Border and Texture Descriptors for Breast MRI Lesions

R. SIVARAMAKRISHNA[1], N. OBUCHOWSKI[2], S. A. ENGLANDER[3], M. SCHNALL[3]
[1]Cleveland Clinic Foundation, Lerner Research Institute, Dept. of Biomedical Engineering, Cleveland, OH USA; [2]Cleveland Clinic Foundation, Biostatistics & Epidemiology, Cleveland, OH USA; [3]University of Pennsylvania, Dept. of Radiology/MRI Ctr., Philadelphia, PA USA;

Introduction

Mammography is the first line imaging technique for the detection and diagnosis of breast cancer. However, it misses about 10% of cancers, especially those in dense breasts. In the recent years, magnetic resonance imaging (MRI) in combination with T1 enhancement using Gd-DTPA as a contrast agent, has emerged as an adjunctive imaging modality to mammography. While it has been widely observed that, in general, malignant lesions enhance quicker and more intensely than benign lesions, it has also been observed that several benign lesions like fibroadenomas enhance equally fast. So, it is not possible to base the diagnosis of a lesion primarily on its contrast uptake characteristics. Although, computer-aided classification of benign versus malignant lesions, using border and texture descriptors has been widely used in mammography, relatively few similar studies have been performed for lesions seen on breast MRI [1-2]. The focus of this paper is to compare the performance of different border and texture descriptors in distinguishing benign from malignant lesions on breast MRI images.

Methods

46 lesions from 44 cases of known pathology (21 benign, 26 malignant) were selected from patient records at the Department of Radiology, University of Pennsylvania. The MRI images were acquired with a 3D fat-suppressed radiofrequency-spoiled gradient echo sequence on a 1.5T system with a General Electric Signa console. The images used in this study were obtained during the first 90 seconds following the delivery of a 20 cc bolus of Gd-DTPA. The resulting sagittal images consisted of 512x512x28 voxels and were obtained from an acquisition matrix of 512x512x32. For each test case, single 2D slices approximately at the center of the lesion were selected from the precontrast and first postcontrast images. Difference images were obtained between the first postcontrast and the precontrast. Since the difference image contained additional information about the lesions, both the difference and first postcontrast images were used for further analysis. An interactive region growing algorithm was used to segment the lesions from the background on the difference images. The obtained mask was used to define the region-of-interest both on the first postcontrast image as well as the difference image. The following border measures were evaluated on the segmented lesions: margin fluctuation (MF), tumor boundary roughness (TBR) [3], temperature and entropy obtained from 2D geometric surface temperature (GST) measurements (T_{GST}, E_{GST}). Two additional measures were the difference between the boundary length of the lesion and the convex hull of the lesion divided by the convex hull boundary length (FCHL), and the difference between the area enclosed by the convex hull of the lesion and the lesion area divided by the convex hull area (FCHA). For these two additional measures, the lesion boundary was filtered to smooth small changes in boundary fluctuations because of region growing prior to their calculation. The following texture measures were evaluated on the region inside the segmented lesions: mean and variance, a selected set of 5 Laws descriptors (variance values of L7L7, E7E7, W707 and R707 as well as the mean of L7E7, a selected set of 4 Haralick's descriptors (correlation, difference entropy, entropy and inertia), temperature and entropy obtained from 3D GST measurements, and a fractal measure using box-counting [4]. In the case of texture measures, values of these descriptors were obtained for both difference images and the postcontrast images. For all descriptors, we computed the mean and standard deviation (SD) by lesion type. For each border descriptor we estimated the area under the ROC curve using maximum likelihood methods. We tested the null hypothesis that the ROC area equals 0.5, versus the alternative hypothesis that the ROC area is different from 0.5 using a z-test (two-tailed). A significance level of 0.05 was used. For descriptors with ROC area significantly greater than 0.5, we performed pairwise comparisons. Lastly, we compared nested logistic regression models using a likelihood ratio test to identify models that were combinations of these border descriptors. To identify the best model for each family of texture descriptors (i.e. Law, Haralick, GST and Fractal), we used logistic regression analysis to build models for each family. Since the sample size for this analysis is small, we used various methods (i.e. Principal component analysis, correlation analysis, and univariate analysis) to reduce each family to at most two descriptors before modeling. Then we estimated the ROC areas using maximum likelihood methods and tested the hypothesis that the ROC area equals 0.51. We investigated the possibility of creating models from combinations of post and difference descriptors, as well as combinations of texture and border descriptors.

Results

The best discriminators are the models fit from the border descriptors (see Table). Amongst border descriptors, the model with margin fluctuation was not statistically significantly different from 0.5. All other descriptors have ROC areas signficantly greater than 0.5. T_{GST}, E_{GST} and FCHL are highly correlated. There is no combination of two border descriptors that gives a statistically significant improvement over just a single descriptor. Amongst texture descriptors obtained using postcontrast images, ROC areas for all models except that obtained with the Haralick descriptor, entropy, (ROC area (SE): 0.687 (0.078)) were not statistically significantly different from 0.5. Amongst texture descriptors obtained using difference images, ROC areas for all models except that obtained with the Haralick descriptor, difference entropy DE_{HAR} (ROC area (SE): 0.717 (0.074)) , were not statistically significantly different from 0.5. No combination texture descriptor model was better than a univariate model using only DE_{HAR}. No combination model with DE_{HAR} and each of T_{GST}, E_{GST} and FCHA proved to be significantly better than the individual border descriptors themselves.

Discussion

Future research will focus on testing more border and texture descriptors than those reported in this study. Future research will also focus on extending the calculation of descriptors over the entire 3D lesion volume. In this analysis only the first postcontrast image was used. Since a typical patient examination may consist of several postcontrast images, use of all these in conjunction could provide better discrimination. Due to the small sample size, we were able to only build models with at most two descriptors. A larger sample size would improve our ability to produce models that are good discriminators.

References

[1] S.Sinha, F.A.Lucas-Quesada, N.D.DeBruhl, J.Sayre, D. Farria, D.P. Gorczyca and L.W. Bassett, "Multifeature analysis of Gd-Enhanced MR images of breast lesions," JMRI, vol. 7, pp. 1016-1026, 1997. [2] K.G.A. Gilhuijs and M.L. Giger, "Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging," Med. Phys., vol 25, pp. 1647-1654, 1998. [3] S.Pohlman, K.A. Powell, N.A. Obuchowski, W.A.Chilcote and S.Grundfest Broniatowski, "Quantitative classification of breast tumors in digitized mammograms," Med. Phys. vol. 23, pp. 1337-1345, 1996. [4] R.Sivaramakrishna, K.A. Powell, N.A. Obuchowski and W.A.Chilcote, "Texture descriptors for mammographic mass characterization," Proc. IEEE-EMBS, Atlanta, Georgia, 2: 1035, 1999.

	ROC area (Standard Error)	P-value
MF	0.5390 (0.894)	0.660
TBR	0.7689 (0.0685)	<0.001
T_{GST}	0.8009 (0.0641)	<0.001
E_{GST}	0.8009 (0.0641)	<0.001
FCHL	0.8009 (0.0641)	<0.001
FCHA	0.7045 (0.0759)	0.007