AD_____

Award Number: DAMD17-94-J-4371

TITLE: Computer Aided Breast cancer Diagnosis

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

CONTRACTING ORGANIZATION: Duke University
Durham, North Carolina 27710

REPORT DATE: October 1999

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, 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 1999	Annual (23 Sep 98 - 23 Sep 99)

4. TITLE AND SUBTITLE
Computer Aided Breast cancer Diagnosis

5. FUNDING NUMBERS
DAMD17-94-J-4371

6. AUTHOR(S)
Carey E. Floyd, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Duke University
Durham, North Carolina 27710

E*Mail: cef@deckard.mc.duke.edu

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, 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*

While biopsy is a sensitive and specific test for breast cancer, to achieve a high sensitivity for cancer, many women with mammographic findings due to benign processes undergo biopsy. To improve the specificity o f the decision to recommend biopsy and reduce the number of benign biopsies performed, a computer decision aid is being developed.

ANN systems for prediction have been evaluated using a database of mammographic cases that were sent to biopsy with the results known. A case findings matching algorithm was implemented using a relational database to simplify and speed the coding. A Case-Based Reasoning approach was selected for this study since we wished to examine the cases and the similarity between them. To classify a given test case as benign or malignant, the case is compared to all previous cases, selecting those cases with were similar with regards to their findings. A decision variable was formed as the "malignancy ratio" computed as the ratio of the number of malignant cases to the total number of similar or "matched" cases.

The system performed with an ROC area of 0.77. As described here, the system performs better than chance but poorer than the performance reported for radiologists on these data.

14. SUBJECT TERMS 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.

____ 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.

_X_ 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.

Carey E Gold 25 Feb 90

PI - Signature Date

Table of Contents

Report of the progress on Grant DAMD17-94-J-4371

for the period October 1998 to October 1999.

Progress in this funding period: October 1998 to October 1999 is demonstrated through the 8 publications supported in part by this grant.

The publications included four peer-reviewed journal articles (2 published, 2 in press), three manuscripts in conference proceedings, and two conference abstracts for presentations at the RSNA meeting.

Peer-reviewed Journal Manuscripts

Baker JA, Frederick ED, Lo JY, Kornguth PJ, and Floyd CE Jr. Incorporation of an Artificial Neural Network into Clinical Mammography to Reduce Benign Breast Biopsies. *AJR* 170:84, 1998.

Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr. Effect of Patient History Data on the Prediction of Breast Cancer from Mammographic Findings with Artificial Neural Networks. *Acad Radiol*, 6;10-15; 1999.

Floyd CE Jr, Lo JY, Tourassi GD. Breast Biopsy: Case-Based Reasoning Computer-Aided Using Mammography Findings for the Decision to Biopsy. In Press to *American Journal of Roentgenology* (AJR).

Gavrielides MA, Lo J, Vargas-Voracek R, Floyd CE Jr. Segmentation of suspicious clustered microcalcification in mammograms *Medical Physics.*

Conference Proceedings Manuscripts

Vargas-Voracek R, Floyd CE Jr. Computer-Aided Diagnosis for Early Detection of Breast Cancer from Mammograms. Susan G. Komen Breast Cancer Foundation "Reaching for the Cure" National Grant Conference. (1998).

Vargas-Voracek R, Floyd CE Jr. Hierarchical Markov-Random Field Texture Modeling for Mammographic Structure Segmentation Using Multiple Spatial and Intensity Image Resolutions. Presented at the 1999 Medical Imaging Symposium. International Society for Optical Engineering (SPIE). February 20-26, 1999.

Tourassi GD, Floyd CE Jr, Lo JY. A Constraint Satisfaction Neural Network for Medical Diagnosis. Presented at the 1999 International Conference on Neural Networks (ICNN), Washington, DC.

Conference Abstracts

Vargas-Voracek R, Floyd CE Jr. Markov-Random Field Texture Model for Automatic Breast Parenchyma Characterization. Presented at the 84th Radiological Society of North America (RSNA) Scientific Assembly and Annual Meeting. November 29-December 4, 1998.

Lo JY, Kornguth PJ, Floyd CE Jr. Multi-Institution Evaluation of BIRADS Breast Cancer Prediction Model. *Radiology* 209(P):271, 1998.

Introduction

Biopsy is considered to be a definitive test to rule out breast cancer for those patients who participate in breast screening examinations and whose mammograms are interpreted as having suspicious findings. Excisional biopsy is a sensitive and specific test for breast cancer[1]. If the cost of excisional biopsy were minimal, this would be an ideal test for breast cancer malignancy. Unfortunately, the cost of this procedure in both monetary and emotional terms, is significant[2,3,4]. Unfortunately, to achieve a high sensitivity for detecting cancer, many women with mammographic findings due to benign processes undergo biopsy. The false positive rate for the decision to biopsy is currently between 66% and 90%. The goal of the work described here is to design a decision tool to support the decision to biopsy. This decision aid must maintain the current high detection rate for true cancers while accurately ruling out some of the benign cases and thus avoiding unnecessary biopsies.

The problem of classifying suspicious mammographic lesions as benign or malignant is recognized as a difficult practice. There is considerable variation in the skill with which the task is achieved even within the group of radiologists who specialize on this task. The radiographic manifestation of breast cancer is not well enough understood from a fundamental scientific basis to allow an

accurate theoretical predictive model to be constructed from first principals. There is no accurate deterministic model for relating mammographic findings to biopsy outcomes although some general rules are accepted. Examples of these rules are "Older women are more likely to develop breast cancer than young women." "If the margin of a mass appears well circumscribed, the mass is likely to be benign." Unfortunately, the sensitivity and specificity of those rules that are generally agreed upon is not sufficient to allow a strict implementation over the full range of cases that are encountered in clinical practice. While rule based expert system s have enjoyed success in some medical diagnostic tasks, and there are expert mammographers whose diagnostic performance would qualify them as experts, the construction of rule-ased expert systems for this diagnostic task has met with limited success. This is quite possibly due to the difficulty of describing the logical and analytic process used by the experts in a form that can be used by other mammographers. Atypical difficulty with the expert system approach is the description and encoding of the input data: two radiologists often will use similar, but not exactly the same descriptions for a given lesion. It is often difficult to overcome instability in a model due to this potential ambiguity in the input data. Another difficulty for strict rule-based systems is that the descriptors used as inputs to the model can be nonspecific: two lesions with similar descriptions can have opposite outcomes at biopsy. These arguments indicate that an example-based technique would be more appropriate. This is supported by realizing that radiologists are trained by repeatedly examining sample cases with known outcomes that are maintained in a medical school's teaching files. The focus of this research has been in developing and evaluating data driven models, specifically artificial neural

networks (ANN) for the task of predicting the outcome of biopsy given the description of mammographic lesions as inputs. This work has been facilitated by the growing acceptance of BI-RADS as a standardized lexicon for mammographic case reporting.

ANN systems for the prediction task have been constructed and evaluated using a growing database of mammographic cases that were sent to biopsy with the results known. This work has been successful and has resulted in xxx peer-reviewed publications, xxx invited presentations, xxx competitave presentations, and has served to foster further research efforts for constructing decision aids for the diagnosis of breast cancer as demonstrated by the 11 funded grants that have been awarded since the start of this project. The ideas that developed into these other grants were realized from this work.

In the latter part of this research grant, the investigative emphasis has shifted from algorithm development to clinical evaluation reflecting the shift from specific aim one to specific aim two. After conducting several preliminary sessions with mammographers using the system including three years of presentation of a live computer demonstration version at the Radiological Society of North America InfoRad exhibit, several important question shave been identified regarding the user interface of the system. The first is the question of how the results should be presented to the mamographer. An informal exit interview with mammographers who used the system indicated that 70% preferred a probabilistic output where the mammographer would be given a number between 0 and 100 to indicate the estimated percent probability tha tth ecase in question was malignant. The other 30% of he users did not want a

probability, they wanted a hard decision to biopsy or not to biopsy. For these clinicians, the system threshold would be set to some value and the binary result would be presented. All users, especially those preferring the hardwired decision threshold, desired some indication of the certainty with which the ddecision was presented. Several individuals expressed an interest in being presented with "similar" cases from which the neural network was trained. These reasonable requests initiated the ideas that lead to the development of the CBR. In an effort to provide similar "example cases", it was realized that cases with similar findings would generate similar ANN outputs even though these would not provide a complete or unique subset. A case findings matching algorithm was implemented using a relational database (MicroSoft ACCESS ™) to simplify and speed the coding. It was later found that this implementation also dramatically improved the speed of execution. With this case matching tool, different definitions of what constituted a similar case could be investigated. It was found that depending on how strict the definition of similarity, the existing database could provide between 10 and 100 similar cases for each cases investigated. While beyond the scope of this investigation, given these cases identifications, it would be straightforward to present digital version of the cases on a monitor to provide partial explanation of the ANN result. It was natural to compute the fraction of malignancies to total cases within the set of matched cases. With the use of this fraction as a decision variable, a predictive tool was naturally implemented. While the evolution of this technique proceeded as described above from an effort to provide explanation to the mammographer for the recommendation suggested by the ANN, it was soon recognized that the

resulting algorithm was an instantiation of a simple case-based reasoning system.

A preliminary investigation was performed to better understand the relationship of findings to malignancy within the framework of the BI-RADS reporting lexicon. A Case-Based Reasoning (CBR) approach was selected for this study since we wished to examine the cases and the similarity between them. In this context, a CBR was developed and evaluated by its ability to predict the outcome of biopsy from mammographic findings reported in the BI-RADS lexicon. To classify a given test case as benign or malignant, CBR was implemented by comparing the case to all previous cases, selecting those cases with were similar with regards to their findings and examining the outcomes for those similar cases. A decision variable was formed as the "malignancy ratio: computed as the ratio of the number of malignant cases to the total number of similar or "matched" cases. Performance was evaluated by generating an ROC curve from the true positive fraction and the false positive fraction as the threshold was applied to the malignancy ratio.

The system is implemented as follows. The mammograms are read by clinicians using a standard reporting lexicon (BI-RADS™). These findings are compared to a database of findings from cases with known outcomes (from biopsy). The fraction of similar cases that were malignant is returned. The clinician can then consider this result when making the decision regarding biopsy. . This malignancy fraction is an intuitive measure that can be readily included in the

medical decision. This approach is intuitive. The CBR answers the question "Of all cases that are similar to this one, how many were malignant at biopsy?" This process is similar to that followed by the clinician when considering the same problem.

Methods

The CBR was implemented as a case retrieval engine in a relational database framework. In this context, it functions as a query of a table of cases and outcomes. To predict the outcome for a new test case, the test case is compared to each case in the database through a matching rule. The prediction is the ratio of the number of malignant to the total number of cases that match.

The components of the system include the case encoding and the matching rule. The cases are encoded through a subset of the categorical BI-RADS^{TM} image findings and the patients' age. For the initial experiment, similarity is defined as an exact match of some subset of the findings. The database has been described previously[5] and was restricted for this feasibility study to the first 500 cases since the properties of this set were well understood and numerous previous studies had been conducted on this set.. Of these 500, 232 of the cases described masses, 192 cases described microcalcifications, and 29 cases described masses and microcalcifications associated that were associated with the mass. The remaining 47 cases did not describe either a mass or a calcification but were architectural distortions, asymmetric breast density, focal asymmetric density, and/or asymmetric breast tissue. Malignant outcomes were reported for 174

(35%) of the biopsies while 326 (65%) were found to be benign resulting in a Positive Predictive Value (PPV) of 35%.

The input features were restricted to those that had been found to have the highest independent predictive power in our earlier studies and included the patient age, and , for masses, the mass margin, mass size, mass density, and mass shape, while for calcifications included calcification description, calcification number, calcification distribution, and special cases/associated findings.

Performance of the predictive system was evaluated through a round-robin technique in which: a test case is selected from the dataset, the database is formed from all of the other remaining cases, all of these remaining cases are compared to the test case and those that match are selected. The malignancey fraction is found for the set of matching cases. The testing example is replaced in the set and another is removed, the resulting system is evaluated and this is repeated until all examples have been used as testing cases. A threshold is applied to the set of malignancy fractions and the sensitivity and false positive fraction are plotted as the threshold is applied at each value of the malignancy fraction. A Receiver Operating Characteristic ROC curve is plotted from these ordered pairs. The ROC areas were computed from the resulting curve using Newton's method of integration. Used as a performance measure, ROC area gives equal significance to the sensitivity and specificity resulting from the application of a specific threshold. It is clear that sensitivity has higher priority than specificity for this problem since, while there is a need to reduce the number of benign biopsies, there is a greater cost incurred by missing a malignancy than by performing a biopsy on a benign lesion. As a consequence, a more appropriate measure could be formed by concentrating on the performance at

high sensitivity. To concentrate on this region, three other measures are presented: the partial ROC area reported for sensitivity greater than 90% , and , the specificity at sensitivities of 98% and 100%.

Results

In these terms, the performance is presented in table 1 and described below. From a previous study on the predictive power of the findings using linear discrimanent analysis (LDA) the following six findings were found to contribute significantly: Age, Mass Margin, Mass Density, Calcification Description, Calcification Distribution, and Associated Findings. Requiring an exact match on all six features resulted in an ROC area of 0.77 but with very poor (<1%) specificity at high sensitivities of 90% and higher.

Table 1

Performance of Case Based Reasoning				
Matching Rule	ROC Az	Partial ROC Az	Specificity at 100% Sensitivity	Specificity at 98% Sensitivity
6 findings	0.77	0.016	<0.01	0.012

Table 1 CBR performance.

The histogram for the malignant (positive) and benign (negative) cases as a function of malignancy fraction is shown in fig. 1. The striped boxes indicate the negative cases while the solid boxes show the positive cases.

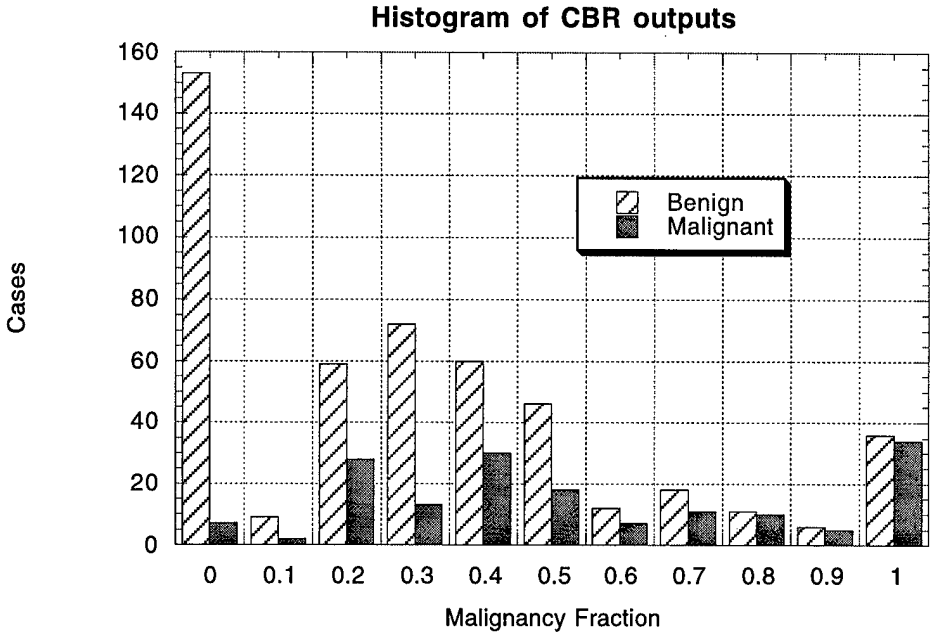

Figure 1 Histogram of CBR outputs.

The ROC curve is shown in fig.2.

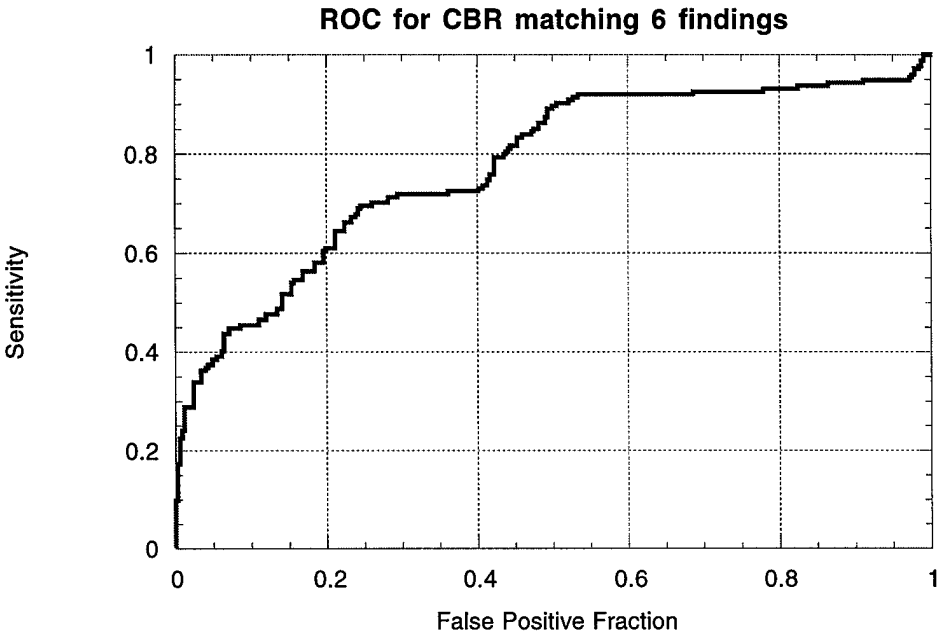

Figure 2 ROC curve for CBR with exact match of 6 findings and age within 5 years.

Less than 0.12 seconds are required to predict the malignancy ratio for a new case with the system running in a non-optimized ACCESS™ (Microsoft Inc., Redmond Washington) database language on a Pentium II 300Mhz personal computer.

Discussion

The simple system described above was developed as a natural result of our efforts to develop a user interface for the clinical evaluation of the artificial neural network for reducing benign biopsies. As described here, it performs better than chance but poorer than the performance reported for radiologists on these data[5]. No optimization was performed to refine the matching rule. Future

work will examine other matching rules and will compare the performance to that of the artificial neural networks.

References

1. Kopans DB. The positive predictive value of mammography. *AJR. American Journal of Roentgenology* **1992**; 158: 521-526
2. Dixon JM and John TG. Morbidity after breast biopsy for benign disease in a screened population. *Lancet* **1992**; 1: 128
3. Helvie MA, Ikeda DM, and Adler DD. Localization and needle aspiration of breast lesions: complications in 370 cases. *AJR. American Journal of Roentgenology* **1991**; 157: 711-714
4. Schwartz GF, Carter DL, Conant EF, Gannon FH, Finkel GC, and Feig SA. Mammographically detected breast cancer: nonpalpable is not a synonym for inconsequential. *Cancer* **1994**; 73: 1660-1665
5. Lo JY, Baker JA, Kornguth PJ, and Floyd CE, Jr. Effect of patient history findings on predicting breast cancer from mammograms using artificial neural networks. *Academic Radiology* **1998**; 7.