An Observer Study for a Computer-Aided Reading Protocol (CARP) in the Screening Environment for Digital Mammography

Paymann Moin, MD, Ruchi Deshpande, MS, Jim Sayre, PhD, Ellen Messer, BS, Sangeeta Gupte, Harlan Romsdahl, Akira Hasegawa, PhD, Brent J. Liu, PhD

Rationale and Objectives: The aims of this study were to investigate improving work flow efficiency by shortening the reading time of digital mammograms using a computer-aided reading protocol (CARP) in the screening environment and to increase detection sensitivity using CARP, compared to the current protocol, commonly referred to as the quadrant view (QV).

Materials and Methods: A total of 200 cases were selected for a receiver-operating characteristic (ROC) study to evaluate two image display work flows, CARP and QV, in the screening environment. A Web-based tool was developed for scoring, reporting, and statistical analysis. Cases were scored for and stratified by difficulty. A total of six radiologists of differing levels of training ranging from dedicated mammographers to senior radiology residents participated. Each was timed while interpreting the 200 cases in groups of 50, first using QV and then, after a washout period, using CARP. The data were analyzed using ROC and κ analysis. Interpretation times were also assessed.

Results: Using QV, readers' average area under the ROC curve was 0.68 (range, 0.54–0.73). Using CARP, readers' average area under the ROC curve was 0.71 (range, 0.66–0.75). There was no statistically significant difference in reader performance using either work flow. However, there was a statistically significant reduction in the average interpretation time of negative cases from 64.7 seconds using QV to 58.8 seconds using CARP.

Conclusions: CARP determines the display order of regions of interest depending on computer-aided detection findings. This is a variation of traditional computer-aided detection for digital mammography that has the potential to reduce interpretation times of studies with negative findings without significantly affecting sensitivity, thus allowing improved work flow efficiency in the screening environment, in which, in most settings, the majority of cases are negative.

Key Words: ROC observer study; mammography; microcalcifications; computer-aided detection; work flow.

Mammography has been and continues to be valuable in the detection of breast cancer. Because the early detection of a potential malignancy could improve the chance of survival, Breast Imaging Reporting and Data System (BI-RADS) categories 4 and 5 usually proceed to biopsy (1–4).

Changes in technology, most notably the advent and rapid adoption of full-field digital mammography, have improved the practice of mammography. However, the move to digital mammography does not come without drawbacks. In the setting of screening mammography interpretation, it has been demonstrated that digital mammograms take more time to read than similar film-screen mammograms (5–7).

A limitation of soft-copy monitors is the maximum image size that can be displayed at full spatial resolution of 5 megapixels, less than the acquired digital spatial resolution as well as traditional screen-film mammography. To search for subtle microcalcifications, a digital mammogram must be displayed in at least four sections to view the entirety of the image at the acquired spatial resolution. A screening mammogram consists of four images, thus accounting for at least 16 sections to be viewed. The evaluation of a case in which comparisons are available only adds to the complexity of interpreting a study. This added time to the interpretation process can significantly slow the work flow of a practice. Given the resolution of the monitor and the limited total monitor display space, this can make the process of reading a digital mammogram inefficient, which could also result in a decrease in performance as defined by both time and sensitivity.

Acad Radiol 2011; 18:1420–1429

From the Image Processing and Informatics Laboratory, Biomedical Engineering Department, Viterbi School of Engineering, University of Southern California, 734 W Adams Blvd, KER-201, Los Angeles, CA 90089-7725 (P.M., R.D., E.M., B.J.L.); the Department of Biostatistics, School of Public Health, University of California, Los Angeles, Los Angeles, California (J.S.); and FUJIFILM Medical Systems, USA, Inc, San Jose, California (S.G., H.R., A.H.). Received April 20, 2011; accepted July 14, 2011. This study was funded by FUJIFILM Medical Systems, USA, Inc. **Address correspondence to:** P.M. e-mail: paymann.moin@gmail.com

doi:10.1016/j.acra.2011.07.003

Most current applications of commercially available computer-aided detection (CAD) in digital mammography display regions of interest (ROIs) for added review by the interpreting radiologist, with the intent of increasing the sensitivity of the detection and the accuracy of the characterization of radiographic abnormalities (8). It has been reported that reading time increases when mammographic CAD is used as a second reader, an intended use during design by vendors and regulatory review and approval by government agencies such as the US Food and Drug Administration (FDA) (9). Such regulatory agencies might be concerned with the potential off-label use of traditional CAD, whereby an interpreting radiologist might review CAD results before or in lieu of interpreting a mammogram, essentially as a "prescreening" tool, thus potentially examining only ROIs prompted by a CAD system (9,10). This use of CAD, to view only regions of an exam marked by CAD, may result in undiagnosed lesions, because not all suspicious lesions will be detected and marked by CAD. Thus, it is of extreme importance for the interpreting radiologist to review an entire exam for suspicious regions, not only the CAD-highlighted regions of interest.

To address these issues without lessening the potential value of CAD as an adjunct diagnostic tool, we propose another use of CAD in mammography, which we have termed the computer-aided reading protocol (CARP). CARP is a reading protocol wherein the reading order of ROIs, depending on CAD findings, is marked by a proprietary algorithm, but the radiologist is forced to review the entirety of a study before the next case can be reviewed. This CARP methodology addresses the concerns raised by the FDA regarding the unintended use of CAD. The CARP method first displays ROIs found by CAD, when the radiologist's attention is the highest (11), and then successively displays each area of the breast. The user is thus forced to see all imaged areas of the breast before the next case can be reviewed.

The purpose of our study was to investigate the possibility of improving work flow efficiency by shortening the reading time of a standard full-field digital mammographic study. A secondary objective was to increase the sensitivity in the detection of microcalcifications by using CARP compared to currently practiced reading work flow using quadrant view (QV).

MATERIALS AND METHODS

Data Set

The case data set was collected from Fujifilm collaboration sites from around the world. The radiologists reading at these sites were board-certified, Mammography Quality Standards Act (MQSA) or equivalent certified mammographers who had interpreted >2000 mammograms in each of the past 3 years. Each case was acquired on a commercial mammographic x-ray machine using a Fujifilm FCRm system (FUJIFILM Medical Systems, USA, Inc, San Jose, CA). Each acquisition center participating in the study maintained FDA MQSA certification or equivalent for each country for each FCRm system used during the image acquisitions. Prior to study initiation, institutional review board approval was obtained for the retrospective use of the data set, and the study protocol was reviewed to ensure compliance with the Health Insurance Portability and Accountability Act.

Study cases consisting of the four standard views were deidentified and had unique identifiers assigned. These images were stored in Digital Imaging and Communications in Medicine format by each acquisition center involved in the study. Fujifilm provided storage media or a dedicated computer for this purpose. Deidentified image data were periodically transmitted to Fujifilm to form the study data set.

A case was defined as a minimum of a two-view unilateral standard mammographic study. Such cases consisted of subjects with prior mastectomy with a return to standard screening mammography. The maximum, and most prevalent, number of mammographic images for each case comprised a four-view bilateral standard mammographic study. The four views consisted of a craniocaudal and mediolateral oblique for each breast.

From the original data set of 550 cases, 205 were reported by the on-site reader as benign, 191 were reported as malignant, and 154 were reported as negative. Nearly all of the cases designated with BI-RADS scores of $\geq$3 were pathology-proven cases. Benign cases were either assigned BI-RADS scores of $\leq$3 or were pathology-proven cases of benign entities for cases with BI-RADS scores of $\geq$3. Negative cases were assigned BI-RADS scores of $\leq$3 by the on-site reader and did not proceed to biopsy. All malignancies were pathology proven.

Exclusion criteria for the use of a case in this data set consisted of patients who, at the time of acquisition, were pregnant or believed that they may have been pregnant, images with inadequate technical quality such as insufficient anatomic coverage or motion artifacts, the presence of a palpable mass only visible on a modality other than mammography or by clinical exam, or images obtained from patients who were incarcerated at the time of the exam. In addition, cases with BI-RADS scores of 0 were excluded from data collection.

All cases from the original 550-case data set were reviewed, again assigned BI-RADS scores, and stratified by difficulty level by a radiologist who was not a member of one of the original reporting sites or a member of the six selected observers for this receiver-operating characteristic (ROC) study. For the purposes of this study, cases with findings were categorized as demonstrating microcalcifications, masses, or both. The BI-RADS definition for each finding was used in the selection of these cases.

The 550 cases were stratified on the basis of the scoring of case difficulty as defined by a point scale in the following categories: breast density, abnormality features, and the presence of associated features. The scores in each category were summed to obtain a total score to grade the difficulty of