

Comparing Performance of the CADstream and the DynaCAD Breast MRI CAD Systems

Joann Pan · Basak E. Dogan · Selin Carkaci ·
Lumarie Santiago · Elsa Arribas · Scott B. Cantor ·
Wei Wei · R. Jason Stafford · Gary J. Whitman

Published online: 16 April 2013

© Society for Imaging Informatics in Medicine 2013

Abstract Computer-aided diagnosis (CAD) systems are software programs that use algorithms to find patterns associated with breast cancer on breast magnetic resonance imaging (MRI). The most commonly used CAD systems in the USA are CADstream (CS) (Merge Healthcare Inc., Chicago, IL) and DynaCAD for Breast (DC) (Invivo, Gainesville, FL). Our primary objective in this study was to compare the CS and DC breast MRI CAD systems for diagnostic accuracy and postprocessed image quality. Our secondary objective was to compare the evaluation times of radiologists using each system. Three radiologists evaluated 30 biopsy-proven malignant lesions and 29 benign lesions on CS and DC and rated the lesions' malignancy status using the

Breast Imaging Reporting and Data System. Image quality was ranked on a 0–5 scale, and mean reading times were also recorded. CS detected 70 % of the malignant and 32 % of the benign lesions while DC detected 81 % of the malignant lesions and 34 % of the benign lesions. Analysis of the area under the receiver operating characteristic curve revealed that the difference in diagnostic performance was not statistically significant. On image quality scores, CS had significantly higher volume rendering (VR) ($p < 0.0001$) and motion correction (MC) scores ($p < 0.0001$). There were no statistically significant differences in the remaining image quality scores. Differences in evaluation times between DC and CS were also not statistically significant. We conclude that both CS and DC perform similarly in aiding detection of breast cancer on MRI. MRI CAD selection will likely be based on other factors, such as user interface and image quality preferences, including MC and VR.

Keywords Breast · Breast diseases · Computer-assisted detection · Computer-aided diagnosis (CAD) · Diagnostic imaging · MR imaging

J. Pan

Department of Biochemistry and Cell Biology, Rice University,
P.O. Box 1892, MS-140,
Houston, TX 77251-1892, USA
e-mail: jlp7@rice.edu

B. E. Dogan (✉) · L. Santiago · E. Arribas · G. J. Whitman
Department of Diagnostic Radiology, The University of Texas MD
Anderson Cancer Center, 1515 Holcombe Blvd.,
Houston, TX 77030, USA
e-mail: basak.dogan@mdanderson.org

S. Carkaci
Department of Diagnostic Radiology, Ohio State University
Wexner Medical Center, Breast Imaging, 395 W. 12th Ave.,
Columbus, OH 43210, USA

S. B. Cantor · W. Wei
Department of Biostatistics, The University of Texas MD
Anderson Cancer Center, 1400 Pressler St.,
Houston, TX 77030, USA

R. J. Stafford
Department of Imaging Physics, The University of Texas MD
Anderson Cancer Center, 1400 Pressler St.,
Houston, TX 77030, USA

Introduction

Magnetic resonance imaging (MRI), a powerful tool for the identification and staging of primary breast cancer, performs especially well when used for specific indications such as screening for breast cancer in a high-risk population, staging newly diagnosed breast cancer and concurrently screening for contralateral breast cancer, monitoring response to neoadjuvant chemotherapy, and evaluating patients with axillary metastases from an unknown primary tumor [1]. Advances in dynamic contrast-enhanced breast MRI over the past decade have led to significant improvements in the detection of early breast cancer and to higher accuracy in the

detection of multifocal and multicentric disease, and these improvements have resulted in the increased application of breast MRI in the diagnosis and management of breast cancer globally [2–7].

Despite its increased use worldwide in breast cancer, MRI is not without some limitations. MRI is often compared to other techniques for detecting breast cancer, such as mammography and breast ultrasound. Although a highly sensitive technique for detecting invasive breast cancer (94–99 %) [8], MRI is generally accepted as being less sensitive for detecting in situ breast cancer (50–92 %) [8, 9]. Furthermore, MRI has low and variable specificity (37–97 %) [10–15], which hinders its use as a routine imaging technique in breast cancer patients. Benefits of breast MRI include better cancer detection rates in high-risk women and providing more information regarding the extent of disease in women with known breast cancer.

To improve the accuracy of breast MRI for detecting breast cancer, radiologists use computer-aided diagnosis (CAD) systems, software programs that use algorithms to find patterns associated with breast cancer. Used after the radiologist's initial reading of the MRI study, CAD systems do not diminish the role of the radiologist. Instead, they improve the accuracy of radiologists' diagnoses by indicating signs of possible breast cancer that may be missed on the initial reading or by emphasizing benign features that might appear suspicious for malignancy, especially benign areas of enhancement [16].

With many CAD systems now available commercially, it is important to evaluate which system can best help detect signs of breast cancer on breast MRI. The most commonly used CAD systems in the USA are CADstream (CS) (Merge Healthcare Inc., Chicago, IL) and DynaCAD for Breast (DC) (Invivo, Gainesville, FL). Our primary objective in this study was to compare the CS and DC breast MRI CAD systems for diagnostic accuracy and postprocessed image quality. Our secondary objective was to compare the interpretation times of radiologists using each system, as speed of interpretation may also affect the selection of an MRI CAD system.

Materials and Methods

A retrospective search of our institutions breast MRI database was performed by two radiologists with 10 and 7 years of experience in reading breast MRI evaluated 177 lesions in 175 consecutive patients who underwent second-look ultrasound guided biopsy or MRI-guided biopsy. Both radiologists reviewed cases independently and reached a consensus on evaluation of lesion. All MRI studies had been performed with the patients lying prone in a 1.5-T scanner (Signa EXCITE, General Electric) and using a dedicated

breast array coil (MRI Devices Corporation, Pewaukee, WI). The breast MRI protocol consisted of a unilateral T1-weighted and a T2-weighted fat-suppressed sagittal sequence followed by a single precontrast and five serial dynamic VIBRANT sagittal image sets obtained before and immediately after rapid intravenous bolus infusion of 0.1 mmol/kg of a gadopentetate dimeglumine contrast medium (Magnevist; Bayer Healthcare Pharmaceuticals, Inc., Wayne, NJ) at a rate of 3 ml/s with a power injector (Spectris Solaris MR Injector; MEDRAD, Warrendale, PA). Delayed postcontrast three-dimensional fast spoiled gradient echo images with fat suppression in axial plane had also been obtained.

For use in our study, 30 malignant lesions and 29 benign lesions were classified as “difficult to categorize” and had to have histopathologic verification of benign or malignant status. Excluded from this study were high-risk breast lesions (i.e., atypical ductal hyperplasia, atypical lobular hyperplasia, and lobular carcinoma in situ), as MRI features and management options for these lesions are not well defined. After studying the images using the CS and then the DC systems, three reading radiologists with 6, 5, and 4 years of experience reading breast MRI studies rated each lesion according to the Breast Imaging Reporting and Data System (BI-RADS) criteria using only BI-RADS categories 2, 4, and 5 (Table 1) [17].

Radiologists were aware of which breast and of the location of the lesion within the breast being assessed but were blinded to all other patient data. Histopathologic results that were obtained following MRI-guided or second-look ultrasound-guided biopsy and then confirmed with follow-up studies after a mean of 34 months (range, 24–50 months) were considered the gold standard of diagnosis. Malignant lesions were either diagnosed at percutaneous biopsy ($n=27$) at excision ($n=3$). CAD MRI versions used were DynaCAD 2.1 and CADstream 4.1 that included maximum intensity projection, reformat, motion correction, color parametric map, time–intensity curve, and volume rendering features. For this study, the quality of the images

Table 1 BI-RADS ranking

0	Additional imaging is needed
1	No MRI abnormalities
2	Benign
3	Probably benign; short-term follow-up is needed to confirm stability (probability of malignancy estimated to be ≤ 2 %)
4	Lesion is suspicious and should be biopsied
5	Lesion is highly suspicious (malignant until proven otherwise)
6	Lesion is confirmed to be malignant

was evaluated on a scale of 0–5, with higher scores indicating better quality (Table 2).

Readers were asked to rate the following qualitative image quality features of each CAD system: motion correction, maximum intensity projection, three-dimensional reformatting, volume rendering, time–intensity curve, and color parametric map. Ratings of each system were performed at least 6 weeks apart (range, 6–18 weeks). Overall evaluation times were also recorded. The evaluation time is the amount of time for the reading radiologist to interpret and rate the lesion using BI-RADS categories.

Summary statistics in the form of frequency tables and percentages were provided for diagnostic scores for the CS and the DC CAD systems. Image quality scores, processing time, and the differences in attribute scores between CS and DC were summarized using mean, standard deviation, and range. A signed-rank test was used to assess whether any image quality scores or interpretation times differed from zero between the CS and the DC systems. Only *p* values less than 0.05 were considered statistically significant for image quality comparisons. A Bonferroni correction was used to control the type I error rate across all comparisons of image quality scores. A logistic regression model with the generalized estimating equation method was used to estimate and compare the diagnostic performances of CS and DC. The generalized estimating equation model was fit with a random effects model to account for correlations between scores from the same lesion. A mean diagnostic score (across all radiologists) was calculated for each patient by each software system, and receiver operating characteristic (ROC) analysis was carried out using the mean diagnostic scores. The areas under the ROC curves for the CS and the DC systems were estimated and compared. Kappa statistics were calculated to evaluate agreement between pairs of raters. All tests were two-sided, and *p* values of 0.05 or less were considered statistically significant for comparisons of ROC curves and interpretation times. Statistical analyses were carried out using SAS version 9 (SAS Institute Inc., Cary, NC) and S-PLUS version 7 (TIBCO Software Inc., Seattle, WA). Time–intensity kinetic assessments of all readers were similar in CS and DC. Therefore,

Table 2 Image quality ranking

0	Not used/not necessary
1	Not helpful
2	Poor
3	Good
4	Very good, but not used in the clinical assessment/report
5	Excellent, high impact on ultimate BI-RADS impression, should be used on follow-up

kinetic assessment results reflect assessments by both softwares.

The research procedures were conducted with approval from our hospital’s Institutional Review Board and in accordance with the Helsinki Declaration of 1975, as revised in 2000. A waiver of informed consent was obtained from the Institutional Review Board since this study did not involve any therapeutic or diagnostic interventions to the patients.

Results

The mean patient age was 53.8 years (range, 23–77 years). In agreement with the standard of care in our institution, the lesion size was obtained by measuring the longest diameter of the lesion. The mean lesion size was 2.73 cm (malignant, 2.52 cm; benign, 3.15 cm; range, 0.4–10 cm). Table 3 contains distribution of lesion MRI characteristics and percentages of lesions characterized by morphology and time–intensity kinetic parameters.

Comparison of Diagnostic Accuracy

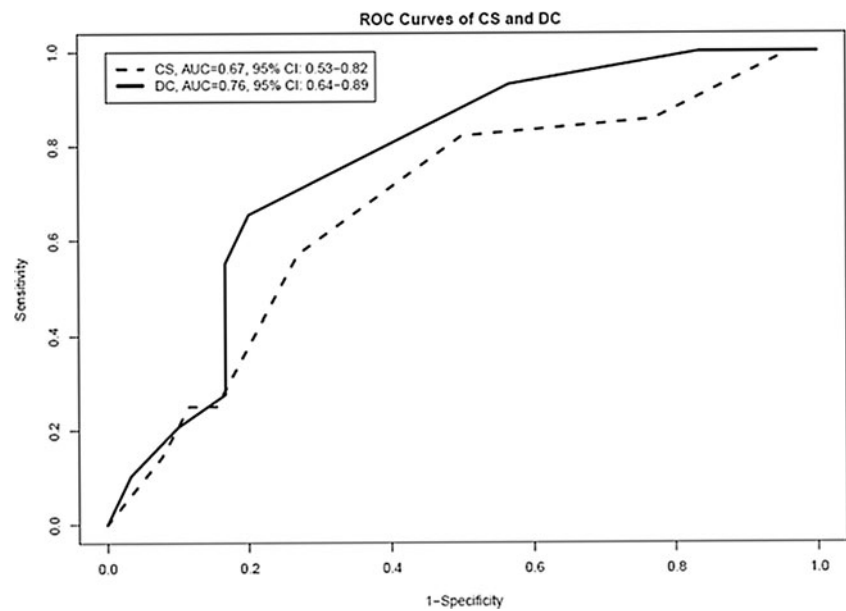
Using CS, readers rated a mean of 70 % (range, 63–80 %) of malignant lesions and 63 % (range, 48–72 %) of benign lesions as “suspicious” (BI-RADS 4 or 5). They rated 32 % (range, 24–45 %) of benign lesions and 17 % (range, 7–23 %) of malignant lesions as “benign” (BI-RADS 2). Using DC, readers rated a mean of 81 % (range, 70–93 %) of malignant lesions and 63 % (range, 45–76 %) of benign lesions as “suspicious” (BI-RADS 4 or 5) and 34 % (range, 21–52 %) of benign lesions and 13 % (range, 7–17 %) of malignant lesions as “benign” (BI-RADS 2).

Agreement between pairs of raters was fair (kappa range, 0.15–0.46). Logistic regression did not reveal significant differences in diagnostic accuracy between the two CAD systems (*p*=0.16) or between raters (*p*=0.62). Lesions with

Table 3 Distribution of lesion MRI characteristics

	N (%)
Morphology	
Mass	20 (34)
Nonmass	39 (66)
Total	59 (100)
Time–intensity	
Washout	21 (35.5)
Plateau	22 (37)
Continuous	16 (27)
Total	59 (100)

Fig. 1 ROC curves of the mean scores for the CADstream and the DynaCAD for Breast systems. The figure shows the ROC curves based on the two mean diagnostic scores of the three raters for each software system. *CS* CADstream, *DC* DynaCAD for Breast, *AUC* area under the curve, *ROC* receiver operating characteristic



higher diagnostic scores were significantly more likely to be malignant ($p=0.009$).

Figure 1 shows the ROC curves based on the two mean diagnostic scores of the three raters for each software system. The area under the ROC curve \pm standard error was 0.67 ± 0.07 (95 % confidence interval, 0.53–0.82) with CS and 0.76 ± 0.06 (95 % confidence interval, 0.64–0.89) with DC; these two areas did not significantly differ ($p=0.18$) (Table 4).

Comparison of Image Quality Scores

Table 5 compares the image quality scores and rankings between CS and DC. CS yielded significantly higher scores in motion correction and volume rendering compared to DC ($p < 0.0001$ by signed-rank test for both). Scores for the other image quality features did not differ significantly between the two systems.

Comparison of Evaluation Times

The mean image evaluation times were 12.0 min (range, 6–20 min) with DC and 12.7 min (range, 5–45 min) with CS. A Wilcoxon rank-sum test did not reveal a

significant difference in evaluation times between DC and CS.

Discussion

We did not find any significant differences between the diagnostic performances of CS and DC. The systems had similar sensitivity and specificity (CS had 70 % sensitivity and 32 % specificity whereas DC had 81 % sensitivity and 34 % specificity). Both CS and DC had a high sensitivity for detecting malignant lesions on breast MRI; however, neither system significantly improved specificity for the diagnosis of benign lesions.

In order to utilize a CAD system, a radiologist must evaluate many parameters that have proven to be effective in detecting malignancy. For example, the washout sign on breast MRI CAD parametric color overlay images is one of the parameters that radiologists use to predict malignancy [18]. In a 2011 study, two additional CAD-related variables, the predominant enhancement type and the most suspicious enhancement type, were compared with the central washout sign for association with malignancy and benignity. All three variables can be determined only with the help of a CAD system on MR images. Eighty-seven percent of malignant lesions and 11 % of benign lesions were associated with a central washout sign [16]. Previous CAD MRI studies showed that a CAD-detected washout pattern was a reliable indicator of malignancy [18]. In our study, our predominant objective was to evaluate the contribution of CAD to morphometric evaluation of breast lesions, and hence, only 35.5 % of the lesions evaluated had a washout pattern.

Table 4 Comparison of performance of the CADstream and the DynaCAD for breast MRI CAD systems in 59 patients

	Area under the curve	Standard error	95 % confidence interval
CS mean score	0.67	0.07	0.53–0.82
DC mean score	0.76	0.06	0.64–0.89

Table 5 Comparison of image quality scores and processing times between the CADstream and the DynaCAD for breast systems

Score categories	Number of observations	Mean	Std	Min	Max	<i>p</i> value
Maximum intensity projection score						
CS	163	3.05	0.58	1	4	0.58
DC	172	2.97	0.94	0	4	
Reformat score						
CS	163	2.38	1.08	0	4	0.05
DC	171	2.26	1.16	0	4	
Motion correction score						
CS	163	2.26	1.26	0	4	<0.0001
DC	172	1.29	1.19	0	3	
Color parametric map score						
CS	163	3.14	0.55	1	4	0.24
DC	172	3.08	0.5	0	4	
Time–intensity curve score						
CS	163	3.09	0.51	0	4	0.59
DC	172	3.06	0.47	0	4	
Volume rendering score						
CS	163	2.94	1.32	0	5	<0.0001
DC	172	1.28	1.54	0	5	
Overall score						
CS	165	3.16	0.74	1	5	0.08
DC	172	3.04	0.31	2	4	
Reader interpretation time						
CS	163	12.69	4.55	5	45	0.11
DC	172	11.98	2.98	6	20	

Std standard deviation, *Min* minimum, *Max* maximum, *CS* CADstream, *DC* DynaCAD for Breast

The maximum number of observations is 177. *p* values are based on signed-rank test. Using the Bonferroni correction for multiple comparisons (six comparisons for image quality), the threshold for significance at a 5 % level is 0.05/9=0.006

The reliability of certain CAD systems also has been studied. In a 2006 study, the accuracy of diagnoses made using CS was tested by comparing the probability of malignancy at three thresholds based on the presence of software-defined significant enhancement [19]. CS helped analyze whether significant enhancement was present and classified enhancements by percentage volumes of washout, plateau, and persistent enhancement. CS was able to accurately show significant enhancement for all malignant lesions at all thresholds and show nonsignificant enhancement for 12/24 benign lesions [19]. In other words, 50 % of benign lesions were ranked as suspicious lesions using the CAD thresholds. Our findings agree with prior results as CS did not improve specificity in our study, and the MRI readers in our study rated 63 % of the benign lesions as suspicious (BI-RADS 4 or 5) using CS.

Our results agree with those of a similar study that compared three CAD systems [20]. The data sets of breast MRI from 21 patients were retrospectively evaluated by three investigators with all three software systems with the use of BI-RADS classifications. Twenty-five image quality criteria were also assessed and ranked on a scale of 0–5. This study concluded that all three software programs showed adequate sensitivity of breast MRI and that there was no difference in evaluating MRI data sets. There was

also no difference in image quality criteria between the three systems.

One limitation of our study could be the low specificity of both CAD systems. A possible explanation for this finding is that the cases we used were preselected and categorized as difficult to diagnose by experienced radiologists and that knowledge may have influenced radiologists to rank the chosen lesions higher on the BI-RADS. In other words, lesions are more likely to be ranked higher, because the lesions are harder to categorize as malignant or benign and radiologist have a greater chance of interpreting them as malignant. Another limitation could be that the order depending on which of the two CAD systems the three reading radiologists used to read the breast MRI cases and rate the lesions first could have affected their interpretations of the lesions.

Strengths of this study include obtaining lesions from hard to diagnose cases, which excluded those lesions showing obvious signs of malignancy or benignity. This helped create a “real-life” environment where CAD systems have a true contribution to practicing radiologists in their decision making. Another strength of our study is the fact that all biopsy-proven lesions had a minimum of 24 months of MRI follow-up to confirm imaging–pathology correlation. To our knowledge, this is the first study comparing CS and DC,

which are the two CAD systems used on a large scale in the USA.

From our findings, it can be inferred that interpretation times for CS and DC did not significantly differ. The mean reading times for CS and DC were similar, yet a reader's experience could play an important role in reading time. For this study, all three readers were trained to read CS, whereas they had all been using DC in breast MRI reading for the past 4 years.

Conclusions

The two most commonly used breast MRI CAD systems in the USA, CS and DC, did not differ significantly in sensitivity or specificity. Even though both CS and DC did demonstrate adequate sensitivity, they did not significantly improve specificity. CS and DC were similar in diagnostic performance. Therefore, we hypothesize that breast MRI CAD system selection could be driven by user interface factors, user preferences, and user experience.

References

- Loo CE, Straver ME, Rodenhuis S, Muller S, Wesseling J, Peeters M, Gilhuijs K: Magnetic resonance imaging response monitoring of breast cancer during neoadjuvant chemotherapy: relevance of breast cancer subtype. *J Clin Oncol* 29:660–666, 2011
- Dorrius MD, Jansen-van der Weide MC, van Ooijen PM, et al: Computer-aided detection in breast MRI: a systematic review and meta-analysis. *Eur Radiol* 21(8):1600–1608, 2011
- Demartini WB, Lehman CD, Peacock S, et al: Computer-aided detection applied to breast MRI: assessment of CAD-generated enhancement and tumor sizes in breast cancers before and after neoadjuvant chemotherapy. *Acad Radiol* 12(7):806–14, 2005
- Lehman CD, Peacock S, DeMartini WB, et al: A new automated software system to evaluate breast MR examinations: improved specificity without decreased sensitivity. *AJR Am J Roentgenol* 187(1):51–6, 2006
- Williams TC, DeMartini WB, Partridge SC, et al: Breast MR imaging: computer-aided evaluation program for discriminating benign from malignant lesions. *Radiology* 244(1):94–103, 2007
- Meeuwis C, van de Ven SM, Stapper G, et al: Computer-aided detection (CAD) for breast MRI: evaluation of efficacy at 3.0 T. *Eur Radiol* 20(3):522–8, 2010
- Arazi-Kleinman T, Causer PA, Jong RA, et al: Can breast MRI computer-aided detection (CAD) improve radiologist accuracy for lesions detected at MRI screening and recommended for biopsy in a high-risk population? *Clin Radiol* 64(12):1166–74, 2009
- Sardanelli F, Giuseppetti GM, Panizza P, et al: Sensitivity of MRI versus mammography for detecting foci of multifocal, multicentric breast cancer in fatty and dense breasts using the whole-breast pathologic examination as a gold standard. *Am J Roentgenol* 183:1149–1157, 2004
- Kuhl CK, Schrading S, Bieling HB, et al: MRI for diagnosis of pure ductal carcinoma in situ: a prospective observational study. *Lancet* 370:485–492, 2007
- Partridge SC, Mullins CD, Kurland BF, Allain MD, DeMartini WB, Eby PR, Lehman CD: Apparent diffusion coefficient values for discriminating benign and malignant breast MRI lesions: effects of lesion type and size. *Am J Roentgenol* 194:1664–1673, 2010
- Kuhl CK, Mielcareck P, Klaschik S, et al: Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology* 211(1):101–110, 1999
- Boetes C, Strijk SP, Holland R, Barentsz JO, Van Der Sluis RF, Ruijs JH: False negative MR imaging of malignant breast tumors. *Eur Radiol* 7(8):1231–1234, 1997
- American College of Radiology. ACR practice guideline for the performance of contrast-enhanced magnetic resonance imaging (MRI) of the breast. http://www.acr.org/secondarymainmenucategories/quality_safety/guidelines/breast/MRI-breast.aspx. Accessed August 16, 2010.
- Kuhl CK, Schmutzler RK, Leutner CC, et al: Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. *Radiology* 215(1):267–279, 2000
- Kriege M, Brekelmans CTM, Boetes C, et al: Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N Engl J Med* 351(5):427–437, 2004
- Doi K: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph* 31:198–211, 2007
- American College of Radiology (ACR). ACR BI-RADS-magnetic resonance imaging. In: ACR Breast Imaging Reporting and Data System, Breast Imaging Atlas. Reston: American College of Radiology; 2003.
- Ko ES, Choi HY, Lee BH, Noh W, Kim RB: Central washout sign in computer-aided evaluation of breast MRI: preliminary results. *Acta Radiol* 52:256–263, 2011
- Lehman CD, Peacock S, DeMartini WB, Chen X: A new automated software system to evaluate breast MR examinations: improved specificity without decreased sensitivity. *Am J Roentgenol* 187:51–56, 2006
- Kurz KD, Steinhaus D, Klar V, et al: Assessment of three different software systems in the evaluation of dynamic MRI of the breast. *Eur J Radiol* 69:300–307, 2009