Advanced Materials Research Vol. 896 (2014) pp 672-675
Online available since 2014/Feb/19 at www.scientific.net

doi:10.4028/www.scientific.net/AMR.896.672

Comparison between Automatic and Semiautomatic Thresholding Method for Mammographic Density Classification

Shofwatul 'Uyun[1,a], Sri Hartati[2,b] , Agus Harjoko[2,c] , Subanar[3,d], Lina Choridah[4,e]

[1]Department of Informatics, Faculty of Science and Technology, Sunan Kalijaga State Islamic University, Yogyakarta, Indonesia

[2]Department of Computer Science and Electronics, [3]Department of Mathematics,

[2,3]Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Yogyakarta, Indonesia

[4]Department of Radiology, Faculty of Medicine, Gadjah Mada University, Yogyakarta, Indonesia

[a]shofwatul.uyun@uin-suka.ac.id, {[b]shartati, [c]aharjoko, [e]linachoridah}@ugm.ac.id, [d]subanar@yahoo.com.

Keywords: mammographic density classification, breast cancer risk, automatic threshold, semi automatic threshold

Abstract. Mammographic density is a novel independent risk factor of breast cancer that reflects the amount of fibroglandular tissue. Breast Imaging Reporting and Data System (BIRADS) density is one of the mammographic density classification schemes which are most widely used by radiologists. Initially, the method used for assessing mammographic density was subjective and qualitative. Recently however, the measurement of mammographic density is more objective and quantitative. In this paper, we propose an alternative model of breast cancer risk factor assessment based on a quantitative approach of density mammogram. This model consists of pre-processing, breast area counting, fibroglandular tissue area counting that uses maximum entropy and multilevel thresholds, and finally breast density counting to determine the risk classification of breast cancer. The proposed model has been tested on a private database from Oncology Clinic Kotabaru, Yogyakarta, Indonesia consisting of 30 mammograms and has been analyzed by some radiologists using the semiautomatic threshold. The result shows that percentage of mammographic density counted by maximum entropy threshold method has the accuracy, sensitivity and specificity of about 67%, 50% and 75% respectively compared to the semiautomatic thresholding method. On the other hand, the accuracy, sensitivity and specificity resulted from using multilevel threshold is about 93%, 87% and 95% respectively. The obtained results suggest that multilevel threshold is perfectly suited for getting quantitative measurement of mammographic density as one of the strongest risk factors for breast cancer.

Introduction

Breast cancer is the most common cancer cases found in women worldwide. Mammographic density is a novel independent risk factor of breast cancer that reflects the amount of fibroglandular tissue. It has potential clinical applications in screening, in research on breast cancer prevention, and in risk prediction in individuals. There are qualitative and quantitative methods to assess mammographic density. Both assessments can be categorized as (1) extremely fatty <25% of dense tissue; (2) scattered density 25–50%; (3) heterogeneous density 51–75%; and (4) extreme density >75%) [1]. The quantative assessment is carried out by calculating ratio of fibroglandular and breast tissue area. Therefore, to determine these areas, a segmentation process needs to be carried out. The segmentation methods used for mammogram include thresholding, statistics approach, region growing, markov random field, watershed transformation and clustering. For the purpose of computing fibroglandular and breast tissues, the area threshold method is used here. Thresholding methods for getting the breast area have been proposed by [2], including row by row threshold (RRT) and average row threshold (ART). The results show that the ART performance is much

better than the RRT. Other thresholding methods to determine fibroglandular tissue area have been proposed by [3,4]; namely cross entropy threshold and gaussian mixture modeling threshold. In addition, multi-level thresholding is applied to segment color images in the cells of breast cancer [5]. The paper aims to compare between automatic and semiautomatic thresholding methods for classifying mammographic densities. Three methods of automatic threshold used in this work were the triangle threshold, the maximum entropy threshold, and the multilevel threshold.

Materials and methods

The data used in this work were taken from the private database of Oncology Clinic Kotabaru, Yogyakarta, Indonesia that consisted of 30 mammograms of 30 patients with CC view (Cranio Caudal), which was the result of digitizing scanner with a resolution of 50 microns and the BMP (BitMaP format) with a certain different size. The data were analyzed and classified by the radiologists into four risk levels of breast cancer using semiautomatic threshold. The method proposed in this research consists of pre-processing, breast area counting, fibroglandular tissue area counting, and breast density counting to determine the risk factors of breast cancer, as shown in Figure 1.

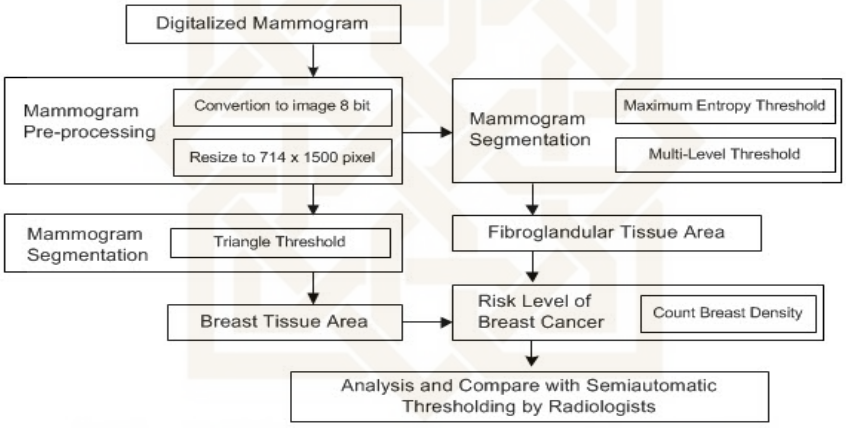


Figure 1. The proposed model

Pre-processing; The pre-processing step was carried out for converting RGB to grayscale mammograms that resulted on digitizing scanner and resizing an image of a mammogram to be 714 x 1500 pixels.

Breast tissue area counting; Segmentation of mammogram was carried out to separate the breast tissue area from the background using triangle threshold that has been introduced by [6]. This technique was particularly effective when the object pixels produced a weak peak in the histogram. A line was constructed between the maximum of the histogram at brightness b_{max} and the lowest value b_{min} =*(p=0)%* in the image. The distance d between the line and the histogram h[b] was computed for all values of b from $b=b_{min}$ to $b=b_{max}$. The brightness value b_0 in which the distance between $h[b_0]$ and the line was maximum was the threshold value.

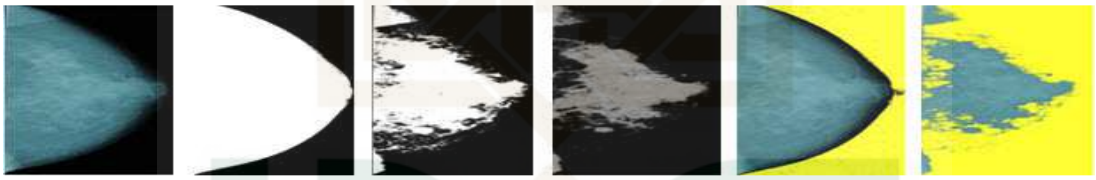
Fibroglanduar tissue area counting. Two methods used here to determine fibroglandular tissue area were maximum entropy threshold and multilevel threshold. The difference between those two methods lied in the amount of threshold being used. For the maximum enthrophy, the threshold value was used to separate histogram into two classes based on the enthrophy value. On the other hand, multilevel thresholding used three threshold values to divide histogram into four classes : uncompressed fatty tissues, fatty tissues, non-uniform density tissues and high density tissues as fibroglandular tissue area based on the discriminant value. This research only limited the use of fibroglandular tissue area resulted by multilevel thresholding to in advance count the breast density. Firstly, maximum entropy was an automatic threshold method that has been described by [7] in the

literature. The optimal threshold value can be obtained by taking the maximum value of foreground and background classes. By applying this method, a precise threshold value would be enhanced if the pixel value between fibroglandular tissue and its background was quite different. General algorithm of maximum entropy; suppose that h(i) is a value in a normalized histogram, counting the entropy of black pixels, then of white pixels and the optimal threshold can be selected by maximizing the sum of foreground and background entropies. Secondly, multilevel thresholding was another threshold method that has been introduced by [8], this method can be used for image segmentation that used more than one value threshold which was for finding a modified between-class variance.

Breast density counting; Based on quantitative BIRADS, the risk level of breast cancer could be calculated based on its breast density. The breast density is a comparison between fibroglandular and breast tissue area of mammograms. The result of calculating breast density in the percentage form showed that the risk of breast cancer on BIRADS that was classified into four risk classes.

Experimental Result

The proposed alternative model for calculating breast density has been tested using 30 mammograms. The same data have been analyzed by radiologists using semiautomatic threshold for getting both fibroglandular and breast tissue area. Figure 2 shows an example of analyzed mammogram, in which Figure 2(a) shows the original image, Figure 2(b) and 2(e) show the breast area obtained using the triangle threshold and semiautomatic threshold respectively, while Figure 2(c), 2(d) and 2(f) show the fibroglandular tissue area obtained using maximum entropy, multilevel threshold and semiautomatic threshold respectively.



(a) (b) (c) (d) (e) (f)

Figure 2. Mammogram sample (a) Original image. Breast tissue area obtained using (b) triangle threshold and (e) semiautomatic threshold. Fibroglandular tissue area obtained using (c) maximum entropy threshold, (d) multilevel threshold and (f) semiautomatic threshold

The breast density was found by calculating the ratio between fibroglandular tissue area and breast area. The experimental results of calculating breast density are shown in Figure 3 and labeled by  ,  and  where the x-axis shows the mammograms that are used in this experiment whereas the y-axis shows the breast density.

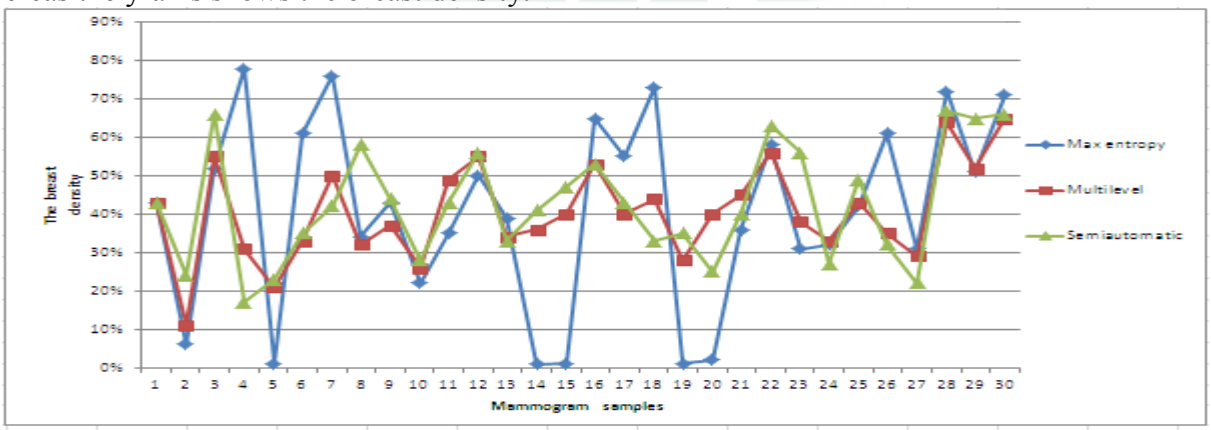


Figure 3. The results of calculating breast density using maximum entropy, multilevel and semiautomatic thresholding

After calculating the breast density, the next process was to classify the breast into four risk factor classes based on BIRADS scheme in which the lowest class had less risk than the higher

class. The result of classification was further compared with classification proposed by the radiologists that used semiautomatic threshold as shown in Table 1. For example, number 2 in the third row and the second column shows that there are two mammograms having risk factor of breast cancer in the level one that have breast density <25%, according to the maximum entropy threshold.

Table 1. Assessment matriks between automatic (two methods) and semiautomatic threshold

Assesment by radiologists	Maximum entropy				Multilevel threshold			
	I	II	III	IV	I	II	III	IV
I	2	1	0	1	2	2	0	0
II	5	7	4	1	0	17	0	0
III	0	3	6	0	0	2	7	0
IV	1	0	0	0	0	0	0	0

The performance of each method was analyzed statistically using the measurement of sensitivity, specificity and accuracy. For this reason, it was needed to calculate the value of four parameters of TP (true positive), FP (false positive), TN (true negative), and FN (false negative). The result shows that percentage mammographic density counted by maximum entropy threshold method has the accuracy, sensitivity, and specificity of about 67%, 50%, and 75% respectively. On the other hand, the use of multilevel threshold results in accuracy, sensitivity and specificity of about 93%, 87% and 95% respectively.

Conclusion

Two threshold methods were proposed based on mammograms characteristics. The two methods were maximum entropy and multilevel thresholds implemented on a number of mammogram images. A comparison analysis was carried out between maximum entropy and multilevel thresholds and it was found that the multilevel threshold performance was much better than maximum entropy threshold that had accuracy, sensitivity, and specificity of about 93%, 87%, and 96% respectively. The obtained results suggest that multilevel threshold was better suited for getting quantitative measurement of mammographic density as one of the strongest risk factors for breast cancer.

References

[1] K.E. Martin, M.A. Helvie, C. Zhou, M.A. Roubidoux, J.E. Bailey, C. Paramagul, C.E. Blane, K.A. Klein, S.S. Sonnad and H.P. Chan, Radiology, 240 (3), (2006) 656-665.

[2] A. Abubaker, R.S. Qahwaji, M. Aqel & M. Saleh, Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the. IEEE, (2006) 3288–3291.

[3] C. Olsén, A. Mukhdoomi, Automatic Segmentation of Fibroglandular Tissue, Image Analysis. Springer Berlin Heidelberg, (2007) 679–688.

[4] R.J. Ferrari, R.M. Rangayyan, R. Borges & A.F. Frère, Medical and Biological Engineering and Computing, 42(3), (2004) 378–387.

[5] R. Harrabi, and E.B. Braiek, EURASIP Journal on Image and Video Processing 2012, 1(2012) 1-11

[6] G.W. Zack, W.E. Rogers, and S.A. Latt, Journal of Histochemistry & Cytochemistry, 25(7), (1977) 741-753.

[7] S. Jarek, Maximum Entropy Thresholding, (2004) [Online] Avalible from: http://ij-plugins.sourceforge.net/plugins/ segmentation/Maximum_Entropy_Thresholding.pdf

[8] P.S. Liao, T.S. Cheen and P.C. Chung, Journal of Information Science and Engineering, 17(5), (2001) 713-727.

Advanced Materials Science and Technology
10.4028/www.scientific.net/AMR.896

Comparison between Automatic and Semiautomatic Thresholding Method for Mammographic Density Classification

10.4028/www.scientific.net/AMR.896.672