

AD _____

Award Number: DAMD17-94-J-4076

TITLE: Development of a Common Database for Digital Mammography
Research

PRINCIPAL INVESTIGATOR: Robert M. Nishikawa, Ph.D.

CONTRACTING ORGANIZATION: The University of Chicago
Chicago, Illinois 60637

REPORT DATE: October 1999

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for public release;

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

20010620 173

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE October 1999	3. REPORT TYPE AND DATES COVERED Annual (15 Sep 98 - 14 Sep 99)	
4. TITLE AND SUBTITLE Development of a Common Database for Digital Mammography Research			5. FUNDING NUMBERS DAMD17-94-J-4076	
6. AUTHOR(S) Robert M. Nishikawa, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Chicago Chicago, Illinois 60637 E-MAIL: r-nishikawa@uchicago.edu			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited				12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 Words) The purpose of this infrastructure project is to develop a large database of digitized mammograms that will be distributed free of charge to researchers working in all aspects of digital mammography. This database will facilitate and promote rapid development in digital mammography research. The database will consist of 1000 cases subdivided into 5 categories, 4 containing different breast lesions -- masses, microcalcifications, architectural distortions, asymmetric densities (both benign and malignant) -- and one containing normal mammograms. The mammograms will be collected and digitized (0.05-mm pixel size) at two sites: the Universities of Chicago and North Carolina. The database will be stored at the two sites and will be available over internet, and by mail on CD, tape and magneto-optical disks. To date 611 cases have been digitized. Each case consists of index and previous exams (each having four standard views) and up to two special-view mammograms (e.g., magnification views). The computer systems for the database have been assembled and are connected to the network. The first release of 50 cases with clustered microcalcifications will be made shortly after true location of the cluster is determined. This release will be followed by another 50 cases of microcalcifications and 100 cases with masses.				
14. SUBJECT TERMS Digital mammography, database, information systems, image analysis, computer-aided diagnosis, image processing.			15. NUMBER OF PAGES 9	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

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.

N/A 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).

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

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

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

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

Kurt A. [Signature] 10/13/99

4. TABLE OF CONTENTS

Front Cover.....	1
SF 298 Report Documentation	2
Foreword	3
Table of Contents.....	4
Introduction	5
Body	5
Key Research Accomplishments.....	7
Reportable Outcomes.....	7
Conclusions.....	7
References	8

5. INTRODUCTION

The purpose of this project is to develop a database of digital mammograms that can be used by researchers who (1) are trying to determine the image quality requirements of detectors for digital mammography; (2) are developing image processing techniques to optimize the displayed digital mammogram; (3) are developing computerized methods for analyzing mammograms; (4) are studying the effects of image compression methods on image quality; (5) are developing methods for remote transmission of mammograms; and (6) are studying the relationship between image quality and diagnostic accuracy. This database also could be used as a resource for teaching radiology residents and for testing the performance levels of mammographers.

The specific aims of this proposal are:

1. Collect and digitize 200 cases in each of 5 different categories, mammograms exhibiting: (i) clustered microcalcifications, (ii) masses, (iii) architectural distortions, (iv) asymmetric densities, and (v) no lesions (i.e. normals).
2. Make these cases available to other researchers either over computer network (Internet) or by sending images on computer tape or CD. The database will be distributed as widely as possible so that comparisons of different computerized analysis techniques can be standardized.

6. BODY

This research is being funded as an infrastructure award and as such it does not represent a research project per se. That is, there is no hypothesis that we are trying to prove. Therefore, this report is structured slightly different than a normal scientific research report -- heavy on the method and light on actual results. In this project, the procedure is the most important component, which is applied continuously in a straightforward manner to achieve the goal of creating the database of mammograms.

Task 1: Collect and digitize mammograms. (See Figure 1.)

Until recently (July), the mammography section had been severely understaffed and without a section chief. Two new radiologists have been hired, but they have been preoccupied with getting the clinical side of the section running smoothly again. Therefore, it has been difficult to add more cases to the database. Specifically, we need a radiologist to review the case to ensure that it meets the requirements for inclusion in the database, review the pathology report, and mark the location of the lesion(s). If one of the two new radiologists cannot devote some time to the project, then I will enlist Carl Vyborny, who is a radiologist with expertise in mammography, to help out on the project. It is necessary to add him to the IRB protocol, before he can begin. This is being done presently.

We now have 611 cases digitized (see Table I, at the end of the report). We performed a per- release of 50 cases of clustered microcalcifications this year to test the database. It was found that the truth data needed to be specified more exactly. Initially we had just indicated the center of the lesion, but not its extent. We are now specifying the four corners of a rectangle that completely encloses the cluster of calcifications. Currently, because we are still short staffed in the radiology department, it has been difficult to get a radiologist to mark the truth. We did hire a consultant radiologist for a 2-month period, but she is currently unavailable. Once the truth has been recorded, we will release the 50 cases. We will follow this initial release with an additional 50 cases of microcalcifications and 100 cases of masses. As more cases accrue, further releases will be made in batches of 50 or 100. We will accelerate our case accrual to meet our goal of 1000 cases.

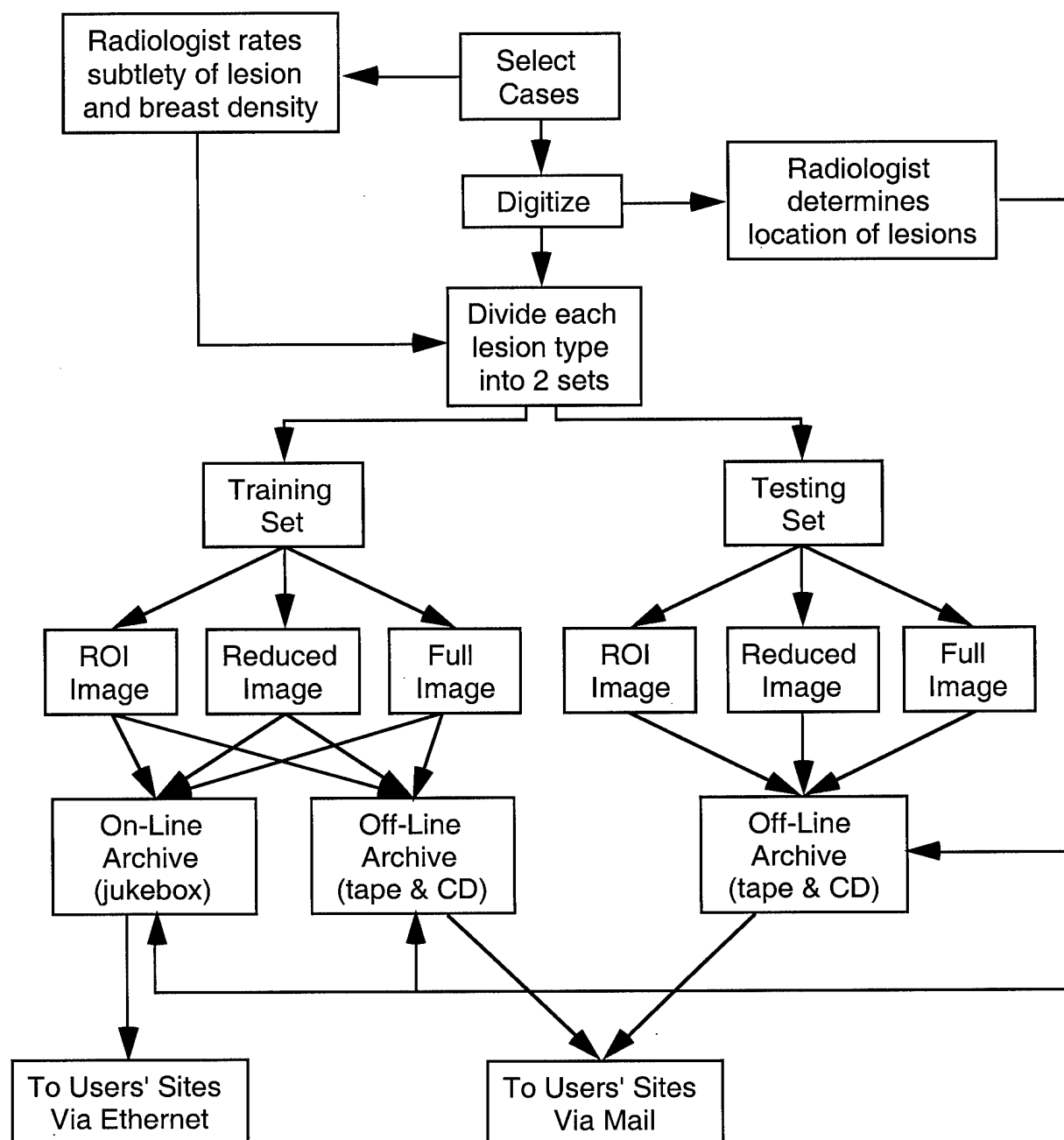


Figure 1. A flowchart of the steps required to collect, digitize, archive, and distribute the mammographic database. The 'Full Image' is the whole digitized mammogram at full resolution. The 'Reduced Image' is a minified version (reduced resolution) of the full image. The 'ROI Image' is a portion of the full image at full resolution.

It is important that this database be used in such a way so as comparisons between different algorithms can be made. This is the main motivation for creating such a database. To allow for valid comparisons to be made, two things are needed: (i) the exact same cases need to be used to measure performance, and (ii) the same criteria for scoring the results need to be used. To this end, we have divided the database into testing and training cases. Furthermore, we are preparing recommendations for scoring the results based on some preliminary results in our lab. [1] This information will be released with the database. No other database offers such instructions for its use and thus comparisons of different techniques is still difficult to do.

Task 2: Establish protocol for transmitting database

We originally considered the ACR/NEMA (DICOM) image format for our database. However, when we began our work, the ACR/NEMA format did not have a module for mammography, and it would have been an extensive project to develop one at that time. Recently a digital mammography module has been approved. We will consider whether it is appropriate to use this file format, since many users of our database may not have a DICOM reader.

Task 3: Maintain database and distribute cases

Maintenance of the database and distribution of the database are at a minimum currently. These tasks will become important shortly as cases go "on-line". Cases are being archived on 4 mm tape and DVD.

7. KEY RESEARCH ACCOMPLISHMENTS

- Collection of 611 mammographic cases

8. REPORTABLE OUTCOMES

Presentations and Manuscripts:

1. Nishikawa RM, Wolverton DE, Schmidt RE, Johnson RE, Pisano, ED, Hemminger BM: A common database of mammograms for research in digital mammography. U.S. Army Medical Research and Materiel Command Breast Cancer Research Program: An Era of Hope, November, 1997, Washington, DC.
2. Nishikawa RM, Wolverton DE, Schmidt RA, Pisano ED, Hemminger BM, Moody J: A common database of mammograms for research in digital mammography,. In: Doi K, Giger ML, Nishikawa RM, and Schmidt RA (eds.), Digital Mammography '96. (Amsterdam: Elsevier Science) 435-438, 1996.
3. Nishikawa RM: Mammographic databases. Breast Disease 10 137-150, 1998.

9. CONCLUSIONS

The development of a common database of mammograms for digital mammography research is underway. We performed a prerelease of 50 cases of clustered microcalcifications to test the database. It was found that the truth data needed to be specified more exactly. Accural of cases has been hampered because of short staffing in our mammography section. If necessary, we

will hire an outside radiologist, Carl Vyborny, to assist us. We will follow the initial release with an additional 50 cases of microcalcifications and 100 cases of masses. As more cases accrue, further releases will be made. Cases will be released with instructions that will allow all users of the database to compare their results in a meaningful manner. Such comparisons are currently not possible or are problematic with any existing database.

10. REFERENCES

1. Nishikawa RM, Yarusso LM: Variations in measured performance of CAD schemes due to database composition and scoring protocol. Proc. SPIE 3338: 840-844, 1998.

Table I. Breakdown of cases in the database as of October 1/99.

Type of Lesion	Pathology	# of Cases
Mass	Malignant	116
Mass	Benign	75
Microcalcifications	Malignant	115
Microcalcifications	Benign	87
Asymmetric Density	Malignant	18
Asymmetric Density	Benign	4
Architectural Distortion	Malignant	25
Architectural Distortion	Benign	3
Normal		168
Total		611