Texture Based Computer Aided Malignant Lesion Segmentation of MR Mammography Images Compared With Majority of Manual Segmentations

B. Woods[1], B. Clymer[1,2], S. Sammet[3], R. Koch[3], R. Stevens[3], and M. V. Knopp[3]

[1]Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, United States, [2]Department of Biomedical Engineering, The Ohio State University, Columbus, OH, United States, [3]Department of Radiology, The Ohio State University, Columbus, OH, United States

Purpose

Dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) utilizes a contrast agent along with four-dimensional MR imaging to provide radiologists with the ability to detect malignant lesions and perform a virtual biopsy whereby the size, location, and type of tumor present may be analyzed. Although DCE-MRI has shown to be an effective diagnostic tool, manual segmentation of malignant lesions can be time consuming and suffer from inter- and intra-observer variability. This study details a model-free computer-aided voxel-by-voxel diagnostic tool for segmenting malignant lesions. For seven patient scans, the automated segmented lesions are compared with six segmentations per study produced manually using the pharmacokinetic two compartment model. It is the goal of this study to show that the classifier can be used to emulate the majority of manual segmentations produced by human observers for each patient.

Materials and Methods

MRI Acquisition. This study retrospectively examines DCE-MRI data from scans of seven female patients known to have malignant carcinoma of the breast. All acquisitions were performed on a 1.5 Tesla General Electric MR system at the National Institute of Health Clinical Center. T1-weighted axial images were acquired using a gradient echo sequence (TR/TE = 7.8/4.2 ms; field of view = 300 x 300 mm^2; matrix = 256 x 256; number of slices = 32; slice thickness = 5 mm). A gadopentetate dimeglumine contrast agent (MAGNEVIST; Berlex Laboratories, Wayne, New Jersey, USA) was used via a constant bolus injection at a dosage of 0.2 ml/kg bodyweight. A total of 28 volumes were acquired over 12 minutes for each subject. All acquisitions and analysis were performed following the approved practices of the Internal Review Board for use with human subjects.

Manual Segmentations by Radiologists. Manual segmentation was performed on each dataset by three trained radiologists using a pharmacokinetic two-compartment (1) based analysis tool implemented in IDL (Interactive Data Language, Boulder, Colorado, USA). Each radiologist performed the segmentation process two times each two weeks apart. Previous studies have shown that two weeks is enough time for the radiologists to forget the previous segmentation of the dataset (2). Thus, six segmentations exist for each of the seven studies. Because the segmentations suffer from both inter- and intra-observer variations, we use a gold standard that uses a majority scheme. If at least four out of six manual segmentations mark a voxel as malignant then that voxel is considered malignant in the gold standard segmentation.

Classification System. Co-occurrence based texture analysis was performed on each of the seven DCE-MRI datasets (3). Co-occurrence matrices were constructed by scanning the 0,0,0,1 (x,y,z,t) direction and comparing adjacent (distance = 1) voxels. Thus only temporal voxel variations were analyzed in this study. Different textures were computed depending on scanning window size (5x5x1x2: Contrast; 5x5x1x7: Angular Second Moment, Contrast, Difference Entropy, Difference Variance, Information Correlation 2, 5x5x1x23: Difference Entropy). The computed textures were set as inputs to a neural network based classifier developed in Matlab (MathWorks Inc., Natick, Massachusetts, NJ). The neural network was trained using selected voxels from three of the seven datasets. The four out of six majority segmentation (the gold standard) was used to determine if a voxel should be classified as malignant or nonmalignant. Once trained the neural network can classify a voxel based on the texture inputs for the test voxel. Output from the neural network is the probability that a particular voxel is malignant. Using thresholds (0 to 100%) we can assign outputs below the threshold to be nonmalignant and outputs greater than or equal to the threshold to be malignant.

Evaluation of Classifier Performance. The neural network output performance is evaluated in two different ways. First color coded results can be superimposed on anatomical images from the baseline DCE-MRI scan. The following color scheme was used: True Positive (TP) = yellow, True Negative (TN) = Black/Gray Scaled, False Positive (FP) = Green, and False Negative (FN) = Red. Statistical results can also be obtained. Using a receiver operating characteristics (ROC) curve we plot the true positive fraction (TPF) versus the false positive fraction (FPF). We note that voxels used for training are excluded from statistical analysis.

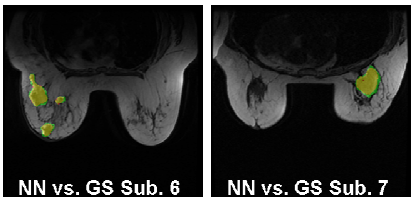

Figure 1: Performance results superimposed on anatomy for subjects six and seven. Yellow: TP, Black/Gray Scaled: TN, Green: FP, Red: FN. Neural network classification results using a threshold of 50% compared with gold standard majority segmentations.

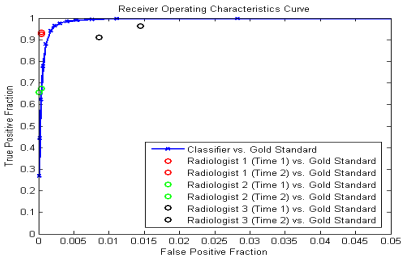

Figure 2: The ROC performance curve produced by comparing the neural network output with the gold standard majority segmentations. The circles 'o' denote the performance of manual segmentations compared with the gold standard majority segmentations.

Results

After training the neural network, all voxels from all seven datasets were classified and both color-coded images and statistical results were obtained. Figure 1 shows the color-coded images comparing the classifier output with the gold standard majority segmentations. From the images we notice that the majority of misclassifications (FP,FN) occur near the border of the lesion where partial volume effects may influence results. Figure 2 shows the ROC analysis based on the TPF and FPF calculated from all voxels in all seven studies. In addition to classifier performance the plot also indicates how each radiologist performed at each of the two times (two weeks apart) compared with the gold standard majority segmentations. These results indicate inter- and intra-observer variations relative to the neural network performance.

A paired *t*-test at significance level of 0.05 was performed using the TPF and FPF values obtained from each of the seven studies. The classifier performance relative to the gold standard was tested against each of the manual segmentations relative to the gold standard. Forty-two total paired samples were obtained (six segmentations per subject times seven subjects). The *t*-test results show that at thresholds between 30% to 60%, the classifier performs statistically equal to or better on average than the manual segmentations relative to the gold standard majority segmentations. At a threshold of 50%, the classifier performs statistically better on average than the manual segmentations relative to the gold standard majority segmentations.

Conclusion

Texture analysis along with a neural network based classifier can be used to detect malignant lesions in DCE-MRI datasets. The results of this study show that the classifier can be trained to emulate the segmentation performance of the majority of a group of radiologists. Further it is shown that once trained the classifier can on average perform on par or better in the statistical (TPF,FPF) sense than manual segmentations. This is because the classifier gives consistent computational results whereas inter- and intra-observer variations occur within the manual segmentation process. When used as a tool, the classifier gives the radiologists a first guess approximation of malignant lesion size and location. From there the radiologist may adjust the segmentation to his or her preference.

References

(1) Brix G., Semmler W., Port R., Schad L., Layer G., Lorenz W.. Pharmacokinetic Parameters in CNS Gd-DTPA Enhanced MR Imaging. *Journal of Computer Assisted Tomography*. 1991;15:621–627.

(2) Chun H. *Auditory Sensory Feedback Tool to Supplement Visual Data Perception in Radiologic Imaging – A Demonstration Using MR Mammography* [Dissertation]. The Ohio State University; 2006

(3) Haralick R.M., Shanmugam K., Dinstein I.. Textural Features for Image Classification. *IEEE Trans. on Systems, Man, and Cybernetics*. 1973;3:610–621.