Impact of CAD on Radiologist's Discrimination of Benign from Malignant Breast Lesions on MR

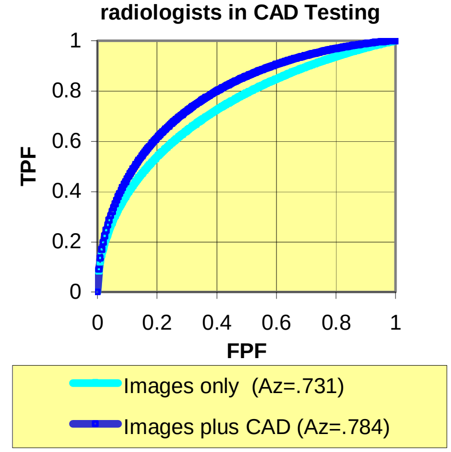
A. I. Penn[1], S. F. Thompson[1], R. F. Brem[2], G. M. Newstead[3], E. D. Pisano[4], P. Weatherall[5], S. M. Ascher[6], E. F. Conant[7], M. D. Schnall[7], C. Lehman[8]

[1]Alan Penn & Associates, Inc., Rockville, Maryland, United States, [2]Radiology, The George Washington University Medical Center, Washington, D.C., United States, [3]Radiology, University of Chicago, Chicago, IL, United States, [4]Radiology, UNC School of Medicine, Chapel Hill, NC, United States, [5]Rogers MR Center, The University of Texas Southwestern Medical Center, Dallas, TX, United States, [6]Radiology, Georgetown University Hospital, Washington, D.C., United States, [7]Radiology, University of Pennsylvania Medical Center, Philadelphia, PA, United States, [8]Radiology, University of Washington Medical Center, Seattle, WA, United States

Introduction: MR can be a valuable tool for detecting, staging, and treating breast cancer. The number of breast MR procedures in North America was estimated to be 87,500 in 2003, and to be growing by 24%-50% per year.[1] Differences in timing sequences, spatial resolution, and other factors have led to multiple interpretation guidelines for assessing the likelihood that an enhancing region is cancer. Differences in protocols and interpretation guidelines can impact clinical outcome: Residents trained to interpret images from one protocol may have difficulty interpreting images acquired with a different protocol, and radiologists may need to interpret breast MR images acquired using multiple protocols. A CAD system is being developed to provide a computerized 2nd opinion of likelihood of malignancy on breast MR images that have been acquired using a variety of protocols and MR imaging systems. The 2nd opinion of likelihood of malignancy is based on a combination of reader interpretations and computer analysis.

Methods: 169 breast MR image sets were acquired from 6 different 1.5T breast MR systems: U of Wash (UW/GE), U of Texas SW (UT/Philips), U of NC (UNC/Siemens), U of Pa (UP/GE and UP/Siemens), and Memorial Sloan Kettering Cancer Center (MSKCC/GE). The CAD system was trained using data from a reader study of 15 radiologists, using 134 of the cases. CAD training consisted of tuning computer analysis features that had been developed in a prior study and refining a Bayesian discriminator. Images from 3 of the 6 imaging systems (UP-GE; UW-GE; UT-Philips) were used for training. The CAD system was tested using an independent set of 20 radiologists, using the 134 training cases plus 35 additional cases. CAD training and testing sets contained approximately ½ cancers and ½ benigns. Readers included 5 experts and 15 non-specialists from 5 cities. Each reader interpreted 60 cases in 2 sessions, first without CAD support and then with CAD support. Reader performance was measured using: (1) area under the ROC curve (Az),[2] (2) partial area under the ROC curve for sensitivity > .90 (pAz)[3] and (3) fraction of readers whose Az changed more than twice estimated intra-observer standard deviation.[4] Az was based on a 0-100 likelihood of cancer assessment. Change in Az and pAz was evaluated using fraction of readers whose Az change between sessions was statistically significant (2-tailed: $p<.05$) or nearly significant ($.05<p<.10$). Inter-observer variability among readers was evaluated using Fleiss-Cohen weighted Kappa[5] for each of the sessions. Mean Kappa was computed using pairs of readers who interpreted the same cases, using reader assessment on a 5-category scale: {Definitely Cancer; Probably Cancer; Uncertain; Probably Benign; Definitely Benign}. Reader interpretations used in computing the 2nd opinion were: (A) specification of rectangular region-of-interest around lesion; (B) categorical description of shape; (C) existence or absence of rim enhancement. The computer features used in the 2nd opinion were: Intensity of lesion, kinetic analysis, and fractal analysis of enhancement pattern adjacent to lesion.

Figure 1
Mean ROC curves for 20 radiologists in CAD Testing



Results: Results are Session 2 (CAD) compared to Session 1 (images only). <u>ROC Area:</u> 2 (10%) of the radiologists showed statistically significant improvement and 3 (15%) of the radiologists showed near statistically significant improvement. <u>Partial (TPF>.90) ROC Area:</u> 5 (25%) of the radiologists showed statistically significant improvement and 1 (5%) of the radiologists showed near statistically significant improvement. <u>Change in Az greater than 2 σ of intra-observer variation:</u> 9 (45%) of readers improved and one (5%) reader worsened > 2σ. Reader that worsened showed increased sensitivity from .72 to .97 and decreased specificity from .65 to .19. Figure 1 shows mean ROC curves for readers in two sessions of study. Mean weighted Kappa among readers increased from .387 ("Fair Agreement") to .513 ("Moderate Agreement").[5]

Conclusion: Our analysis shows that CAD 2nd opinion of likelihood of cancer, which is computed using a combination of reader interpretations and computer image analysis, has the potential for improving discrimination of benign from malignant breast lesions for a substantial fraction of practicing radiologists. The CAD 2nd opinion also shows promise of reducing inter-observer variability.

---

[1] R.D. Barlow. Imaging Technology News. July/August 2004. Citing research from Frost & Sullivan.
[2] C.E. Metz. Rockit 0.9B, Beta Version.
[3] Y Jiung, C.E. Metz, R.M. Nishikawa. Radiology, Vol 201(3): 745-50.
[4] N.A. Obuchowski. AJR, Vol 175: 603-608.
[5] L. Cyr, K. Francis. Comput Biol Med, Vol 22(4): 239-246.