

# Stochastic Model and Probabilistic Decision-Based Classifier for Mass Detection in Digital Mammography

Huai Li<sup>1,2,3</sup> K. J. Ray Liu<sup>2</sup> Shih-Chung B. Lo<sup>3</sup> Yue Wang<sup>3,4</sup>

<sup>1</sup>Odyssey Technologies LLC, 10620 Guilford Road, Suite 206, Jessup, MD 20794

<sup>2</sup>Department of Electrical Engineering, University of Maryland at College Park, College Park, MD 20742

<sup>3</sup>Department of Radiology, Georgetown University Medical Center, Washington, DC 20007

<sup>4</sup>Department of Electrical Engineering, The Catholic University of America, Washington, DC 20064

## Abstract

*In this paper, we have developed a combined method utilizing morphological operations, a finite generalized Gaussian mixture (FGGM) modeling, and a contextual Bayesian relaxation labeling technique (CBRL) to enhance and extract suspicious masses. A feature space is constructed based on multiple feature extraction from the regions of interest (ROIs). Finally, a multi-modular probabilistic decision-based classifier is employed to distinguish true masses from non-masses.*

## 1 Introduction

The detection of masses is considered a difficult task for radiologists because of the subtle difference between local dense tissues and masses. In recent years, some computer-aided diagnosis (CADx) schemes for mass detection and classification have been developed by several researchers [1, 2, 3]. Most CADx systems consist of three main procedures: preprocessing, feature extraction, and classification. In the preprocessing step, segmentation of suspected masses from breast tissue background is of great importance since all subtle masses should not be missed at the first step. In addition, the accurate feature extraction mainly depends on the performance of the segmentation. Stochastic model-based image segmentation technique has received a considerable attention in medical image segmentation. In our previous work, we showed that a good segmentation result would depend on the suitable model selection and the enhancement of the patterns of interest [4]. At the end of the preprocessing step, all suspected mass regions as well as some normal dense tissues with brighter intensities are located. The latter should be eliminated from the true masses through feature discrimination. In the proposed mass detection approach, some intensity, geometric, and texture features are extracted and

investigated based on the segmented regions. These features usually possess clinical significance and are widely used in most CADx systems. Finally, a multi-modular probabilistic decision-based classifier is employed to distinguish true masses from non-masses [5, 6]. The classifier can solve classification problems by dividing the input feature space into regions and learning simplified problems within each region. It is a powerful classifier in statistical pattern recognition and can often lead to simple and efficient learning algorithms.

## 2 Methods

The framework of our proposed method is illustrated in Fig. 1. For the purpose of enhancing mass signals, we developed a dual morphological operation approach as follows:

$$f_1(i, j) = \max(0, [(f \circ B_2)(i, j) - (f \circ B_1)(i, j)]), \quad (1)$$

where  $f(i, j)$  and  $f_1(i, j)$  are the original and enhanced images.  $B_1$  and  $B_2$  are two specified structuring elements. This operation can remove the background noise and the structure noise inside the suspected mass patterns.

The finite generalized Gaussian mixture (FGGM) model is used to model the histogram of the image. The FGGM pdf of gray-level of each pixel is given by  $p(x_i) = \sum_{k=1}^K \pi_k p_k(x_i)$ . The generalized Gaussian probability density function given region  $k$  is defined by

$$p_k(x_i) = \frac{\alpha \beta_k}{2\Gamma(1/\alpha)} \exp[-|\beta_k(x_i - \mu_k)|^\alpha], \quad \alpha > 0 \quad (2)$$

With different model parameter  $\alpha$ , the model probability density function represents different distributions. Therefore, the generalized Gaussian model is

a good model for those images which statistical properties are unknown. The model parameters can be estimated using a expectation-maximization (EM) algorithm [4, 7]. The number of image regions  $K$  in the FGGM model can be determined by Akaike information criterion (AIC) and minimum description length (MDL). Once  $K$  is known, we use global relative entropy (GRE) between the histogram and the estimated FGGM distribution as a measure of the estimation bias. The appropriate value of  $\alpha$  can be chosen when GRE is minimum. Given the FGGM model, a contextual Bayesian relaxation labeling technique is employed to perform image segmentation. Finally, binary morphological opening and closing operations are used to reduce all small objects which, as we knew previously, were not masses.

In our feature extraction step, several reasonable features which include geometric features, intensity features, and texture features were extracted based on the segmented regions as shown in Fig. 3. The histogram pairs of these features extracted from true and false mass regions were investigated and the features which can better separate the true and false mass regions are selected to further study. In this study, we chose area, two kinds of compactness (circularity), and difference entropy as useful features for classifying masses and non-masses. The definitions of these features are as follows: Compactness definition 1:  $C_1 = \frac{A_1}{A}$ , where  $A$  is the area of the actual suspected region, and  $A_1$  is the area of the overlapped region of  $A$  and the effective circle  $A_c$ , which is defined as the circle whose area is equal to  $A$  and is centered about the corresponding centroid of  $A$ . Compactness definition 2:  $C_2 = \frac{P}{4\pi A}$  where  $P$  is the boundary perimeter, and  $A$  is the area of region. Given a second-order joint probability matrix  $p_{d,\theta}(i, j)$  which is the joint gray level distribution of a pixel pair with the distance  $d$  and in the direction  $\theta$ , the difference entropy  $DH_{d,\theta}$  is defined as

$$DH_{d,\theta} = - \sum_{k=0}^{L-1} p_{x-y}(k) \log p_{x-y}(k) \quad (3)$$

where  $p_{x-y}(k) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{d,\theta}(i, j)$ ,  $|i - j| = k$ . According to our investigation, these features have the better separation (discrimination) in true mass class and false mass class. They are also not correlated to each other.

The structure of the classifier used in this study is shown in Fig. 2. The classifier is a simple version of mixtures of experts. One modular expert is trained to detect true masses, another is trained to detect false

masses. After training, the feature vectors extracted from ROIs are input to this network to classify true or false masses. In both training and testing processes, the feature vectors  $\vec{x}$  in class  $i$  ( $i = 1, \dots, M$ ) are assumed to be mixtures of multi-dimensional Gaussian distribution, i.e.,  $p(\vec{x}|i) = \sum_{k=1}^K \pi_{ki} p_k(\vec{x}|i)$  and  $p_k(\vec{x}|i) = N(\vec{\mu}_{ki}, \Gamma_{ki})$ . The unsupervised EM algorithm is employed to train the local modular experts separately. We summarize the EM algorithm as follows.

**1. E-Step:** for training sample  $\vec{x}^{(t)}$ ,  $t = 1, \dots, N$ , compute the probabilistic membership

$$h_{ki}^{(m)}(t) = \frac{\pi_{ki}^{(m)} p_k^{(m)}(\vec{x}^{(t)}|i)}{\sum_{k=1}^K \pi_{ki}^{(m)} p_k^{(m)}(\vec{x}^{(t)}|i)} \quad (4)$$

**2. M-Step:** compute the updated parameter estimates

$$\pi_{ki}^{(m+1)} = \frac{1}{N} \sum_{t=1}^N h_{ki}^{(m)}(t) \quad (5)$$

$$\vec{\mu}_{ki}^{(m+1)} = \frac{1}{N\pi_{ki}^{(m+1)}} \sum_{t=1}^N h_{ki}^{(m)}(t) \vec{x}^{(t)} \quad (6)$$

$$\Gamma_{ki}^{(m+1)} = \frac{1}{N\pi_{ki}^{(m+1)}} \sum_{t=1}^N h_{ki}^{(m)}(t) A A^T \quad (7)$$

where  $A = [\vec{x}^{(t)} - \vec{\mu}_{ki}^{(m+1)}]$ . When the EM iteration converges, it should obtain the maximum likelihood estimation of the feature distribution, i.e., the estimation of the weighting factor  $\pi_{ki}$ , the mean vector  $\vec{\mu}_{ki}$ , and the covariance matrix  $\Gamma_{ki}$  of class  $i$ .

### 3 Results

Fifty mammograms with biopsy proven masses and 50 normal cases were selected from the data set for training. The mammogram set used for testing contained 46 single-view mammograms: 23 normal cases and 23 with biopsy proven masses. All selected mammograms were digitized with an image resolution of  $100\mu\text{m} \times 100\mu\text{m}$  per pixel by the laser film digitizer (Model: Lumiscan 150). The image sizes are  $1792 \times 2560 \times 12$  bpp. For this study, we shrunk the digital mammograms with the resolution of  $400\mu\text{m}$  by averaging  $4 \times 4$  pixels into one pixel. According to radiologists, the size of small masses is  $3 - 15\text{mm}$ . A  $3\text{mm}$  object will occupy the range of about  $7 - 8$  pixels in the shrunk mammogram. The object with the size of 7 pixels is expected to be detectable by any computer

algorithm. Therefore, the shrinking step is applicable for mass cases and can save computation time. For enhanced mammograms, we found that  $K = 4$  is the optimal choice using the AIC and MDL criteria. We also found that GRE achieved a minimum distance when the FGGM parameter  $\alpha = 3.0$ . This indicated that the FGGM model is better than the finite normal mixture model ( $\alpha = 2.0$ ), which has been mostly chosen in stochastic model-based segmentation.

After the segmentation, the area index was first used to eliminate the non-mass regions. In our study, we set  $A_1 = 7 \times 7$  pixels and  $A_2 = 75 \times 75$  pixels as the thresholds.  $A_1$  corresponds to the smallest size of masses ( $3mm$ ), and an object with a area of  $75 \times 75$  pixels corresponds to  $30mm$  in the original mammogram. This indicates that the scheme can detect all masses with sizes up to  $30mm$ . Masses larger than  $30mm$  are rare cases in the clinical setting. When the segmented region satisfied the condition  $A_1 \leq A \leq A_2$ , the region was considered to be suspicious masses. The feature vector contained two features: compactness and difference entropy. The feature vector can extend multi-dimensionally. But a large number of features will increase the complexity of the learning procedure of the classifier. A training feature vector set was constructed from 50 true mass ROIs and 50 false mass ROIs. The training set was used to train two modular probabilistic decision-based neural networks separately. Fig. 4 shows the classification of two classes with compactness definition 1. Fig. 5 shows the classification of two classes with compactness definition 2. According to our experience, the values of compactness with definition 1 are more reliable than those of compactness with definition 2. In order to have more accurate texture information, the computation of the second-order joint probability matrix  $p_{d,\theta}(i, j)$  is only based on the segmented region of the original mammogram. For the shrunk mammograms, we found that the difference entropy had better discrimination with  $d = 1$ . The difference entropy used in this study was the average of values at  $\theta = 0^\circ, 45^\circ, 90^\circ$ , and  $135^\circ$ .

After the pre-scan process (i.e., suspected region segmentation), 6 – 15 suspected masses per mammogram were detected and required further evaluation. We found that the classifier can reduce the number of suspicious masses with a sensitivity of 84% at a specificity of 82% (1.6 false positive findings per mammogram) based on the database containing 46 mammograms (23 of them have biopsy proven masses). Fig. 6 shows the mass detection result on one mammogram with a spiculated mass. After the enhancement, ten

regions with brightest intensity were segmented. Using the area criterion, too large and too small regions were eliminated first and the rest regions were submitted to the classifier for further evaluation. The results indicated that the spiculated mass was detected.

## 4 Conclusions

In this study, we proposed an unified method for mass detection on mammographic images. The results demonstrated that the morphological dual operation enhancement combined with the FGGM model-based segmentation is an effective way to extract mammographic suspicious mass patterns. The proposed classifier can effectively reduce the number of false positive cases and detect the true masses. Compared with conventional neural networks, the probabilistic decision-based classifier can lead to a more efficient learning algorithm and can provide more understanding in the analysis of the distribution patterns of multiple features extracted from the suspicious masses.

## References

- [1] M. L. Giger, C. J. Vyborny, and R. A. Schmidt, "Computerized Characterization of Mammographic Masses: Analysis of Spiculation," *Cancer Letters*, Vol. 77, pp. 201-211, 1994.
- [2] H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Alder, M. M. Goodsitt, and N. Petrick, "Computer-Aided Classification of Mammographic Masses and Normal Tissue: Linear Discriminant Analysis in Texture Feature Space," *Phys. Med. Biol.*, Vol. 40, pp. 857-876, 1995.
- [3] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, "Markov Random Field for Tumor Detection in Digital Mammography," *IEEE Trans. on Med. Imaging*, Vol. 14, No. 43 pp. 565-576, 1995.
- [4] H. Li, K. J. R. Liu, Y. Wang, and S. C. Lo, "Morphological Filtering and Stochastic Modeling-Based Segmentation of Masses on Mammographic Images," *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 1792-1796, 1996.
- [5] L. Xu, M. I. Jordan, and G. E. Hinton, "A Modified Gating Network for the Mixtures of Experts Architecture," *Proc. World Congress on Neural Networks*, Vol. II, pp. 405-410, 1994.
- [6] S. H. Lin and S. Y. Kung, "Probabilistic DBNN via Expectation-Maximization with Multi-Sensor Classification Applications," *Proc. Int. Conf. on Image Processing*, Vol. III, pp. 236-239, 1995.
- [7] Y. Wang, T. Adah, and T. Lei, "Unsupervised Medical Image Analysis by Multiscale FNM Modeling and MRF Relaxation Labeling," *Proc. IEEE Information Theory Workshop on Information Theory and Statistics*, pp. 101-103, 1994.

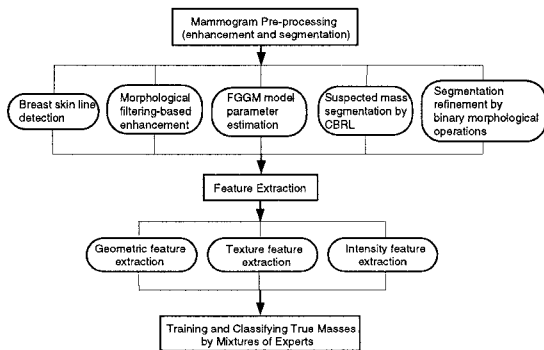


Figure 1: The flow diagram of the mass detection in digital mammograms.

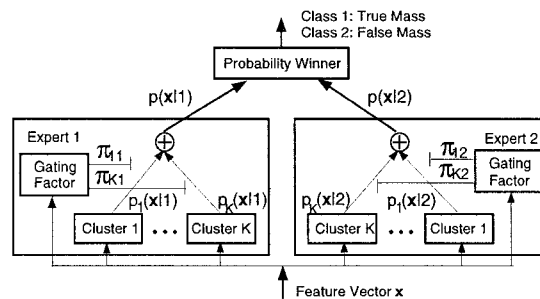
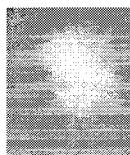
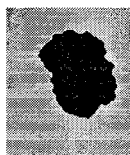


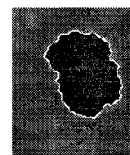
Figure 2: The structure of the mixtures of experts network.



(a) mass patch;



(b) segmentation;



(c) boundary extraction

Figure 3: One example of the mass segmentation and boundary extraction.

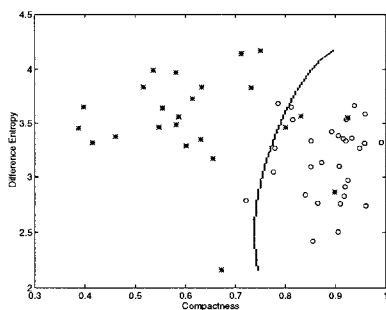


Figure 4: The classification using compactness definition 1. -o- denotes true mass cases; -\*- denotes false mass cases.

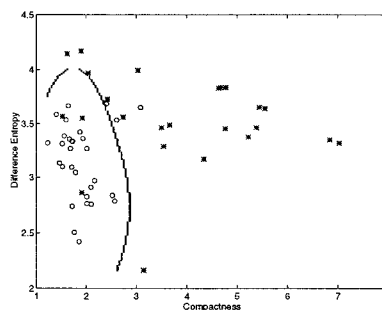
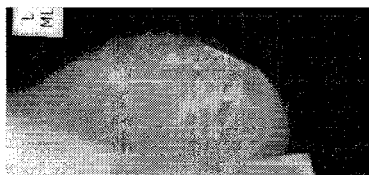
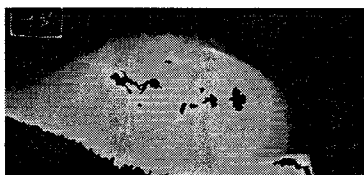


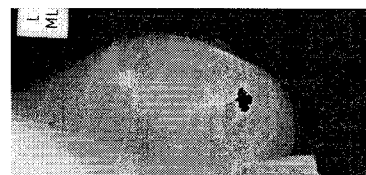
Figure 5: The classification using compactness definition 2. -o- denotes true mass cases; -\*- denotes false mass cases.



(a) original mammogram



(b) suspected mass segmentation



(c) true mass detection

Figure 6: The mass detection results using the proposed approach.