Journal of Breast Imaging, 2020, Vol. 2, Issue 1, 22–28
doi:10.1093/jbi/wbz075
Original Research
Received: August 28, 2019; Editorial Acceptance: October 18, 2019
Published Online: December 21, 2019

Original Research

Impact of Colorized Display of Mammograms on Lesion Detection

Emily E. Knippa, MD*, Erica Berg, MD, Samuel Richard, PhD, Yuan Lin, PhD, Kingshuk Roy Choudhury, PhD, Ehsan Samei, PhD, Jay A. Baker, MD

University of Texas Southwestern Medical Center, Department of Radiology, Dallas, TX (E.E.K.); Wisconsin Radiology Specialists, Milwaukee, WI (E.B.); Carestream Health, Rochester, NY (S.R.); OPPO US Research Center, AI Department, Palo Alto, CA (Y.L.); Duke University, Department of Biostatistics and Bioinformatics, Durham, NC (K.R.C.); Duke University, Departments of Radiology, Physics, and Biomedical Engineering, and Electrical and Computer Engineering, Durham, NC (E.S.); Duke University Hospital, Department of Radiology, Durham, NC (J.A.B.)

*Address correspondence to E.E.K. (e-mail: emily.knippa@utsouthwestern.edu)

Abstract

Objective: To assess the effect of the colorized display of digital mammograms on observer detection of subtle breast lesions.

Methods: Three separate observer studies compared detection performance using grayscale versus color display of 1) low-contrast mass-like objects in a standardized mammography phantom; 2) simulated microcalcifications in a background of normal breast parenchyma; and 3) standard-of-care clinical digital mammograms with subtle calcifications and masses. Colorization of the images was done by displaying each image pixel in blue, green, and red hues, or gray, maintaining DICOM–calibrated luminance scale and consistent luminance range. For the simulated calcifications and clinical mammogram studies, comparison of detection rates was computed using McNemar's test for paired differences.

Results: For the phantom study, mass-like object detection was significantly better using a green colormap than grayscale (73.3% vs 70.8%, $P = .009$), with no significant improvement using blue or red colormaps (72.6% and 72.5%, respectively). For simulated microcalcifications, no significant difference was noted in detection using the green colormap, as compared with grayscale. For clinical digital screening mammograms, no significant difference was noted between gray and green colormaps for detection of microcalcifications. Green color display, however, resulted in decreased sensitivity for detection of subtle masses (63% vs 69%, $P = .03$).

Conclusion: Although modest improvement was demonstrated for a detection task using colorized display of a standard mammography phantom, no significant improvement was demonstrated using a color display for a simulated clinical detection task, and actual clinical performance was worse for colorized display of mammograms in comparison to standard grayscale display.

Key words: mammography; phantom; color display.

Introduction

Conventional two-dimensional mammography has traditionally been viewed in grayscale. This was originally because of the limitations of film-screen technology and later because of the lack of diagnostic-grade, high-luminance color liquid crystal display (LCD) monitors. Such LCD monitors are now available, and studies have demonstrated comparable performance of monochrome cathode rate tube (CRT) and color LCD displays for the detection

Key Messages

- Colorized display of a standard mammography phantom resulted in modest improvement in lesion detection compared with grayscale display.
- No significant improvement in lesion detection was demonstrated using a color display of simulated microcalcifications.
- Actual clinical performance was worse for color than standard grayscale display of mammograms, resulting in a statistically significant *decrease* in sensitivity for detecting masses.

of mammographic lesions (1). What remains to be investigated is whether postprocessing techniques such as the addition of color can improve the detection of subtle mammographic lesions.

Traditional grayscale display conveys a natural sense of order (from bright to dark); however, a grayscale display has a limited perceived dynamic range. Using a grayscale display, the human eye can distinguish only 60–90 "just noticeable differences" in the state of fixed adaptation when the human visual system is adapted to a wide range of luminance values (2). This limitation of perception may limit use of information that is otherwise available in the image. Tailoring the display to maximize the "just noticeable differences" the eye can detect may increase the sensitivity of lesion detection.

Color perception by the human retina is the result of the interaction of three types of cones whose maximal sensitivities are to blue, green, and red wavelengths of light. The maximal sensitivity of the fovea, the portion of the retina responsible for fine vision, is at the wavelength of 555 nm, within the green spectral range of light (3). When using polychromatic (white) light displays (ie, traditional grayscale), a large portion of the photons are not efficiently detected by the observer's eyes. Displaying an image at wavelengths of light to which the eye is most sensitive (ie, green, 555 nm) may result in improved photon-detection efficiency and reduced perceptual noise. This, in turn, may result in improved detection of subtle lesions.

Digital mammography appears to be an ideal modality to test this hypothesis. Screening mammography is a challenging task, with up to a 50% or 60% false negative rate, depending on tissue density (4, 5). Clinical outcomes may therefore be improved by increasing the sensitivity of breast imaging through technical advances such as computer-aided detection (6–9), digital breast tomosynthesis (10, 11), and, potentially, colorized display of the images.

In breast imaging, little has been published previously to evaluate the use of color in display and interpretation of mammograms. Almost three decades ago, color overlays were described in dual-energy mammography as a technique for displaying a colorized "calcium image" superimposed on the conventional mammogram image (12); however, a review of the literature yields no published studies regarding color display of standard digital mammograms. The purpose of this study was to evaluate the use of colorized display of phantom, simulated clinical, and actual clinical mammographic images to determine whether color display can improve the detectability of subtle breast lesions.

Methods

Three observer studies were conducted using 1) a mammography phantom, 2) simulated microcalcifications, and 3) clinical mammograms with subtle lesions. This study was approved by the institutional review board, who granted a waiver of consent. The study was compliant with the Health Insurance Portability and Accountability Act.

Contrast Detail Mammography Phantom (CDMAM) Study

A digital image was obtained of a contrast detail mammography phantom (CDMAM, Artinis CDMAM 3.4) using a clinical digital mammography system (Essential, GE, Milwaukee, WI) and standard mammographic technique. The CDMAM phantom is comprised of a matrix of cells arranged in 16 rows by 16 columns containing gold discs of varying thicknesses and diameters. Within a row, the disc diameter is constant, with logarithmically increasing thickness; within a column the thickness of the discs is constant with logarithmically increasing diameter. Each cell contains two identical discs (same thickness and diameter), one in the center and one randomly placed in one of the four corners. To enhance the efficiency of the study and to maximize the ability to detect small differences between the colormaps, only selected low-conspicuity cells around the barely perceptible threshold (ie, cells within the central portion of the CDMAM phantom (Figure 1) were used. Cells in which the target lesion was obvious or imperceptible were excluded to decrease the number of cases shown to each observer.

The mass-like low-conspicuity cells from the CDMAM phantom were displayed individually in four colormaps—gray, green, blue, and red. Equal luminance scale (in cd/m^2) and DICOM calibration as would be measured by a luminance meter was maintained between the four colormaps (13). The observer detection task was to determine the corner of the cell containing the low-conspicuity disc (ie, a four-alternative forced-choice task (14, 15).

This phantom study consisted of 18 observers including three fellowship-trained breast imaging radiologists with 4–17 years of experience, one breast imaging fellow, 12 diagnostic radiology residents, and two medical physicists (PhDs). All observers, for this and the three subsequent studies described below, were administered the Ishihara Color Vision Test before the study, and none were colorblind (16). All images were displayed on a high-resolution, high-luminance three-megapixel color liquid crystal display (LCD) monitor (Eizo, Radiforce RX320, St Louis, MO). The room light was set to the same levels for all readers and all colormaps.

Each observer reviewed 200 training cases comprised of 50 cases for each of the four colormaps. Feedback was provided indicating a correct or incorrect response. Subsequently, each observer reviewed

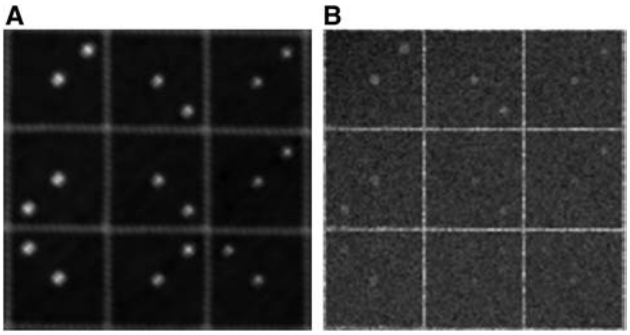

Figure 1. Contrast detail mammography (CDMAM) phantom. **A:** Photograph of nine selected low-conspicuity cells from the CDMAM phantom used for the observer task. **B:** Radiograph of the CDMAM phantom obtained with standard mammographic technique. In the center of each cell is the central reference object, indicating the object to detect in one of the four corners.

a different set of 520 test cases. For the test cases, 26 detection tasks (low-conspicuity cells from the CDMAM phantom) were randomly displayed for each of the four colormaps (Figure 2). Each image was shown five times in random orientation, yielding a total of 520 cases per observer. Customized software was used to display images and record observer response.

Statistical Analysis

A pairwise comparison was performed of the pooled data for all 18 observers. A regression model based on a generalized estimating-equations approach was also used to account for correlation of multiple measurements made on the same case. Based on results from the contrast phantom study (detailed in the Results section), the follow-up studies below were conducted with grayscale and green colormaps only.

Simulated Microcalcifications Study

The four-alternative forced-choice methodology was also used for the simulated microcalcifications observer study. Using a previously described technique (17), simulated microcalcifications were digitally embedded in square regions of normal tissue obtained from mammographic images from normal subjects. For the simulated microcalcification study, four squares containing background breast parenchyma were displayed in gray and green colormaps, maintaining equal luminance. Simulated microcalcifications were embedded in one of the four squares (Figure 3), and the observer detection task was to select which of the four squares contained the simulated microcalcifications within the center of the square (Figure 4).

Each observer reviewed 30 training cases in each colormap (60 cases total) with immediate feedback indicating correct or incorrect responses. Ten observers (five fellowship-trained breast imaging radiologists with 4–18 years of experience, one breast imaging fellow, and four radiology residents) were shown a total of 117 cases per colormap, for a total of 234 cases per observer.

Statistical Analysis

Comparison of detection rates for grayscale and green colormaps and the significance of differences for each observer was computed using McNemar's test for paired differences. A comparison of pooled observers was also performed using McNemar's test for paired differences. Furthermore, a logistic regression model was applied for the outcome using the color scheme and observer as predictors.

Clinical Mammogram Study

We selected 120 single-view mammogram images (craniocaudal (CC) or mediolateral oblique (MLO) images): 40 from patients with negative mammograms; 40 from patients with subtle masses, asymmetries, or architectural distortion, which were all subsequently biopsy-proven invasive cancers; and 40 from patient with subtle microcalcifications, which were all biopsy-proven atypical ductal hyperplasia, DCIS, or invasive carcinoma. Single-view mammogram images were selected

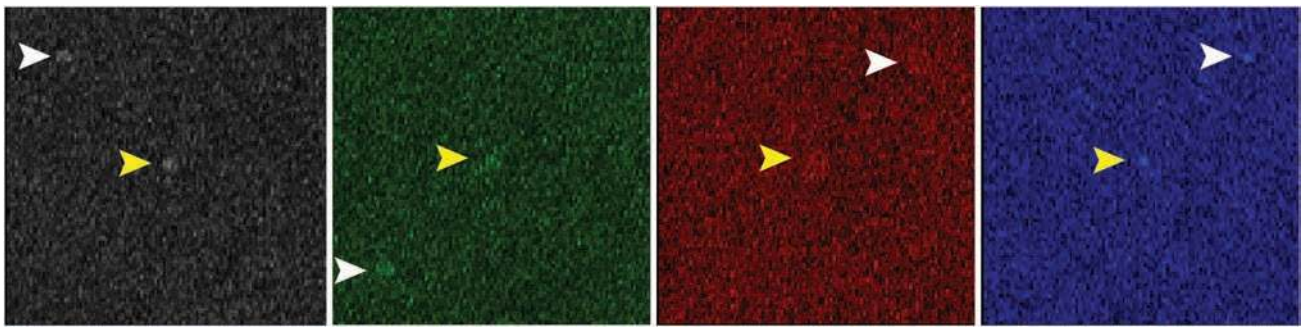

Figure 2. Contrast detail mammography (CDMAM) observer task. Representative examples of four different phantom images that were individually displayed to observers in each of the four colormaps (gray, green, red, and blue). In the center is the central reference object, indicating the object to detect (yellow arrowheads). In 1 of 4 corners is the second object to detect, identical to the central reference object (white arrowheads).

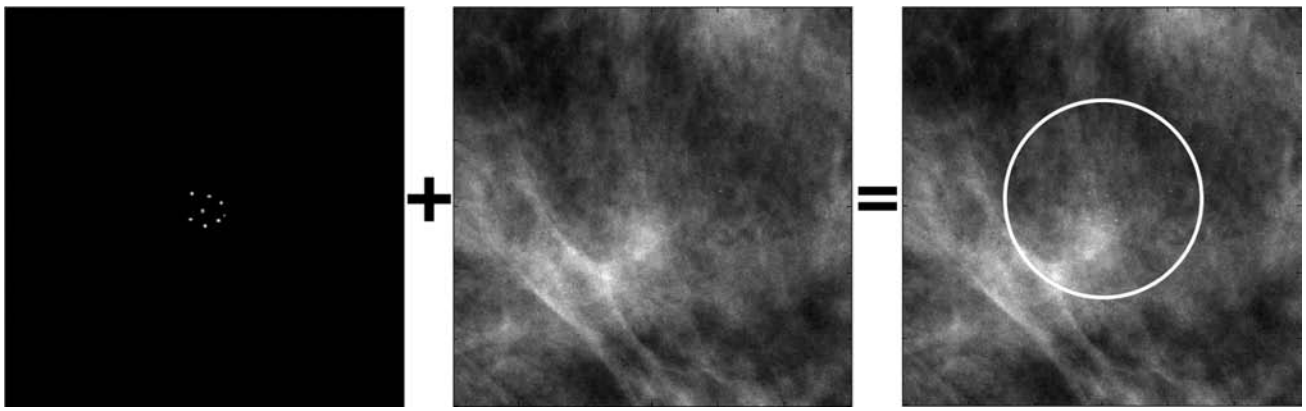

Figure 3. Simulated microcalcifications. Simulated clusters of microcalcifications were embedded within regions of interest obtained from normal mammograms, yielding a fused image with subtle microcalcifications in the center of the image (circle).

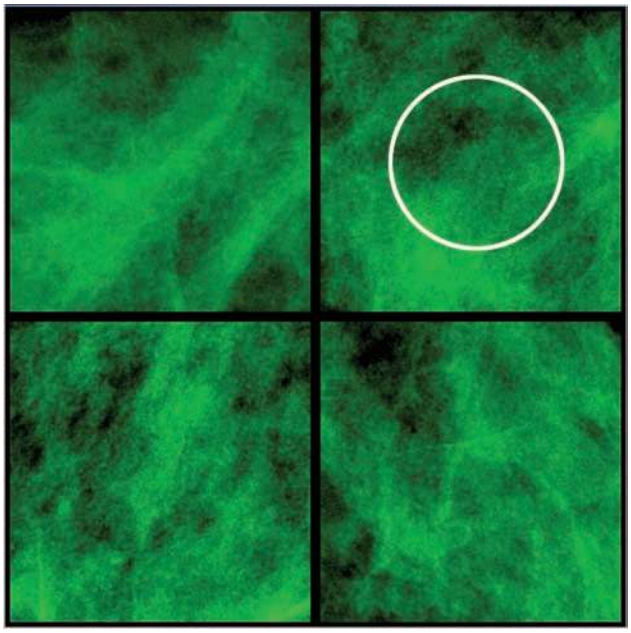

Figure 4. Simulated microcalcifications observer task. Using the four-alternative forced- choice methodology, four regions of interest were displayed at a time. The observer task was to select which of the four squares contains the microcalcifications in the center (circle).

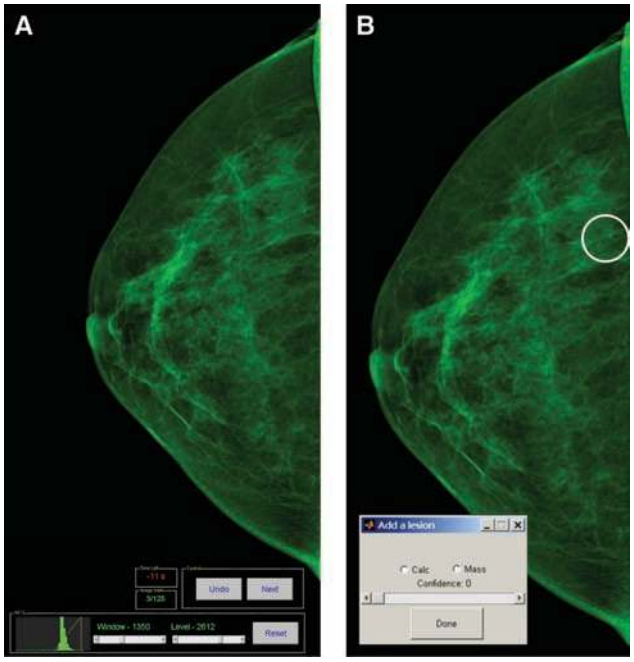

Figure 5. Clinical mammogram images. **A:** Graphical user interface demonstrating the display of a single-view mammogram in the green colormap. Observers were able to window and/or level the image, reset the default window or level setting, select a lesion, undo a lesion selection, and advance to the next case. A counter indicated time spent per case, although there was no time limit imposed. **B:** Graphical user interface for lesion selection. The observer clicked on the mammogram image to generate a round region of interest (circle) and indicate whether the lesion was a microcalcification or mass, as well as a confidence level.

to simplify lesion localization and scoring. All findings were visible on the single view. Cases were chosen retrospectively, and the study readers were not the original readers of the case, to avoid bias.

A single radiologist who had access to all relevant clinical imaging, surgical notes, and pathology reports (ie, the study radiologist) placed a region-of-interest (ROI key) around the known region of atypia or malignancy. Subsequently, seven observers—each a fellowship-trained breast imaging radiologist with 4–18 years of experience—reviewed each case. Customized software was used to display images and record observer responses (Figure 5). For each case, observers indicated the location of a suspected breast lesion, and this location was recorded and compared with the location of the ROI key. Each case had only one suspicious lesion, and if the observer marked a lesion within the ROI key, the case was scored as a true positive. Any selections outside the ROI key were scored as false positives.

Each observer reviewed the same collection of cases in the two colormaps—gray and green—in two sessions, separated by a 3-week washout period. Four observers reviewed cases in grayscale first, and the remaining three observers initially reviewed cases in the green colormap.

Statistical Analysis

Comparison of detection rates for gray- and green-colormap displays and the significance of differences for each observer was computed using McNemar's test for paired differences. Sensitivity and specificity for detection of microcalcifications and masses for both gray and green colormaps were also calculated. A secondary endpoint, time per case, was also recorded. A paired t-test for differences and a distribution-free Wilcoxon test was applied to the time data.

Results

Contrast Detail Mammography Phantom (CDMAM) Study

The CDMAM phantom study assessed for the detection of select low-conspicuity mass-like objects on a standard mammography phantom. A pairwise comparison of the pooled data for all 18 observers demonstrated a statistically significant improvement in sensitivity (lesion-detection rate) for the green colormap compared to gray (73.3% vs 70.8%, P = .009) (Table 1). Observers' sensitivity using blue (72.6%, P = .091) and red (72.5%, P = .141) colormaps trended toward improvement, but these differences were not statistically significant.

Although only the green colormap demonstrated statistically significantly improved sensitivity over grayscale display for the pooled results of all 18 observers, individual observers performed best with several different colormaps. Six individual observers demonstrated better sensitivity using a green colormap; six using a red colormap; and four using a blue colormap. Two observers performed equally well with green or blue. None of the 18 observers performed best with the grayscale display.

Simulated Microcalcifications Study

In this four-alternative forced-choice study, barely perceptible simulated microcalcifications were embedded into a background of breast parenchyma taken from normal mammograms. No significant difference was noted in detection rate for microcalcifications for pooled observers between green and gray colormaps (Table 2). For pooled results of all 10 observers, sensitivity was nearly equal, at 65.7% for gray- and 65% for green-colormap display (McNemar's test P-value = .62). Only one observer demonstrated a statistically significant difference in detection rate with improved detection in grayscale compared to the green colormap (McNemar's test P-value = .025).

Clinical Mammogram Study

There was no statistically significant difference in the detection of microcalcifications when clinical mammogram images were displayed using a green colormap compared with a standard grayscale

Table 1. Pairwise Comparison of Mammography Phantom Task in Blue, Green, Red, and Gray Colormaps

Color	Correct Mean			
Blue	0.7256			
Gray	0.7081			
Green	0.7329			
Red	0.7248			
Comparison	**Difference**	**Mean Lower CL**	**Mean Upper CL**	***P*-value**
Blue vs gray	0.0175	-0.0028	0.0378	0.0910
Blue vs green	-0.0073	-0.0260	0.0115	0.4480
Blue vs red	0.0009	-0.0169	0.0186	0.9247
Gray vs green	-0.0248	-0.0434	-0.0062	0.0091
Gray vs red	-0.0167	-0.0389	0.0055	0.1412
Green vs red	0.0081	-0.0095	0.0258	0.3667

Table 2. Pairwise Comparison of Detection of Simulated Microcalcifications in Gray versus Green Colormaps. Comparison of detected microcalcifications versus undetected microcalcifications for pooled observers and for the single observer (observer 9) with a statistically different detection rate (gray > green, McNemar's test *P*-value = .025)

Pooled Observers[a]

		Green	
		Undetected	**Detected**
Gray	Undetected	272	129
	Detected	138	631

Observer 9[b]

		Green	
		Undetected	**Detected**
Gray	Undetected	26	6
	Detected	18	67

[a]McNemar's test *P*-value = **.62**.
[b]McNemar's test *P*-value = **.025**.

display (sensitivity 73% vs 71%, McNemar's test *P*-value = .71). In contrast, use of a green-colormap display led to a statistically significant *decrease* in sensitivity for detecting masses compared with use of a standard grayscale display (63% vs 69%, McNemar's test *P*-value = .03) (Table 3).

Overall, observers spent significantly less time looking at images displayed in the green colormap than in grayscale, with a median difference of 2.12 seconds less time spent on green images (*P*-value < .0001) (Figure 6). The greatest difference in response time was for cases with a suspicious mass (median difference 3.52 seconds less, *P*-value < .0001) (Table 4). The median difference in time spent on microcalcification cases was 1.9 seconds less with the green colormap than the gray (25.9 s vs 28.5 s, *P* = .001).

Discussion

While medical imaging has traditionally been displayed in grayscale, the availability of high-resolution color displays has resulted in increasing utilization of color. The use of color in medical imaging has primarily focused on "pseudocolor" display rather than displaying the intrinsic color of the object being imaged. Pseudocolor is the application of color-coded scalar imaging data to highlight quantitative data (18, 19). For example, in nuclear medicine, color displays are used to display cardiac perfusion data. In ultrasound, color overlays are used to display Doppler data. In magnetic resonance imaging and computed tomography, color overlays are used to display perfusion data.

What remains to be investigated is whether color display of the images, rather than color overlays of quantitative data, can improve lesion detectability. Because the maximal spectral response of the human retina is at a wavelength of 555 nm, a green spectral range, we hypothesized that colorized display of mammography images in green would result in increased photon-detection efficiency by the retina, and an increased lesion-detection rate.

Few studies have evaluated the use of color in display and interpretation of mammograms. Color has been used as an adjunct, or overlay, such as using dual-energy mammography to displaying a colorized "calcium image" superimposed on the conventional mammogram image (12). The aim of this study was to determine if color display of a mammography phantom, anatomic model, and clinical mammogram images would result in an increased lesion-detection rate.

The CDMAM phantom-detection task assesses an observer's ability to detect objects with very low contrast and small diameter within a digital display system, similar to a difficult-to-detect breast mass. Detection in this model is limited by the background quantum noise of the mammography system. The initial CDMAM phantom study demonstrated increased sensitivity of mass-like-object detection with a green display compared to gray display, with lesser improvements in sensitivity for the blue and red colormap display.

This increased sensitivity noted in the simple phantom-detection task, however, did not carry through to testing using an anatomic model with simulated microcalcifications. Instead, sensitivity of calcification detection was found to be nearly identical for displays using grayscale and green colormaps. Limitations of this anatomic model study include the relatively small number of cases shown to each observer, which may limit confirmation of very small differences between the two displays, as well as the limited range of the luminance values available in the color monitor.

Finally, the observer study of subtle calcifications and masses in actual clinical mammograms confirmed the results of the anatomic model with no improvement in sensitivity. In fact, although sensitivity for detection of microcalcifications was nearly equal for the green colormap and grayscale, the detection of masses was statistically significantly worse for images displayed in the green colormap

Table 3. Sensitivity for Detection of Microcalcifications and Masses in the Clinical Mammogram Study. No statistically significant difference was seen in sensitivity for microcalcifications between the gray and green colormap (71% versus 73%, McNemar *P*-value .71). A statistically significant decrease in sensitivity for masses in the green colormap versus gray was seen (63% versus 69%, McNemar *P*-value .03)

	Sensitivity-Gray	Sensitivity-Green	McNemar *P*-value
Microcalcifications	71% [66%–77%]	73% [67%–78%]	.71
Masses	69% [63%–74%]	63% [57%–68%]	.03

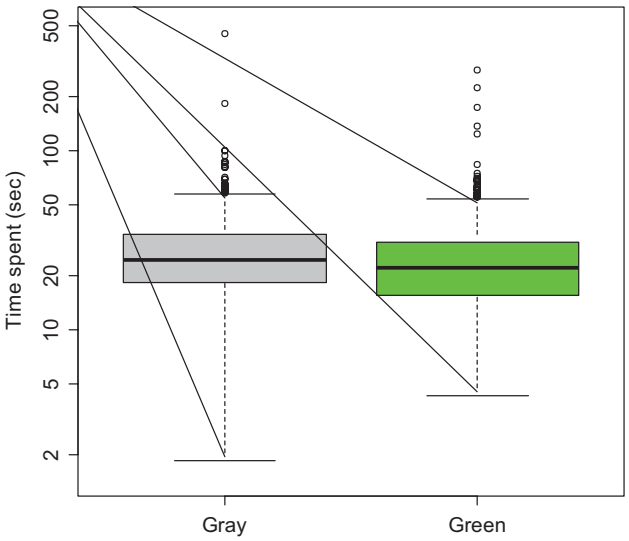

Figure 6. Comparison of time spent by case, aggregated across readers. Observers spent less time evaluating green mammogram images compared to gray. The distribution of time spent was heavily skewed, so a log transformation was applied for display. A median difference of 2.12 seconds less time was spent on green images, and a paired t-test for differences was highly statistically significant (*P*-value < .00001). A distribution-free Wilcoxon test confirms this result (*P*-value < .0001).

Table 4. Comparison of Time Spent by Case, Aggregated Across Readers. Observers spent 3.52 seconds less analyzing green cases with masses, compared with 1.91 seconds less on cases with calcifications and 1.94 seconds less on normal cases. This difference for mass cases was highly significant (*P*-value < .0001).

	t-test *P*-value	Wilcox *P*-value	Median Difference
Calcifications	.001	.01	1.91
Masses	<.0001	<.0001	3.52
Normal	.01	.07	1.94

compared with grayscale. Potential explanations include the observers' lack of familiarity with the green colormap. Observers spent less time analyzing the green mammograms compared with the gray mammograms, and this time difference, though small, was statistically significant. Given that there was no statistically significant difference in detection rate for calcifications, this suggests that observers are able to detect microcalcifications displayed in green slightly faster with an equal sensitivity. This improved efficiency of calcification detection, however, is at the expense of a lower sensitivity for masses.

A limitation of the clinical mammogram study is the use of single-view mammograms, whereas two-view mammograms are interpreted in clinical practice. While all findings were visible on the single-view mammograms displayed, lack of a second view may have reduced diagnostic confidence, particularly for masses, which by definition must be seen in two views at mammography. Several observers commented that viewing the green images was unpleasant, causing eyestrain. This perceived unpleasantness and the unfamiliarity of viewing green images may have resulted in the shorter observation time. This decrease in observation time may account for the decreased sensitivity noted for masses with the green colormap. Additional training with green-colormap images may have improved observer performance.

Dark adaptation in the eye is less affected by red light (at lower luminance) than blue or green light. This could contribute to the "unpleasantness" of the green colormap reported by observers. Furthermore, in the CDMAM phantom study, images were displayed in a single colormap at a time, with random rotation of colormaps between each case (eg, green, red, blue, gray, blue, green, gray, etc.). Random alternation of colormaps may have made it difficult for the eyes to adapt, which may have had an effect on lesion detection for some colormaps. We further acknowledge that the application of different color scales can potentially influence the luminance and structured noise of the display, degrading the quality of nonmonochromatic renditions. Although this effect is potentially very small, our study did not explicitly control for it, and the potential influence cannot be ruled out.

In summary, this study of colorized display of mammograms demonstrates that in a quantum-noise-limited phantom task, the green colormap can offer improved detection. However, that advantage is negated when images contain anatomical variability that dwarfs the limiting influence of quantum noise and the added advantage of improved retinal sensitivity. In that scenario, sensitivity with the grayscale for detection of microcalcifications was equivalent to that with a green colormap. The decreased sensitivity with the green colormap for the detection of masses seen in the clinical mammogram study suggests that a green display would be inferior to grayscale in a clinical setting. Future studies could investigate the effect of color displays after greater training, using a larger study size to detect smaller differences in detection rates, and across a range of display luminance values. Additionally, colorized displays could be assessed as an adjunct to grayscale display, similar to computer-aided detection.

Conflict of interest statement

Samuel Richard is currently employed by Carestream Health. Ehsan Samei discloses relationships with the following entities unrelated to the present publication: GE, Siemens, Bracco, Imalogix, 12Sigma, SunNuclear, Metis Health Analytics, Cambridge University Press, and Wiley and Sons. The remaining authors declare no conflicts of interest.

References

1. Samei E, Poolla A, Ulissey MJ, Lewin JM. Digital mammography: comparative performance of color LCD and monochrome CRT displays. *Acad Radiol* 2007;14(5):539–546.

2. Levkowitz H, Herman GT. Color scales for image data. *IEEE Comput Graph* 1992;12(1):72–80.
3. Vos JJ. Colorimetric and photometric properties of a 2-deg fundamental observer. *Color Res Appl* 1978;3:125–128.
4. Birdwell RL, Ikeda DM, O'Shaughnessy KF, Sickles EA. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology* 2001;219(1):192–202.
5. Pisano ED, Gatsonis C, Hendrick E, et al; Digital Mammographic Imaging Screening Trial (DMIST) Investigators Group. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 2005;353(17):1773–1783.
6. Cole EB, Zhang Z, Marques HS, Edward Hendrick R, Yaffe MJ, Pisano ED. Impact of computer-aided detection systems on radiologist accuracy with digital mammography. *AJR Am J Roentgenol* 2014;203(4):909–916.
7. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007;356(14):1399–1409.
8. Gilbert FJ, Astley SM, Gillan MG, et al.; CADET II Group. Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008;359(16):1675–1684.
9. James JJ, Gilbert FJ, Wallis MG, et al. Mammographic features of breast cancers at single reading with computer-aided detection and at double reading in a large multicenter prospective trial of computer-aided detection: CADET II. *Radiology* 2010;256(2):379–386.
10. Rafferty EA, Park JM, Philpotts LE, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology* 2013;266(1):104–113.
11. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;311(24):2499–2507.
12. Boone JM. Color mammography. Image generation and receiver operating characteristic evaluation. *Invest Radiol* 1991;26(6):521–527.
13. Samei E, Badano A, Chakraborty D, et al.; AAPM TG18. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. *Med Phys* 2005;32(4):1205–1225.
14. Jäkel F, Wichmann FA. Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers. *J Vis* 2006;6(11):1307–1322.
15. Eckstein MP, Whiting JS. Why do anatomic backgrounds reduce lesion detectability? *Invest Radiol* 1998;33(4):203–208.
16. Ishihara. *Ishihara's Tests for Color Deficiency 2010. 14 Plate Book Concise Edition*. 1st ed. Tokyo: Graham-Field; 2010.
17. Saunders R, Samei E, Baker J, Delong D. Simulation of mammographic lesions. *Acad Radiol* 2006;13(7):860–870.
18. Zabala-Travers S, Choi M, Cheng WC, Badano A. Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images. *Med Phys* 2015;42(6):2942–2954.
19. Li H, Burgess AE. Evaluation of signal detection performance with pseudocolour display and lumpy backgrounds. In: Kundel HL, ed. *Image Perception: Medical Imaging*. Vol 3036. Bellingham: SPIE-Int Soc Optical Engineering, 1997.