positive and false-positive mass detections, will also be used.

#### Results to Date: Ultrasound

We are developing computerized analyses of breast lesions in ultrasound images to aid in the discrimination between malignant and benign lesions (24). We extracted and calculated features related to lesion margin, shape, homogeneity (texture) and the nature of the posterior acoustic attenuation pattern in ultrasound images of the breast. Our database contained 184 digitized ultrasound images from 58 patients with 78 lesions. Benign lesions were confirmed by biopsy, cyst aspiration, or image interpretation alone, while malignant lesions were confirmed by biopsy. ROC analysis was used to study the performance of the various individual features and the output from linear discriminant analysis in distinguishing benign from malignant lesions. From ROC analysis, the feature characterizing the margin yielded  $A_z$  values of 0.85 and 0.75, in the task of distinguishing between benign and malignant lesions in the entire database and in an equivocal database, respectively. The "equivocal" database contained lesions that had been proven to be benign or malignant by either cyst aspiration or biopsy. Linear discriminant analysis round-robin runs yielded  $A_z$  values of 0.94 and 0.87 in the task of

distinguishing between benign and malignant lesions in the entire database and the equivocal database, respectively.

We are currently evaluating the method on a database of ultrasound images from Northwestern University. The database of over 400 cases includes pathology truth as well as radiologists BI-RADS ratings with all cases having gone to biopsy or aspiration. Our previous method required radiologists' manually-drawn lesion contours as input to the computerized classification scheme. The current method, however, involves automatic segmentation of the lesion contour from the ultrasound image data. Of the 410 cases, 126 were complex cysts, 186 were benign solid lesions, and 98 were malignant lesions. Features related to lesion margin, shape, echogenicity (texture) and posterior acoustic attenuation were automatically extracted. To evaluate the performance of the computer alone, the entire database was divided into training and testing groups. The independent linear discriminant analysis yielded a validation result of an  $A_z$  of 0.89 and a partial  $A_z$  value at 0.90 sensitivity of 0.52. In addition, in order to evaluate the performance of the computer relative to that of the radiologists, 125 cases were assessed for suspicion by an expert sonographer. Round-robin analysis in the task of distinguishing malignant from benign lesions yielded  $A_z$  values of 0.88 and 0.92 for the computer and the radiologist, respectively.

We have submitted two manuscripts to Medical Physics -- one on the computerized segmentation method and one on the computer-extracted ultrasound features. These preprints are in the appendix.

#### Results to Date: MRI

We are developing computerized analyses of breast lesions in MR images to aid in the discrimination between malignant and benign lesions (25). Dynamic MR data was obtained from 27 patients by a T1-weighted sequence, using 64 coronal slices, a typical slice thickness of 2 mm, and a pixel size of 1.25 mm. After injection of GDTA contrast, 4 to 6 scans of both breasts were obtained at 90 sec. time intervals. The database contained 13 benign and 15 malignant lesions. Our computerized classification method includes temporal features of normalized speed and inhomogeneity of uptake, and spatial features of margin descriptors such as circularity and irregularity. Our results indicate that classification based on temporal and spatial features combined can yield a positive predictive value of 94%, and has the potential to reduce the number of unnecessary biopsies by approximately 92%.

We have developed a new method for automatically extracting the lesion from the 3D image set of the breast. Our previous results were based on the use of manually-drawn lesion contours in the various slices of the MR data. The new segmentation method involves the use of an encompassing shell to limit the region for local thresholding. ROC analysis yielded  $A_z$  values of 0.90 when the manual segmentation was used in the classification and 0.93 when automatic segmentation was included.

We are currently evaluating the method on 362 cases from the University of Pennsylvania as well as the cases from the University of Muenster and University of Berlin. The UPENN images differ from our initial database in that these cases are sagittal and had fat suppression applied. Also, the UPENN dataset uses fat suppression and thus a modification in the automatic lesion extraction method was made. For the evaluation, we developed a new interface for the human delineation of the lesion margin in 3D to serve as "margin truth". While outlining the margin in a slice, the observer can see their outline in other views. One performance of index is an overlap calculation in which, in terms of voxels, we calculate the intersection of the human and computer margins divided by the union. We now have this margin truth for roughly 200 cases and we are now running the overlap comparison to determine if the computer outlines similar to the human. We will also do the comparison in

terms of the performance of the features extracted from the lesion in the task of distinguishing malignant and benign lesions.

### **3. Evaluation in the task of distinguishing between malignant and benign lesions.**

In order to test the capability of the neural networks to learn the features of malignant and benign lesions, a consistency test will be conducted in which the network is first trained with all the cases in the database and then tested with the same cases used in the training. A consistency test indicates that the network is able to "remember" all of the input types that were used for training. However, it is more important to test if the network can learn a generalized set of inputs from the examples provided and if it can then make a correct prediction for new cases that were not included in the training. Thus, a round-robin method will be employed to test the network's generalizing ability. With the jack-knife method, all but one of the cases are selected randomly from the database for training of the network, and the remaining one case is used for testing the network. The output values are then compared to the "truth" data. Various combinations of training and testing pairs will be selected by using a random number generator and the results will be analyzed using ROC analysis. ROC curves will be obtained by fitting continuous output data from the neural networks using the LABROC4 program (26). The area under the ROC curve ( $A_z$ ) will be used as an indicator of performance. In order to determine the structure of the neural network as well as the necessary number of training iterations, we will analyze the consistency results and the round-robin results in terms of  $A_z$  as a function of number of iterations, momentum, learning rate and number of hidden units. We use  $A_z$  as an indicator of performance since it includes information on both the sensitivity and specificity of the measures.

The proposed techniques are expected to yield measures about the likelihood of malignancy. Receiver Operating Characteristic (ROC) analysis (26) will be employed in evaluating the performance of the measures. We have used ROC analysis successfully in both evaluating the performance of human observers as well as that of computerized schemes. The task in which the image features will be evaluated will be in their ability to determine an estimate of the likelihood of malignancy. The decision variable for the ROC analysis will be each individual feature as well as combined measures within a modality and combined measures from multiple modalities (x-ray, MR, and ultrasound).

We expect that 500 lesions and their ultrasound images will be available for testing. Note that here the measure of performance will be the  $A_z$  value (from ROC analysis) obtained in the task of distinguishing between malignant and benign lesions. To obtain an estimate of the number of lesions needed for adequate statistical power in testing differences in  $A_z$  values, we assume only a correlation of 0.60 between the estimates of  $A_z$  that are found for our current method involving the computerized analysis of mammographic lesions ( $A_z=0.87$ ) and that for the expected improved method ( $A_z=0.92$ ). With  $N_{\text{pos}}$  patients who have a malignant lesion and  $N_{\text{neg}}$  patients who have a benign lesion, the standard error of the resulting estimate can be approximated (Eqn. 9 in Ref. 27) by the expression  $\{[2A_z - (1-f)(1-A_z)](1-A_z)/3N_{\text{pos}}\}^{1/2}$ , where  $f$  represents the ratio  $N_{\text{pos}}/N_{\text{neg}}$ . Thus, with  $f = 1$ , the statistical power at a critical significance level of  $\alpha = 0.05$  for 500 mass lesions is 94%.

### **Results to Date**

The results from the evaluation of the methods is described in the preliminary studies described above. We have submitted two manuscripts to Medical Physics -- one on the computerized segmentation method and one on the computer-extracted ultrasound features. These preprints are in the appendix. We also presented preliminary results on the combination

of mammographic and sonographic features on the common database as SPIE Medical Imaging 2001.

## KEY RESEARCH ACCOMPLISHMENTS

- Development of an automatic method for segmenting lesions on ultrasound.
- Development of robust features for characterizing lesions in ultrasound images of the breast.
- Development of robust features for characterizing lesions in MRI images of the breast.
- Investigation and development of methods for segmentation in 2D for ultrasound images and in 2D and 3D for MRI image datasets.
- Preliminary investigation of merging mammographic and sonographic features of lesions on a common database.

## REPORTABLE OUTCOMES

1. Gilhuijs KGA, Giger ML, Bick U: Automated analysis of breast lesions in three dimensions using dynamic magnetic resonance imaging. Medical Physics 25:1647-1654, 1998.
2. Giger ML, Al-Hallaq H, Huo A, Moran C, Wolverton DE, Chan CW, Zhong W: Computerized analysis of lesions in ultrasound images of the breast. Academic Radiology 6: 665-674, 1999. (also being reprinted in the Yearbook of Radiology)
3. Horsch K, Giger ML, Venta LA, Huo Z, Vyborny CJ: Computer-aided diagnosis of breast lesions on ultrasound. Proceedings, International Workshop on Digital Mammography. Toronto, Canada, June, 2000.
4. Horsch K, Giger ML, Venta LA, Vyborny CJ: Automatic segmentation of breast lesions on ultrasound. Medical Physics (in press).
5. Horsch K, Giger ML, Venta LA, Vyborny CJ: Computerized diagnosis of breast lesions on ultrasound. Medical Physics (accepted with revision).
6. Giger ML, Huo Z, Horsch K, Hendrick E, Venta L, Vyborny CJ: Computer-aided diagnosis of lesions on multimodality images of the breast. Proc. SPIE 2001 (in press).

## CONCLUSIONS

We have made great strides in the development of methods for the classification of lesions in ultrasound and MR images of the breast. We are retrospectively collecting large datasets of ultrasound and MR cases with solid pathology truth and radiologists' ratings. These cases include malignant lesions, benign solid masses, and complex cysts. We are developing noninvasive, computerized methods for characterizing the lesions to yield an output related to the probability of malignancy and plan to evaluate the efficacies of the new image analysis

methods in the task of distinguishing between malignant and benign lesions. It is expected that the results from this research will aid radiologists in determining the likelihood of malignancy and in reducing the number of benign cases sent to biopsy.

## REFERENCES

1. American Cancer Society: Cancer Facts and Figures -- 1995. Atlanta: American Cancer Society, 1995.
2. Feig SA: Decreased breast cancer mortality through mammographic screening: Results of clinical trials. Radiology 167:659-665, 1988.
3. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O: Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am 30:187-210, 1992.
4. Smart CR, Hendrick RE, Rutledge JH, Smith RA: Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials. Cancer 75:1619-26, 1995.
5. Bassett LW, Gold RH: Breast Cancer Detection: Mammography and Other Methods in Breast Imaging. New York: Grune and Stratton, 1987.
6. Kopans DB: Breast Imaging. Philadelphia: JB Lippincott, 1989.
7. Brown ML, Houn F, Sickles EA, Kessler LG: Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. AJR 165:1373-1377, 1995.
8. Jackson VP: The role of US in breast imaging. Radiology 177:305-311, 1990.
9. Hilton SW, Leopold GR, Olson LK, Wilson SA: Real-time breast sonography: application in 300 consecutive patients. AJR 147:479-486, 1986.
10. Tohno E, Cosgrove DO, Sloane JP: Ultrasound Diagnosis of Breast Diseases. Churchill Livingstone, Edinburgh, 1994, pp. 50-73.
11. Fornace BD, Lorigan JG, Andry E: Fibroadenoma of the breast: sonographic appearance. Radiology 172:671-675, 1989.
12. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA: Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology 196:123-134, 1995.
13. Muller-Schimpfle M, Stoll P, Stern W. et al.: Do mammography, sonography, and MR mammography have a diagnostic benefit compared with mammography and sonography? AJR 168: 1323-1329, 1997.
14. Brinck U, Fischer U, Korabiowska M, et al.: The variability of fibroadenoma in contrast-enhanced dynamic MR mammography. AJR 168: 1331-1334, 1997.
15. Adams AH, Brookeman JR, Merickel MB: Breast lesion discrimination using statistical analysis and shape measures on magnetic resonance imagery. Comp Med Imaging and Graphics 15: 339-349, 1991.
16. Huber S, Delorme S, Knopp MV, Junkermann H, Zuna I, von Fournier D, van Kaick G: Breast tumors: computer-assisted quantitative assessment with color Doppler US. Radiology 192:797-801, 1994.
17. Giger ML: Computer-aided diagnosis. In: Syllabus: A Categorical Course on the Technical Aspects of Breast Imaging, edited by Haus A, Yaffe M. Oak Brook, IL: RSNA Publications, 1993, pp. 272-298.
18. Vyborny CJ, Giger ML: Computer vision and artificial intelligence in mammography. AJR 162:699-708, 1994.
19. Gale AG, Roebuck EJ, Riley P, Worthington BS, et al.: Computer aids to mammographic diagnosis. British Journal of Radiology 60: 887-891, 1987.
20. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA: Enhanced interpretation of diagnostic images. Invest. Radiol. 23: 240-252, 1988.

21. Swett HA, Miller PA: ICON: A computer-based approach to differential diagnosis in radiology. Radiology 163: 555-558, 1987.
22. Huo Z, Giger ML, Vyborny CJ, Bick U, Lu P, Wolverton DE, Schmidt RA: Analysis of spiculation in the computerized classification of mammographic masses" Medical Physics 22:1569-1579, 1995.
23. Jiang Y, Nishikawa RM, Wolverton DE, Giger ML, Doi K, Schmidt RA, Vyborny CJ: Automated feature analysis and classification of malignant and benign clustered microcalcifications. Radiology 198(3):671-678, 1996.
24. Giger ML, Al-Hallaq H, Huo A, Moran C, Wolverton DE, Chan CW, Zhong W: Computerized analysis of lesions in ultrasound images of the breast. Academic Radiology (in press).
25. Gilhuijs KGA, Giger ML, Bick U: Automated analysis of breast lesions in three dimensions using dynamic magnetic resonance imaging. Medical Physics 25:1647-1654, 1998.
26. Metz CE: Some practical issues of experimental design and data analysis in radiological ROC studies. Invest. Radiol. 24: 234-245, 1989.
27. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating graph. J. Math Psych 12: 387-415, 1975.

# Computerized Diagnosis of Breast Lesions on Ultrasound

Karla Horsch\*, Maryellen L. Giger\*, Luz A. Venta<sup>†‡</sup>, Carl J. Vyborny\*

June 12, 2001

---

\*Department of Radiology, University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637

<sup>†</sup>Currently: Baylor Methodist Breast Care Center, 6550 Fannin St., suite 701, Houston, TX 77030

<sup>‡</sup>Formally: Lynn Sage Breast Center, Northwestern University, Chicago, IL 60611

### Abstract

We present a computer-aided diagnosis (CAD) method for breast lesions on ultrasound that is based on the automatic segmentation of lesions and the automatic extraction of four features related to the lesion shape, margin, texture and posterior acoustic behavior. Using a database of 400 cases (94 malignant lesions, 124 complex cysts and 182 benign solid lesions), we investigate the marginal benefit of each feature in our CAD method and the performance of our CAD method in distinguishing malignant lesions from various classes of benign lesions. Finally, independent validation is performed on our CAD method. Eleven independent trials yielded an average  $A_z$  value of 0.87 in the task of distinguishing malignant from benign lesions.

**Keywords:** Breast sonography, computer-aided diagnosis.



# 1 Introduction

Although ultrasound is currently used to diagnose simple cysts with a reported accuracy of 96-100% [1], it is not currently used to differentiate solid lesions by most radiologists. Biopsy or aspiration is the usual management for those lesions that are not diagnosed as clearly cystic during the ultrasound exam. In general, for masses undergoing surgical biopsy, only 10 to 31% are actually cancerous. As biopsy is associated with greater cost and patient anxiety, researchers are exploring the diagnostic capability of breast sonography in differentiating malignant from benign solid masses. In a recent study, Stavros et al [2] developed a classification scheme, using various human-extracted sonographic features, that achieved a sensitivity of 98.4% and a negative predictive value of 99.5% on a data set of 750 solid breast masses.

Computer-aided diagnosis (CAD) in breast ultrasound is being explored by various researchers. Giger et al. have developed a computer-aided diagnosis scheme that uses clinically-motivated, computer-extracted sonographic features to quantify breast lesion shape, margin, texture and posterior acoustic behavior [3]. Other researchers have concentrated on computer-extracted texture features [4], [5] or RF signal characteristics [6]. Sahiner et al [7] has explored computerized characterization of breast masses using texture features extracted from three-dimensional ultrasound images.

We present a CAD method for breast lesions on ultrasound that performs automatic feature extraction on automatically-segmented lesions. The computer-extracted features are then merged through linear discriminant analysis. Three studies were performed on a large clinical database of 400 cases: 1) evaluation of the marginal benefit of each feature to our CAD method, 2) determination of the performance of our CAD method in distinguishing carcinomas from different types of benign lesions, and 3) independent validation of the method using 11 independent trials.

## 2 Material and Methods

### 2.1 Database

The database in our study consists of 400 consecutive ultrasound examinations acquired during diagnostic breast evaluations at the Lynn Sage Breast Center of Northwestern Memorial Hospital, in which lesions were detected and either biopsied or aspirated. Of the 400 cases, 94 were malignant solid lesions, 124 were complex cysts, 107 were benign tumors (fibroadenomas and papillomas), 65 were fibrocystic disease, and 10 had other benign causes. The database contains no simple cysts. The number of each lesion type and information on their size is listed in Table 1. Examples of a malignant mass, a complex cyst, fibrocystic disease and a fibroadenoma are shown in Figure 1. The 757 images in our study were obtained with an ATL 3000 unit using a 5 MHz transducer and were captured directly from the 8-bit video signal. The number of images per cases varied from one to six.

### 2.2 Notation

In what follows, the image gray level data is denoted by  $I(m, n)$  where  $m = 0, 1, \dots, M_I - 1$  and  $n = 0, 1, \dots, N_I - 1$ . Here,  $M_I$  is the number of pixels in the lateral direction of the image and  $N_I$  is the number of pixels in the depth direction of the image. The gradient image is denoted by  $\nabla I$  and is computed using Sobel filters. The gray level data of a subimage, or region of interest (ROI), is denoted by  $R(m, n)$  where  $m = 0, 1, \dots, M_R - 1$  and  $n = 0, 1, \dots, N_R - 1$ . Again,  $M_R$  is the number of pixels in the lateral direction of the ROI and  $N_R$  is the number of pixels in the depth direction of the ROI. The points on the lesion margin have  $x$  and  $y$  coordinates  $(\gamma_1(j), \gamma_2(j))$  where the index  $j = 0, 1, \dots, J - 1$  and  $J$  is the number of points in the margin. We will also require a vector  $\hat{r}(m, n)$  of unit length in the radial direction from the geometric center of the lesion to the point indexed by  $(m, n)$ .

The geometric center of the lesion is computed by

$$(m_c, n_c) = \left( \frac{\sum_m \sum_n L(m, n)m}{A}, \frac{\sum_m \sum_n L(m, n)n}{A} \right) \quad (1)$$

where  $L(m, n)$  is the lesion mask, a binary image having value 1 within the image and 0 elsewhere.  $A$  is the area of the lesion.

## 2.3 Lesion Segmentation

For each image, lesions were segmented both manually and automatically from the normal breast tissue.

Manual segmentation involved displaying each ultrasound image on an IBM PowerDisplay20 monitor, and having a mammographer or medical physicist outline the lesion margin. Of the 757 images, 360 were outlined by a mammographer and 397 were outlined by a medical physicist. In another paper [8], we performed an analysis to compare the lesion margins of a medical physicist with the lesion margins of an experienced mammographer. Each of 113 images was outlined by two mammographers and a medical physicist, and the overlap [8] between the first and second mammographers was compared to the overlap between the first mammographer and the medical physicist. In this way, we demonstrated that the medical physicist performs similarly to an experienced mammographer in the task of outlining lesion margins on ultrasound images.

Our automatic lesion segmentation algorithm involves the following steps [8]: (1) pre-processing by cropping and median filtering the image, (2) multiplication with a Gaussian constraint function, (3) determination of potential lesion margins through gray-value thresholding, and (4) maximization of a utility function for the potential lesion margins. The Gaussian constraint function is centered at the manually defined lesion center, which is determined by computing the geometric center of the manually segmented lesion margins. (For description of how the geometric center is computed, see Section 2.2.)

We use a utility function called the Average Radial Derivative (ARD), which gives the average directional derivative in the radial direction

$$ARD = \frac{1}{J} \sum_{j=0}^{J-1} \nabla I(\gamma_1(j), \gamma_2(j)) \cdot \hat{r}(\gamma_1(j), \gamma_2(j)), \quad (2)$$

where, as defined in Section 2.2,  $I$  is the image gray level data,  $\nabla I$  is the gradient image,  $(\gamma_1, \gamma_2)$  is the discretized lesion margin,  $J$  is the number of points in the discretized margin, and  $\hat{r}(\gamma_1, \gamma_2)$  is the unit vector in the radial direction from the geometric center of the lesion to the point  $(\gamma_1, \gamma_2)$ . Note that the center of the lesion is the only information defined manually for the automatic segmentation algorithm.

## 2.4 Automatic Feature Extraction

In the clinical evaluation of breast lesions on ultrasound, radiologists consider features that include lesion shape, margin definition, echogenic texture, posterior acoustic enhancement or shadowing [9]. Benign lesions tend to demonstrate a lesion shape that is wider than taller, well-defined, smooth margins, and posterior acoustic enhancement. Benign solid lesions tend to be hyperechoic while benign cysts tend to be anechoic. Malignant lesions, on the other hand, tend to be taller than wider with ill-defined, angular margins while also manifesting hypoechogenicity, and posterior acoustic shadowing. We will consider computer-extracted features which quantify these clinically-used features.

The shape feature that we consider is the depth-to-width ratio (DWR) of the lesion, which is defined by

$$DWR = \frac{\text{Depth}}{\text{Width}} = \frac{\max_j(\gamma_2(j)) - \min_j(\gamma_2(j))}{\max_j(\gamma_1(j)) - \min_j(\gamma_1(j))}, \quad (3)$$

where  $j = 0, 1, \dots, J - 1$ . (See Section 2.2 for the definition of  $\gamma_1$ ,  $\gamma_2$  and  $J$ .) Cysts and benign solids tend to be wider than deep and thus, benign lesions tend to yield smaller values for the DWR than malignant lesions.

The margin feature that we consider is the normalized radial gradient (NRG), which is a measure of the average orientation of the gray level gradients along the margin. It is given by [10, 11]

$$NRG = \frac{\sum_{j=0}^{J-1} \nabla I(\gamma_1(j), \gamma_2(j)) \cdot \hat{r}(\gamma_1(j), \gamma_2(j))}{\sum_{j=0}^{J-1} \|\nabla I(\gamma_1(j), \gamma_2(j))\|}. \quad (4)$$

(See Section 2.2 for the definition of  $\nabla I$ ,  $\hat{r}$ ,  $\gamma_1$ ,  $\gamma_2$  and  $J$ . In general, the NRG varies between -1 and 1, being near 1 when the gradients tend on average to point radially outward and near -1 when the gradients tend on average to point radially inward. For ultrasound, however, the NRG tends to be greater than zero. This is because almost all US lesions of significance tend to be darker (i.e. less echogenic) than the surrounding tissue, and therefore, the gradients along the lesion margin tend, on average, to point outward toward increasing gray-level values. The benign lesions tend to yield larger values of the NRG. Observe that the NRG contains no information about the magnitude of the gradient along the margin.

To quantify texture, the autocorrelation in depth of  $R$ , the gray level values in the minimal rectangular ROI containing the lesion, is used to define

$$COR = \frac{\sum_{n=0}^{N_R-1} \overline{C_y}(n)}{\overline{C_y}(0)}, \quad (5)$$

where the autocorrelation in depth and its sum in the lateral direction are

$$\begin{aligned} C_y(m, n) &= \sum_{p=0}^{N_R-1-n} R^2(m, n+p) R^2(m, p), \\ \overline{C_y}(n) &= \sum_{m=0}^{M_R-1} C_y(m, n). \end{aligned}$$

A picture of the minimal rectangular ROI for an example lesion is shown in Figure 2. Observe that the COR is a sum and not an average. Thus, COR includes not only texture information, but also size information.

Posterior acoustic behavior is quantified by comparing the gray-level values posterior to the lesion to those in adjacent tissue at the same depth. This comparisons is accomplished

by considering differences in the average gray level values of the appropriate region of interest (ROI). To avoid edge shadows, define an ROI which is the lesion itself minus a portion of its lateral sides. This is done to avoid edge shadows. The left, post, and right ROIs are rectangular with the same width and area as the ROI which is the lesion itself minus a portion of its lateral sides. These ROIs are shown schematically in Figure 3. The posterior acoustic behavior feature considered is the minimum side difference ( $MSD$ ). To understand why the minimum is chosen, observe that the difference in the average gray-level posterior to the lesion and that in adjacent tissue at the same depth tends to be negative for malignant lesions because of posterior acoustic shadowing. Choosing the minimum thus errs on the side of malignancy. The posterior acoustic behavior feature is defined as

$$MSD = \min(A_{post} - A_{left}, A_{post} - A_{right}), \quad (6)$$

where  $A_{post}$ ,  $A_{left}$  and  $A_{right}$  is the average gray-level value over the appropriate ROI.

The above features are computed for each image in both the manually-segmented and computer-segmented approaches. A particular feature value for a given lesion and segmentation is the average of that feature over all the views available for the lesion, with each lesion being represented by one to six images.

The computer-extracted features of shape, margin, texture and posterior acoustic behavior are then merged through linear discriminant analysis (LDA) [12] for automatic classification.

## 2.5 Evaluation

In order to investigate the marginal difference of adding a feature to the LDA, combinations of two and three features are merged in addition to merging all four features. In the case of merging all four features, both consistency and round robin evaluations are performed in the task of distinguishing malignant and benign lesions. In a consistency LDA [12], each case is

classified according to a classifier trained with *all of the cases*. In a round robin LDA [12], one of the cases is removed from the data and that case is classified according to a classifier trained with *the remaining cases*. This process is then repeated for each lesion. Since we are using a *linear* classifier, we do not expect much change in the performance of the consistency and round robin evaluations, and indeed, we found this to be the case.

It is of interest to compare how our CAD method performs in differentiating carcinomas from the different types of benign lesions. In particular, we use the two, three and four-feature LDA classifiers to differentiate malignant lesions from benign lesions for the following data subsets: (A) the entire database, (B) carcinomas and benign solid lesions (all benign cases except for complex cysts), (C) carcinomas and complex cysts, (D) carcinomas and benign tumors (fibroadenomas and papillomas), (E) carcinomas, complex cysts and benign tumors, and (F) carcinomas and fibrocystic disease. Data subset (B) is important as there is considerable clinical importance in the differentiation of malignant and solid benign lesions. The reason for considering the data subset (E) is that first, complex cysts and benign tumors are the most represented benign lesion types in our database. The second reason is that complex cysts and benign tumors tend to have well-defined margins, and are thus more easily differentiated from other types of lesions by our CAD method. Data subset (F) is interesting because many of the cases of fibrocystic disease are difficult for radiologists to see on our images, much less to diagnose.

For the two, three and four-feature classifiers, the LDA was trained on the entire database and then tested on the data subset containing only the particular class of benign lesions of interest. We emphasize that for a given set of features, the classifier is not retrained for each data subset, but rather the same classifier (trained on the entire database) is used for each of the different data subsets. For the round robin evaluation of the four-feature classifier, one of the cases from a particular data subset is removed and that case is classified according to the classifier trained with the remaining cases from the entire database. This process is then repeated for each case in the particular data subset.

Finally, independent validation was also performed on our CAD method. We randomly selected half the cysts, half the benign solid lesions and half the malignant lesions to be our training database of 200 cases. The remaining 200 cases formed the validation database (see Figure 4). The random selection was performed 11 times. For each of the 11 randomizations, LDA was used to merge the four computer-extracted features in the training database, and the resulting classifiers were evaluated on the validation database.

Receiver operating characteristic (ROC) analysis [13] was used to evaluate (by case, not by image) the performance of the individual computer-extracted features and the various LDA classifiers in the task of distinguishing malignant lesions from various classes of benign lesions. The area under the ROC curve, or  $A_z$  value, and the partial area at 0.90 sensitivity, or  $_{0.9}A_z$  value [14], were used as indicators of merit.

### 3 Results and Discussion

Figures 5 and 6 show scatter plots of the four computer-extracted features derived from the automatically-defined lesion margins for the entire database (400 cases). As anticipated, malignant lesions tend to demonstrate a larger depth-to-width ratio, a smaller normalized radial gradient value and a more negative minimum side difference than benign lesions. The autocorrelation based feature demonstrates more overlap in the values of benign and malignant cases.

Combinations of two, three and four of the computer-extracted features were merged with LDA using the entire database and the performance of the resulting classifiers tested on each data subset using ROC analysis. The  $A_z$  and partial  $_{0.9}A_z$  values, as well as the standard estimated deviations on the  $A_z$  values (computed by LABROC4 [15]), for each of the individual computer-extracted features are shown in Tables 2, 3, and 4 for data subsets A and B, data subsets C and D, and data subsets E and F, respectively. Also shown are the  $A_z$  and partial  $_{0.9}A_z$  values for the combination of all four computer-extracted features, for both



the consistency and round robin evaluations. When tested on the entire database, the best performing combination of two computer-extracted features was the depth-to-width ratio and the normalized radial gradient. The best performing combination of three computer-extracted features was the depth-to-width ratio, the normalized radial gradient and the autocorrelation feature. In Tables 2, 3, and 4, we report only the performance of these strongest two and three feature classifiers.

Considering the round robin evaluation of all four features, we see that the classifier has the best performance when differentiating malignant lesions from complex cysts, with  $A_z$  and  $_{0.9}A_z$  values of 0.95 and 0.78, respectively, in the case of manual segmentation, and  $A_z$  and  $_{0.9}A_z$  values of 0.94 and 0.71, respectively, in the case of automatic segmentation. The worst performance was demonstrated by the classifier when differentiating malignant lesions from fibrocystic disease with  $A_z$  and  $_{0.9}A_z$  values of 0.80 and 0.37, respectively, in the case of manual segmentation, and  $A_z$  and  $_{0.9}A_z$  values of 0.70 and 0.19, respectively, in the case of automatic segmentation. ROC curves of the round robin evaluations of the LDA using all four features for data subsets A, B and C are shown in Figure 7.

It should be noted that for a given segmentation method and benign class, the performances of the strongest two-feature, strongest three-feature and four-features classifiers are fairly similar. To determine whether the differences in performance are statistically significant, univariate z-score tests of the differences in the  $A_z$  and partial  $A_z$  values were performed using the program CLABROC [16]. Using a p-value less than 0.05 as the cut-off, we failed to show a statistically significant difference between the performances of the strongest three-feature and four-feature classifiers in differentiating carcinomas from any class of benign lesions, using either manual and automatic segmentation. This indicates that for our database, the posterior acoustic behavior feature, MSD, does not add significantly to the performance of the four-feature LDA. The correlation coefficients between the individual features were all less than 0.42, with the smallest correlation coefficient being  $-0.06$  between the depth-to-width ratio and the normalized radial gradient.

When comparing the performances of the strongest two-feature classifier to the strongest three-feature classifier, a statistically significant difference in  $A_z$  values was found in three situations. The first is that of using lesion margins delineated through automatic segmentation and testing the LDA on the entire database. Here the  $A_z$  values of the strongest two-feature and three-feature classifiers are 0.87 and 0.88, respectively, with a p-value of 0.03. The second is that of using lesion margins delineated through manual segmentation and using the LDA classifiers to differentiate malignant lesions from complex cysts. Here the  $A_z$  values of the strongest two-feature and three-feature classifiers are  $0.93 \pm 0.02$  and  $0.95 \pm 0.01$ , respectively, with a p-value of 0.001. The third is that of using lesion margins delineated through manual segmentation and using the LDA classifiers to differentiate malignant from complex cysts and benign tumors. Here the  $A_z$  values of the strongest two-feature and three-feature classifiers are  $0.93 \pm 0.02$  and  $0.95 \pm 0.01$ , respectively, with a p-value of 0.02.

Independent validation was performed 11 times by splitting the entire database randomly into two equal parts, as schematically shown in Figure 4. For each of the 11 independent trials, LDA was used to determine two classifiers: one by merging the four computer-extracted features derived from manually-defined lesions margins and the other by merging the four computer-extracted features derived from the automatically-defined lesion margins. Then, for each of the 11 independent trials, these two classifiers were tested on the validation database, again using features derived from both manually-defined and automatically-defined lesion margins. Table 5 lists the low, high and average  $A_z$  and  $0.9A_z$  values resulting from ROC analysis of each of the 11 independent trials in the task of distinguishing malignant from benign lesions. Shown in Figure 8, for the case of using manually segmented lesions for both training and validation, are the average  $A_z$  values for the first  $n$  trials, where  $n = 2, 3, \dots, 11$ . Also shown are the error bars, which are the  $A_z$  values plus and minus one standard deviation. As Figure 8 demonstrates, the  $A_z$  values plateau after about 8 randomizations, indicating that 11 randomizations are sufficient. Shown in Figure 9 are the average

ROC curves for each training/validation pair. The average ROC curves were obtained by averaging the  $a$  and  $b$  ROC curve parameters [17].

## 4 Summary

In summary, we have developed a CAD method for the classification breast lesions on ultrasound and performed three studies on a database of 400 cases. First, to investigate the marginal benefit of adding a feature to our CAD method, LDA was used to merge combinations of two, three and four of the computer-extracted features. In the task of distinguishing malignant from benign lesions, the best two feature classifier merges the depth-to-width and normalized radial gradient features to yield an  $A_z$  value of 0.90 using manual segmentation and an  $A_z$  values of 0.88 using automatic segmentation. At a p-value cutoff of 0.05, we fail to show a statistically significant difference between the best two-feature classifier and the four-feature classifier.

Second, the performance of our CAD method in distinguishing carcinomas from different types of benign lesions was determined. Our CAD method yielded the best performance in distinguishing carcinomas from complex cysts ( $A_z = 0.95$ , round robin evaluation using automatic segmentation) and the worst performance in distinguishing carcinomas from fibrocystic disease ( $A_z = 0.70$ , round robin evaluation, using automatic segmentation). The four-feature classifier using automatically-delineated lesion margins yielded a high performance in the task of distinguishing carcinomas from complex cysts and benign tumors ( $A_z = 0.92$ , round robin evaluation).

Finally, 11 independent trials were performed on the entire database to obtain validation results. Using computer-extracted features derived from automatically-delineated lesion margins for both the training and validation, a mean  $A_z$  of  $0.87 \pm 0.02$  was obtained.

The results of this study warrant further investigation and in the future, an observer study will be performed to evaluate the potential of our CAD method in improving physician

performance in the task of differentiating malignant from benign breast lesions on ultrasound.

**Acknowledgement** The authors would like to thank Ioana Bonta, MD for useful discussions. This work was supported in parts by grants from the US Army Medical Research and Materiel Command (DAMD 17-98-18194) and USPHS grants RR11459 and T32 CA09649.

M. L. Giger and C. J. Vyborny are shareholders in R2 Technology, Inc. (Los Altos, CA). It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest which would reasonably appear to be directly and significantly affected by the research activities.

## References

- [1] Jackson VP. The role of ultrasound in breast imaging. Radiology, 177:305-311, 1990.
- [2] Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, and Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology, 196:123-134, 1995.
- [3] Giger ML, Al-Hallaq H, Huo Z, Moran C, Wolverton DE, Chan CW, and Zhong W. Computerized analysis of lesions in us images of the breast. Acad Radiol, 6:665-674, 1999.
- [4] Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, and Zeman RK. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. Ultrason Imaging, 15:267-285, 1993.
- [5] Chen DR, Chang RF, and Huang YL. Computer-aided diagnosis applied to us of solid breast nodules by using neural networks. Radiology, 213:407-412, 1999.
- [6] Golub RM, Parsons RE, Sigel B, and et el. Differentiation of breast tumors by ultrasonic tissue characterization. J Ultrasound Med, 12:601-608, 1993.

- [7] Sahiner B, LeCarpentier GL, Chan HP, and et el. Computerized characterization of breast masses using three-dimensional ultrasound images. In Proceedings of the SPIE, volume 3338, pages 301-312, 1998.
- [8] Horsch K, Giger ML, Venta LA, and Vyborny CJ. Automatic segmentation of breast lesions on ultrasound. Submitted to Medical Physics (under review), 2000.
- [9] Tohno E, Cosgrove DO, and Sloane JP. Ultrasound Diagnosis of Breast Disease. Churchill Livingstone, Edinburgh, Scotland, 1994.
- [10] Bick U, Giger ML, Schmidt RA, and Doi K. A new single-image method for computer-aided detection of small mammographic masses. In Prodeedings of CAR '95, pages 357-363, 1995.
- [11] Huo Z, Giger ML, Vyborny CJ, and et el. Analysis of spiculation in the computerized classification of mammographic masses. Med Phys, 22:1569-1579, 1995.
- [12] Lachenbruch PL. Discriminant Analysis. Hafner, London, England, 1975.
- [13] Metz CE. Some practical issues of experimental design and data analysis in radiological roc studies. Invest Radiol, 24:234-245, 1989.
- [14] Jiang Y, Metz CE, and Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. Radiology, 201:745-750, 1996.
- [15] Metz CE. LABROC4. Computer Program.
- [16] Metz CE. CLABROC. Computer Program.
- [17] Metz CE, Herman BA, and Shen JH. Maximum likelihood estimation of receiver operating characteristic (roc) curves from continuously-distributed data. Statistics in Medicine, 17:1033-1053, 1998.

Table 1: The number of each lesion type, as well as size information.

Lesion Type	Number of Cases	Minimum Size	Maximum Size	Average Size
Carcinomas	94	4 mm	37 mm	11 mm
Complex Cysts	124	4 mm	28 mm	9 mm
Benign Solid Lesions (All benign cases except complex cysts)	182	3 mm	28 mm	10 mm
Benign Tumors (Fibroadenomas and Papillomas)	107	4 mm	26 mm	10 mm
Fibroadenomas	100	4 mm	26 mm	10 mm
Papillomas	7	3 mm	19 mm	11 mm
Fibrocystic Disease	65	3 mm	23 mm	10 mm
Inflammation	2	7 mm	8 mm	7 mm
Infection	1	9 mm	9 mm	9 mm
No Abnormality	3	5 mm	14 mm	11 mm
Radial Scar	3	14 mm	28 mm	19 mm
Intramammary Lymph Node	1	5 mm	5 mm	5 mm

Analysis	Data Subset A Malignant Lesions (94) vs. All Benign Lesions (306)				Data Subset B Malignant Lesions (94) vs. Benign Solid Lesions (182)			
	Manual Segmentation		Automatic Segmentation		Manual Segmentation		Automatic Segmentation	
	$A_z$	$(\sigma)$	$0.9A_z$		$A_z$	$(\sigma)$	$0.9A_z$	
DWR	0.84	(0.02)	0.41		0.82	(0.02)	0.40	
NRG	0.76	(0.03)	0.27		0.75	(0.03)	0.34	
COR	0.81	(0.03)	0.42		0.70	(0.03)	0.22	
MSD	0.73	(0.02)	0.20		0.75	(0.03)	0.30	
DWR and NRG	0.73	(0.02)	0.20		0.68	(0.03)	0.14	
DWR, NRG and COR	0.90	(0.02)	0.55		0.88	(0.02)	0.55	
All four features	0.90	(0.02)	0.55		0.88	(0.02)	0.52	
Round Robin: All four features	0.91	(0.02)	0.62		0.89	(0.02)	0.57	
	0.91	(0.02)	0.63		0.88	(0.02)	0.53	
	0.91	(0.02)	0.61		0.87	(0.02)	0.51	
	0.83	(0.02)	0.45		0.83	(0.02)	0.40	
	0.83	(0.02)	0.42		0.83	(0.02)	0.42	
	0.82	(0.02)	0.40		0.82	(0.02)	0.40	

Table 2: Performance in terms of  $A_z$  and  $0.9A_z$  values of the LDA for combinations of the individual computer-extracted features for both manual and automatic segmentation. The standard deviations on the  $A_z$  values are given in parentheses. The LDA classifiers were tested in differentiating malignant lesions from all benign lesions, and in differentiating malignant lesions from benign solid lesions (all benign cases except the complex cysts).

Analysis	Data Subset C Malignant Lesions (94) vs. Complex Cysts (124)				Data Subset D Malignant Lesions (94) vs. Benign Tumors (107)			
	Manual Segmentation		Automatic Segmentation		Manual Segmentation		Automatic Segmentation	
	$A_z$	$(\sigma)$	$0.9A_z$		$A_z$	$(\sigma)$	$0.9A_z$	
DWR	0.83	(0.03)	0.33		0.85	(0.03)	0.29	
NRG	0.86	(0.02)	0.39		0.75	(0.03)	0.24	
COR	0.91	(0.02)	0.62		0.81	(0.03)	0.30	
MSD	0.81	(0.03)	0.30		0.73	(0.03)	0.19	
DWR and NRG	0.93	(0.02)	0.56		0.94	(0.01)	0.66	
DWR, NRG and COR	0.95	(0.01)	0.71		0.93	(0.02)	0.61	
All four features	0.95	(0.01)	0.72		0.94	(0.01)	0.69	
Round Robin: All four features	0.95	(0.01)	0.70		0.93	(0.02)	0.61	

Table 3: Performance in terms of  $A_z$  and  $0.9A_z$  values of the LDA for combinations of the individual computer-extracted features for both manual and automatic segmentation. The standard deviations on the  $A_z$  values are given in parentheses. The LDA classifier were tested in differentiating malignant lesions from complex cysts, and in differentiating malignant lesions from benign tumors.



Analysis	Data Subset E Malignant Lesions (94) vs. Complex Cysts and Benign Tumors (231)				Data Subset F Malignant Lesions (94) vs. Fibrocystic Disease (65)							
	Manual Segmentation		Automatic Segmentation		Manual Segmentation		Automatic Segmentation					
	$A_z$	$(\sigma)$	$0.9A_z$	$A_z$	$(\sigma)$	$0.9A_z$	$A_z$	$(\sigma)$	$0.9A_z$			
DWR	0.86	(0.02)	0.44	0.86	(0.02)	0.38	0.78	(0.04)	0.33	0.72	(0.04)	0.28
NRG	0.81	(0.03)	0.33	0.79	(0.03)	0.39	0.63	(0.04)	0.13	0.62	(0.04)	0.16
COR	0.85	(0.02)	0.49	0.73	(0.03)	0.23	0.72	(0.04)	0.27	0.61	(0.04)	0.17
MSD	0.77	(0.03)	0.25	0.78	(0.03)	0.28	0.61	(0.04)	0.08	0.60	(0.05)	0.08
DWR and NRG	0.93	(0.01)	0.61	0.93	(0.02)	0.59	0.80	(0.04)	0.32	0.74	(0.04)	0.29
DWR, NRG and COR	0.95	(0.01)	0.70	0.92	(0.02)	0.57	0.81	(0.03)	0.38	0.74	(0.04)	0.28
All four features	0.95	(0.01)	0.70	0.93	(0.02)	0.56	0.81	(0.03)	0.38	0.73	(0.04)	0.26
Round Robin: All four features	0.94	(0.01)	0.69	0.92	(0.02)	0.39	0.80	(0.04)	0.36	0.72	(0.04)	0.25

Table 4: Performance in terms of  $A_z$  and  $0.9A_z$  values of the LDA for combinations of the individual computer-extracted features for both manual and automatic segmentation. The standard deviations on the  $A_z$  values are given in parentheses. The LDA classifiers were tested in differentiating malignant lesions from complex cysts and benign tumors, and in differentiating malignant lesions from fibrocystic disease.

Segmentation Used for Training	Segmentation Used for Validation	Low		High		Average	
		$A_z$	$(\sigma)$	$0.9A_z$	$A_z$	$(\sigma)$	$0.9A_z$
Manual	Manual	0.87	0.50	0.94	0.72	$0.91 \pm 0.02$	$0.62 \pm 0.07$
Manual	Automatic	0.85	0.51	0.90	0.70	$0.87 \pm 0.02$	$0.60 \pm 0.05$
Automatic	Manual	0.82	0.28	0.93	0.62	$0.88 \pm 0.04$	$0.47 \pm 0.15$
Automatic	Automatic	0.82	0.36	0.92	0.70	$0.87 \pm 0.02$	$0.52 \pm 0.11$

Table 5: Low, high and average  $A_z$  and  $0.9A_z$  values of the LDA for the 11 independent trials.

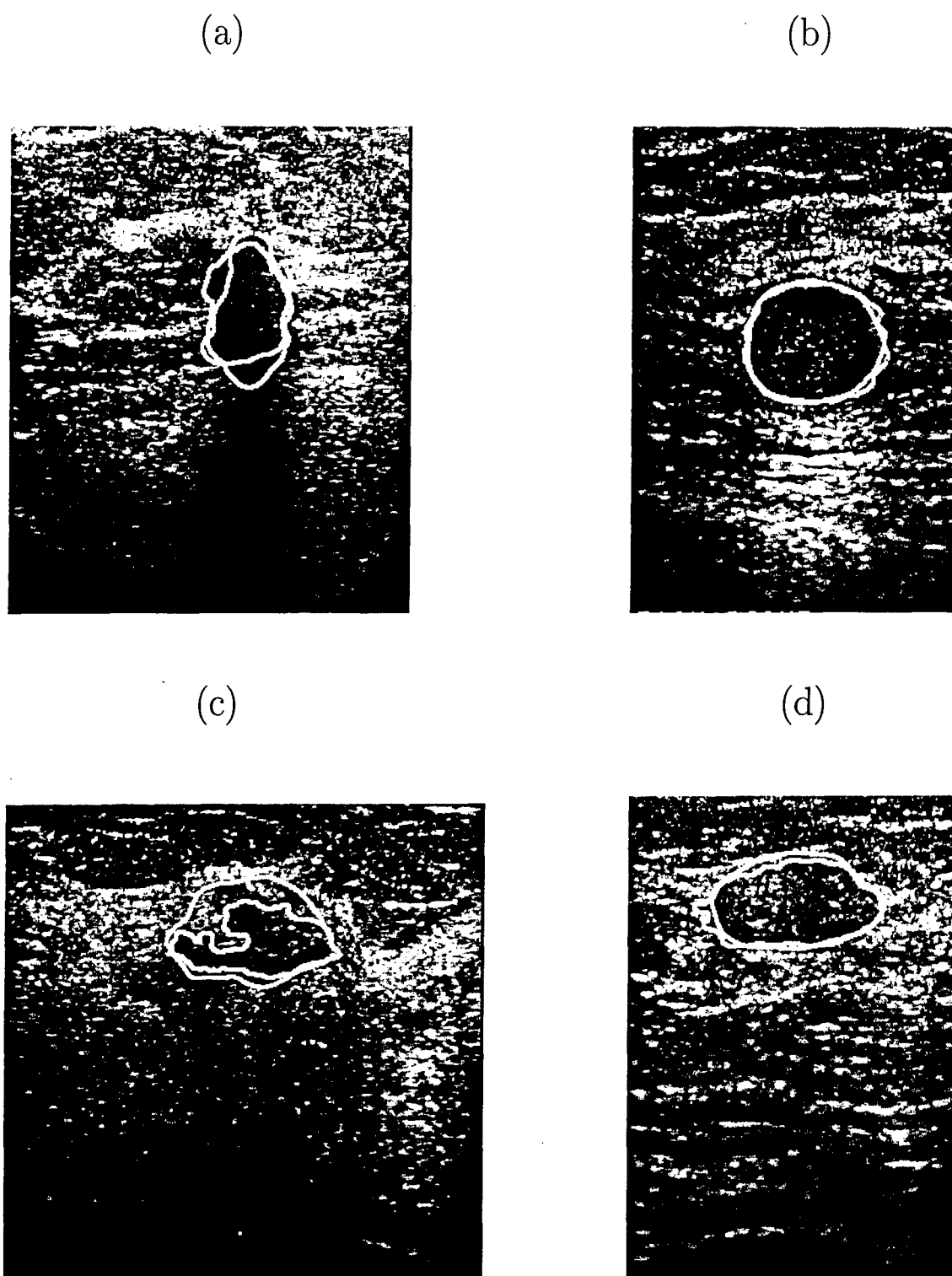


Figure 1: Examples of (a) a malignant lesion, (b) a complex cyst, (c) fibrocystic disease and (d) a fibroadenoma. The manually-delineated margin is given in gray and the computer-delineated margin in white.

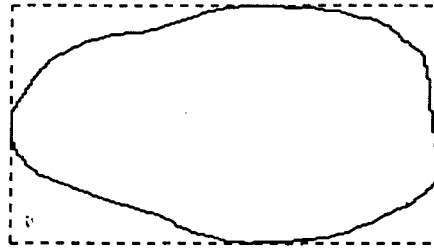


Figure 2: The ROI used to define the autocorrelation feature. The lesion is outlined with a solid line and the ROI is outlined with a dashed line.

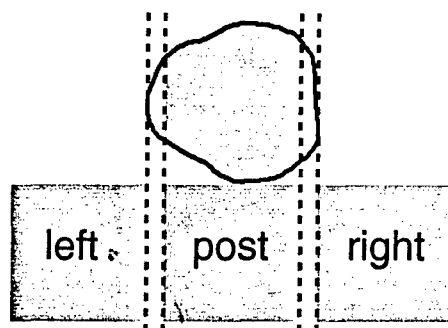


Figure 3: ROIs used to define the posterior acoustic behavior feature.

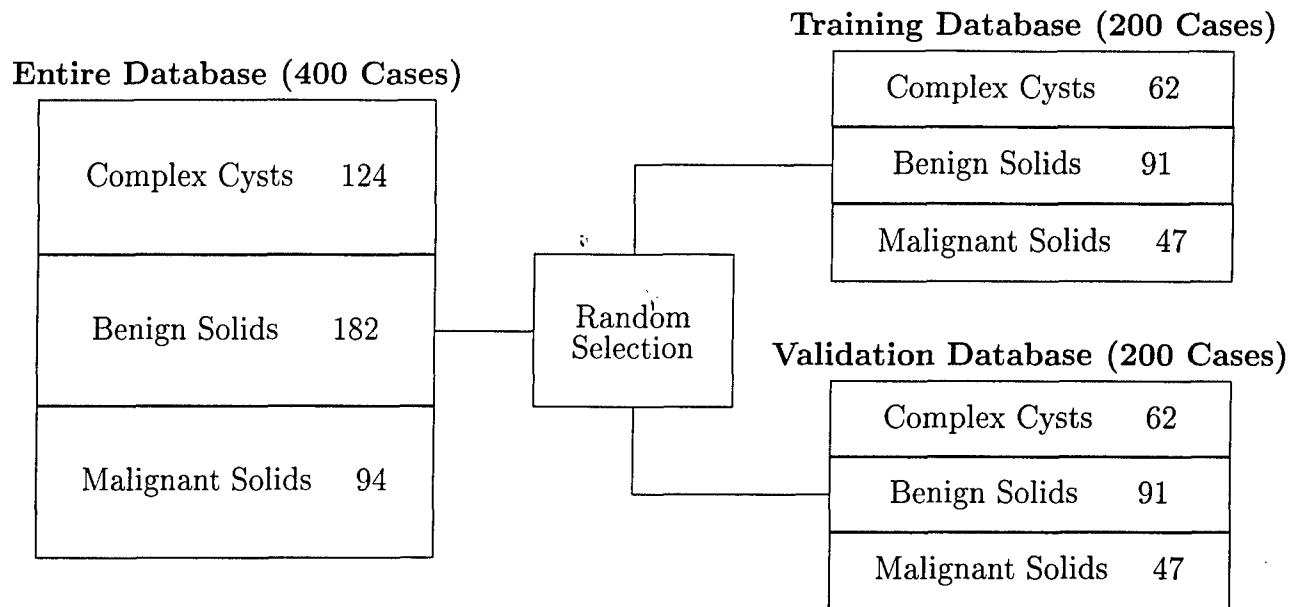


Figure 4: Random selection splitting the entire database in two: the training and validation databases.

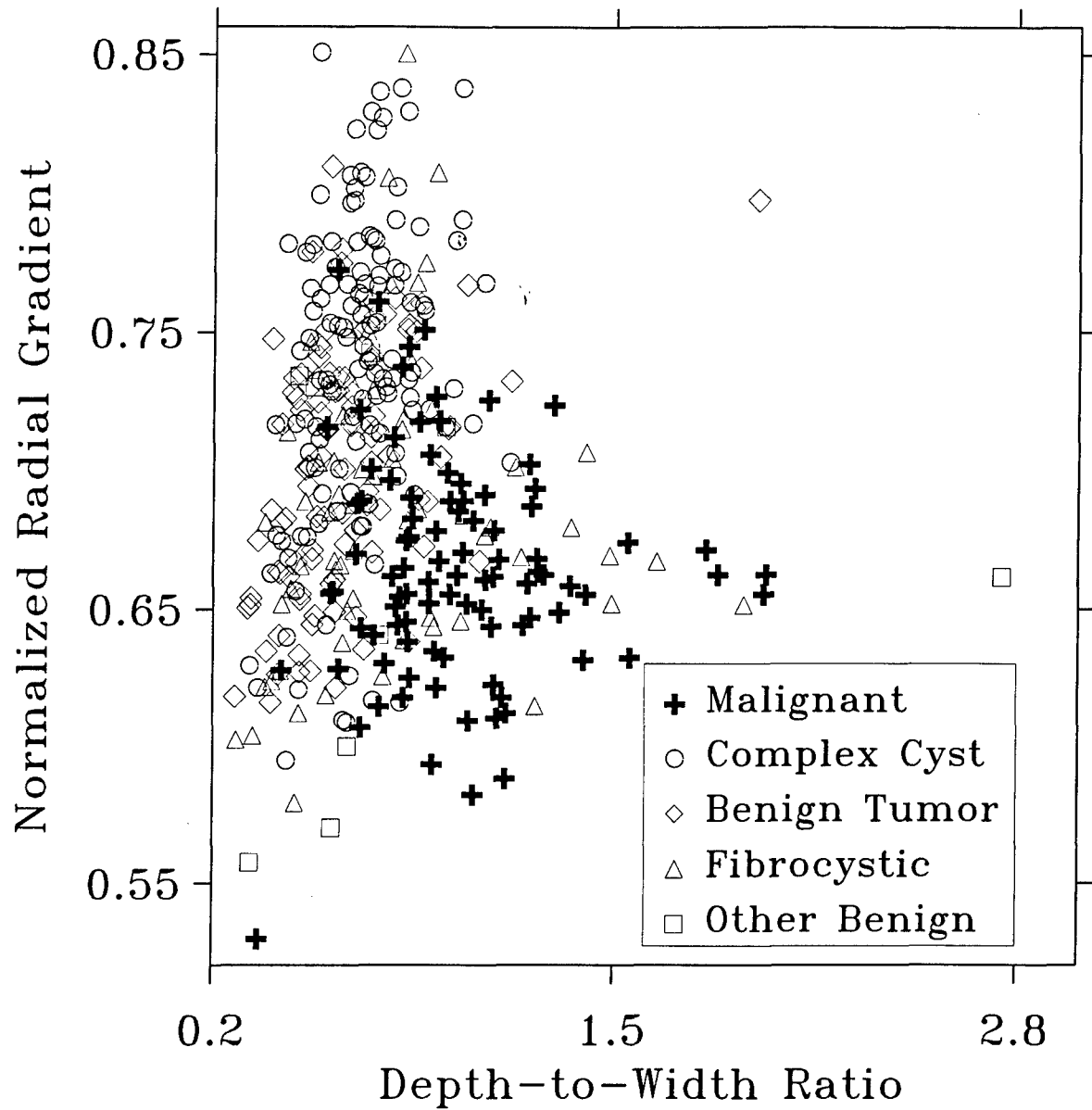


Figure 5: The scatter plots indicate values for the depth-to-width and normalized radial gradient features for the entire database. Margins defined via automatic segmentation were used.

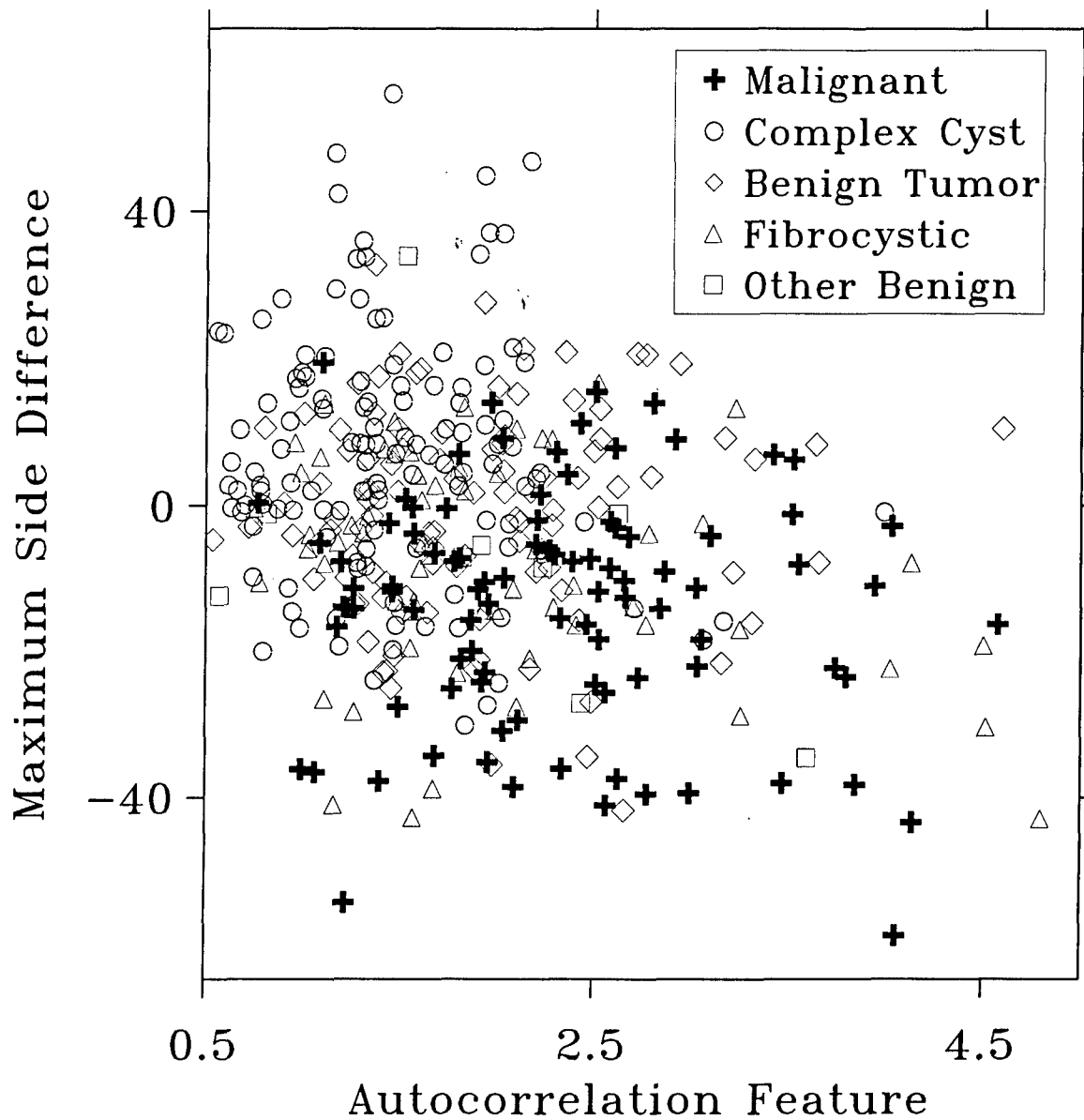


Figure 6: The scatter plots indicate values for the auto-correlation based feature and the minimum side difference for the entire database. Margins defined via automatic segmentation were used.



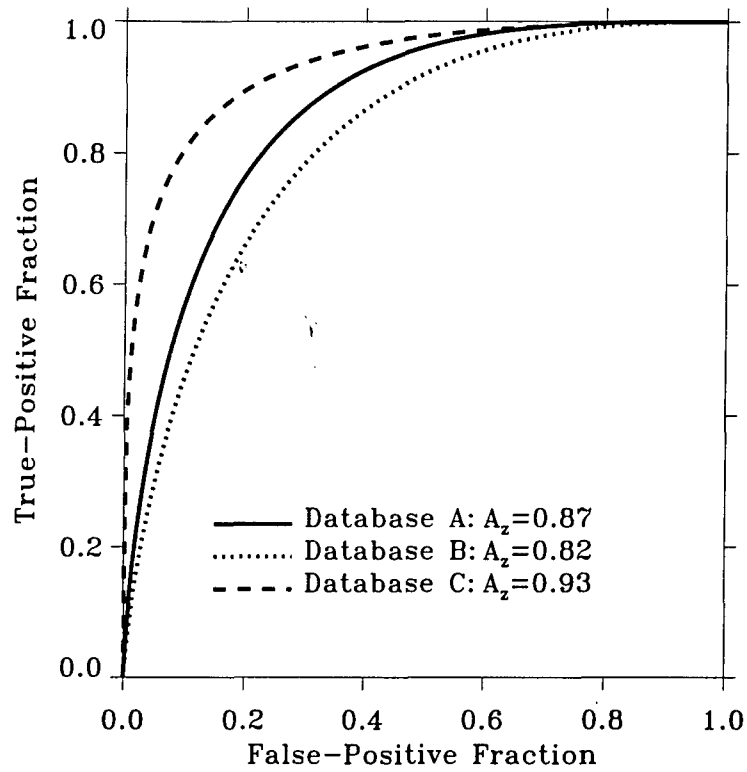


Figure 7: ROC curves of round robin evaluations for data subset A (the entire database), data subset B (carcinomas and benign solid lesions) and data subset C (carcinomas and complex cysts). Margins defined via automatic segmentation were used.

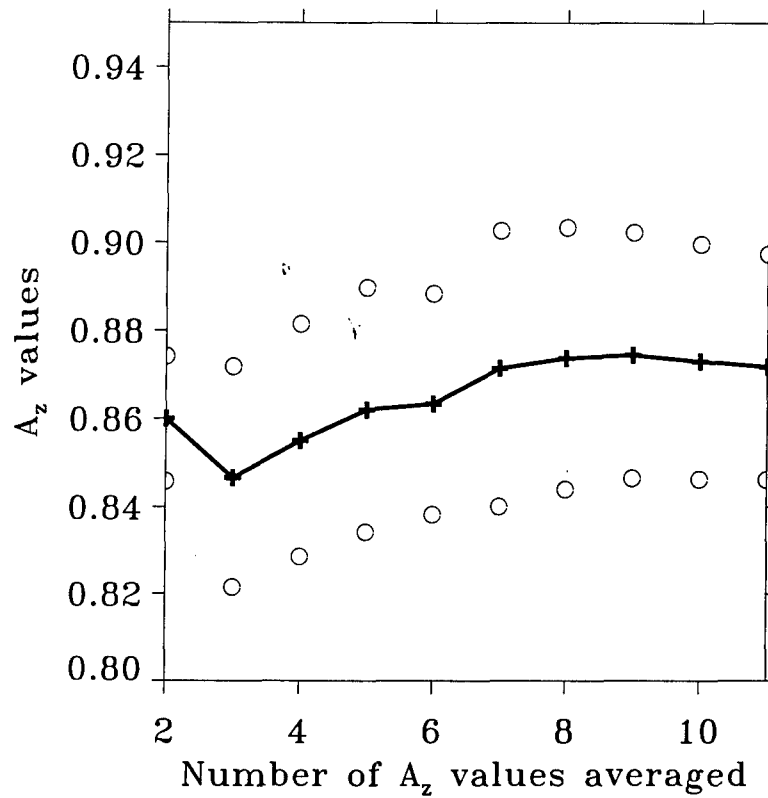


Figure 8: For the manually trained and tested case, the average  $A_z$  values and error bars for the first  $n$  trials where  $n = 2, 3, \dots, 11$ .

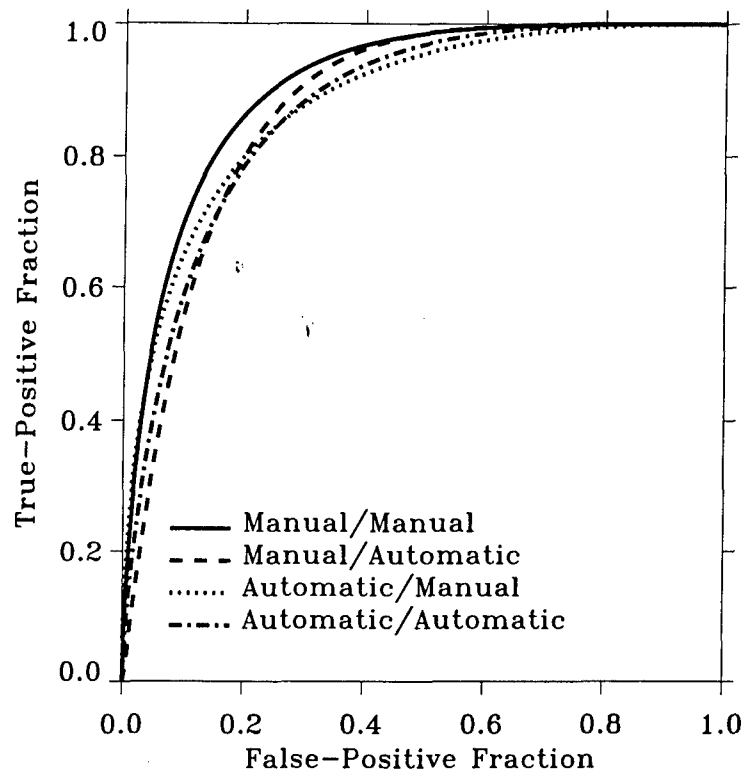


Figure 9: Average ROC curves for the 11 independent trials.

# Automatic Segmentation of Breast Lesions on Ultrasound

Karla Horsch\*, Maryellen L. Giger\*, Luz A. Venta<sup>†‡</sup>, Carl J. Vyborny\*

May 10, 2001

## Abstract

This paper presents a simple and computationally-efficient segmentation algorithm for breast masses on sonography that is based on maximizing a utility function over partition margins defined through gray-value thresholding of a preprocessed image. The performance of the segmentation algorithm is evaluated on a database of 400 cases in two ways. Of the 400 cases, 124 were complex cysts, 182 were benign solid lesions and 94 were malignant lesions. In the first evaluation, the computer-delineated margins were compared to manually-delineated margins. At an overlap threshold of 0.40, the segmentation algorithm correctly delineated 94% of the lesions. In the second evaluation, the performance of our computer-aided diagnosis method on the computer-delineated margins was compared to the performance of our method on the manually-delineated margins. Round robin evaluation yielded  $A_z$  values of 0.90 and 0.87 on the manually-delineated margins and the computer-delineated margins, respectively, in the task of distinguishing between malignant and non-malignant lesions.

**Keywords:** Lesion segmentation, breast sonography, computer-aided diagnosis.

---

\*Department of Radiology, University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637

<sup>†</sup>Currently: Baylor Methodist Breast Care Center, 6550 Fannin St., suite 701, Houston, TX 77030

<sup>‡</sup>Formally: Lynn Sage Breast Center, Northwestern University, Chicago, IL 60611

# 1 Introduction

Ultrasound is currently used to diagnosis simple cysts of the breast with a reported accuracy of 96-100% [1]. However, due to the large overlap in the sonographic appearance of malignant and benign solid lesions, most radiologists feel uncomfortable relying on ultrasound to differentiate solid masses. This results in the utilization of biopsy procedures for most solid breast lesions interpreted with sonography. In a recent study, Stavros et al [2] developed a classification scheme which used various sonographic features identified by radiologists to achieve a sensitivity of 98.4% and a negative predictive value of 99.5% on a data set of 750 solid breast masses. The use of specific sonographic feature hold the potential for accurate classification of solid breast masses using ultrasound, thereby allowing a decrease in the number of biopsies performed for benign solid lesions. The identification of sonographic features can also be potentially automated.

Computer-aided diagnosis (CAD) methods on breast ultrasound are being explored by various researchers [3, 4, 5, 6, 7]. Lesion segmentation is often an important step in computer-aided diagnosis schemes. In this paper, we propose a simple and computationally efficient segmentation algorithm for breast sonography that is based on maximizing a utility function over partition margins defined through gray-value thresholding of a preprocessed image. The key step in the image processing involves multiplication by a constraint function whose level surfaces are ellipses. When gray-value thresholding is applied to an image so processed, the result is potential lesion margins (partition margins) that are deformations of ellipses, or "lesion-like". A gradient-based utility function is then used to choose the lesion margin from the potential margins. The performance of this algorithm is evaluated on a large database in two ways: (1) by comparing computer-delineated margins to manually-delineated margins, and (2) by comparing the performance of our CAD scheme on the computer-delineated margins to the performance of our CAD scheme on the manually-delineated margins.

## 2 Material and Methods

### 2.1 Database

Our database consists of 400 consecutive ultrasound cases, being represented by 757 images. These images were acquired during diagnostic breast exams at the Lynn Sage Breast Center of Northwestern Memorial Hospital. The cases were collected retrospectively and all had been either biopsied or aspirated. Of the 400 cases, 124 were complex cysts, 182 were benign solid lesions and 94 were malignant solid lesions. Note that the database does not contain any simple cysts. The images were obtained with an ATL 3000 unit and were captured directly from the 8-bit video signal. The number of images per cases varied from one to six. Size information for each of the lesion types is given in Table 1.

Table 1: Size information for complex cysts, benign solid lesions and malignant lesions.

Lesion Type	Minimum Size	Maximum Size	Average Size
Complex Cysts	4 mm	28 mm	9 mm
Benign Solid Lesions	3 mm	28 mm	10 mm
Malignant Lesions	4 mm	37 mm	11 mm

### 2.2 Lesion Segmentation

In each image, lesions were both manually and automatically segmented from normal breast tissue.

Manual segmentation involved displaying each ultrasound image on an IBM PowerDisplay20 monitor (Armonk, NY), and having a mammographer or medical physicist outline the lesion margin using software designed for that purpose. The geometric centers of these manually outlined lesions are then used as input to the automatic segmentation algorithm.

The automatic lesion segmentation algorithm involves (1) preprocessing by cropping and

median filtering, (2) multiplication with a Gaussian constraint function, (3) determination of potential lesion margins through gray-value thresholding, and (4) maximization of a utility function on the potential lesion margins.

The segmentation method begins with preprocessing the image. The subcutaneous fat is removed by cropping the top of the image by 35 pixels. We found that for this database, cropping in this manner was sufficient for our purpose. In the future, instead of cropping each image by a fixed number of pixels, the edge of the subcutaneous fat could be detected in each image and used to estimate the appropriate number of pixels to crop. After removing the subcutaneous fat, a 10 by 10 median filter is used to suppress the ultrasound speckle. An example of a preprocessed image is shown in Figure 1b.

The next step involves multiplying by a constraint function centered on the lesion center. Kupinski and Giger used this method for lesions in mammography with the effect of suppressing distant pixel values and encouraging potential lesion margins to be more "lesion-like" [8]. A similar technique may be applied to ultrasound images by inverting the gray-scale of the preprocessed image before multiplying by a constraint function. If  $C$  is the constraint function, then the resulting image is

$$J(\hat{P}) = C(\hat{P}) * \left( 1 - \frac{\tilde{I}(\hat{P})}{\max_{\hat{P}}(\tilde{I}(\hat{P}))} \right), \quad (1)$$

where  $\hat{P}$  is the pixel location. Inverting the image changes the lesion from dark (low gray values) to light (high gray values). The constraint function should have higher gray values in the region of the lesion and gray values near zero far from the lesion. An example of an inverted image is shown in Figure 1c. In this study, a Gaussian was used as the constraint function. The Gaussian is centered at the manually defined lesion center,  $\hat{\mu}$ :

$$C(\hat{P}) \equiv N(\hat{P}; \hat{\mu}, \hat{\sigma}) = \frac{\exp(-\frac{1}{2}(\hat{P} - \hat{\mu})^T K^{-1}(\hat{P} - \hat{\mu}))}{2\pi\sqrt{\det K}}. \quad (2)$$

Here the covariance matrix is assumed diagonal,

$$K = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}, \quad (3)$$

where  $\sigma_x^2$  and  $\sigma_y^2$  are the variances in the lateral and depth directions, respectively. These variances are chosen as

$$\sigma_x = \frac{w}{2}, \quad \sigma_y = \frac{h}{2}, \quad (4)$$

with  $w$  being the estimated lesion width and  $h$  being the estimated lesion height (or depth). An example of the preprocessed image multiplied by a Gaussian constraint function is shown in Figure 1d.

In order to study the sensitivity of the segmentation algorithm on the choice of variance, both manual and automatic width and height estimation were performed. In this paper, the segmentation algorithms using manually and automatically-estimated lesion width and height are referred to as partially-automatic and fully-automatic, respectively.

In the *partially-automatic segmentation*, manual estimation of the lesion width and height is achieved using the manually-delineated lesion margin. If  $\gamma(i) = (\gamma_1(i), \gamma_2(i))$  is a discrete parameterization of the manually-delineated margin with  $\gamma_1$  and  $\gamma_2$  being the coordinates in the lateral and depth directions, respectively, then we define

$$w_{\text{manual}} = \max_i(\gamma_1(i)) - \min_i(\gamma_1(i)), \quad (5)$$

$$h_{\text{manual}} = \max_i(\gamma_2(i)) - \min_i(\gamma_2(i)). \quad (6)$$

In the *fully-automatic lesion segmentation*, estimations of the lesion width and height are determined through Sobel edge detection. The Sobel filtered images are defined by

$$\tilde{I}_x = F_x * \tilde{I},$$

$$\tilde{I}_y = F_y * \tilde{I},$$



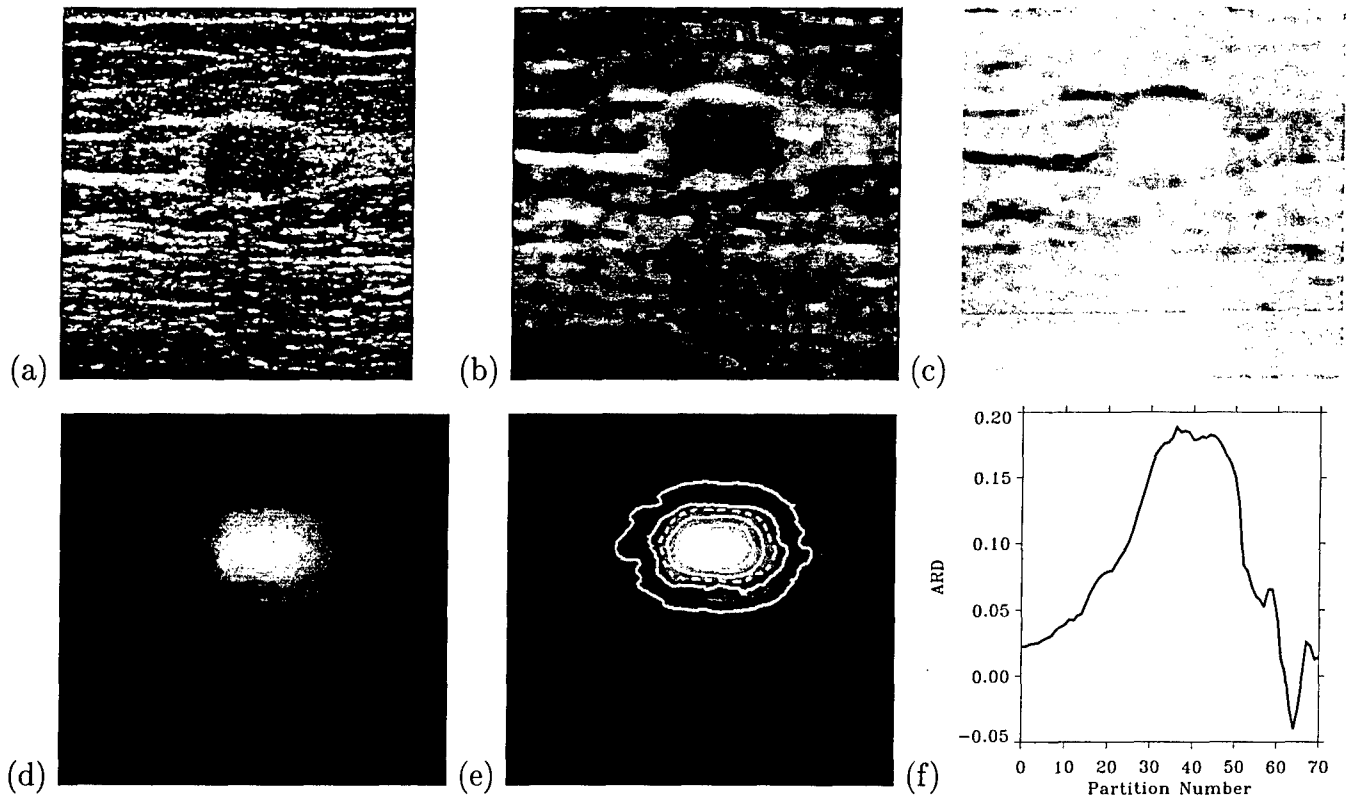


Figure 1: The results of the segmentation processing steps on an example image: (a) the original image, (b) the preprocessed image (cropped and median filtered), (c) the inverted preprocessed image, (d) the inverted preprocessed image multiplied by a Gaussian, (e) the partitions resulting from gray-value thresholding, and (f) the ARD as a function of partition number. In (f), the smaller the partition number, the smaller the area enclosed by the partition. For this particular example, the computer-chosen partition is number 35 and is shown as a dashed line on the image in (e).

where  $\tilde{I}$  is the preprocessed image,  $*$  is the convolution operator, and  $F_x$  and  $F_y$  are 3 by 3 Sobel filters in the lateral and depth directions, respectively,

$$F_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \quad F_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}. \quad (7)$$

Estimations of the location of the lesion edge along horizontal and vertical lines through the lesion center are given by

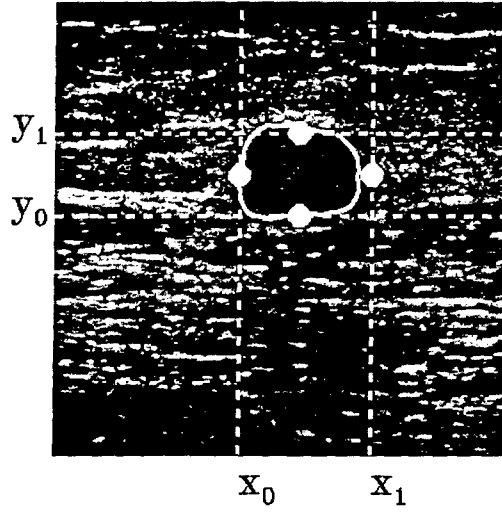


Figure 2: Automatic estimations of edge points. The manually-delineated lesion is shown for reference.

$$\begin{aligned}
 x_0 &= \arg \left( \min_{i \in [1, \mu_x]} \tilde{I}_x(i, \mu_y) \right), \\
 x_1 &= \arg \left( \max_{i \in [\mu_x, N_x]} \tilde{I}_x(i, \mu_y) \right), \\
 y_0 &= \arg \left( \min_{i \in [1, \mu_y]} \tilde{I}_y(\mu_x, i) \right), \\
 y_1 &= \arg \left( \max_{i \in [\mu_y, N_y]} \tilde{I}_y(\mu_x, i) \right).
 \end{aligned}$$

An example of these locations is shown in Figure 2. The estimated location of lesion edges are then used to estimate the lesion width and height by

$$w_{automatic} = 2 * \min(\mu_x - x_0, x_1 - \mu_x), \quad (8)$$

$$h_{automatic} = 2 * \min(\mu_y - y_0, y_1 - \mu_y). \quad (9)$$

Note that for the width, instead of using the length between the left and right edges, we use twice the minimum of the lengths between the lesion center and the left and right edges. This is done to avoid the overestimation which may result when distant pixels are mistaken

for the lesion edge. Similar comments apply to the automatic lesion height estimation. The lesion segmentation which results from using such estimations will err on the side of “under growing” rather than “over growing”.

When  $w_{automatic}$  and  $h_{automatic}$  are used in equation (2), the lesion center is the only information defined manually that is needed by the segmentation algorithm.

We emphasize that the variances in the width and depth directions for the Gaussian constraint function are varied adaptively and automatically *for each image*. This differs from the study done by Kupinski and Giger, in which a single variance is used.

After applying the Gaussian constraint function to the inverted preprocessed image, gray-value thresholding defines partitions whose margins are potential lesion margins. The potential margin that maximizes the utility function on the preprocessed image then defines the lesion margin. The utility function used in our segmentation algorithm is the Average Radial Derivative (ARD), which gives the average directional derivative in the radial direction along the margin,

$$ARD(\Gamma) = \frac{1}{N} \sum_{\hat{P} \in \Gamma} \nabla \tilde{I}(\hat{P}) \cdot \hat{r}(\hat{P}), \quad (10)$$

where  $\Gamma$  is the discretized potential lesion margin,  $N$  is the number of points in  $\Gamma$ ,  $\hat{r}(\hat{P})$  is the unit vector in the radial direction from the geometric center of the partition to the point  $\hat{P} = (x, y)$ , and  $\cdot$  is the dot product between vectors. An example of potential lesion margins resulting from gray-value thresholding and an example of the ARD as a function of partition number are shown in Figures 1e and 1f. Note that this utility function differs from that used by Kupinski and Giger for mammographic lesions. Their technique, based on a utility function called the Normalized Radial Gradient, evaluates the average orientation of the gray level gradients along the margin [8]. The Normalized Radial Gradient is used elsewhere in this paper (see Equation 12).

Manual, partially-automatic and fully-automatic segmentation were performed on each ultrasound image in the database. Examples of each type of segmentation are shown in

Figure 3.

We observe that on average, both the partially-automatic and the fully-automatic segmentation algorithms tend to result in smaller lesions than those defined manually. One reason for this may be that radiologists seem to “overdraw” lesion margins. This remark is based on observing many radiologists outline lesions. In addition, the lesions segmented by the fully-automatic algorithm tend to be smaller than those segmented by the partially-automatic algorithm. This is in part because the lesion height and width estimations for fully-automatic segmentation tend to be less than those for partially-automatic segmentation. The fully-automatic width estimation is twice the *minimum* of the lengths from the lesion center to the left and right lesion edges (see Equation 9). The partially-automatic width estimation is the *maximum* horizontal length in the manually outlined margin (see Equation 6). Similar definitions apply to the height estimations.

## 2.3 Performance Evaluation

The performance of the segmentation algorithm can be assessed by comparing the computer-delineated outlines against the the outlines drawn by human observers. For a particular lesion, the overlap,  $O$ , between the computer-segmentation and the manual-segmentation is given by

$$O = \frac{\text{Area}(\mathcal{M} \cap \mathcal{C})}{\text{Area}(\mathcal{M} \cup \mathcal{C})}, \quad (11)$$

where  $\mathcal{M}$  is the set of points in the manually-segmented lesion and  $\mathcal{C}$  the set of points in the computer-segmented lesion (either partially or fully-automatic). The overlap ranges between zero and one, being zero in the case of no overlap and one in the case of exact overlap. To study the overlap for the entire database, overlap thresholds are set. At each threshold, the number of lesions “correctly” segmented is given by the number of lesions with  $O$  greater than the threshold.

Ultimately, we are concerned with computer-aided diagnosis. Therefore, the performance

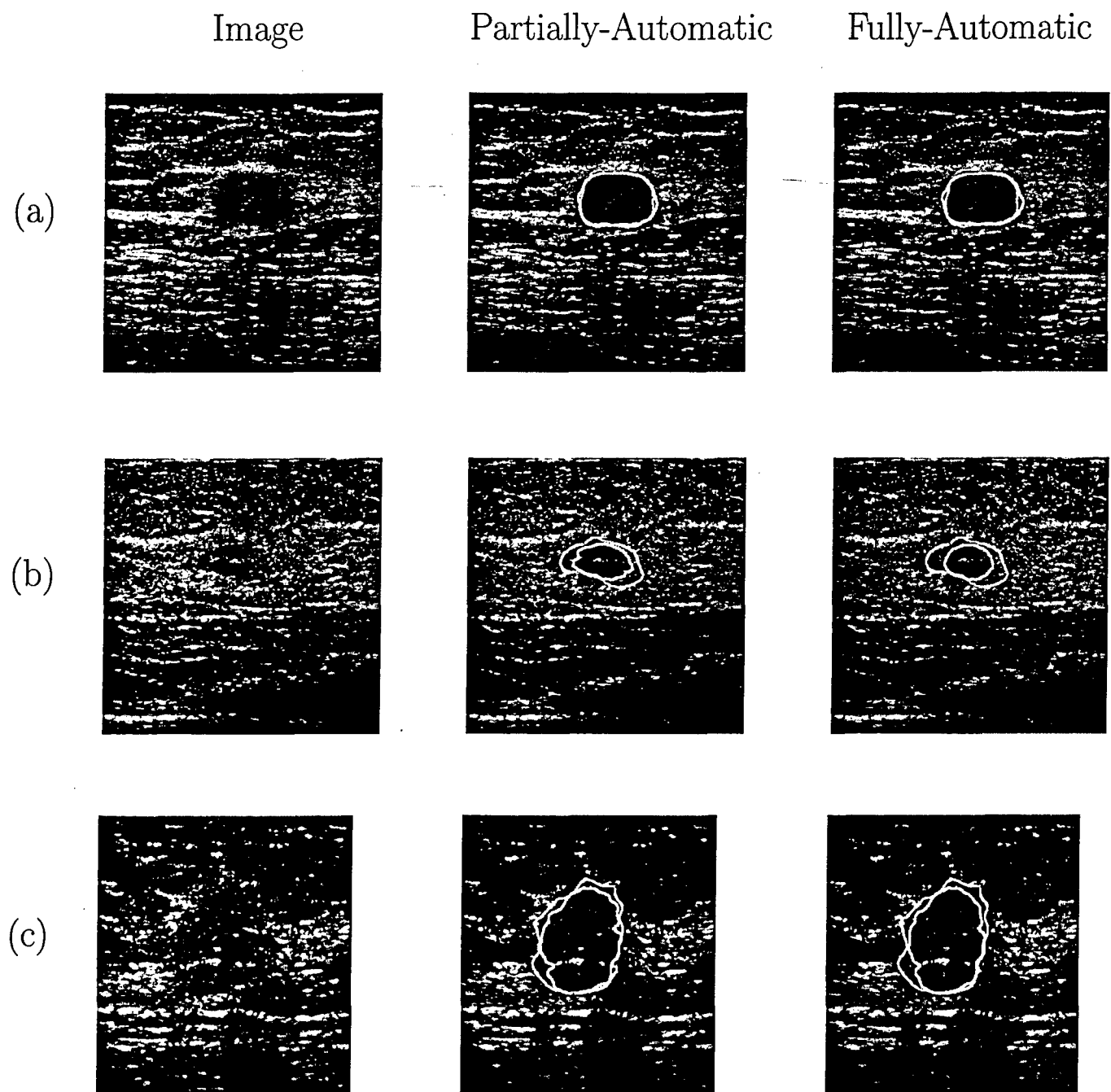


Figure 3: Examples of both the partially and fully-automatic segmentation results. The manually-delineated margin is given in gray and the computer-delineated margin in white. A complex cyst (a), a benign solid (b) and a malignant solid (c).

of the segmentation algorithm can be measured through the performance of our automatic classifier. The classifier uses linear discriminant analysis (LDA) to merge four computer-extracted features. The classifier and the extracted features are described in the next section and the appendix. Receiver operating characteristic (ROC) analysis [9] was used to evaluate (by case, not by image) the performance of the individual computer-extracted features and the LDA classifier in the task of distinguishing benign from malignant lesions. The  $A_z$  [9] and partial  $A_z$  values [10] are used as indicators of merit. The  $A_z$  value is the area under the ROC curve and the partial  $A_z$  value is the area under the ROC curve but above the 0.90 sensitivity line [9], [10].

## 2.4 Automatic Feature Extraction

Features that radiologists use clinically in the evaluation of breast masses on sonograms include margin definition, echogenic texture, posterior acoustic enhancement or shadowing and lesion shape [11]. Benign lesions tend to demonstrate well-defined, smooth margins, posterior acoustic enhancement, and a lesion shape that is wider rather than taller. Benign solid lesions can be hypoechoic or hyperechoic. Malignant lesions, on the other hand, tend to demonstrate ill-defined, angular and irregular margins, marked hypoechogenicity, and posterior acoustic shadowing.

Four characteristics were studied here: margin, echogenicity, posterior acoustic behavior and shape. These were automatically quantified using the normalized radial gradient [12, 13], the autocorrelation, a comparison of gray levels and the depth-to-width ratio. These are briefly described in the appendix and discussed in detail elsewhere [14].

The computer-extracted features were computed for each image and for each of the segmentation methods described earlier: manual, partially-automatic and fully-automatic. A particular feature value for a given lesion (case) and segmentation was taken to be the average of that feature over all the views available for the lesion, each lesion being represented by one to six images. Linear discriminant analysis (LDA) [15] was used to merge the

computer-extracted margin, echogenicity, posterior acoustic behavior and shape features. Both consistency and round-robin runs were performed. In a consistency LDA [15], each lesion is classified according to a classifier trained with *all of the lesions*. In a round robin LDA [15], one of the lesions is removed from the data and that lesion is classified according to a classifier trained with *the remaining lesions*. This process is then repeated for each lesion.

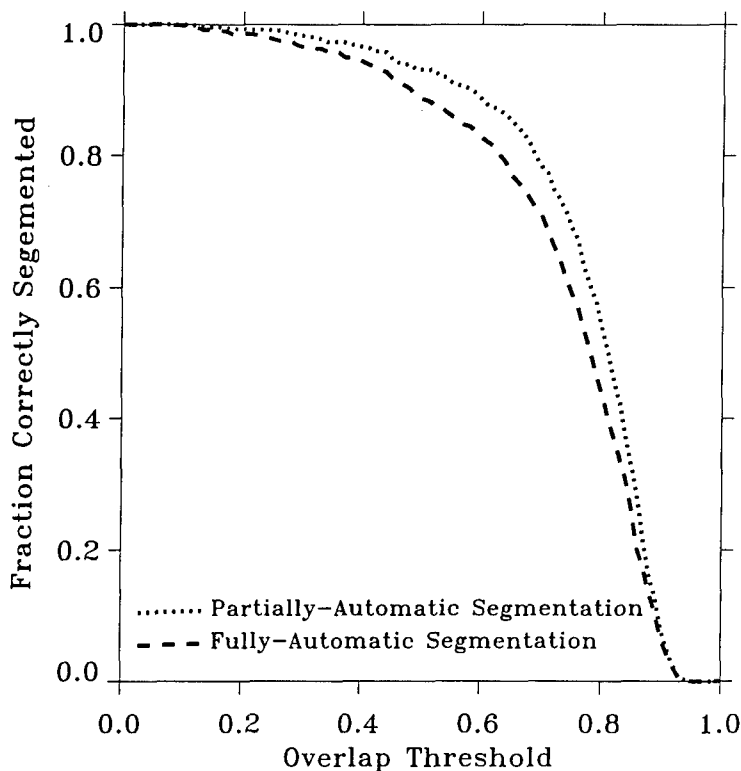


Figure 4: The overlap results for the partially and fully-automatic segmentation on the entire database (757 images).

### 3 Results and Discussion

#### 3.1 Segmentation Overlap

The overlap results for the entire database are shown in Figure 4. The overlap of the lesions defined through partially-automatic segmentation was slightly greater than the over-

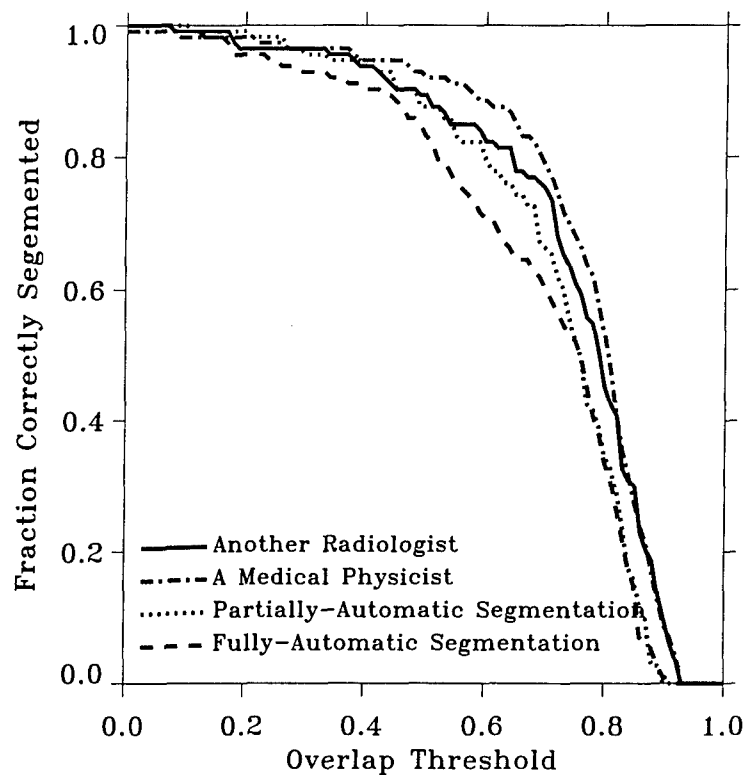


Figure 5: The overlap results on 113 images for the segmentation methods in addition to the overlap results of another radiologist's outlines and a medical physicist's outlines.

lap of those defined through fully-automatic segmentation. This is not surprising as the partially-automatic segmentation algorithm uses height and width estimations derived from the manual margin outlines and so their overlap with the manually segmented lesions should be greater. At a overlap threshold of 0.4, the fraction of images "correctly" segmented is 0.97 for the partially-automatic method, and 0.94 for the fully-automatic method. The mean overlap for the partially-automatic segmentation method is 0.77 and for the fully-automatic segmentation method is 0.73. A paired t-test for the overlap measures yields a p-value of less than 0.0001.

In order to study variability in margin definition between different human observers, 113 images representing 55 cases were outlined also by a second radiologist and a medical physicist. Figure 5 shows the overlap results for these 113 images. Table 2 gives the p-values resulting from paired t-tests for means on the overlap measures from the various segmentation methods. For example, the null hypothesis of the first row is that the overlap



Table 2: The p-values resulting from paired t-tests for means on the overlap measures from the various segmentation methods.

Segmentation Method	Mean	Segmentation Method	Mean	p-value
Second Radiologist	0.74	Medial Physicist	0.76	0.18
Partially-Automatic	0.71	Fully-Automatic	0.68	0.014
Partially-Automatic	0.71	Second Radiologist	0.74	0.18
Partially-Automatic	0.71	Medical Physicist	0.76	0.016
Fully-Automatic	0.68	Second Radiologist	0.74	0.0041
Fully-Automatic	0.68	Medical Physicist	0.76	0.0003

sample means of both data sets (the overlap for the 113 images from the second radiologist and those from the medical physicist) are equal. A p-value less than 0.05 is commonly used as the cutoff indicating a statistically significant difference. The overlap of the medical physicist and the second radiologist are similar (p-value  $> 0.05$ , indicating a failure to show a statistically significant difference), showing that the variability in margin definition between two radiologists is similar to the variability between a radiologist and a medical physicist. This provides some justification for using the medical physicist's outlines for part of the database. In general, the overlap of the lesions defined by another human observer is similar to the overlap of the lesions defined by partially-automatic segmentation and slightly better than the overlap of lesions defined by fully-automatic segmentation.

### 3.2 Segmentation and Computer-Aided Diagnosis

It is important to consider how changes in segmentation affect the performance of individual computer-extracted features in the task of differentiating benign and malignant lesions. The  $A_z$  values for the various computer-extracted features and for the LDA consistency and round robin runs are given in Table 3. In Table 4 are shown the p-values associated with univariate z-score tests of the differences in the  $A_z$  and partial  $A_z$  values for each individual feature, as well as for the LDA consistency and round robin runs. The null hypothesis assumes that

the data in different columns arose from binomial ROC curves having equal  $A_z$  values.

For the individual computer-extracted features, the  $A_z$  values ranged from 0.70 to 0.86 in distinguishing benign from malignant lesions. The depth-to-width ratio (DWR) performed best in this study, with an  $A_z$  of 0.84 for the manual segmentation, 0.86 for the partially-automatic segmentation, and 0.82 for the fully-automatic segmentation. At a significance level of 0.05, we see a statistically significant difference in the  $A_z$  value of the depth-to-width feature when changing the segmentation from manual to partially-automatic (from  $A_z = .84$  to  $A_z = .86$ ) or from partially-automatic to fully-automatic (from  $A_z = .86$  to  $A_z = .82$ ). The correlation feature performs well in the manual segmentation case, with an  $A_z$  of 0.81. The performance in the partially and fully-automatic cases is not as high, with  $A_z$  values of 0.74 and 0.70, respectively. This difference in performance of the correlation feature when changing the segmentation from manual to either partially-automatic or fully-automatic is statistically significant. Both the normalized radial gradient and the maximum side difference perform similarly with all three segmentation methods. This indicates that, for our database, these features are robust to small changes in segmentation.

Table 3: Performance in terms of  $A_z$  and  $0.9A_z$  values of individual computer-extracted features as well as the LDA for manual, partially-automatic and fully-automatic segmentation.

Analysis	Manual		Partially-Automatic		Fully-Automatic	
	$A_z$	$0.9A_z$	$A_z$	$0.9A_z$	$A_z$	$0.9A_z$
Depth-to-Width Ratio	0.84	0.41	0.86	0.47	0.82	0.40
Autocorrelation Based	0.81	0.42	0.74	0.28	0.70	0.22
Normalized Radial Gradient	0.76	0.27	0.76	0.34	0.75	0.34
Maximum Side Difference	0.73	0.20	0.73	0.23	0.74	0.24
LDA Consistency	0.91	0.63	0.90	0.61	0.88	0.53
LDA Round Robin	0.91	0.61	0.89	0.58	0.87	0.51

For the LDA consistency and round robin runs, the change in the  $A_z$  and partial  $A_z$  are statistically significant when changing from either manual segmentation or partially-

Table 4: P-values for Table 3.

Analysis	Manual vs. Partially-Automatic		Manual vs. Fully-Automatic		Partially-Automatic vs. Fully Automatic	
	$A_z$	$0.9A_z$	$A_z$	$0.9A_z$	$A_z$	$0.9A_z$
Depth-to-Width Ratio	0.006	0.035	0.47	0.26	0.033	0.036
Autocorrelation Based	<0.0001	0.01	<0.0001	0.0015	0.11	0.87
Normalized Radial Gradient	0.80	0.39	0.43	0.042	0.42	0.11
Maximum Side Difference	0.49	0.92	0.70	0.61	0.99	0.99
LDA Consistency	0.21	0.14	0.0017	0.0045	.0053	.033
LDA Round Robin	0.21	0.18	0.0034	0.0066	.0094	.030

automatic segmentation to fully-automatic segmentation. However, we fail to demonstrate a statistical difference between the manual and partially-automatic cases indicating that, given good estimates of the lesion width and height, our segmentation algorithm performs as well as manual segmentation in conjunction with our automatic classifier. Figure 6 shows the performance of the LDA in terms of ROC curves for each type of segmentation in the task of distinguishing malignant from benign lesions.

## 4 Summary

We have developed and tested a segmentation method for breast lesions on ultrasound. One of the advantages of the method is that it tends to produce margins that are “lesion-like”. On the other hand, the segmented margins delineate the general shape of the lesions and may not depict margin details such as spiculation or high irregularity [8]. However, segmentation of the general lesion shape appears sufficient for the features chosen in our experiment, as indicated by the performance of our classifier on the lesions segmented with the partially-automatic method.

In conclusion, our automatic classifier yielded  $A_z$  values of 0.91 and 0.87 in distinguishing malignant from benign lesions when using fully-automatic segmentation and manual segmen-

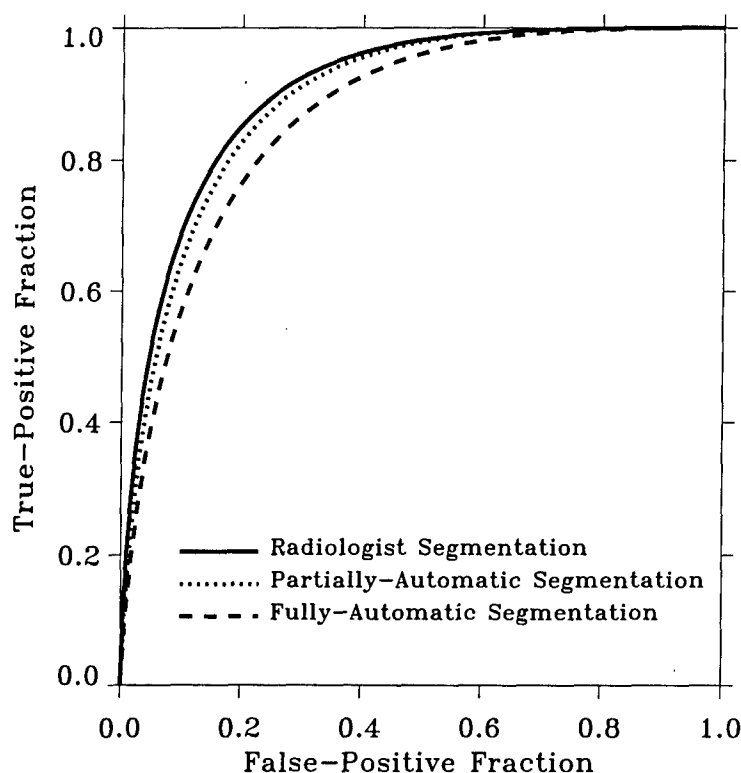


Figure 6: Performance of discriminant scores in distinguishing malignant from benign lesions for manual, partially-automatic and fully-automatic segmentations. Results are from round-robin analyses.

tation, respectively. Our results indicate that when used in conjunction with our automatic classifier, our automatic segmentation algorithm of breast lesions on ultrasound performs similarly to manual segmentation.

**Acknowledgement** This work was supported in parts by grants from the US Army Medical Research and Materiel Command (DAMD 17-98-18194) and USPHS grants RR11459 and T32 CA09649.

M. L. Giger and C. J. Vyborny are shareholders in R2 Technology, Inc. (Los Altos, CA). It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest which would reasonably appear to be directly and significantly affected by the research activities.

## 5 Appendix

To quantify the margin, we consider the normalized radial gradient (NRG) [12, 13], which is a measure of the average orientation of the gray level gradients along the margin. It is given by

$$NRG = \frac{\sum_{\hat{P} \in \Gamma} \nabla I(\hat{P}) \cdot r(\hat{P})}{\sum_{\hat{P} \in \Gamma} \|\nabla I(\hat{P})\|}. \quad (12)$$

where  $I$  is the image,  $P = (x, y)$  is the pixel location,  $\Gamma$  is the discretized lesion margin,  $\hat{r}(\hat{P})$  is the unit vector in the radial direction from the geometric center of the lesion to the point  $\hat{P}$ , and  $\cdot$  is the dot product between vectors. For ultrasound, the NRG is bound between zero and one.

The posterior acoustic behavior is quantified by comparing the gray-level values posterior to the lesion to those in adjacent tissue at the same depth. Define  $A_p$  as the average gray-level of a region of interest (ROI) posterior to the lesion. Similarly, let  $A_l$  be the average gray-level of the ROI to the left of the lesion at the same depth and  $A_r$ , the average gray-level of the ROI too the right of the lesion at the same depth. Then the minimum side difference (MSD) is

$$MSD = \min(A_p - A_l, A_p - A_r). \quad (13)$$

The shape feature that we consider is the depth-to-width ratio of the lesion [3]. Let  $\gamma(i) = (\gamma_1(i), \gamma_2(i))$  be a discrete parameterization of the margin with  $\gamma_1$  and  $\gamma_2$  the coordinates in the lateral and depth directions respectively. Then

$$DWR = \frac{\text{Depth}}{\text{Width}} = \frac{\max_i(\gamma_2(i)) - \min_i(\gamma_2(i))}{\max_i(\gamma_1(i)) - \min_i(\gamma_1(i))}. \quad (14)$$

To quantify texture, the autocorrelation in depth of the minimal rectangular ROI  $R$  containing the lesion is used to define

$$COR = \sum_n \frac{\overline{C_y(n)}}{\overline{C_y(0)}}, \quad (15)$$

where the autocorrelation in depth and its sum in the lateral direction are

$$C_y(m, n) = \sum_p R^2(m, n + p) R^2(m, p),$$

$$\overline{C_y}(n) = \sum_m C_y(m, n).$$

## References

- [1] Jackson VP. The role of ultrasound in breast imaging. *Radiology*, 177:305–311, 1990.
- [2] Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, and Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology*, 196:123–134, 1995.
- [3] Giger ML, Al-Hallaq H, Huo Z, Moran C, Wolverton DE, Chan CW, and Zhong W. Computerized analysis of lesions in us images of the breast. *Acad Radiol*, 6:665–674, 1999.
- [4] Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, and Zeman RK. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrason Imaging*, 15:267–285, 1993.
- [5] Chen DR, Chang RF, and Huang YL. Computer-aided diagnosis applied to us of solid breast nodules by using neural networks. *Radiology*, 213:407–412, 1999.
- [6] Golub RM, Parsons RE, Sigel B, and et el. Differentiation of breast tumors by ultrasonic tissue characterization. *J Ultrasound Med*, 12:601–608, 1993.
- [7] Sahiner B, LeCarpentier GL, Chan HP, and et el. Computerized characterization of breast masses using three-dimensional ultrasound images. In *Proceedings of the SPIE*, volume 3338, pages 301–312, 1998.

- [8] Kupinski MA and Giger ML. Automated seeded lesion segmentation on digital mammograms. *IEEE Trans Med Im*, 17:510–517, 1998.
- [9] Metz CE. Some practical issues of experimental design and data analysis in radiological roc studies. *Invest Radiol*, 24:234–245, 1989.
- [10] Jiang Y, Metz CE, and Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology*, 201:745–750, 1996.
- [11] Tohno E, Cosgrove DO, and Sloane JP. *Ultrasound Diagnosis of Breast Disease*. Churchill Livingstone, Edinburgh, Scotland, 1994.
- [12] Bick U, Giger ML, Schmidt RA, and Doi K. A new single-image method for computer-aided detection of small mammographic masses. In *Prodeedings of CAR '95*, pages 357–363, 1995.
- [13] Huo Z, Giger ML, Vyborny CJ, and et el. Analysis of spiculation in the computerized classification of mammographic masses. *Med Phys*, 22:1569–1579, 1995.
- [14] Horsch K, Giger ML, Venta LA, and Vyborny CJ. Computerized analysis of breast lesions on ultrasound. *To be submitted*, 2000.
- [15] Lachenbruch PL. *Discriminant Analysis*. Hafner, London, England, 1975.

# Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging

Kenneth G. A. Gilhuijs<sup>a)</sup> and Maryellen L. Giger

*The Kurt Rossmann Laboratories for Radiologic Image Research, The University of Chicago, Department of Radiology, MC 2026, 5841 South Maryland Avenue, Chicago, Illinois 60637*

Ulrich Bick<sup>b)</sup>

*The University of Muenster, Department of Radiology, Albert-Schweitzer-Strasse 33, D-48129 Muenster, Germany*

(Received 3 November 1997; accepted for publication 30 June 1998)

Contrast-enhanced magnetic resonance imaging (MRI) of the breast is known to reveal breast cancer with higher sensitivity than mammography alone. The specificity is, however, compromised by the observation that several benign masses take up contrast agent in addition to malignant lesions. The aim of this study is to increase the objectivity of breast cancer diagnosis in contrast-enhanced MRI by developing automated methods for computer-aided diagnosis. Our database consists of 27 MR studies from 27 patients. In each study, at least four MR series of both breasts are obtained using FLASH three-dimensional (3D) acquisition at 90 s time intervals after injection of Gadopentetate dimeglumine (Gd-DTPA) contrast agent. Each series consists of 64 coronal slices with a typical thickness of 2 mm, and a pixel size of 1.25 mm. The study contains 13 benign and 15 malignant lesions from which features are automatically extracted in 3D. These features include margin descriptors and radial gradient analysis as a function of time and space. Stepwise multiple regression is employed to obtain an effective subset of combined features. A final estimate of likelihood of malignancy is determined by linear discriminant analysis, and the performance of classification by round-robin testing and receiver operating characteristics (ROC) analysis. To assess the efficacy of 3D analysis, the study is repeated in two-dimensions (2D) using a representative slice through the middle of the lesion. In 2D and in 3D, radial gradient analysis and analysis of margin sharpness were found to be an effective combination to distinguish between benign and malignant masses (resulting area under the ROC curve: 0.96). Feature analysis in 3D was found to result in higher performance of lesion characterization than 2D feature analysis for the majority of single and combined features. In conclusion, automated feature extraction and classification has the potential to complement the interpretation of radiologists in an objective, consistent, and accurate way. © 1998 American Association of Physicists in Medicine. [S0094-2405(98)01509-0]

**Key words:** breast imaging, magnetic resonance imaging (MRI), computer-aided diagnosis, ROC analysis, contrast agent

## I. INTRODUCTION

Breast cancer is a major cause of death among women in most western countries. Although mammography has demonstrated to be the most efficient tool for early detection of breast cancer, the technique may result in a missed fraction of cancers as high as 9%.<sup>1</sup> In addition, the fraction of lesions found by mammography that is sent to biopsy and proves to be malignant can be as low as 10%–20%.<sup>2</sup> Accurate examination of mammograms is particularly difficult in dense breasts, because lesions may be occluded by dense tissue. Consequently, complementary information by ultrasound or biopsy is often obtained.

Magnetic resonance imaging (MRI) is a promising complementary technique to mammography because of its inherent three-dimensional (3D) nature. In addition to possible improvement of diagnostic accuracy from dense breasts, MRI has shown superior potential for quantification of tumor volume, and detection of multifocal and multicentric disease.<sup>3,4</sup> These issues are of interest in the consider-

ation of breast-conserving therapy. From current consensus, MR is particularly suited for specific problem cases, such as patients who have high risk of developing breast cancer, patients with implants, postoperative scars, or clinical evidence of breast cancer that cannot be detected by conventional diagnostic methods.<sup>4–6</sup>

MRI has become practically useful for breast imaging since the introduction of contrast agents that alter the spin-lattice ( $T_1$ ) relaxation time.<sup>7,8</sup> Due to increased vascularity and capillary permeability of tumors,<sup>9</sup> contrast-enhanced MRI shows better distinction between lesions and normal tissue than conventional MRI alone. Nonetheless, contrast-enhanced MRI is known to enhance both malignant as well as some benign types of masses, thus compromising the specificity of the technique. In general, the sensitivity reported for diagnosis of breast cancer in MR images is larger than 90%,<sup>10</sup> but the reported specificity varies considerably and may be substantially lower.<sup>6,10</sup> The majority of these studies are solely based on enhancement as a function of



time. To improve the specificity without reducing the sensitivity, the morphology of enhancement has been studied as well.<sup>11,12</sup> Most of these techniques are, however, based on slice by slice assessment of the morphology in 2D. The performance is likely to improve when one takes full advantage of the 3D nature of the MR data.

An important aspect that may contribute to varying specificity is interobserver variation in the interpretation of the MR images. Nearly all studies presented to date are based on visual assessment by one or multiple radiologists. Musurakis *et al.*<sup>13</sup> report significant variability in the assessment of lesions in MR data by human readers and stress the importance of standardized terminology. Heywang-Köbrunner *et al.*<sup>6</sup> indicate that differences in interpretation guidelines will influence the accuracy of contrast-enhanced MRI. Attempts to increase the objectivity of the interpretation have recently been reported using quantitative rating of features such as spiculation by a radiologist, followed by merging of these ratings using an interpretation model.<sup>12</sup> In this scheme, the classification stage is objective, but the rating of the features is still subjective to the interpretation of the radiologist.

Automated quantification and classification of features to discriminate between benign and malignant lesions has been pursued in other diagnostic areas such as mammography in the context of computer-aided diagnosis (CAD).<sup>14</sup> Several investigators have successfully developed methods for computerized detection<sup>15-17</sup> and computerized classification<sup>18-20</sup> for ultimate use in CAD as "second readers" for radiologists. In addition to objective analysis, computerized analysis can take full advantage of information across slices in 3D data sets which is difficult to assess visually from individual images.

The aim of this study is to increase the objectivity of breast cancer diagnosis in contrast-enhanced MRI. This aim is pursued by automated extraction of features that quantify spatial properties of contrast enhancement in 3D, and by merging different features into an estimate of malignancy using automated classification. The ultimate objective of this feasibility study is to reduce the number of biopsies of benign lesions and to increase the sensitivity for cancer cases.

## II. MATERIAL AND METHODS

### A. Image and patient data

The images in this study were obtained using fast low-angle shot (FLASH) 3D acquisition at field strength of 1.0 T (Siemens Impact, Siemens, Erlangen, Germany). The acquisition parameters were:  $TR = 14.0$  ms,  $TE = 7.0$  ms, and flip angle of  $25^\circ$ . Fat suppression was not employed. The patients were scanned in prone position using a standard double-breast coil. In total, 141 preoperative MR series were acquired from 27 patients. Each series contains 64 coronal slices with a typical field of view of  $32 \times 16$  cm<sup>2</sup>. Each slice contains  $256 \times 256$  pixels of  $1.25 \times 1.25$  mm<sup>2</sup> and has a typical thickness of 2 mm. There are no gaps in between the slices. Gadopentetate dimeglumine (Gd-DTPA) contrast agent was injected intravenously by power injection after acquisition of the precontrast MR series. At least four series

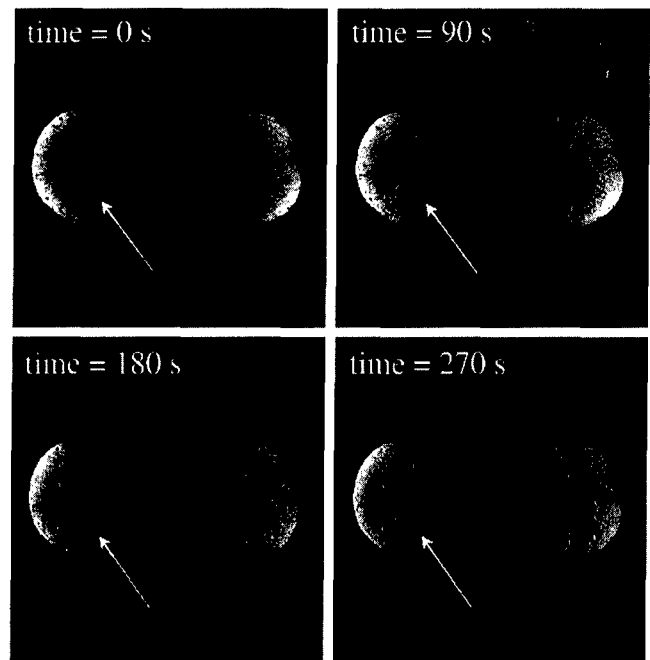


FIG. 1. Example of contrast enhancement in a malignant lesion, illustrated on a dynamic series of a single MR slice. Note the irregular "donut" shape of the lesion (arrow) as it enhances in time before merging with the background.

were taken per patient at 90 s intervals. Figure 1 shows an example of a dynamic MR sequence on a single slice through a malignant lesion.

The database in this study contains 28 lesions: 13 benign and 15 malignant masses. Histology in 27 out of 28 lesions was confirmed by open excisional biopsy, one case was benign based on core biopsy and four-year follow up. The distribution of the size of the lesions is shown in Fig. 2. The relative position of the lesions in the breast varies, some are close to the skin and near the chest wall. Benign masses include fibroadenoma (6/13), papilloma (2/13), and benign mastopathy (5/13). Malignant cases include papillary (1/15),

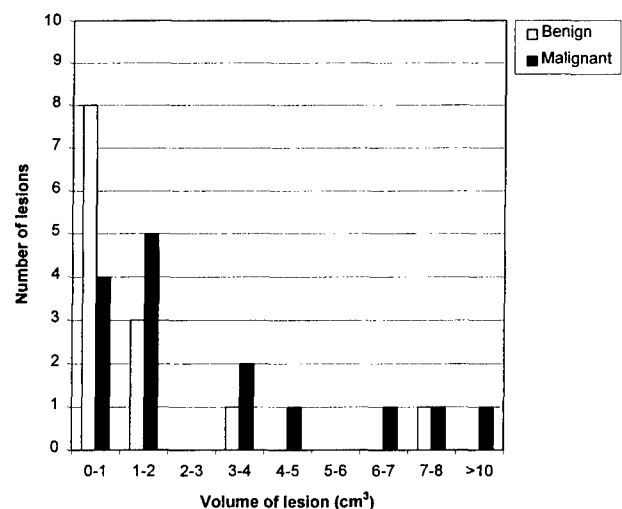


FIG. 2. Distribution of the size of benign and malignant lesions in our database.

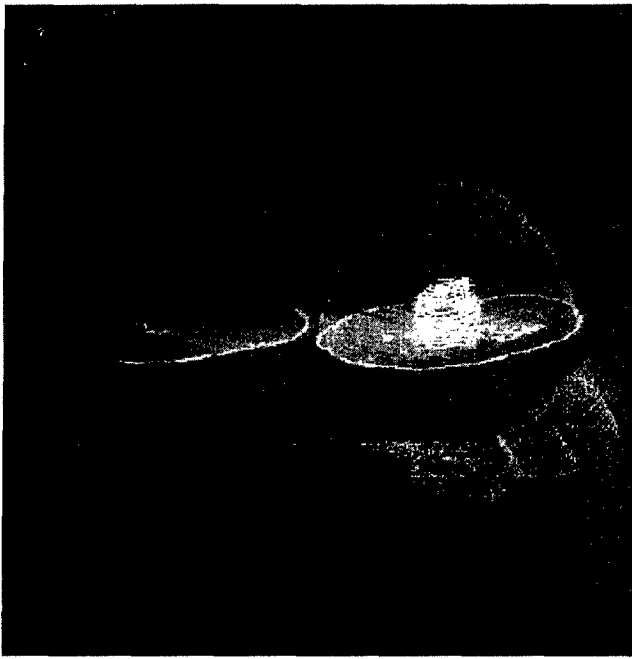


FIG. 3. Visualization of lesion shape, size, and orientation in 3D with respect to reference landmarks. The MR data is acquired in slices in coronal orientation, but the computerized analysis of the lesion is performed in all three dimensions simultaneously.

tubular (2/15), and medullary carcinoma (1/15), invasive lobular (3/15) and ductal (7/15) cancer, and ductal carcinoma *in situ* (DCIS) (1/15).

## B. Feature extraction

In this preliminary study, suspect masses were delineated by a radiologist (U.B.) experienced in MR-mammography and blinded concerning the histological diagnosis. This segmentation was performed in the subtraction images (postcontrast–precontrast) by contouring the enhanced tumor area in each slice that intersected the lesion. All available subtraction images were used for this purpose. As an additional reference, the radiologist had access to the original (nonsubtracted) MR images as well. All other stages of the scheme, though, as described below, are fully automated.

The proposed strategy for computerized analysis of dynamic MR data in 3D consists of two consecutive stages: Feature extraction and classification. The feature extraction stage is aimed at quantification of spatial properties of enhancement in suspicious lesions. Feature extraction has two parts: Extraction of the breast volume, and quantification of spatial properties. Although the MR data is obtained in slices, the analysis of the lesion is performed in 3D, taking all directions into account (Fig. 3). The volume of the breast is extracted from the MR data by global segmentation of pixel values at a threshold that maximizes the interclass variance between two pixel-value regions.<sup>21</sup> All slices in the data contribute to the computation of a single threshold value. The result of the segmentation is a 3D binary mask in which breast voxels are labeled with value “1,” and background voxels with value “0.” Remaining gaps in the mask are

removed by morphological closing operations:<sup>22</sup> Morphological erosion<sup>22</sup> is employed to remove an empirically established margin of two voxels from the external surface of the breast mask. This step is required to avoid strong voxel-value gradients near the borders of the breast to be included in the computation of gradient-based lesion features. Subsequent computation of features in the original MR data is restricted to voxels that have value “1” at the corresponding locations in the breast mask.

Features investigated in this study concern the inhomogeneity of uptake in the lesion [Eqs. (1) and (2)], sharpness of the lesion margin [Eqs. (3) and (4)], analysis of the shape of the lesion [Eqs. (5) and (6)], and radial gradient analysis [Eq. (7)].

If the set of voxel values in the lesion at time frame “*i*” is given by  $F_l(\mathbf{r}, i)$ , where vectors  $\mathbf{r}$  point to the lesion, and index “*i*” runs from frame 0 (i.e., the frame before injection of contrast) to  $M-1$  (where  $M$  is the total number of time frames), then the inhomogeneity of contrast uptake is characterized by two features which are defined by

$$\max_{i=0, \dots, M-1} \left\{ \frac{\text{variance}_{\mathbf{r}} F_l(\mathbf{r}, i)}{\text{variance}_{\mathbf{r}} F_l(\mathbf{r}, 0)} \right\}, \quad (1)$$

referred to here as variance of uptake

and

$$\min_{i=0, \dots, M-2} \left\{ \frac{\text{variance}_{\mathbf{r}} F_l(\mathbf{r}, i)}{\text{variance}_{\mathbf{r}} F_l(\mathbf{r}, i+1)} \right\}, \quad (2)$$

referred to here as change in variance of uptake,

where  $\text{variance}_{\mathbf{r}} F_l(\mathbf{r}, i)$  denotes the computation of the variance of the voxel values at all  $\mathbf{r}$  in the lesion at fixed time frame “*i*.”

The sharpness of the lesion margins is characterized by two features as well. The first feature is given by

$$\max_{i=0, \dots, M-1} \left\{ \frac{\text{mean}_{\mathbf{r}} \|\nabla [F_m(\mathbf{r}, i) - F_m(\mathbf{r}, 0)]\|}{\text{mean}_{\mathbf{r}} F_m(\mathbf{r}, i)} \right\}, \quad (3)$$

referred to here as margin gradient,

where  $\nabla [F_m(\mathbf{r}, i) - F_m(\mathbf{r}, 0)]$  denotes the set of voxel-value gradients at the margin of the suspect lesion in the difference images of time frame “*i*” and precontrast frame “0.” Thus, the sharpness of the uptake of contrast is computed at the lesion margin. The range of vectors  $\mathbf{r}$  in  $F_m$  is limited to a shell—three voxels thick—centered on the surface of the lesion. The shell is employed to account for small inaccuracies that may occur in the delineation of the lesion outlines.

The second feature related to margin sharpness is defined by

$$\frac{\text{variance}_{\mathbf{r}} \|\nabla [F_m(\mathbf{r}, i) - F_m(\mathbf{r}, 0)]\|}{[\text{mean}_{\mathbf{r}} F_m(\mathbf{r}, i)]^2}, \quad (4)$$

referred to here as variance of margin gradient,

and is only computed from the subtraction frames of “*i*” and “0” where the margin gradient [Eq. (3)] is maximum.

In Eqs. (3) and (4), computation of the spatial gradient is accomplished in 3D by convolution with the components of a  $3 \times 3 \times 3$  Sobel filter<sup>23</sup> in three orthogonal directions. Note that this approach takes information on lesion margins across slices into account.

Circularity of the shape of the lesion in 3D is given by

$$\frac{\text{volume of lesion within sphere of effective diameter}}{\text{volume of lesion}}, \quad (5)$$

and irregularity in 3D by

$$1 - \frac{\pi \cdot \text{effective diameter}^2}{\text{surface of lesion}}, \quad (6)$$

where the effective diameter is defined by

$$2 \cdot \sqrt[3]{\frac{3 \cdot \text{volume of lesion}}{4\pi}}.$$

The volume and the surface of the lesion are estimated from the contours of the segmented masses. For this purpose, a set of binary images is created in which the pixels at and enclosed by the contours are set to value "1" (object pixels), and remaining pixels to value "0" (background pixels). The volume of the lesion is determined by multiplying the number of object pixels with the volume of one voxel in world coordinates. The surface of the lesion is computed by combining the set of 2D binary images into a 3D binary representation of the lesion. Next, the faces of the object voxels that are exposed to background voxels in the 3D binary volume are identified by examining the value of the neighboring voxels in the  $x$ ,  $y$ , and  $z$  directions. The face of an object voxel is exposed to the background if the neighboring voxel has value "0." The surface of the lesion is subsequently determined by calculating the sum of the areas of the faces exposed to the background in world coordinates. Note that the circularity and irregularity in 3D are computed from the volume and the surface of the lesion in world coordinates—rather than in voxel coordinates—to account for the differences between pixel size and slice thickness (i.e., the anisotropic voxel shapes).

Radial gradient analysis is based on examination of the angles between voxel-value gradients and lines intersecting a single point near the center of the suspect lesion (i.e., lines in radial directions). Radial gradient values are given by the dot product of the gradient direction and the radial direction. The histogram of radial gradient values—quantifying the frequency of occurrence of the dot products in a given region of interest (Fig. 4)—is called the radial gradient histogram (RGH). Analysis of the RGH yields

$$\max_{i=0, \dots, M-1} \{ \text{variance}_p H(p) \}, \quad p > 0, \quad (7)$$

referred to here as the variance of RGH values.

In this relationship,  $H(p)$  denotes the normalized RGH, and variable  $p$  is given by the normalized dot product

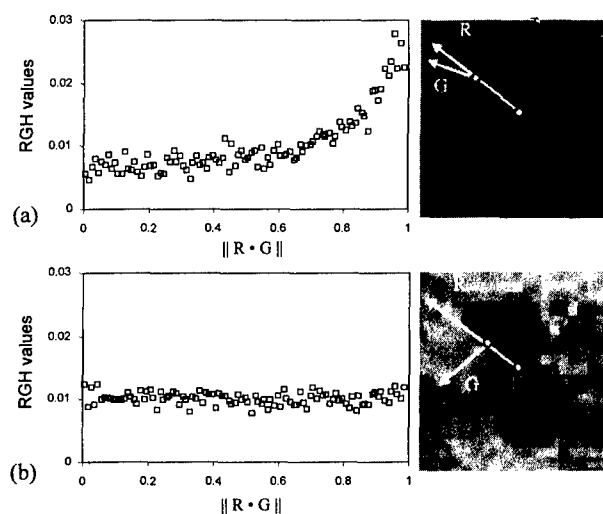


FIG. 4. The radial gradient histogram (RGH) of a volume of interest (VOI) with a benign lesion (a) and a malignant lesion (b). Shown are images of representative cross sections through the lesions. The radial vector ( $R$ ) originates in the center of the VOI. The gradient vector ( $G$ ) indicates the local direction of the voxel-value gradient. The RGH maps the dot product of  $R$  and  $G$  against the frequency of occurrence (RGH values). In benign lesions,  $R$  and  $G$  tend to point in comparable directions within the VOI, yielding a peak in the RGH around 1.0 (a). Malignant lesions typically extend in less spherical patterns resulting in a flat RGH (b). The variance of RGH values is used to quantify the flatness of the RGH.

$$p = \frac{|\nabla[F_b(\mathbf{r}, i) - F_b(\mathbf{r}, 0)] \cdot (\mathbf{r} - \mathbf{r}_c)|}{\|\nabla[F_b(\mathbf{r}, i) - F_b(\mathbf{r}, 0)]\| \cdot \|\mathbf{r} - \mathbf{r}_c\|},$$

where  $\nabla[F_b(\mathbf{r}, i) - F_b(\mathbf{r}, 0)]$  indicates the set of voxel-value gradients in a rectangular box of interest at the subtracted time frames " $i$ " and " $0$ ." The box encompasses the suspect lesion with an additional margin of three voxels along all sides. Vector  $\mathbf{r}_c$  points to the center of this rectangular box. Thus,  $\mathbf{r}_c$  generally will not (and does not need to) point to the exact center of the lesion. This aspect will be reviewed in more detail in the Discussion section.

In essence, above equations quantify the observation that malignant lesions take up contrast agent in a less homogeneous pattern than benign masses, have less sharp boundaries and are more irregularly shaped. "Circularity" quantifies how well the lesion conforms to a spherical shape [Eq. (5)], and "irregularity" indicates the roughness of the surface of the lesion [Eq. (6)]. Radial gradient analysis was previously applied to mammograms to quantify spiculation of projected masses.<sup>18</sup> The analysis provides a measure that indicates how well the image structures in a region of interest (ROI) extend in a radial pattern originating from the center of the ROI. Round and well-defined masses produce different measures than irregular and spiculated lesions. In the current study, radial gradient analysis is extended to 3D. The feature "Variance of RGH values" quantifies how well the image structures in a volume of interest (VOI) extend in a spherical pattern originating from the center of the VOI (Fig. 4).

With the exception of circularity and irregularity, which are computed from the coordinates of the segmented lesions, all other features are extracted from the data at each available

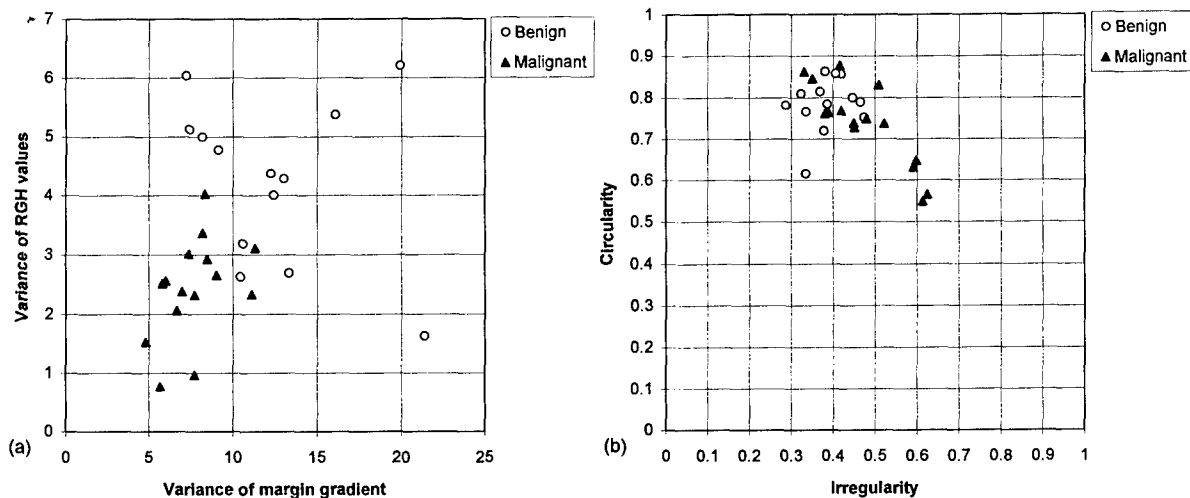


FIG. 5. Relationships of various features for the database of benign and malignant lesions (a) margin gradient analysis and radial gradient analysis, (b) shape analysis.

time frame and combined by minimum or maximum operations, such as described by Eqs. (1)–(4) and (7).

To assess the efficacy of 3D analysis in comparison with conventional 2D techniques, the feature extraction and analysis procedures were repeated, although in 2D, on a single representative slice through the middle of the lesions.

### C. Feature selection and classification

From the total set of seven features, stepwise multiple regression<sup>24</sup> produced a selection that performs efficiently in distinguishing between benign and malignant lesions. The technique involves adding and removing features to obtain a limited subset that provides statistically significant separation in the estimated likelihood of malignancy. Linear discriminant analysis<sup>25</sup> is employed to estimate this likelihood of malignancy from single or combined features.

### D. Evaluation

The performance of the computerized method in classification (distinguishing between benign and malignant lesions) is quantified by receiver operating characteristics (ROC) analysis.<sup>26</sup> In particular the area under the ROC curve ( $A_z$ )—which maps the fraction of false positives to the fraction of true positives—is used as a measure of performance in this study. Sensitivity is defined as the true-positive fraction, specificity as one minus the false-positive fraction. The area under the ROC curve at true-positive fractions larger than 0.9 (partial  $A_z$ ) is employed to rate the performance of computerized analysis at high sensitivity levels.<sup>27</sup>

The general performance of the computerized method is estimated by round-robin testing<sup>28</sup> on our current database. This “leave-one-out” technique involves estimating the likelihood of malignancy from all cases but one, testing classification on that single case, and repeating the procedure until each case has been tested individually.

## III. RESULTS

All features investigated in this study show potential for distinguishing between benign and malignant lesions (Fig. 5, Table I). As expected, benign masses were found to extend more along spherical patterns than malignant lesions, and the margins of benign masses were found to be sharper on average than the margins of malignant lesions. An interesting observation is, however, that the variance of sharpness along the margin of the lesions is larger on average for benign than for malignant masses [Fig. 5(a)]. A possible explanation is offered in the discussion section. Less surprising was the result that some malignant lesions tend to be more irregularly shaped than benign masses [Fig. 5(b)]. Circularity was, however, not found to be a strong feature to distinguish between

TABLE I. Area under the ROC curves ( $A_z$ ) using 2D and 3D analysis of individual and combined features. The standard deviations (1 SD) are shown in parentheses.

Feature	$A_z$ (2D)	$A_z$ (3D)
Inhomogeneity of uptake		
Variance of uptake	0.54 (0.11)	0.72 (0.11)
Change in variance of uptake	0.59 (0.11)	0.77 (0.10)
Sharpness		
Margin gradient	0.83 (0.07)	0.88 (0.07)
Variance of margin gradient	0.71 (0.10)	0.86 (0.07)
Shape		
Circularity	0.67 (0.10)	0.65 (0.10)
Irregularity	0.66 (0.10)	0.80 (0.08)
Radial gradient analysis		
Variance of RGH values	0.80 (0.08)	0.88 (0.07)
Combinations of features		
Variance of RGH values and margin gradient	0.87 (0.11)	0.92 (0.05)
variance of RGH values and variance of margin gradient	0.86 (0.08)	0.96 (0.03)

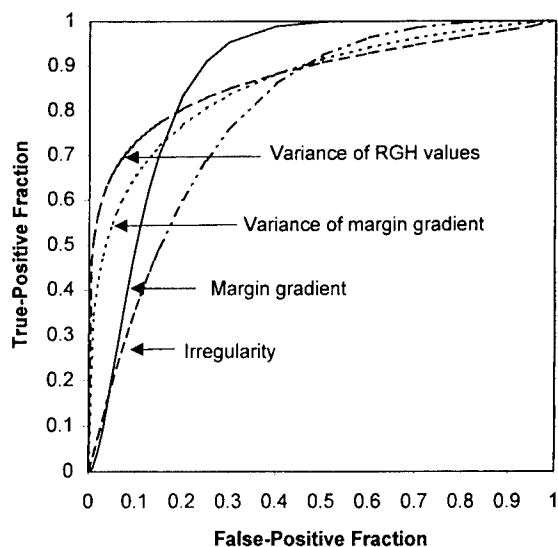


FIG. 6. ROC curves showing the performance of the best single features in the task of distinguishing between benign and malignant lesions. (RGH=Radial gradient histogram.)

benign and malignant in our database (Table I). Using single features only, the highest performances were obtained with radial gradient analysis ( $A_z=0.88$ ), sharpness ( $A_z=0.88$  and  $0.86$ ), and shape analysis of the lesion ( $A_z=0.80$ ). The corresponding ROC curves are shown in Fig. 6. For operation at sensitivities larger than 0.9, the highest performance was achieved with the "margin gradient" feature, yielding a partial  $A_z$  value of 0.70. Note that although the "margin gradient" and "variance of margin gradient" have comparable  $A_z$  values (Table I), their ROC curves are differently shaped (Fig. 6). The curve of the "margin gradient" feature is steeper, indicating that higher specificity can be achieved at high sensitivity. In addition, "variance of RGH values" and "margin gradient" have comparable  $A_z$  values, although the shapes of their ROC curves differ. The "margin gradient" feature was found to perform better at higher sensitivity than the "variance of RGH values" (Fig. 6).

Stepwise multiple regression at a confidence region of 95% resulted in combination of two features: "Variance of RGH values" [eq. (7)] and "variance of margin gradient" [eq. (4)]. Their combined performance in distinguishing between benign and malignant lesions resulted in an  $A_z$  value of 0.96. The corresponding ROC curve is shown in Fig. 7. Some combinations that have more than two features yielded slightly better results, but this increase in performance was not found to be statistically significant.

Based on round-robin testing on our database, the computerized scheme achieved 77% specificity (10/13) at 100% sensitivity. Note that all lesions in our database had been biopsied. In other words, all benign lesions were basically misclassified before biopsy, whereas the computerized method misclassified only 3 of the 13 benign lesions without misclassifying any malignant cases. These preliminary results indicate that the computerized method has the potential to reduce the number of biopsies of benign lesions.

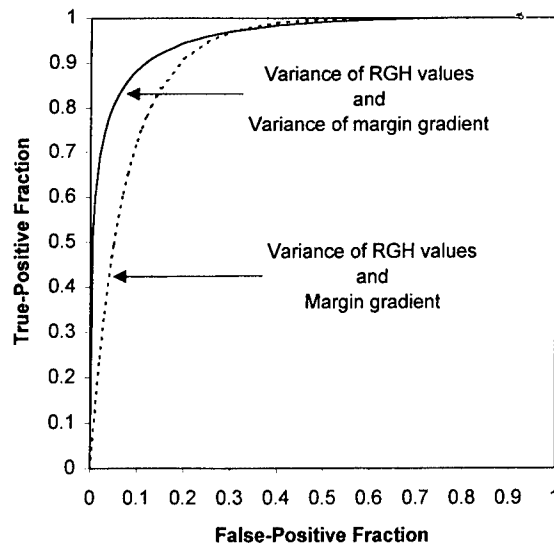


FIG. 7. ROC curves showing the performance of effective combined features in the task of distinguishing between benign and malignant lesions (using round-robin testing). (RGH=Radial gradient histogram.)

Although statistical significance of the differences in performance between 3D and 2D analysis of features could not be ascertained given the current size of our database, a consistent superior performance using 3D analysis was found for nearly all single and all combined features (Table I).

#### IV. DISCUSSION

Automated extraction of mathematically defined features in 3D yields encouraging results in distinguishing benign from malignant lesions ( $A_z=0.96$ ).  $A_z$  values of classification achieved by manual rating of features by radiologists have been reported<sup>12</sup> to be around 0.86. Direct comparison with results reported in the literature is, however, difficult due to differences in database used. In addition, most reports indicate only a single operating point for sensitivity and specificity on the ROC curve. In our study, the operating point must be tuned to a desired trade-off between sensitivity and specificity. This trade-off would be determined clinically by a cost-benefit analysis. In our current study, a sensitivity of 100% can be obtained at specificity of 77%. At this operating point, all malignant lesions, including one case of DCIS, are successfully identified as malignant. One of the next steps of research is to study the accuracy of breast cancer diagnosis done by radiologists when assisted by the automated technique as a second opinion.

We found that the margins of benign lesions are typically sharper in appearance than the margins of malignant lesions, which is consistent with observations from other studies. When the average sharpness of the lesion margins is high, deviations from this average caused by anatomical morphology, partial volume effect, or inaccuracies in segmentation of the lesion, will result in higher variance values than would occur with a small average sharpness of lesion margins. Thus, the variation of sharpness along the margin is expected

to be larger for benign than for malignant cases. Our preliminary results indicate that this observation yields good potential to discriminate between benign and malignant lesions [Fig. 5(a), Table I].

In the current study, we found that spatial features are effective to distinguish between benign and malignant lesions, in particular the combination of radial gradient analysis and analysis of margin sharpness. Most studies of contrast-enhanced MRI of the breast are based on analysis of temporal features of uptake only—such as speed of uptake—and report varying specificity. It is possible, however, that consistent high performance can be obtained from temporal features when the temporal resolution of the data is high. Boetes *et al.*<sup>29</sup> report encouraging results from temporal features in data obtained at high temporal resolution at the expense of spatial resolution. Physiological models of contrast uptake and washout have also been applied to increase the specificity of the diagnosis from temporal features, e.g., Tofts *et al.*<sup>30</sup> The preliminary results from our study indicate that good distinction between benign and malignant lesions can be obtained from spatial features without extremities in temporal or spatial resolution. An important aspect is, however, 3D acquisition and analysis of the spatial features. The results from the current study indicate that it is beneficial to analyze spatial features in 3D rather than in 2D to distinguish between benign and malignant lesions.

In addition to differences in image acquisition, other aspects may also influence the specificity of the diagnosis. Differences in bolus size and particular hemodynamic characteristics of each patient as well as hormonal factors, may cause variable enhancement.<sup>4,6</sup> Image artifacts can be caused by inhomogeneity of the magnetic field and by patient motion. To reduce the effect of some of these aspects, features have been normalized within or across time frames in the same examination. Gradient artifacts caused by inhomogeneity of the magnetic field typically occur at much lower spatial frequency than the lesion margins, and were found to be of less importance in this study. Patient movement is estimated to be about 2 mm on average in our data set. Because the voxel dimensions are  $1.25 \times 1.25 \times 2.0 \text{ mm}^3$ , the motion causes some blurring of the lesions rather than an actual displacement of image structures. To avoid image artifacts due to motion of the heart, the MR slices were obtained in coronal orientation. Different slice thickness may also result in differences in performance. To take the anisotropic voxel shapes in the MR data into account, features related to the shape of the lesion are calculated in world coordinates, rather than in voxel coordinates. Lesion sharpness is calculated, however, using a  $3 \times 3 \times 3$  Sobel filter, which does not account for the anisotropic voxel shape. An intuitive approach would be to resample the MR voxels into a uniform coordinate grid by linear interpolation. Such linear modeling of the discontinuity at the edges of the lesions may, however, lead to underestimation of sharp lesion margins, thus compromising the benefit of the correction. In addition, the approach does not take other effects of deviating slice thickness into account, such as differences in partial volume effect. These aspects are topic of future study.

Automated segmentation of the lesions is likely to further improve the objectivity of diagnosis. Techniques for this purpose have recently been investigated, e.g., Lucas-Quesada *et al.*,<sup>31</sup> and will remain a subject of future research. In our study, the sharpness-related features are computed in a shell around the indicated outline of the tumor to account for small inaccuracies in the segmentation. Radial gradient analysis does not require accurate delineation of the margins of the lesion: The region of interest is a rectangular box positioned roughly around the lesion in the subtracted images. Spurious gradients associated with background noise are expected to be randomly distributed with respect to the radial directions of the noise voxels, thus adding a small constant offset to the RGH values. Consequently, a region of interest somewhat larger than the size of the actual lesion is not expected to have much effect on the variance of the RGH values. The radial lines intersect the center of the rectangular bounding box positioned around the lesion. This intersection point will usually not coincide with the exact center of the lesion. Nevertheless, since the shape of the lesions is generally not perfectly symmetrical and regular, the center of the lesion is expected to vary with the shape of the lesions in a similar way as the center of the rectangular bounding box is expected to vary with the shape of the lesions. Consequently, neither definition of the center is expected to yield superior performance compared to the other.

It is likely that the ability to distinguish between benign and malignant lesions will decrease for smaller tumor sizes. The smallest lesion in our database has a volume of  $0.1 \text{ cm}^3$  (benign case) and was correctly classified at no loss of malignant cases. At this operating point, 3 of the 13 benign lesions were incorrectly classified as malignant. These lesions had volumes of 0.2, 0.9, and  $3.4 \text{ cm}^3$ , respectively (papilloma and benign mastopathy). Other benign lesions with similar histology and volumes were, however, correctly classified, as well as all malignant lesions, which have sizes ranging from  $0.1 \text{ cm}^3$  to larger than  $10 \text{ cm}^3$  (Fig. 2). In conclusion, our database did not show an obvious correlation between accuracy of the performance of the computerized diagnosis and lesion size, nor between accuracy and histology. Evaluation of the computerized analysis technique on larger databases is, however, required, and may warrant the use of more advanced classification methods, such as artificial neural networks.

Once a mass is suspected to be malignant, localization for accurate biopsy is a next step. Techniques for MRI-directed biopsy are being developed and evaluated.<sup>32,33</sup> The 3D nature of the MR data may allow useful complementary information to visualize the size and shape of the lesion as well as its location relative to the nipple position and pectoralis muscle (Fig. 3).

## V. CONCLUSIONS

A technique aimed at computer-aided diagnosis of suspect lesions in contrast-enhanced MRI of the breast has been developed to increase the objectivity of breast cancer diagnosis. Initial results of analysis of spatial features in 3D indicate good accuracy of classification ( $A_z = 0.96$ ), and higher per-

formance than analysis in 2D for the majority of single and combined features. Consequently, automated extraction of features that quantify the spatial properties of contrast enhancement has the potential to complement the interpretation of radiologists in an objective, consistent and accurate way. In addition, the computerized analysis technique shows potential to reduce the fraction of biopsies of benign lesions.

## ACKNOWLEDGMENTS

The authors thank K. Doi, G. S. Karczmar, Z. Huo, H. Al-Hallaq, C. E. Metz, and M. Kupinski for helpful discussions. This work was supported in part by the Netherlands Cancer Institute (NKI/AvL), U.S. Army Grants No. DAMD17-96-1-6058 and No. DAMD17-98-1-8194, and USPHS Grants No. CA 24806 and No. RR 11459. M. L. Giger is a shareholder of R2 Technology, Inc., Los Altos, CA. It is the policy of the University of Chicago that investigators disclose publicly actual or potential significant financial interests that may appear to be affected by the research activities.

<sup>a</sup>Author to whom correspondence should be addressed. Also at: The Netherlands Cancer Institute/Antoni van Leeuwenhoek Huis, Radiotherapy Department, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands; Electronic mail: kengil@nki.nl

<sup>b</sup>Currently at: University Clinics Charite, Department of Radiology, Schumannstrasse 20121 10117 Berlin, Germany.

<sup>1</sup>L. H. Baker, "Breast cancer detection demonstration project: five year summary report," *Cancer* **32**, 194–225 (1982).

<sup>2</sup>L. W. Bassett and R. H. Gold, *Breast Cancer Detection: Mammography and Other Methods in Breast imaging* (Grune & Stratton, New York, 1987).

<sup>3</sup>P. L. Davis, M. J. Staiger, K. B. Harris, M. A. Gannot, J. Klementaviciene, K. S. McCarty, Jr., and H. Tobon, "Breast cancer measurements with magnetic resonance imaging, ultrasonography, and mammography," *Breast Cancer Res. Treatment* **37**, 1–9 (1996).

<sup>4</sup>S. E. Harms, "MRI in breast cancer diagnosis and treatment," *Curr. Probl. Diagn. Radiol.* **25**, 193–215 (1996).

<sup>5</sup>S. E. Harms, D. P. Flaming, K. L. Hesley, M. D. Meiches, R. A. Jensen, W. P. Evans, D. A. Savino, and R. V. Wells, "MR imaging of the breast with rotating delivery of excitation off resonance: clinical experience with pathologic correlation," *Radiology* **187**, 493–501 (1993).

<sup>6</sup>S. H. Heywang-Köbrunner, P. Viehweg, A. Heinig, and Ch. Küchler, "Contrast-enhanced MRI of the breast: accuracy, value, controversies, solutions," *Eur. J. Radiol.* **24**, 94–108 (1997).

<sup>7</sup>S. H. Heywang, D. Hahn, H. Schmid, I. Krischke, W. Eiermann, R. Bassermann, and J. Lissner, "MR imaging of the breast using gadolinium-DTPA," *J. Comput. Assist. Tomogr.* **10**, 199–204 (1986).

<sup>8</sup>W. A. Kaiser and E. Zeitler, "MR imaging of the breast: fast imaging sequences with and without Gd-DTPA," *Radiology* **170**, 681–686 (1989).

<sup>9</sup>D. Revel, R. Brasch, H. Paaanen, W. Rosenau, W. Grodd, B. Engelstad, P. Fox, and J. Winkelhake, "Gd-DTPA contrast enhancement and tissue differentiation in MR imaging of experimental breast carcinoma," *Radiology* **158**, 319–323 (1986).

<sup>10</sup>F. Kelcz and G. Santyr, "Gadolinium-Enhanced breast MRI," *Crit. Rev. Diagn. Imaging* **36**, 287–338 (1995).

<sup>11</sup>A. H. Adams, J. R. Brookeman, and M. B. Merickel, "Breast lesion discrimination using statistical analysis and shape measures on magnetic resonance imagery," *Comp. Med. Imaging Graphics* **5**, 339–349 (1991).

<sup>12</sup>L. W. Nunes, M. D. Schnall, S. G. Orel, M. G. Hochman, C.-P. Langlotz, C. A. Reynolds, and M. H. Torosian, "Breast MR imaging: Interpretation Model," *Radiology* **202**, 833–841 (1997).

<sup>13</sup>S. Mussurakis, D. L. Buckley, A. M. Coady, L. W. Turnbull, and A. Horseman, "Observer variability in the interpretation of contrast enhanced MRI of the breast," *Br. J. Radiol.* **69**, 1009–1016 (1996).

<sup>14</sup>M. Giger and H. MacMahon, "Image processing and computer-aided diagnosis," *Radiol. Clin. N. Am.* **34**, 565–596 (1996).

<sup>15</sup>H. P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. MacMahon, and P. M. Jokich, "Image feature analysis and computer-aided diagnosis in digital radiography: 1. Automated detection of microcalcifications in mammography," *Med. Phys.* **14**, 538–548 (1987).

<sup>16</sup>W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," *Radiology* **191**, 331–337 (1994).

<sup>17</sup>N. Karssemeijer, "Recognition of stellate lesions in digital mammograms," *Digital Mammography* (Elsevier Science B.V., Amsterdam, The Netherlands, 1994).

<sup>18</sup>Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, P. Lu, D. E. Wolverton, and R. A. Schmidt, "Analysis of spiculation in the computerized classification of mammographic masses," *Med. Phys.* **22**, 1569–1579 (1995).

<sup>19</sup>Y. Jiang, R. M. Nishikawa, D. E. Wolverton, C. E. Metz, M. L. Giger, R. A. Schmidt, C. J. Vyborny, and K. Doi, "Automated feature analysis and classification of malignant and benign clustered microcalcifications," *Radiology* **198**, 671–678 (1996).

<sup>20</sup>R. M. Rangayyan, N. El-Faramawy, J. E. L. Desautels, and O. A. Alim, "Discrimination between benign and malignant breast tumors using a region-based measure of edge profile acutance," *Digital Mammography* (Elsevier Science B.V., Amsterdam, The Netherlands, 1996).

<sup>21</sup>N. Otsu, "A threshold selection method from gray-value histograms," *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).

<sup>22</sup>J. Serra, *Image Analysis and Mathematical Morphology*, 2nd ed. (Academic, San Diego, California, 1988), Vol. 1.

<sup>23</sup>R. C. Gonzales and P. Wintz, *Digital Image Processing*, 2nd ed. (Addison-Wesley, Reading, Massachusetts, 1987).

<sup>24</sup>J. Neter, W. Wasserman, and M. H. Kutner, *Applied Linear Statistical Models Regression, Analysis of Variance, and Experimental Designs*, 2nd ed. (Irwin, Homewood, 1985).

<sup>25</sup>R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 3rd ed. (Prentice-Hall, Englewood Cliffs, New Jersey, 1992).

<sup>26</sup>C. E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies," *Invest. Radiol.* **24**, 234–245 (1989).

<sup>27</sup>Y. Jiang, C. E. Metz, and R. M. Nishikawa, "A receiver operating characteristic partial area index for highly sensitive diagnostic tests," *Radiology* **201**, 745–750 (1996).

<sup>28</sup>G. Gong, "Cross-validation, the Jackknife, and the Bootstrap: Excess Error Estimation in Forward Logistic Regression," *J. Am. Stat. Assoc.* **81**, 108–113 (1986).

<sup>29</sup>C. Boetes, J. O. Barentsz, R. D. Mus, R. F. van der Sluis, L. J. T. O. van Erning, J. H. C. L. Hendriks, R. Holland, and S. H. J. Ruys, "MR characterization of suspicious breast lesions with a Gadolinium-enhanced TurboFLASH subtraction technique," *Radiology* **193**, 777–781 (1994).

<sup>30</sup>P. S. Tofts, B. Berkowitz, and M. D. Schnall, "Quantitative analysis of dynamic Gd-DTPA enhancement in breast tumors using a permeability model," *Magn. Reson. Med.* **33**, 564–568 (1995).

<sup>31</sup>F. A. Lucas-Quesada, U. Sinha, and S. Sinha, "Segmentation strategies for breast tumors from dynamic MR images," *J. Magn. Reson. Imaging* **6**, 753–763 (1996).

<sup>32</sup>S. H. Heywang-Köbrunner, A. T. Huynh, P. Viehweg, W. Hanke, H. Requardt, and I. Paprosch, "Prototype breast coil for MR-guided needle localization," *J. Comput. Assist. Tomogr.* **18**, 876–881 (1994).

<sup>33</sup>S. G. Orel, M. D. Schnall, R. W. Newman, C. M. Powell, M. H. Torosian, and E. F. Rosato, "MR imaging—guided localization and biopsy of breast lesions: initial experience," *Radiology* **193**, 97–102 (1994).

# Computerized Analysis of Lesions in US Images of the Breast<sup>1</sup>

Maryellen L. Giger, PhD, Hania Al-Hallaq, BS, Zhimin Huo, PhD, Catherine Moran, BA  
Dulcy E. Wolverton, MD, Chun Wai Chan, MS, Weiming Zhong, MS

**Rationale and Objectives.** Breast sonography is not routinely used to distinguish benign from malignant solid masses because of considerable overlap in their sonographic appearances. The purpose of this study was to investigate the computerized analyses of breast lesions in ultrasonographic (US) images in order to ultimately aid in the task of discriminating between malignant and benign lesions.

**Materials and Methods.** Features related to lesion margin, shape, homogeneity (texture), and posterior acoustic attenuation pattern in US images of the breast were extracted and calculated. The study database contained 184 digitized US images from 58 patients with 78 lesions. Benign lesions were confirmed at biopsy or cyst aspiration or with image interpretation alone; malignant lesions were confirmed at biopsy. Performance of the various individual features and output from linear discriminant analysis in distinguishing benign from malignant lesions was studied by using receiver operating characteristic (ROC) analysis.

**Results.** At ROC analysis, the feature characterizing the margin yielded  $A_z$  values (area under the ROC curve) of 0.85 and 0.75 in distinguishing between benign and malignant lesions for the entire database and for an "equivocal" database, respectively. The equivocal database contained lesions that had been proved to be benign or malignant at cyst aspiration or biopsy. Linear discriminant analysis round-robin runs yielded  $A_z$  values of 0.94 and 0.87 in distinguishing benign from malignant lesions for the entire database and for the equivocal database, respectively.

**Conclusion.** Computerized analysis of US images has the potential to increase the specificity of breast sonography.

**Key Words.** Artificial intelligence; breast imaging; computer-aided diagnosis; computer vision; differential diagnosis; US imaging.

Breast cancer is a leading cause of death in women, causing an estimated 44,000 deaths per year (1). Mammography is the most effective method for early detection of breast cancer, and periodic screening of asymptomatic women reduces the mortality rate (2–4). Many breast cancers are detected, and these patients are referred for biopsy on the basis of a radiographically observed mass lesion or cluster of microcalcifications. General rules for the differentiation of benign from malignant mammographically identified breast lesions exist (5,6), but considerable misclassification of these lesions still occurs. On average, less than 30% of masses referred for surgical breast biopsy are actually malignant (7).

Breast sonography is an important adjunct to diagnostic mammography, and it is typically performed on palpable and/or mammographically identified masses to determine their cystic or solid nature. The accuracy rate of ultrasonography (US) has been reported to be 96%–100% in the diagnosis of simple benign cysts (8), and masses so characterized do not require further evaluation. US has not been used for screening purposes, however, because of relatively high false-negative and false-positive rates. Even so, US is being evaluated as a potential screening method in women with dense breasts (9). Physicians at

---

*Acad Radiol* 1999; 6:665–674

<sup>1</sup> From the Department of Radiology, MC 2026, Kurt Rossmann Laboratories for Radiologic Image Research, University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637. Received December 31, 1998; revision requested February 2, 1999; revision received June 10; accepted June 15. Supported in part by National Institutes of Health grant P20 CA66132 and U.S. Army Medical Research and Materiel Command grant DAMD17-98-1-8194. Address reprint requests to M.L.G.

M.L.G. is a shareholder in R2 Technology, Inc (Los Altos, Calif). It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest that would reasonably appear to be directly and significantly affected by the research activities.

© AUR, 1999



some centers are successful at visually distinguishing benign from malignant masses by using US, but physicians at most facilities are unable to rely on breast US to avoid biopsy because of the considerable overlap in the sonographic appearances of these masses.

With the advent of modern, high-frequency transducers that have improved spatial and contrast resolution, however, several sonographic features have emerged as potential indicators of malignancy and others as potential indicators of benign masses (10,11). Benign features include hyperechogenicity, ellipsoid shape, mild lobulation, and a thin, echogenic pseudocapsule. Malignant features include spiculation, angular margins, marked hypoechogenicity, posterior acoustic shadowing, and a depth-to-width ratio greater than 0.8.

Stavros et al (12) used various features to characterize masses as being either benign, indeterminate, or malignant. Their classification scheme had a sensitivity of 98.4% and a negative predictive value of 99.5%. The sonographic evaluation described by these investigators, however, is much more extensive and complex than that usually performed at most breast-imaging centers. US is a notoriously operator-dependent modality, and until these encouraging results are corroborated through additional studies by other investigators, it is unclear how widely applicable or reliable such sonographic classification schemes truly are.

Computer-aided techniques have been applied to color Doppler US evaluation of breast masses with promising results (13). Color Doppler imaging is a technique that focuses on the vascularity of lesions. Not all sonographically visible cancers have demonstrable neovascularity, however, and benign lesions can be vascular. Therefore, the sensitivity and specificity of this technique are inherently somewhat limited. These limitations have been demonstrated in power Doppler imaging of solid breast masses (14).

Comprehensive summaries of investigations regarding mammographic computer-aided diagnosis have been published (15,16). During the 1960s and 1970s, several investigators attempted to analyze mammographic abnormalities by using computers (17–24). These investigators demonstrated the potential capability of computers in the detection of mammographic abnormalities. Gale et al (17) and Getty et al (18) both reported on computer-based classifiers that take diagnostically relevant features obtained from radiologists' readings of breast images as input. Getty et al found that with use of this classifier, community radiologists performed as well as unaided

expert mammographers in differentiating benign from malignant lesions. In addition, Swett and Miller (19) developed an expert system to provide both visual and cognitive feedback to radiologists by using a critiquing approach combined with an expert system. At the University of Chicago, we have shown that computerized analysis of mass lesions (21,23) and clustered microcalcifications (22,24) as shown on digitized mammograms yields performance rates similar to those of expert mammographers and significantly better ( $P < .05$ ) than those of average radiologists in distinguishing malignant from benign lesions.

US is a digital modality that is amenable to application of computer-aided diagnosis techniques that could ultimately be used in a real-time fashion (at the time of examination) to improve diagnostic accuracy. Given that sonographic interpretation is a subjective process, however, and that criteria have been developed that may allow for differentiation of benign from malignant solid breast masses, it is reasonable to assume that computer-aided diagnosis techniques applied to sonographic images would also improve radiologists' performance, particularly when this method is combined with corresponding mammographic data (25). Recently, Garra et al (26) showed promising results with the use of computer-extracted features derived from co-occurrence matrices of images of breast lesions.

In this study, we attempted to determine if computer analysis of breast lesions in gray-scale, US images could be used to discriminate malignant from benign lesions.

## MATERIALS AND METHODS

### Database

Masses were viewed sonographically by filming representative images in orthogonal planes. The US examinations were performed with an Ultramark 9 with High Definition Imaging (HDI) from Advanced Technology Laboratories (Bothell, Wash) with a high-frequency, 7.5-MHz, electronically focused, near-field imaging probe. The static images of lesions that did not contain overlaid cursors or color Doppler signals were used in this study.

The US film images were retrospectively collected and then digitized with a laser film scanner (KFDR-S; Konica, Tokyo, Japan) with a scanner pixel size of 0.1 mm and 10-bit quantization. Each multiformat film contained only one US image. Film digitization is not the optimal approach to acquiring digital US data, but it was the only one available for this initial study.

The 58 patients in the study ranged in age from 35 to 89 years (mean age, 53 years). They had 78 masses as shown on 184 digitized US images. Benign lesions were confirmed at biopsy or cyst aspiration or with image interpretation alone, whereas malignant lesions were confirmed at biopsy. Of the total 184 images, 144 were from 43 patients with 62 benign lesions, and 40 were from 15 patients with 16 malignant lesions. Benign lesions included simple cysts, complex cysts, and solid masses. Of the 62 benign lesions, 19 (all solid) were proved at biopsy, five (1 solid lesion and four complex cysts) were proved at cyst aspiration, and 38 (four solid lesions and 34 cysts) were deemed to be benign on the basis of visual interpretation of the US images alone. All 16 malignant lesions were proved at biopsy.

Lesions were further subcategorized into an "equivocal" category on the basis of the necessity of performing an interventional procedure to determine their status. The 24 benign lesions that were proved at biopsy or cyst aspiration and the 16 malignant lesions made up the equivocal database (total, 40 lesions). The purpose of this subcategorization was to determine the ability of the computer features to distinguish benign from malignant lesions that required an interventional procedure (cyst aspiration or biopsy) for definitive diagnosis.

### Manual Lesion Segmentation and Region-of-Interest Selection

Once digitized, the US images were displayed on an IBM monitor, and a breast-imaging radiologist (D.E.W.) outlined the approximate margins of each lesion. Figure 1 shows US images of breast lesions with outlined margins. Regions of interest (ROIs) of  $32 \times 32$  pixels were selected from regions within and around the lesions. Features were calculated on the basis of the manually extracted lesion margin or the  $32 \times 32$ -pixel ROI.

### Automated Feature Extraction

Four types of lesion characteristics were investigated: margin, shape, homogeneity (texture), and posterior acoustic attenuation. Table 1 lists these characteristics and their relationships to benign and malignant lesions.

The lesion characteristics were quantified by using various computer-extracted features, and Table 2 lists the computer-extracted features used in distinguishing malignant from benign lesions. Features were calculated either along or within the lesion margin or within the  $32 \times 32$ -pixel ROI (placed within the central portion of or posterior to the lesion).

To quantify the lesion margin characteristics, a gradient analysis was performed along a computer-expanded margin of the lesion. In this analysis, the manually extracted margin was first expanded by using morphologic filtering. Next, this region was processed by using a Sobel filter to obtain the gradient and its direction at each pixel. The normalized radial gradient was then calculated to quantify the margin sharpness and degree of irregularity (shape) (21,27). The normalized radial gradient (21,27) is given by the equation

$$\text{normalized radial gradient} = \frac{\sum_{P \in \text{margin}} \cos \phi \sqrt{D_x^2 + D_y^2}}{\sum_{P \in \text{margin}} \sqrt{D_x^2 + D_y^2}},$$

where  $D_x$  is the gradient along the x axis,  $D_y$  is the gradient along the y axis, and  $\phi$  is the angle between the gradient vector and the radial gradient. A lower value for the normalized radial gradient indicates a less distinct margin.

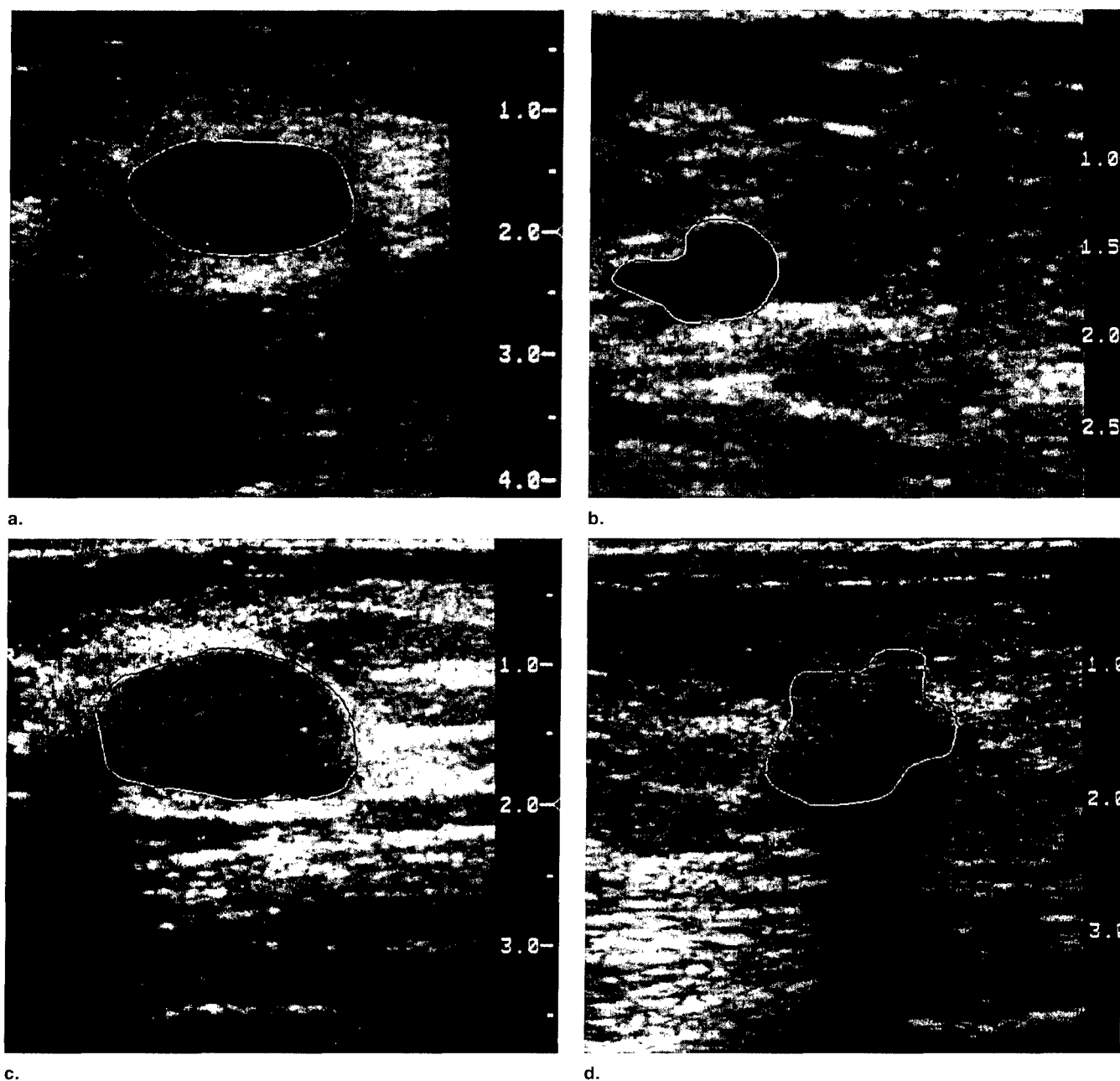
The geometric measure of shape in terms of a short-to-long axis ratio for each lesion was determined by using the image data along the margin. Note here that the short-to-long axis ratio corresponds to a depth-to-width ratio to extract the orientation of the long axis. Cysts tend to be ellipsoid, thereby resulting in a depth-to-width ratio of much less than 1, whereas malignant lesions tend to have a vertical or round axial orientation (28).

Texture can be described through spatial relationships between image pixels by using changes in the intensity patterns and gray levels. Texture characteristics of the homogeneity within the lesion were determined by using a measure of coarseness (29). The texture measure of coarseness (local uniformity) is given by the equation

$$\text{coarseness} = \left[ \sum_i^{G_h} p_i s(i) \right]^{-1},$$

where  $G_h$  is the highest gray-level value in the ROI and  $p_i$  is the probability of occurrence for gray level  $i$ . Thus, if  $N$  is the width of the ROI ( $N = 32$ ),  $d$  is the neighboring size (half the operating kernel size  $W$ ), the  $i$ th entry of  $s$  is given by

$$s(i) = \begin{cases} \sum |i - A_i| & \text{for } i \in \{N_i\} \text{ if } N_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$



**Figure 1.** (a) US image of a simple cyst. (b) US image of a complex cyst. (c) US image of a benign solid lesion. (d) US image of a malignant lesion, along with radiologist-drawn lesion margin.

where  $\{N_i\}$  is the set of pixels having gray level  $i$ ,

$$A_i = \frac{1}{W-1} \sum_{p=-d}^d \sum_{q=-d}^d f(x+p, y+q) \quad (p, q) \neq (0,0) \text{ to exclude } (x,y)$$

$$W = (2d+1)^2 \quad (d=3).$$

Thus, a lower value of coarseness corresponds to a finer visual texture.

The computerized assessment of posterior acoustic attenuation or enhancement associated with different lesions was determined in two ways: (a) by comparing the gray-level values within the lesion with those posterior to that lesion and (b) by comparing the gray-level values posterior to the lesion with those in adjacent tissue at the same depth. These calculations were performed to quantify

**Table 1**  
**Lesion Characteristics and Their Relationship to Benign and Malignant Lesions**

Characteristic	Benign (Cystic or Solid)	Malignant
Lesion margin	Smooth borders	Angular margins, spiculation
Lesion shape	Ellipsoid, mildly lobulated	Irregular; depth-to-width ratio, >0.8
Texture within lesion	Anechoic, hyperechoic, reverberation artifacts	Hypoechoic
Posterior acoustic attenuation	Posterior enhancement	Posterior shadowing

the amount of any posterior acoustic shadowing or enhancement. For example, benign lesions are often associated with posterior enhancement, whereas malignant lesions are often associated with posterior shadowing. Simple cysts that are anechoic produce less attenuation of the US waves than surrounding parenchyma produces and, thereby, cause relative hyperechogenicity posterior to the lesion. In this analysis,  $32 \times 32$ -pixel ROIs were placed within the lesion, posterior to the lesion, and in the adjacent tissue at the same depth, and the differences in average gray levels were then calculated to quantify the posterior acoustic attenuation.

The feature value for a given lesion was obtained by averaging that feature value over all views of the lesion. Each lesion had from two to five images available from one clinical examination.

Linear discriminant analysis (LDA) was used to merge the four individual, computer-extracted features into a single index related to an estimate for the likelihood of malignancy. In LDA, the discriminant function is formulated by using a linear combination of the individual features (30). Both consistency and round-robin runs were performed. In round-robin analysis, the discriminant function is trained on all but one case and is then tested on that remaining case; this process is repeated until all cases have been individually tested.

### Evaluation

Receiver operating characteristic (ROC) analysis (31) was used to evaluate (by case, not by image) the performance of the individual computer-extracted features in distinguishing benign from malignant lesions. The decision variable for the ROC analysis was each individual feature. The area under the ROC curve ( $A_z$ ) was used as an indicator of performance. Specificity at high sensitivity is relevant clinically, because the cost of missing a cancer is greater than the cost of performing an interventional procedure for a benign lesion. Therefore, we also calculated the performance of the features in the high-sensitivity range (true-positive fraction [TPF<sub>0</sub>], >0.90) by using the partial area

index ( ${}_{\text{TPF}_0}A_z'$ ), which is the portion of the area under the ROC curve that lies above the true-positive fraction divided by the constant  $(1 - \text{TPF}_0)$  (32). Both  $A_z$  and partial  $A_z$  values were calculated for the entire database and for the equivocal database.

### RESULTS

The  $A_z$  values for the various computer-extracted US features ranged from 0.54 to 0.85 in distinguishing benign from malignant lesions. Table 2 provides these values for both the entire database and the equivocal database. Because missing a cancer is more important clinically than performing an interventional procedure for a benign lesion, we used the partial area index to quantify performance of the features at a high-sensitivity level (32). Table 2 provides these  ${}_{\text{TPF}_0}A_z'$  values for the entire database and the equivocal database.

When the first four features (listed in Table 2) were used, LDA consistency runs yielded  $A_z$  values of 0.95 and 0.93 in distinguishing benign from malignant lesions for the entire database and the equivocal database, respectively, and the round-robin runs yielded  $A_z$  values of 0.94 and 0.87 in distinguishing benign from malignant lesions for the entire database and the equivocal database, respectively (Table 2). When the second posterior acoustic attenuation feature was used, LDA consistency runs yielded  $A_z$  values of 0.94 and 0.93 in distinguishing benign from malignant lesions for the entire database and the equivocal database, respectively, and the round-robin runs yielded  $A_z$  values of 0.92 and 0.86 in distinguishing benign from malignant lesions for the entire database and the equivocal database, respectively (Table 2).

### DISCUSSION

Figure 2 shows cluster plots of the coarseness and margin features for malignant and benign lesions in the entire database and the equivocal database. Figure 3 shows clus-

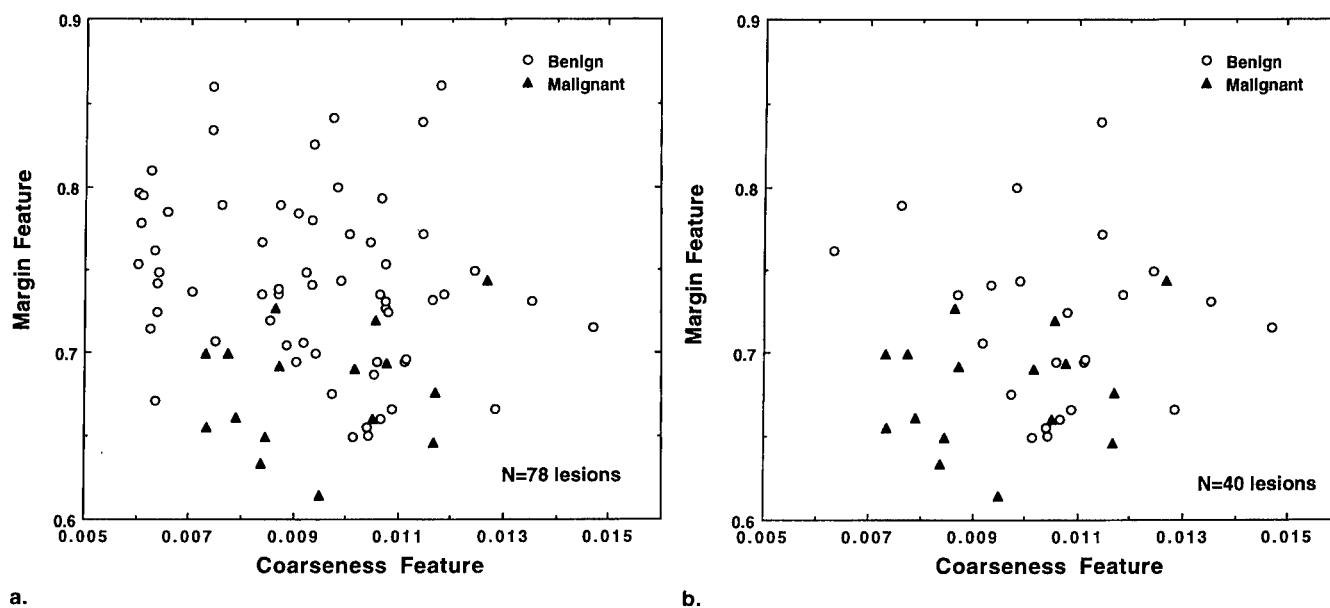
**Table 2**  
Computer-extracted Features Used to Quantify Lesion Characteristics

Analysis	Region for Database	Entire Database*		Equivocal Database†	
		$A_z$	$_{0.90}A_z$	$A_z$	$_{0.90}A_z$
Lesion margin					
1. Normalized radial gradient	Margin	0.85	0.46	0.75	0.28
Shape					
2. Depth-to-width ratio	Margin	0.67	0.20	0.75	0.29
Texture within the lesion					
3. Coarseness	ROI	0.54	0.12	0.67	0.14
Posterior acoustic attenuation					
4. Difference in gray level between "within lesion" and posterior to lesion	ROIs	0.77	0.29	0.72	0.27
5. Difference in gray level between "posterior to lesion" and adjacent tissue at same depth	ROIs	0.83	0.35	0.72	0.17
Linear discriminant analysis (features 1, 2, 3, and 4)					
Consistency analysis		0.95	0.78	0.93	0.70
Round-robin analysis		0.94	0.76	0.87	0.56
Linear discriminant analysis (features 1, 2, 3, and 5)					
Consistency analysis		0.94	0.68	0.93	0.58
Round-robin analysis		0.92	0.59	0.86	0.38

Note.—Performance is given in terms of  $A_z$  and partial  $A_z$  for the entire database and the "equivocal" database in distinguishing malignant from benign lesions.

\* $n = 78$  cases.

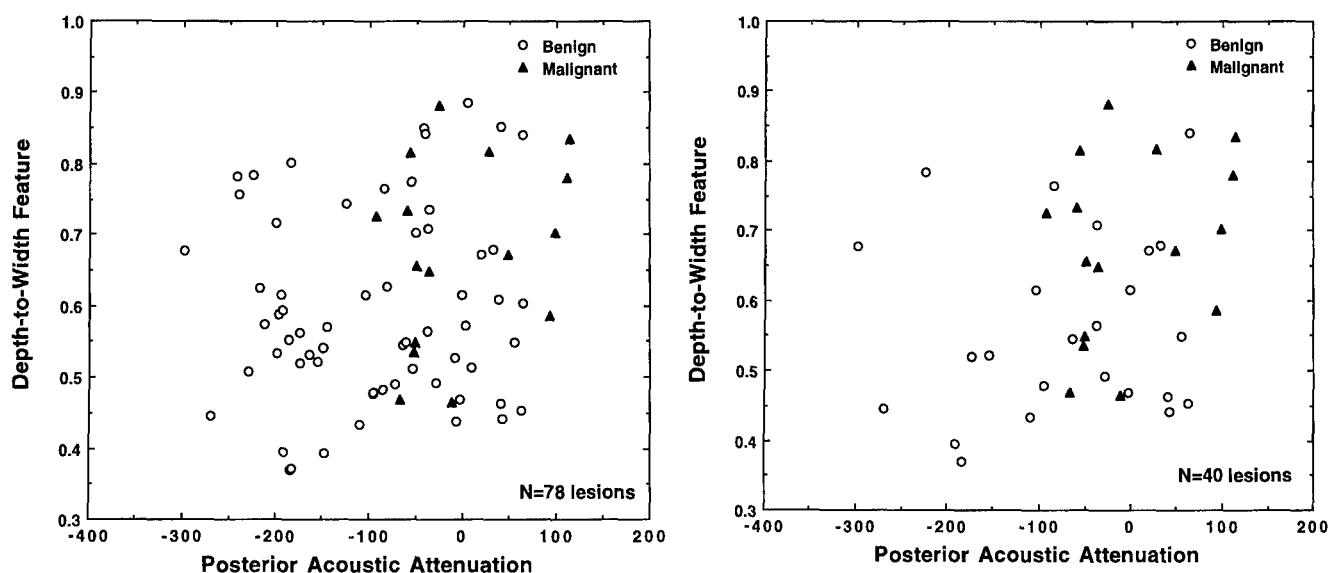
† $n = 40$  cases.



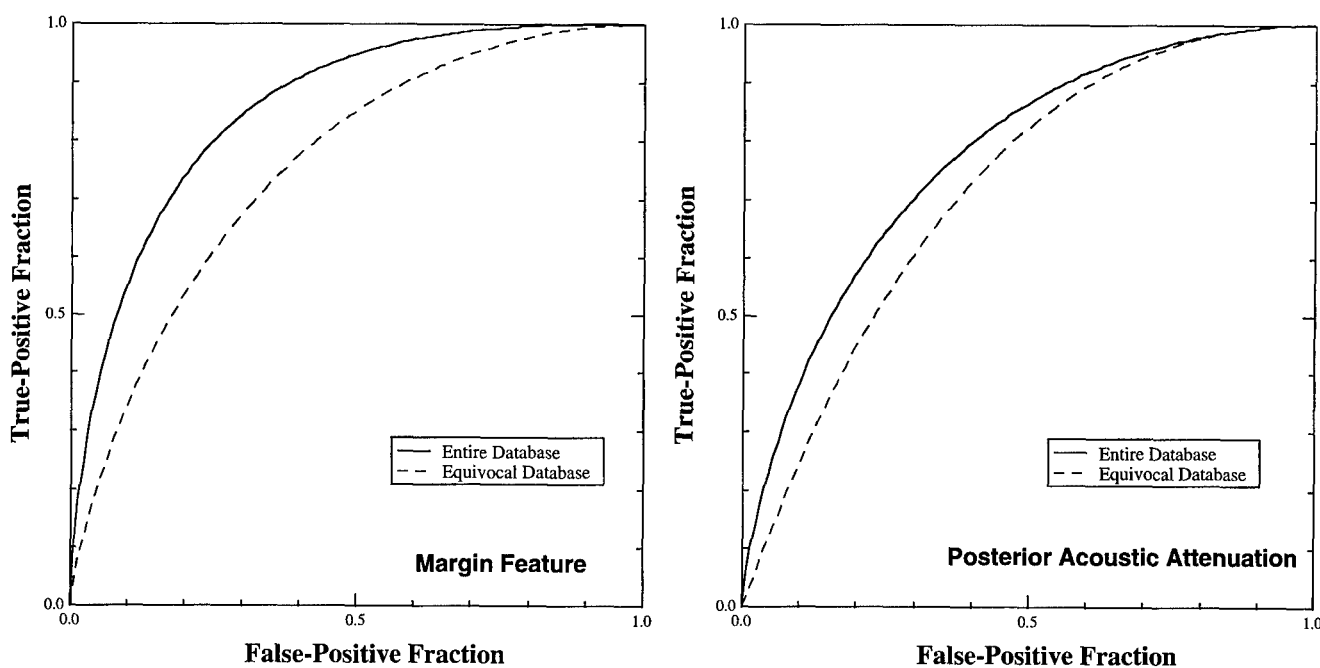
**Figure 2.** Cluster plots indicate feature values for margin and texture. (a) Values for the entire database. (b) Values for the equivocal database.

ter plots of the depth-to-width ratio and the (first) posterior acoustic attenuation feature for malignant and benign lesions in the entire database and the equivocal database. As these figures show, malignant lesions tend to exhibit less

distinct margins and more posterior shadowing than benign lesions as documented by using visual US criterion (10,12,26,28). It is interesting that many benign lesions not in the equivocal database of this study had a very



a. **Figure 3.** Cluster plots indicate feature values for lesion shape and posterior acoustic attenuation. (a) Values for the entire database. (b) Values for the equivocal database.

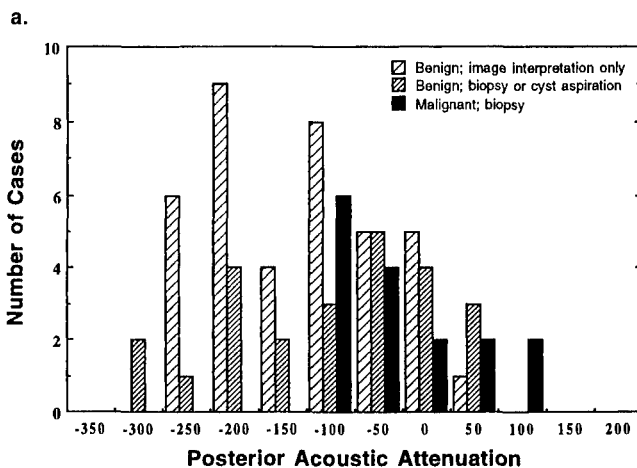
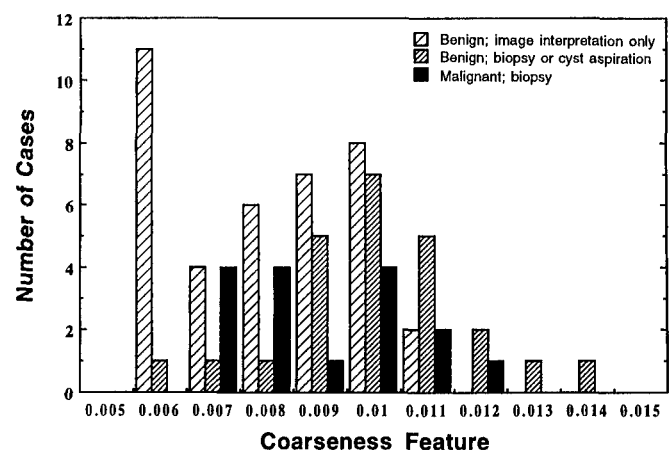
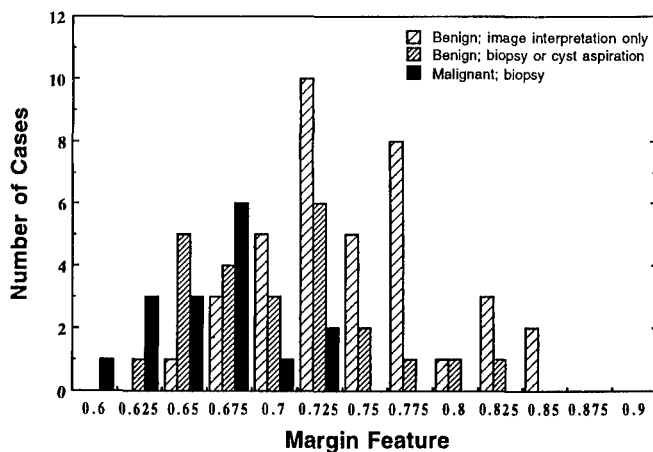


a. **Figure 4.** Performance of the computer-extracted features in distinguishing malignant from benign lesions for the entire database and for the equivocal database. (a) Margin. (b) Posterior acoustic attenuation.

“fine” texture (a low coarseness feature), because many were cysts and anechoic. In contrast, the benign solid lesions tended to have a coarse texture.

Figure 4 shows performance in terms of ROC curves for the features characterizing margin and acoustic attenuation in distinguishing malignant from benign lesions for

the entire database and for the equivocal database. The ROC curves are lower for the equivocal database than for the entire database, thereby indicating that as for radiologists, benign lesions in the equivocal database are more difficult to distinguish from malignant lesions. Figure 5, which shows histograms of the feature values (for malig-



b.

**Figure 5.** Histograms of the computer-extracted features of the benign nonequivocal database, benign equivocal database, and malignant database. (a) Margin. (b) Texture. (c) Posterior acoustic attenuation.

ing our digital US database, but it was the only one available for this study. Even with the limited image quality obtained from digitization of the multifilm US images, however, we could observe computer features capable of distinguishing malignant from benign lesions. Computerized analysis of direct digital US data is expected to improve the discriminatory value of the various features, especially for texture features of the lesion interior. Our future database collection will include direct digital data.

The purpose of this study was to determine if computer-extracted features on US images of breast lesions have the potential to discriminate malignant from benign lesions and, thus, ultimately help to reduce the number of unnecessary biopsies performed. This potential has been shown even though the number of lesions in the database is small. In addition, the features chosen for this study agree well with those used by radiologists when interpreting breast US images. It should be noted, however, that the computer-extracted features were obtained from radiologist-drawn margins. In the future, the subjectiveness of human-drawn margins will be eliminated with use of computer segmentation.

US images of the breast can yield information on the interior of the lesion (homogeneity), as well as on the interface of the lesion with its surroundings. This is why US is used to distinguish solid from cystic lesions. Gradient analysis of the margin yields information on the lesion margin, including its sharpness and shape. Geometric features relating to the depth-to-width ratio of the lesion are also useful, because even though many solid lesions may be ellipsoid, the orientation of the ellipse regarding the

nant lesions, benign lesions in the equivocal database, and the remaining benign lesions) for margin, texture, and posterior acoustic attenuation, further illustrates this point. As shown, malignant lesions tend to exhibit a coarse texture, less distinct margins, and posterior shadowing. There is, however, substantial overlap in the features of malignant and features of benign lesions in the equivocal database.

Table 2 and Figure 6 show performance of the linear discriminant function in distinguishing malignant from benign lesions for the entire database and the equivocal database. Of note are the partial  $A_c$  values and the shape of the ROC curves. Extraction from the fitted ROC curves for the equivocal database (in which all lesions underwent a clinical procedure [cyst aspiration or biopsy]) indicates that at a high sensitivity level (90%) for malignant cases, 30% of the benign cases were classified as benign and, thus, could potentially have avoided biopsy. Therefore, combined use of the four computer-extracted features yields superior performance.

Film digitization was not the optimal approach to acquir-

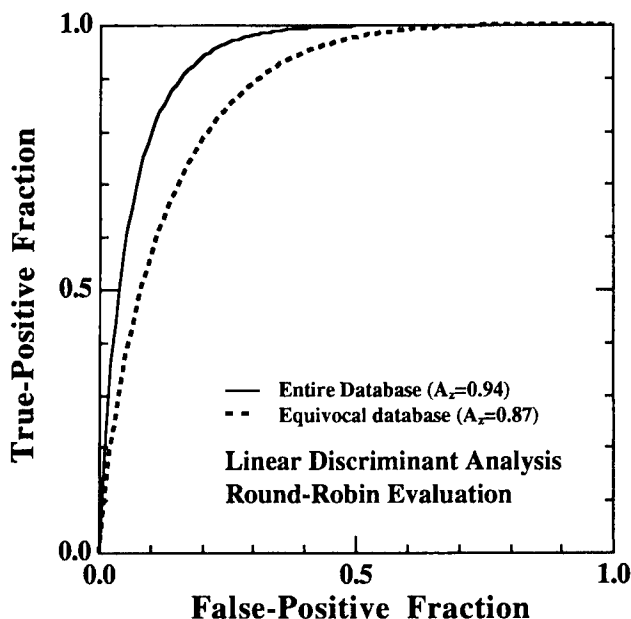


Figure 6. Performance of discriminant scores in distinguishing malignant from benign lesions for the entire database and the equivocal database. Results are from round-robin analyses.

skin is important in distinguishing benign from malignant lesions. In addition, computerized analysis allows for an objective assessment of posterior acoustic shadowing and enhancement, which can also aid in distinguishing benign from malignant lesions. It should be noted, however, that because shadowing depends on gain settings, scanning parameters should be set more automatically than they are at present.

Use of LDA as a classifier with which to merge the four computer-extracted features was useful for improving performance in distinguishing malignant from benign lesions for both the entire database and the equivocal database. As the overall database increases, other classifiers such as artificial neural networks will be investigated as means with which to merge these features into an estimate of the likelihood of malignancy.

In conclusion, we have developed methods for the computer analysis of breast lesions in US images. In this study, we automatically extracted and calculated features related to lesion margin, shape, texture (homogeneity) within the lesion, and posterior acoustic attenuation. From ROC analyses of the computer-extracted features, the features based on margin characteristics yielded  $A_z$  values of 0.85 and 0.75 in distinguishing benign from malignant lesions for the entire database and the equivocal database, respectively. The equivocal database consisted of lesions that had been proved to be benign or

malignant at either cyst aspiration or biopsy. LDA consistency runs yielded  $A_z$  values of 0.95 and 0.93 in distinguishing benign from malignant lesions for the entire database and the equivocal database, respectively, and the round-robin runs yielded  $A_z$  values of 0.94 and 0.87 in distinguishing benign from malignant lesions for the entire database and the equivocal database, respectively. Our results indicate that computerized analysis of US images has the potential to increase the specificity of breast sonography. These promising results warrant further development and testing on a large, direct-digital database.

#### ACKNOWLEDGMENTS

The authors thank Fred Winsberg, MD, for contributing to the database and Ulrich Bick, MD, Jason Rubenstein, Michael Chinander, and Samuel Armato III, PhD, for useful discussions.

#### REFERENCES

1. American Cancer Society. Cancer facts and figures—1998. New York, NY: American Cancer Society, 1998.
2. Feig SA. Decreased breast cancer mortality through mammographic screening: results of clinical trials. *Radiology* 1988; 167:659–665.
3. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* 1992; 30:187–210.
4. Smart CR, Hendrick RE, Rutledge JH, Smith RA. Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials. *Cancer* 1995; 75:1619–1626.
5. Bassett LW, Jackson VP, Jahan R, Yao SF, Gold RH. Diagnosis of diseases of the breast. Philadelphia, Pa: Saunders, 1997.
6. Kopans DB. Breast imaging. Philadelphia, Pa: Lippincott, 1998.
7. Brown ML, Houn F, Sickles EA, Kessler LG. Screening mammography in community practice: positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. *AJR* 1995; 165:1373–1377.
8. Jackson VP. The role of US in breast imaging. *Radiology* 1990; 177:305–311.
9. Kolb TM, Lichy J, Newhouse JH. Occult cancer in women with dense breasts: detection with screening US—diagnostic yield and tumor characteristics. *Radiology* 1998; 207:191–199.
10. Tohno E, Cosgrove DO, Sloane JP. Ultrasound diagnosis of breast diseases. Edinburgh, Scotland: Churchill Livingstone, 1994; 50–73.
11. Fornage BD, Lorigan JG, Andry E. Fibroadenoma of the breast: sonographic appearance. *Radiology* 1989; 172:671–675.
12. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology* 1995; 196:123–134.
13. Huber S, Delorme S, Knopp MV, et al. Breast tumors: computer-assisted quantitative assessment with color Doppler US. *Radiology* 1994; 192:797–801.
14. Birdwell RL, Ikeda DM, Jeffrey SS, Jeffrey RB Jr. Preliminary experience with power Doppler imaging of solid breast masses. *AJR* 1997; 169:703–707.
15. Giger ML. Computer-aided diagnosis. In: Haus A, Yaffe M, eds. *Syllabus: a categorical course on the technical aspects of breast imaging*. 2nd ed. Oak Brook, Ill: Radiological Society of North America, 1993; 272–298.
16. Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography. *AJR* 1994; 162:699–708.
17. Gale AG, Roebuck EJ, Riley P, et al. Computer aids to mammographic diagnosis. *Br J Radiol* 1987; 60:887–891.
18. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA. Enhanced interpretation of



- diagnostic images. *Invest Radiol* 1988; 23:240-252.
19. Swett HA, Miller PA. ICON: a computer-based approach to differential diagnosis in radiology. *Radiology* 1987; 163:555-558.
  20. Giger ML, Vyborny CJ, Schmidt RA. Computerized characterization of mammographic masses: analysis of spiculation. *Cancer Lett* 1994; 77: 201-211.
  21. Huo Z, Giger ML, Vyborny CJ, et al. Analysis of spiculation in the computerized classification of mammographic masses. *Med Phys* 1995; 22: 1569-1579.
  22. Jiang Y, Nishikawa RM, Wolverton DE, et al. Automated feature analysis and classification of malignant and benign clustered microcalcifications. *Radiology* 1996; 198:671-678.
  23. Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Schmidt RA, Doi K. Automated computerized classification of malignant and benign mass lesions on digitized mammograms. *Acad Radiol* 1998; 5:155-168.
  24. Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiol* 1999; 6:22-33.
  25. Giger ML, Huo Z, Wolverton DE, et al. Computer-aided diagnosis of digital mammographic and ultrasound images of breast mass lesions. In: Karssemeijer N, Thijssen M, Hendriks J, van Erning L, eds. *Digital mammography* 1998. Dordrecht, the Netherlands: Kluwer, 1998; 143-147.
  26. Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, Zeman RK. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrason Imaging* 1993; 15: 267-285.
  27. Bick U, Giger ML, Schmidt RA, Doi K. A new single-image method for computer-aided detection of small mammographic masses. In: Lemke HU, Inamura K, Jaffe CC, Vannier MW, eds. *Proceedings of CAR '95*. 1995; 357-363.
  28. Sohn C, Blohmer JU, Hamper UM. *Breast ultrasound: a systematic approach to technique and image interpretation*. New York, NY: Thieme, 1998.
  29. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 1989; 19:1264-1274.
  30. Lachenbruch PL. *Discriminant analysis*. London, England: Hafner, 1975.
  31. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* 1989; 24:234-245.
  32. Jiang Y, Metz CE, Nishikawa RM. A receiver operating characteristics partial area index for highly sensitive diagnostic tests. *Radiology* 1996; 201:745-750.



DEPARTMENT OF THE ARMY

US ARMY MEDICAL RESEARCH AND MATERIEL COMMAND  
504 SCOTT STREET  
FORT DETRICK, MARYLAND 21702-5012

REPLY TO  
ATTENTION OF:

MCMR-RMI-S (70-1y)

28 Aug 02

MEMORANDUM FOR Administrator, Defense Technical Information  
Center (DTIC-OCA), 8725 John J. Kingman Road, Fort Belvoir,  
VA 22060-6218


SUBJECT: Request Change in Distribution Statement

1. The U.S. Army Medical Research and Materiel Command has reexamined the need for the limitation assigned to technical reports written for this Command. Request the limited distribution statement for the enclosed accession numbers be changed to "Approved for public release; distribution unlimited." These reports should be released to the National Technical Information Service.

2. Point of contact for this request is Ms. Kristin Morrow at DSN 343-7327 or by e-mail at Kristin.Morrow@det.amedd.army.mil.

FOR THE COMMANDER:

Encl

  
PHYLIS M. RINEHART  
Deputy Chief of Staff for  
Information Management

ADB231838  
ADB240253  
ADB251610  
ADB275099  
ADB253637  
ADB261538  
ADB275186  
ADB264648  
ADB275102  
ADB241899  
ADB259033  
ADB266113  
ADB275663  
ADB254489  
ADB262700  
ADB276708  
ADB274345  
ADB274844  
ADB275154  
ADB275535  
ADB275101  
ADB275451  
ADB274597  
ADB273871  
ADB275145  
ADB274505  
ADB275851  
ADB274459  
ADB277942  
ADB277404  
ADB277494  
ADB277536