THE UNIVERSITY OF CHICAGO

INVESTIGATION OF SIMILARITY MEASURES FOR SELECTION OF SIMILAR IMAGES IN COMPUTER-AIDED DIAGNOSIS OF BREAST LESIONS ON MAMMOGRAMS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
COMMITTEE ON MEDICAL PHYSICS

BY

CHISAKO MURAMATSU

CHICAGO, ILLINOIS
MARCH 2008

ABSTRACT

Mammography is considered a useful screening method for early detection of breast cancer. However, the interpretation of mammograms can be difficult, and some cancers are missed at screening. Computer-aided diagnosis (CAD) for detection has a potential to improve the diagnostic accuracy, and some studies have reported that the CAD can increase the radiologists' sensitivity for detecting breast cancer. Even when lesions are detected by a radiologist, some of them might be misclassified as benign lesions and not sent for additional examinations or to biopsy. In addition, many benign lesions are also sent for biopsy. Therefore, a computer-aided diagnosis for differential diagnosis can also be useful for classification of breast lesions. Because radiologists learn and accumulate their diagnostic skills by experience with previous similar cases, the presentation of images with known pathology similar to that of an unknown case can be expected to be useful for radiologists. The objective of this study was to investigate similarity measures for the selection of similar images as a diagnostic aid in the distinction between benign and malignant lesions on mammograms.

The images of masses and clustered microcalcifications used in this study were obtained from the Digital Database for Screening Mammographyof the University of South Florida. We obtained regions of interest with 728 malignant and 840 benign masses, and with 457 malignant and 644 benign clustered microcalcifications which were proved by biopsy. In order to select similar images that are really similar and useful for radiologists, subjective similarity ratings by radiologists for 300 pairs of masses and 300 pairs of clustered microcalcifications were obtained. Ten breast radiologists provided the similarity ratings between 0 and 1, corresponding to "not similar at all" and "almost identical," respectively, based on the overall impression for diagnosis. The average ratings were employed as a "gold standard" for this study.

Four and three similarity measures were investigated for the mass pairs and for the calcification pairs, respectively: Similarity measures based on the pixel-value correlations of images (masses only), distances in feature space, the likelihood of malignancy of lesions, and the output from an artificial neural network (ANN). If an objective similarity measure would agree well with the radiologists' subjective similarity ratings, that measure would be useful for automatic selection of "visually" similar images.

The result of this study indicated that the similarity measures determined on the basis of pixel-value correlation of images and likelihood of malignancy were not very useful. The similarity measures based on the distances in the selected feature space provided moderate correlations with the gold standard for both mass pairs and calcification pairs. The correlation coefficients were improved when similarity measures, called psychophysical similarity measures, were determined by use of the ANN trained with the selected image features and the radiologists' subjective similarity ratings as input data and teacher data, respectively.

The usefulness of similar images in the diagnosis of breast lesions was evaluated in an observer study. Sixty cases, including 30 benign and 30 malignant masses, were selected as unknown cases. A set of benign and malignant images similar to each unknown image was selected based on the sizes of the unknown and known lesions and their similarity measures. Eleven radiologists provided their confidence level regarding the malignancy of the lesions without and with the similar images. The result indicated that the observers' performances without and with similar images were comparable in terms of the area under the curve (AUC) of the receiver operating characteristic analysis. However, in terms of the change in confidence level of malignancy, there were many cases in which the similar images had a beneficial effect to the observers. On average, the number of cases in which the confidence level was changed beneficially was larger than the number of cases in which the confidence level was changed detrimentally. Although not apparent in AUC, the presentation of similar images has a potential to increase radiologists' confidence and improve their diagnostic performance.

This observer study had several limitations: the number of cases included was small, the fraction of extremely difficult cases and cases with small masses was large, and the size criterion for selecting similar images was too strict. For similar images to be useful to radiologists, the similar-image database and the selection scheme must be further improved and reevaluated in the future.