Computer-aidied Diagnosis of Solid Breast Nodules on Ultrasound with Digital Image Processing and Artificial Neural Network

Segyeong Joo[1], Woo Kyung Moon[2] and Hee Chan Kim[3]

[1]Interdisciplinary Program-Biomedical Engineering Major,
[2]Department of Radiology, and
[3]Department of Biomedical Engineering College of Medicine, Seoul National University, Korea

Abstract—**A computer-aided diagnosis algorithm identifying breast nodule malignancy using multiple ultrasonography features and aritificial neural network classifier was developed from a database of 584 hitologically-confirmed cases containing 300 benign and 284 malignant breast nodules. The features were extracted from sonographic images through digital image processing. An artificial neural network then distinguished malignant nodules based on those features. The trained artificial neural network showed the normalized area under the receiver operating characteristic curve of 0.95.**

Keywords—**Computer-aided diagnosis, digital image processing, artificial neural network**

I. INTRODUCTION

Breast cancer is one of the most frequent forms of cancer among women all over the world, and the early detection of the cancer provides the better chance of a proper treatment [1]. The emphasis on the early detection of breast cancer, the desire not to miss a malignant lesion in the early stage of disease, and the current medical environment encourage an aggressive biopsy approach to breast problems. Although well tolerated, biopsy is a typical invasive procedure having some risk, inducing patient discomfort and anxiety, and increasing costs in terms of both patient recovery and overall health care expense. Moreover, the positive biopsy rate for cancer is low, between 10% and 31%, which means 70%-90% of breast biopsies are performed in women with benign disease [2]. Therefore, both mammographic and sonographic methods have been used in attempts to reduce the negative-to-positive biopsy ratio, and therefore, the cost to society by improving feature analysis and refining criteria for recommendation for biopsy [3], [4].

Diagnostic breast ultrasonography (US) is a convenient and safe tool in the classification of tumors. In a benchmark study [5], Stavros *et al.* described a US classification model with a reported 99.5% (424 of 426 cases) negative predictive value and 98.4% (123 of 125 cases) sensitivity. The model based on 20 specific sonographic features of breast masses, including morphologic descriptors of the shape, margin, texture of a mass, acoustic properties such as sonographic sound transmission and mass echogenicity. But subsequent studies [6], [7] showed substantial variability in identification of the specific sonographic features, which could yield varying conclusions on the use of US for characterization of solid breast nodules. Interpreter variability in US differentiation of solid breast nodule was emphasized.

The purpose of our study is to develop a CAD system to identify a subgroup of solid nodules with definitive benign sonographic characteristics suggested by Stavros [5] so as to obviate unnecessary biopsies in clinical situations. The rationale of our scheme for achieving the goal is based on three factors: First, sonographic features are extracted automatically from an image by digital image processing techniques, thereby, inter- and intraobserver variability problem will be solved. Second, through providing multiple sonographic feature values to the ANN, the performance of the decision algorithm can be improved. Third, considering the fact that users of CAD system can locate the suspicious lesion area which covers only small portion of each original image, we started with the manually segmented region of interest (ROI) of the lesion area. This scheme was expected to save processing time as well as to minimize false detection even with a simple edge detection algorithm.

II. METHODOLOGY

A. Ultrasonic image database

The US images used in this paper were provided by the Seoul National University Hospital, Seoul, Korea. The database contained 584 digital ultrasonic images, composed of 300 benign and 284 malignant images. The images were prospectively collected consecutive series of solid masses seen at US and confirmed histologically. Solid mass without a histological confirmation and simple cysts and complicated cysts were excluded. US was performed by one breast radiologist by using a HDI 3000 (Advanced Technology Laboratories, Bothell, WA) or a Voluson 730D (GE Medical Systems, Milwaukee, WI) machine with a 10 MHz transducer and freeze-frame capability. Images used in this study were from the plane showing the longest diameter of the mass.

B. Image processing

1. Median Filtering

The ultrasonic images suffer from speckle noise due to interference of back-scattered signal, and this noise significantly degrades the image quality and hinders to discriminate the fine details. Therefore, ROI was

preprocessed with 4 × 4 median filter to reduce the speckle noise and to enhance features.

2. Unsharp Masking

Unsharp masking is a well-known technique to enhance the edges of the structures in image [8]. To improve the perceptibility of edges of breast nodule, unsharp masking was used after median filtering. 3 × 3 unsharp filter was constructed using the negative of two-dimensional Laplacian filter. The elements with meaningful signal level were emphasized and the contrast between nodule and background was enhanced.

3. Binary Thresholding

After contrast enhancement, ROI was converted to binary image using binary thresholding. The threshold was also determined by the histogram of ROI. If a valley of histogram between 33% and 66% of pixel population can be found, this intensity value was selected as threshold. If there existed no such valley in that range, the intensity of 50% pixel population was selected as threshold value. Another 4 × 4 median filtration removed noises generated due to binary thresholding.

4. Edge Detection

Binary thresholding produces a relatively large object of the nodule together with many separated or interconnected islands. After applying an image opening operation with a disk-shaped (radius = 3 pixels) structuring element [8], we removed small islands having area smaller than 500 pixels (actual area ≈ 6 mm^2) from ROI. Among remaining objects, one closest to the center of ROI was then automatically selected as the nodule. Then, holes inside the nodule were filled. Finally, we obtained the selected nodule's connected boundary pixels by removing interior pixels. The whole image processing procedure is illustrated in Fig. 1.
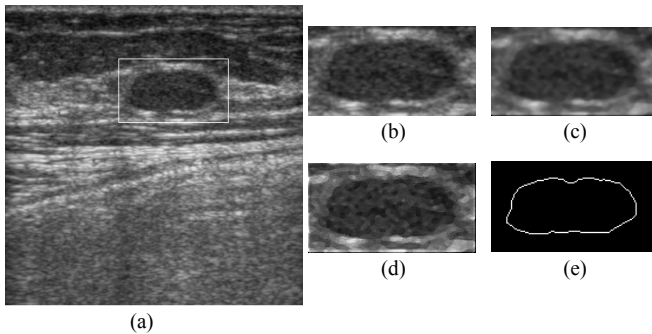

Fig. 1. The whole image processing procedure:
(a) original image captured by US device, (b) ROI image, (c) median filtered image, (d) image after unsharp masking and (e) detected edge.

C. Feature extraction

1. Spiculation

Spiculation showed highest odds ratio (5.5) for malignant sonographic characteristics versus malignant histologic finding in Stavros' study [5]. Spiculation consists of alternating hyperechoic straight lines that radiate perpendicularly from the surface of solid nodule and is a characteristic of malignant nodule. To obtain numeric values representing this feature, the polar coordinates (r,θ) of boundary pixels of nodule images were calculated with the origin at their center of mass. As a result, the boundary of nodule can be represented by the trajectory of its radial distance, $r(\theta)$, as θ spans from 0 to 2π. Finally, the ratio of low frequency component (area under the graph $|R(\omega)|$ from $\omega = 0$ to $\omega = \pi/4$) to high frequency component (area under the graph $|R(\omega)|$ from $\omega = \pi/4$ to $\omega = \pi$) was selected as the feature representing spiculation characteristic.

2. Ellipsoid shape

Ellipsoid shape means that an object has greater sagittal and transverse dimensions than that of anteroposterior. This is a feature of benign nodule [5]. If any part or all of a nodule has greater anteroposterior dimension than that of either sagittal or transverse, the nodule can be considered malignant. Therefore, the ratio of the maximum height to the maximum width of nodule's edge was calculated as a feature representing elliptic shape.

3. Branch pattern

Branch pattern is defined as multiple projections from the nodule within or around ducts extending away from the nipple [5]. This malignant finding can be represented by the number of local extrema in the low pass filtered radial distance graph.

4. Relative brightness of nodule

Malignant nodules are darker when compared with the surrounding isoechoic fat or hyperechoic parenchyma [5]. To calculate the value of brightness, the method of dilation was used. From the detected edge of the ROI, we thickened the boundary of the image. 10 pixel wide layer of boundary was used for the surroundings. The ratio of average gray level of detected region area to the average gray level of surroundings represents the relative brightness value.

5. Number of lobulations

In solid breast nodules, gentle lobulations defined as fewer than four lobulations have been regarded as a sign of benignancy at US [5]. For the detection of peak value and the graph pattern while excluding insignificant local peaks and high frequency characteristics of the edge due to speculations or noise, the radial distance graph was median filtered and curve-fitted to 15th order polynomials. The number of local minima and maxima of the fitted curve is a numerical feature indicating whether the nodule has gentle lobulations or not.

D. Neural network classification

To identify whether a breast nodule is benign or malignant, an ANN classification system was used. This ANN is a general multilayer perceptron (MLP) neural network and the back propagation learning rule was used [9].

1. Network Topology Determination

Network topology was systematically determined in terms of mean and standard deviation of classification accuracy. For all devisable network topologies, classification results were calculated using *k*-fold cross-validation method with $k = 10$ [10]. An ANN with topology having 5 input nodes to receive the extracted feature values of nodule, 8 neurons in the first hidden layer, 8 neurons in the second hidden layer and one output node which indicates whether a nodule is benign or malignant was selected.

2. Classifier Performance Test

With the determined topology, the neural network was trained with half of the 300 benign and 284 malignant images, selected at random. Then with sweeping cutoff value at the output neuron in full scale, ROC analysis was performed to show the efficiency in true positive (TP)/false positive (FP) tradeoff [11], [12].

III. RESULTS

A. Feature Extraction

Fig. 2 and Fig. 3 display the selected ROI on the original image, detected edge, normalized radial distance trajectory $r(\theta)$, log amplitude of Fourier transform (calculated by fast Fourier transform), $R(\omega)$, low pass filtered $r(\theta)$ and curve-fitted $r(\theta)$ with its local extrema of typical benign and malignant nodule, respectively. The radial trajectory of the benign nodule shows smoother curve than that of the malignant nodule. As a result, the amplitude spectrum of the benign nodule contains smaller high frequency components than that of the malignant nodule.
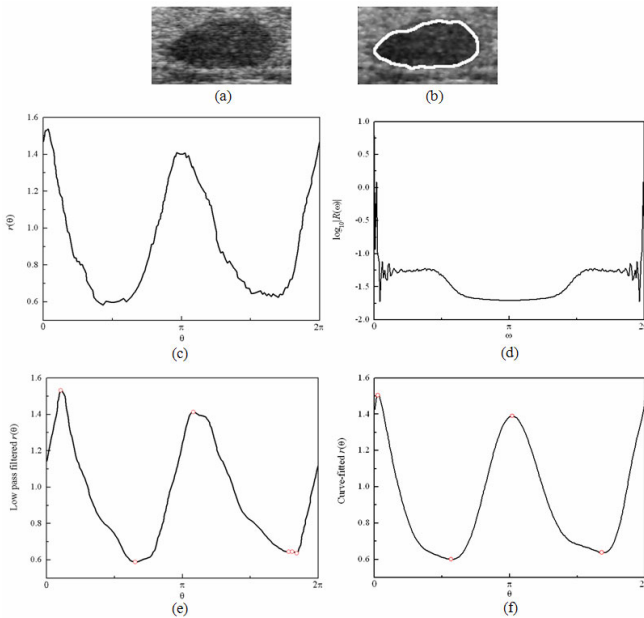

Fig. 2. Typical example of benign nodule with (a) Selected ROI on the original image, (b) detected edge, (c) normalized radial distance trajectory $r(\theta)$, (d) log amplitude of $R(\omega)$, (e) low pass filtered $r(\theta)$ and (f) curve-fitted $r(\theta)$.
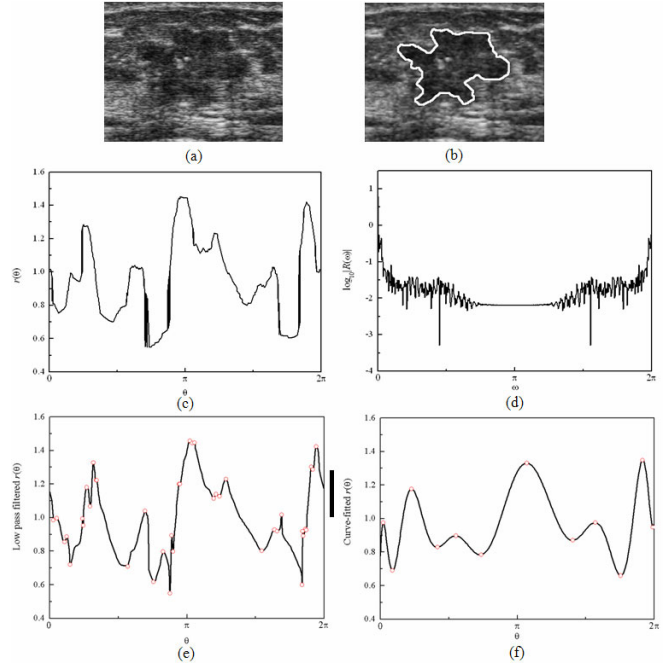

Fig. 3. Typical example of malignant nodule with (a) Selected ROI on the original image, (b) detected edge, (c) normalized radial distance trajectory $r(\theta)$, (d) log amplitude of $R(\omega)$, (e) low pass filtered $r(\theta)$ and (f) curve-fitted $r(\theta)$.

Low pass filtered radial distance graph of the benign nodule shows fewer number of local maxima and minima than that of the malignant nodule. Curve-fitted radial distance graph shows similar trend to low pass filtered radial distance graph.

B. ANN Classification

As a typical result with the cutoff at half of the output scale, the trained ANN showed 100% accuracy for the training set and 91.4% accuracy for the test set. In detail, the sensitivity and the specificity of the trained ANN was 92.3%
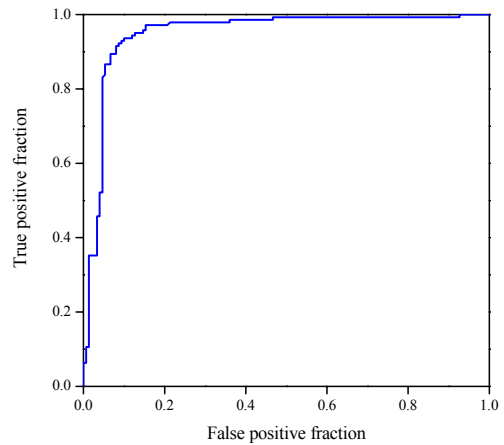

Fig. 4. ROC curve of the developed classifier with automatically detected boundary data from half of the 584 patient images.

(131/142) and 90.7% (136/150), respectively. Fig. 4 is the ROC curve of the trained ANN obtained by sweeping the cutoff value, whose normalized area under the curve (Az) is 0.95. By adjusting the cutoff level, the sensitivity increased to 99.3% (141/142) and 100% (142/142), while the specificity decreased to 53.3% (136/150) and 7.3% (11/150), respectively.

IV. CONCLUSION

We developed a CAD program to determine breast nodule malignancy using digital image processing and ANN based on multiple sonographic features. The typical accuracy for classifying benign and malignant tumors on US was 91.4% with 92.3% sensitivity and 90.7% specificity. The Az value was 0.95. In addition, our results indicate that we could potentially avoid 53.3% of biopsies on benign nodules with 99.3% sensitivity. This performance of our CAD system is comparable to the clinical study by Stavros' in the similar patient populations.

REFERENCES

[1] P.Pisani, D. M. Perkin, and J. Ferlay, "Estimates of the worldwide mortality from eighteen major cancers in 1985: implications for prevention and projections of future burden," *Int'l. J. Cancer*, vol. 55, pp. 891-903, 1993.

[2] J. E. Mayer, T. J. Eberlein, P. C. Stomper, and M. R. Sonnenfeld, "Biopsy of occult breast lesions: analysis of 1261 abnormalities," *J. Am. Med. Assoc.*, vol. 263, pp. 2341-2343, 1990.

[3] E. A. Sickles, "Periodic mammographic follow-up of probably benign lesions: results in 3,184 consecutive cases," *Radiology*, vol. 179, pp. 463-468, 1991.

[4] W. K. Moon, D. Y. Noh, and J. G. Im, "Multifocal, multicentric, and contralateral breast cancers: bilateral whole-breast US in the preoperative evaluation of patients," *Radiology*, vol. 224, pp. 569-576, 2002.

[5] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, "Solid breast nodules: use of sonography to distinguish between benign and malignant lesions," *Radiology*, vol. 196, pp.123-134, 1995.

[6] G. Rahbar, A. C. Sie, G. C. Hansen, J. S. Prince, M. L. Melany, H. E. Reynolds, V. P. Jackson, J. W. Sayre, and L. W. Bassett, "Benign versus malignant solid breast masses: US differentiation," *Radiology*, vol. 213, pp.889-894, 1999.

[7] J. A. Baker, P. J. Kornguth, M. S. Soo, R. Walsh, and P. Mengoni, "Sonography of solid breast lesions: Observer variability of lesion description and assessment," *Am. J. Roentgenology*, vol. 172, pp.1621-1625, 1999.

[8] A. K. Jain, *Fundamentals of digital image processing*, Englewood Cliff, NJ: Prentice Hall, 1986.

[9] J. A. Freeman and D. M. Skapura, *Neural Networks: Algorythms, Applications, and Programming Techniques*, New York, NY: Addison-Wesley, 1992.

[10] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd edition, Englewood Cliff, NJ: Prentice Hall, 1999.

[11] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. VIII, pp.283-298, 1978.

[12] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, New York, NY: Academic, 1990.