Artificial Intelligence Support for Mammography: In-Practice Clinical Experience

Marie Tartar, MD, Lina Le, MD, Alyssa T. Watanabe, MD, Adam Enomoto, BS

BRIEF DESCRIPTION OF THE PROBLEM

Mammographic results can be delayed for many reasons, including physician shortages. In our practice, most delays occur when patients fail to bring their prior outside studies when they present for mammography. In those instances, screening mammographic examinations were held for several weeks until prior studies were received or were ultimately interpreted without comparison to satisfy Mammography Quality Standards Act requirements that reports and lay letters to patients be provided within 30 days of the acquisition of a mammogram [1]. Many women experience anxiety waiting for their mammographic results, with 97% of women in one study reporting that immediate results would lower anxiety [2]. We improved our mammographic interpretation process through use of an artificial intelligence (AI)–based computer-aided detection (CAD) and triage software suite and share here our 2-year experience.

WHAT WAS DONE

An AI-based mammography triage platform (cmTriage; CureMetrix, San Diego, California) and AI-based CAD software (AI-CAD) (cmAssist; CureMetrix) were integrated into the PACS (eRAD, Greenville, South Carolina) at one of our outpatient imaging centers in June 2019, which performed 2-D digital mammography. This site's experience is highlighted because it has had the longest clinical experience with this software, which was deployed subsequently to other sites. All screening mammograms (full-field digital) were evaluated using the AI-CAD and triage software immediately after completion of each examination. All examinations that were triaged as suspicious displayed with a notification on the sortable work list.

IMPLEMENTATION AND CHALLENGES

Radiologists

Challenges included getting buy-in from the radiologists (two with more than 20 years of experience and one with less than 5, all Mammography Quality Standards Act qualified), whose initial reactions included fear of being replaced, skepticism, and resistance to learning new software.

AI-PACS Integration

Installation of AI-CAD and the triage platform with the PACS was unproblematic. Both AI software algorithms were implemented using a cloud-based software-as-a-service solution (Fig. 1) rather than an on-premises installation. No hardware installation or equipment calibration was needed on site. All personal health information stays on site via the use of local software that anonymizes images before sending them to the cloud-based AI software for analysis. This PACS supports review of multiple CAD vendor results. The AI-CAD was run in parallel with the existing traditional CAD (ImageChecker version 10.0; Hologic, Sunnyvale, California) so that radiologists could gain familiarity with the AI-CAD in a real-time "head-to-head" comparison. The radiologists were trained on how to choose which CAD they wanted to review and how to view the AI-CAD quantitative neuScore of flagged lesions. The neuScore ranges from 0 to 100, with 100 being the highest suspicion level. The cmTriage case-based notification allows a simple sorting of examinations with suspicious findings in the "Urgency" column of the study browser work list. The radiologist work lists could be easily rearranged to prioritize examinations triaged as suspicious to the top of the work list.

To assess the impact of AI-PACS integration, we measured workflow modification and flag reduction with AI-CAD.

Workflow Modification. Before the installation of the triage platform, screening patients new to this center with known prior outside studies had their mammograms withheld from immediate review by the radiologists. Because mammography was a relatively new modality at this location, a substantial number of examinations were held awaiting prior studies. With the use of triage , examinations with suspicious findings were expedited for

1546-1440/21/$36.00 ■ https://doi.org/10.1016/j.jacr.2021.09.016

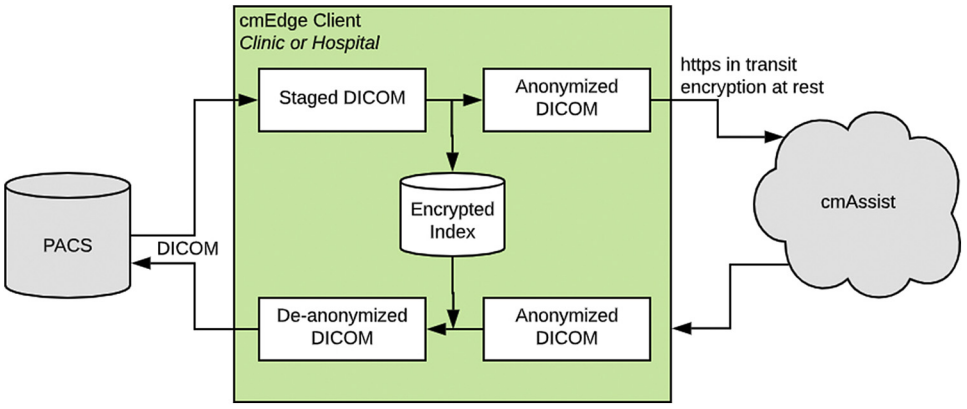

Fig. 1. Artificial intelligence (AI) support cloud-based pathway. Personal health information never leaves the facility. All images are anonymized using an edge client, and the AI results are then matched to the examinations in the PACS within minutes.

review. To assess the impact of AI on workflow, we compared average turnaround time (TAT) from examination completion to final reporting in samples performed before and after AI installation (from January 22 to February 5, 2019, and from July 23 to August 3, 2021).

Flag Reduction With AI-CAD. To compare the AI-CAD with the traditional CAD, we assessed the frequency of flags per examination using AI-CAD versus traditional CAD on 150 full-field digital screening examinations obtained with a Hologic Selenia unit from December 13, 2019, to February 3,2020.

OUTCOMES AND LIMITATIONS

Radiologists

Despite initial skepticism, a verbal survey of the interpreting radiologists performed 2 years after implementation showed universal preference for the AI-CAD compared with traditional CAD, the value of which has been questioned [3]. Furthermore, the use of triage is now seen as the preferred way to manage their work lists. A subjective benefit noted by readers after the implementation of cmTriage was a perception of greater ease of reading batched screening mammograms, particularly those that were prescreened as not suspicious and flag free, resulting in elimination of the extra step of checking for CAD flags.

Workflow Modification

Before AI implementation, average TAT was 9.6 days, on the basis of the 2019 sampling. There was significant reduction in TAT to 3.9 days on the basis of the 2021 sampling ($P < .05$ using two-sample Wilcoxon's rank-sum test) [4]. To further assess the impact of AI on workflow, TAT was tracked on samples of BI-RADS® category 0 patients. In 2019, for 26 BI-RADS category 0 cases, the average TAT was 9.4 days (range, 1-33 days). We tracked 29 BI-RADS category 0 examinations performed in 2021, with average TAT reduction to 4.7 days (range, 0-22 days) ($P < .05$).

Flag Reduction

The 150-mammogram sample comprised women ranging in age from 42 to 84 years (mean, 62 years). Ninety-three percent of the patients were Asian and the rest were Hispanic, with breast tissue density breakdown as follows: 3% fatty, 29% scattered fibroglandular density, 63% heterogeneously dense, and 6% extremely dense.

We found a 71% reduction in flags per examination using the AI-CAD versus the traditional CAD (Table 1), with 2.26 traditional CAD flags per examination versus 0.65 AI-CAD flags per examination (95% confidence interval [CI], 1.19-2.04). Flag reduction was comparable and significant for both masses (72% reduction; 95% CI, [0.80-1.32) and calcifications (70% reduction; 95% CI, 0.27-0.84). Significance was measured using Wilcoxon's rank-sum

Table 1. Flag reduction with AI-CAD versus CAD in 150 examinations

Lesion Type	CAD Flags	AI-CAD Flags	Flag Reduction With AI-CAD
Calcifications	119	36	70%
Density	220	61	72%
Total	339	97	71%

Note: There were 71% fewer flags per examination with AI-CAD versus CAD. The reduction was consistent with both calcifications and masses. AI = artificial intelligence; CAD = computer-aided detection.

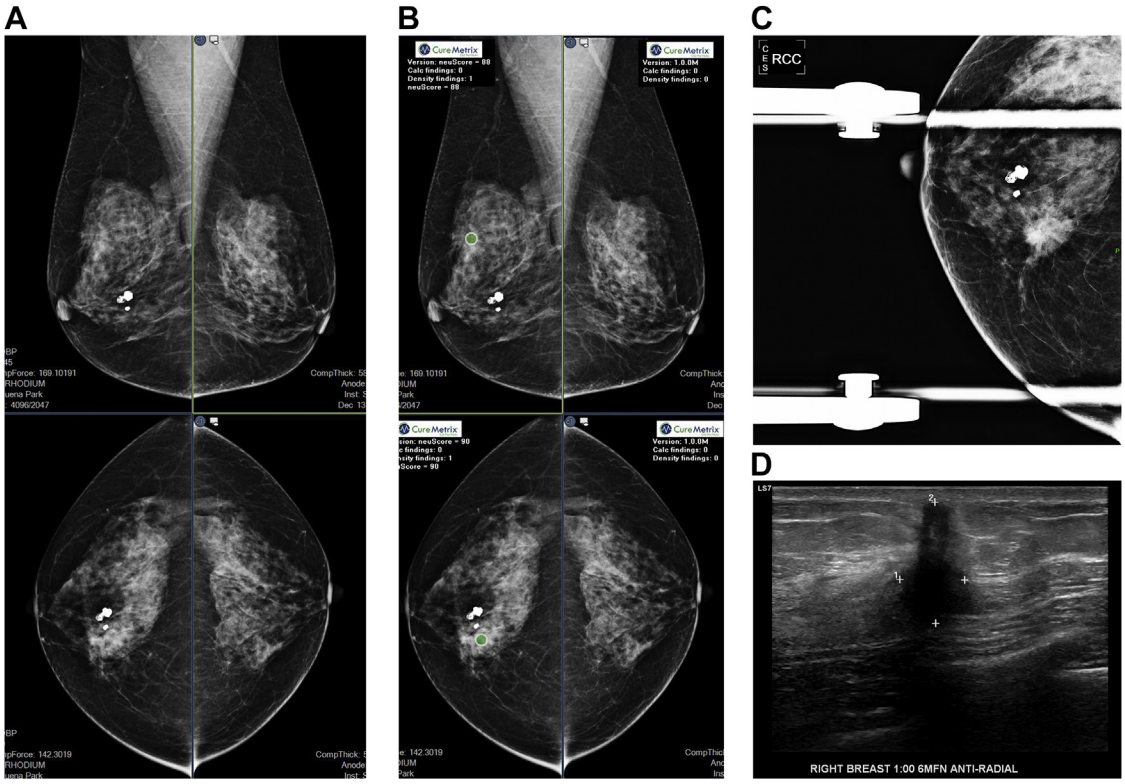

Fig. 2. Biopsy-confirmed cancer triaged as suspicious. (a) Mediolateral oblique (MLO) and craniocaudal (CC) screening digital mammograms show heterogeneously dense breast parenchyma. On the right, a partially obscured asymmetric density was seen in the upper inner quadrant. (b) CureMetrix cmAssist artificial intelligence–based computer-aided detection software applied. The lesion is flagged in both projections and has a high neuScore (88 on the MLO image and 90 on the CC image). Note that there are no false flags. (c) Spot compression confirms a mass with architectural distortion. (D) Ultrasound shows a hypoechoic, shadowing, irregular mass confirmed to be an infiltrating ductal carcinoma.

test, and all P values were $<.05$. In addition, 62% of examinations were flag free with AI-CAD, whereas only 26% of traditional CAD examinations were flag free. This difference was also significant, as measured by the Pearson χ^2 test ($P < .001$) [5]. Because these data are based on recent experience, a complete tracking of examination outcomes could not be performed. On the basis of available follow-up data to date, there were no examinations with false-negative findings.

An example of an examination flagged as suspicious by cmTriage at the case level is presented in Figure 2. The mammogram shows an asymmetric density, which was flagged by cmAssist AI-CAD in both projections with high neuScores at the lesion level and also triaged as suspicious at the case level. The patient was recalled and underwent diagnostic evaluation, with confirmation of infiltrating ductal carcinoma at biopsy.

In summary, triage of screening mammograms resulted in significant improvement in reporting of recalled patients, thereby expediting workup. Examinations with more suspicious findings were interpreted within 24 hours, with fewer examinations held for outside comparisons. There were significantly fewer flags with the AI-CAD compared with the traditional CAD. Subjectively, the radiologist experience over 2 years was improved by having fewer, more meaningful flags to evaluate and the perceived benefit of a sorted screening mammography work list.

Limitations

This case study is based on our early and initial use of an AI-CAD and triage platform at a single center. Because this clinical AI implementation is recent, we do not have long-term follow-up in many of these patients beyond 1 year. Large prospective and long-term studies will be needed to demonstrate how AI support tools will generalize in other clinical scenarios.

ACKNOWLEDGMENT

We thank Huye Zhou for assistance with statistical analysis.

REFERENCES

1. US Food and Drug Administration. The Mammography Quality Standards Act final regulations: modifications and additions to policy guidance help system #13. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/mammography-quality-standards-act-final-regulations-modifications-and-additions-policy-guidance-help-0. Accessed October 3, 2021.
2. Lee J, Hardest LA, Kunzler NM, et al. Direct interactive public education by breast radiologists about screening mammography: impact on anxiety and empowerment. J Am Coll Radiol 2016;13:12-20.
3. Lehman CD, Wellman RD, Buist DS, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 2015;175:1828-37.
4. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 1947;18:50-60.
5. Cochran WG. The χ^2 test of goodness of fit. Ann Math Stat 1952;23:315-45.

Marie Tartar, MD, is from Scripps Clinic Medical Group-Scripps Green Hospital, La Jolla, California. Lina Le, MD, is from Golden West Radiology Medical Group, Anaheim, California. Alyssa T. Watanabe, MD, is from the University of Southern California Keck School of Medicine, Los Angeles, California. Adam Enomoto, BS, is from the University of California, Los Angeles, Medical Center, Los Angeles, California.

Dr Tartar is vice president of medical affairs for CureMetrix. Dr Le is a shareholder in CureMetrix. Dr Watanabe is chief medical officer of CureMetrix. Mr Enomoto states that he has no conflict of interest related to the material discussed in this article. Drs XXX are XXX.

Marie Tartar, MD: Scripps Clinic Medical Group-Scripps Green Hospital, 10666 N Torrey Pines Road, La Jolla, CA 92037; e-mail: tartar.marie@scrippshealth.org.