Identifying the Optimal Segmentors for Mass Classification in Mammograms

Yu Zhang*[a], Noriko Tomuro[b], Jacob Furst[b] , Daniela Stan Raicu[b]
[a] Dept. of Computer Information Science, East-West University, 816 Michigan Ave., Chicago, IL USA 60640; [b] College of Computing and Digital Media, DePaul University, 243 S. Wabash Ave., Chicago, IL USA 60604

ABSTRACT

In this paper, we present the results of our investigation on identifying the optimal segmentor(s) from an ensemble of weak segmentors, used in a Computer-Aided Diagnosis (CADx) system which classifies suspicious masses in mammograms as benign or malignant. This is an extension of our previous work, where we used various parameter settings of image enhancement techniques to each suspicious mass (region of interest (ROI)) to obtain several enhanced images, then applied segmentation to each image to obtain several contours of a given mass. Each segmentation in this ensemble is essentially a "weak segmentor" because no single segmentation can produce the optimal result for all images. Then after shape features are computed from the segmented contours, the final classification model was built using logistic regression. The work in this paper focuses on identifying the optimal segmentor(s) from an ensemble mix of weak segmentors. For our purpose, optimal segmentors are those in the ensemble mix which contribute the most to the overall classification rather than the ones that produced high precision segmentation. To measure the segmentors' contribution, we examined weights on the features in the derived logistic regression model and computed the average feature weight for each segmentor. The result showed that, while in general the segmentors with higher segmentation success rates had higher feature weights, some segmentors with lower segmentation rates had high classification feature weights as well.

Keywords: Mammogram, Mass Segmentation, Segmentation Evaluation, Mass Classification

1. INTRODUCTION

Breast cancer is the second leading cause of cancer-related deaths for women in the U.S. after lung cancer. In 2014, it was estimated that 232,670 new cases of breast cancer would be diagnosed among women in the United States; and there were an estimated 40,000 women deaths from breast cancer [1].

Breast cancer is treatable when it is discovered early, so early detection is the best protection. At present, the most effective method for early detection is mammography screening. The National Cancer Institute recommends that women age 40 and older should have mammograms every 1 to 2 years. However, the problem with mammogram screening is that the error rate is high. A study showed that, among the women who had mammogram screenings and were recommended to have a biopsy for further confirmation, only 25% of the cases were found to have breast cancer [2]. Some breast biopsies can be avoided if radiologists can more accurately diagnose those abnormalities in mammograms.

A computer-aided diagnosis (CADx) system automatically extracts a variety of shape and texture features, identifies the most distinguished features by feature selection, and then uses these features for prediction of diagnosis. Previous studies have shown that CADx systems can assist radiologists as a second opinion to analyze a suspected abnormality in a mammogram, and improve the diagnosis accuracy [3, 4]. Mass and microcalcification are the two most common types of abnormalities associated with breast cancer in mammograms.

*jzhang@eastwest.edu; phone 1 312 939-0111 ext 2211; fax 1 312 939-0083; eastwest.edu

Our research focuses on mass diagnosis. The research presented in this paper is part of an ongoing project aiming to develop an image-based CADx system for classifying suspicious masses in mammograms as malignant or benign. The proposed CADx consists of five parts: 1) mass ROI extraction, 2) mass segmentation, 3) mass feature extraction, 4) feature selection and 5) classification. This paper is focus on the mass segmentation evaluations.

Most mammogram images have varied intensity contrast ranges, and patients also have different breast density levels. With those variability factors, it is impossible to apply the same image enhancement, and produce the optimal mass segmentation for all images. In our previous research, we proposed to build multiple segmentors for each mass ROI to adapt the various ranges of image intensity contrast. We call each of segmentations a "weak segmentor" because no single segmentation can produce the optimal results for all images [5].

In our previous study, we used segmentation success rates to evaluate the performance of each weak segmentor. However, by using the success rates, the quality of the segmentations is not fully revealed, and the effectiveness of segmentations for mass classification has to be further measured and assessed. In this research, we propose to use the shape feature weights of a classification model to measure the effectiveness of each weak segmentor, and to identify the optimal segmentors for mass classification.

For this study, we built six edge-based weak segmentors, and extracted shape features from detected mass contours. We then combine the shape features from six segmentors and feed them to a linear logistic regression model for classifying the masses. From the built regression model, the average feature weights from each segmentor were used to evaluate the effectiveness of segmentation for mass classification. Figure 1 displays the framework for this research.

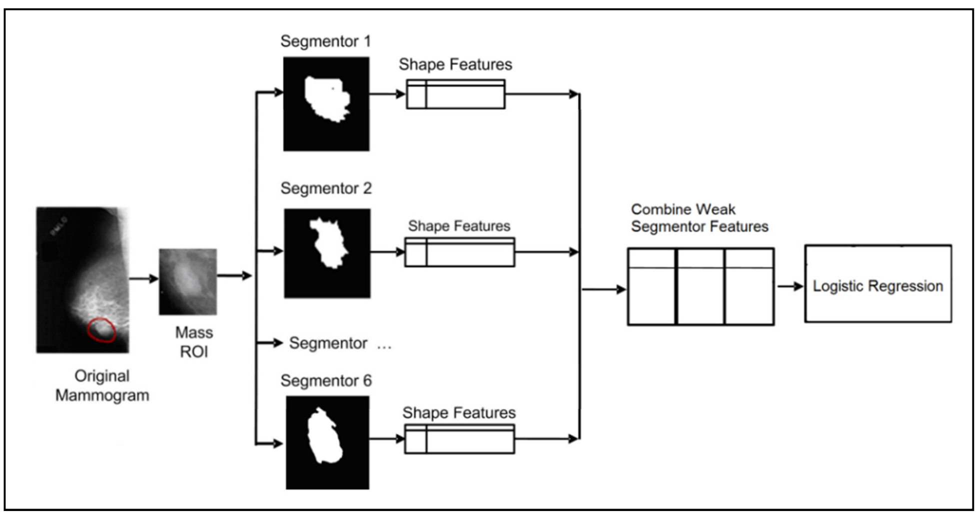


Figure 1.Overall Framework

This paper is organized as following. In Section 2, we review related works of mass segmentation and evaluations. In Section 3, we present the data and methodology. In Section 4, the mass segmentation evaluations are presented. In Section 5, the conclusion and future work are discussed.

2. MASS SEGMENTATIONS AND SEGMENTATION EVALUATIONS

Masses are thickenings of breast tissue which appear as lesions in mammograms. For radiologists, the shape and margin (spiculation level) of masses are the two most important criteria in distinguishing malignant from benign masses. A mass with poorly defined shape is more likely to be malignant than a well-circumscribed mass; and a mass with ill-defined margins or spiculated lesions is much more likely to be malignant than a mass with smoothed margins [6].

In a CADx system for a suspicious mass, segmentation separates a mass region from its background and captures the shape and boundary of the suspicious mass. After segmentation, the contour of a mass is identified; and the shape features and spiculation level can be computed for classifying the mass as benign or malignant. Previous studies have shown that it is critical for a CADx system to extract accurate shape and margin information for suspicious masses; and that improving the mass segmentation can significantly improve the accuracy of mass diagnosis [7, 8].

Many mass segmentation methods for mammogram images have been developed. Mass segmentation methods can be grouped into two basic approaches: 1) similarity approach: region-based methods and 2) discontinuity approach: edge-based methods. In region-based methods, the similarity can be measured with different types of properties, such as pixel intensity or computed texture features. In edge-based methods, segmentation is based on edge detection. Usually image enhancement is applied to ROI images to enhance relevant edges prior to the edge detection stage.

Mencattini et al. [8] modified a region-growing mass segmentation procedure in their Computer Aided Detection (CAD) system. Their segmentation method adjusted mass image contrast by applying a non-linear operator, and applied a region growing segmentation, where pixel intensity was used for similarity measurement. To evaluate the result of segmentation, they computed the completeness and correctness by comparing the area segmented by their algorithm with the ground-truth area, where the completeness was measured as the percentage of the ground-truth region that is detected by the segmented region, and the correctness was measured as the percentage of correctly extracted breast region of the segmented region.

Jiang et al. [9] applied a gamma correction and a Gaussian filter to mass ROI images to improve contrast and remove image noise. Then, using the principle of maximum entropy, optimal thresholds were selected to obtain initial segmented mass regions. In their study, they computed the spiculation index (SI) to assess a segmented mass spiculation level quantitatively.

Song et al. [10] developed an edge-base segmentation method using the plane fitting and dynamic programming techniques to find the "optimal" contour of a mass. The segmentation method was evaluated by comparing the annotations marked manually by radiologists against the segmented region. Yuan et al. [11] developed a method, where a radial gradient index (RGI)-based segmentation was applied to yield an initial mass contour. In their research, the area overlap ratio between the computer segmentation and the manual segmentation by an expert radiologist was computed to evaluate the segmentation.

Yong et al. [12] proposed region growing segmentation. In their research, the procedure of region growing was represented as a growing tree, in which the root was the selected seed. The segmented mass regions were evaluated by the areas overlapping with the annotations made by the radiologists. Byrd et al. [13] presented a comprehensive analysis to evaluate the performance of three existing digital mammography segmentation algorithms against the manual segmentation results produced by two radiologists. In this study, four measures (overlap, accuracy, sensitivity and specificity) were computed to evaluate the three segmentation algorithms.

Petrick et al. [14] developed a region-based mass segmentation algorithm. In their system, the suspicious areas were adjusted using adaptive enhancement method for variability of the image contrast ranges and patients' breast density levels. Ball et al. [15] developed a method that segmented suspicious masses in the polar domain and adaptively adjusted the border threshold at each angle. In their study, the segmentation performance was analyzed by visual inspection and quantitatively measured by the classification accuracies.

Zhang et al. [16] developed an edge-based segmentation method which built energy images and identify the edges of mass contours. In this study, the segmentation was evaluated based on the overlapping ratio -- the proportion of the segmented region over the ROI central area. Zhang et al. [5] proposed to build multiple "weak segmentors" and combine them to achieve a "strong segmentor" for mass segmentations, where the method was evaluated by the overall segmentation rate of the strong segmentor.

3. DATA AND METHOD

3.1 Data Description and Mass ROI Extraction

In this work, all mass ROI images were extracted from the Digital Database for Screening Mammography (DDSM) from the University of South Florida [17]. DDSM is the largest publicly available resource for the mammogram analysis research community. In DDSM images, suspicious regions (ROIs) are marked by experienced radiologists, and BI-RADS information is annotated for each abnormal region [18]. However, in DDSM, most mass ROIs are marked for the mass location, and do not trace the detailed outline of the mass.

In DDSM, mammogram images are digitized by different scanners with different resolutions. In this research, for data consistency purposes, all mass ROI images are collected from the same type of scanner and resolution. We chose the scanner type LUMSYSYS because the largest number of cases are digitized by this type in DDSM. In our experiment, we first extracted all mass ROIs from the mammogram images digitized by LUMSYSYS. Then, we removed the instances with extreme digitization artifacts (e.g. incorrectly ordered scan lines) and of extremely large size (over 2000 x 2000 pixels). We also removed instances with mixed BI-RADS descriptors and those masses which displayed only a portion of a mass. After removing those instances, a total of 543 mass ROI images were left for this study, where 272 instances were benign and 271 instances were malignant.

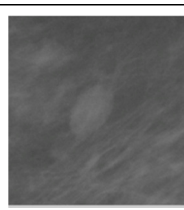
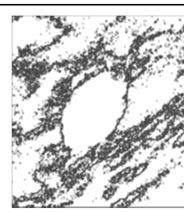
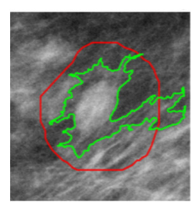
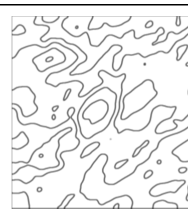
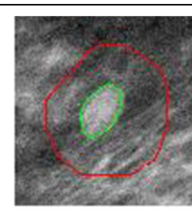
3.2 Building Multiple Edge-based Mass Segmentors

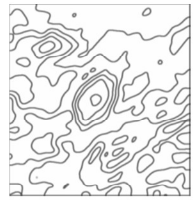
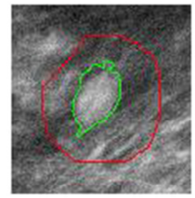
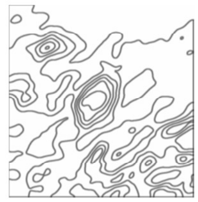
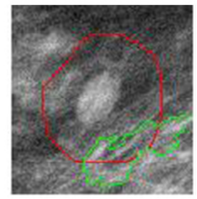
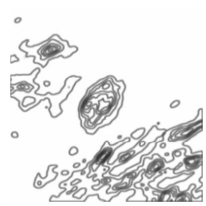
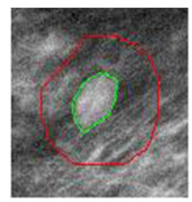
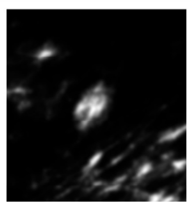
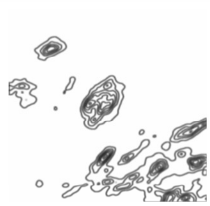
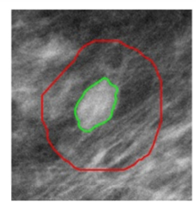
Mammogram images have varied intensity contrast ranges. Those variability factors make it difficult to build single segmentor for all images. In our previous research, we proposed to apply different image enhancements for each mass ROI, and built multiple segmentors for each mass ROI image. Each segmentor is a “weak segmentor”. By combining multiple weak segmentors, we constructed a "strong segmentor" [5].

In this study, we built six edge-based segmentors as weak segmentors. We first applied histogram equalization to each mass ROI to adjust image intensity range; then used five different gamma corrections (γ = 0.5, 1, 2, 5, 7.5) to increase the image contrast and used a Gaussian filter (σ = 5) to suppress the enhanced image noise. Using the original images and their five enhanced images, we computed their energy texture images. Finally, we used our edge-based segmentation method to identify the mass contours [16]. A total of six segmentation results were generated for each mass ROI.

Table 1 shows examples of the original mass ROI, and its enhanced images, mass texture images and six contours detected by the edge-based segmentors from the same ROI. In the examples, the green line is the mass contour identified by our segmentation, while the red line is the mass outline marked by a radiologist.

Table 1 Examples from Six Edge-based Segmentors

	Enhanced Images (a)	Texture Images (b)	Segmentation Results (c)
Weak Segmentor 1: Original ROI Image			
Weak Segmentor 2: Image Enhancement γ = 0.5			

Weak Segmentor 3: Image Enhancement $\gamma = 1$			
Weak Segmentor 4: Image Enhancement $\gamma = 2$			
Weak Segmentor 5: Image Enhancement $\gamma = 5$			
Weak Segmentor 6: Image Enhancement $\gamma = 7.5$			

In this study, each segmentation result is evaluated as successful or unsuccessful based on the overlapping ratio with 25% threshold -- the proportion of the segmented region over the ROI central area [16]. A segmentation is counted as success if its overlapping ratio was above a 25% threshold. For example, in Table 1, mass contours from the segmentors 1, 2, 3, 5 and 6 are evaluated as success, while the segmentor 4 is evaluated as unsuccessful.

Each weak segmentor has individual segmentation rate. A segmentor with a higher segmentation rate indicates that more mass contours have been successfully detected by the segmentor. However, the segmentation rates cannot fully reveal the quality of the segmentations. For example, in Table 1, the mass contours detected from segmentor 1 and 2 both are evaluated as successful segmentations; but by visual inspection, two segmented mass contours are very different.

The purpose of the mass segmentation is to extract shape features for mass classification. Therefore, in the next step, we use their shape features to measure the effectiveness of each segmentor for mass diagnosis.

3.3 Mass Shape Feature Extraction

In this step, we compute shape features from mass contours detected by six weak segmentors. The shape features include: area, convex, perimeter, circularity, compactness, solidity, convex, roughness, equivalent diameter, elongation, major axis length, minor axis length, eccentricity and extent [19]. In addition, we also include a binary feature – the segmentor success indicator, where value of 1 represents success, and 0 indicates failure of the segmentor. So there are 15 features from each weak segmentor.

For each ROI image, we concatenate the shape features from the six weak segmentors and represent each mass instance by a total of 15 x 6 shape features:

$$\{\{f_{1\_1}, f_{1\_2}, \ldots, f_{1\_m}, \ldots, f_{1\_15}\}, \{f_{2\_1}, f_{2\_2}, \ldots, f_{2\_m}, \ldots, f_{2\_15}\}, \ldots, \{f_{k\_1}, f_{k\_2}, \ldots, f_{k\,m}, \ldots, f_{k\_15}\}\}$$

where f_{k_w} ($1 <= k <= 6$, $1 <= w <= 15$) denotes a value of the wth shape feature produced by the kth segmentor. Thus, a total of 90 shape features were computed for feature extraction (6 weak segmentors x 15 shape descriptors). For some mass ROIs, no shape features can be computed because of the unsuccessful segmentation. We set those feature values to an average value so that they will have no influence in classification.

3.4 Logistic Regression

In this step, we feed the shape features extracted from six weak segmentors to a logistic regression model for mass classification. Then, we use the feature weights from each segmentor to measure the effectiveness of the segmentation for the mass classification.

Logistic regression (LR) is used for prediction of the probability of binary class to a logistic function. A logistic regression resulting model is:

$$Pr\,[y=1|x_1, x_2,\ldots, x_k] = 1\,/\,(\,1 + exp\,(- w_0 - w_1x_1 - w_2x_2 - \ldots - w_kx_k)\,)$$

where x_k is the kth input feature and w_k is its regression weight, and y is the outcome variable. In this work, we used a widely used machine learning software Weka [20] to build a logistic regression model [21].

In our experiment, all input features are normalized as the same value range, therefore the weighs on the features in the derived logistic models reflect the relevancy of the features in classification, i.e., features with larger weights (positive or negative) have more influence on the classification. Using the square root of the absolute value of the feature weights in the derived logistic model; we computed the average feature weight for each segmentor and examined the relative contribution in the overall classification.

4. RESULTS AND DISCUSSION

In the experiment, using one original mass ROI image and five enhanced images, we built a total of six edge-based segmentors. Table 2 shows the success rate and the average shape feature weight of each segmentor.

Generally speaking, the results show that segmentors with higher segmentation success rates also have higher feature weights, which indicates that higher-precision segmentors are more effective and useful for mass classification. For example, Segmentor-5 (with $\gamma = 5$) had the best segmentation rate (76.6%) as well as the highest average weight (0.63), and Segmentor-6 (with $\gamma = 7.5$) has the second-best segmentation rate (75.0%) and the third-highest average weight (0.59). However, notably Segmentor-1 (the original ROI) had the lowest segmentation rate (47.7%) but the highest average weight (76.6%; tied with Segmentor-5). This suggests the segmentation precision is not the full indicator of effectiveness of a segmentor, because even low-precision segmentors were contributing in classification.

Table 2 Segmentation Results by Each Weak Segmentor

	Segmentor 1	Segmentor 2	Segmentor 3	Segmentor 4	Segmentor 5	Segmentor 6
Image Enhancement	None	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 5$	$\gamma = 7.5$
Successful Segmentation Rate	47.7%	60.4%	66.3%	72.6%	76.6%	75.0%
Average Feature Weights	0.63	0.52	0.57	0.42	0.63	0.59
Percentage of Total Weights	**18.8%**	15.3%	16.9%	12.6%	**18.8%**	17.6%

Essentially what we observed is the synergy from ensemble -- a mix of weak segmentors of various precision complemented each other in classification. In supporting our conclusion, we have noticed by visual inspection that Segmentor-1 produced more detailed mass contours than other segmentors, and that might have played a supplemental role to other high-precision segmentors.

In Table 2, we also list the percentages of the total weights from each segmentors, which can be used as the measurement of the segmentors' contributions for mass classification. We can identify the top contributors which are segmentor-1 and segmentor-5 as the optimal segmentors for mass classification.

Figure 2 shows the comparison of segmentor average feature weights vs. and segmentation rate from six segmentors. The result shows that for the enhanced images, the segmentation rates are going up from 60.4% to 76.6%, and then going down slightly; they all have much higher segmentation rate than the original image. And we notice that the segmentor-4 (γ = 2) with 72.6% success rate, has a much lower regression weight (0.42). In our future work, we will further investigate this segmentor.

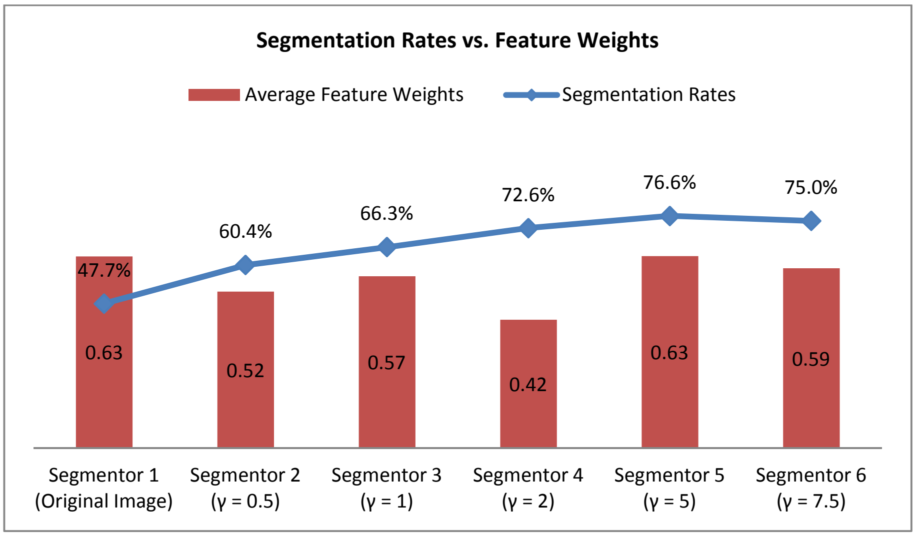


Figure 2.Segmentation Rates vs. Feature Weights

5. CONCLUSIONS AND FUTURE WORK

In the DDSM database there are no detailed mass contours for suspicious masses. Since there is no solid ground truth, it is difficult to evaluate our segmentation results in that aspect. In previous studies we used successful segmentation rates to evaluate the performance of each segmentor. However, by using this evaluation method, the quality of the segmentation is not fully revealed.

In this study, we built six weak segmentors for a given ROI by using various image enhancements. And, each weak segmentor is further analyzed by their extracted shape features for mass classification. Using this method, we can identify the optimal segmentors which have more contributions for mass classification than other segmentors. In a CADx system, by using only optimal segmentors, we expect that the processing time can speed up and the classification accuracy can possibly be improved.

In our future work, we will extract shape features from each segmentor for classification, and use the classification performance to identify the optimal segmentors for mass diagnosis. We may also explore other evaluation methods for mass segmentation.

REFERENCES

[1] American Cancer Society(ASC), "American Cancer Society, Cancer Facts & Figures 2014," http://www.cancer.org, (2014).

[2] Ries, LAG, Harkins, D., Krapcho, M., et al., "SEER Cancer Statistics Review, 1975-2003," National Cancer Institute (2006).

[3] Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P, et al., "Approaches for Automated Detection and Classification of Masses in Mammograms," Pattern Recognition, 39(4), 646-668 (2006).

[4] Mu, T., Nandi, A. and Rangayyan, R., "Classification of Breast Masses Using Selected Shape, Edge-sharpness, and Texture Features with Linear and Kernel-based Classifiers," Journal of Digital Imaging Volume 21, 153-169 (2008).

[5] Zhang, Y., Tomuro, N., Furst, J. and Raicu, D.S. "Building Multiple Weak Segmentors for Strong Mass Segmentation in Mammogram." In SPIE Medical Imaging, 796233-796233 (2011).

[6] Winchester, D. J., Winchester, D. P., Hudis C. A. and Norton, L., [Breast Cancer(Second Edition)], Springer, (2007).

[7] Domínguez, A. R. and Nandi, A. K., "Toward Breast Cancer Diagnosis Based on Automated Segmentation of Masses in Mammograms," Pattern Recognition Volume 42, 1138-1148 (2009).

[8] Mencattini, A., Rabottino, G., Salmeri, M., Lojacono, R. and Colini, E., "Breast Mass Segmentation in Mammographic Image by an Effective Region Growing Algorithm," Advanced Concepts for Intelligent Vision Systems Conference Volume 5259, 948-957 (2008).

[9] Jiang, L., Song, E., Xu, X., Ma, G. and Zhang, B., "Automated Detection of Breast Mass Spiculation Levels and Evaluation of Scheme Performance," Acad Radiol 15(12), 1534-1544 (2008).

[10] Song, E., Jiang, L., Jin, R., Zhang, L., Yuan, Y. and Li, Q., "Breast Mass Segmentation in Mammography Using Plane Fitting and Dynamic Programming," Academic radiology 16(7), 826-35 (2009).

[11] Yuan, Y., Giger, M. L., Li, H., Suzuki, K. and Sennett, C., "A Dual-stage Method for Lesion Segmentation on Digital Mammograms," Med. Phys. Volume 34, 4180-4193 (2007).

[12] Yong, C., Li, L. and Robert, A.C., "Graph-based Region Growing for Mass Segmentation in Digital Mammography," Proceedings of SPIE Vol.4684, 1690 (2002)

[13] Byrd, K. A., Zeng, J. and Chouikha, M., "A Validation model for segmentation algorithms of digital mammography images," Journal of Applied Science & Engineering Technology, 1 (2007).

[14] Petrick, N., Chan, H.P., Sahiner, B. and Helvie, M.A., "Combined Adaptive Enhancement and Region-growing Segmentation of Breast Masses on Digitized Mammograms," Med. Phys. 26, 1642 (1999).

[15] Ball, J.E. and Bruce, L.M., "Digital Mammographic Computer Aided Diagnosis (CAD) Using Adaptive Level Set Segmentation," Engineering in Medicine and Biology Society, 4973-4978 (2007).

[16] Zhang,Y., Tomuro,N., Raicu, D.S. and Furst, J., "Image Enhancement and Edge-based Mass Segmentation in Mammogram," In SPIE Medical Imaging, 76234-76234 (2010).

[17] DDSM, "The Digital Database for Screening Mammography, " http://marathon.csee.usf.edu/Mammography/DDSM/ (2015)

[18] Heath, M., Bowyer, K., Kopans, D., Moore R. and Kegelmeyer, W. P., "The Digital Database for Screening Mammography," Proceedings of the Fifth International Workshop on Digital Mammography, Medical Physics Publishing, 212-218 (2001).

[19] Choras, R.S., "Shape and Texture Feature Extraction for Retrieval Mammogram in Databases," In Information Technologies in Biomedicine, 121-128 (2008).

[20] Witten, I. and Frank, E., Data Mining, [Practical Machine Learning Tools and Techniques (Second edition)], Morgan Kaufmann (2005).

[21] Sumner, M., Frank, E., Hall, M., "Speeding up logistic model tree induction." Knowledge Discovery in Databases: PKDD 2005. Springer Berlin Heidelberg, 675-683 (2005).