AD_____

GRANT NUMBER DAMD17-96-1-6226

TITLE: Computer-Aided Diagnosis of Breast Cancer: A Multi-Center Demonstrator

PRINCIPAL INVESTIGATOR: Carey E. Floyd, Ph.D.

CONTRACTING ORGANIZATION: Duke University Medical Center
Durham, North Carolina 27710

REPORT DATE: October 1998

TYPE OF REPORT: Annual

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

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

4. TITLE AND SUBTITLE

Computer-Aided Diagnosis of Breast Cancer:
A Multi-Center Demonstrator

5. FUNDING NUMBERS

DAMD17-96-1-6226

6. AUTHOR(S)

Carey E. Floyd, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Duke University Medical Center
Durham, North Carolina 27710

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, MD 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*

14. SUBJECT TERMS Mammography, Radiology, Computer-Aided Diagnosis, Neural Networks, Malignancy Prediction, Artificial Intelligence, Breast Cancer

15. NUMBER OF PAGES 16

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.

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

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

CJ 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.

NA 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).

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

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

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

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

Carey E Ellyll 20 Sep 98

PI - Signature Date

Computer Aided Diagnosis of Breast Cancer: A Multi-Center Demonstration.
PI: Carey E. Floyd Jr.

Contents

Computer Aided Diagnosis of Breast Cancer: A Multi-Center Demonstration.
PI: Carey E. Floyd Jr.

Statement of Work

(months 1-36)

1)Acquire diagnostic mammography cases from mammography providers distributed over a wide geographical area using the BI-RADS™ findings reporting criteria.

(months 1-6) Develop tools for managing the database and generating reports)

Cases will be acquired from each site and entered into the database as a continual effort.

(months 1-36)

2)Test the existing CAD system on biopsy cases from other mammographic facilities (external to Duke). This testing will be performed on a monthly schedule. The results will be summarized at the end of the first six months and periodically through the project.

3)Develop an ANN to predict biopsy outcome from BI-RADS™ mammographic and history findings for the individual and combined datasets from other mammographic facilities.

(months 1-6) Develop tools for importing cases from the database into the artificial neural network systems.

(months 6-12) Refine the coding of the ANNs to facilitate use with large datasets.

(months 6 -36) Examine the behavior of the different training techniques: cross-validation, bootstrap, and round-robin as the datasets grow in size.

4)Evaluate the difference between the individual and combined networks.

(months 6-36) This work will begin in the first year as the data and tools become available. It will continue throughout the project.

Progress in the second period (months 12-24)

(months 1-36)

1)Acquire diagnostic mammography cases from mammography providers distributed over a wide geographical area using the BI-RADS™ findings reporting criteria.

Progress has been made toward this aim and the progress is on target.

Specifically,

In the second year of this project we have:

1 Ported the database from the FOXPRO database language into ACCESS since the commercial support for FOXPRO has diminished.

2 Searched the Tumor Registry of the Duke University Medical Center Comprehensive Cancer Center to attempt to find cases initially read as benign that turned out to be malignant from the 700 cases from Duke.

3 acquired 500 cases from Sloan-Ketering.

4 acquired 500 cases from U of Maryland.

Cases will be acquired from each site and entered into the database as a continual effort.

(months 1-36)

These cases have been entered into the database and have been examined for completeness and accuracy. About 8% of the records were contradictory or

incomplete and a portion of these were recovered after iteration with the contributing sites.

2)Test the existing CAD system on biopsy cases from other mammographic facilities (external to Duke). This testing will be performed on a monthly schedule. The results will be summarized at the end of the first six months and periodically through the project.

This work has been performed for the 1000 cases from Penn and is reported below.

3)Develop an ANN to predict biopsy outcome from BI-RADS™ mammographic and history findings for the individual and combined datasets from other mammographic facilities.

Done (Reported below)

(months 1-6) Develop tools for importing cases from the database into the artificial neural network systems.

(months 6-12) Refine the coding of the ANNs to facilitate use with large datasets.

(months 6 -36) Examine the behavior of the different training techniques: cross-validation, bootstrap, and round robin as the datasets grow in size.

4)Evaluate the difference between the individual and combined networks.

(months 6-36) This work will begin in the first year as the data and tools become available. It will continue throughout the project.

Publication

In the current period, we published 1 manuscript and 5 abstracts describing work funded in whole or in part by this grant.

Body of Report

We describe an Artificial Neural Network (ANN) approach to computer aided diagnosis of breast cancer from mammographic findings. An ANN has been developed to provide support for the clinical decision to perform breast biopsy. The system is designed to aid in the decision to biopsy those patients who have suspicious mammographic findings. The decision to biopsy can be viewed as a two stage process: 1)the mammographer views the mammogram and determines the presence or absence of image features such as calcifications and masses, 2) the presence and description of these features and the patient's medical history are merged to form a diagnosis. The ANN system is an aid to the second step and is motivated by the large fraction of biopsies that are benign.

While mammography is a sensitive procedure for detecting breast cancer, the positive predictive value (PPV) is low. Only 10-34% of women who undergo biopsy for mammographically suspicious nonpalpable lesions actually are found to have malignancy(Kopans 1992) Between 0.5 - 2.0% of all mammographic exams result in biopsy; several hundreds of thousands of biopsies are performed on benign lesions each year. The women undergoing biopsy for a benign finding are unnecessarily subjected to the discomfort, expense, potential complications, change in cosmetic appearance, and anxiety that can accompany breast

biopsy(Helvie, Ikeda et al. 1991; Dixon and John 1992; Kopans 1992; Schwartz, Carter et al. 1994). In addition, the financial burden of these procedures (between $3000 and $5000 per biopsy) is significant in the present political and economic effort to reduce expenditures. Our system may significantly improve this performance through an ANN approach that utilizes a large database of cases with known outcomes. In clinical practice, this system can be easily integrated into the mammographers' work flow through a computerized reporting system. The clinician reads a mammogram and records the findings into a computer using a standard reporting lexicon (BI-RADS™). The categorical findings for the case are encoded as numerical values and are presented to the ANN as inputs. The ANN produces an output that is associated with the likelihood of malignancy. This fraction is referred to as the malignancy fraction and is an intuitive response that the woman's health care team can then include in the medical decision for biopsy.

In this report we describe in detail the comparison of the model developed on one dataset and evaluated on another. This was the primary goal of the project.

Previous studies at Duke University Medical Center resulted in a computer model to predict breast lesion malignancy based upon BI-RADS mammographic features and the patient age(Floyd, Lo et al. 1994; Baker, Kornguth et al. 1995; Baker, Kornguth et al. 1995; Floyd, Lo et al. 1995; Lo, Baker et al. 1995; Lo, Baker et al. 1995; Lo, Baker et al. 1995; Lo, Baydush et al. 1995; Lo, Grisson et al. 1995; Baker, Kornguth et al. 1996; Lo, Baker et al. 1996; Lo and Floyd 1996; Lo, Kim et al. 1996; Lo, Baker et al. 1997; Lo, Baker et al. 1997; Lo and

Floyd 1997; Lo, Baker et al. 1998). This study evaluates performance of this artificial neural network (ANN) model for cases from an independent institution, the University of Pennsylvania.

At both institutions, consecutive cases of nonpalpable breast lesions which underwent excisional biopsy were selected, resulting in a set of 500 cases from Duke and 1000 cases from Pennsylvania. For each lesion, ten BI-RADS descriptors and the patient age were recorded by expert mammographers. An original ANN was trained and tested on the Duke cases to predict biopsy outcome. With no further adaptation, this ANN was then evaluated on the Pennsylvania cases. The hypothesis was that a network trained on cases from one institution could generalize and accurately predict the outcomes for cases from another institution. To test this hypothesis, The ANN that had been trained on the Duke cases was then evaluated on the cases from Penn. For comparison, another ANN was trained and tested on the Penn data alone. The performance of these three evaluations is presented below.

The ANN that was trained and tested on the Duke cases alone performed with ROC area of 0.86 ± 0.02. the ROC curve for this network is shown as the solid line in fig. 1. The ANN that was trained and tested on the Penn cases alone performed with ROC area of 0.82 ± 0.02. The ROC curve for this network is shown as the long dashed curve in fig. 1. These results suggest that the Pennsylvania cases alone are more challenging to describe with an ANN model than the Duke cases. When the network trained on Duke cases was evaluated on the Penn cases, an ROC area of 0.79 ± 0.01 was obtained. This curve is plotted as the short dashed curve in Fig. 1.

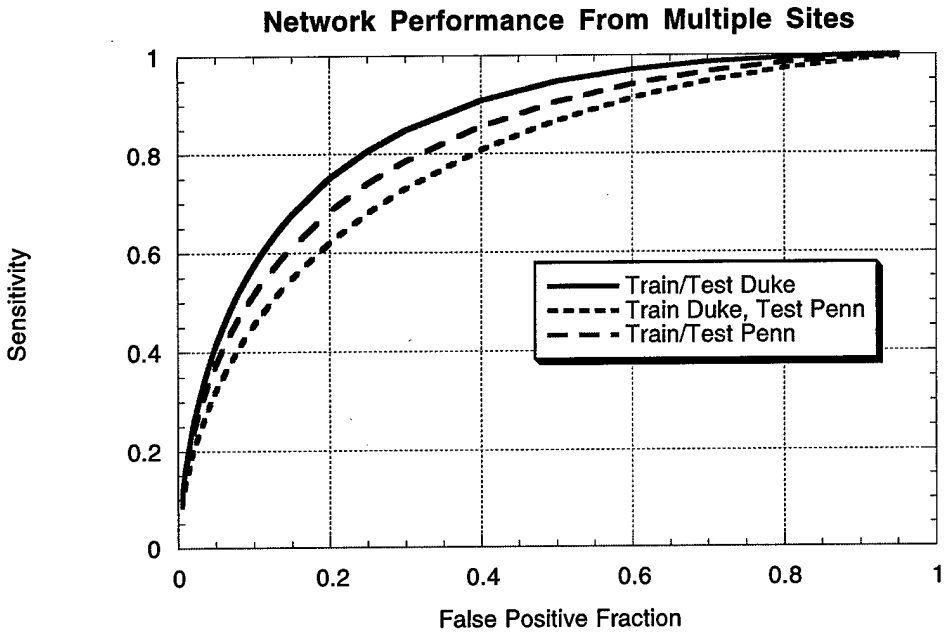

Fig. 1 ROC comparison of the three ANN evaluations.

While ROC area is the most common criteria for comparing two diagnostic systems, using this criteria assumes an equal "cost" for miss-classifying a positive and a negative case. For the medical decision of whether to biopsy a suspicious region in a breast, the cost of missing a true cancer is higher than the cost of performing a biopsy on a benign region. While a cost-benefit analysis is the best technique for evaluating such a problem, this is beyond the scope of this project. In clinical practice, no decision aid will be accepted that performs with less than a high sensitivity. The performance of the ANNs was evaluated by comparing the specificity at a high fixed level of sensitivity of 98%. The performance of the three systems is compared in table 1 where is shown the ROC are (Az), the specificity at 98% sensitivity, the Positive predictive value

(PPV), the number of malignancies missed and the number of benign biopsies saved.

Table 1

Comparison of the performance of the systems.

train / test	spec. at 98% sens	cancers missed	biopsies obviated	PPV	Az
Duke MDs	12%	4 / 174	39 / 326	35%	0.82 ± 0.02
Duke / Duke	42%	3 / 174	136 / 326	47%	0.87 ± 0.02
Penn / Penn	15%	7 / 396	90 / 604	43%	0.82 ± 0.01
Duke / Penn	18%	7 / 396	107 / 604	44%	0.79 ± 0.01

Table 1 Comparison of the performance of the systems.

Row 1:

The "Duke network" (trained on Duke 500, tested on Duke 500) improved specificity at 98% sensitivity over Duke MDs dramatically, from 12% to 42%. There was also an improvement in PPV from 35% to 47% and in Az from 0.82 to 0.87 (p=0.08).

Row 2:

The "Penn network" (trained on Penn 1k, tested on Penn 1k) simulating effect of customizing an ANN just for Penn showed much lower specificity (15%) and somewhat lower PPV (43%) and Az (0.82) compared to the Duke net. This is all

consistent with the Penn data set being inherently more challenging as noted above.

Row 3:

The "Cross network" (trained on Duke 500, tested on Penn 1k) simulating effect of cross-institution application showed similarly poorer performance. In particular, specificity was 18% and Az only 0.79. It should be noted however that the performance was almost identical to that of the Penn net. No matter if the ANNs were trained on Duke or Penn cases, both performed equally poorly on the Penn cases. In other words, the limiting factor may be the inherent difficulty of the Penn cases, not the ANN's inability to generalize. This is encouraging because nothing was gained by customizing the ANN specifically for Penn. If we can learn how to characterize the Penn cases better, we can probably make an ANN that will generalize better as well. Another encouraging observation: the cross net maintained the same high PPV as the duke net and Penn net (all in 40's). All 3 PPVs were much higher than that of original Duke MDs.

CONCLUSION: The ANN that was trained on Duke cases alone generalized successfully to a relatively large, independent data set. The performance was comparable to or better than that of the radiologists at that institution, and only slightly worse than a new ANN specifically optimized for the new cases. This breast cancer prediction model thus shows potential to be applied in other institutions which also utilize the standardized BI-RADS mammography lexicon, and it may help reduce the number of unnecessary biopsies.

References

Baker, J. A., P. J. Kornguth, et al. (1996). "Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description." AJR. American Journal of Roentgenology **in press**.

Baker, J. A., P. J. Kornguth, et al. (1995). "Artificial neural network: improving the quality of breast biopsy recommendations." Radiology **198**: 131-135.

Baker, J. A., P. J. Kornguth, et al. (1995). "Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon." Radiology **196**(3): 817-822.

Dixon, J. M. and T. G. John (1992). "Morbidity after breast biopsy for benign disease in a screened population." Lancet **1**: 128.

Floyd, C. E., Jr, J. Y. Lo, et al. (1995). "Interactive computer-aided diagnosis of breast cancer." Radiology **197(P)**: 533.

Floyd, C. E., Jr, J. Y. Lo, et al. (1994). "Prediction of breast cancer malignancy using an artificial neural network." Cancer **74**(11): 2944-2948.

Helvie, M. A., D. M. Ikeda, et al. (1991). "Localization and needle aspiration of breast lesions: complications in 370 cases." AJR. American Journal of Roentgenology **157**: 711-714.

Kopans, D. B. (1992). "The positive predictive value of mammography." AJR. American Journal of Roentgenology **158**: 521-526.

Lo, J. Y., J. A. Baker, et al. (1996). "Artificial neural networks for the prediction of breast cancer invasiveness by using Breast Imaging and Reporting Data System mammography lexicon." Radiology **201(P)**: 370.

Lo, J. Y., J. A. Baker, et al. (1997). "Predicting breast lesion malignancy and invasion using the BI-RADS mammography lexicon." Radiology **205(P)**: 447.

Lo, J. Y., J. A. Baker, et al. (1995). "Application of artificial neural networks to the interpretation of mammograms based on the radiologist impression and optimized BI-RADS™ image features." Radiology **197(P)**: 242.

Lo, J. Y., J. A. Baker, et al. (1995). "Computer-aided diagnosis of breast cancer: artificial neural network approach for optimized merging of mammographic features." Academic Radiology **2**(10): 841-850.

Lo, J. Y., J. A. Baker, et al. (1995). Computer-aided diagnosis of mammography: Artificial neural networks for optimized merging of standardized BIRADS features. World Congress on Neural Networks 95 (International Neural Network Society Annual Meeting), Washington, D.C.

Lo, J. Y., J. A. Baker, et al. (1998). "Effect of patient history findings on predicting breast cancer from mammograms using artificial neural networks." Academic Radiology **submitted**.

Lo, J. Y., J. A. Baker, et al. (1997). "Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features." Radiology **203**(1): 159-163.

Lo, J. Y., A. H. Baydush, et al. (1995). "Computer-aided diagnosis of breast mass malignancy using automated feature extraction and artificial neural networks." Radiology **197(P)**: 425.

Lo, J. Y. and C. E. Floyd, Jr (1996). Analysis of error surfaces of neural network applied to computer-aided diagnosis in mammography. World Congress

on Neural Networks '96 (International Neural Network Society 1996 Annual Meeting), San Diego, CA, Lawrence Erlbaum Associates, Inc.

Lo, J. Y. and C. E. Floyd, Jr (1997). Self-organizing maps for analyzing mammographic findings. IEEE International Conference on Neural Networks, Houston, TX, IEEE.

Lo, J. Y., A. T. Grisson, et al. (1995). Computer-aided diagnosis of mammograms using an artificial neural network: merging of standardized input features from the ACR lexicon. SPIE Medical Imaging 1995: Image Processing.

Lo, J. Y., J. Kim, et al. (1996). Computer-aided diagnosis of mammography using an artificial neural network: Predicting the invasiveness of breast cancers from image features. SPIE Medical Imaging 1996: Image Processing.

Schwartz, G. F., D. L. Carter, et al. (1994). "Mammographically detected breast cancer: nonpalpable is not a synonym for inconsequential." Cancer **73**: 1660-1665.