AD ________________

AWARD NUMBER DAMD17-96-1-6012

TITLE: Computer-Aided Classification of Malignant and Benign Lesions on Mammograms

PRINCIPAL INVESTIGATOR: Berkman Sahiner, M.D.,Ph.D.

CONTRACTING ORGANIZATION: Regents of the University of Michigan
Ann Arbor, Michigan 48109-1274

REPORT DATE: May 1998

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1998	Annual (1 May 97 - 30 Apr 98)

4. TITLE AND SUBTITLE
Computer-Aided Classification of Malignant and Benign Lesions on Mammograms

5. FUNDING NUMBERS
DAMD17-96-1-6012

6. AUTHOR(S)
Sahiner, Berkman, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Regents of the University of Michigan
Ann Arbor, Michigan 48109-1274

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this project is to develop computerized classification methods for mammographic abnormalities which will aid radiologists in deciding whether a patient should be biopsied. The regions of interest (ROIs) will be identified by radiologists, and the features to be used in classification will be computer-extracted image features.

In the second year of our project, we have made progress in all five major objectives of the proposal. We have digitized over 600 new films for computerized analysis. We have investigated new segmentation and morphological feature extraction methods for classification of mammographic masses. We have investigated a new high-sensitivity training algorithm for artificial network, and designed a new hierarchical classifier for improved separation between malignant and benign cases. The mass classification algorithm that was developed in the first year of the project was evaluated with six radiologists in an observer study. The results of the observer study indicate that their classification accuracy was significantly improved when they were aided by our computer classification scores. Finally, using an artificial neural network classifier, texture features were evaluated for classification of 86 ROIs containing microcalcification clusters.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES 12

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

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

_X_ 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).

_X_ 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.

PI - Signature ____________ Jun 10, 98
Date

TABLE OF CONTENTS

INTRODUCTION

Treatment of the breast cancer at an early stage is the most significant means of improving the survival rate of the patients. Mammography is currently the most sensitive method for detecting early breast cancer, and it is also the most practical for screening. However, the positive predictive value of mammographic diagnosis is only about 15%-30%. As the number of patients who undergo mammography increases, it will be increasingly important to improve the positive predictive value of mammography in order to reduce costs and patient discomfort.

In this proposal, our goal is to investigate the problem of classifying mammographic lesions as malignant or benign using computer vision, automatic feature extraction, statistical classification, and artificial intelligence techniques. We hypothesize that a second opinion provided by computerized analysis would increase the positive predictive value of mammography, reduce the number of unnecessary biopsies without increasing the number of missed carcinomas, and reduce both cost and patient discomfort.

Our efforts are concentrated on the computer-aided classification of two kinds of breast abnormalities, masses and microcalcifications, which are the primary mammographic signs of malignancy. We are investigating computerized extraction of useful features for the differentiation of malignant and benign cases for both abnormalities, and the application of classical statistical classifiers and newly developed paradigms such as neural networks and genetic algorithms for the classification task. Our purposes are to i) improve existing techniques, devise new methods, and identify the preferred approaches for the classification of mammographic lesions, ii) show that computerized classification of mammographic lesions is feasible, and iii) develop a computerized program that can subsequently be shown to improve radiologists' classification of mammographic abnormalities.

BODY

The progress made so far in the development of the five technical objectives of this project are summarized below. The implications of these results are summarized in the conclusion section.

Technical Objective 1: Database collection

We have continued the collection of mammograms in the second year of this proposal. We have digitized over 600 new films from over 100 patients where each case contained either a biopsy proven mass or a biopsy proven microcalcification cluster. The expert mammographer in this project, Dr. Mark Helvie, is currently reading these films, which involves the identification of the biopsied lesion, and its rating for malignancy and visibility. To date, he has read films of 30 new patients. The digitized mass cases will be used as an independent test set in year three for the evaluation of the classification algorithms developed in this project.

Technical Objective 2: Feature Extraction for Masses:

Segmentation of masses:

Computerized segmentation of mammographic masses is a very important step in our project because all the subsequent processing steps depend on the segmentation results. A clustering algorithm was developed for the segmentation of breast masses in year 1 of the project. In year 2, we evaluated two additional methods for the segmentation of mammographic masses. These methods were (i) a neural network (NN) based segmentation method [1], and (ii) a method based on Gaussian Markov Random Fields (GMRF) [2]. The purpose of both methods was to incorporate neighborhood information into the segmentation process.

NN-based segmentation:

In the neural network-based segmentation method, we formulated the segmentation problem as an optimization problem, and proposed to solve the optimization problem using a Hopfield neural network. In our model, each neuron represented a pixel of the image. A neuron that fired represented a pixel of the segmented mass, and a neuron that did not fire represented a background pixel. The net input to each neuron was modeled as a bias input, plus a constant λ times the sum of the outputs of neighboring neurons. Our formulation was such that when the constant λ was chosen as $\lambda=0$, the segmentation solution coincided with our previous clustering algorithm operating only on one feature (the input image). A larger value of λ meant more interaction of neighboring pixels into the segmentation of a particular pixel.

The segmentation algorithm was tested on a data set of 33 mammograms [1]. The accuracy of the algorithm was evaluated by comparing the computer segmentation with hand segmentations obtained using the expertise of radiologists. Three quantitative measures were examined: Hausdorff distance measure (HD), area overlap measure (AO), and perimeter to area ratio (PAR). The results showed that, compared to clustering, the neural network segmentation provided superior HD and PAR, but inferior AO. We believe that the neural network segmentation is a promising approach. In our initial investigation [1], a one-dimensional feature vector was used for NN-based segmentation. The use of a multi-dimensional feature vector will be investigated in the future.

GMRF-based segmentation

The GMRF-based segmentation technique first estimates texture parameters in the region of interest using the assumption that the textures are Gaussian and fit GMRF models. To compute the texture parameters, the image is divided into overlapping subimages, and sample correlations, which are known to be sufficient statistics under GMRF models, are computed. These parameters constitute the feature vectors for each pixel. Pixels with similar feature vectors are assigned to the same class using a clustering algorithm.

The segmentation algorithm was tested on a data set of 249 mammograms [2]. The accuracy of the algorithm was evaluated by comparing the computer segmentation with hand segmentations obtained using the expertise of radiologists. The AO measure, as well as the fractional background tissue (FBT) measure was used for quantitative analysis. Our results indicated that, compared to the clustering method, the GMRF-based segmentation technique yielded superior AO measure, but inferior FBT measure. The average RMS error between the hand-segmentation and GMRF segmentation was 2.3mm, and the RMS error between the hand-segmentation and clustering was 2.4mm. Although GMRF segmentation seemed to be marginally superior, the difference between the two methods was not statistically significant.

The radiologists who participated in this study were also asked to rate the size of the computerized classification on a scale 1 to 5. A rating of 1 meant 25% or more undersized computer segmentation; 2: 25%-10% undersized; 3: between 10% undersized and 10% oversized; 4: 25%-10% oversized; 5: 25% or more oversized. It was determined that for both methods, less than 10% of the masses received a rating of 1 or 5. This shows that both methods are satisfactory in determining the mass size.

Both methods developed in year 2 for segmentation of mammographic masses yielded satisfactory performance. However, the advantages of these methods did not seem to be significant enough to replace the clustering-based segmentation algorithm in our automatic classification method. We will continue searching for a significantly better segmentation method in year 3.

Morphological feature extraction

In year 1 of the project, a classification method based on texture features was developed. The shape of the mass, and morphological features extracted from this shape are also known to be good indicators of malignancy. In year 2, we developed morphological feature extraction methods in order to take advantage of these indicators.

For morphological feature extraction, boundaries of the masses were manually delineated by two MQSA-approved radiologists. The morphological features evaluated in this study included Fourier descriptors, convexity measures, normalized radial length statistics, contrast, circularity, area, perimeter, and the perimeter-to-area ratio. Our data set included 205 biopsy-proven masses, of which 100 were malignant and 105 were benign. The best two morphological features were the Fourier descriptor summary feature (A_Z=0.87) and the convex hull area measure (A_Z=0.84). When the Fourier descriptor summary feature and four texture features were combined in a linear discriminant classifier, the area under the ROC curve was 0.91 using leave-one-case-out test scores. In comparison, for the classification of the same set of masses, the accuracy of the two radiologists were A_Z= 0.91 and 0.88.

From this study, we conclude that the morphological features extracted from the mass shapes were effective for classification of the masses as malignant or benign. The use of texture features in addition to morphological features in a linear classifier will likely improve the classification accuracy.

To be used in an automatic classification algorithm, the shapes of the masses have to be automatically segmented. In year 3, we will evaluate morphological features extracted from mass shapes obtained by our current segmentation algorithm, and investigate new segmentation methods to improve the classification accuracy of these features.

Technical Objective 3: Feature extraction for microcalcifications

In year 2 of our project, we investigated the use of texture features extracted from spatial gray-level dependence (SGLD) matrices for classification of microcalcifications as malignant and benign.

A region of interest (ROI) containing the microcalcification cluster was identified by an expert radiologist so that only true microcalcification clusters were analyzed. An ROI of 1024 X 1024 pixels (corresponding to 3.58 cm X 3.58 cm on the film) was extracted for analysis. The low-frequency background was subtracted using the background subtraction technique presented in the original proposal. After background correction, SGLD texture matrices were computed as discussed in our original proposal. Forty SGLD matrices were derived from each ROI at different 10 distances and 4 directions. Thirteen features were extracted from each SGLD matrix, and features extracted at axial directions and diagonal directions were averaged. The final texture space therefore contained 260 features for each ROI. The classifier used with these features is described in the discussion of Technical Objective 4, and accuracy of the designed classifier is described in the discussion of Technical Objective 5.

Technical Objective 4: Development of Classifiers

<u>Neural Network Design for Optimization of the Partial Area under the Receiver Operating Characteristic Curve</u>

In classification of mammographic lesions, the cost of missing a malignant case is much larger than that of misclassifying a normal case. The decision threshold therefore cannot be determined without a well-designed cost-benefit analysis. Receiver Operating Characteristic (ROC) analysis is a commonly-used methodology for representing the tradeoff between the true-positive fraction (TPF) and the false-positive fraction (FPF) in a two-group classification task. The area A_{TPF_0} above a sensitivity level TPF_0 under the ROC curve represents the average specificity above that sensitivity level. By maximizing A_{TPF_0}, where TPF_0 is close to 1, one can design a classifier that has good specificity at high sensitivity. In year 1 of the project, a GA-based high sensitivity classifier was designed using this idea. In year 2, we developed a methodology for training a backpropagation neural network (BPN) by maximizing the same criterion.

The training of a BPN via the gradient-descent rule involves the computation of the partial derivatives of the network error with respect to the network weights. For training a high-sensitivity BPN, the network error to be minimized is defined as $-A_{TPF_0}$. With this new error criterion, the partial derivatives cannot be computed analytically. However, using a judicious approximation [3], we were able to represent the partial derivatives fairly accurately.

To test our new BPN training algorithm, we used a randomly-generated Gaussian data set. We studied the effect of the number of hidden layer nodes on the conventional and high-sensitivity training algorithms by varying the number of hidden-layer nodes between 2 and 9. For both training algorithms, the BPN was trained starting from the same initial condition. After 5000 training iterations, test outputs were obtained by applying independent test samples to the trained network. Our results indicated that for small number of hidden-layer nodes, the new training algorithm achieved the goal of decreasing false-positives for a TPF of 0.8 and above. When the number of hidden-layer nodes was increased, the difference between the two training algorithms diminished. This simulation study therefore demonstrated that the new training algorithm would be useful if the number of hidden layer nodes has to be small. This is a frequently encountered condition in practical situations, because the limited number of training samples confines us to use a simpler classifier with fewer hidden-layer nodes.

Development of a hierarchical classifier

A hierarchical classifier, which combines an unsupervised adaptive resonance network (ART2) and a supervised linear discriminant classifier (LDA) was developed for the classification of mammographic masses as malignant or benign. At the first stage, the ART2 network separated the masses based on the similarity of the input vectors. At the second stage, a separate LDA model was formulated within each class to classify the masses as malignant or benign. In a preliminary study to examine the utility of this approach, the ART2 network was presented with texture features that were useful in classifying the masses as spiculated or non-spiculated. The ART2 network classified the masses into three classes, one of which contained predominantly spiculated masses. For each class, stepwise feature selection was used to determine the optimal feature subset for classification of malignant and benign masses using LDA. The areas under the ROC curve for the three classes were 0.94, 0.86, and 0.95. Approximately 48% percent of the benign masses could correctly be identified without missing a malignant mass, compared to 41% with LDA alone.

BPN for microcalcification classification

The texture features described in Technical Objective 4 were used in a backpropagation neural network (BPN) for classification of microcalcification clusters as malignant or benign. First, stepwise feature selection was used to select effective features for classification. Then, several BPN structures were tested their classification accuracy. The BPNs employed a modified delta-bar-delta rule to improve the convergence rate. The number of hidden nodes in the BPNs varied between 1 and 10. In order to make efficient use of the relatively small number of training samples, a leave-one-case-out methodology was used for testing the BPN. The classification accuracy was evaluated by ROC methodology.

Technical Objective 5: Evaluation of classification methods

Classification of masses: Observer study

In year 1 of the project, an algorithm was developed and tested for computerized classification of mammographic masses as malignant or benign [6-8]. In year 2, the effect of this algorithm on radiologist' classification was evaluated using an observer study [9]. Of the 255 films that were used in our previous studies, 15 were used for training the radiologists to use the computer estimation for malignancy, and the remaining 240 were used for the actual evaluation. Six MQSA-approved radiologists assessed the probability of malignancy of the masses with and without CAD. Two experiments, one with single view and another with two views were conducted. The classification accuracy was quantified by the area, A_z, under the ROC curve.

The computer classifier alone distinguished the malignant and benign masses with a test Az of 0.92. The radiologists' Az ranged from 0.78 to 0.91 without CAD and were improved to 0.91 to 0.97 with CAD. For a subset of 77 matched paired views, the radiologists' Az ranged from 0.87 to 0.93 without CAD and were improved to 0.91 to 0.97 with CAD. The improvements were statistically significant with p=0.02 and 0.01, respectively.

Classification of microcalcifications: Computer performance with texture features

The BPN classifiers described in Technical Objectives 4 were tested on a database of 86 mammograms from 54 cases[10]. The Az value obtained with different BPN structures varied between 0.88 and 0.86 with the best feature set. An analysis of the dependence of the classification accuracy on BPN architecture indicated that the BPN with one hidden node provided the best classification accuracy. Since a BPN with a single hidden node is equivalent to a linear classifier, this result appears to indicate that a linear classifier may be optimal with this data set and training samples. However, it has to be emphasized that this observation may not apply when the classifiers are trained with large number of samples. The reduction in classification accuracy with increased number of hidden layer nodes in our current study could have been caused by overtraining with a small sample size.

Using the best classifier in this study, 11 out of 45 benign films could correctly be classified without any false negatives (a sensitivity of 100% and a specificity of 24%). Because some of the films are from the same patient, it is reasonable to make the malignant or benign decision on a case-by-case basis. In this case, 11 of the 28 benign cases could correctly be identified without missing any malignancies (a sensitivity of 100% and a specificity of 39%).

CONCLUSION

In the second year of the proposal, we have made progress in all five major objectives of this proposal:

- Over 600 new films from over 100 patients were digitized for our database.
- Two new methods have been investigated for the segmentation of masses on mammograms [1,2].
- A number of morphological features were developed for classification of mammographic masses. Using mass boundaries delineated by radiologists, it was

shown that these features are effective in classifying masses as malignant and benign [3].

- A neural network training algorithm was designed for training high-sensitivity classifiers. Using simulated data, it was shown that the new algorithm could be more effective than traditional neural network training for improved specificity at high sensitivity [4].
- A hierarchical classifier which combines an unsupervised adaptive resonance network (ART2) and a supervised linear discriminant classifier (LDA) was developed for the classification of mammographic masses as malignant or benign [5].
- The effect of the mass classification algorithm on radiologists' classification was evaluated using an observer study [9]. Using a database of 240 mammograms, it was shown that the radiologists' classification was significantly improved when they were aided by the computerized classification scores.
- Texture features extracted from spatial gray-level dependence matrices were evaluated for classification of microcalcifications as malignant and benign. Using a backpropagation neural network for classification, the area under the receiver operating characteristic curve was 0.88 for a database of 86 films [10].

REFERENCES

[1] S.S. Gopal, B. Sahiner, H.-P. Chan, and N. Petrick, "Neural network based segmentation using *a priori* image models," *Proc. IEEE International Conference on Neural Networks 1997,* 4:2455-2459.

[2] S.S. Gopal, B. Sahiner, H.-P. Chan, N. Petrick and M.A. Helvie, "Evaluation of automated methods for the segmentation of lesion boundaries in mammograms for computer aided diagnosis.," *83rd Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, December 1997.

[3] B. Sahiner, H.-P. Chan, M.A. Helvie, T.E. Wilson, S.S. Gopal, N, Petrick, "Computerized classification of mammographic masses using morphological features," submitted to *84th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, December 1998.

[4] B. Sahiner, H.-P. Chan, N. Petrick, S.S. Gopal, and M.M. Goodsitt, "Neural network design for optimization of the partial area under the receiver operating characteristic curve," *Proc. IEEE International Conference on Neural Networks 1997,* 4:2468-2471.

[5] L. M Hadjiiski, B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie, and M.M. Goodsitt, "Characterization of malignant and benign masses on mammograms based on a hierarchical classifier," submitted to *84th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, December 1998.

[6] B. Sahiner, H.-P. Chan, N. Petrick, M.M. Goodsitt, and M.A. Helvie, "Characterization of masses on mammograms: Significance of the use of the rubber band straightening transform," *Proc. SPIE Medical Imaging 1997*, 3034:491-500.

[7] B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie, and M.M. Goodsitt, "Design of a high-sensitivity classifier based on a genetic algorithm: Application to computer-aided diagnosis," accepted for publication in *Physics of Medicine and Biology,* 1998.

[8] B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie, and M.M. Goodsitt, "Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis," *Medical Physics,* 1998, 25:516-526.

[9] H.-P. Chan, B. Sahiner, M.A. Helvie, C. Paramagul, J. Newman, S. Sanjay-Gopal, N. Petrick, D.D. Adler, M. Roubidoux and T. Wilson, "Effects of computer-aided diagnosis (CAD) on radiologists' classification of malignant and benign masses on mammograms: an ROC study," *83rd Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, December 1997.

[10] H.-P. Chan, B. Sahiner, N. Petrick, M.A. Helvie, K.L. Lam, D.D. Adler, and M.M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network," *Physics of Medicine and Biology,* 1997, 42:549-567.