AD_____

Award Number: DAMD17-00-1-0410

TITLE: Remote Patient Management in a Mammographic Screening Environment in Underserved Areas

PRINCIPAL INVESTIGATOR: David Gur, Sc.D.

CONTRACTING ORGANIZATION: University of Pittsburgh
Pittsburgh, Pennsylvania 15260

REPORT DATE: September 2003

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. 074-0188



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | September 2003 | Annual (1 Sep 2002 - 31 Aug 2003) |

**4. TITLE AND SUBTITLE**
Remote Patient Management in a Mammorgraphic Screening Environment in Underserved Areas

**5. FUNDING NUMBERS**
DAMD17-00-1-0410

**6. AUTHOR(S)**
David Gur, Sc.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

E-Mail: gurd@msx.upmc.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)**
Early detection of breast cancer is of significant interest to our society. Mammographic screening is gradually moving toward a "distributed acquisition - centralized review" approach. Unfortunately, a relatively high recall rate using this approach increases patient anxiety as well as the cost and complexity of the diagnostic process. The purpose of this project is to evaluate in a multi-phase project the possible impact of a unique telemammography system that utilizes common carriers with wavelet-based data compression for image transmission on the recall rate in remote locations where physicians are not available during mammographic procedures. The initial phase of the project encompasses the design, assembly, and technical testing of a multi-site telemammography system that enables the digitization, transmission, and display of wavelet compressed images as well as associated text documents of a case in less than 15 minutes. The impact of such a system with and without the incorporation of CAD results will be evaluated during a step-by-step assessment in a multi-site study.

**14. SUBJECT TERMS**
Breast cancer, telemammography, detection, CAD

**15. NUMBER OF PAGES**
43

**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 |

## Table of Contents

## Introduction

Periodic mass screening of asymptomatic women is rapidly gaining approval and acceptance, and the population segment recommended for screening is increasing due to both longer life expectancy as well as earlier recommended age for initial examination [1-3]. The large variability in a number of important aspects related to mammography, as practiced in the U.S., resulted in the enactment of the Mammography Quality Standards Act, which mandates accreditation of each program (facility, technical, and professional) [4,5]. Shortages of expert mammographers in many locations, combined with the desire to make it convenient for the patient to undergo the procedure, suggest that there may be a need for high-quality telemammography systems that enable a distributed acquisition-centralized expert review type solution to the problem, particularly in underserved areas [6,7]. The relatively high recall rates (5-15%) of screened women to supplement information that was not ascertained during the initial visit (e.g. magnification views) also make it desirable to enable physician "monitoring" and "management" of remote locations so that patient-management decisions can be made while the patient remains in the clinic [8-11]. In addition, a technologist who observes a possible abnormality during the performance of the study could benefit from the ability to communicate her/ his suspicion, and an expert mammographer could review the specific case, together with the technologist's observation, resulting in an improved and perhaps a more timely diagnosis. Current practices result in increased patient anxiety and added practice complexity and cost. Early attempts to develop and implement a practical telemammography solution to this problem failed due to several significant technical problems associated with acquisition, transmission, management, and display of the images and other related information [12-14]. Many of these technical issues have been resolved in recent years, but some remain [14-18]. Although an adequate communication infrastructure for high-quality telemammography is available within some urban regions, the fact remains that where it may be needed most (i.e. remote, non-urban locations), enabling (two-way) communication systems remain limited to lower level communication capabilities (e.g., the Plain Old Telephone System (POTS)). Other communication technologies, such as satellites, are being evaluated for this purpose, but it is not likely that these will displace lower level communication technologies in many underserved areas for quite some time [19-23]. Hence, the problem of cost effective, timely remote patient monitoring and management in many underserved areas is not a simple one.

As a part of this project, we are assembling and evaluating a unique telemammography system that enables improved communication between remote sites where physicians are not always available during the mammographic acquisition process and a central location where experts can review the acquired images shortly after acquisition and assess whether or not additional procedures (e.g., spot compression views) are needed [24,25]. The system we are assembling is based on prior preliminary experience acquired in our group during ten years of research in this general area. It includes the use of a common carrier for communication (Plain Old Telephone System, POTS) and other "low level" communication capabilities, wavelet-based image compression for data reduction, and the optional incorporation of other text information and CAD results into the transmitted information. The main goal is to assess in a step-by-step approach whether the use of such a system could significantly reduce recall rates in the remote sites. Other secondary objectives regarding ways to improve communication and creating an environment for "more active" participation of the technologist in the diagnostic process are also being explored.

## Body:

Since the initiation of the project on September 1, 2000, we have been progressing methodologically step by step on the tasks listed in the Statement of Work (page 5 of the proposal), as originally submitted. It should be noted that the project is, for the most part, on track, schedule wise, despite the fact that the Imaging Research group was relocated during November and December 2000 from Scaife Hall of the University of Pittsburgh to Magee Womens Hospital of the University of Pittsburgh Medical Center Health System. While this move resulted in minor interruptions, schedule wise, in the long run, the project is benefiting significantly from such a move, since the group is now located where much of the project is being carried out and evaluated. As will be explained in the body of the report, our initial findings necessitated the addition of several technical tasks that are all being successfully performed in order to maximize our ability to learn about the applications being investigated in this project. During year three of the project, work was performed in three different areas listed under Task 1 (Redesign and Assemble System), Task 3 (Clinical System's Evaluation), and Task 4 (Incorporation of CAD Results) in the original proposal. We have also begun planning for Task 5. As we explain in the body of the report, several new additions (capabilities) were added to the system as a result of our operational and preliminary clinically simulated evaluation tasks. These were designed to significantly improve the communication capabilities in an efficient and more concise manner between the remote and central sites.

**Under Task 1, we performed the following:**

With the exception of some upgrades we are currently testing, we completed all originally proposed tasks under this category. We assembled and tested a multi-site telemammography system that meets (and in several important aspects exceeds) our proposed specifications. While we have implemented a way to upload acquired FFDM images onto the telemammography system (Task 1.c in the proposal), as we indicated last year, our health care delivery system (UPMC) has decided to continue to use the GE FFDM system only at the central site and mainly for diagnostic mammography purposes (not screening). The main reason is the cost of these systems and the fact that to date, there are no conclusive results demonstrating that the use of such systems lead to more accurate diagnosis in the screening environment. Hence, our clinical application assessment tasks continue to be carried out on digitized films. The status of the tasks described under this category is as follows:

a) **Select and Purchase Equipment**: **Completed.**

b) **Convert Software to Windows Based:** **Completed.**

c) **Develop Interface to FFDM Acquisition System:** Technical task completed. We enabled the system to accept cases generated by our FFDM system. We have been actually transferring FFDM images to a server. From this server the images are uploaded onto the telemammography workstation at the central site. Hence, the technical task has been completed and the interface has been verified. However, the FFDM field has been progressing rapidly from an acquisition technology point of view (in that several companies are now offering high-quality systems), and the specific systems that may be ultimately

implemented in the future in our remote sites have not been determined. As important, the cost associated with such implementation is quite high, and we are finding that in most remote underserved sites (ours as well as others), there is a reluctance to move rapidly into digital acquisition (FFDM). Hence, when it makes sense we will incorporate the needed interface to an FFDM system. At this time, we continue to focus our efforts on film digitization. It is not clear that the use of FFDM devices in remote "underserved" sites for screening purposes is likely to be common or appropriate in the near future.

**d) Develop a New User Interface for the Acquisition Sites: Completed and tested.**

A remote site user interface was completed and tested, both subjectively and objectively. After minor modifications that were based on users' comments, our data entry and case-sending routines were refined and finalized.

**e) Complete Data Compression Software Module: Completed and tested.**

A compression software scheme was finalized and tested. The scheme allows for a site-specific selectable level of compression to be used.

**f) Develop and Refine Measures of Image Fidelity that can be used to Automatically Monitor and Adjust (if needed) Compression Levels on an Image-by-Image Basis:** Completed and tested. Based on two independent tests (see evaluation section below), at two compression levels, 50:1 and 75:1, we enabled a "dial-up" compression capability in the system. However, we are finding out that the high level of acceptance of either compression level practically eliminates the need for this option. Therefore, we are currently using the system with a fixed level of compression (75:1). We believe that we have achieved high-quality images at such high compression levels that second-order image-specific adjustments are not needed for all practical purposes.

**g) Integrate all Software Modules**: Completed and tested. All software modules were successfully integrated.

**h) Develop Display Protocols for the Workstation**: Completed and tested. User-friendly display protocols have been developed and tested extensively (see system evaluation section).

**i) Assemble System**: Completed and tested. The system was assembled as proposed.

**j) Test System in Laboratory:** Completed and tested. The system has been tested in the laboratory. To enable us to continue development efforts without undue interruptions in the clinically simulated assessment tasks, we have assembled a second laboratory system at no additional cost to the project. This system is used for development and pilot testing modifications of and improvements in the "clinical" system.

**k) Trouble Shoot, Refine, and Finalize System**: Completed and tested. Through refinements, we increased the operational ease-of-use and reliability of the system and finalized the base configuration for implementation.

**l) Prepare Clinical Sites for Implementation**: Completed and tested. All three remote sites were prepared for system implementation as required.

**New improvements that were added to the system:**

Our technical efforts during the last year were guided by the radiologists and technologists' actual reaction to the workflow implemented by the system. For example, the radiologists feel more comfortable if the technologists mark the location of suspicious findings prior to sending the cases. Hence, we are working on enabling them to do so. However, because we do not want to mark the images permanently on the digitized images, we are implementing a module that will allow the marking of suspicious regions as well as the removal of the marking on the display. This task is underway and expected to be completed in the very near future.

In addition, we noticed that radiologists' "comfort level" in making decisions is quite dependent on the amount of information they have access to during the review. Hence, we enabled a module the allows them to view prior reports side by side with the images. To do so we needed to add a third monitor to the workstation (and write the needed software to control the three monitors as a single unit), and we did so without any additional cost to the project. This task was completed, tested, and is currently in routine operations.

Last, although not a major point, radiologists felt that on some occasion they wish to document (record) the images they view while making decisions. Hence, we enabled a function to print the images on a high-resolution laser printer.

**Under Task 2, we performed the following:**

**a) Move Hardware. Completed.** All needed equipment was moved to the appropriate locations at the three remote sites. At each location, the equipment (send station and digitizer) is located at an easily accessible place. At the central site, we placed the "receive" workstation in a "screening" reading room at a central location within the Breast Center. This required some construction that was completed at no cost to the project.

**b) Re-assemble System. Completed.** The complete system was reassembled on location.

**c) Re-test Technical Performance Levels. Completed.** Technical and operational performance levels were retested on site.

**d) Develop and Test Initial Protocols. Completed.** Different evaluation protocols for initial system evaluations were developed and implemented.

1) 100 cases were randomly selected at each remote site and transmitted to a central site to assess ease-of-use, reliability, reproducibility, and cycle times. The results clearly indicate that cases from all sites at 15, 20, and 90 miles away can be transmitted with a full duty cycle time (from data entry at remote site to display) that easily meets our proposed specifications. A four-image case can be completed in less than seven minutes using 75:1 compression, which is less than half the time we originally specified.

2) We performed a multi-reader subjective assessment of image quality, and all participating radiologists rated the quality as acceptable or better for the task at hand.

3) We evaluated differences in image quality on film and soft display at zero (no), 50:1, and 75:1 compression ratios and found that only under extreme magnification, the 75:1 level can be identified (recognized), but image quality is not significantly degraded for all practical purposes. Note that an additional related study was performed this year and is described under Task 3.

4) In order to comply with HIPAA regulations, we moved one of the three sites to a facility approximately 15 miles away from our central site, in which physicians are not generally present during four of its six operating days. The move enabled us to continue simultaneous operations at three sites that are clinically (formally) interpreted by the same group of radiologists. This change was performed with minimal interference, indicating the ease of performing this task at remote locations.

**Under Task 3, we performed the following:**
**Note that a significant fraction of the effort during the last year was carried out under Task 3. We wish to emphasize that this is a step-by-step iterative process that leads to incremental changes and adjustments as we proceed. The main effort lies in improving our understanding of the utilization process (work flow) and how it could be potentially improved with the aid of the telemammography system. Hence, there are many issues that need to be addressed; the most important of which is perhaps the human machine interaction aspect of the project. We are "breaking ground" in several respects that include but are not limited to the involvement of technologists in the decision-making process (e.g., which cases to send over to the central site) and the increasing "reliance" of the radiologists on the technologists' judgments. These are becoming some of the more important and exciting aspects of the project, but they are not easy to study or resolve.**

**a. Collect information on clinical performance levels without the system: Partially completed.** We continue to analyze the data available in our databases concerning patient distributions and process-related information. This includes the recall rate by physician, site, type, and reason for recall. We have also obtained patient satisfaction survey results as ascertained from internal and external surveys, which had been performed by our institution for other purposes outside this project. Last, we reviewed records concerning the cycle time from the initial examination to a definitive diagnosis for cases that were not being recalled, as well as cases that were. This analysis is performed for the different sites in which we operate. This effort continues throughout the project as data are collected and analyzed regarding the above-mentioned variables (mainly for clinical monitoring purposes). The effort described here constitutes the initial baseline (reference) information for comparison purposes. One of the more interesting (and relevant) finding in this regard is the long delays in scheduling (average > 20 days) between the patient's call for an appointment due to recall and the actual date of examination, underlying the potential benefit of the use of telemammography to reduce recall rates.

b. During the last year we completed a large study to assess the recall rates and detection rates of our ten highest volume radiologists. One of the issues that was raised in our group was the issue of correlations (if any) between the recall and detection rates of radiologists. This is an important point since there is a significant pressure on radiologists to reduce their individual recall rates to below ten percent. While we recognize the tremendous value of

reducing recall rates without a substantial degradation in detection rates (sensitivity), the question arises as to whether or not higher recall rates are also generally associated with higher detection rates. This issue has not been well studied. We reviewed 98,668 mammograms interpreted by 10 radiologists over a period of three years. Screening mammography examinations performed in our facilities at Magee-Womens Hospital of Pittsburgh and our satellite breast-imaging clinics during 2000, 2001, and 2002 were reviewed under an IRB-approved protocol. Mammograms that had been interpreted by our ten highest volume mammographers during this period were included in the study.

These ten radiologists interpreted a total of 98,668 cases during this time and detected 368 cancers as a result of recommendations for recall in this group. A wide range of recall rates (from 7.7% to 17.2%) and detection rates (from 2.6 to 5.4 per 1000 mammograms) was observed. Despite the low number of readers (10), when we compared recall and detection rates using the parametric Pearson (r), the correlation between recall and detection rates was significant ($r=0.76$, $p=0.01$). Similarly, a significant correlation in our group of readers was observed using the nonparametric Spearman ($rho = 0.72$, $p=0.02$). Despite significant inter-reader variability, the slope indicates an average of 0.22 additional cancer detections for one percent increase in recall rates (the 95 percent confidence limits on the slope are 0.068 to 0.378). These results are currently under review for publication. The important point we learned is that reducing recall rates through improved communication and the use of technology for this purpose may be more important than doing so by sheer pressure (or "decree"). In a similar study on the effect of CAD on diagnosis, we found out that the use of CAD may be extremely important for the purpose of determining the need for recall (as we envisioned in this project), but ultimately the improvement in actual final diagnosis is somewhat limited in our environment. This work is also under review for publication.

**c) Perform a simulated prospective study: Partially completed.**
Last year we reported on our initial study that indicated the possibility of significantly reducing actual recall rates, but at the cost of a large number of additional procedures that would need to be performed during the initial visit. As a result of our initial experience, we added the following capabilities to the system and evaluated their impact on performance.

1) **Real-time "chat"** – To facilitate effective communication between the technologists in the remote sites and experienced radiologists, we have implemented a "chat" box type function. The chat box provides a real-time interactive capability. Chat boxes on both sides contain: patient demographics; message area; pull-down menus; and a free typing text area. Typical communication includes the technologist sending a chat dialog with each case indicating: breast, left or right; view, cradiocaudal and/or mediolateral oblique; finding, mass or calcifications; comparison with prior exam, baseline, new, or increased; and possible additional procedure, additional views and/or ultrasound. The radiologists can reply after reviewing the case to do recommended procedure as suggested; no additional procedures are necessary; and do not do suggested procedure, but do X, Y, and Z.

2) **Case folder enables more than four images** – We enabled the "case folder" to include scanned reports (text) as well as more than four images (e.g., prior examination).

3) **The transmission of prior reports** -- For a period of several months we requested that technologists at the remote sites send us several cases from each site they encounter during screening days that they believe would be eventually recalled by the radiologist for

additional procedures. The cases were sent with a "chat" message regarding the reason for their suspicion. These were reviewed at the central site, and a simulated response from a radiologist was sent back (off line). The study was successful technically, but radiologists indicated that they would like to have more information on the prior examination when available. As a result, we upgraded the system (see Task 1, Section l), and the study is being repeated with the transmission of the prior report associated with each case as well. We anticipate that this study will be completed in late November or early December 2003. The initial reaction is that this is a notable improvement over the prior functionality.

**d) Assess the differences between conventional and telemammography supported operations: Partially completed.** As already indicated, subjective feelings and personal confidence levels are important for acceptance of new concepts and practices in this field. We have been frequently discussing these issues with the radiologists. There is no doubt that they believe that increased communication between radiologist and technologists is an advantage. At the same time, radiologists feel that they would rather operate on the "conservative" side if they do not feel comfortable with the technologists' recommendations or when they feel they would like additional information to make a clinical decision. As we indicated in our last year's report, this resulted in a significant "over-reading" during our first study. The reasons for the over-readings were several, but the indications were that first, the radiologists wanted messages from the technologists as to the reasons the case was sent for review. This was addressed by adding the "chat" capability to the system. We were concerned that this will not reduce recommended "additional procedures," because the radiologists will now identify their own reasons, plus take into account some that they did not identify but the technologist did. A second study was conducted, and while there was some difference in the number of recommendations for additional views, the main problem of a high fraction of additional procedures during the initial visit was not resolved (see Task 2, Section b.1). The second reason stated by the radiologists for the high "recommendation level for additional procedures" was the availability of prior reports. As a result, we enabled this function, and a second study is underway to assess its effect.

In one study to address this question, 169 cases, 69 of which (40.8%) were actually recalled clinically, were included. This was a more difficult set than our initial study in that a large number (57 cases) had subtle benign findings. Four radiologists recommended additional procedures in the majority of these difficult cases (average 82%), as expected. When we compared the recommendation with and without "chat messaging," the results were comparable on this set of cases. As in the clinical environment, we observed a large inter-reader variability, and those who tend to have a higher recall in the clinic exhibited the same pattern in the study. A subset of these cases (99) is now being read with the availability of prior reports. It should be noted that these results are encouraging in that our previous experience with the technologists making a decision for additional views resulted in approximately 60 percent of ALL women receiving additional procedures in a typical screening population (unlike this difficult set). Projections based on this experiment to the clinical environment would result in approximately 25 - 30 percent of women under this category with the use of remote consultation. As important, it is estimated that the use of this approach could reduce the number of women actually being recalled for a second visit from 11.4 percent to approximately 7 percent, which would be a substantial reduction. Since, a high number of women will receive additional procedures during their initial screening visit to achieve this type of reduction, we are focusing our efforts on reducing this number with

the aid of prior reports (the current study is underway) and possibly the use of CAD (the last technical study we anticipate before the high-volume demonstration).

**d, e) Technical and Clinical System Evaluations – Objective measures: Partially Completed.** We continue to record our performance levels throughout the project. One of the areas of initial concern was the use of highly compressed images at the central site. To assess this issue, we conducted the following experiment during the third year of the project. The purpose of the study was to evaluate the ability of radiologists to identify high-levels of image compression applied to digitized mammographic images and displayed on high-resolution, grayscale monitors. Mammography films were digitized at 50-micron pixel dimensions using a high-resolution laser film digitizer. The image data were compressed using the irreversible (lossy), wavelet-based JPEG 2000 method. Twenty images were randomly presented in pairs (one image per monitor) in three modes: mode 1, no compression versus 50:1; mode 2, no compression versus 75:1; and mode 3, 50:1 compression versus 75:1 with 20 random pairs presented twice to evaluate intra-observer variability (80 pairs total). Six radiologists were "forced" (2-AFC experiment) to choose which image had the lower level of data compression. The average percent correct across the six radiologists for modes 1, 2 and 3 were 56% (+/- 8), 55% (+/- 14), and 59% (+/- 8), respectively. The percent of correct choices identified on the left monitor was statistically greater compared to the right monitor for mode 2 ($p = 0.048$). Intra-observer percent agreement ranged from 10 to 50% and Kappa from -0.78 to -0.19. Kappa for inter-observer agreement ranged from -0.47 to 0.37. In this controlled evaluation, radiologists did not accurately or reliably distinguish between non-compressed and compressed images. Intra-observer agreement was poor. We conclude that either 50:1 or 75:1 image compression levels should be acceptable for displaying digitized mammograms in a telemammography system. Interestingly, although both carefully calibrated, the "monitor effect" (left versus right) was of the same order of magnitude as the effect of image compression.

**f) Analyze the Performance using FFDM (see comment in Task 1):** To date we used digitized films only for evaluation of high levels of data compression displayed on high-resolution monitors, since all of our radiologists prefer to use the workstation for clinical review purposes. The use of films for selected difficult cases (particularly those with possible subtle microcalcification clusters) has been completed. The FFDM utilization for this purpose was previously addressed. Although we are ready to implement this capability, we do not anticipate that FFDM will play a significant role in this project. All of our observations to date are relevant to an FFDM based environment, but we do not believe it will be applicable to underserved areas in most situations in the near future.

**Under Task 4, we performed the following:**

**a. CAD Software Module: Completed.**

**b. <u>CAD Incorporation</u>: Completed.** During the third year of the project we completed the design and implementation and testing of a modular software set of routines that enable the incorporation of CAD into the telemammography system at the remote (sending) sites and transmit the results to the central site.

**c. CAD Technical Performance Evaluation: Completed.** The system was tested technically using over 100 cases, and after de-bugging, we incorporated the module into the

operations. Currently all transmitted cases are processed by the CAD scheme and can be displayed on the workstation at the operator's discretion (with or without the CAD results).

**d. CAD Operational and Clinical Use:** The operational use of CAD results was tested using a retrospective clinical review and found acceptable. The clinical aspects of this added feature are currently being evaluated. The impact of the added feature on radiologists' ability to make better calls in regard to the need for additional procedures in specific cases will be evaluated during the next year of the project.

**Under Task 5, we performed the following:**

There is only one significant effort under this category; namely, it is the "high volume" demonstration of the transmission of a volume of suspected cases at the remote sites and a simulated response from the central site in "almost real time." This task is proposed for initiation later in year four. We wish to correct one issue regarding this task in that although possible technically, we did not intend to send all screening cases from the remote sites. Rather, in this "high-volume" demonstration, we intend to transmit a high volume of cases that the technologists consider candidates for recall (which in our experience is technologist dependent and amounts to approximately 20 - 50 percent of all cases.). Because of the operational issues associated with this task, we have already begun to address some scheduling concerns. We are routinely testing the system's ability to handle a reasonably high volume of cases from all sites.

## Key (Research) Accomplishments:

During the first three years of the project, we have been progressing according to the original plan and addressed a large number of the technical tasks and operational issues associated with the design, implementation, technical, and simulated clinical testing of the multi-site telemammography system. The key accomplishments for the first three years were:

- We carried out a comprehensive review of the performance of our radiologists in terms of recall and detection rates.
- We upgraded the initial system twice in response to radiologist preferences during the performance of the task the telemammography system was designed for.
- We successfully and reliably transmitted over 1500 cases from three remote sites to the central site.
- We successfully reviewed a large number of cases on the workstation and generated a "chat" response in a clinically simulated environment.
- We completed two observer performance studies to assess agreement levels between the technologists and radiologists on suspicious cases.
- We are increasing the communication level between technologists and physicians in regard to decision-making processes, and we are engaged in discussions concerning a more extensive use of technologists as physician extenders in several areas
- We have been able to coherently engage a large team of administrative, technical, clinical (i.e., technologist), and physician personnel in a large and complicated project.

- We demonstrated that in principle one can achieve a significant reduction in actual recall rates for a second visit, albeit at this time, at the cost of a substantial increase in the number of women who would receive additional procedures during their initial screening visit. Our current focus is on reducing this number.

**Reportable Outcomes:**

The nature of this project is such that much of the work performed to date does not result in a large number of significant reportable outcomes. However, as we developed and tested the system, several reportable tasks have been performed for which partial support (albeit quite limited) is provided by this project. For example, we developed a software package that incorporated CAD results into the telemammography system during the third year of the project. The development of our CAD schemes continue, and the performance seems to be improving as we progress in optimizing step-by-step the different schemes we have developed. In addition, our comprehensive assessment of the actual performance in our clinical operations as it relates to recall and detection rates was partially supported (again to a limited extent) by this project. These efforts have led to important developments and observations that may have a significant impact on this field. Therefore, several of our scientific reports acknowledge this project.


- Zheng B, Ganott MA, Britton CA, Hakim CM, Hardesty LA, Chang TS, Rockette HE, Gur D. Soft-copy mammographic readings with different computer-assisted diagnosis cuing environments: Preliminary findings. Radiology 2001; 221:663-640
- Zheng B, Chang Y-H, Good WF, Gur D. Performance gain in computer-assisted detection schemes by averaging scores generated from artificial neural networks with adaptive filtering. Med Phys 2001; 28: 2302-2308
- Drescher JM, Maitz GS, Leader JK, Sumkin JH, Poller WR, Klaman H, Zheng B, Gur D. Design considerations for a multi-site, POTS-based telemammography system. Proc SPIE 2002; 4685:416-421
- Zheng B, Shah R, Wallace L, Hakim C, Ganott MA, Gur D. Computer-aided detection in mammography: An assessment of performance on current and prior images. Acad Radiol 2002; 9:1245-1250
- Leader JK, Sumkin JH, Drescher JM, Maitz GS, Zheng B, Wallace L, Hakim C, Hertzberg TM, Hardesty L, Shah R, Clearfield R, Sneddon C, Lindeman S, Craig D, Pugliese F, Duffner D, Lockhart J, Traylor C, Gur D. A multi-site telemammography system: technical challenges, operational issues, and preliminary clinical evaluation. Presented at the Department of Defense "Era of Hope" meeting, September 25, 2002.
- Drescher JM, Maitz GS, Traylor C, Leader JK, Clearfield RJ, Shah R, Ganott MA, Pugliese F, Duffner D, Lockhart J, Gur D. A multi-site telemammography system: preliminary assessment of technical and operational issues. Proc SPIE 2003;5033:360-369
- Leader JK, Wallace LP, Hakim CM, Hertzberg TM, Hardesty LA, Sumkin JH, Cohen C, Sneddon C, Lindeman S, Craig D, and Drescher JM. Preliminary clinical evaluation of a multi-site telemammography system in a screening mammography environment. Proc SPIE 2003: 5033:273-280

- Zheng B, Wang XH, Wallace L, Cohen C, Hardesty LA, Hakim CM, Abrams G, Sumkin J, Gur D. Improving CAD performance in detecting masses depicted on prior images. Proc SPIE 2003; 5032:215-221

We anticipate that additional results of the system upgrades and the simulated clinical testing will continue to be reported at upcoming national meetings (e.g., SPIE) and others will be published in refereed journals.

**Conclusions:**

There are several technical, clinical, and assessment tasks listed in the Statement of Work of this project. During the first three years, we undertook a large number of technical and application-based tasks associated with the design, implementation, and preliminary evaluation of a multi-site telemammography system. We overcame many of the technical problems and assembled a multi-site system that exceeds several of the performance goals we originally proposed. The system has been undergoing a comprehensive step-by-step evaluation (and refinement as deemed appropriate), and the goal is to establish and test an environment with improved communications' capabilities between remote (and often underserved) facilities and a central site. Our main observation to date is that the general concept was verified and the actual implementation resulted in an appreciation for the importance of the "comfort level" of the team (physicians and technologists) in operating and using such a system for the stated purpose. As a result of our experience, we have been improving the system performance to meet the operational and clinical needs as suggested by many members of the professional team involved in this project. Most important perhaps is the demonstration that in principle, one can achieve a significant reduction in actual recall rates for a second visit. At this time, it can be done at the cost of a substantial increase in the number of women who would receive additional procedures (e.g., views) during their initial screening visit, and we currently focus on investigating different ways to reduce this number.

**So What?**

The main goal of this project is to evaluate how the use of an "almost real-time" telemammography system (with or without the use of CAD results) may impact the diagnostic process in terms of complete cycle time and patients' recall rate. At this stage, when we focus on system implementation, improvements, and clinically simulated evaluations, it is premature to consider any impact statements that are relevant to the actual clinical environment. This task (Task 5) is planned as the last major effort for this project. The nature of this project necessitates that the evaluation requires a careful multi-step approach; hence, actual clinically simulated results can only be realized at a later date. Success of this project will enable a comprehensive demonstration of different ways to increase communication between remote (and potentially underserved) sites and a central site. Our hope is that by using this approach, one may be able to provide better, more timely and cost-effective service at these sites, and in the process substantially reduce actual recall rates in these remote facilities. Despite significant advances in our understanding of the many issues and alternatives surrounding the "optimal" screening environment, many of our current clinical practice guidelines are based on limited subjective assessments and anecdotal experiences, and a significant fraction is related to operational matters in busy urban

environments that are staffed by experienced radiologists. The area of optimizing remote, underserved practices has been studied only in a cursory manner. Our project is but one attempt to improve our understanding of the technical, operational, and clinical issues facing these facilities and implementing technology-based solutions that may help them provide a better service to the populations they serve.

**Background References:**


1. S Pelikan, M Moskowitz, "Effects of lead time length bias, and false-negative assurance on screening for breast cancer," Cancer 71, 1998-2005 (1993).
2. L Tabar, G Fagerberg, HH Chen, SW Duffy, CR Smart, A Gad, RA Smith, "Efficacy of breast cancer screening by age: New results from the Swedish Two-Country Trial," Cancer **75**, 2507-2517 (1995).
3. F Houn, ML Brown, "Current practice of screening mammography in the United States: Data from the national survey of mammography facilities," Radiology **190,** 209-215 (1994).
4. CA Beam, PM Layde, DC Sullivan, "Variability in the interpretation of screening mammograms by US radiologists," Arch Intern Med **156**, 209-213 (1996).
5. Food and Drug Administration, "Mammography facilities: requirements for accrediting bodies and quality standards and certification requirements-interim rules," Federal Register **58**, 67558-72. (CFR21, Part 900) (1993).
6. JG Elmore, CK Wells, CH Lee, DH Howard, AR Feinstein, "Variability in radiologists' interpretations of mammograms," N Engl J Med **331**, 1493-1499 (1994).
7. RML Warren, SW Duffy, "Comparison of single reading with double reading of mammograms and change in effectiveness with experience," Br J Radiol **68**, 958-962 (1995).
8. CJ Wright, CB Mueller, "Screening mammography and public health policy: The need for perspective," Lancet **346**, 29-32 (1995).
9. LW Bassett, RE Hendrick, TL Bassford, PF Butler, D Carter, M DeBor, CJ D'Orsi, CJ Garlinghouse, RF Jones, AS Langer, JL Lichtenfeld, JR Osuch, LN Reynolds, ES de Paredes, RE Williams, "Responsibilities of the mammography facility," In: Quality determinants of mammography, clinical practice guideline. Number 13. Washington, DC: US Department of Health and Human Services, AHCPR publication no. 95-0632 (1994).
10. JG Elmore, MB Barton, VM Moceri, S Polk, PJ Arena, SW Fletcher, "Ten-year risk of false-positive screening mammograms and clinical breast examinations," N Engl J Med **338**, 1089-1096 (1998).
11. DS May, NC Lee, MR Nadel, RM Henson, DS Miller, "The National Breast and Cervical Cancer Early Detection Program: Report of the First 4 Years of Mammography Provided to Medically Underserved Women," AJR **170**, 97-104 (1998).
12. SA Feig, MJ Yaffe, "Digital mammography, computer-aided diagnosis, and telemammography," Radiol Clin North Am **33**, 1205-1228 (1995).
13. LL Fajardo, MT Yoshino, GW Seeley, R Hunt, TB Hunter, R Friedman, D Cardenas, R Boyle, "Detection of breast abnormalities on teleradiology transmitted mammograms," Invest Radiol **25**, 1111-1115 (1990).
14. MA Goldberg, "Telemammography: Implementation issues," Telemedicine Journal **1,** 215-226 (1995).

15. HK Huang, SL Lou, E Sickles, D Hoogstrate, M Jahangiri, F Cao, J Wang, "Technical issues in full-field direct digital telemammography," [Chapter] In: Computer Assisted Radiology and Surgery. Lemke HU, Inamura K, Editors. Elsevier Science B.V., 662-667 (1997).
16. HK Huang, "Digital Mammography: A Missing Link in a Totally Digital Radiology Department," Presented at the EuroPACS 97 Meeting; PISA, Italy. September 25-27, (1997).
17. JM Murphy, NJ O'Hare, D Wheat, PA McCarthy, A Dowling, R Hayes, H Bowmer, GF Wilson, MP Molloy, "Digitized mammograms: a preliminary clinical evaluation and the potential for telemammography," Journal of Telemedicine and Telecare **5,** 193-197 (1999).
18. SL Lou, HD Lin, KP Lin, D Hoogstrate, "Automatic breast region extraction from digital mammograms for PACS and telemammography applications," Computerized Medical Imaging and Graphics **24,** 205-220 (2000).
19. S Dwyer, Private communications. See also "Telemedicine Targets Mammographic Services" in Biophotonics International Nov/Dec 1997. Page 10.
20. SL Lou, EA Sickles, HK Huang, D Hoogstrate, F Cao, J Wang, M Jahangiri, "Full-field direct digital telemammography: Technical components, study protocols, and preliminary results," IEEE Trans Info Technology in Biomedicine **1,** 270-278 (1997).
21. SL Lou, HK Huang, E Sickles, D Hoogstrate, F Cao, J Wang, "Full-field direct digital telemammography: system implementation," Proc SPIE **3339**, 156-164 (1998).
22. Wu M, Zheng Y, North M, Pisano E. NLM tele-educational application for radiologists to interpret mammography. Proc AMIA Symposium, 2002, pg 909-913
23. Sheybani EO, Sankar R. ATMTN: a telemammography network architecture. IEEE Trans Biomed Eng 2002; 49:1438-1443
24. GS Maitz, TS Chang, JH Sumkin, PW Wintz, CM Johns, M Ganott, BL Holbert, CM Hakim, KM Harris, D Gur, JM Herron, "Preliminary clinical evaluation of a high-resolution telemammography System," Invest Radiol **32**, 236-240 (1997).
25. JM Holbert, M Staiger, TS Chang, JD Towers, CA Britton, "Selection of processing algorithms for digital image compression: A rank-order study," Acad Radiol **2,** 273-276 (1995).

## Appendix

See Attached.

PROGRESS IN BIOMEDICAL OPTICS AND IMAGING

SPIE—The International Society for Optical Engineering

# Multi-site telemammography system: preliminary assessment of technical and operational issues

J. M. Drescher, G. S. Maitz, C. Traylor, J. K. Leader, R. J. Clearfield, R. Shah, M. A. Ganott, F. Pugliese, D. Duffner, J. Lockhart, D. Gur

**Reprinted from**

*Medical Imaging 2003*

***PACS and Integrated Medical Information Systems: Design and Evaluation***

SPIE P PROCEEDINGS SERIES

**Volume 5033**

**18–20 February 2003**
**San Diego, California, USA**

# A multi-site telemammography system: preliminary assessment of technical and operational issues

John M. Drescher*[a], Glenn S. Maitz[a], Christopher Traylor[a], J. Ken Leader[a], Ronald J. Clearfield[ab], Ratan Shah[ab], Marie A. Ganott[ab], Francine Pugliese[b], Dian Duffner[b], Janet Lockhart[b], David Gur[a]
[a]University of Pittsburgh, Pittsburgh, PA USA 15213
[b]Magee-Womens Hospital, Pittsburgh, PA USA 15213

## ABSTRACT

Our goal was to develop an inexpensive, high-quality, multi-site telemammography system, implemented with low-level data connections that provided a communication link for an "almost real-time" response from a radiologist (central site) to remote "underserved" sites. The remote sites digitize mammographic films using high-resolution, laser digitizers. Images are automatically cropped, compressed (wavelet-based), and encrypted prior to transmission. At the central site images are decrypted, decompressed, unsharp masked, and displayed using automatically determined LUTs. The sites communicate instantly via a "chat box." Remote sites 1, 2, and 3 are 15, 20, and 90 miles from the central site, respectively, and connected by POTS (sites 1 and 2) and LAN (site 3). Only minimal noticeable difference at compression levels of 50:1 and 75:1 could be identified unless magnified to extreme levels. Two experienced observers rated the LUTs for 200 images as "acceptable" to "excellent." Average cycle times to digitize, transmit and receive cases (four films each) at 75:1 compression were 5.97, 6.85, and 5.77 min/case from sites 1, 2, and 3, respectively. Unique data-handling schemes significantly decrease the image file size and allow successful transmission in a reliable, timely manner. Over 1000 cases have been transmitted to date. Messaging was found to be easy to use.

**Keywords:** Teleradiology, breast cancer screening, image decision making, mammography.

## 1. INTRODUCTION

The benefits of breast cancer screening mammography of asymptomatic women have been extensively studied and reported in the recent literature.[1-6] Mammographic screening will continue to be widely used worldwide, despite periodic reports of limited or no benefits from such practices.[7-9] Management of mammographic screening in terms of public perception and compliance,[10-12] radiologist's practice and performance,[13-15] and personnel shortages[11,16] could be improved in both rural and urban clinics. The use of teleradiology is one approach that could assist in this regard.

The high-spatial resolution required by mammography necessitates the use of commercial digitizers and high-resolution monitors to sufficiently preserve image quality.[17] Transmission time of large amounts of mammographic image data (35-55 MBytes per image) is frequently dependent on the communication link. Low-level data connections (i.e., Plain Old Telephone System (POTS)) may require data processing to decrease the image file size to enable transmission of large amounts of data in a timely manner.

This manuscript presents preliminary assessment of technical and operational issues regarding a multi-site telemammography system using low-level data connections. This study is a continuation of an ongoing effort over the past several years.[18,19] The system was designed on the concept of distributed acquisition/centralized review and to facilitate communication between a radiologist at a central site and a technologist at a remote "underserved" site. For the purpose of this project, "underserved" means a location where a physician is not physically present when the screening examinations are conducted. The technical features described were designed and implemented using a low-cost approach to transmit data across low-level data connections in a timely manner and maintain a high-level of image quality. Issues evaluated included: look-up table settings (window and level), image cropping, image compression,

* drescherjm@msx.upmc.edu; phone (412) 641-2563; fax (412) 641-2582, University of Pittsburgh, Magee-Womens Hospital, 300 Halket Street., Suite 4200, Pittsburgh, PA 15213



Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation, H. K. Huang, Osman M. Ratib, Editors, Proceedings of SPIE Vol. 5033 (2003)

transmission time, and workstation display features. We expect to demonstrate that the combination of efficient data handling, intelligent image processing, and easy to use messaging can be implemented to produce an inexpensive, high quality telemammography system capable of an "almost real-time" response from the central site radiologist to remote site technologist.

## 2. METHODS

### 2.1 Central and remote sites

The central site is staffed by experienced radiologists and located at Magee-Womens Hospital, Pittsburgh, PA, USA. The telemammography workstation at the central site is powered by a dual 1.2 GHz multi-processor (Athlon MP, Advanced Micro Device, Sunnyvale CA, USA) with 2 GB of RAM operating under Microsoft Windows 2000 Server (Microsoft Corporation, Redmond, WA, USA). The workstation display consists of three high-resolution (2048 x 2560) 8-bit grayscale portrait monitors at a nominal setting of 80 ftL (DS5100P, Clinton Electronics, Rockford, IL, USA). For data communication, the workstation uses 56K hardware modems (U.S. Robotics, Rolling Meadows, IL, USA) and ethernet network cards (OfficeConnect 10/100 NIC, 3COM, Santa Clara, CA, USA). A Kodak Dryview film printer (Eastman Kodak, Rochester, NY, USA) is connected to the workstation for film printing as necessary (Fig. 1).

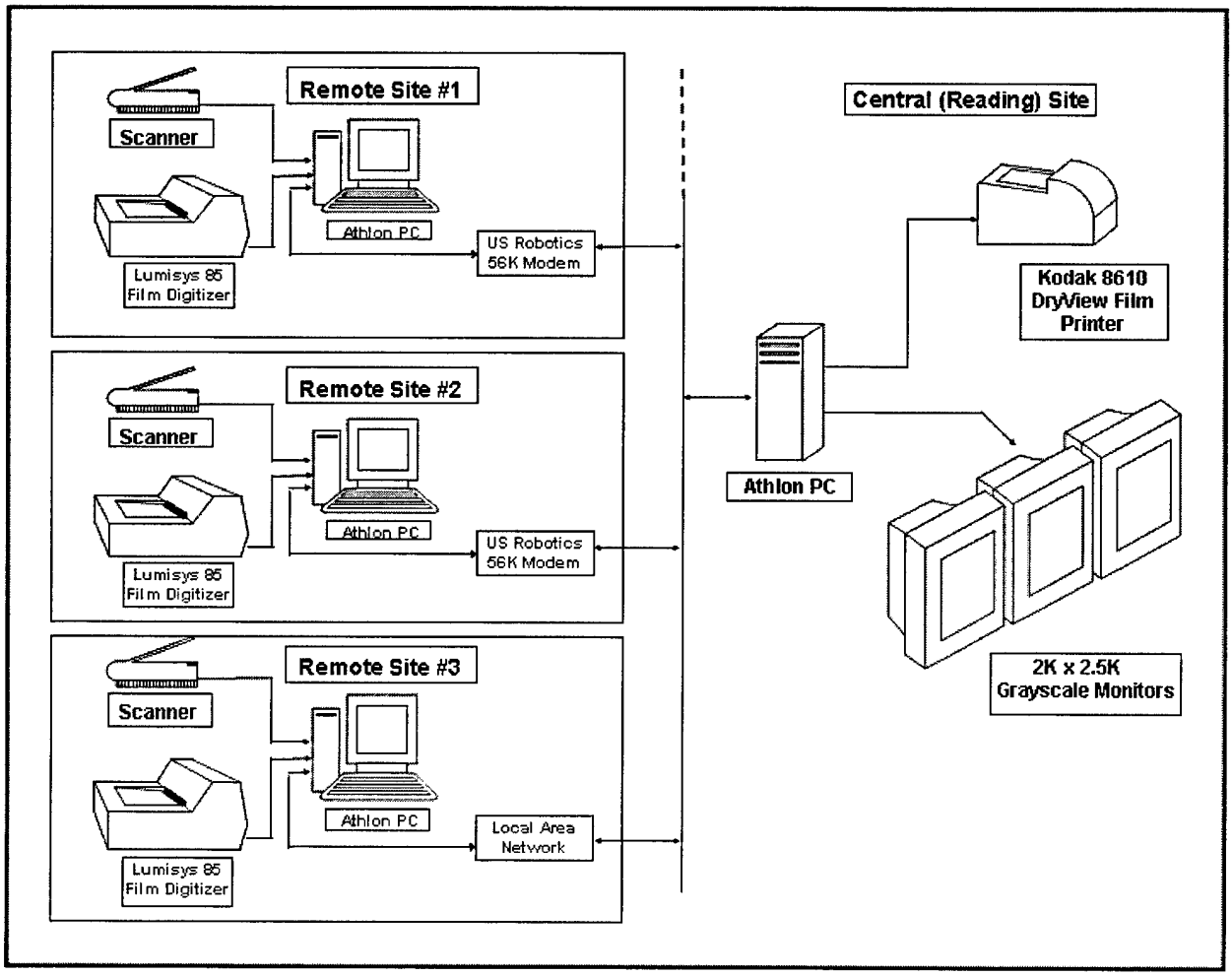


Fig. 1. Multi-site telemammography system schematic diagram of the remote and central sites.

The remote sites are staffed by mammography technologists. The computer hardware at the remote sites operates under Microsoft Windows 2000 Workstation powered by a 900 MHz processor (Athlon 900, Advanced Micro Device, Sunnyvale CA, USA) with 512 MB of RAM. High-resolution, laser film digitizers (Lumiscan 85, Eastman Kodak, Rochester, NY, USA) are connected to the remote computers via SCSI interface and equipped with a film feeder capable of holding six films as large as 10 x 12 inches. Mammographic films are digitized at 50 micron pixel dimensions and 12-bit grayscale. The remote site computers also have 56K hardware modems and ethernet network cards (Integrated PRO/100 S Desktop Adapter, Intel Corporation, Santa Clara, CA, USA) for data communication. Prior patient reports or history are transmitted along with the images by inserting them into an attached page scanner (hp scanjet 5490C, Hewlett-Packard, Palo Alto, CA, USA). Sites 1 and 2 transmit data across Plain Old Telephone System (POTS) lines and are located 15 and 20 miles from the central site, respectively (Fig.1). Site 3 is 90 miles from the central site and transmitted data across a Local Area Network (LAN).

### 2.2 Software Design

The software architecture at the central and remote sites is a multithreading design that allows independent task assignment with simultaneous response to user input. A message dispatch mechanism synchronizes bi-directional communication between all the main threads, except for the Time Manager (Fig. 2). The Time Manager periodically dispatches elapsed time messages to the other main threads without receiving messages. Each main thread acts on only messages associated with its function and may spawn subordinate (worker) threads that share data objects to accomplish tasks. A Reader/Writer lock, derived from Microsoft Windows synchronization primitives, prevents corruption of the shared data. The Reader/Writer lock permits access to the shared data to any number of readers simultaneously.

Central site main threads:

- Time manager - periodically indicates elapsed time.
- Archive manager - manages disk space by loading images, saving images, managing cases, and deleting archived cases when disk space is limited.
- Case manager - creates cases, assigns data, and performs database functions.
- Display manager - displays images and forwards messages to the main application window.
- Distribution manager - receives, transmits, and processes data.

Remote site main threads:

- Digitization manager - manages film digitizing.
- Case manager
- Display manager
- Distribution manager

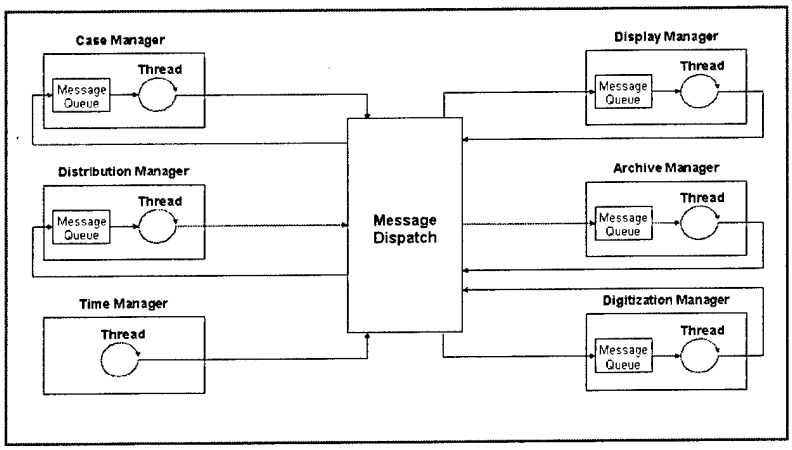


Fig. 2. Main threads and intra-process communication. Time manager does not receive messages.

## 2.3 Image processing

The first step in the series of the image processing procedures is designed to automatically crop each image to decrease the non-tissue area surrounding the breast (Fig. 3). The automated cropping algorithm begins by sub-sampling the image at an 8:1 ratio. The standard deviation (STD) of a 7 x 7 pixel mask is calculated at each sub-sampled pixel (STD of the sub-sampled image). Next, a threshold is applied to the STD image to separate tissue and non-tissue regions where a high STD indicated tissue regions. A region growing algorithm based on 4-neighbor connectivity is used to identify breast tissue as the largest region in the image. Finally, rudimentary logic is used to determine the cropping parameters based on the orientation of the tissue regions which is applied to the original image.

Following image cropping, the image data are compressed using the irreversible (lossy), 9/7 transform, wavelet-based JPEG 2000 method. Prior to transmission from the remote sites, the data packets are encrypted using strong 128 bit Microsoft Point-to-Point Encryption (MPPE) with version 2 authenticate Microsoft Challenge Handshake Authenticate Protocol (CHAP). The first steps at the central site are decryption and decompression of the image data.

Image display on the workstation monitors at the central site is enhanced by minimal unsharp masking of the decompressed image data prior to display. To begin unsharp masking, the image data are first smoothed with a 2-D 129 mean kernel. The weighted (0.10) smoothed image is subtracted from the decompressed image. The resulting pixel values of the image data are then re-scaled from 0 to 4095.

To minimize the need for manual adjustment during image viewing, default look-up table (LUT) values are automatically calculated based on the pixel value distribution (histogram). The typical pixel value distribution is bimodal. The window value (contrast) is set as the span of the two modes, and the level value (brightness) is set as the center between the two modes. The final stage of the image processing prior to image display is to pad (fill) the images to restore the full height of the image, but not the full width (Fig. 3).

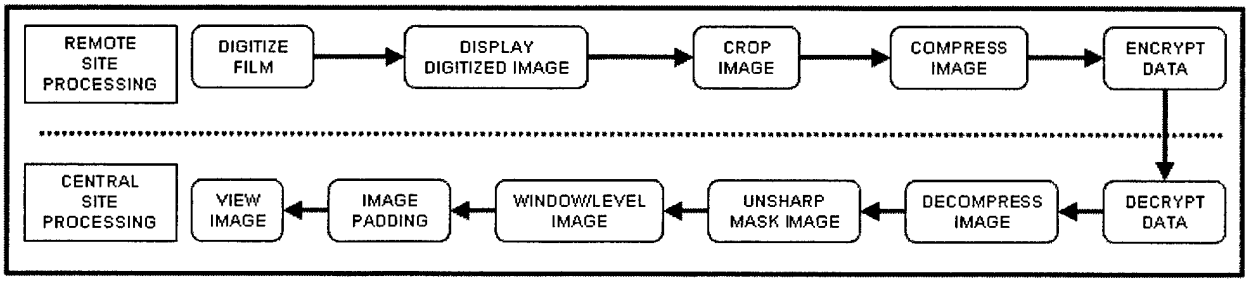


Fig. 3. Data flow of the telemammography system illustrating the order of the image processing tasks and where (remote or central) the process is performed.

## 2.4 Workstation display functions and features

To allow user-specific preferences to be used during case review, display options on the workstation are flexible with all features being mouse-driven. The default display is left and right craniocaudal views (LCC & RCC) on the left monitor, and left and right mediolateral oblique views (LMLO & RMLO) on the center monitor to be similar to our conventional clinical film presentation (Fig. 4). However, a large number of display options are available to users. If a film is digitized in an incorrect orientation, the user has the ability to flip images (top to bottom or left to right) and rotate images 180 degrees. Communication from the remote site is displayed on the right monitor (Fig. 4).

Two forms of image magnification are available on the workstation display. Typically, the normal display scale with a single image per monitor is approximately 100 micron pixel dimensions and with two images per monitor it is approximately 200 micron pixel dimensions. A scrollable image magnification box provides a true 1:1 presentation (monitor pixel:digitized pixel) resulting in 50 micron pixel dimensions. The size of the box varies from 511 x 566 pixels for one image/monitor, and 408 x 566 pixels for two images/monitor, and 204 x 266 pixels for four images/monitor. It is also possible to pan across the image quadrant-by-quadrant.

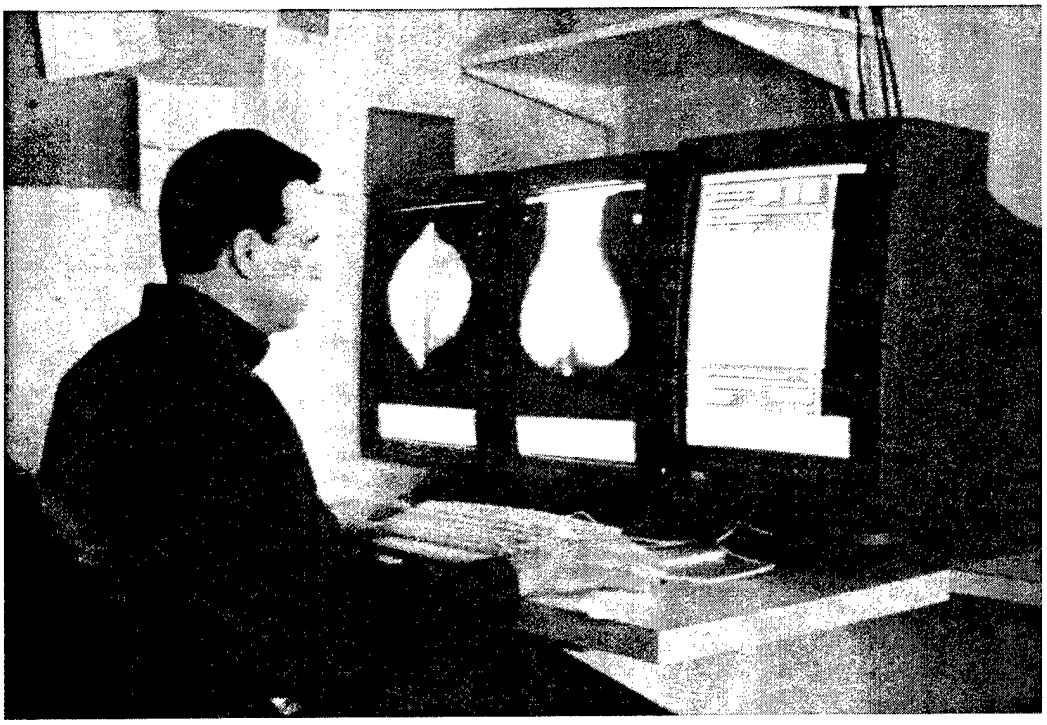

Fig. 4. Telemammography workstation at the central site pictured in the default image display format.

The automated LUT values can be manually adjusted per observer's preference. The window and level values are determined based on the mouse position (movement), and the image display is instantly updated as the mouse is moved. Once the desired values are determined, these can be applied to the individual image or all images associated with the case. The LUT values can be reset to the automated (default) values at anytime during viewing.

### 2.5 Inter-site communication

To facilitate effective communication between the technologists (remote site) and radiologists (central site), a "chat box" type messaging function was implemented. The "chat message" can be sent with each case and it provides a real-time, interactive communication tool between the sites. During the initial phase of evaluating the system, communication is performed in one cycle. The technologist sends a chat message with each case, and the radiologist responds directly to the message. The chat boxes on both sides contained four general areas: (1) patient demographics, (2) message display area, (3) pull-down menus, and (4) free text area (Fig. 5). There are five pull-down menus on the technologist chat box to focus communication on possible actionable items. These indicate: (1) breast: left or right; (2) view: craniocaudal and/or mediolateral oblique; (3) finding: mass or calcifications; (4) comparison with prior exam: baseline, new, or change in findings; and (5) possible additional procedure needed: additional views and/or ultrasound. The radiologists can reply after reviewing each case. His/her response includes: (1) do recommended procedure as suggested; (2) no additional procedures necessary; and (3) do not do the procedure recommended, but do X, Y, and Z.

### 2.6 Technical and operational evaluation

In the preliminary technical assessment phase, three processes of the telemammography system were evaluated. First to assess the user's acceptance of the automated LUT values for image review without the need to adjust display parameters, 50 cases (200 images) sent from all sites were subjectively rated by two experienced observers on a scale of 1 to 4. The experiment was designed to assess acceptability of default values for the purpose of reviewing each case and determining the need (or not) for additional procedures. In all of our studies we evaluated the system under normal operating conditions. As a result, intra- and inter-site measured variability reflect what could be expected in an "on-line" clinical operation. Second, the implementation of high-level image compression in mammographic imaging was evaluated during subjective Just Noticeable Difference (JND) studies. The studies compared images at no compression,

50:1, and 75:1 compression levels. Third, the average cycle time from initiation of digitization to availability for display at the central site was evaluated. This involved transmission of a series of four cases (back to back) each consisting of four images per case (all images were 8 x 10 inches).

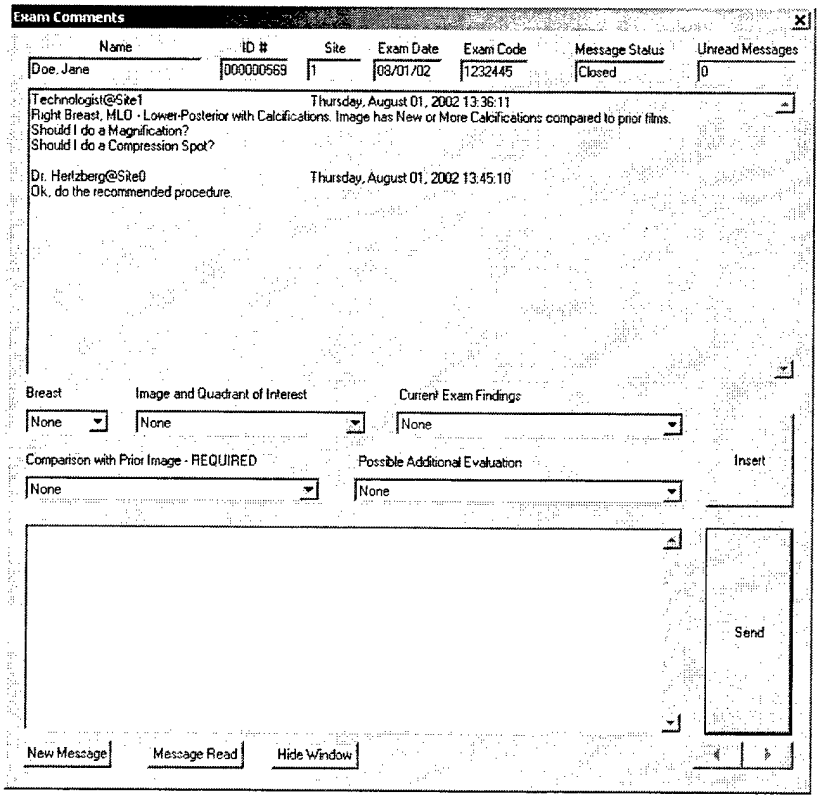


Fig. 5. "Chat box" for the remote site technologists.

## 3. RESULTS

The evaluation of the technical and operational processes was favorable in all areas. The automated LUT settings, the image cropping, the high level of image compression, and the cycle time to transmit and receive cases were all acceptable for implementation of the telemammography system for the designed purpose. The initial impressions of the inter-site communication, "chat messaging," indicate that it can facilitate effective communication between the technologist at remote sites and the radiologist at the central site. Although the technical issues with regard to scanning and transmitting patient reports with each case have been resolved, the practice of has not been implemented to date.

The automatically calculated LUT settings were reported as "acceptable" to "excellent" by two experienced mammography researchers. On a scale of 1 to 4 (1 = unusable, 2 = need minor adjustments, 3 = acceptable, and 4 = excellent), the two observers had mean ratings for 200 automatically computed LUT settings of 2.64 (STD = 0.57) and 3.51 (STD = 0.53). After minor adjustments were made as the result of the above experiment, all observers including clinicians using the workstation to test different aspects of the system accepted automatically set values in over 90% of cases. Consequently, window and/or level manipulations are being performed in less than 10% of cases during retrospective and simulated prospective case reviews.

For review of non-magnified or moderately magnified images, 50:1 and 75:1 data compression levels were comparable and acceptable when evaluated on either laser-printed films or the telemammography workstation. Subjective JND studies were conducted using laser-printed films as well as images displayed side-by-side on workstation monitors. The studies indicated that at extreme magnifications, differences were detected, but did not necessarily result in degradation

of perceived diagnostic quality. For example, the "visibility" and "clarity" of microcalcifications in the digital images were judged as "almost equivalent" between the full-scale, non-compressed images and images compressed at a 75:1 ratio (Figs. 6 and 8). Comparable results were obtained with magnification (Figs. 7 and 9). The automated cropping did not remove breast tissue in any of our cases to date, and it produces "aesthetically pleasing" images.

The time to transmit and receive four films (8 x 10 inches each) was reliably less than 7 minutes/case for each site using 75:1 data compression (Table 1). The combination of image cropping and 75:1 data compression ratio decreased image file size to allow cycle times that were adequate for implementation of the telemammography concept and met our planned technical specifications. Sites 1 and 2 were connected via 56K modems that dialed a four digit telephone number (i.e., connected via an in-house telephone line) and a ten digit telephone number (i.e., connected via an outside telephone line), respectively. Consistent bandwidths of sites 1 and 2 were approximately 33 Kbits/second and 21 Kbits/second, respectively. The digitization process (approximately 50 seconds/film) was the limiting factor at site 3 which was connected via LAN. Site 2 had communication problems (decreased bandwidth) during the first measurement that have been largely resolved.

**TABLE 1**
**Experimentally Measured Average Cycle Time for Digitizing, Transmitting and Receiving a Case with 4 Films (8 x 10 inches each)**

| Image format | Site 1 - POTS* (min/case) | Site 2 - POTS (min/case) | Site 3 - LAN (min/case) |
|---|---|---|---|
| 50:1 compression, not cropped, and not encrypted | 13.22 | 24.42 | 5.38 |
| 50:1 compression, cropped, and encrypted | 6.47 | 13.13 | 5.65 |
| 75:1 compression, cropped, and encrypted | 5.97 | 6.85 | 5.77 |

*in-house POTS

## 4. DISCUSSION

The "proof of concept" to design an inexpensive, high-quality, multi-site telemammography system implemented with low-level data connections has been established to facilitate the concept of "almost real-time" distributed acquisition/centralized review. The technical feasibility of the concept was demonstrated by: (1) the digitization of films acquired during clinical breast cancer screening mammography; (2) the timely transmission of the digitized images across low-level data connections (less than 7 minutes/case); and (3) the efficient archiving, retrieving, and viewing of image data at the central site. The short cycle time of the system was realized because of the image file size reduction due to automated image cropping and image data compression and the efficient multi-tasking software approach based on a synchronized multi-threading design. Image processing methods were fundamental to the success of the telemammography system. The automated cropping and compression produced images without a significant degradation of the diagnostic image quality, which were well-received by the radiologists. Although the automated window and level calculations were found to be acceptable, in approximately ten percent of cases, radiologists manually employed window and level settings during an individual case review. The high-resolution image display of the telemammography workstation was rated acceptable for reviewing screening mammographic images for the purpose of determining the need for additional procedures.

To date, over 1000 screening exams have been successfully transmitted using the telemammography system. The preliminary results suggest that the telemammography system could accomplish the goals to increase effective communication between remote "underserved" sites and the central location, and permit experienced radiologists to remotely monitor and facilitate some decision making while the patient remains in the clinic. The addition of two key components to the telemammography system should improve the system's capability and effective utilization. First, scanned prior patient reports will be added to the information transmitted with each case. Second, Computer Aided Detection (CAD) schemes will be incorporated into the system and the results will be displayed at the central site.

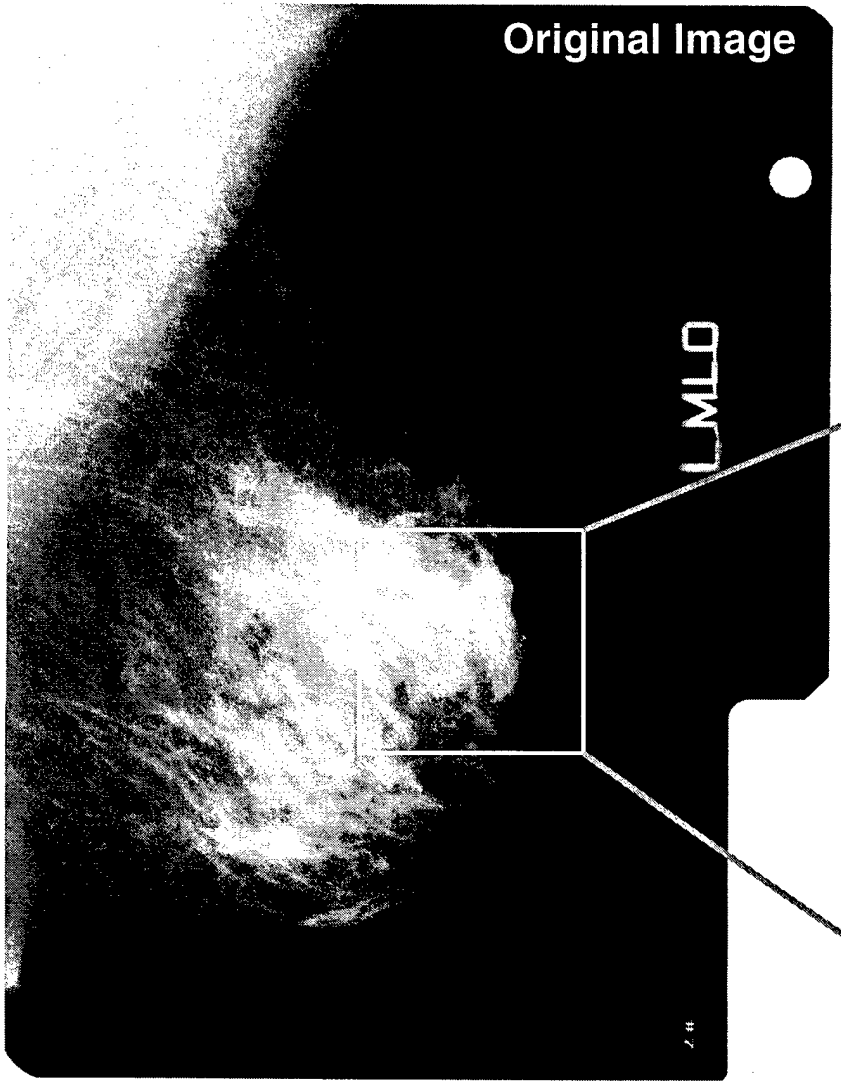


Fig. 6. Original left medial lateral oblique image of patient #569. Image is not cropped, compressed, or unsharp masked.

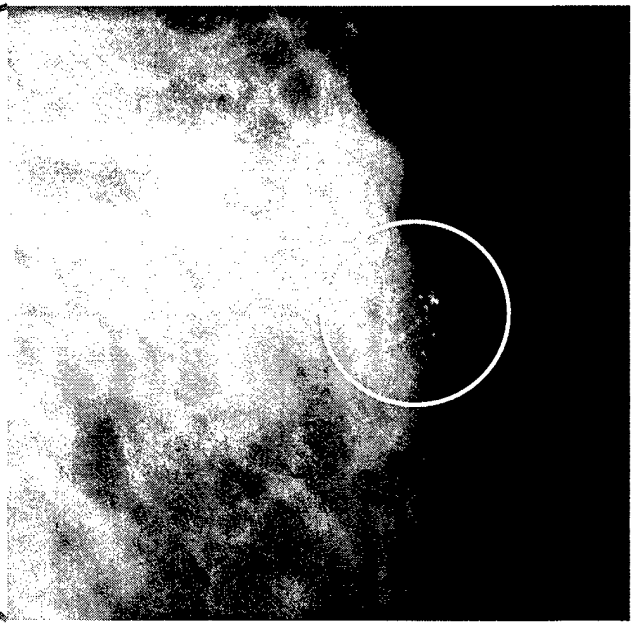

Fig. 7. Magnified region of Fig. 6 with microcalcifications marked.

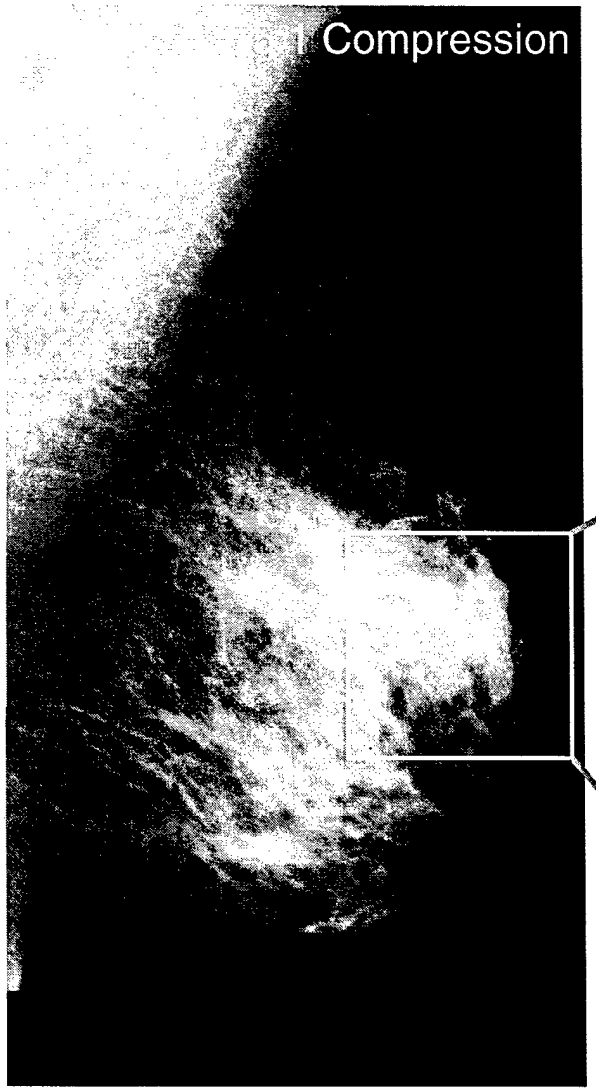


Fig. 8. Processed left medial lateral oblique image of patient #569. Image is cropped, compressed at a 75:1 ratio, and unsharp masked.

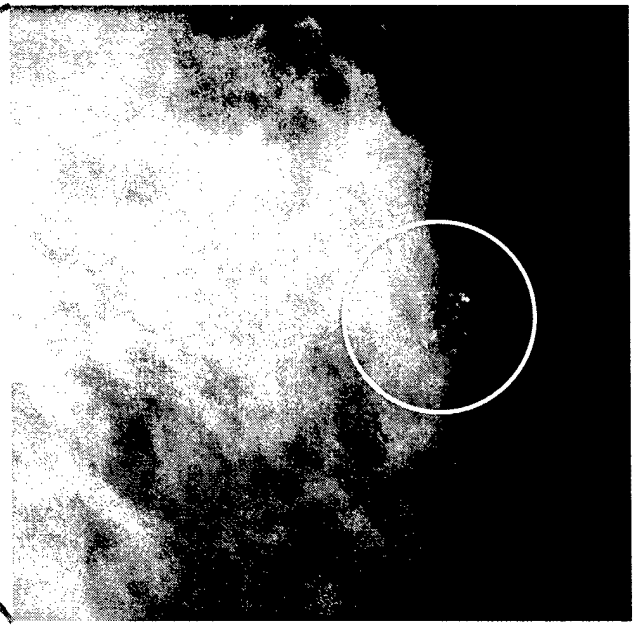

Fig. 9. Magnified region of Fig. 8 with microcalcifications marked.

## ACKNOWLEDGEMENTS

This work is supported in part by the US Army Medical Research Acquisition Center, 820 Chandler Street, Fort Detrick, MD 21702-5014 under contract DAMD17-00-1-0410. The content of the information contained herein does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred.

## REFERENCES

1. Duffy SW, Tabar L, Chen HH, Holmqvist M, Yen MF, Abdsalah S, Epstein B, Frodis E, et al. "The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish Counties." *Cancer* **95(3)**:458-469, 2002.
2. Feig SA. Current status of screening mammography. *Obstet Gynecol Clin North Am* 2002;29(1):123-136.
3. Huphrey LL, Helfand M, Chan BKS, and Woolf SH. "Breast cancer screening: a summary of the evidence for the U.S. preventive services task force." *Ann Intern Med* **137(5, Part 1)**:E347-E367, 2002.
4. Lee CH. "Screening mammography: proven benefit, continued controversy." *Radiol Clin North Am* **40(3)**:395-407, 2002.
5. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, and Rutqvist LE. "Long-term effects of mammography screening: update overview of the Swedish randomised trials." *Lancet* **359(9310)**:909-919, 2002.
6. Tabar L, Vitak B, Chen HHT, Yen MF, Duffy SW, and Smith RA. "Beyond randomized clinical trials: organized mammographic screening substantially reduces breast carcinoma mortality." *Cancer* **91(9)**:1724-1731, 2002.
7. Gotzsche PC and Olsen O. "Is screening for breast cancer with mammography justifiable?" *Lancet* **355**:129-134, 2000.
8. Olsen O and Gotzsche PC. "Cochrane review on screening for breast cancer with mammography." *Lancet* **358**:1340-1342, 2001.
9. Miller AB, To T, Baines CJ, and Wall C. "The Canadian Nation Breast Screening Study-1: nreast cancer mortality after 11 to 16 years of follow-up: a randomized screening trial of mammography in women age 40 to 49 years." *Ann Intern Med* **137(5, Part 1)**:E305-E315, 2002.
10. Chamot E and Perneger TV. "Misconception about efficacy of mammography screening: a public health dilemma." *J Epidemiol Community Health* **55(11)**:799-803, 2001.
11. Coughlin SS, Thompson TD, Hall HI, Logan P, and Uhler RJ. "Breast and cervical carcinoma screening practices among women in rural and nonrural areas of the United States, 1998-1999." *Cancer* **94(11)**:2801-2812, 2002.
12. Michaelson J, Satija S, Moore R, Weber G. Halpern E, Garland A, Puri D, and Kopans DB. "The pattern of breast cancer screening utilization and its consequences." *Cancer* **94(1)**:37-43, 2002.
13. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. "Variability in radiologists' interpretations of mammograms." *N Engl J Med* **331(22)**:1493-1499, 1994.
14. Warren RML and Duffy SW. "Comparison of single reading with double reading of mammograms and change in effectiveness with experience." *Br J Radiol* **68(813)**:958-962, 1995.
15. Hulka CA, Slanetz PJ, Halpern EF, Hall DA, McCarthy KA, Moore R, Boutin S, and Kopans DB. "Patients' opinion of mammography screening services: immediate results versus delayed results due to interpretation by two observers." *AJR Am J Roentgenol* **168**:1085-1089, 1997.
16. Yawn B, Krein S, Christianson J, Hartley D, and Moscovice I. "Rural radiology: who is producing images and who is reading them?" *J Rural Health* **13(2)**:136-144, 1997.
17. Lou SL, Lin HD, Lin KP, and Hoogstrate D. "Automatic breast region extraction from digital mammograms for PACS and telemammography applications." *Comput Med Imaging Graph* **24(4)**:205-220, 2000.
18. Maitz GS, Chang TS, Sumkin JH, Wintz PW, Johns CM, Ganott M, Holbert BL, Hakim CM, Harris KM, Gur D, and Herron JM. "Preliminary clinical evaluation of a high-resolution telemammography System." *Invest Radiol* **32(4)**:236-240, 1997.
19. Drescher JM, Maitz GS, Leader JK, Sumkin JH, Poller WR, Klaman H, Zheng B, and Gur D. "Design considerations for a multi-site, POTS-based telemammography system." *Proceedings of SPIE Medical Imaging 2002: PACS and Integrated Medical Information Systems: Design and Evaluation*, 4685:416-421, San Diego, CA, February 2002.

# Improving CAD performance in detecting masses depicted on prior images

Bin Zheng[a], Xiao-Hui Wang[a], Luisa Wallace[b], Cathy Cohen[b], Lara A. Hardesty[b], Christiane M. Hakim[b], Gordon Abran[b], Jules Sumkin[b], and David Gur[a].
[a] Department of Radiology, University of Pittsburgh, Pittsburgh, PA, 15261
[b] Magee Womens Hospital, 300 Halket Street, Suite 4200, Pittsburgh, PA 15213

## ABSTRACT

We investigated a new approach to improve the performance of a computer-aided detection (CAD) scheme in identifying masses depicted on images acquired earlier ("prior"). The scheme was trained using a dataset with simulated mass features. From a database with images acquired during two consecutive examinations, 100 locations matched pairs of malignant mass regions were selected in both the "current" and the most recent "prior" images. While reviewing the current images, mass regions were identified and as a result biopsies were ultimately performed. Prior images were not identified as suspicious by radiologists during the original interpretation. The same number of false-positive regions was also selected in both current and prior images. The selected regions were then randomly divided into training and testing datasets with 50 true-positive and 50 false-positive regions in each. For each selected region, five features; area, contrast, circularity, normalized standard deviation of radial length, and conspicuity; were computed. The ratios of the average difference of five feature values between current and prior mass regions in the training datasets were also computed. Multiplying these ratios by the computed values in current mass regions, we generated a new dataset of simulated features of "prior" mass regions. Three artificial neural networks (ANN) were trained. ANN-1 and ANN-2 were trained using training datasets of current and prior regions, respectively. ANN-3 was trained using simulated "prior" dataset. The performance of three ANNs was then evaluated using the testing dataset of prior images. Areas under ROC curves ($A_z$) were 0.613 ± 0.026 for ANN-1, 0.678 ± 0.029 for ANN-2, and 0.667 ± 0.029 for ANN-3, respectively. This preliminary study demonstrated that one could estimate an average change of feature values over time and "adjust" CAD performance for better detection of masses at an earlier stage.

**Keywords:** Computer-aided detection, Mammography, Mass detection, Artificial neural network

## 1. INTRODUCTION

Computer-Aided Detection (CAD) systems are currently used in a large number of medical institutions around the world to assist radiologists in reading and interpreting mammograms in the screening environment [1-3]. A large number of studies have been conducted to assess the possible impact of CAD systems on radiologists' performance. Although there is no general agreement on whether and how CAD systems help radiologists improve their diagnostic accuracy [3-6], several studies demonstrated that the performance of the CAD scheme itself might be an important factor to increase radiologists' confidence to accept and act on the CAD cues and help to improve their diagnostic accuracy when using such tools [6-8].

Current guidelines recommend periodic mammography screening for women over the age of 40 [9]. As compliance increases in the general population, a large fraction of patients will have undergone series of consecutive mammographic examinations. As a result, detected breast cancers will in time, "shift" on the average toward an earlier stage. In fact, retrospective review have indicated that a large fraction of breast cancers that are identified by radiologists were also visible in prior images [10]. It is expected that comparison with prior images could over time help radiologist detect

Medical Imaging 2003: Image Processing, Milan Sonka, J. Michael Fitzpatrick, Editors,
Proceedings of SPIE Vol. 5032 (2003) © 2003 SPIE · 1605-7422/03/$15.00

more subtle cancers [11,12], hence, more subtle cancers will be considered "visible" or detectable on routine mammograms. In such a changing environment, maintaining "optimal" performance of CAD schemes becomes a challenge. Although CAD schemes can detect a large number of true-positive abnormalities (e.g., masses and microcalcification clusters) depicted on prior images [7,12,13], current CAD schemes that had been optimized using a large fraction of "easy" cancers are unlikely to achieve "optimal" performance in detecting "earlier" or more "subtle" cancers. This is due to several factors: (1) performance of CAD schemes that use a feature-based machine-learning classifier heavily depends on the characteristics of training database [14,15] and (2) a large number of image features used to train CAD schemes varies differently for abnormalities as depicted on the current images as compared with prior images [16]. Several studies have demonstrated that in order to achieve optimal performance in detecting suspicious masses as depicted on prior images, a different set of image features should be selected for re-optimization of CAD schemes [17,18].

In previous studies [17,18] optimal performance in detecting masses depicted on prior images was achieved by re-training the scheme using a set of mass regions extracted from prior images. This requires a significant effort. Since there is a training database available for each CAD scheme, this database could potentially be used to re-optimize the scheme after a computational adjustment of some feature values. For this purpose, we investigated a new method to generate a simulated training database and used it to re-optimize our CAD scheme. A detailed description of our approach and preliminary experimental results follow.

## 2. MATERIALS AND METHODS

From an image database established in our laboratory, we selected 100 matched pairs of digitized mammograms from two consecutive (the most recent or "current" and the latest previous or "prior") examinations. There is a verified mass region depicted in each case. During the current examination, these 100 mass regions were identified by radiologists as suspicious and as a result biopsies were ultimately performed. Although in a retrospective review and with the support of available source documents, an experienced observer could identify some indication of the presence of a "mass" in the corresponding locations on prior images, these regions had not been identified as suspicious by radiologists during the original interpretation. All 100 mass regions selected for this study were associated with biopsy-proven malignancies. The locations of all masses depicted on current images and the corresponding locations on prior images were visually identified. The centers (x, y coordinate) of all verified mass regions were marked manually and saved in a reference (or "truth") file.

All 200 images (100 from current and 100 from prior examination) were processed by a CAD scheme developed previously in our laboratory [19]. To detect suspicious masses, each image is first subsampled (pixel-averaged) in both dimensions to increase pixel size from original 50 μm × 50 μm (or in some cases 100 μm × 100 μm) to 400 μm × 400 μm. The CAD scheme then uses three stages to identify suspicious regions. In the first stage, the scheme uses image subtraction and threshold results after processing by two Gaussian filters with a large difference in the kernel sizes (7 and 51 pixels) to search for the initial set of "suspicious" regions, which usually generates in the range of 10 to 30 initial "suspicious" regions per image. In the second stage, based on local contrast measurement the scheme uses an adaptive region growth algorithm to define three topographic layers. After simple intra-layer based threshold conditions on growth ratio and shape factor, this stage typically eliminates approximately 85% of regions identified in stage one, while maintaining a very high sensitivity. A set of features is computed for each detected region. During stage three the detected regions are classified based on scores (likelihood of being true-positive) generated by a nonlinear multi-layer feature-based classifier (e.g., an artificial neural network) [20]. To determine whether a detected region represents a true-positive or false-positive mass region in this study, the following criterion was used. If the distance between the center of gravity of a detected region and the center of the mass as recorded in the reference file was shorter than the radius of the longest axis of the detected region, it was considered as a true-positive identification. Otherwise, the region was considered a false-positive identification. In this experiment, all suspicious mass regions identified after the second stage of the CAD scheme became candidates for the study (namely, the classification scores in the third stage were ignored). One hundred true-positive mass regions from current images and 100 mass regions from prior images were selected. The CAD scheme detected 187 and 202 false-positive mass regions in the current and prior images as well. From these, 200 false-positive regions were randomly selected (100 from current images and 100 from prior images). Hence, 400

suspicious mass regions were selected for the study. The regions were then divided (block randomization) into training and testing datasets for both current and prior images. Each dataset included 50 true-positive and 50 false-positive mass regions.

For each region the following five features were computed:

1. Region area ($F_1 = 0.16 \times N_T$): This feature is computed by counting the number of pixels in the growth region ($N_T$) and then multiplying it by the size unit of each pixel (0.16 $mm^2$).
2. Average contrast ($F_2 = \frac{1}{N_T}\sum_{i=1}^{N_T} I_i - \frac{1}{N_B}\sum_{j=1}^{N_B} I_j$): This feature is computed by the average pixel value ($I$) difference between the growth region and its surrounding background.
3. Circularity ($F_3 = \frac{N_C}{N_T}$): To compute this feature, CAD scheme first computes the area of a growth region ($N_T$) and calculates an equivalent circle originating at the center of gravity of the region. For a circle with the same size as the growth region, the number of pixels that are located inside the growth region contour and the circle ($N_C$) is computed. Circularity is defined as the fraction of the growth region pixels covered by the circle.
4. Normalized standard deviation of radial length ($F_4 = \sqrt{\frac{1}{N_b}\sum_{i=1}^{N_b}(\frac{r_i - m_r}{m_r})^2}$): The radial length $r_i$ is defined as the distance between the region center and a point ($i$) located on the perimeter of the region. $m_r$ is the mean value of radial length over all points $N_b$ in the region boundary. This feature indicates the changes in the shape of region boundary.
5. Conspicuity ($F_5 = \frac{F_2}{C_B}$): This feature is defined as "region contrast" ($F_2$) divided by "surrounding complexity" ($C_B$); where $C_B = \frac{1}{N_B}\sum_{i=1}^{N_B} |Max(I_i - I_F)|$ and $Max(I_i - I_F)$ is the maximum pixel value difference between background pixel ($i$) and its neighboring pixels (e.g., 24 pixels in a $5 \times 5$ square window).

Using these features, three artificial neural networks (ANN) were constructed to classify suspicious regions. The topology of all ANNs was the same. It involved five input neurons (each represented by one feature), three hidden neurons, and one output neuron. The ANN was trained using 500 iterations. The training momentum and learning rate were 0.8 and 0.01, respectively.

ANN-1 and ANN-2 were trained using training dataset of current and prior images, respectively. ANN-3 was trained using a set of simulated "prior" mass regions. To generate a simulated dataset, we computed the ratio of the average feature values for each of five features between 50 pairs of true-positive mass regions as extracted from current and prior images. Ratios were computed as follows:

$$D_k = \frac{\frac{1}{N}\sum_{i=1}^{N} F_{k,i}^{Prior}}{\frac{1}{N}\sum_{j=1}^{N} F_{k,j}^{Current}}, \quad k = 1,2,3,4,5. \text{ and } N = 50.$$

Each feature of true-positive mass region in the current training dataset was then multiplied by the ratio, such as $F'_{k,j} = F^{Current}_{k,j} \times D_k$. Hence, a set of new feature values was generated to represent each of 50 "simulated true-positive mass regions." Using these data combined with feature values of 50 original false-positive regions extracted from the current images, ANN-3 was trained. Although the 50 simulated mass regions (used in ANN-3) and 50 original prior mass regions (used in ANN-2) have identical mean values for each of the five features, the feature values for a specific region are different (i.e., $F'_{k,j} \neq F^{Prior}_{k,j}, k = 1,2,\ldots,5$). In other word, the simulated set of "prior" features does not simply duplicate the actual feature set in prior images.

The performances of three ANNs were evaluated separately using testing datasets of 50 current and 50 prior images. For each test region, the ANN generates a classification score ranged from 0 to 1, where the larger the score, the higher the computed likelihood of being a true-positive mass region. The classification scores generated for all test regions were used as input data in the ROCFIT program that generates a receiver operating characteristic (ROC) curve and computes the area under the ROC curve ($A_z$ value) [21]. We compared performance levels when using the three ANNs to classify an independent set of suspicious mass regions as depicted on prior images.

## 3. RESULTS

Table 1 shows the averages of the five feature values in the two training datasets extracted from the current and prior images. Using paired chi-square test to examine the mean values of each of the five features between 50 pairs of training mass regions, the significant difference ($p < 0.05$) was found in the average value of each of the five features. Table 2 summarizes the areas under ROC curves ($A_z$ values) for all three ANNs during training and testing. Figure 1 demonstrates three ROC curves generated by applying three ANNs to the prior testing dataset. ANN-1 yields the best performance in testing current dataset ($A_z$ = 0.781 ± 0.019) and the worst performance in prior testing dataset ($A_z$ = 0.613 ± 0.026) as shown in table 2. Both ANN-2 and ANN-3 yield significantly better performance than ANN-1 in classifying mass regions on prior testing dataset ($p < 0.05$). $A_z$ values were increased by 10.6% (from 0.613 to 0.678) in ANN-2 and 8.8% in ANN-3 (from 0.613 to 0.667), respectively. The experimental results also demonstrated that there was no significant performance difference between ANN-2 and ANN-3 in testing prior dataset ($p = 0.15$).

Table 1: Average feature values and their difference ratios between 50 pairs of mass regions depicted on current and prior images.

| Feature: | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ |
|---|---|---|---|---|---|
| Average value (prior images): | 78.60 | 34.90 | 0.24 | 0.78 | 4.25 |
| Average value (current images): | 122.67 | 42.68 | 0.21 | 0.83 | 5.07 |
| Ratio: | 0.70 | 0.82 | 1.14 | 0.94 | 0.84 |

Table 2: Areas under ROC curves ($A_z$ values) of three ANNs during training and testing.

| Network | Training | Testing current images | Testing prior images |
|---|---|---|---|
| ANN-1 | 0.873 ± 0.016 | 0.781 ± 0.019 | 0.613 ± 0.026 |
| ANN-2 | 0.761 ± 0.021 | 0.709 ± 0.026 | 0.678 ± 0.029 |
| ANN-3 | 0.779 ± 0.019 | 0.736 ± 0.028 | 0.667 ± 0.029 |

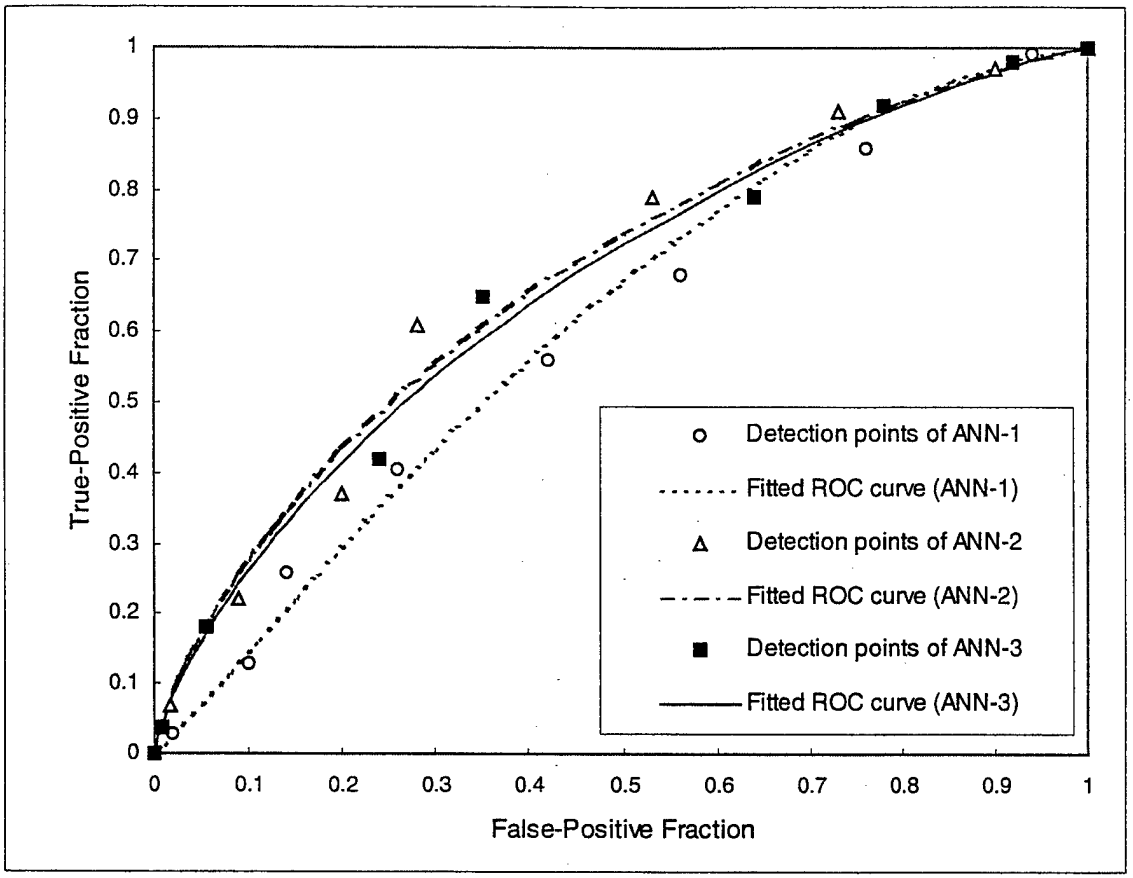


Figure 1: ROC curves of testing results when applying three ANNs to the test dataset of prior images.

## 4. DISCUSSION

With improvements of diagnostic technologies and increase in screening compliance of the general population, radiologists have to detect increasingly more subtle abnormalities as depicted on mammograms. As a result, CAD systems that currently provide satisfactory cueing results could face deterioration in performance over time due to a general shift in the subtleness of and stage at detection. Feature-based machine learning classifiers, such as ANNs, are widely used in final stage of the CAD schemes for identifying masses and microcalcification clusters. Since these classifiers are trained to generate "global" functions that cover the entire instance space, CAD performances heavily depend on the training databases [22]. This is true, in particular, in mammography where the size and diversity of training datasets is generally limited [14,15]. A single CAD scheme that achieves high sensitivity on both "subtle" and relatively "easy" masses at an acceptable false-positive rate can be developed, however, in reality, it is a very difficult task because image features are substantially different for suspicious mass regions extracted from the current and prior images [16,17]. In order to improve CAD performance in detecting subtle masses in an earlier stage, the schemes should be trained (or optimized) using databases involving a large fraction of subtle mass regions (e.g., new cases that had been rated originally as negative and later proven to be positive) [17,18].

However, it is a very difficult and time-consuming task to collect a large number of diverse subtle cases (e.g., the false-negative cases). This study demonstrated an alternative approach to collectively simulate such cases. By systematically adjusting the feature values extracted from current images, we generated a set of simulated "prior" mass

regions. Our results demonstrated that (1) an ANN trained using simulated prior mass regions could achieve significantly better performance in detecting the masses at an earlier stage than an ANN trained using current mass regions and (2) there is no significant difference in the performance between the ANNs trained using either real or simulated prior mass regions. As a result, by estimating the change over time of some important features, one can adjust CAD performance for better detection of masses at an earlier stage. Since this is a very preliminary study involving a limited database and a small set of features, the concept need to be further investigated. If this approach is validated with significantly larger image databases and larger number of features, it may provide a simple and efficient method to periodically update (or re-optimize) CAD schemes.

## ACKNOWLEDGEMENTS

This work is supported in part by Grants CA77850, CA85241 and CA80836 from the National Cancer Institute of National Institutes of Health and also by the US Army Medical Research Acquisition Center, 820 Chandler Street, Fort Detrick, MD 21702-5014 under Contract DAMD17-00-1-0410. The content of the information contained here does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred.

## REFERENCES

1. A. Malich, C. Marx, M. Facius, T. Boehm, M. Fleck, W.A. Kaiser, "Tumour detection rate of a new commercially available computer-aided detection system," *Eur Radiol* 11:2454-2459 (2001).
2. M. Lechner, M. Nelson, E. Elvecrog, "Comparison of two commercially available computer-aided detection (CAD) systems," *Appl Radiol* 31:31-35 (2002).
3. T.W. Freer, M.J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* 220:781-786 (2001).
4. K. Moberg, N. Bjurstam, B. Wilczek, L. Rostgard, E. Egge, C. Muren, "Computed assisted detection of interval breast cancers," *Eur J. Radiology* 39:104-110 (2001).
5. W.R. Hendee, "In the next decade automated computer analysis will be an accepted sole method to separate "normal" from "abnormal" radiological images," *Med Phys* 26:1-4 (1999).
6. B. Zheng, M.A. Ganott, C.A. Britton, C.M. Hakim, L.A. Hardesty, T.S. Chang, H.E. Rockette, D. Gur, "Soft-copy mammographic reading with different computer-assisted detection cueing environments: Preliminary findings," *Radiology* 221:633-640 (2001).
7. G.M. te Brake, N. Karssemeijer, J.H. Hendriks, "Automated detection of breast carcinomas not detected in a screening program," *Radiology* 207:465-471 (1998).
8. A. Malich, T. Azhari, T. Bohm, M. Fleck, W.A. Kaiser, "Reproducibility – an important factor determining the quality of computer aided detection (CAD) systems," *Eur J. Radiology* 36:170-174 (2000).
9. S.A. Feig, C.J. D'Orsi, R.E. Hendrick, "American College of Radiology guidelines for breast cancer screening," *Am J Roentgenol* 171:29-33 (1998).
10. L.J. Warren Burhenne, S.A. Wood, C.J. D'Orsi, S.A. Feig, D.B. Kopans, K.F. O'Shaughnessy, E.A. Sickles, L. Tabar, C.J. Vyborny, R.A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* 215:554-562 (2000).
11. L.W. Bassett, B. Shayestehfar, I. Hirbawi, "Obtaining previous mammograms for comparison: usefulness and cost," *Am J Roentgenol* 163:1083-1086 (1994).
12. M.P. Callaway, C.R. Boggis, S.A. Astley, "Influence of pervious films on screening mammographic interpretation and detection of breast carcinoma," *Clin Radiol* 52:527-529 (1997).
13. R.L. Birdwell, D.M. Ikeda, K.F. O'Shaughnessy, E.A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology* 219:192-202 (2001).
14. B. Zheng, Y.H. Chang, W.F. Good, D. Gur, "Adequacy testing of training set sample sizes in the development of a computer-assisted diagnosis scheme," *Acad Radiol* 4:497-502 (1997).
15. M.A. Kupinski, M.L. Giger, "Feature selection with limited database," *Med Phys* 26:2176-2182 (1999).

16. B. Zheng, R. Shah, L. Wallace, C. Hakim, M.A. Ganott, D. Gur, "Computer-aided detection in mammography: An assessment of performance on current and prior images," *Acad Radiol* in press (2002).
17. L. Hadjiiski, B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvie, M. Gurcan, "Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses," *Med Phys* 28:2309-2317 (2001).
18. B. Zheng, W.F. Good, D.R. Armfield, C. Cohen, T. Hertzberg, J.H. Sumkin, D. Gur, "Performance changes of mammographic CAD schemes optimized using most recent and prior image databases," *Acad Radiol* in review (2002).
19. B. Zheng, Y.H. Chang, D Gur, "Computerized detection of masses in digitized mammograms using single image segmentation and a multi-layer topographic feature analysis," *Acad Radiol* 2:959-966 (1995).
20. B. Zheng, J.H. Sumkin, W.F. Good, G.S. Maitz, Y.H. Chang, D. Gur, "Applying computer-assisted detection schemes to digitized mammograms after JPEG data compression: An assessment," *Acad Radiol* 7:595-602 (2000).
21. C.E. Metz, B.A. Herman, J.H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Stat Med* 17:1033-1053 (1998).
22. T.M. Mitchell, *Machine learning,* WCB/McGraw-Hill, Boston, MA, (1997).

PROGRESS IN BIOMEDICAL OPTICS AND IMAGING

SPIE—The International Society for Optical Engineering

# Preliminary clinical evaluation of a multi-site telemammography system in a screening mammography environment

J. K. Leader, L. P. Wallace, C. M. Hakim, T. M. Hertzberg, L. A. Hardesty, J. H. Sumkin, C. Cohen, C. Sneddon, S. Lindeman, D. Craig, J. M. Drescher

**Reprinted from**

*Medical Imaging 2003*

***PACS and Integrated Medical Information Systems: Design and Evaluation***

SPIE P PROCEEDINGS SERIES

**Volume 5033**

**18–20 February 2003**
**San Diego, California, USA**

# Preliminary clinical evaluation of a multi-site telemammography system in a screening mammography environment

J. Ken Leader[a], Luisa P. Wallace[ab], Christiane M. Hakim[ab], Todd M. Hertzberg[ab], Lara A. Hardesty[ab], Jules H Sumkin[ab], Cathy Cohen[ab], Colleen Sneddon[b], Shirley Lindeman[b], Deborah Craig[c], John M. Drescher[a]
[a]University of Pittsburgh, PA USA 15213
[b]Magee-Womens Hospital, Pittsburgh, PA USA 15213
[c]Lee Regional Hospital, Johnstown, PA USA 15904

## ABSTRACT

We evaluated a telemammography system for reviewing and rating screening mammography in a clinical setting. Three remote sites transmitted 306 exams to a central site. Films were digitized at 50 micron pixel dimensions and compressed at a 50:1 ratio. At the central site images were displayed on a workstation with two high-resolution monitors. Five radiologists reviewed and rated the screens without the availability of prior images or additional information indicating: 1) if additional procedures were needed, 2) which breast was involved, and 3) when appropriate, the recommended additional procedures. During the actual clinical interpretation 13.7% (42 cases) of the patients were recalled for additional procedures. During the retrospective review radiologists 1, 2, 3, 4, and 5 recommended additional procedures for 26.1%, 29.1%, 36.3%, 45.1%, and 54.2% of the cases, respectively. The agreements between the clinical interpretation and radiologists 1, 2, 3, 4, and 5 were 77.8%, 76.1%, 69.0%, 62.7%, and 53.6%, respectively. The exceedingly high percentage of recommended additional procedures using the workstation was attributed to lack of prior images or additional information, the knowledge that case management was not affected, and the observers' expectation for an enriched case mix.

**Keywords:** Teleradiology, human performance, recall rate, breast cancer screening, mammography.

## 1. INTRODUCTION

Teleradiology can challenge typical radiology practices in areas ranging from personnel assignments to data management. In remote or underserved clinics in may be necessary to evaluate personnel qualifications in regards to deciding if teleradiology is appropriate and the necessary radiographic procedures.[1-3] Many teleradiology systems employ image processing techniques to manage the digital image data in terms of data acquistion,[4-9] transmission time (e.g., compression,[4,10,11,12] cropping,[13] image selection[14]), and image display.[7,8,10,11,13,14,15] The effects of data management techniques on diagnostic image quality are application specific. Comparisons between film-based and digitized image-based (film digitization) diagnostic radiographic interpretation have produced mixed results. In some laboratory studies the area under the receiver operating characteristic (ROC) curve, sensitivity, and accuracy have been shown to be slightly greater for film-based interpretation,[4,7,16,17] but the differences were generally not statistically significant. Reported specificity has been relatively equivalent for the two interpretation methods.[4,7,16,17]

The high-spatial resolution necessary to interpret mammographic images presents unique challenges when designing and implementing a telemammography system. Improvements in image quality of x-ray film mammography have been associated with improvements in breast cancer detection.[18-23] Therefore, it is important that the image processing techniques of a telemammography system do not degrade the diagnostic image quality of the digital (full-field digital mammography (FFDM)) or digitized (film digitization) mammographic images.

* jklst3@pitt.edu; phone (412) 641-2572; fax (412) 641-2582, University of Pittsburgh, Magee-Womens Hospital, 300 Halket Street., Suite 4200, Pittsburgh, PA 15213



Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation,
H. K. Huang, Osman M. Ratib, Editors, Proceedings of SPIE Vol. 5033 (2003)

Mammography interpretation has been reported as relatively equivalent for film mammography and digitized mammographic images. Fajardo et al.[24] (1990) found film mammography statistically superior for detecting skin and nipple abnormalities compared to digitized mammography in an ROC study, but found the two methods equivalent for detecting microcalcifications and masses. An ROC study performed by Nab et al.[25] (1992) found that the diagnostic performance of film and digitized mammography were comparable. Powell et al.[26] (1999) reported that film mammography was slightly superior to digitized mammography in several diagnostic measures (i.e., accuracy, false-positive rates, and callback rates for mammograms with normal and malignant findings), but only the callback rates for normal findings were statistically different. The callback rates for benign findings were slightly better for digitized mammography. A follow-up study by Powell et al.[27] (2000) compared film mammography to wavelet-compressed digitized mammographic images. The only statistically significant finding was that the false positive rate was lower for compressed digitized images compared to film mammography. Compressed digitized images were also slightly better (though not statistically) in terms of callback for mammograms with normal and benign findings. Film mammography was slightly better (though not statistically) for callback rates for depicting malignant abnormalities.

This manuscript presents a preliminary, retrospective clinical evaluation of an inexpensive, high-quality, multi-site telemammography system[28,29] for the review of screening mammography examinations. The study was designed to assess the effectiveness of the system for the review of breast cancer screening mammography with the objective to assess its possible use in determining the need for additional procedures (rather than primary diagnosis). The limited retrospective review was conducted using only digitized mammographic images without the benefit of prior images or any additional information. Five radiologists reviewed and rated screening exams using the telemammography system, and their results were compared to the actual clinical interpretations of the same cases regarding the need for additional procedures. It was anticipated that in this experimental protocol the number of cases recommended for additional procedures would be greater during the limited telemammography review compared to the clinical interpretation.

## 2. METHODS

### 2.1 Case selection

The 306 cases retrospectively evaluated in this study originated from patients who underwent breast cancer screening mammography at three woman's imaging centers. The mammography technologists at these centers were instructed to select an approximately equal number of cases they (the technologists) believed may and may not need additional imaging procedures for complete evaluations. Cases were selected by the technologists in a prospective mode and they did not know at the time of selection whether or not the patient would actually be recalled for additional procedures during the clinical interpretation. The mean patient age was 53.8 years ranging from 35 to 88 years old. The actual, subsequent clinical interpretation categorized each case using the Breast Imaging Reporting and Data System (BIRADS) (Table 1). The four routine screening mammographic films of the left and right craniocaudal views (LCC & RCC), and left and right mediolateral oblique views (LMLO & RMLO) were used to review and rate cases in this study.

**Table 1**
**Distribution of BIRADS categories as a result of clinical interpretation of the cases**

| BIRADS Category | 0 | 1 | 2 | total |
|---|---|---|---|---|
| Number of cases | 42 | 206 | 58 | 306 |

### 2.2 Telemammography system

The cases for this study were transmitted from the three centers (remote sites) to Magee-Womens Hospital, Pittsburgh, PA, USA (central site) using an inexpensive, high-quality, multi-site telemammography system. The operation of the system including digitization the mammographic films, digital image processing, data transmission, and image display were conducted under routine operating procedures and are described in detail by Drescher et al.[29] (2003). A brief description, as relevant to this study is provided below.

#### 2.2.1 Central and remotes sites

The central site telemammography workstation is connected to two high-resolution (2048 x 2560) 8-bit grayscale portrait monitors at a nominal setting of 80 ftL (DS5100P, Clinton Electronics, Rockford, IL, USA). A dual 1.2 GHz

multi-processor (Athlon MP, Advanced Micro Device, Sunnyvale CA, USA) with 2 GB of RAM powers the workstation which operating under Microsoft Windows 2000 Server (Microsoft Corporation, Redmond, WA, USA). The workstation is equipped with 56K hardware modems (U.S. Robotics, Rolling Meadows, IL, USA) and an ethernet network cards (OfficeConnect 10/100 NIC, 3COM, Santa Clara, CA, USA) for communication with the remote sites.

The computers at the remote sites operate under Microsoft Windows 2000 Workstation powered by a 900MHz processor (Athlon 900, Advanced Micro Device, Sunnyvale CA, USA) with 512 MB of RAM. The mammographic films are digitized using a high-resolution, laser film digitizers (Lumiscan 85, Eastman Kodak, Rochester, NY, USA) at 50 micron pixel dimensions and 12-bit grayscale. Data communication from the remote site computers is conducted via 56K hardware modems and ethernet network cards (Integrated PRO/100 S Desktop Adapter, Intel Corporation, Santa Clara, CA, USA). Sites 1 and 2 are 15 and 20 miles from the central site, respectively, and transmit data across Plain Old Telephone System (POTS) lines. Site 3 is 90 miles from the central site and transmits data across a Local Area Network (LAN).

### 2.2.2 Image processing

The first image processing step was to perform an automated cropping that removed the non-tissue area surrounding the breast. Next, the image data were compressed using the irreversible (lossy), 9/7 transform, wavelet-based JPEG 2000 method at a 50:1 compression ratio. Prior to transmission from the remote sites, the data packets were encrypted using strong 128 bit Microsoft Point-to-Point Encryption (MPPE) with Microsoft Challenge Handshake Authenticate Protocol (CHAP) version 2.

Upon arrival to the central site the image data were decrypted and decompressed. The decompressed images data were minimally unsharp masked to enhance display on the workstation monitors. The image data range was maximized for display by re-scaling the image data from 0 to 4095. To facilitate image viewing default look-up table (LUT) values were automatically calculated based on the typically bimodal pixel value distribution (histogram). The images were restored to full height, but not the full width, by padding (filling) prior to image display.

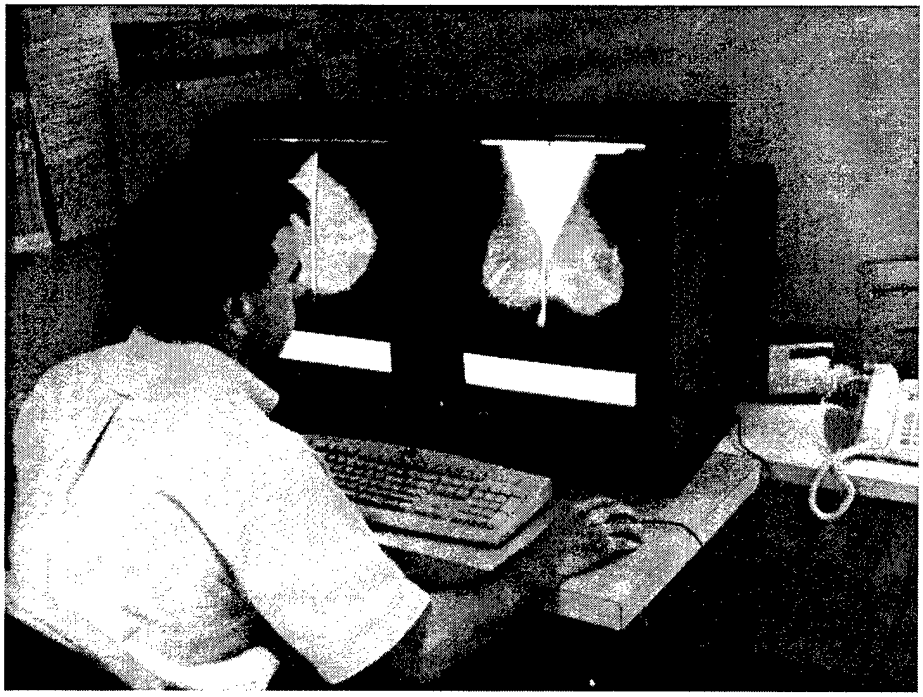

Fig. 1. Telemammography workstation at the central site pictured in the default image display format.

### 2.2.3 Central site image display

There are several mouse-driven image display features on the central site workstation available to the user during case review. Image display formats possible included: one image/monitor, two images/ monitor, or four images/monitor. To duplicate our standard film presentation LCC and RCC are displayed on the left monitor, and LMLO and RMLO on the right monitor as the default presentation (Fig. 1).

The typical display resolution was approximately 100 micron pixel dimensions for one image/monitor and 200 micron pixel dimensions for two images/monitor. Images can be magnified by a free-moving magnification box or quadrant panning. The magnification box size varied dependent on the image display format; for one image/monitor the box was 511 x 566 pixels and for two images/monitor the box was 204 x 266 pixels. The LUT settings could be adjusted by the user by moving the mouse horizontally or vertically. Selected LUT settings could be applied (at user's option) to all images associated with the case and could be reset to the default (automated) values at any point.

### 2.3 Reviewing and rating cases

Five experienced radiologists (each reading over 2000 mammograms per year) reviewed and rated each case on the telemammography workstation. Cases were randomly presented in each session. The rating form for each case was presented on the workstation monitors and completed using the computer mouse (Fig. 2). The computerized scoring form recorded: (1) if additional procedures were indicated, (2) use of prior images (disabled for this study), (3) which breast was involved, and (4) when appropriate, the specific recommended procedure. The radiologists' reviews were conducted based entirely on the four mammographic views (LCC, RCC, LMLO, & RMLO), without additional, potentially relevant information (e.g., prior images, prior reports, patient history). The radiologists were informed of the case origination, but not the case selection criteria. The written instructions to observers regarding case review were:

> In this phase of testing our telemammography system, we would like you to review cases and take a few seconds to *quickly* decide whether or not the case should be recalled for additional procedures. These cases are routine screening mammograms. You will fill out a computer form to indicate if a case should be recalled. If you choose to recall the case you must check off which additional procedures you would recommend for each breast. A "done" button on the bottom of the form will bring up the next case. The computer will automatically track the cases that you have completed and load your remaining cases; the count will be in the bottom of the right screen.

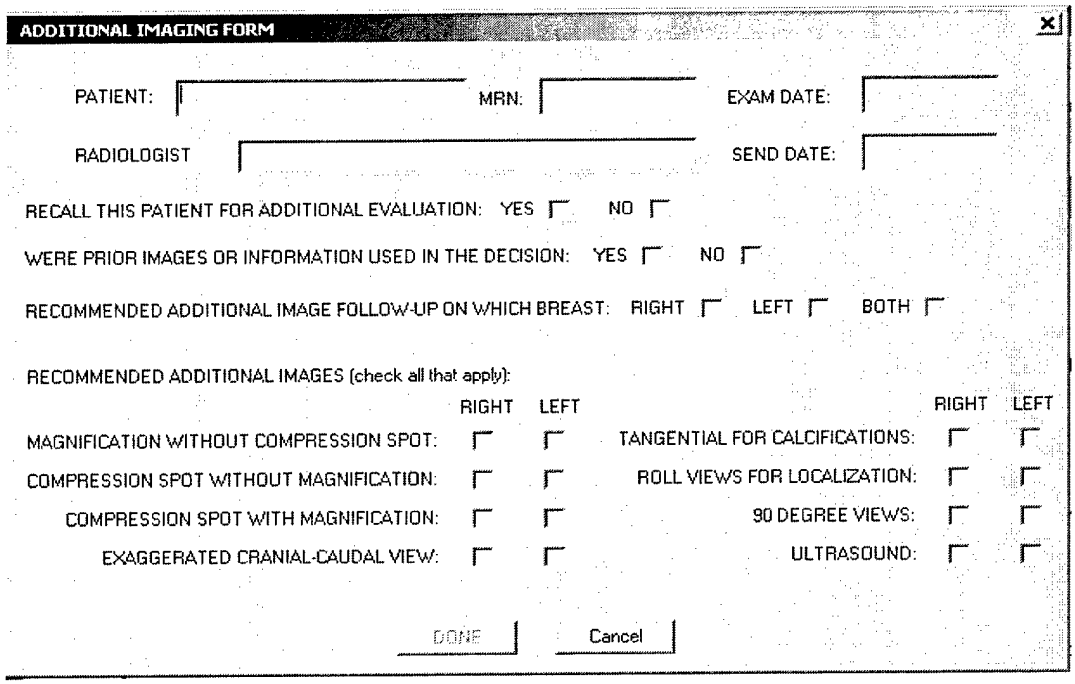


Fig. 2. Computer scoring form complete by the radiologists for each case.

### 2.4 Data analysis

The radiologists' recommendations using the telemammography workstation were compared with the actual clinical interpretation during the original clinical review. The comparisons were done using agreement/disagreement measures. The disagreements when clinical interpretation indicated no-recall and telemammography interpretation indicated recall were further evaluated based on the actual BIRADS ratings during the clinical interpretation.

## 3. RESULTS

Image quality, effects of the image processing, and features of the multi-site telemammography system were subjectively reported as more than adequate for reviewing screening mammography examinations and generally were well-received by the radiologists. The cropped images retained all breast tissue areas and were visibly appealing for image review. The automated LUT settings were normally acceptable and were changed in approximately 10% of the cases during review. Magnification allowed detailed review of the breast tissue patterns, particularly microcalcifications. Although there were some detectable differences at extremely high magnifications between non-compressed and compressed images at a 50:1 compression ratio, the images were subjectively judged to "not affect the diagnostic quality."

The preliminary assessment of the limited case review (i.e., no prior images, prior reports, or patient history) of screening exams using the multi-site telemammography system resulted in an exceedingly high recommended recall rates and modest agreement between the actual clinical interpretation and the radiologists' recommendations using the telemammography system. During the actual clinical interpretation 13.7% (42) of the cases were recalled (BIRADS = 0). Radiologists 1, 2, 3, 4, and 5 recall rates were 26.1% (80), 29.1% (89), 36.3% (111), 45.8% (138), and 54.2% (166), respectively, when using the telemammography system to determine the need for additional procedures (Table 2). The overall agreement between the clinical interpretation and the recommendations of radiologists 1, 2, 3, 4, and 5 were 77.8%, 76.1%, 69.0%, 62.7%, and 53.6%, respectively. Kappa for radiologists 1, 2, 3, 4, and 5 were 0.32, 0.32, 0.22, 0.20, and 0.13, respectively.

**Table 2**
**Reviewing and rating screening mammography exams, telemammography workstation recommendations versus clinical interpretation**

| Telemammography recommendations | Clinical interpretation recall (n = 42) | no-recall (n = 264) | Total |
|---|---|---|---|
| Radiologist 1 | | | |
| recall | 8.8% (27) | 17.3% (53) | 26.1% (80) |
| no-recall | 4.9% (15) | 69.0% (211) | 73.9% (226) |
| Radiologist 2 | | | |
| recall | 9.5% (29) | 19.6% (60) | 29.1% (89) |
| no-recall | 4.2% (13) | 66.7% (204) | 70.9% (217) |
| Radiologist 3 | | | |
| recall | 9.5% (29) | 26.8% (82) | 36.3% (111) |
| no-recall | 4.2% (13) | 59.5% (182) | 63.7% (195) |
| Radiologist 4 | | | |
| recall | 10.8% (33) | 34.3% (105) | 45.1% (138) |
| no-recall | 2.9% (9) | 52.0% (159) | 54.9% (168) |
| Radiologist 5 | | | |
| recall | 10.8% (33) | 43.5% (133) | 54.2% (166) |
| no-recall | 2.9% (9) | 42.8% (131) | 45.8% (140) |

The cases when the recommendation using the telemammography system was "recall" and the clinical interpretation indicated "no-recall" represented a large percentage of the disagreement, and nearly one half had some type of findings reported during the clinical review. The disagreement when the clinical interpretation indicated "no-recall" and the telemammography indicated "recall" accounted for 77.9%, 82.2%, 86.3%, 92.1%, and 93.7% of the total disagreement for radiologists 1, 2, 3, 4, and 5, respectively (Table 2). Further evaluation of these disagreement cases revealed that

cases with a BIRADS category of 2 during the clinical interpretation accounted for 49.1%, 53.3%, 51.2%, 34.3%, and 36.1% of the disagreement cases for radiologists 1, 2, 3, 4, and 5, respectively (Table 3).

**Table 3**
**Disagreement cases when the clinical interpretation was no-recall and the telemammography recommendation was recall for different BIRADS ratings during the clinical interpretation**

| | BIRADS category | |
|---|---|---|
| Disagreement cases | 1 (n = 206) | 2 (n = 58) |
| Radiologist 1 (n = 53) | 50.9% (27) | 49.1% (26) |
| Radiologist 2 (n = 60) | 46.7% (28) | 53.3% (32) |
| Radiologist 3 (n = 82) | 48.8% (40) | 51.2% (42) |
| Radiologist 4 (n = 105) | 65.7% (69) | 34.3% (36) |
| Radiologist 5 (n = 133) | 63.9% (85) | 36.1% (48) |
| Average (n = 86.6) | 55.2% (49.8) | 44.8% (36.8) |

## 4. DISCUSSION

The review of breast cancer screening mammography by five experienced radiologists using the telemammography system demonstrated that the system was adequate for reviewing the mammographic image data. The limited, retrospective review of screens using the telemammography system with only mammographic image data (i.e., no prior images, prior reports, or patient history) produced modest agreement with the actual clinical interpretation. The agreement between the limited telemammography review and clinical interpretation for five radiologists ranged from 53.6% to 77.8% and Kappa ranged from 0.13 to 0.32. On average the radiologists recommended additional procedures using the limited telemammography system in 38.2% of cases which was exceedingly high compared with 13.7 % of patients actually recalled in this group during the clinical interpretation.

The majority of the disagreement between the two review formats occurred when the telemammography review resulted in a recommendation for additional procedures and the clinical interpretation did not, accounting for an average of 86.4% of the disagreement cases for the five radiologists. Of these disagreement cases (clinical no-recall and telemammography recall), on average across the radiologists 44.8% of the patients had a clinical BIRADS category of 2. That is, when findings were detected using the telemammography system under restricted conditions, but the history of the findings (i.e., new, increased, or unchanged) was unavailable, the radiologists tended to recommend additional procedures. Another potential partial explanation for the high recall rate was the radiologists' expectation of an "enriched" sample population because of their knowledge that this is a laboratory study. In addition, the mere fact that patient recall does not affect clinical management tends to produce over reading.

High recall rates were similarly observed by Elmore et al.[30] (1994), where 11-65% of patients without cancer were recommended for immediate workup. In the Elmore study, prior images were not available for any of the cases reviewed and clinical history was not available for every case. They also attributed the high recall rates to the radiologists' knowledge of an "enriched" sample population and study participation.

Although the limited, retrospective review using the telemammography system produced modest agreement with the actual clinical interpretation, the feasibility of the system use for such a review was clearly demonstrated and well-received by the radiologists. Current efforts have begun to add information such as text communication between the technologist (remote site) and radiologist (central site) to the information transmitted with each case.

## ACKNOWLEDGEMENTS

This work is supported in part by the US Army Medical Research Acquisition Center, 820 Chandler Street, Fort Detrick, MD 21702-5014 under contract DAMD17-00-1-0410. The content of the information contained herein does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred.

## REFERENCES


1. Yawn B, Krein S, Christianson J, Hartley D, and Moscovice I. "Rural radiology: who is producing images and who is reading them?" *J Rural Health* **13(2)**:136-144, 1997.
2. Benger JR. "Can nurses working in remote units accurately request and interpret radiographs?" *Emerg Med J* **19(1)**:68-70, 2002.
3. Coughlin SS, Thompson TD, Hall HI, Logan P, and Uhler RJ. "Breast and cervical carcinoma screening practices among women in rural and nonrural areas of the United States, 1998-1999". *Cancer* **94(11)**:2801-2812, 2002.
4. Bolle SR, Sund T, and Stormer J. "Receiver operating characteristic study of image processing for teleradiology and digital workstations." *J Digit Imaging* **10(4)**:152-157, 1997.
5. O'Reilly S, Spedding R, Dearden C. and Loane M. "Can *x* rays be accurately interpreted using a low cost telemedicine system?" *J Accid Emerg Med* **15(5)**:312-314, 1998.
6. Pysher L and Harlow C. "Teleradiology using low-cost consumer-oriented computer hardware and software." *AJR Am J Roentgenol* **172(5)**:1181-1184, 1999.
7. Bancroft LW, Berquist TH, Morin RL, Pietan JH, Knudson JM, and Williams HJ. "Fracture interpretation using electronic presentation: a comparison." *J Digit Imaging* **13(1)**:13-18, 2000.
8. Burgul R, Gilbert FJ, and Undrill PE. "Methods of measurement of image quality in teleultrasound." *Br J Radiol* **73**:1306-1312, 2000.
9. Davidson HC, Johnston DJ, Christian ME, and Harnsberger HR. "Comparison of radiographic image quality from four digitization devices as viewed on computer monitors." *J Digit Imaging* **14(1)**:24-29, 2001.
10. Maitz GS, Chang TS, Sumkin JH, Wintz PW, Johns CM, Ganott M, Holbert BL, Hakim CM, Harris KM, Gur D, and Herron JM. "Preliminary clinical evaluation of a high-resolution telemammography System." *Invest Radiol* **32(4)**:236-240, 1997.
11. Mitra S, Yang S, and Kustov V. "Wavelet-based vector quantization for high-fidelity compression and fast transmission of medical images." *J Digit Imaging* **11(4 Suppl 2)**:24-30, 1998.
12. Kalyanpur A, Neklesa VP, Taylor CR, Daftary AR, and Brink JA. "Evaluation of JPEG and wavelet compression of body CT images for direct digital teleradiology transmission." *Radiology* **217(3)**:772-779, 2000.
13. Lou SL, Lin HD, Lin KP, and Hoogstrate D. "Automatic breast region extraction from digital mammograms for PACS and telemammography applications." *Comput Med Imaging Graph* **24(4)**:205-220, 2000.
14. Ludwig K, Bick U, Oelerich M, Schuierer G, Puskas Z, Nicolas K, Koch A, and Lenzen H. "Is image selection a useful strategy to decrease the transmission time in teleradiology? A study of 100 emergency cranial CTs." *Eur Radiol* **8(9)**:1719-1721, 1998.
15. Lou SL, Sickles EA, Huang HK, Hoogstrate D, Cao F, Wang J, and Jahangiri M. "Full-field direct digital telemammography: Technical components, study protocols, and preliminary results." *IEEE Trans Info Technol Biomed* **1(4)**:270-278, 1997.
16. O'Sullivan DC, Averch TD, Cadeddu JA, Moore RG, Beser N, Breitenbach C, Khazan R, and Kavoussi LR. "Teleradiology in urology: comparison of digital image quality with original radiographic films to detect urinary calculi." *J Urol* **158(6)**:2216-2220, 1997.
17. Youmans DC, Don S, Hildebolt C, Shackelford GD, Luker GD, and McAllister WH. "Skeletal surveys for child abuse: comparison of interpretation using digitized images and screen-film radiographs." *AJR Am J Roentgenol* **171(5)**:1415-1419, 1998.
18. Roberts MM, Alexander FE, Anderson TJ, Chetty U, Donnan PT, Forrest P, Hepburn W, Huggins A, Kirkpatrick AE, Lamb J, Muir BB, and Prescott RJ. "Edinburgh trial of screening for breast cancer: mortality at seven years." *Lancet* **335(8684)**:241-246, 1990.
19. Sickles EA and Kopans DB. "Deficiencies in the analysis of breast cancer screening data." *J Natl Cancer Inst* **85(20)**:1621-1624, 1993.
20. Young KC, Wallis MG, and Ramsdale ML. "Mammographic film density and detection of small breast cancers." *Clin Radiol* **49(7)**:461-465, 1994.
21. Young KC, Wallis MG, Blanks RG, and Moss SM. "Influence of numbers of view and mammographic film density on the detection of invasive cancers: results from the NHS Breast Cancer Screening Programme." *Br J Radiol* **70(833)**:482-488, 1997.
22. Feig SA. "Image quality of screening mammography: effect on clinical outcome." *AJR Am J Roentgenol* **178(4)**:805-807, 2002.

23. Taplin SH, Rutter CM, Finder C, Mandelson MT, Houn F, and White E. "Screening mammography: clinical quality and the risk of interval breast cancer." *AJR Am J Roentgenol* **178(4)**:797-803, 2002.
24. Fajardo LL, Yoshino MT, Seeley GW, Hunt R, Hunter TB, Friedman R, Cardenas D, and Boyle R. "Detection of breast abnormalities on teleradiology transmitted mammograms." *Invest Radiol* **25(10)**:1111-1115, 1990.
25. Nab HW, Karssemeijer N, Van Erning LJ, and Hendriks JH. "Comparison of digital and conventional mammography: a ROC study of 270 mammograms." *Med Inform* **17(2)**:125-131, 1992.
26. Powell KA, Obuchowski NA, Chilcote WA, Barry MM, Ganobcik SN, and Cardenosa G. "Film-screen versus digitized mammography: assessment of clinical equivalence." *AJR Am J Roentgenol* **173(4)**:889-894, 1999.
27. Powell KA, Mallasch PG, Obuchowski NA, Kerczewski RJ, Ganobcik SN, Cardenosa G, and Chilcote W. "Clinical evaluation of wavelet-compressed digitized screen-film mammography." *Acad Radiol* **7(5)**:311-316, 2000.
28. Drescher JM, Maitz GS, Leader JK, Sumkin JH, Poller WR, Klaman H, Zheng B, and Gur D. "Design considerations for a multi-site, POTS-based telemammography system." *Proceedings of SPIE Medical Imaging 2002: PACS and Integrated Medical Information Systems: Design and Evaluation*, 4685:416-421, San Diego, CA, February 2002.
29. Drescher JM, Maitz GS, Traylor C, Leader JK, Clearfield RJ, Shah R, Ganott MA, Pugliese F, Duffner D, Lockhart J, and Gur D. "A multi-site telemammography system: preliminary assessment of technical and operational issues." *Proceedings of SPIE Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation*, In Press, San Diego, CA, February 2003.
30. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. "Variability in radiologists' interpretations of mammograms." *N Engl J Med* **331(22)**:1493-1499, 1994.