UNCLASSIFIED

AD NUMBER

ADB210734

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies only; Proprietary Info.; Mar 96. Other requests shall be referred to Commander, Army Medical Research and Materiel Command, Attn: MCMR-RMI-S, Fort Detrick, Frederick, MD 21702-5012.

AUTHORITY

USAMRMC ltr dtd 21 Apr 97

THIS PAGE IS UNCLASSIFIED

AD_____

GRANT NUMBER: DAMD17-93-J-3021

TITLE: Development of Methods for Computer-Assisted
Interpretations of Digital Mammograms for Early Breast Cancer
Detection

PRINCIPAL INVESTIGATOR: Maryellen L. Giger, Ph.D.

CONTRACTING ORGANIZATION: University of Chicago
                            Chicago, Illinois   60637

REPORT DATE: March 1996

TYPE OF REPORT: Final

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

19960531 072

DISCLAIMER NOTICE

UNCLASSIFIED

Technical Report
distributed by

DEFENSE
TECHNICAL
INFORMATION
CENTER

DTIC Acquiring Information-
Imparting Knowledge

Alexandria, Virginia 22304-6145

UNCLASSIFIED

THIS DOCUMENT IS BEST
QUALITY AVAILABLE. THE
COPY FURNISHED TO DTIC
CONTAINED A SIGNIFICANT
NUMBER OF PAGES WHICH DO
NOT REPRODUCE LEGIBLY.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

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 the 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 March 1996	3. REPORT TYPE AND DATES COVERED Final (1 Mar 93- 28 Feb 96)

4. TITLE AND SUBTITLE Development of Methods for Computer-Assisted Interpretations of Digital Mammograms for Early Breast Cancer Detection	5. FUNDING NUMBERS DAMD17-93-J-3021

6. AUTHOR(S)

Maryellen L. Giger, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Chicago Chicago, Illinois 60637	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick Frederick, 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, Mar 96). Other requests for this document shall be referred to Commander, U.S. Army Medical Research and Materiel Command, ATTN: MCMR-RMI-S, Fort Detrick, Frederick, MD 21702-5012.	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The goal of the research is to develop a computer-vision module as an aid to radiologists. Specifically, we have: (1) Further developed our computerized schemes for the detection and classification of masses and microcalcifications in digitized mammograms. Using new lesion features and genetic algorithms, we have improved mass detection by reducing the number of false positives per image from 2.6 to 1.5. Shift-invariant artificial neural networks have been used to reduce the number of false-positive detections in microcalcification detection. Computerized classification methods for both masses and clustered microcalcifications have demonstrated performance levels that were higher than that of "average radiologists" in the task of classifying lesions as malignant or benign. We have also developed a dedicated intelligent workstation for screening mammography, which has been used in our clinical mammography reading area for over a year. We evaluated the dedicated "clinical" workstation using the first 1149 mammographic screening cases. The computer found 6 of the 7 biopsy-confirmed cancers with false-positives rates of 0.9 per image for clustered microcalcifications and 1.4 per image for mass lesions. The significance of this research is that if the detectability of cancers can be increased by employing a computer to aid the radiologist's diagnosis, then the treatment of patients with cancer can be initiated earlier and thus chance of survival improved.

14. SUBJECT TERMS Computer-aided diagnosis, computer vision, digital mammography 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 Limited

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to *stay within the lines* to meet *optical scanning requirements*.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	-	Contract	PR	- Project
G	-	Grant	TA	- Task
PE	-	Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - Leave blank.
NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

__√__ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

_____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

_____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

_____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Maryellen Sigie 3-27-96
PI - Signature   Date

Table of Contents

INTRODUCTION

Nature of the problem

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 the early detection of breast cancer (2-5) and it has been shown that periodic screening of asymptomatic women does reduce mortality (6-11). Various medical organizations have recommended the use of mammographic screening for the early detection of breast cancer (3). Thus, mammography is becoming one of the largest volume x-ray procedures routinely interpreted by radiologists.

It has been reported that between 30 to 50% of breast carcinomas detected mammographically demonstrate clusters of microcalcifications (12-14), although about 80% of breast carcinomas reveal microcalcifications upon microscopic examination (15-18). In addition, studies indicate that 26% of nonpalpable cancers present mammographically as a mass while 18% present both with a mass and microcalcifications (19). Although mammography is currently the best method for the detection of breast cancer, between 10-30% of women who have breast cancer and undergo mammography have negative mammograms (20-24). In approximately two-thirds of these false-negative mammograms, the radiologist failed to detect the cancer that was evident retrospectively (23-26). Low conspicuity of the lesion, eye fatigue and inattentiveness are possible causes for these misses. We believe that the effectiveness (early detection) and efficiency (rapid diagnosis) of screening procedures could be increased substantially by use of a computer system that successfully aids the radiologist by indicating locations of suspicious abnormalities in mammograms.

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 breast lesions exist (20,27), considerable misclassification of lesions occurs with the current methods. On average, only 10-30% of masses referred for surgical breast biopsy are actually malignant (20,28). Surgical biopsy is an invasive technique that is an expensive and traumatic experience for the patient and leaves physical scars that may hinder later diagnoses (to the

extent of requiring repeat biopsies for a radiographic tumor-simulating scar). A computerized method capable of detecting and analyzing the characteristics of benign and malignant masses, in an objective manner, should aid radiologists by reducing the numbers of false-positive diagnoses of malignancies, thereby decreasing patient morbidity as well as the number of surgical biopsies performed and their associated complications.

The development of computer methods to assist radiologists is a timely project in the sense that digital radiography is on the threshold of widespread clinical use. The arrival of digital radiographic systems allows for the acquisition of image data in a format accessible to computerized schemes. The potential significance of this research project lies in the fact that if the detectability of cancers can be increased by employing a computer to aid the radiologist's diagnosis, then the treatment of patients with cancer can be initiated earlier and their chance of survival improved.

The systematic and gradual introduction of computer-assisted interpretation to radiologists that is presented in this proposal is very important in that it allows for a mode of presentation with minimum modification to the current reading habits of radiologists and does not require a "digital" department in which reading must be done from a CRT screen. These two issues are of concern since (1) some radiologists are not comfortable with computer-based methods and (2) primary diagnosis from a CRT display is still controversial. However, the introduction of computer vision to radiologists presented in this proposal is not affected by either concern. In addition, when filmless image acquisition and/or digital (PACS) radiology departments are commonplace in the future, the computer-vision module can be immediately interfaced to electronic, filmless imaging and reading areas.

Background of previous work

In the 1960's and 70's, several investigators attempted to analyze mammographic abnormalities with computers. Winsberg et al. (29), in an early study, examined areas of increased density in contralateral breasts. They felt that their results demonstrated the feasibility for future computer interpretation of mammograms. Spiesberger (30) developed various feature-extraction techniques and a two-view verification method involving medio-lateral oblique and cranio-caudal views to detect

microcalcifications. Kimme et al. (31) developed a computerized method for the detection of suspicious abnormalities in mammograms based on the statistical measures of textural features. They tested their algorithm on 7 patient cases. A similar approach using texture analysis and bilateral comparison was also employed by Hand et al (32) and Semmlow (33) in the computerized localization of suspicious abnormal areas of breasts. Their results yielded a 66% true-positive rate with approximately 26 false suspicious areas per image. With regard to classification methods, Ackerman et al. (34), using digital xeroradiographs, devised four measures of malignancy: calcification, spiculation, roughness and shape, to perform classification on specific areas selected by human observers. The authors viewed their research as only a small step toward the automated reading of xeroradiographs and appeared to discontinue prematurely their computer vision work. The same group (35) did, however, attempt to improve diagnosis by using 36 radiographic properties which were evaluated semi-quantitatively by a radiologist for input to a computer decision tree. Wee et al. (36) and Fox et al. (37) performed preliminary studies on the classification of microcalcifications. These previous studies demonstrated the potential capability of using a computer in the detection of mammographic abnormalities. Their results, however, yielded a large number of false-positives and were based on small data sets.

Computer-aided diagnosis, in general, has attracted little attention during the last decade, perhaps due to the inconvenience involved in obtaining a radiograph in digital format. Recent work, though, shows a promising future. Magnin et al. (38) and Caldwell (39) used texture analysis to evaluate the breast's parenchymal pattern as an indicator of cancer risk. These preliminary studies raised many unanswered questions regarding topics ranging from the digital recording process to the type of numerical risk coefficient employed. Thus, further studies using texture analysis are indicated. The work by Fam and Olson (40,41) on the computer analysis of mammograms is encouraging; however, their method has only been tested on 20 mammographic regions of interest (each roughly half a mammogram). Davies and Dance (42) have reported on their automatic method for the detection of clustered calcifications using local gray-level thresholding and also a clustering rule. Their results yielded a true-positive rate of 96%; however, no indications of the subtlety and size of the calcifications were given. Astley et al. (43), Grimaud et al. (44) and Jin et al. (45) recently reported on their methods

for the detection of breast lesions. Karssemeijer (46) has described a stochastic method based on Bayesian decision theory that appears promising. Lai et al. (47) and Brzakovic et al. (48) are also developing techniques for the detection of mass lesions. The actual performance level and difficulty of the databases, however, are unknown. Gale et al. (49) and Getty et al. (50) 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. (51,52) are developing an expert system to provide visual and cognitive feedback to the radiologist using a critiquing approach combined with an expert system. The system has been demonstrated, though not tested.

We in the Kurt Rossmann Laboratories for Radiologic Image Research at The University of Chicago have vast experience in developing various computer-aided diagnosis (CAD) methods in mammography, chest radiography, and angiography (53-66). We believe that our CAD methods in digital mammography, which include the computerized detection of microcalcifications and masses, have achieved levels of sensitivity and specificity that warrant testing in a clinical environment.

Our detection scheme for clustered microcalcifications includes a preprocessing step referred to as a difference-image approach (53,54). Basically, the original digital mammogram is spatially filtered twice: once to enhance the signal-to-noise ratios of the microcalcifications and a second time to suppress them. The difference between the two resulting processed images yields an image (a difference image) in which the variations in background density are largely removed. Microcalcifications are then segmented from the difference image using global gray-level thresholding and local thresholding techniques. The segmented image is next subjected to feature-extraction techniques in order to remove signals that likely arise from structures other than microcalcifications. An area filter (56), based on mathematical morphology, is used to eliminate small features. Next, each region of interest that contains remaining features is subjected to low-frequency background correction and is characterized by the first moment of its power spectrum, defined as the weighted average of radial spatial frequency over the two-dimensional power spectrum (55). A clustering filter

(57) is next used so that only clusters that contain more than a preselected number of signals within a region of preselected size are retained by the computer. The computerized scheme, using 78 mammograms (39 normal and 39 abnormals) in which most clusters were quite subtle, the scheme yielded a sensitivity of 85% with approximately 2.5 false-positive detections per image (58).

The computerized scheme for detection of clustered microcalcifications (55) developed at The University of Chicago has been tested as an aid to radiologic diagnosis. Using a database of 60 clinical mammograms, half of which contained subtle clusters of microcalcifications, a human observer study was conducted in order to examine the effect of the computer-vision aid on radiologists' performance in a situation that simulated rapid interpretation of screening mammograms. The computer scheme attained an 87% true-positive detection rate with an average of four false-positive clusters per image. The effect of the number of false-positive detections on radiologist performance was also examined by simulating a computer performance level of 87% sensitivity with one false-positive detection per image. Radiologist detection performance was evaluated using ROC (receiver operating characteristic) methodology (68). It was found from the ROC analysis that there was a statistically significant improvement in the radiologists' accuracy when they were given the computer-generated diagnostic information (at either false-positive level), compared with their accuracy obtained without the computer output.

Our scheme for the detection of mammographic masses is based on deviations from the architectural symmetry of normal right and left breasts, with asymmetries indicating potential masses (60,61). The input to the computerized scheme, for a given patient, are the four conventional mammograms obtained in a routine screening examination: the right cranio-caudal (CC) view, the left CC view, the right medio-lateral-oblique (MLO) view, and the left MLO view. After automatic registration of corresponding left and right breast images, a nonlinear subtraction technique is employed in which gray-level thresholding is performed on the individual mammograms prior to subtraction. Ten images thresholded with different cutoff gray levels are obtained from the right breast image, and ten are obtained from the left breast image. Next, subtraction of the corresponding right and left breast images is performed to generate ten bilateral-subtraction images. Run-length

analysis is then used to link the data in the various subtracted images. This linking process accumulates the information from a set of 10 subtraction images into two images that contain locations of suspected masses for the left and right breasts. Next, feature-extraction techniques, which include morphological filtering and analysis of size, shape and distance from border, are used to reduce the number of false-positive detections. Currently, using 150 pairs of clinical mammograms (from 75 cases), the approach achieves a true-positive detection rate of approximately 85% with 3 to 4 false-positive detections per image (62).

We have also investigated the application of artificial neural networks to the detection and classification of mammographic lesions. We used an artificial neural network (ANN) to extract microcalcification image data from digital mammograms (59). The ANN, which was supplied with the power spectra of remaining suspected regions (from the CAD scheme) as input, distinguished actual clustered microcalcifications from false-positive regions and was able to eliminate many of the false positives. Also, we are applying ANNs to the decision-making task in mammography (63). Three-layer, feed-forward neural networks with a back-propagation algorithm were trained for the interpretation of mammograms based on features extracted from mammograms by experienced radiologists. The database for input to the ANN consisted of features extracted from 133 textbook cases and 60 clinical cases. Performance of the ANN was evaluated by ROC analysis. In tests, using 43 initial image features (related to masses, microcalcifications and secondary abnormalities) that were later reduced to 14 features, the performance of the neural network was found to be higher than the average performance of attending and resident radiologists in classifying benign and malignant lesions. At an optimal threshold for the ANN output value, the ANN achieved a classification sensitivity of 100% for malignant cases with a false-positive rate of only 41%, whereas the average radiologist yielded a sensitivity of only 89% with a false-positive rate for classification of 60%.

We are also developing computer-aided methods for the interpretation of digital chest radiographs, such as in the detection of pulmonary nodules, interstitial infiltrates, pneumothorax and cardiomegaly (67,69-75). The computer-vision scheme for the detection of lung nodules is based on a difference-image approach, which (like the one described above for detection of clustered

microcalcifications) is novel in that it attempts to remove the structured anatomic background before applying feature-extraction techniques. After the difference between the signal-enhanced image and the signal-suppressed image is obtained, gray-level thresholding and feature-extraction techniques (involving the size, contrast and shape of the detected features) are performed by the computer to identify the locations of possible nodules. More recently, false-positive detections have been reduced by adding nonlinear filters to the difference-image step and additional feature-extraction techniques based on detailed analyses of the false positives.

The research team in the Rossmann Lab also has considerable experience in evaluation of factors affecting image quality and diagnostic accuracy in digital radiography. We have investigated basic imaging properties including the characteristic system response, spatial resolution properties and noise properties of various types of digital radiographic imaging systems (76-86). The effects of various physical parameters, such as detector system, sampling aperture, pixel size, number of quantization levels, exposure level and display aperture, were examined at various stages of the digital imaging chain (87-91). Knowledge gained in this research will be useful in understanding the effect of spatial resolution and noise on the performance of computer-assisted interpretation.

In developing methods for computer-assisted interpretations, it is crucial to employ appropriate means for evaluation. We have carried out various observer performance studies in comparing the detection capability of new techniques both with regard to simulated and clinical images. 18-alternative forced-choice observer studies were employed to examine the effect of pixel size on the threshold contrast of simple objects digitally superimposed on uniform background noise (92-94) and the effect of structured background on the detectability of simulated stenotic lesions (95). In an observer study with radiologists using clinical images, ROC analysis was employed in order to examine the effects of different display modalities (film and CRT) on diagnostic accuracy in digital chest radiography (96). Similar studies were performed to investigate the effect of data compression ratios on detectability (97), the comparison of computed radiography with conventional screen/film imaging (98), and the utility of computer-assisted interpretation in mammography (55) and chest (71). In addition, we have used ROC and FROC analyses to evaluate the performance level of the computerized schemes and the artificial

neural networks (99). This broad experience will provide the basis for developing similar methodology to evaluate the computer-vision modules for mammography proposed in this application.

Purpose of the present work

The main hypothesis to be tested is that given a dedicated computer-vision module for the computer-assisted interpretation of mammograms, the diagnostic accuracy for `mammographic` interpretation will be improved, yielding earlier detection of breast cancer (i.e., a reduction in the number of missed lesions) and a reduction in the number of benign cases sent to biopsy.

Computer-aided diagnosis (CAD) can be defined as a diagnosis made by a radiologist who takes into consideration the results of a computerized analysis of radiographic images and uses them as a "second opinion" in detecting lesions and in making diagnostic decisions. The final diagnosis would be made by the radiologist. Although mammography is currently the best method for the detection of breast cancer, between 10-30% of women who have breast cancer and undergo mammography have negative mammograms (20-24). It has been suggested that double reading (by two radiologists) may increase sensitivity (100-102). Thus, one aim of CAD is to increase the efficiency and effectiveness of screening procedures by using a computer system, as a "second opinion or second reading," to aid the radiologist by indicating locations of suspicious abnormalities in mammograms.

If a suspicious region is detected by a radiologist, he or she must then visually extract various radiographic characteristics. Using these features, the radiologist then decides if the abnormality is likely to be malignant or benign, and what course of action should be recommended (i.e., return to screening, return for follow-up or send for biopsy). Many patients are referred for surgical biopsy on the basis of a radiographically detected mass lesion or cluster of microcalcifications. On average, only 10-20% of masses referred for surgical breast biopsy are actually malignant (20,28). Thus, another aim of CAD is to extract and analyze the characteristics of benign and malignant lesions in an objective manner in order to aid the radiologist by reducing the numbers of false-positive diagnoses of malignancies, thereby decreasing patient morbidity as well as the number of surgical biopsies performed and their associated complications.

Methods of approach

The objective of the proposed research is to develop a dedicated computer-vision module for use in mammography in order to increase the diagnostic decision accuracy of radiologists and to aid in mammographic screening programs. The computer-aided diagnostic module will incorporate various novel computer-vision and artificial intelligence schemes already under development in the Rossmann Laboratories at the University of Chicago.

The specific objectives of the research to be addressed are:

(1) Further development of advanced computerized schemes for the detection and classification of masses and microcalcifications in digital mammograms. This part of the research involves quantitative analysis of the radiographic characteristics of masses and microcalcifications, and the decision-making processes used by radiologists in making a decision with respect to the likelihood of malignancy and in choosing the appropriate course of action.

(a) Further development of an advanced computerized detection scheme for masses that uses bilateral-subtraction techniques, gray-level thresholding, and analysis of various image features.

(b) Further development of an advanced computerized detection scheme for microcalcifications that uses linear and nonlinear spatial filters, spectral content analysis and various morphological filters for size, contrast and cluster analyses.

(c) Further development of advanced computerized classification schemes for masses and microcalcifications that use computer-vision techniques and artificial-intelligence techniques to calculate a probability of malignancy.

(2) Development of a dedicated module with man-machine interfaces appropriate for the effective and efficient use of the CAD schemes. Final diagnostic decisions will remain with the radiologists.

(a) Optimization of the CAD software.

(b) Examination of various methods of presenting the computer's results to the radiologist.

(c) Development of a prototype intelligent modular workstation using a high-speed (fast CPU & large-capacity memory) computer and a high-resolution, filmless CRT display.

(3) Evaluation of the efficacy and efficiency of the dedicated computer-vision module for mammography using a large clinical database. This part will use both film and filmless media for image acquisition and display.

BODY: Experimental methods and results to date

(1)   Development of the computerized schemes for the detection and classification of masses and microcalcifications in digital mammograms.

Experimental methods

The computerized schemes for detection and classification are at various levels of development. These schemes will be used as aids by radiologists in the interpretation of mammograms. For the development and testing of these algorithms, we will collect 500 mammographic cases from the Department of Radiology.

(a)   Development of the computerized detection scheme for masses.

The computer-vision scheme is based on deviations from the architectural symmetry of normal right and left breasts, with asymmetries indicating potential masses (60-62). Thus, we will continue investigating subtraction techniques as a means to increase the conspicuity of masses in mammograms. These techniques will be combined with analysis of individual mammograms. The input to the computerized scheme, for a given patient, are the four conventional breast images obtained in a routine screening examination: the right CC view, the left CC view, the right MLO view, and the left MLO view. Mammograms will be digitized using a laser scanner digitizer (2K by 2K matrix). In the initial detection stage, the digital image can be reduced to a 512 by 512 matrix (with an effective pixel size of 0.4 mm) due to the large size of masses relative to the pixel size. An automated alignment technique, which we have developed, will be used to align corresponding left and right breast images and also

images of the same breast obtained over some time period. The automated alignment of two corresponding breast images will be performed in three stages: image segmentation, image feature selection and image registration. During image segmentation, the breast area will be isolated from the exterior region using a technique which combines multiple gray-level thresholding and morphological filtering. With image-feature selection, landmarks on each breast image will be determined. These landmarks are the breast border and the nipple position. Since the image features around the nipple often include a thicker skin line and greater subcutaneous parenchymal opacity, a band signature method will be employed to identify the nipple position along the breast border. During image registration, translation and rotation of one of the breast images relative to the other will be determined using a partial-border matching technique.

Once the two images are aligned relative to each other, the detection of possible asymmetries between the border-matched right and left breast images is achieved by correlation of the two mammograms, using a bilateral-subtraction technique. We are investigating linear and nonlinear subtraction methods. With linear subtraction, the two breast images are subtracted (using a left-minus-right convention) and then gray-level thresholding is performed in order to segment the image into possible locations of suspect masses. With the nonlinear technique, gray-level thresholding is performed prior to subtraction. This initial thresholding eliminates some normal anatomic background from further analysis. A selected number of images thresholded with different cutoff gray levels is obtained from the right breast image, and a corresponding number is obtained from the left breast image. Subtraction of ten sets of corresponding right and left breast images, each thresholded at ten different levels, is performed to generate ten bilateral-subtraction images (containing information on suspicious masses in the two original mammograms). A linking process then accumulates the information into two images, called runlength images, where the value of each pixel in each image indicates how often the corresponding location in the set of 10 subtraction images has gray levels above or below a particular cutoff gray value. These images are next thresholded to yield the suspicious areas and submitted for feature extraction.

Feature-extraction techniques will be performed on both the runlength images and the original mammograms to reduce the number of false-positive detections. Initially, a morphological closing operation followed by a opening operation will be used to eliminate isolated pixels and merge small neighboring features. Next a size test will be used to eliminate features that are smaller than a predetermined cutoff size. A border test will be used to eliminate artifact features arising from any border misalignment that occurred during digitization and registration. On the original images, suspected regions will be subjected to region-growing techniques and then examined with respect to size, shape and contrast, in order to eliminate features of elongated shape and diffuse connective tissue.

In addition to comparing the right and left breast images of a given view obtained at a given time, comparisons will be made between images of the same breast obtained at the same projection but at different times in order to note changes in the breast. This follows the methodology employed by mammographers when interpreting a case with previous examinations available. Similar subtraction techniques and feature-extraction methods will be employed. Use of histogram specification methods (103), however, may be necessary in order to match the gray-level distributions of the two images (that were obtained at different times) when there exists a large variation in the exposure techniques employed.

Results to date

Currently, the scheme employs two pairs of conventional screen-film mammograms (the right and left MLO views and CC views), which are digitized. After the right and left breast images in each pair are aligned, a nonlinear bilateral-subtraction technique is employed that involves linking multiple subtracted images to locate initial candidate masses. Various features are then extracted and merged using an artificial neural network in order to reduce false-positive detections resulting from the bilateral subtraction.

The features extracted from each suspected mass lesion include geometric measures, gradient-based measures and intensity-based measures. The geometric measures are lesion size, lesion circularity, margin irregularity, and lesion compactness. The gradient-based measures are the average

gradient (based on a 3 by 3 Sobel operator) and its standard deviation calculated within the specified region of interest. The intensity-based measures are local contrast, average gray value, standard deviation of the gray values within the lesion, and the ratio of the average to the standard deviation. The features were normalized between 0 and 1 and input to the a back-propagation, feed-forward neural network. The ANN's structure consisted of 10 input units, one hidden layer with 7 hidden units and one output unit. In this task, the output unit ranged from 0 to 1, where 1 corresponded to the suspected lesion being an actual mass (i.e., a true-positive detection) and 0 corresponded to the suspected lesion being a false-positive detection (and thus, allowed to be eliminated as a suspect lesion-candidate). Based on the performances of the ANN as a function of iteration, in terms of self-consistency and round robin analyses, the optimal number of training iterations was determined.

ROC (receiver operating characteristic) analysis was applied to evaluate the output of the ANN in terms of its ability to distinguish between actual mass lesions and false-positive detections. The output values from the ANN for actual masses and for false-positive detections were used in the ROC analysis as the decision variable. Basically, the ROC curve represents the true-positive fraction and the false-positive fraction at various thresholds of the ANN output. ROC analysis was used as an index of performance in determining the "optimal" number of input features, the "optimal" number of hidden units, and the "optimal" number of training iterations of the ANN.

In the self-consistency analysis, the ANN achieved an Az of 1.0 and in the round-robin analysis, the ANN achieved an Az of 0.92 in distinguishing actual masses (true positives) from false-positive detections. In an evaluation study using the 154 pairs of clinical mammograms (90 pairs with masses and 64 pairs without), the detection scheme yielded a sensitivity of 95% at an average of 2.5 false-positive detections per image. This was a substantial improvement from the previous year's performance of 85% sensitivity and 4 false-positive detections per image.

We have even further reduced the number of false positives per image by expanding the types of gradient-based measures, and using them in addition to the features discussed above. In the feature extraction stage, the potential lesion was extracted from the parenchymal background using region growing techniques yielding the margin of the suspect mass. The gradient-based measures were

calculated by first processing the region with a 3 by 3 Sobel filter yielding the maximum gradient and the angle of this gradient relative to the radial direction and a fixed (x-axis) at each pixel location. Cumulated gradient-weighted histograms were calculated for the maximum gradients across the various angles. From each histogram, various measures were calculated including full-width at half-maximum, average values, minima, heights, and standard deviations, which gave information such as the amount of spiculation and shape.

A three-level, feed-forward neural network, which utilizes a generalized delta rule in the training, was employed in this study. Fifteen features were chosen from   91 initial features by analyzing the differences in the average and standard deviations of true positives (i.e., actual lesions) and false-positive detections. In addition, receiver operating characteristic (ROC) analysis was used to evaluate the individual performance of each feature in the task of distinguishing true positives from false-positive detections. The fifteen features included the three geometric measures and the three intensity-based measures, as well as nine of the gradient-based measures.

The parameters of the ANN, such as the number of hidden units, the learning rate, and the necessary number of training iterations, were determined empirically by evaluating the performance of the ANN as a function of each of the parameters. Area under the ROC curve was used to indicate performance. Both self consistency and round robin testing was employed (111).

Analysis of the ANN in distinguishing true positives (actual masses) from false positive detections yielded an Az of 0.99 and an Az of 0.97 in the consistency and robin round tests, respectively. This yielded a sensitivity of 90% at less than two false positives per image for the overall mass detection scheme using a database of 110 pairs of digital mammograms containing a total of 102 masses (54 malignant and 48 benign) (112).

Also, a new method for segmentation of the breast region in a mammogram was developed (113). The algorithm identifies unexposed and direct exposure image regions and generates a border surrounding the valid breast region, which can then be used as input for further image analysis and input to the CAD schemes. The program was tested on 740 digitized mammograms with the segmentation results being evaluated by two experts on mammograms and two medical physicists. In 97%

of the mammograms, the segmentation results were rated acceptable for use in computer-aided diagnosis schemes. Segmentation problems encountered in the remaining 22 images (3%) were most often due to digitization artifacts or poor mammographic technique. The developed algorithm is a valuable component of an "intelligent" workstation for computer-aided diagnosis.

Further feature selection was performed using genetic algorithms. A genetic algorithm is an optimization or search method loosely based on natural selection and the survival of the fittest. We used genetic algorithms since we have 91 features to describe true-positive and false-positives detections and we want to   choose a subset of 10-15 features for input to a useful neural network (114). In the genetic algorithm, each string (having chances for mutation, deletion, etc.) represented a set of input features to the ANN. The fitness of each string was defined as the performance of the ANN with that set of input features (using the area under the ROC curve as the performance index) in the task of distinguishing between true-positive and false-positive detections. From FROC analysis, we found that use of the genetic algorithm to select the "optimal" features resulted in the false-positive rate to decrease from 2.6 to 1.5 per image while retaining the sensitivity level of 90%.

We are developing a method based on the Hough spectrum as an effective way to detect spiculated lesions and architectural distortions in digitized mammograms (115). In the Hough spectrum geometric texture analysis technique, the mammogram is analyzed ROI by ROI. Each ROI is transformed into its Hough spectrum and then thresholding is performed with its threshold level based on the statistical properties of the spectrum. ROIs with strong signals of spiculation are then screened out as regions of potential lesions. In a preliminary study, 32 images containing spiculated lesions/architectural distortions (biopsy confirmed) were analyzed using information extracted from the Hough spectrum. Our preliminary studies, using only the Hough spectrum based technique without further feature analyses to reduce false positives, yielded sensitivities of 81% for spiculated masses and 67% for architectural distortions at false positives rates of 0.97 and 2.2 per image, respectively. The results are promising and we expect the false positive rate to decrease upon the incorporation of feature analysis into the overall detection scheme, as we have seen with our other detection methods.

    We are also developing another single-image method for detection of small invasive breast

cancers (116). Localized density peaks on mammograms are identified using a specially designed

gradient filter. Lesion contours are generated by matching a deformable template onto a second

derivative edge map. In a preliminary study (without further feature analyses to reduce false

positives) using 45 non-palpable invasive breast cancers, all with a size less than 1 cm (median size of

7 mm), 82% of the cancers were detected with an average false-positive rate of 2.8 per image. We

expect both the sensitivity and specificity to increase with improved feature analyses.

(b)   Development of the computerized detection scheme for microcalcifications.

    Microcalcifications are a primary indicator of cancer and are often visible in the mammogram before

a palpable tumor can be detected. Initially, clinical screen/film mammograms will be digitized using the

laser scanner and analyzed in the 2048 by 2048 matrix format in order to retain the high spatial-

frequency content of the microcalcifications. First, the original mammograms will be processed to

enhance and suppress the signal of the microcalcifications, followed by calculation of a difference

image. Both linear and nonlinear filters will be investigated for enhancement and suppression.

Previous use of both linear and nonlinear filters in detecting lung nodules in digital chest images has

shown that while both types of filters tended to detect nodules, locations of false positives differed.

Thus, a combination of the results from each processing technique has the potential to yield high

sensitivity and reduce the number of false-positive detections. Examples of filters for signal

enhancement include a linear "matched" filter that matches the profile of a typical microcalcification and

a morphological open filter (to enlarge the appearance of microcalcifications). Morphological filtering

(104) is basically a nonlinear filtering method that calculates the logical AND (erosion function) or OR

(dilation function) of pixels within a kernel of some given size and shape. When extended to gray-scale

images, the logical AND and OR operations can be replaced by minimum and maximum operations. By

appropriately choosing the size and shape of the kernels, as well as the sequences of the AND and the

OR, the filters can eliminate groups of pixels of limited size or merge neighboring pixels. Examples of

filters for signal suppression include ring-shaped filters that yield either the average or median value of the surrounding normal anatomic background (54).

The difference image will then be subjected to various feature-extraction techniques to reduce further the number of false-positive detections. These techniques will test for size, contrast and spectral content of neighboring features. New methods for analyzing these features will involve the use of morphological filters. For example, we have found that the use of asymmetric morphological filters to eliminate features less than 3 pixels in size are more effective and efficient than use of a point-by-point analysis that involves counting the number of pixels in each remaining feature and comparing it to a size cutoff. In addition, the presence of clustering of the microcalcifications will be examined since singular microcalcifications are usually not cancerous. The morphological kernel for the clustering test will correspond to the size of a typical cluster (approximately 6 mm in diameter).

Results to date

The microcalcification detection scheme consists of three steps. First, the image is filtered so that the signal-to-noise ratio of microcalcifications is increased by suppression of the normal background structure of the breast. Second, potential microcalcifications are extracted from the filtered image with a series of three different techniques: a global thresholding based on the grey-level histogram of the full filtered image, an erosion operator for eliminating very small signals, and a local adaptive grey-level thresholding. Third, some false-positive signals are eliminated by means of a texture analysis technique, and a nonlinear clustering algorithm is then used for grouping the remaining signals.

In our computer detection scheme it is neccesary to group or cluster microcalcifications, since clustered microcalcifications are more clinically significant than are isolated microcalcifications. In the past we used a "growing" technique in which signals (possible microcalcifications) were clustered by grouping those that were within some predefined distance from the center of the growing cluster. In this research, we introduced a new technique for grouping signals, which consists of two steps (117). First, signals that may be several pixels in area are reduced to single pixels by means of a recursive transformation. Second, the number of signals (non-zero pixels) within a small region, typically

3.2x3.2 mm, are counted. Only if three or more signals are present within such a region are they preserved in the output image. In this way, isolated signals are eliminated. Furthermore, this method can eliminate falsely detected clusters, which were identified by our previous detection scheme, based on the spatial distribution of signals within the cluster. The differences in performance of our CAD scheme for detecting clustered microcalcifications using the old and new clustering techniques was measured using 78 mammograms, containing 41 clusters. The new clustering technique improved our detection scheme by reducing the false-positive detection rate while maintaining a sensitivity of approximately 85%.

We also applied artificial neural networks to the differentiation of actual "true" clusters of microcalcifications from normal parenchymal patterns and from false positive detections as reported by a computerized scheme. The differentiation was carried out in both the spatial and spatial frequency domains (59). In the spatial domain, the performance of the neural networks was evaluated quantitatively by means of ROC analysis. We found that the networks could distinguish clustered microcalcifications from normal nonclustered areas in the frequency domain, and that they could eliminate approximately 50% of false-positive clusters of microcalcifications while preserving 95% of the positive clusters.

The number of false-positive detections was even further reduced when a shift-invariant artificial neural network (SIANN) was used to analyze the remaining suspected locations (118). The SIANN is a multilayer back-propagation neural network with local, shift-invariant interconnections. The advantage of the SIANN is that the result of the network is not dependent on the locations of the clustered microcalcifications in the input layer. The performance of the SIANN was evaluated by means of a jack-knife method and ROC analysis using a database of 168 regions as reported by the CAD scheme. Approximately 55% of the false positives were eliminated without loss of any of the true-positive detections. This technique led to a performance of 85% sensitivity with less than 0.6 false-positive detections per image. In this study, we also examined the effect of the network structure on the performance of the SIANN.

Modifications were made to improve the performance of the SIANN (119). First, the preprocessing was removed because the result of background-rend correction is affected by the size of ROIs. Second, image-feature analysis was employed to the output of the SIANN in an effort to eliminate more of the false detections. In order to train the SIANN to detect microcalcifications and also to extract image features of microcalcifications, zero-mean-weight constraint and training-free-zone techniques were developed. A cross-validation training method was also applied to avoid the over-training problem. The performance of the SIANN was evaluated by means of ROC analysis using a database of 39 mammograms for training and 50 different mammograms for testing. The analysis yielded an average area under the ROC curve (Az) of 0.90 for the testing set. Approximately 62% of false-positive clusters detected by the rule-based scheme were eliminated without any loss of the true-positives clusters by using the improved SIANN with image feature analysis techniques.

(c)   Development of computerized classification schemes.

Various feature-extraction techniques and artificial intelligence schemes will be investigated in order to distinguish malignant masses and/or microcalcifications from benign masses and/or microcalcifications. The database for this investigation will be obtained from the conventional four screening breast images, as well as special views such as spot compression.

In our previous work, we compiled a list of features that radiologists use in distinguishing between malignant and benign masses. These features include: margin spiculation (number of spiculations, length of spiculation, and difference between spicules and local linear features), shape (linear to spherical, geometrical to diffuse, and existence of satellite lesions), size (mean diameter), margin characteristics (complete to inseparable from surround, well-defined to indistinct, and presence of halo sign), and pattern of interior (uniformity, presence of well-defined lucencies, and opacity relative to size). The analysis of spiculation will be based on a novel computer-vision method involving the Fourier analysis of the fluctuations around the margin of the mass in question (60). The computer-extracted margin used in the analysis for spiculation also contains information related to the number and length of spiculations. Also, prior to the analysis of spiculation, the mass is extracted from the normal

anatomic background of the breast parenchyma. Currently, region-growing techniques are employed for this extraction. Once extracted, the shape and size of the mass can be easily calculated. The size will be defined as the effective diameter of a circle that has the same area as the extracted mass. The shape will be expressed by a degree of circularity, which will be defined as the ratio of the area of the mass within the equivalent circle to the total area of the mass. Masses with ill-defined margins are more likely to be malignant than those with relatively well-defined margins. Thus, a margin gradient test will be developed to measure the sharpness of the margin. This sharpness will be defined as the degree of density change across the margin and will be measured perpendicular to the margin at all points along the margin. The pattern of the interior will be quantitatively determined from the spectral content of the interior.

Features related to the classification of microcalcifications include: the shape of the individual microcalcifications (rounded to irregular, linear, and branched), uniformity of microcalcifications within a cluster (uniformity in size, shape, and density), distribution of the microcalcifications (diffuseness and shape of cluster) and presence of macrocalcifications. The size and shape of the individual microcalcifications will be determined by the computer using an effective diameter and a circularity measure, respectively, as described earlier. Uniformity within a cluster will be assessed by calculating the spread of values for a particular characteristic such as size. Once a cluster has been defined, its diffuseness will be given by the number of microcalcifications per unit area and its shape will be defined using a circularity measure.

These various computer-determined quantitative measures describing the mass or cluster of microcalcifications in question will be input to an artificial neural network that will merge the features into a probability of malignancy for use by radiologists. As mentioned in the Background section, our work with a neural network in merging human-reported mammographic features into a malignant/benign decision has been extremely promising. The input data (corresponding to the computer-extracted features of the masses and microcalcifications) will be represented by numbers ranging from 0 to 1 and will be supplied to the input units of the neural network. The output data from the neural network is then provided from output units through two 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. Two different basic processes are involved in a neural network; namely, a training process and a testing process. The neural network will be trained by a back-propagation algorithm (105) using input data (i.e., computer-reported features) and the desired corresponding output data (i.e., biopsy or follow-up proven truth of the malignant or benign status of the mass or microcalcifications in question), for a variety of cases. Once trained, the neural network will accept computer-reported features of the mass or microcalcifications in question and output a value from 0 to 1 where 0 is definitely benign and 1 is definitely malignant. Based on the distribution of these values for various known cases, we will be able to determine what course of action (e.g., biopsy, follow-up or return to normal screening) should be recommended to the radiologist.

Results to date
Classification of masses

Our earlier work showed that a back-propagation, feed-forward artificial neural network could merge human-extracted features of mammographic lesions into a likelihood of malignancy at a similar level of that of an expert mammographer. In the study presented here, however, ANN is used to merge computer-extracted features of mass lesions into a likelihood of malignancy.

The method takes as input the center location of a mass lesion in question. Next, the lesion is segmented from the breast parenchyma (background) using an automatic region growing technique and various features of the lesion are extracted. The automatic lesion segmentation involves the analysis of the size of the grown region as a function of the gray-level interval used for the region growing. Many of the extracted features are determined from a cumulative edge-gradient-orientation histogram analysis modified for orientation relative to a radial angle (120). Input to an ANN consists of four features from the gradient analysis along with the average gray value within the grown lesion. The gradient measures include the FWHM (full width at half max) of the cumulative edge-gradient-orientation histogram calculated from pixels within the lesion and its neighboring surround, and from

just pixels along the lesion margin (see reprint in the appendix). These measures correspond to the presence of spiculation, which is a sign of malignancy in the visual interpretation of mammographic masses. The ANN's structure consisted of 5 input units, one hidden layer with 4 hidden units and one output unit. In this task, the output unit ranged from 0 to 1, where 1 corresponded to the lesion being malignant and 0 corresponded to the lesion being benign. Use of ROC analysis with self-consistency testing and round-robin testing was employed as discussed in the previous section. We found that using a combination of the measurements from the four neigborhoods is superior in the classification of mammographic mass lesions.

We have also incorporated additional features of masses into the computerized classification scheme (121). The classification method was evaluated using a pathologically-confirmed database of 95 masses (57 malignant and 38 benign), of which all but one had been sent to biospy. The mammograms in the database had been digitized to a pixel size of 0.1 mm. Various features related to the margin, shape and density of each were extracted automatically from the neighborhoods of the computer-identified mass regions. Selected features were merged into an estimated probability of malignancy using three different automatic classifiers. The performance of the three classifiers in distinguishing between benign and malignant masses were evaluted by ROC analysis and compared with those of an experienced mammographer and five general radiologists. The computerized classification scheme yielded an Az value of 0.94, similar to that of an experienced mammographer (Az=0.90) and substantially higher that the average performance of the general radiologists (Az=0.81). With the database we have, the computer scheme achieved a positive predictive value of 83% at 100% sensitivity, which was 12.1% higher than that of the experienced mammographer and 21.5% higher than that of the average performance of the general radiologists at a p-value < 0.001. We found that use of a rule based on spiculation prior to use of ANN (i.e., a hybrid system) was superior to use of just an ANN in merging the various features. The reason for this was that spiculation is a dominant rule used by both the computer and radiologists in distinguishing between malignant and benign masses. Thus, when one has a limited database, it appears beneficial to first

use well-known rules prior to ANN training. The computerized classification scheme is expected to be useful in helping radiologists distinguish between benign and malignant masses.

Results to date
Classification of microcalcifications

The analysis of microcalcifications can be difficult to perform consistently for human observers leading to the poor positive predictive value. We have been investigating methods to identify computer-extracted quantitative features of microcalcifications and their clusters that can be used to classify malignant and benign clustered microcalcifications, and, to examine if a computer can make accurate differential diagnoses based on computer-extracted features. In this study, features of the microcalcifications and their clusters were automatically extracted from digitized conventional mammograms.

The microcalcifications were segmented using the following method, which is described in detail elsewhere. A third-degree polynomial was fitted to the pixel-value distribution in a ROI (region of interest) of the digitized mammogram in both horizontal and vertical directions to reduce the background structure of the breast parenchyma. The microcalcification was then delineated by region growing. The effective thickness of the microcalcification (physical dimension along x-ray projection line) was estimated from signal contrast (mean pixel value above background) of the isolated microcalcification. This was done by first converting signal contrast in terms of optical density to contrast in terms of exposure using knowledge of the H&D curve of the screen-film system, and secondly converting contrast in terms of exposure to physical dimension using the exponential attenuation law assuming a "standard" model of the breast and the microcalcification. The standard model assumes (i) a 4-cm compressed breast composed of 50% adipose and 50% glandular tissues; (ii) a microcalcification composed of calcium hydroxyapatite with physical density of 3.06 g/mm^3; and (iii) a 20-keV monochromatic x-ray beam. Two contrast corrections were applied for better

accuracy: compensation for blurring caused by the screen-film system and the digitization process, and compensation for x-ray scatter.

The usefulness of the features were evaluated using the distributions of the benign and malignant populations. Features capable of showing separation between benign clusters from the malignant population were chosen for the automated classification. Extracted features were based on the size, shape, contrast, and uniformity of individual microcalcifications; and the size and shape of microcalcification clusters. An artificial neural network was used to classify benign versus malignant clusters of microcalcifications using 8 computer-extracted features. The database consisted of 100 images, digitized at 100-mm pixel size and 10-bit grey-scale resolution, from 53 patients biopsied for suspicion of breast cancer based on clustered microcalcifications. The neural network correctly identified 69% of the benign patients, all of whom had biopsies, and 100% of the malignant patients.

An observer study was performed which indicated that for the cases used, the performance of the computer method was statistically higher than that of five radiologists (122). Comparison between ROC curves (computer and radiologists) was done using a new partial area index. In clinical practice, operating ranges with low sensitivity are unacceptable, and the portion of the ROC curve at high sensitivity above a preselected threshold is most important. Thus, in the comparison the portion of the area under each ROC curve above a sensitivity of 0.90 was calculated. A partial area of 0.082 was calculated for the computerized method, whereas a partial area of 0.042 was calculated for the five radiologists. Results of the Student t test for paired data showed this difference to be statistically significant (p = 0.03). This computerized classification technique is expected to be helpful to radiologists in reducing the number of false-positive biopsy findings.

(2) Development of a dedicated CAD module for use by radiologists.

Experimental methods

The various computer-vision and artificial intelligence schemes will be incorporated into a dedicated computer system (module) equipped with a high-speed computer and a digital image interface. The digital image interface will initially be to a film digitizer in order to test the CAD schemes using the large database of clinical mammograms available in our Radiology Department. Later, mammographic images will be obtained using the CR system or the CCD-based digital biopsy unit. The intelligent modular workstation will need to have sufficient computer power (CPU and large capacity memory) and display capabilities to allow for "real-time" computation and viewing of the computer-vision results. Thus, we plan to upgrade our current computer hardware and optimize our software to achieve high-speed and efficient computation of CAD results. Our target is to reduce the CPU time required for CAD computations from the current level of about 5 minutes per image to a few seconds. Also, appropriate man-machine interfaces will be needed for effective and efficient computer-assisted interpretations. This part of the research will involve the examination of various methods of presenting the computer-determined results to the radiologists. Important parameters include (a) the shape and size of the markers of the computer output that could represent the severity or confidence level (probability) of the lesion, (b) the optimal operating point of the CAD schemes (high sensitivity with an acceptable number of false positives), (c) the timing and duration of displaying the computer output, (d) the selection of the minimum number of inputs required for radiologists and (e) the user-friendliness of instructions and input entries.

The development of the prototype modular system will be achieved in stages. In Phase 1, the introduction of the computer-vision aid to the radiologists will be implemented with minimum change in the current radiologist method of operating. This will allow for a gradual introduction in order to minimize any resistance to change. Thus, only computer-reported <u>detection</u> results will be presented to the radiologist, leaving all of the interpretation to the radiologist. Basically, the computer will serve as a "second opinion" indicating suspicious areas without critique as to their degree of malignancy. Original

films will be digitized (2048 by 2048 digitization matrix) and analyzed, with the computer output then printed on either film or thermal paper. Radiologists will perform their normal reading using the original image and the computer results. It is believed that this introduction of CAD to radiologists will cause minimum modification to their normal reading patterns, thus allowing for a smooth and effective transition. During Phase 2, results from the classification schemes also will be included, using the methodology described for Phase 1. However, in this second phase the computer will serve as a "second opinion" for both the location and the interpretation of breast lesions.

During the first two phases, we will investigate the best markers for use by radiologists, who may prefer arrows or circles (icon-type symbols). It should be noted that the implementation of computer vision in mammographic screening using the methods described above is not limited to fully digital (PACS) departments but can be incorporated in a general film-based radiology department or in a mobile, filmless mammography unit (i.e., a limited PACS environment).

Once the use of computer vision is shown to be useful, beneficial and efficient, we will incorporate high-resolution, state-of-the-art monitors into the dedicated computer system (Phase 3). The "intelligent" module will be interfaced to our department's RIS (radiology information system) to link the demographic and medical history information with the CAD output. In order for the radiologist to examine the entire breast image, the display monitor will need to have 2K by 2K capability. In mammography, each breast image usually can be digitized adequately into a 2K by 1K image. Thus, in order to view all four breast images (left and right CC views and left and right MLO views), two high-resolution 2K by 2K monitors are needed. However, in the practice of radiology, films (images) from previous examinations play an important role in the current exam due to the need for comparison in order to detect subtle changes. Thus, the display requirements are four 2K by 2K monitors (in a 2 by 2 arrangement), allowing the top two monitors to be used for sequencing through previous exams of the patient in question. In this phase, the radiologists will do their reading of the mammographic cases from the high-resolution monitors. Due to the dynamic nature of the display, the computer-reported results can be presented in a toggle format where the radiologist can press a button to either show or remove the computer-reported results. In addition, the computerized schemes can be configured to

allow for the radiologist to control the tradeoff between the sensitivity and specificity of the computer output, because more true-positive detections always can be achieved at the cost of a larger number of false-positive findings, and vice versa. This tradeoff would be adjusted by the radiologist, depending on the nature of the case material and personal preference. For example, a radiologist might choose a computer output with high sensitivity for examining high-risk patients, whereas a lower sensitivity and correspondingly lower false-positive rate might be preferred for patients at low risk for cancer. It should be noted, however, that increasing the number of interactive choices available to the radiologist will lengthen the reading time per case. Therefore, we will investigate optimization of the module's human interface by studying the relationship between achievable diagnostic accuracy and required reading time.

Results to date

The computerized image analysis software has been integrated into a user friendly interface based on UNIX, XWINDOWS and Motif and operated on an IBM RISC 6000 Series 570 computer workstation. The prototype (hardware & software) was demonstrated at the 1994 annual meeting of the Radiological Society of North America (RSNA) and was well received by the many radiologists in attendance. Currently, arrows (red for masses and yellow for clustered microcalcifications) are used to indicate the computer-detected location of lesions. The input to the system can be either a film that is digitized and then analysed automatically or a computer file containing a digital image. The prototype system is interfaced to a Konica laser film digitizer which enables digitization of the mammograms to approximately 2K by 2K matrices. Video output of the IBM monitor is connected to a low-resolution thermal printer (approximately 1K by 1K) for hardcopy reporting of the CAD results.

Our prototype workstation was placed in the clinical mammography reading area of the Department of Radiology. Since Nov. 8, 1994, we have analyzed over 5000 screening cases. Results are discussed in the next section.

(3)   Evaluation procedure using large clinical databases

Experimental methods

As described in the previous section, the computer-vision methods for mammography will be developed in phases. Plans include testing the computer-vision system at the end of each phase in order to demonstrate the effect of the various modes of presentation on the accuracy, efficiency and acceptability of the mammographic aid. The system will be evaluated using clinical mammograms obtained from both a low-risk population and a high-risk population. The low-risk population will be obtained from The University of Chicago mammography screening program. The high-risk population will be drawn from examinations referred to our Department of Radiology, since The University of Chicago serves as a tertiary medical center. Initially, performance studies will be done using a database of preselected mammographic cases that have a distribution of subtle cases of normal, benign and malignant areas of either masses or microcalcifications. Later studies will be performed using a more representative database of consecutive mammographic cases obtained from four weeks worth of screening. "Truth" concerning the presence and malignancy of masses and microcalcifications will be established with the aid of expert mammographers, follow-up reports and surgical biopsy reports. Normal cases will be selected from patients who have had normal follow-up exams. Performance studies will be done using cases involving the four conventional mammograms (left and right CC views, and left and right MLO views), since these are the usual images obtained in screening.

At the detection stage of the computer-vision system, performance will be examined by calculating the fraction of lesions detected (true-positive rate) and the number of falsely-reported areas per case. At the classification stage of the computer-vision system, performance will be examined by calculating the fraction of malignant cases correctly classified (true-positive classification rate) and the number of benign cases that are reported by the computer as being malignant (false-positive classification rate). The clinical database for these performance evaluations will contain 180 cases (60 normal, 30 with benign masses, 30 with malignant masses, 30 with benign microcalcifications, and 30 with malignant microcalcifications).

Observer studies will be performed to examine the usefulness of the computer-assisted interpretation process in enhancing radiologists' performance levels, as compared to the unaided performance by radiologists. During phases 1 and 2, the database cases will be printed with the computer-vision results on each film. These database cases will then be used in observer performance studies. Stratified sampling (106) will be used in choosing subtle cases in order to avoid problems associated with either "too easy" or "too difficult" cases (107). Twelve attending radiologists and senior residents will act as observers. Then, for the 180 cases in the database, three "reading methods" will be tested; (a) the original cases without the computer-vision aid, (b) the cases with the detection-results reported (phase 1 computer locations of suspicious areas) and (c) the cases with both the detection and classification results reported (phase 2 computer locations with probability of malignancy). Each observer will be asked to perform two tasks: (1) locate and rate suspicious areas as to the presence of an abnormality (rating scale of 0 to 100) and (2) indicate an overall level of certainty as to the presence of cancer using a 5-point rating scale where 1=definitely benign and 5=definitely malignant. This five-point scale is the same as that being recommended by the American College of Radiology for routine use by clinical mammographers. The dual-task observer study will allow for evaluation of the utility of both the computer-vision detection and classification results. (In addition, questionnaires will be given to each observer in order to obtain subjective information with regard to the efficiency and acceptability of the computer-vision mammography system.) In the analysis of the observer study results, maximum likelihood estimation (108) will be used to fit a binormal ROC (receiver operating characteristics) curve to each observer's confidence-rating data from each diagnostic method. The index Az, which represents the area under a binormal ROC curve, will be calculated for each fitted curve. To represent the average performance of the observers for each diagnostic method, the composite ROC curves will be calculated by averaging the slope and intercept parameters of the individual observer-specific ROC curves. The statistical significance of apparent differences between pairs of diagnostic methods will then be analyzed by applying a "two-tailed" t-test for paired data to the observer-specific Az index values.

Free-response ROC (FROC) analysis (109) and FROC-AFROC analysis (110) will be used in analyzing the data pertaining to localization of the abnormality. The ordinates of both FROC curves and AFROC curves are the fraction of lesions (masses or microcalcifications) that are correctly localized by the observer. However, the abscissa of an FROC curve is the average number of false positives per image, whereas the abscissa of an AFROC curve is the probability of obtaining a false-positive image (i.e., an image containing one or more false-positive responses).

After phase 3, another observer study will be performed in which four weeks' worth of mammographic cases will collected and interpreted by six radiologists with and without the computer-vision results of location and classification. Although this database lacks the control over the subtlety of the cases that the earlier mentioned study has, it represents a more typical clinical situation. Half of the radiologists will read the first two weeks of cases without aid and the second two weeks of cases with the mammographic aid; and the other half of the radiologists will read the first two weeks of cases with the aid and the second two weeks of cases without the aid. Rating methods and analyses will be the same as mentioned above.

Results to date

FROC analysis and ROC analysis have been used extensively for the intermediate testing results of the various detection and classification methods. Constant collection of the database is ongoing. Investigators have developed a case reporting sheet for organizing the new cases on a Macintosh computer using FileMakerPro software. The various databases being collected include pathologically-proven mass and clustered microcalcification cases. In addition, a "missed lesion" database is being digitized in order to test the detection methods in the upcoming grant period. This database includes lesions that were seen in retrospect, i.e., after the cancer was detected at a later date. This database will demonstrate the ability of the detection schemes to increase the sensitivity of detection in a screening program. In a preliminary study (presented at the RSNA 94) in which 26 "missed lesion" cases were analyzed, the computerized detection schemes achieved a sensitivity of 50%. (Note that

these "missed lesion" cases can be thought of yielding a sensitivity of 0% when they had been read by the radiologists).

We have been tabulating the performance of the clinical intelligent mammography workstation. In this prospective study, the results of the computer output have been quite promising (123). Since the study is prospective, we do not know all "truth" yet, although we are currently following the workups and biopsy results. Approximately 70% of the cases deemed suspicious by the study radiologist have been detected by the computer. In two cases, a cluster of microcalcifications was located by the computer but not by the radiologists. A large number of screening cases need to be analyzed by the workstation prior to assessment of its performance and contribution in the mammographic interpretation process, since with screening mammography, only 5 to 10 cancers are found for every 1000 patients.

We are currently collecting follow-up data on the patients with abnormal mammograms. Follow up has been performed on the first 1149 screening cases (124). These screenees resulted in 154 abnormal interpretations and in the detection of six cancers. The sensitivity of the computer schemes to detect cancer was 83.3% with a false positive rate of 0.91 false clusters and 1.4 false masses per image. Many of the false clusters are due to calcified vascular structures and many of the false masses are due to nodular-like structures. We found that the study radiologist can easily learn to recognize typical false positives and disregard them in her assessment of the presence of a lesion.

Since April of 1995, the intelligent workstation has been used routinely by the attending mammographers in the clinical reading area of the Department of Radiology. Over 5,000 screening cases have been analyzed by the computer. The radiologists perform their initial interpretation of the mammographic case and then look at the computer results that are printed on thermal paper. The radiologists rate the case from -2 to +2, with +2 meaning the computer output was quite useful. The weekly average rating given to the cases (over both normal and abnormal mammograms) so far ranges from 0.11 to -0.73. Note that for normal mammograms, the highest rating that can be given is zero, and that most of the cases read are normal (for a screening population). For normal mammograms, -1 or -2 are given for too many false positive detections. For abnormal cases, -1 or -2 are given if the

computer fails to point to a location deemed suspicious by the attending radiologist -- whether it is malignant or not. We are currently, analyzing the rating separately for normal and abnormal cases; however it is important to accumulate a sufficient number of cases due to the low incidence of abnormal mammograms in a screening population.

CONCLUSIONS

Substantial improvements in the performances of the computer-aided diagnosis methods for the detection of masses and clustered microcalcifications have been achieved during the past funding period. For the detection of masses, the sensitivity remained constant, while the false-positive rate per image reduced to less than 2 per image. For the detection of clustered microcalcifications, the false-positive rate was reduced from 2 per image to approximately 0.7 per image, without loss in sensitivity. Constant collection of the database is ongoing. Investigators have developed a case reporting sheet for organizing the new cases on a Macintosh computer using FileMakerPro software. The various databases being collected include pathologically-proven mass and clustered microcalcification cases. Databases for both mammograms containing mass lesions and mammograms containing microcalcifications have both increased in size and some have been digitized on more than one digitizer in order to observe the effect of digitization on detection performance. In addition, a "missed lesion" database is being digitized in order to test the detection methods in the upcoming grant period. This database includes lesions that were seen in retrospect, i.e., after the cancer was detected at a later date. This database will demonstrate the ability of the detection schemes to increase the sensitivity of detection in a screening program. In a preliminary study (presented at the RSNA 94) in which 26 "missed lesion" cases were analyzed, the computerized detection schemes achieved a sensitivity of 50%. (Note that these "missed lesion" cases can be thought of yielding a sensitivity of 0% when they had been read by the radiologists).

With regard to the classification of mammographic lesions as an aid in distinguishing between malignant and benign cases, the initial performances for both masses and microcalcifications has been quite promising. In the classsification of masses, an Az (area under the ROC curve) of 0.90 was obtained from the ROC analysis of the output from the neural network, which was used to merge the

extracted features of the lesions. In the classification of clustered microcalcifications a neural network correctly identified 69% of the benign patients, all of whom had biopsies, and 100% of the malignant patients. We conclude that a computer is capable of distinguishing benign from malignant clustered microcalcifications even at 100-mm pixel size.

The computerized image analysis software has been integrated into a user friendly interface based on UNIX, XWINDOWS and Motif and operated on an IBM RISC 6000 Series 570 computer workstation. The prototype (hardware & software) was demonstrated at the 1994 annual meeting of the Radiological Society of North America (RSNA) and was well received by the many radiologists in attendance. The input to the system can be either a film that is digitized and then analysed automatically or a computer file containing a digital image. The prototype system has been used in the clinical reading area since November, 1994, with the attending radiologists using it routinely since April, 1995.

We are very optimistic about the continuing success of our research. We will continue to improve the detection and classification performance of our algorithms. The mammographers in the clinical reading area of the department are pleased with the prototype. Weekly meetings are held between the basic science and clinical researchers in order to ensure a smooth integration of the workstation in the clinical arena. The results with the clinical prototype are promising.

REFERENCES

1.   Silverberg E, Boring CC, Squires TS: Cancer Statistics, 1990. CA 40: 9-27, 1990.
2.   Tabar L, Dean PB: Basic principles of mammographic diagnosis. Diagn. Imag. Clin. Med. 54: 146-157, 1985.
3.   American Cancer Society: CA Cancer J Clin 33: 255, 1983.
4.   Baker L: CA Cancer J Clin 32: 194, 1982.
5.   NCRP Report No. 85: Mammography (National Council on Radiation Protection: Washington, D.C., 1986).
6.   Shapiro S, Venet WS, Strax PH, et al.: Ten to fourteen-year effect of screening on breast cancer mortality. JNCI 69: 349-355, 1982.
7.   Verbeek ALM, Hendricks JH, Holland R, et al: Reduction of breast cancer mortality through mass screening with modern mammography. Lancet 1: 1222-1224, 1984.
8.   Collette HJA, Day NE, et al.: Evaluation of screening for breast cancer in a non-randomized study (the DOM project) by means of a case-control study. Lancet 1: 1224, 1226, 1984.
9.   Tabar L, Gad A, Holmberg LH, et al.: Reduction in mortality from breast cancer after mass screening with mammography. Randomized trial from the Breast Screening Working Groups of the Swedish National Board of Health and Welfare. Lancet 1: 829-832, 1985.
10.   Andersson I, Aspegren L, Janzon L, et al.: Mammographic screening and mortality from breast cancer: The Malmo mammographic screening trial. Br. Med. J. 297:943,1988.
11.   Feig SA: Decreased breast cancer mortality through mammographic screening: Results of clinical trials. Radiology 167: 659-665, 1988.
12.   Black JW, Young B: A radiological and pathological study of the incidence of calcifications in diseases of the breast and neoplasms of other tissues. Br. J. Radiol. 58: 596-598, 1965.
13.   Wolfe JN: Analysis of 462 breast carcinomas. AJR 121: 846-853, 1974.
14.   Sickles EA: Mammographic detectability of breast microcalcifications. AJR 139: 913-918, 1982.
15.   Fisher ER, Gregorio RM, Fisher B, et al.: The pathology of invasive breast cancer. Cancer 36: 1-84, 1975.
16.   Millis RR, Davis R, Stacey AJ: The detection and significance of calcifications in the breast: A radiological and pathological study. Br. J. Radiol. 49: 12-26, 1976.
17.   Murphy WA, DeSchryver-Kecskemeti K: Isolated clustered microcalcifications in the breast: Radiologic-pathologic correlation. Radiology 127: 335-341, 1978.
18.   Muir BB, Lamb J, Anderson TJ: Microcalcification and its relationship to cancer of the breast: Experience in a screening clinic. Clin. Radiol. 149: 193-200, 1983.
19.   Sickles EA: Mammographic features of 300 consecutive nonpalpable breast cancers. Am J Rad 146: 662-663, 1986.
20.   Bassett LW, Gold RH: Breast cancer detection: Mammography and other methods in breast imaging, Grune and Stratton (New York), 1987.
21.   Baines CJ, Miller AB, Wall C, McFarlane DV, et al.: Sensitivity and specificity of first screen mammography in the Canadian National Breast Screening Study: A preliminary report from five centers. Radiology 160: 295-298, 1986.
22.   Pollei SR, Mettler FA, Bartow SA, Moradian G, Moskowitz M: Occult breast cancer: Prevalence and radiographic detectability. Radiology 163: 459-462, 1987.
23.   Andersson I: What can we learn from interval carcinomas? Recent Results in Cancer Research 90: 161-163, 1984.
24.   Martin JE, Moskowitz M, Milbrath JR: Breast cancers missed by mammography. AJR 132: 737, 1979.
25.   Buchanann JR, Spratt JS, Heuser LS: Tumor growth, doubling times, and the inability of the radiologist to diagnose certain cancers. Radiologic Clinics of North America 21: 115, 1983.
26.   Holland T, Mrvunac M, Hendriks JHCL, et al.: So-called interval cancers of the breast. Pathologic and radiographic analysis. Cancer 49:2527, 1982.
27.   Tabar L, Dean PB: Teaching Atlas of Mammography, George Thieme Verlag (Stuttgart, New York), 1983.

28. Moskowitz M: Screening for breast cancer: How effective are our tests? A critical review. <u>Ca-A Cancer Journal for Clinicians</u> 33: 26-39, 1983.
29. Winsberg F, Elkin M, Macy J, Bordaz V, Weymouth W: Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. <u>Radiology</u> 89: 211-215, 1967.
30. Spiesberger W: Mammogram inspection by computer. <u>IEEE Transactions on Biomedical Engineering</u> 26: 213-219, 1979.
31. Kimme C, O'Loughlin BJ, Sklansky J: Automatic detection of suspicious abnormalities in breast radiographs. In: <u>Data Structures, Computer Graphics, and Pattern Recognition</u>, edited by A. Klinger, K. S. Fu, T. L. Kunii (Academic Press, New York), 1975, pp. 427-447.
32. Hand W, Semmlow JL, Ackerman LV, Alcorn FS: Computer screening of xeromammograms: A technique for defining suspicious areas of the breast. <u>Computers and Biomedical Research</u> 12: 445-460, 1979.
33. Semmlow JL, Shadagoppan A, Ackerman LV, Hand W, Alcorn FS: A fully automated system for screening xeromammograms. <u>Computers and Biomedical Research</u> 13: 350-362, 1980.
34. Ackerman LV, Gose EE: Breast lesion classification by computer and xeroradiography. <u>Cancer</u> 30: 1025-1035, 1972.
35. Ackerman LV, Mucciardi AN, Gose EE, Alcorn FS: Classification of benign and malignant breast tumors on the basis of 36 radiographic properties. <u>Cancer</u> 31: 342-352, 1973.
36. Wee WG, Moskowitz M, Chang N-C, Ting Y-C, Pemmeraju S: Evaluation of mammographic calcifications using a computer program. <u>Radiology</u> 116:717-720, 1975.
37. Fox SH, Pujare UM, Wee WG, Moskowitz M, Hutter RVP: A computer analysis of mammographic microcalcifications: Global approach. <u>Proc. IEEE 5th International Conf. on Pattern Recognition</u>: 624-631, 1980.
38. Magnin IE, Cluzeau F, Odet CL, Bremond A: Mammographic texture analysis: an evaluation of risk for developing breast cancer. <u>Optical Engineering</u> 25: 780-784, 1986.
39. Caldwell CB, Stapleton SJ, Holdsworth DW, Jong RA, Weiser WJ, Cooke G, Yaffe MJ: Characterization of mammographic parenchymal pattern by fractal dimensions. <u>Phys. Med. Biol.</u> 35: 235-247, 1990.
40. Fam BW, Olson SL, Winter PF, Scholz FJ: Algorithm for the detection of fine clustered calcifications on film mammograms. <u>Radiology</u> 169: 333-337, 1988.
41. Olson SL, Fam BW, Winter PF, Scholz FJ, Lee AK, Gordon SE: Breast calcifications: Analysis of imaging properties. <u>Radiology</u> 169: 329-331, 1988.
42. Davies DH, Dance DR: Automatic computer detection of clustered calcifications in digital mammograms. <u>Phys. Med. Biol.</u> 35: 111-118, 1990.
43. Astley S, Taylor C, Boggis C, Wilson M, Ellison T: Automated detection of abnormalities on screening mammograms. <u>Radiology</u> 177(P):288, 1990.
44. Grimaud M, Muller S, Meyer F: Automated detection of microcalcifications in mammograms. <u>Radiology</u> 177(P):288, 1990.
45. Jin H-R, Matsumoto K, Kobatake H: Automatic diagnosis of breast cancer with structural line analysis and mathematical morphology. <u>Radiology</u> 177(P) 319, 1990.
46. Karssemeijer N: A stochastic method for automated detection of microcalcifications in digital mammograms. <u>Information Processing in Medical Imaging</u>, Springer-Verlag (New York), pp.227-238, 1991.
47. Lai SM, Li X, Bischof WF: On techniques for detecting circumscribed masses in mammograms. <u>IEEE Transactions on Medical Imaging</u> 8: 377-386, 1989.
48. Brzakovic D, Luo XM, Brzakovic P: An approach to automated detection of tumors in mammograms. <u>IEEE Transactions on Medical Imaging</u> 9: 233-241, 1990.
49. Gale AG, Roebuck EJ, Riley P, Worthington BS, et al.: Computer aids to mammographic diagnosis. <u>British Journal of Radiology</u> 60: 887-891, 1987.
50. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA: Enhanced interpretation of diagnostic images. <u>Invest. Radiol.</u> 23: 240-252, 1988.
51. Swett HA, Miller PA: ICON: A computer-based approach to differential diagnosis in radiology. <u>Radiology</u> 163: 555-558, 1987.

52. Swett HA, Fisher PR, Cohn AI, Miller PI, Mutalik PG: Expert system controlled image display. Radiology 172: 487-493, 1989.

53. Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM: Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography. Med Phys 14: 538-548, 1987.

54. Chan HP, Doi K, Vyborny CJ, Lam KL, Schmidt RA: Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study. Invest Radiol 23: 664-671, 1988.

55. Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, Ogura T, Wu Y, MacMahon H: Improvement in radiologists' detection of clustered microcalcifications on mammograms: The Potential of computer-aided diagnosis. Invest Radiol 25: 1102-1110, 1990.

56. Nishikawa RM, Doi K, Giger ML, Yoshimura H, Wu Y, Vyborny CJ, Schmidt RA, Chan HP: Use of morphological filters in the computerized detection of microcalcifications in digital mammograms. Medical Physics 17: 524, 1990.

57. Nishikawa RM, Giger ML, Doi K, Schmidt RA, Vyborny CJ: Automated detection of microcalcifications on mammograms: New feature-extraction techniques with morphologic filters. Radiology 177(P): 288, 1990.

58. Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schmidt RA: Computer-aided detection of clustered microcalcifications on digital mammograms. Medical and Biological Engineering and Computing 33:174-178, 1995.

59. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE: Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. Radiology 187: 81-87, 1993.

60. Giger ML, Yin F-F, Doi K, Metz CE, Schmidt RA, Vyborny CJ: Investigation of methods for the computerized detection and analysis of mammographic masses. Proc. SPIE 1233: 183-184, 1990.

61. Yin F-F, Giger ML, Doi K, Metz CE, Vyborny CJ, Schmidt RA: Computerized detection of masses in digital mammograms: Analysis of bilateral-subtraction images. Medical Physics 18: 955-963, 1991.

62.. Yin FF, Giger ML, Doi K, Vyborny CJ, Schmidt RA: Computerized detection of masses in digital mammograms: Investigation of feature-analysis techniques. Journal of Digital Imaging 7: 18-26, 1994.

63. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE: Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer. Radiology 187: 81-87, 1993.

64. Giger ML, Nishikawa RM, Doi K, Yin FF, Vyborny CJ, Schmidt RA, Metz CE, Wu Y, MacMahon H, Yoshimura H: Development of a "smart" workstation for use in mammography. Proc. SPIE 1445: 101-103, 1991.

65. Doi K, Giger ML, MacMahon H, Hoffmann KR, et al.: Computer-aided diagnosis: development of automated schemes for quantitative analysis of radiographic images. Seminars in Ultrasound, CT and MR 13(2): 140-152, 1992.

66. Giger ML, Doi K, MacMahon H, Nishikawa RM, Hoffmann KR, et al.: An "intelligent" workstation for computer-aided diagnosis. RadioGraphics 13: 647-656, 1993.

67. MacMahon H, Doi K, Chan HP, Giger ML, Katsuragawa S, Nakamori N: Computer-aided diagnosis in chest radiology. J Thoracic Imaging 5: 67-76, 1990.

68. Metz CE: ROC methodology in radiologic imaging. Invest Radiol 21: 720-733, 1986.

69. Giger ML, Doi K, MacMahon H: Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. Med. Phys. 15: 158-166, 1988.

70. Giger ML, Ahn N, Doi K, MacMahon H, Metz CE: Computerized detection of pulmonary nodules in digital chest images: Use of morphological filters in reducing false-positive detections. Med Phys 17: 861-865, 1990.

71. Giger ML, Doi K, MacMahon H, Metz CE, Yin FF: Computer-aided detection of pulmonary nodules in digital chest images. RadioGraphics 10: 41-52, 1990.

72. Yoshimura H, Giger ML, Doi K, MacMahon H, Montner SM: Computerized scheme for the
     detection of pulmonary nodules: A nonlinear filtering technique. Invest Radiol 27: 124-129,
     1992.
73. Katsuragawa K, Doi K, et al.: Image feature analysis and computer-aided diagnosis in digital
     radiography: Effect of digital parameters on the accuracy of computerized analysis of
     interstitial disease in digital chest radiographs. Med Phys 17: 72078, 1990.
74. Katsuragawa S, Doi K, et al.: Quantitative computer-aided analysis of lung texture in chest
     radiographs. RadioGraphics 10: 257-269, 1990.
75. Nakamori N, Doi K, Sabeti V, MacMahon H: Image feature analysis and computer-aided
     diagnosis in digital radiography: Automated analysis of sizes of hearts and lung in digital chest
     images. Med Phys 17: 342-350, 1990.
76. Fujita H, Doi K, Giger ML, Chan HP: Investigation of basic imaging properties in digital
     radiography. 5. Characteristic curves of I.l.-TV digital systems. Medical Physics 13: 13-18,
     1986.
77. Giger ML: Film digitization—technical requirements. Invited for presentation. In: Proceedings
     of Chest Imaging Conference '87, (Eds.) Peppler W, Alter A, Medical Physics Publishing
     Corp. Madison, Wisconsin, pp. 92-100, 1988.
78. Giger ML, Doi K: Investigation of basic imaging properties in digital radiography. 1.
     Modulation transfer function. Medical Physics 11: 287-295, 1984.
79. Fujita H, Doi K, Giger ML: Investigation of basic imaging properties in digital radiography. 6.
     MTFs of I.I.-TV digital systems. Med Phys 12: 713-729, 1985.
80. Fujita H, Giger ML, Doi K: Investigation of basic imaging properties in digital radiography.
     12. Effect of matrix configuration on system resolution. Medical Physics 15: 384-390, 1988.
81. Fraser RG, Sanders C, Barnes GT, MacMahon H, Giger ML, Doi K, Templeton AW, Cox
     GG, Dwyer SJ, Merritt C, Jones J: Digital imaging of the chest: state of the art. Radiology 171:
     297-307, 1989.
82. Cook LT, Giger ML, Batnitzky S, Wetzel LH, Murphey MD: Digitized film radiography.
     Investigative Radiology 24: 910-916, 1989.
83. Montner S, Xu X-W, Tsuzaka M, Doi K, MacMahon H, Yoshimura H, Sanada S, Giger ML,
     Yin F-F: Evaluation of basic imaging properties of a new digital chest system. Proc. SPIE
     1231: 390-393, 1990.
84. Sanada S, Doi K, Xu X-W, Yin F-F, Giger ML, MacMahon H: Comparison of imaging
     properties of a computed radiography system and screen-film systems. Medical Physics 18:
     414-420, 1991.
85. Yin F-F, Giger ML, Doi K: Measurement of the presampling MTF of film digitizers using a curve
     fitting technique. Medical Physics 17: 962-966, 1990.
86. Yin FF, Giger ML, Doi K, Yoshimura H, Xu XW, Nishikawa RM: Evaluation of imaging properties
     of a laser film digitizer. Phys. Med. Biol. 37: 273-280, 1992.
87. Giger ML, Doi K, Metz CE: Investigation of basic imaging properties in digital radiography. 2.
     Noise Wiener Spectrum. Medical Physics 11: 797-805, 1984.
88. Giger ML, Doi K, Fujita H: Investigation of basic imaging properties in digital radiography. 7.
     Noise Wiener spectra of I.I.-TV digital imaging systems. Medical Physics 13: 131-138, 1986.
89. Giger ML, Ohara K, Doi K: Effect of quantization on digitized noise and detection of low-
     contrast objects. Proc. SPIE 626: 214-224, 1986.
90. Kume Y, Doi K, Ohara K, Giger ML: Investigation of basic imaging properties in digital
     radiography. 10. Structure mottle of I.I.-TV digital imaging systems. Medical Physics 13: 843-
     849, 1986.
91. Giger ML: Image quality: Effects of digitization, matrix size and noise. Invited for presentation
     at the 1987 AAPM Annual Summer School, In: Image Communication and Image Analysis.
     (Mulvaney, Haus, Windham, eds.), New York, AIP (in press) .
92. Giger ML, Doi K: Investigation of basic imaging properties in digital radiography. 3. Effect of
     Pixel Size on SNR and Threshold Contrast. Med. Phys. 12: 201-208, 1985.
93. Giger ML, Doi K: Effect of pixel size on detectability of low-contrast signals in digital
     radiography. J of the Optical Society of America A 4: 966-975, 1987.

94.   Ohara K, Chan HP, Doi K, Giger ML, Fujita H: Investigation of basic imaging properties in digital radiography. 8. Detection of simulated low-contrast objects in DSA images. Medical Physics 13: 304-311, 1986.

95.   Ohara K, Doi K, Metz CE, Giger ML: Investigation of basic imaging properties in digital radiography. 13. Effect of structured noise on the detectability of simulated stenotic lesions. Medical Physics 16:14-21, 1989.

96.   MacMahon H, Metz CE, Doi K, Kim T, Giger ML, Chan H-P: The effect of display format on diagnostic accuracy in digital chest radiography: A comparison of hardcopy, video, and reversed grey scale. Radiology 168: 669-673, 1988.

97.   MacMahon H, Doi K, Sanada S, Montner SM, Giger ML, Metz CE, Nakamori N, Yin F-F, Xu X-W, Yonekawa H, Takeuchi H: Data compression: Effect on diagnostic accuracy in digital chest radiography. Radiology 178: 175-179, 1991.

98.   MacMahon H, Sanada S, Doi K, Giger ML, Xu X-W, Yin F-F, Montner SM, Carlin M: Direct comparison of conventional and computed radiography with a dual image recording technique. RadioGraphics 11: 259-268, 1991.

99.   Asada N, Doi K, MacMahon H, Montner S, Giger ML, Abe C, Wu Y: Potential usefulness of artificial neural network for differential diagnosis of interstitial lung diseases: a pilot study. Radiology 177: 857-860, 1990.

100.   Murphy WA Jr, Destouet JM, Monsees BS: Professional quality assurance for mammography screening programs. Radiology 175: 319-320, 1990.

101.   Bird RE: Professional quality assurance for mammography screening programs. Radiology 177: 587, 1990.

102.   Brenner RJ: Medicolegal aspects of breast imaging: variable standards of care relating to different types of practice. AJR 156: 719-723, 1991.

103.   Pratt WK: Digital Image Processing (Wiley, New York, 1978).

104.   Serra J: Image Analysis and Mathematical Morphology, (Academic Press, New York), 1982.

105.   Rumelhart DE, Hinton GE, Williams RJ: Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, PDP Research Group. Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge: MIT Press 1: 318-362, 1986.

106.   Kendall M, Stuart A: The Advanced Theory of Statistics, Vol. 3 (Hufner, New York), 1976.

107.   Metz CE: Some practical issues of experimental design and data analysis in radiological ROC studies. Invest. Radiol. 24: 234-245, 1989.

108.   Dorfman DD, Alf E: Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals-rating method data. J Math Psych 6: 487-498, 1969.

109.   Bunch PC, Hamilton JF, et al.: A free response approach to the measurement and characterization of radiographic observer performance. Proc. SPIE 127: 124-135, 1977.

110.   Chakraborty D: Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. Med. Phys. 16: 561-568, 1989.

111.   Giger ML, Lu P, Huo Z, Bick U, Vyborny CJ, Schmidt RA, Zhang W, Metz CE, Wolverton D, Nishikawa RM, Zouras W, Doi K: CAD in mammography: computerized detection and classification of masses. Proceedings of 2nd International Workshop on Digital Mammography, York, UK, Elsevier Science, pp. 282-287, 1994.

112.   Kupinski M, Giger ML, Lu P, Huo Z: Computerized detection of mammographic lesions: Performance of artificial neural network with enhanced feature extraction. Proc SPIE 2434: 598-605, 1995.

113.   Bick U, Giger ML, Schmidt RA, Nishikawa RM, Wolverton DE, Lu P, Vyborny CJ, Doi K: Automated segmentation of digitized mammograms. Academic Radiology 2: 1-9, 1995.

114.   Kupinski M, Giger ML: Computerized detection of mammographic masses: neural network feature selection with genetic algorithms, (in preparation), 1996.

115.   Zhang M, Giger ML: Automated detection of spiculated lesions and architectural distortions in digitized mammograms. Proc SPIE 2434: 846-854, 1995.

116.   Bick U, Giger ML, Schmidt RA, Nishikawa RM, Doi K: Peripheral density correction of digital mammographs. RadioGraphics (submitted), 1996.

117.   Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schmidt RA: Computer-aided detection of clustered microcalcifications: An improved method for grouping detected signals. <u>Medical Physics</u> 20: 1661-1666, 1993.

118.   Zhang W, Doi K, Giger ML, Wu Y, Nishikawa RM, Schmidt RA: Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network. <u>Medical Physics</u> 21: 517-524, 1994.

119.   Zhang W, Doi K, Giger ML, Nishikawa RM, Schmidt RA: An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms. <u>Medical Physics</u> (in press), 1995.

120.   Huo Z, Giger ML, Vyborny CJ, Bick U, Lu P, Wolverton DE, Schmidt RA: Analysis of spiculation in the computerized classification of mammographic masses" <u>Medical Physics</u> 22:1569-1579, 1995.

121.   Huo Z, Giger ML, Vyborny C, et al: Computerized classification of benign and malignant mass lesions on digital mammograms, (in preparation), 1996.

122.   Jiang Y, Nishikawa RM, Wolverton DE, Metz CE, Giger ML, Schmidt RA, Vyborny CJ, Doi K: Automated feature analysis and classification of malignant and benign clustered microcalcifications. <u>Radiology</u> 198:671-678, 1996.

123.   Nishikawa RM, Haldemann RC, Papaioannou J, Giger ML, Lu P, Schmidt RA, Wolverton DE, Bick U, Doi K: Initial experience with a prototype clinical "intelligent" mammography workstation for computer-aided diagnosis. Proc SPIE 2434: 65-71, 1995.

124.   Haldemann RC, Nishikawa RM, Giger ML, et al.: Screening mammography with computer-aided diagnosis: Initial pre-clinical prospective evaluation (in preparation), 1996.

APPENDIX

Analysis of spiculation in the computerized classification of mammographic masses[a]

Zhimin Huo, Maryellen L. Giger,[b] Carl J. Vyborny, Ulrich Bick, Ping Lu,
Dulcy E. Wolverton, and Robert A. Schmidt
Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology,
5841 South Maryland Avenue, The University of Chicago, Chicago, Illinois 60637

Spiculation is a primary sign of malignancy for masses detected by mammography. In this study, we developed a technique that analyzes patterns and quantifies the degree of spiculation present. Our current approach involves (1) automatic lesion extraction using region growing and (2) feature extraction using radial edge-gradient analysis. Two spiculation measures are obtained from an analysis of radial edge gradients. These measures are evaluated in four different neighborhoods about the extracted mammographic mass. The performance of each of the two measures of spiculation was tested on a database of 95 mammographic masses using ROC analysis that evaluates their individual ability to determine the likelihood of malignancy of a mass. The dependence of the performance of these measures on the choice of neighborhood was analyzed. We have found that it is only necessary to accurately extract an approximate outline of a mass lesion for the purposes of this analysis since the choice of a neighborhood that accommodates the thin spicules at the margin allows for the assessment of margin spiculation with the radial edge-gradient analysis technique. The two measures performed at their highest level when the surrounding periphery of the extracted region is used for feature extraction, yielding A_z values of 0.83 and 0.85, respectively, for the determination of malignancy. These are similar to that achieved when a radiologist's ratings of spiculation (A_z=0.85) are used alone. The maximum value of one of the two spiculation measures (FWHM) from the four neighborhoods yielded an A_z of 0.88 in the classification of mammographic mass lesions.

Key words: spiculation, digital mammogram, radial edge-gradient analysis, ROC analysis, computer-aided diagnosis, computer vision

I. INTRODUCTION

X-ray mammography has been proven to be the most effective method for the detection of early breast cancer. However, mammographic findings of benign and malignant masses often overlap.[1] At many centers, only 10%–20% of detected masses removed by surgical breast biopsy are malignant.[2,3] A computer scheme capable of providing objective information may aid radiologists in their classification of masses, thus preventing unnecessary biopsies. Computer aids have already been shown to improve the detection performance of radiologists.[4,5]

The shape, margin, and density of a mass are used by radiologists to characterize masses.[1,6,7] The margin characteristics of a mass observed mammographically are very important indicators of its benign or malignant status. The margin of a mass can be categorized as circumscribed, lobulated, obscured, indistinct, or spiculated with a spiculated margin being the strongest sign for malignancy.[1,6,7]

Various investigations[8–13] have attempted to classify breast lesions or to detect spiculated masses based on computer-extracted features characterizing either the margin, shape, or density of a mass. Ackerman et al.[8] extracted four features of malignancy, calcification, spiculation, roughness, and shape, from lesions identified by radiologists on xeroradiographs and then merged the four features to classify those lesions. Brzakovic et al.[9] classified detected abnormalities into nontumor, benign tumor, and malignant tumor using measures of size, shape, and intensity change. Kegelmeyer[10] used the analysis of edge orientation histograms to detect stellate lesions. Kilday et al.[11] segmented lesions with a simple thresholding technique and used linear discriminant analysis to merge several shape-related features to distinguish between fibroadenomas, cysts, and carcinomas. Other investigators have used only a single computer-extracted feature related to either margin, shape, or density as an indicator of malignancy. Burdett et al.[12] applied a fractal analysis to quantify the degree of surface roughness as a single indicator of malignancy. Claridge et al.[13] analyzed a small set of malignant lesions by measuring the lesion edge blurriness. In addition, many investigators[14–18] have taken advantage of the ability of radiologists to extract mammographic features, which are subsequently merged by rule-based, discriminant analysis or neural networks into a final determination of the likelihood of malignancy.

Previously we developed a classification method that involved the extraction of lesions using a manual region-growing technique and the extraction of two features containing margin information. These were merged by an artificial neural network to quantify the degree of spiculation.[19] The database in that study contained 28 benign and 25 malignant masses. The result showed that the mammographic features extracted and merged in this way yielded measures of spiculation comparable to those obtained by an expert mammographer.

In this study, we developed a new spiculation-sensitive pattern-recognition technique, "radial edge-gradient analysis." Prior to the feature extraction, we employed an auto-
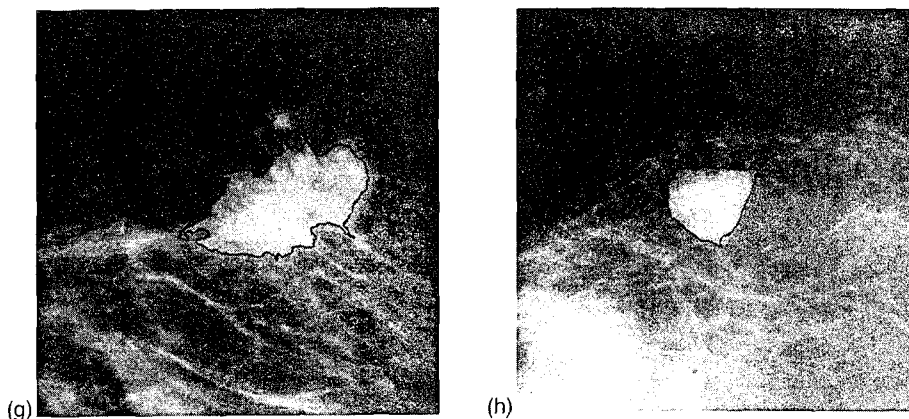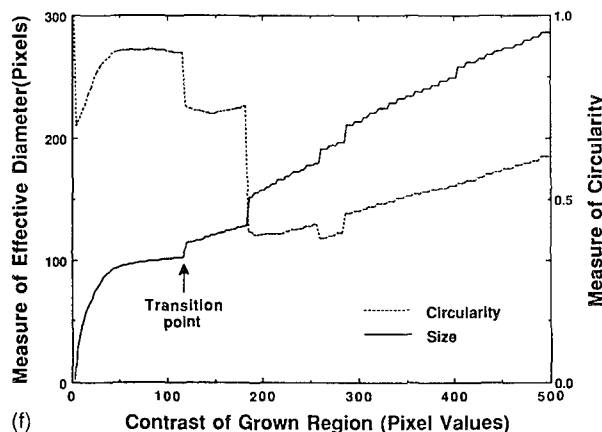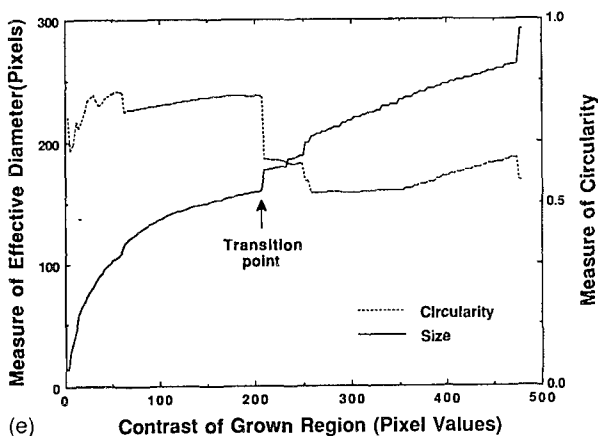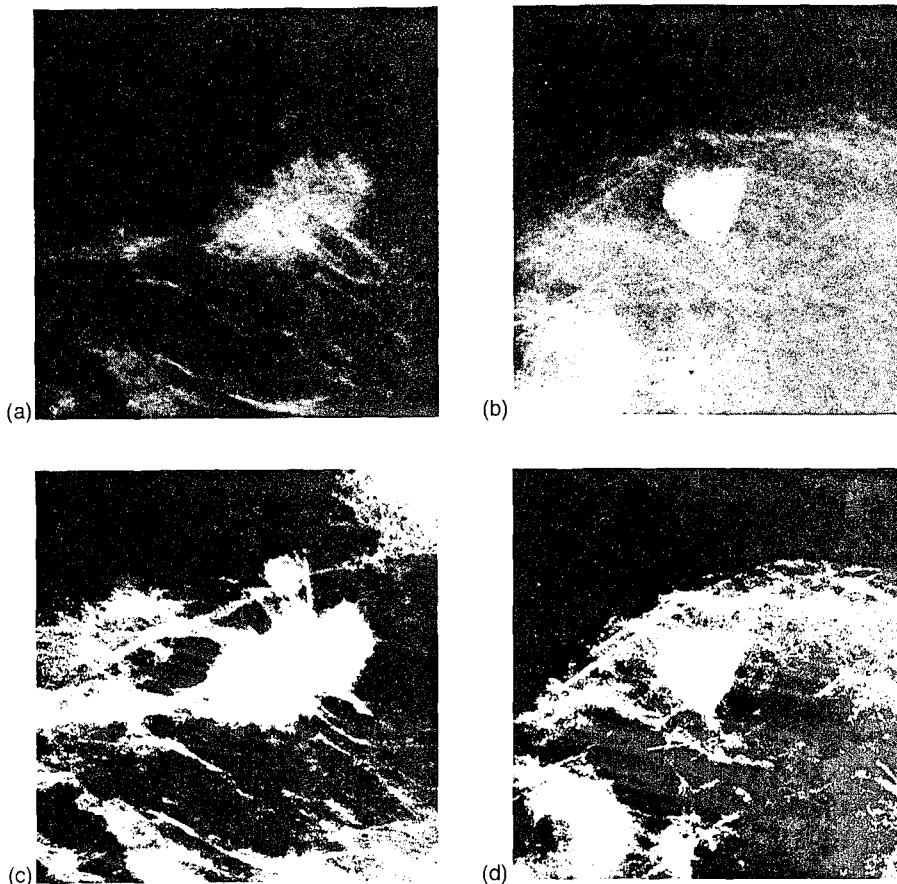
FIG. 2. 512×512 ROIs centered about (a) an original malignant mass and (b) a benign mass. The processed images of the (c) malignant and (d) benign masses after background trend correction and histogram equalization. Diagrams of size and circularity of the grown region as functions of gray-level interval (contrast) with the automatically determined transition point indicated for the (e) malignant and (f) benign masses. The computer-extracted margins overlayed on the (g) malignant and (h) benign masses.
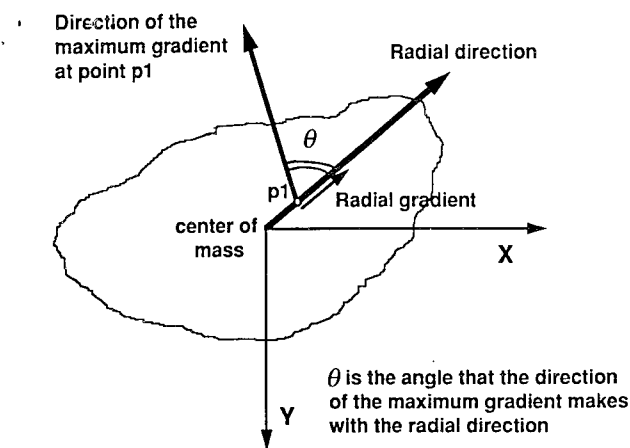
FIG. 4. Illustration defining the radial angle θ as the angle between the direction of the maximum gradient and its radial direction which is the direction pointing from the center of mass to the point $p1$, and the radial gradient as the projection of the maximum gradient along the radial direction.

The radial direction for point $p1$ is the direction pointing from the geometric center of the grown mass to $p1$. The angle θ between the direction of the maximum gradient at the pixel $p1$ and its radial direction is the angle relative to the radial direction or the "radial angle." Note that θ is not the angle the maximum gradient makes with the x direction. Analysis relative to the x axis yields information only on whether a lesion is circular or not,[5,24] i.e., it can only be used to distinguish circular patterns from linear patterns, not circular patterns from spiculated patterns. Rather, our analysis was developed in order to distinguish spiculated masses from circular or oval masses with smooth margins, since spiculation is an important indicator of malignancy.

In each neighborhood, the maximum gradients having the same radial angle are summed for each radial angle, resulting in a cumulated edge-gradient distribution relative to the radial angle. The cumulated edge-gradient distribution is then normalized by the average maximum gradient of the particular neighborhood, enabling comparison of cumulated edge-gradient distributions between various lesions. Normalization is performed such that the area under the normalized distribution curve is one. Representative analyses were performed on a smooth, round benign mass and a spiculated, round malignant mass which are shown in Figs. 5(a) and 5(b), respectively. Figures 5(c) and 5(d) show the corresponding normalized cumulated edge-gradient distributions relative to the radial angle obtained using neighborhood (B) (margin). It should be noted that the benign mass yields a
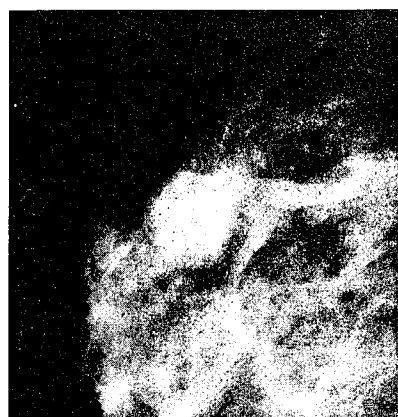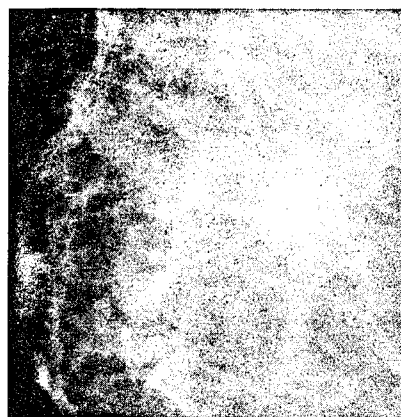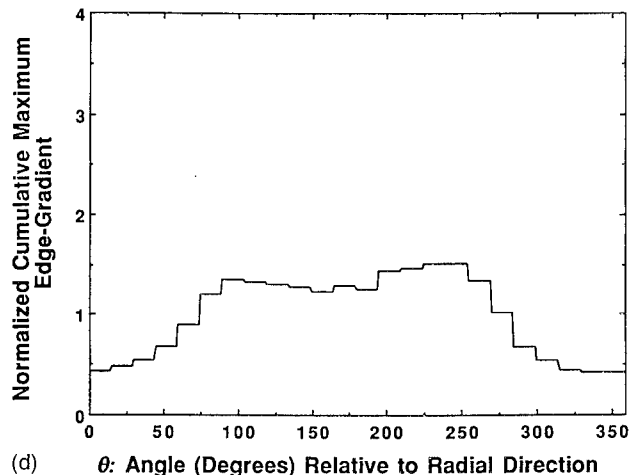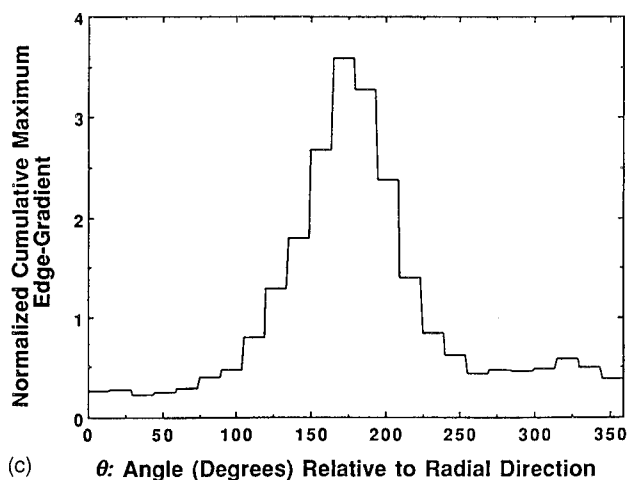
FIG. 5. (a) A mammographic circular, smooth mass and (c) its corresponding normalized cumulated edge-gradient distribution. (b) A mammographic spiculated mass and (d) its corresponding normalized cumulated edge-gradient distribution.
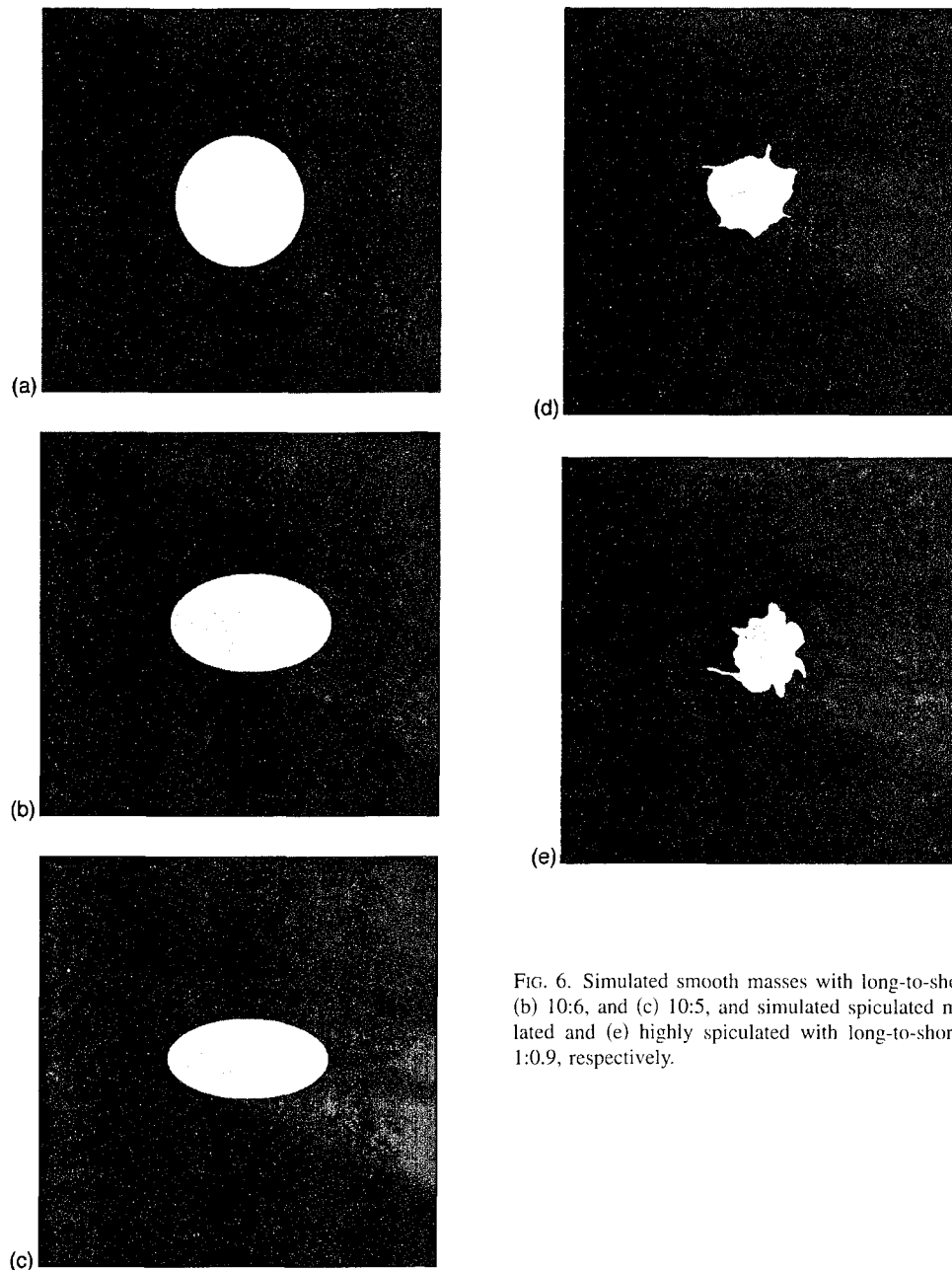
Fig. 6. Simulated smooth masses with long-to-short axis ratios of (a) 1:1, (b) 10:6, and (c) 10:5, and simulated spiculated masses (d) slightly spiculated and (e) highly spiculated with long-to-short axis ratios of 1:1 and 1:0.9, respectively.

lated masses and smooth masses having a long-to-short ratio greater than 1.7. However, to prevent misclassifying spiculated masses as nonspiculated masses by overcorrecting the FWHM measures, the FWHM measure is made only for the masses having a long-to-short axis ratio larger than 1.8. For the same reason, a single value correction of 36° on the FWHM measure for the masses having a long-to-short axis ratio larger than 1.8 is used rather than a correction factor for each individual mass based on its shape.

IV. RESULTS

Figure 8 shows the relationship between the corrected FWHM and the normalized radial gradient measures within the rectangular segment [neighborhood (C)] for the 95 masses. It is apparent that most of the malignant masses have large values of FWHM and small values of normalized radial gradient. For example, by setting a threshold at 160° for the

FWHM measure, 75% of the malignant masses can be correctly identified with only 4 out of 38 benign masses being misclassified (Fig. 8).

ROC analysis[27–29] was undertaken to evaluate the abilities of each of the two spiculation measures determined for the four neighborhoods in distinguishing between benign and malignant masses. The area under the ROC curve (A_z) was calculated as an index for the performance of each feature as shown in Table II. Figures 9(a)–9(c) show the individual performance of the two spiculation measures for each neighborhood type. The performances of the uncorrected FWHM and normalized radial gradient measures in classifying the 95 masses for each neighborhood are similar as shown in Figs. 9(a) and 9(b). It is apparent that the choice of neighborhoods will affect the performance level also as illustrated in Figs. 9(a) and 9(b). The effect of the four neighborhoods on the two spiculation measures shows the same trend for each

TABLE. II. A_z values of the two spiculation features extracted in the four neighborhoods (95 mammographic masses).

Measures	Neighborhoods			
	(A) Margin	(B) Grown region	(C) Encompassing region	(D) Surrounding periphery
Normalized radial gradient	0.70	0.75	0.80	0.80
Uncorrected FWHM	0.70	0.75	0.80	0.83
Corrected FWHM	0.73	0.77	0.83	0.85

(a)

computer-based spiculation measure (FWHM) achieves higher A_z values (A_z=0.88) than that based on the spiculation ratings from an human observer. Of course, with the use of additional mass-related features such as opacity or shape, the performances of both the computer-based measures and human assessment would be expected to improve.

V. DISCUSSION

In order to maximize the extraction of the margin spiculation information from a mass, four different neighborhoods about the grown region were introduced for feature extraction. As described earlier, neighborhoods (A) and (B) rely entirely on the grown region, whereas neighborhoods (C) and (D) introduce regions surrounding the grown mass in order to include thin, short spicules radiating from the margin of a mass, which could not be delineated by the gray-level region growing technique. The size of the region introduced in neighborhoods (C) and (D) is only large enough to accommodate thin, short spicules. Since the four neighborhoods are determined from the grown region, the accuracy of the lesion segmentation prior to feature extraction is important in the success of subsequent feature analysis. However, the regions introduced in neighborhoods (C) and (D) make the subsequent analysis less dependent on the grown region.

Results show that spiculation analysis within neighborhoods (C) and (D) yield higher A_z values than that within neighborhoods (A) and (B). This demonstrates the usefulness of introducing a zone around the extracted lesion to accommodate potential margin spiculation. The A_z values of the two spiculation measures obtained from margin (B) and surrounding periphery (D), which exclude most of the interiors of the grown (A) and encompassing (C) regions, are higher than the A_z values obtained from the grown (A) and encompassing (C) regions themselves, respectively. This demonstrates that mainly using the margin information increases the "signal-to-noise" ratio, and thus optimizes the radial edge-gradient analysis technique in the extraction of the margin spiculation.

Thus, with the radial edge-gradient analysis technique, we found that a lesion can be extracted devoid of its spicules and still be accurately analyzed for spiculation if the proper neighborhood is chosen. That is, by studying the periphery
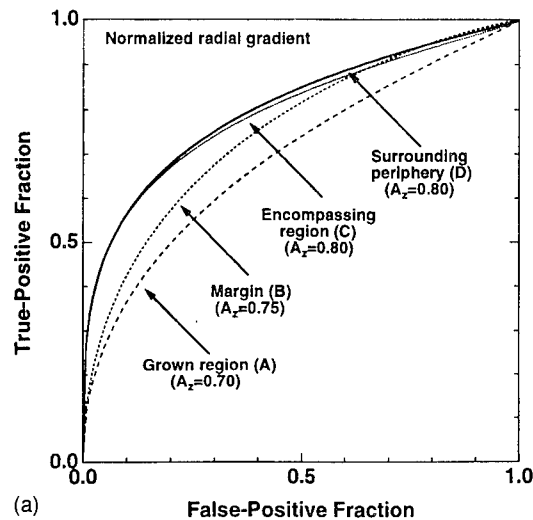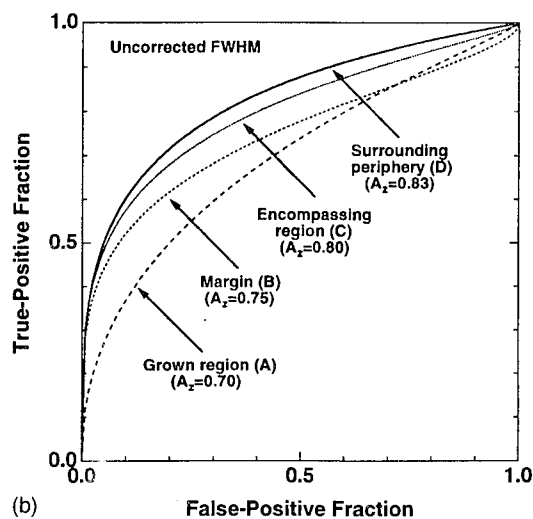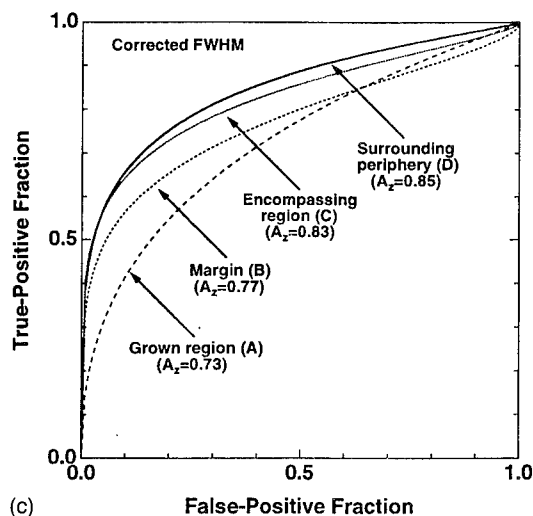
FIG. 9. ROC curves for (a) the normalized radial gradient measures, (b) the FWHM measures, and (c) the corrected FWHM measures on a database of 95 mammographic masses for the four neighborhoods showing the performance in classifying malignant and benign masses.

[neighborhoods (C) and (D)] around a grown mass, it is not necessary to require that the grown region include fine spicules.

In the application of radial edge-gradient analysis in clas-

,the computerized detection of mammographic masses," Invest. Radiol. **28**, 473–481 (1993).

[23]U. Bick, M. L. Giger, R. A. Schmidt, R. M. Nishikawa, D. E. Wolverton, and K. Doi, "Automated segmentation of digitized mammograms," Acad. Radiol. **2**, 1–9 (1995).

[24]T. Matsumoto, H. Yoshimura, K. Doi, M. L. Giger, A. Kano, H. MacMahon, K. Abe, and S. M. Montner, "Image feature analysis of false-positive diagnoses produced by automated detection of lung nodules," Invest. Radiol. **27**, 587–597 (1992).

[25]J. C. Russ, *The Image Processing Handbooks* (CRC, Boca Raton, 1992).

[26]J. Serra, *Image Analysis and Mathematical Morphology* (Academic, New York, 1982).

[27]C. E. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. **21**, 720–733 (1986).

[28]C. E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies," Invest. Radiol. **24**, 234–245 (1989).

[29]C. E. Metz, J. H. Shen, and B. A. Herman, "New methods for estimating a binormal ROC curve from continuously distributed test results," presented at the 1990 Annual Meeting of the American Statistical Association, Anaheim, CA, August 7, 1990.

[30]C. E. Metz, P. L. Wang, and H. B. Kronman, "A new approach for testing the significance of differences between ROC curves measures from correlated data," in *Information Processing in Medical Imaging: Proceedings of the 8th Conference*, edited by F. Deconinck (Martinus Nijhoff, Boston, 1984), pp. 432–445.

[31]Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Feature analysis and classification of mass lesions in digital mammograms," Med. Phys. **22**, 937 (1995).

Reprinted from

Medical Imaging 1995

Image Processing

27 February–2 March 1995
San Diego, California

Volume 2434

rules for determining truth. ANNs "learn" from examples that are presented repeatedly. Neural networks have found popularity in many fields due to their inherent ability to make decisions and draw conclusions when the data presented is complex, noisy, or incomplete. In recent years neural networks have found increased popularity in the field of medical imaging where pattern recognition or classification is important.[15-18] For this study we utilized neural networks to classify (lesion vs. false-positive) the regions of interest based on the features that were extracted from these regions. Figure 1 illustrates the complete process of the mass detection scheme.

2. DATABASE

Artificial neural networks are trained pattern recognition devices, and thus the databases that are used to train and test the ANN are vital in interpreting the performance. If, for example, the computerized method is trained only on lesions that are "easy" to detect, then the method will likely not do well on the more difficult to detect masses. Our database consists of 110 pairs of digital mammograms with 102 masses (54 malignant and 48 benign). In addition, 302 false-positive regions were selected for neural network training. All of the masses in the database were rated subjectively for detection subtlety by experienced radiologists using a five-category scale.[14] Figure 2 shows the distribution of this rating. As the graph shows, a substantial number of the masses in our database were designated by radiologists as "difficult" lesions to detect.

3. FEATURE EXTRACTION

The first step in the feature extraction process is the extraction of the potential lesion by region growing.[19] This process yields a computer-delineated margin around the potential lesion referred to as the grown mass. This grown mass is then used in extracting the features.

A total of 91 features are calculated. Space does not permit discussion of all features, so this paper will focus on only those features eventually selected for input into the ANN. These selected features can be separated into three types: geometric measures, intensity-based measures and gradient-based measures.

3.1. Geometric Measures

Geometric measures pertain to the shape of the lesion. The circularity, effective diameter, and irregularity were the three geometric features used for input to the ANN. Figure 3 gives the definitions of these three features.

3.2. Intensity-based Measures

Intensity-based measures are related to the gray-level values (and thus related to the density of the tissue) within the grown mass. These features include measures of the gray-level difference between the grown mass and its local background, the average of the gray-level values within the grown mass, and the average gray level within a smoothed margin of the grown mass.

3.3. Gradient-based Measures

The gradient-based measures are calculated for four neighborhoods about the potential lesion. These neighborhoods are (1) the margin of the grown mass, (2) the grown region (inside the grown mass), (3) the periphery (outside the grown mass but within the ROI), and (4) the ROI (rectangle encompassing the mass).[20] For gradient-based measures, each of the four regions are processed with a 3x3 Sobel filter to calculate the gradient at each individual pixel. At each location the maximum gradient and the angle of this gradient relative to the radial direction and a fixed axis are calculated. Gradient-weighted histograms are then determined using both the radial angle and the angle with respect to the fixed axis. Measures such as full-width half-maximums (FWHMs), average values, minima, heights and

difference between the output of the ANN and the desired output.[21] Training was performed with an output value of 0.9 representing a true-positive and a value of 0.1 representing a false-positive.

The neural network was tested using both consistency and round robin testing. Consistency testing involves training the ANN on the entire database and then testing it on the same database (see Figure 6A). This will give a measure of how well the network "learned" its training set. In round robin testing, the ANN is trained on all but one of the training set ROIs, which is used for testing. This process is repeated so that each case in the training set is used once as a testing case (see Figure 6B). Since the round-robin tests employ data that the ANN has not been trained on, these trials provide an approximation of the general performance of the ANN.

Receiver operating characteristic (ROC) analysis[22, 23] was employed to evaluate the performance of the ANN in distinguishing true lesions from false-positives. The LABROC4 program developed by Metz et al.[24] was used to fit the data output from the neural networks. The area, A_z, under the ROC curve represents the performance of the ANN. Free-response operating characteristic (FROC) curves, obtained by plotting the sensitivity (lesions detected divided by the actual number of lesions) versus the number of false positives per image, were also used.

The input features were chosen by determining those features that exhibited the greatest one-dimensional separation, but other variables of the neural network were determined empirically during the training process. The number of hidden units in the hidden layer is an important parameter because it is a measure of how complex a separation the ANN can make. The network requires enough hidden units to make the separation between true and false data; however, if a large number of hidden units are arbitrarily added, the ANN will begin to "learn" the training set too well, which decreases round robin performance due to loss of generality. In order to determine the optimal number of hidden units, we investigated the performance of the ANN as a function of the numbers of hidden units and chose the number of hidden units that maximized the round robin performance. We performed similar tests with other parameters of the ANN, such as the learning rate, to optimize the structure of the artificial neural network.

5. RESULTS AND DISCUSSION

Figure 7 shows ROC curves illustrating the consistency and round robin performance levels of our previous and current mass detection scheme at the feature analysis stage. As these plots indicate, there is a substantial increase in the round robin A_z using the enhanced feature extraction. Thus, the new features and the ANN structure have increased the general performance of the computerized scheme.

The sensitivity for detection of the malignant lesions was 100% at 15 false positives per image at the bilateral subtraction stage and 89% at 2 false positives per image after the ANN stage (using the round robin results). We can conclude from this information that the gradient-based features, which were initially thought to be useful only in the classification (malignant versus benign) of masses,[20] are also useful features to employ in the detection of mammographic lesions and the elimination of false positives.

We are presently performing additional research using rule-based studies and optimization of the input features to further improve the performance of the mass detection program.
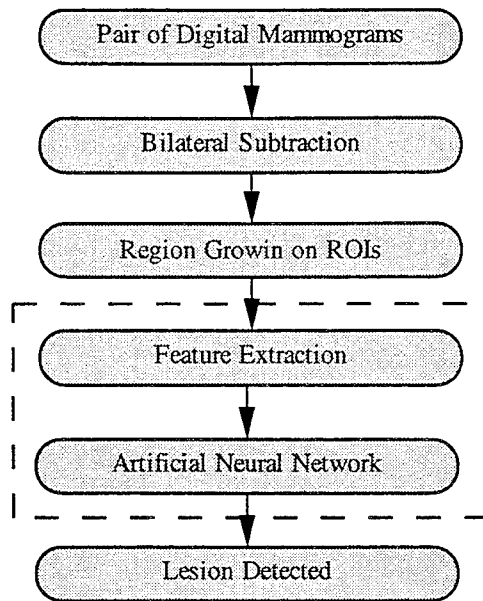
6. ACKNOWLEDGMENTS

Figure 1. Procedural diagram of the mass detection scheme. The dashed box signifies the areas on which this research focused.
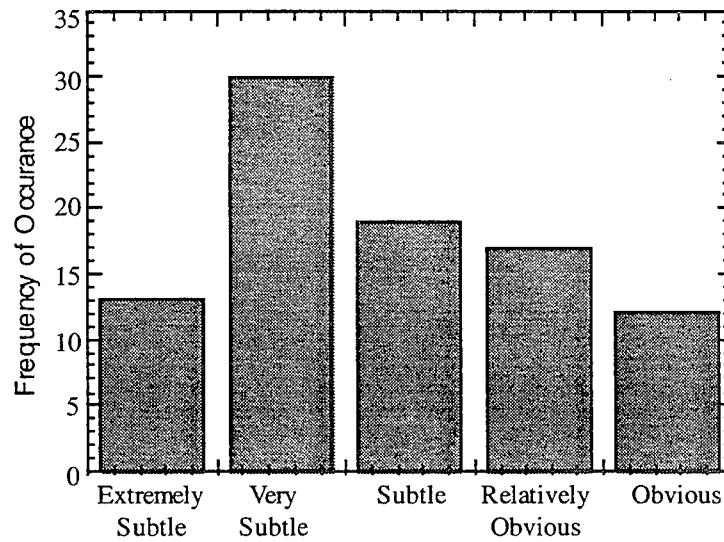

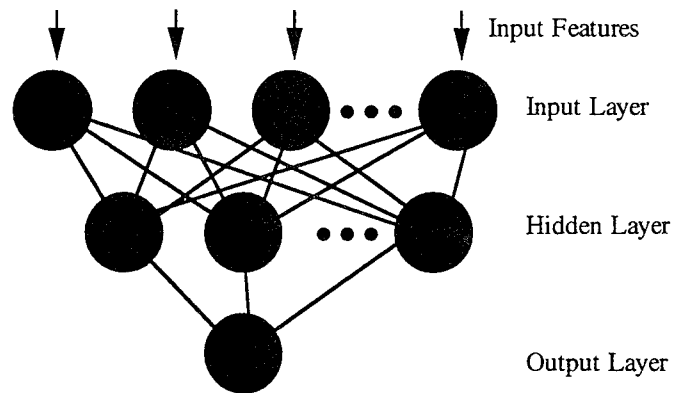Figure 2. Description of database in terms of subtlety for detection as rated by an experienced mammographer.

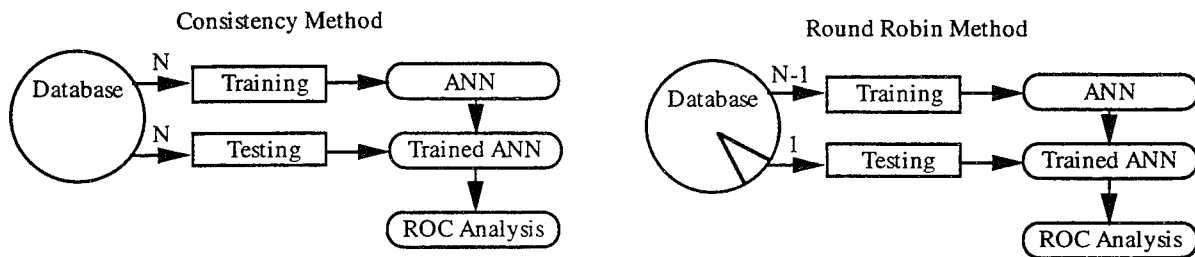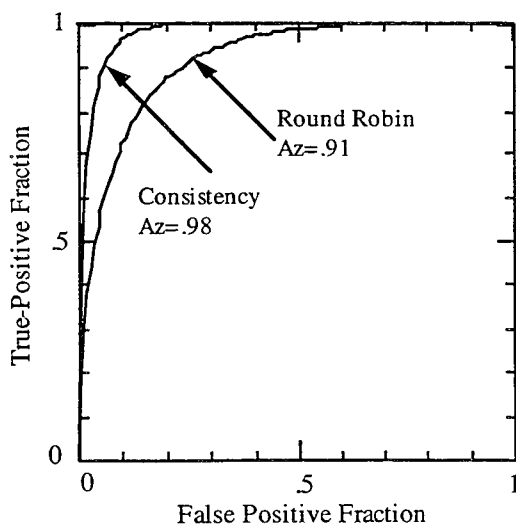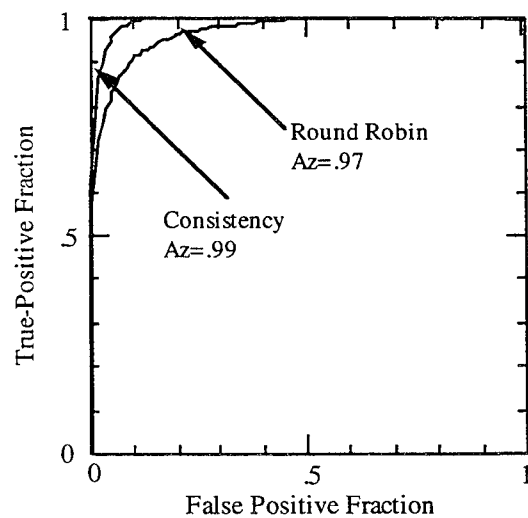Figure 5. Structure of a three-layer, feed-forward neural network.

Figure 6. Schematic diagram of (a) consistency method and (b) round robin method.
N is the number of ROIs in the database.

(a)

(b)

Figure 7. ROC curves indicating the performance of (a) the previous ANN and (b) the current ANN.

Yulei Jiang, MS • Robert M. Nishikawa, PhD • Dulcy E. Wolverton, MD
Charles E. Metz, PhD • Maryellen L. Giger, PhD • Robert A. Schmidt, MD
Carl J. Vyborny, MD, PhD • Kunio Doi, PhD

Malignant and Benign Clustered Microcalcifications: Automated Feature Analysis and Classification[1]

PURPOSE: To develop a method for differentiating malignant from benign clustered microcalcifications in which image features are both extracted and analyzed by a computer.

MATERIALS AND METHODS: One hundred mammograms from 53 patients who had undergone biopsy for suspicious clustered microcalcifications were analyzed by a computer. Eight computer-extracted features of clustered microcalcifications were merged by an artificial neural network. Human input was limited to initial identification of the microcalcifications.

RESULTS: Computer analysis allowed identification of 100% of the patients with breast cancer and 82% of the patients with benign conditions. The accuracy of computer analysis was statistically significantly better than that of five radiologists (*P* = .03).

CONCLUSION: Quantitative features can be extracted and analyzed by a computer to distinguish malignant from benign clustered microcalcifications. This technique may help radiologists reduce the number of false-positive biopsy findings.

ALTHOUGH mammography is highly sensitive (70%–90%) in the early detection of breast cancer, its efficacy is limited by the poor positive predictive value (15%–30%) obtained by human observers (1–3). Since calcifications are commonly seen on mammograms (3), detection of breast cancer prompted by clustered microcalcifications relies on accurate differential analysis (4). However, analysis of microcalcifications is often difficult; this prevents radiologists from obtaining a high positive biopsy yield on the basis of mammography (1).

Several articles describe computerized methods that potentially assist radiologists to more accurately differentiate breast cancer from benign breast disease (5–7). However, image features were manually extracted in these studies and the computer was used only for decision making. The reliance on humans to extract subjective impressions of many image features effectively renders these techniques impractical for routine clinical use. Other investigators have used computer-extracted features to classify malignant and benign clustered microcalcifications (8–11). In all but one of these studies, the features tended to express mathematical forms with no direct correlation to the features identified by a radiologist (1,4,12,13). Furthermore, these features often did not allow successful classification of clustered microcalcifications.

We developed a computerized method to extract features of clus-

tered microcalcifications that qualitatively correlate to those seen by a radiologist, and produced an estimate of the likelihood of malignancy on the basis of these automatically extracted features. The purpose of our study was to develop a computer-aided diagnostic technique to improve radiologists' performance in differentiating malignant from benign clustered microcalcifications (14).

MATERIALS AND METHODS

Database

One hundred digitized standard-view screen-film mammographic images from 53 patients that showed clustered microcalcifications were analyzed. These images were initially selected to study a computer-aided detection scheme. The original selection criteria were that the clusters of microcalcifications were difficult to detect and that biopsy had been performed (15). Nineteen patients had unilateral breast cancer and 34 had benign breast disease. Of the 19 malignancies, three were classified as ductal carcinoma in situ and 16 as infiltrating ductal carcinoma. The 34 benign lesions were classified as fibrocystic changes or disease (FCD) (*n* = 14), sclerosing adenosis (*n* = 5), FCD and papillomatosis (*n* = 4), FCD and fibroadenoma (*n* = 2), papillomatosis (*n* = 2), fibrosis (*n* = 2), adenosis (*n* = 2), fibroadenoma (*n* = 2), and FCD and sclerosing adenosis (*n* = 1). Two patients had multiple groups of clustered microcalcifications on one or both sides, with one benign lesion found at biopsy in each case. The remaining clusters from these two patients were classified as benign, since biopsy was performed for the cluster that appeared most likely to be malignant. The 100 images from these 53 patients showed 107 cases (40 malignant, 67 benign) of clustered microcalcifications (some of the 55 clusters were shown more than once in different views). Mammograms were digitized to a 0.1-mm pixel and 10-bit gray scale with a laser drum scanner.
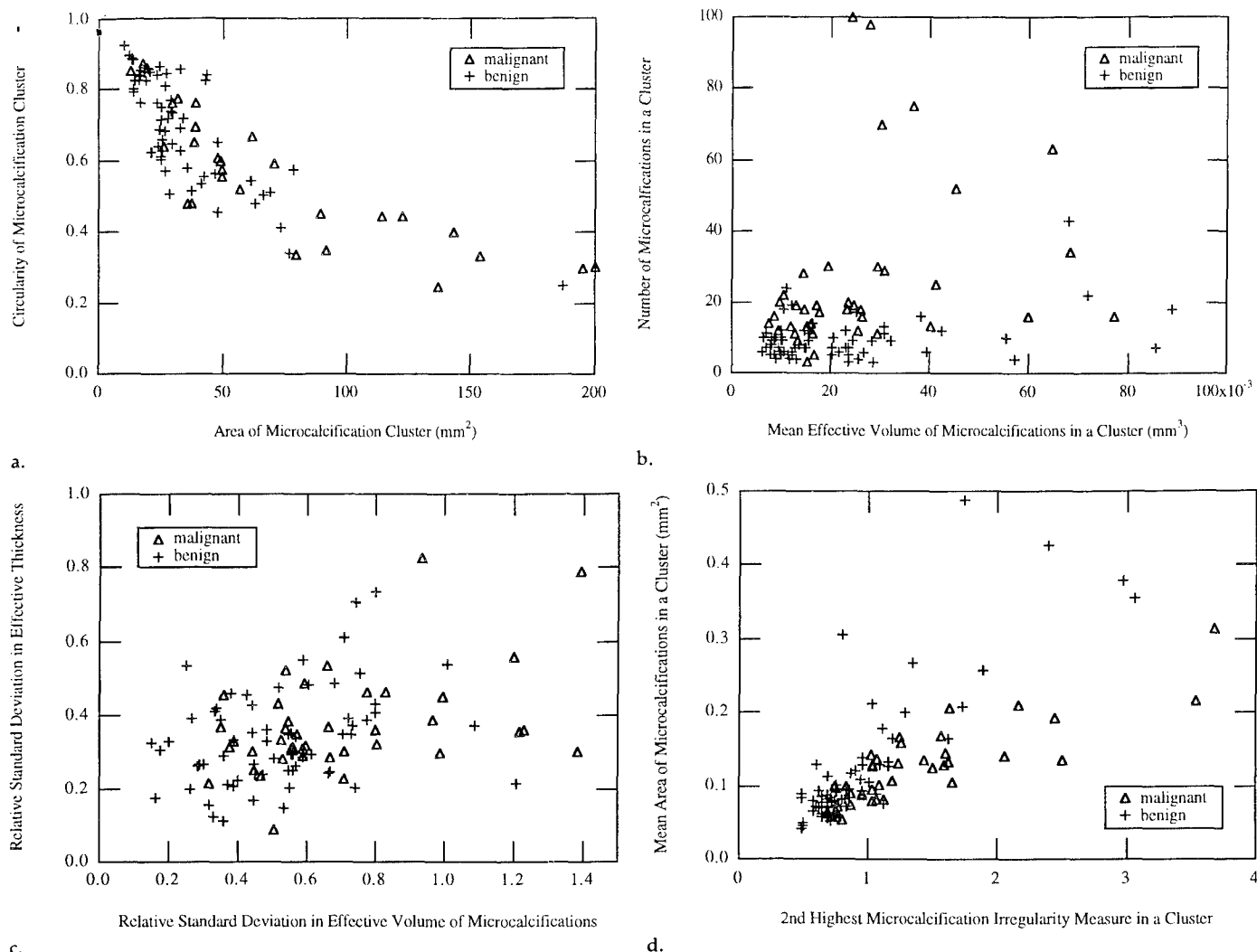
Figure 2. Diagrams of the distributions of malignant and benign clustered microcalcifications in the database for eight arbitrarily paired features. **(a)** Cluster circularity versus cluster area. **(b)** Number of microcalcifications in a cluster versus mean effective microcalcification volume in a cluster. **(c)** Relative standard deviation of effective microcalcification thickness in a cluster versus relative standard deviation of effective microcalcification volume in a cluster. **(d)** Mean microcalcification area in a cluster versus second highest irregularity value of microcalcifications in a cluster.

benign disease only if all clusters on all views were classified as benign.

Observer Study

Five radiologists—three who specialize in mammography and two radiology fellows with some mammographic experience—participated in the observer study. Each observer was presented with one mammogram at a time and asked to estimate the likelihood of malignancy (on a scale of 0–100) on the basis of the clustered microcalcifications. The order of presentation of the 100 images was randomized, except that mammograms from any one patient were carefully separated by other mammograms. The ratings assigned to individual views were treated as independent, and the highest rating among each patient's images was assigned to that patient. ROC curves were generated for each observer for classification of breast cancer or benign breast disease. An additional ROC curve was generated for the five observers as a group by averaging the binor-

mal parameters (*a* and *b*) of the five individual ROC curves (21).

RESULTS

Selected Image Features

Figure 2 shows the distributions of the eight features for all microcalcification clusters in the database. In general, considerable overlap between the malignant and benign clusters was observed. However, in each scatter plot there were some benign or malignant clusters that did not overlap with the others. For example, in Figure 2b there is a group of benign clusters that is closer to the lower left corner of the graph than all malignant clusters. Therefore, these benign clusters can be separated from malignant clusters on the basis of the number of microcalcifications and the mean effective volume of microcalcifications

in a cluster. All eight features could be used to distinguish some benign clusters from malignant clusters. However, the combined effectiveness of the eight features in separating benign from malignant clusters was difficult to visualize graphically, partly because the benign clusters identified with one pair of features were not necessarily the benign clusters identified with other features. The combined usefulness of the eight features could be realized with an artificial neural network.

Performance of Artificial Neural Network

A consistency test for the neural network with use of an identical set of data for both training and testing yielded a perfect ROC curve with an A_z of 1.0. Therefore, 100% of both malignant and benign clusters were cor-
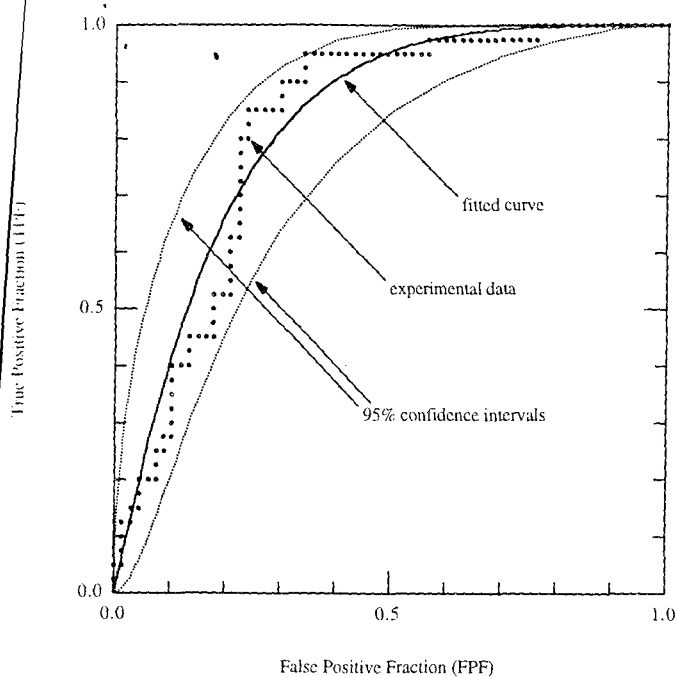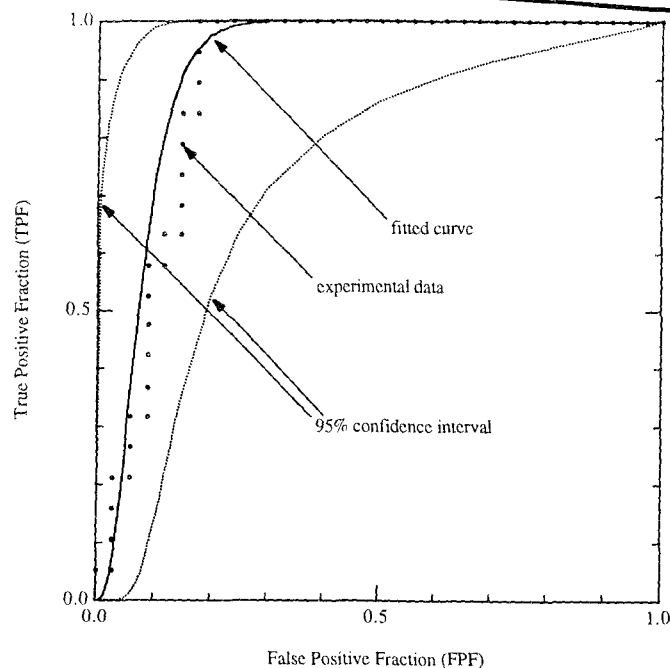
Figures 4, 5. (4) Diagram of the ROC curve for the neural network in the classification of malignant and benign microcalcification clusters. The A_z value for the fitted curve is 0.83. (5) Diagram of the ROC curve for the neural network in the classification of breast cancer versus benign breast disease. The A_z value for the fitted curve is 0.92.
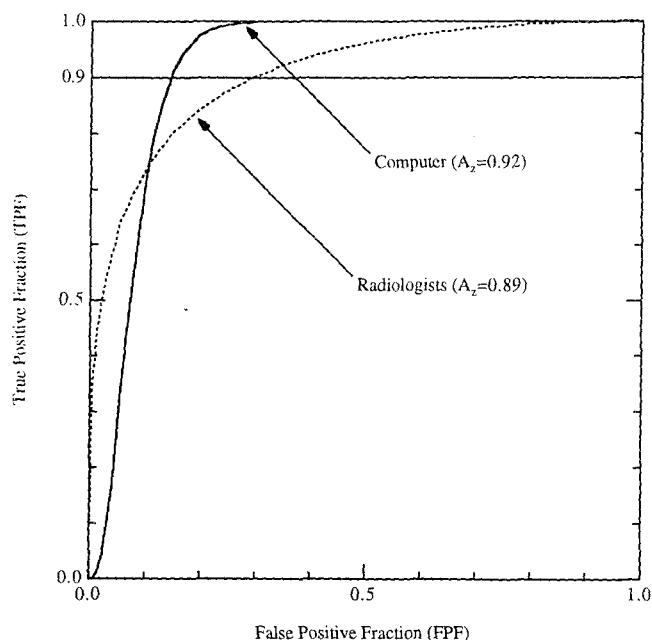

Figure 6. Diagram of the ROC curves for the computerized method and for the combined results of five radiologists reviewing the mammograms retrospectively for classification of breast cancer versus benign breast disease. The areas under the ROC curve above a sensitivity of 0.90 are 0.082 for the computerized method and 0.042 for the five radiologists. Results of the Student t test for the difference in these areas yielded a two-tailed P value of .03.

cision must be made by the radiologist whether to recall the patient for additional examinations, with that decision based only on standard views. Another important point is that although radiologists have become accustomed to making decisions about whether calcifications are benign (not suspicious) or malignant (findings suggest the need for a biopsy) based on magnification views, it would be very helpful to have a method that allows accurate prediction of malignant potential without obtaining additional special views, since these entail an additional appointment, expense, and anxiety for the patient and technical difficulties associated with high-quality magnification imaging. Although radiologists may need magnification views to make these critical decisions, it is evident from our study that the computer does not need these views as much as a radiologist. This is a substantial advantage for computer-aided diagnosis. If computer-aided diagnosis were used at the diagnostic work-up stage, even though the availability of additional views might improve the diagnostic performance of the radiologists, it is likely that this would also improve the performance of the computer. This possibility is being analyzed in another study.

The success of our method is due in part to the choice of features of the microcalcifications and their clusters, which we believe provide good quan-
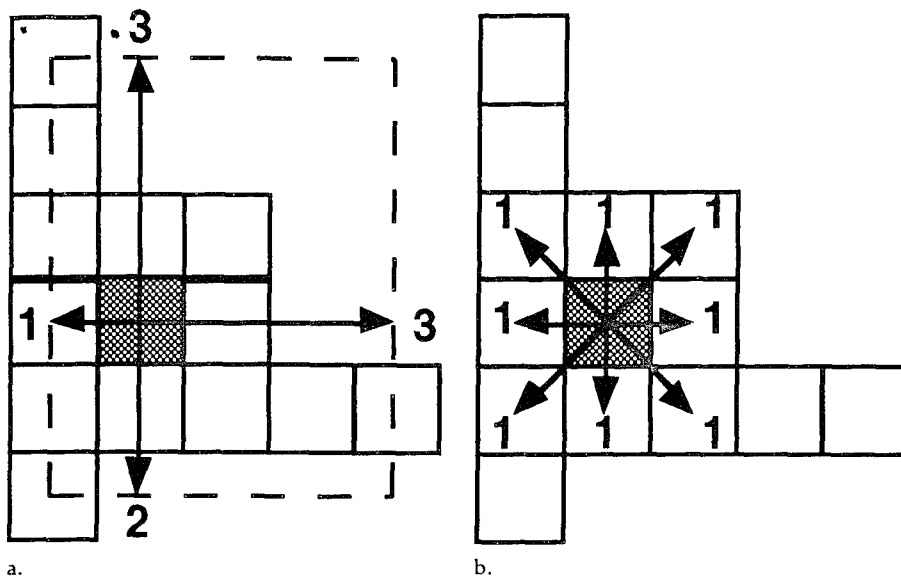
screening setting. It is possible that computer-aided diagnosis might reduce unnecessary callbacks in this setting by accurate identification of findings that are almost certainly benign. Use of the higher estimate of malignancy is clearly the conservative (safer) approach. Our data indicate

that this use may eventually be practical. However, our study primarily demonstrates the efficacy of computer-extracted features to enable accurate mammographic diagnoses made by the computer on the basis of standard-view mammograms. This in itself is a valuable finding, since a de-

a. **b.**

Figure A1. Illustrations of the definition of the shape indexes for individual microcalcifications. **(a)** Four shape indexes are distances (arrows) between the center-of-mass pixel (shaded) of the microcalcification and the boundary of the smallest rectangular box that encloses the microcalcification (dashed lines). **(b)** Eight shape indexes are the maximum lengths of straight lines drawn between the center-of-mass pixel (shaded) and other pixels in the microcalcification in eight directions (arrows).

that encloses all pixels of the microcalcification). Another eight shape indexes were constructed by drawing straight lines in eight directions between the center-of-mass pixel and other pixels within the microcalcification (Fig A1). Each of these eight shape indexes was the maximum length of the line segment drawn in one direction. The standard deviation of these 12 shape indexes was used to identify microcalcifications that are linear or irregular and therefore suggestive of malignancy. For a compact (square) microcalcification, all 12 shape indexes have similar values, and their standard deviation would be small. For an irregular (linear) microcalcification, some of the 12 shape indexes have large values, whereas others have small values. Therefore, their standard deviation would be large.

Cluster Margin

The computer-estimated margin of a cluster was used to calculate the circularity of the cluster, defined as $4\pi A/P^2$, in which A is the area and P is the perimeter of the cluster. This margin was estimated with binary morphologic dilation and erosion operations. A single kernel (constructed from a 5 × 5-pixel square with the 4 corner pixels removed) was used in both the dilation and the erosion operations. In a binary dilation operation, a pixel is assigned the maximum value of its neighboring pixels defined by the dilation kernel. In a binary erosion operation, a pixel is as-

signed the minimum value of its neighboring pixels defined by the erosion kernel. Dilation makes an object larger, whereas erosion shrinks the object. To estimate the margin of a microcalcification cluster, an initial binary image of individual microcalcifications was created: Pixels corresponding to microcalcifications were assigned a value of 1, and background pixels were assigned a value of 0. A sequence of 10 consecutive dilation operations merged the individual microcalcifications into a single object representing the cluster, followed by three erosion operations that reduced the size of the object to a reasonable representation of the cluster's margin. This technique worked well for most clusters in the database. In the exceptional cases, "islands" of microcalcifications did not merge into one cluster because the microcalcifications occupied a large area and were sparsely distributed. This situation was recognized by counting the number of islands within a cluster. If there was more than one island within a cluster, additional dilation operations were automatically performed until a single island was formed. The contours of such clusters deviate from the perceived cluster margins. However, these clusters were both characterized and perceived as having irregular shapes.

Artificial Neural Network

A three-layer, feed-forward, error-back-propagation artificial neural network was used to classify malignant and benign clustered microcalcifica-

tions (26). The neural network had eight input units, a single hidden layer, and one output unit. The input units corresponded to the eight selected features of individual microcalcifications and their clusters. The numerical value of each feature was normalized to the range between 0 and 1 according to the maximum value of the feature in the data set. The optimal number of hidden units was determined empirically. The output of the neural network represented the likelihood of malignancy (0 = benign, 1 = malignant; however, 0.1 and 0.9 were used in training for faster convergence of the neural network). ∎

References

1. Sickles EA. Mammographic features of 300 consecutive nonpalpable breast cancers. AJR 1986; 146:661–663.
2. Hall FM, Storella JM, Silverstone DZ, Wyshak G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. Radiology 1988; 167:353–358.
3. Kopans DB. The positive predictive value of mammography. AJR 1992; 158:521–526.
4. Knutzen AM, Gisvold JJ. Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions. Mayo Clin Proc 1993; 68:454–460.
5. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA. Enhanced interpretation of diagnostic images. Invest Radiol 1988; 23:240–252.
6. Cook HM, Fox MD. Application of expert systems to mammographic image analysis. Am J Physiol Imaging 1989; 4:16–22.
7. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology 1993; 187:81–87.
8. Magnin IE, Alaoui ME, Brémond A. Automatic microcalcifications pattern recognition from x-ray mammographies. Proc SPIE 1989; 1137:170–175.
9. Dhawan AP, Chitre Y, Moskowitz M. Artificial neural network based classification of mammographic microcalcifications using image structure features. Proc SPIE 1993; 1905:820–831.
10. Parker J, Dance DR, Davies DH. Classification of ductal carcinoma in-situ by image analysis of calcifications from mammograms. Proc SPIE 1993; 1905:832–840.
11. Shen L, Rangayyan RM, Leo Desautels JE. Application of shape analysis to mammographic calcifications. IEEE Trans Med Imaging 1994; 13:263–274.

Automated Segmentation of Digitized Mammograms

Ulrich Bick, MD, Maryellen L. Giger, PhD, Robert A. Schmidt, MD,
Robert M. Nishikawa, PhD, Dulcy E. Wolverton, MD, Kunio Doi, PhD

Rationale and Objectives. Fast and reliable segmentation of digital mammograms into breast and nonbreast regions is an important prerequisite for further image analysis. We are developing a segmentation algorithm that is fully automated and can operate independent of type of digitizing system, image orientation, and image projection.

Methods. The algorithm identifies unexposed and direct-exposure image regions and generates a border surrounding the valid breast region, which can then be used as input for further image analysis. The program was tested on 740 digitized mammograms; the segmentation results were evaluated by two expert mammographers and two medical physicists.

Results. In 97% of the mammograms, the segmentation results were rated as acceptable for use in computer-aided diagnostic schemes. Segmentation problems encountered in the remaining 22 images (2.9%) were most often caused by digitization artifacts or poor mammographic technique.

Conclusion. The developed algorithm can serve as a component of an "intelligent" workstation for computer-aided diagnosis in mammography.

Key Words. Computer-aided diagnosis; digital mammography; image segmentation; image processing; digitization.

The advent of digital projection radiography, either as a direct digital modality (e.g., computed radiography) or as film digitization, has opened a variety of new opportunities including digital image processing, digital image storage and transfer, and computer-aided image analysis [1]. For any type of automatic image analysis, it is necessary to first identify a region of interest (ROI; e.g., the breast region in a mammogram). In many previous studies of computer-aided diagnosis (CAD) in mammography, analysis was based on manually selected ROIs [2–6]. Semmlow et al. [7] described a method that automatically detects the breast skin line in xeromammograms with the use of edge detection. However, because of the different image characteristics of xeromammograms, this method is not directly applicable to screen-film mammograms. As part of our CAD scheme in mammography, we previously developed a method for identifying the breast region in mammo-

TABLE 2: Outline and Performance of Segmentation Algorithm

Algorithm Step	CPU Time (sec)[a]
Noise filter	0.3
Calculation of local gray-value range	0.7
Modified global histogram analysis	0.1
Classification of image pixels	0.1
Region growing	0.3
Morphologic filtering	0.2–0.4
Determination of object contour	0.7–1.0
Total performance time	2.4–2.9

CPU = central processing unit.
[a]CPU time on an IBM 570 for 128 × 160 subsampled matrix excluding image data input and output.

(2) directly exposed image region; or (3) potential object (in this case, the breast) pixel (Fig. 1B–D). The local range operator used in our algorithm was based on a 7-pixel-wide ring of 16 pixels. From this neighborhood, the local maximum and minimum pixel values were calculated. A modified, "selective" histogram [16] was constructed including only pixels with a small local range (local maximum minus local minimum), as shown in Figure 2. For a pixel to be classified as a direct-exposure pixel, the following criteria had to be fulfilled: A direct-exposure peak exists in the modified global histogram; the pixel value is close to this direct exposure peak; and
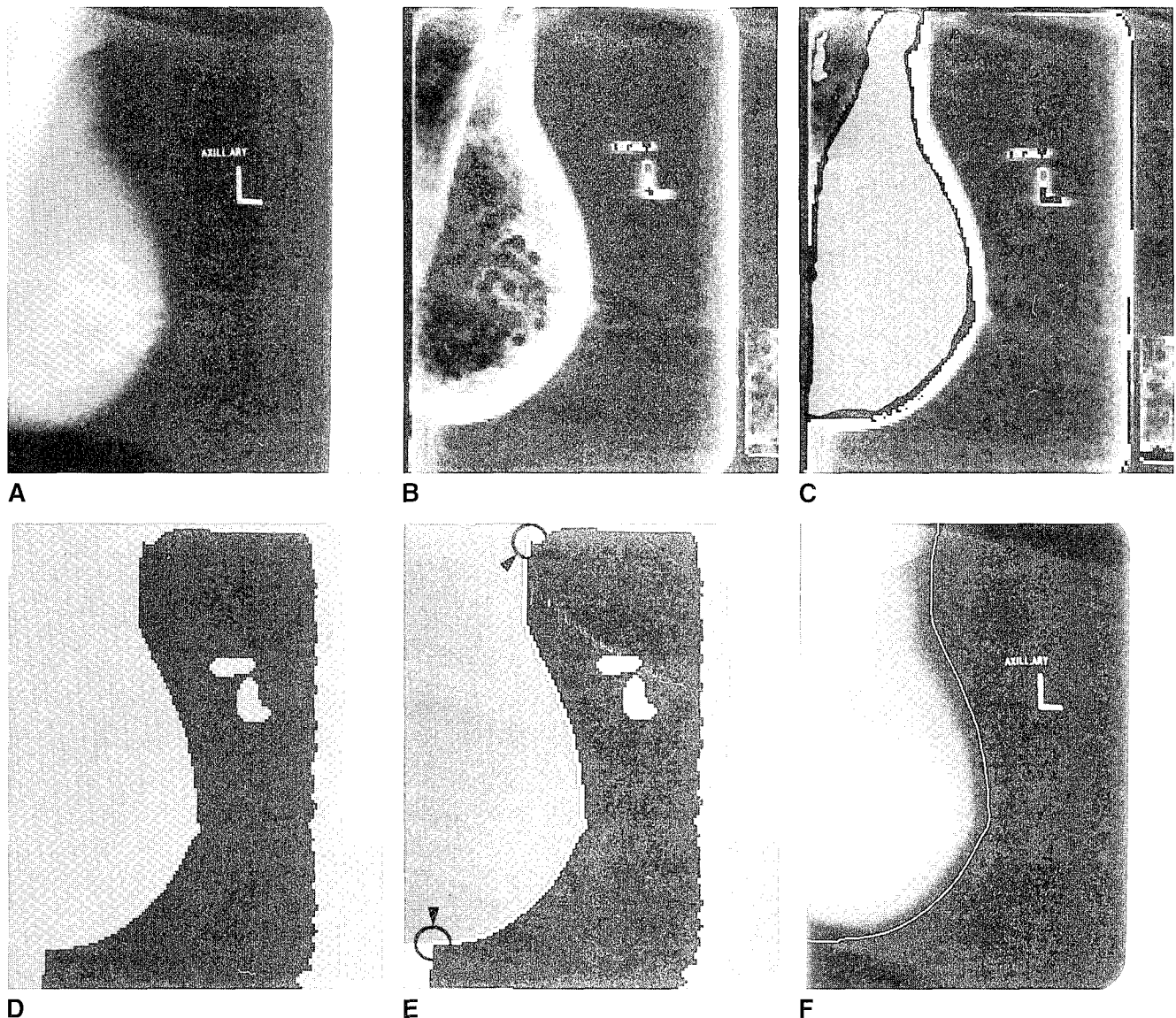

FIGURE 1. Segmentation of digital mammograms. *A*, Original digital mammogram. *B*, Local gray-value range (local maximum minus local minimum) image. *C*, Range image with intermediate density pixels inside the breast already identified as object pixels by the modified global histogram analysis shown as dark gray. *D*, Image after initial pixel classification based on local gray-value range and modified global histogram analysis. *Black* = direct exposure, *gray* = potential object pixels, and *white* = unexposed image region. Note that there is a transition zone of gray potential object pixels along the edge between the direct-exposure and unexposed image region. *E*, Computer-generated breast border. *Arrowheads* mark the connection points from the internal object border (between object and direct-exposure region) to the external object border (between object and unexposed image region). *F*, Computer-generated breast border superimposed on the original image.

In the final step, a closed, 8-point connected border defining the breast region was generated (Fig. 1E). Because, in most cases, a transition zone of intermediate density object pixels is found along the edge between direct-exposure and unexposed image regions, the border generation algorithm has to identify certain "connection" points, where the object border is allowed to connect from the internal object border (between object and direct-exposure region) to the external object border (between object and unexposed region). Potential connection points are identified as points that fulfill the following two criteria: (1) a short connected path of object pixels exists between the connection point along the internal object border and the outside unexposed region and (2) the internal object border forms a concave angle at the connection point, which is smaller than a certain threshold. If more than one isolated object region exists in an image (additional "objects" may represent, for example, letters or the identification label), the breast region can be identified easily as the largest region of connected object pixels. The generated breast border is then expanded by linear interpolation to the original image matrix and smoothed using a running average of the border coordinates. Figure 1F shows the final computer-generated breast border superimposed on the original mammogram.

Evaluation

The testing database consisted of 740 routine clinical screen-film mammograms, including 373 mediolateral oblique and 367 craniocaudal views. One hundred twenty-one images were digitized with the optical drum scanner system A, 350 images with the laser scanner system B, and 269 images with the newer laser scanner system C (for a description of the digitizers, see Table 1). The program was run on all 740 images with a fixed default parameter setting. The computer-generated breast border was superimposed on the original image and displayed on a computer monitor. The segmentation results were subjectively rated by two expert mammographers and two medical physicists and were categorized as follows: (1) optimal—deviations of the computer-generated border from the "true" breast border of less than the sampling distance of 2 mm; (2) minor localized deviations; (3) readily visible deviations—however, results still acceptable for CAD purposes (e.g., no breast parenchymal tissue excluded); (4) substantial deviations—however, overall segmentation is still correct (may influence results of CAD schemes); and (5) complete failure of segmentation (likely to influence CAD results). Examples of minor (category 2) and acceptable (category 3) deviations are shown in Figure 4. During the evaluation, the observers were able to choose between different default window settings as well as manually adjust the window in order to better assess the performance of the segmentation. A chi-square test was used for statistical analysis of the results.

RESULTS

Results, shown in Figure 5, indicate that in more than 97% of the cases, the segmentation results were rated as acceptable for CAD purposes (category 1, 2, or 3). No significant differences in rating ($p = .12$) were found between mammographers and physicists (Fig. 5B). In 22 images (2.9%), the segmentation results were considered unsatisfactory (rated as category 4 or 5 by at least two observers). The most common causes of seg-
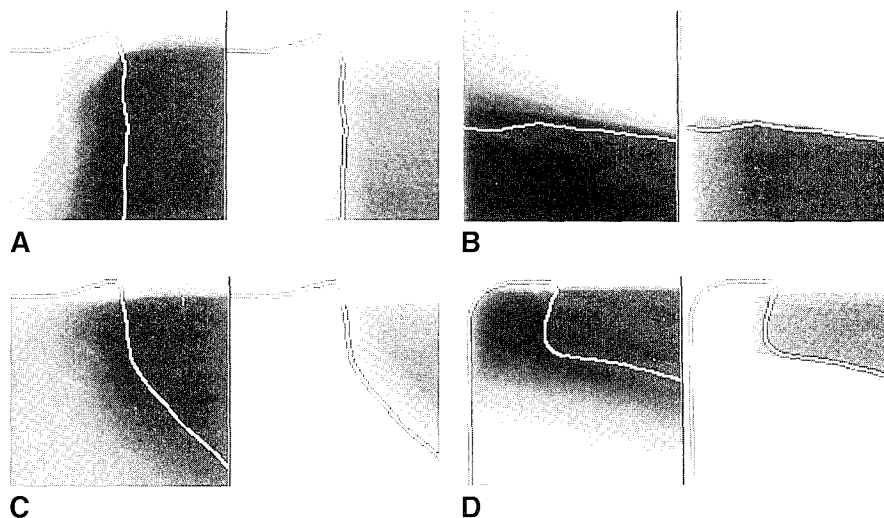

FIGURE 4. Evaluation of segmentation results. **A** and **B**, Examples of minor localized deviations from the "true" breast border (category 2). **C** and **D**, Deviations considered acceptable for computer-aided diagnostic purposes (category 3). All images are displayed in two different window settings with a "normal" wide window (left side) and a second narrow window (right side) showing the dark peripheral breast portion. Note that minor deviations along the skin line can be assessed only on the narrow window image.

dynamic range extending above 3 optical density units, created two typical artifacts that often interfered with the segmentation. In almost all images, a band of pixels with a higher signal intensity, up to 200 pixels (2 cm) in width, was found along the posterior edge of the direct-exposure area (Fig. 7). In some instances, this artifact was so severe that it completely masked the adjacent breast border. This problem was overcome by allowing the final border generation step to connect through such a transition zone of intermediate pixels with a certain maximum width, as described earlier (Fig. 7F). The second artifact was a region of lower signal intensity pixels in the direct-exposure area, which was found only in scanning lines that included the relatively dark identification label (Fig. 3). This led to misclassification of pixels with a large local gray-value range along the edge of this darker direct-exposure region. However, in most instances, these misclassified pixels could be eliminated by the morphologic filtering step (Fig. 3). Such artifacts were not found in the older system B laser scanner or in the system A optical drum scanner. In both of these latter digitizers, however, the small dynamic range often led to poor definition or complete loss of the skin line.

DISCUSSION

To be integrated into an automated, real-time radiographic CAD system, a segmentation algorithm must be fully automated, fast, reliable, and independent of the specific imaging condition (e.g., imaging system, type of image object, image orientation, and exposure conditions). Our proposed algorithm—a combination of a modified global histogram analysis, a gray-value range operator, and region growing—has been shown to fulfill these conditions. With a central processing unit time of

2–3 sec (Table 2), it is fast enough to be implemented in a real-time system. The program does not require any user interaction, and the only prior information necessary for operation is the image pixel size, which is usually included in the image file header after digital image acquisition. In our study of 740 routine clinical mammograms from different sources, 97% of the segmentation results were rated as acceptable for CAD purposes. When analyzing these results, one must remember that the described default program parameters (filter and range operator kernel size, segmentation image matrix), which were held constant throughout the testing, are a compromise between speed and accuracy.

Current mammographic screen-film systems with background optical densities approaching 4 [18, 19] pose a considerable challenge for film digitization systems. Only recently have new laser scanners been developed for medical imaging that are capable of digitizing film with optical densities of 3 or more [20]. In older systems, the dark peripheral parts of the breast and the skin line are often lost or indistinct because of the small dynamic range and a significant increase in digitizer noise in dark image areas [21–24]. Among the digitizer systems used in our study, only the newer system C laser scanner had a dynamic range including optical densities of more than 3 (Table 1). However, this was coupled with typical artifacts in dark image areas, which frequently interfered with the segmentation process (Figs. 3 and 7). These problems may be overcome with new improved digitizer systems [20] or by direct digital mammography [25, 26].

Our algorithm creates an initial raw segmentation of the image and is designed to operate in conjunction with an automatic evaluation of the segmentation results and an optional local contour optimization as shown in Figure 8.

FIGURE 7. Typical example of a system C digitizer artifact. Digitized mammogram displayed in normal (**A**) and narrow direct-exposure (**B**) window setting. Note band of pixels with increased density along the posterior edge of the direct-exposure area. **C–F,** Enlarged region of interest: enlarged original (**C**), border generated with default parameter setting but without being allowed to connect through artifact area (**D**), increased gray-value range threshold (**E**), and after use of the connecting algorithm (**F**). Because of the higher edge strength of the artifact, an increase of the gray-value range threshold led to loss of the skin line (**E**) before the artifact area was eliminated.

7
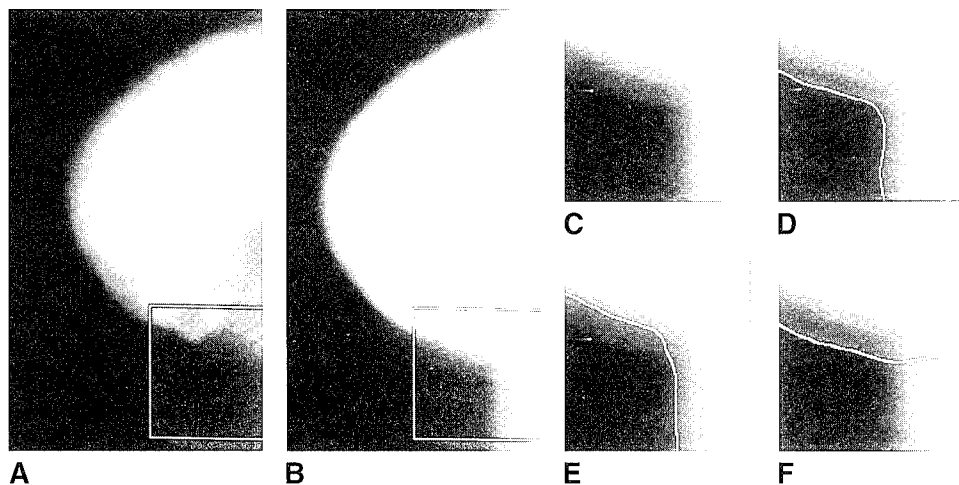
13. Karssemeijer N. Recognition of clustered microcalcifications using a random field model. *Proc Soc Photo-Op Instrum Eng* **1993**;1905:776–786.

14. Ishida M, Kato H, Doi K, Frank PH. Development of a new digital radiographic image processing system. *Proc Soc Photo-Op Instrum Eng* **1982**;347:42–48.

15. Yin F-F, Giger ML, Doi K, Yoshimura H, Xu X-W, Nishikawa RM. Evaluation of imaging properties of a laser film digitizer. *Phys Med Biol* **1992**;37:273–280.

16. Weszka JS. A survey of threshold selection techniques. *Comput Graph Image Process* **1978**;7:259–265.

17. Zucker SW. Region growing: childhood and adolescence. *Comput Graph Image Process* **1976**;5:382–399.

18. Haus A, Gray JE, Daly TR. Evaluation of mammographic viewbox luminance, illuminance, and color. *Med Phys* **1993**;20:819–821.

19. Haus AG. Technologic improvements in screen-film mammography. *Radiology* **1990**;174:628–637.

20. Whiting BR, Muka E, Kocher TE, Flynn MJ. High-resolution, high-performance radiographic film scanner. *Proc Soc Photo-Op Instrum Eng* **1990**;1231:295–305.

21. Fredfeldt KE, Christensen E, Conradsen K, Ersbøll B, Stedstrup S. Automatic screening of plain film mammography. *Semin Ultrasound CT MR* **1992**;13:135–139.

22. Vranckx J, Strul B. Performance evaluation of a laser digitizer. *Proc Soc Photo-Op Instrum Eng* **1987**;767:524–528.

23. Lo S-CB, Gaskill JW, Mun SK, Krasner BH. Contrast information of digital imaging in laser film digitizer and display monitor. *J Digit Imaging* **1990**;3:119–123.

24. Halpern EJ, Esser PD, Nickoloff EL, Alderson PO. A quality-control phantom for digitization of radiographs. *J Digit Imaging* **1990**;3:42–48.

25. Nishikawa RM, Mawdsley GE, Fenster A, Yaffe MJ. Scanned-projection digital mammography. *Med Phys* **1987**;14:717–727.

26. Jarlman O, Samuelsson L, Braw M. Digital luminescence mammography: early clinical experience. *Acta Radiol* **1991**;32:110–113.

27. Pope TL, Read ME, Medsker T, Buschi AJ, Brenbridge AN. Breast skin thickness: normal range and causes of thickening shown on film-screen mammography. *J Can Assoc Radiol* **1984**;35:365–368.

28. Bick U, Giger ML, Huo Z, et al. Automated detection of skin thickening in mammograms. In: Lemke HU, Inamura K, Jaffe CC, Felix R, eds., *Computer assisted radiology: proceedings of the international symposium.* Berlin: Springer, **1993**:461–465.

29. McNitt-Gray MF, Pietka E, Huang HK. Image preprocessing for a picture archiving and communication system. *Invest Radiol* **1992**;27:529–535.

30. Blume H, Roehrig H, Browne M, Ji TL. Comparison of the physical performance of high resolution CRT displays and films recorded by laser image printers and displayed on light-boxes and the need for a display standard. *Proc Soc Photo-Op Instrum Eng* **1990**;1232:97–114.

31. Sezan MI, Tekalp AM, Schaetzing R. Automatic anatomically selective image enhancement in digital chest radiography. *IEEE Trans Med Imaging* **1989**;8:154–162.

32. McAdams HP, Johnson GA, Suddarth SA, Ravin CE. Histogram-directed processing of digital chest images. *Invest Radiol* **1986**;21:253–259.

33. Yoshimura H, Xu X, Doi K, et al. Development of a high quality film duplication system using laser digitizer: comparison with computed radiography. *Med Phys* **1993**;20:51–58.

REPLY TO
ATTENTION OF:

MCMR-RMI-S   (70-1y)   21 Apr 97

MEMORANDUM FOR Administrator, Defense Technical Information
          Center, ATTN: DTIC-OCP, 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 Grant Number DAMD17-93-J-3021. Request the
limited distribution statement for Accession Document Number
ADB210734 be changed to "Approved for public release;
distribution unlimited." This report should be released to the
National Technical Information Service.

2. Point of contact for this request is Ms. Judy Pawlus at
DSN 343-7322.

FOR THE COMMANDER:

                              GARY R. GILBERT
                              Colonel, MS
                              Deputy Chief of Staff for
                                Information Management

Completed 1-10-00