WEB OF SCIENTIST: INTERNATIONAL SCIENTIFIC RESEARCH JOURNAL
ISSN: 2776-0979, Volume 3, Issue 8, Aug., 2022

MAMMOGRAPHY IMAGE SEGMENTATION IN BREAST CANCER IDENTIFICATION USING THE OTSU METHOD

Sh. Kh. Fazilov
Research Institute for Development of Digital Technologies and Artificial Intelligence Tashkent, Uzbekistan
sh.fazilov@mail.ru

O. R. Yusupov
Department of Software Engineering
Samarkand State University Samarkand, Uzbekistan
ozodyusupov@gmail.com

Kh.S. Abdiyeva
Department of Software Engineering
Samarkand State University Samarkand, Uzbekistan
orif.habiba1994@gmail.com

Abstract

Exact medical image segmentation is the first step in assisting medical experts in determining diseases. The desired information from the segmented image helps in diagnosis and treatment planning. Nowadays, some cancer types can be detected by using medical images such as mammograms, computer tomography images, and others. Mammography has become the most effective method for the early detection of breast cancer. A mammogram is a difficult image to interpret with unwanted parts in the background. Hence, a preprocessing phase is very important to standardize and enhance the quality of the image for a CAD system. Binarization based on Otsu's threshold is a main part of all preprocessing steps. Otsu's thresholding method is analyzed and reviewed in this paper.

Keywords—mammogram, CAD system, image, image segmentation, Otsu's thresholding.

INTRODUCTION

Cancer is a main cause of death and reduces people's life expectancies in every country. Breast cancer in women is the most common type of cancer in developing and developed countries worldwide [1]. Breast cancer arises in the lining cells (epithelium) of the ducts (85%) or lobules in the glandular tissue of the breast [1]. In the beginning, the growth of cancerous cells is confined to the duct where it generally causes no symptoms and has minimal potential for spread (metastasis). Over time, in the first stage of the disease, cancer may progress and invade the surrounding breast tissue (invasive breast cancer), then spread to the nearby lymph nodes (called "regional metastasis") or to other organs in the human body (called "distant metastasis"). A woman dies from breast cancer because of widespread metastasis [2]. Treatment of breast cancer can be highly effective, especially when the disease is identified in its early stages. Treatment of this illness often consists of a combination of surgical removal, radiation therapy, and medication (hormonal therapy, chemotherapy, and targeted biological therapy) to treat microscopic cancer that has spread from the breast tumor through the blood [15]. In 2020, there were 2.3 million women diagnosed with breast cancer and 685, 000 deaths globally. As of the end of 2020, there were 7.8 million women alive who had been diagnosed with breast cancer in the past 5 years [1]. Breast cancer strikes women at any age after puberty in every country on the planet, with rates rising as they become older. From the 1930s until the 1970s, breast cancer mortality did not change much. Improvements in survival began in the 1980s in countries with early detection programs combined with different modes of treatment to eradicate invasive disease [1]. According to the latest data from the World Health Organization, the death rate from breast cancer in Uzbekistan in 2018 was 1,449. or 0.92% of the total number of deaths [1], and the incidence of the last 10 ranks first among oncological diseases. The incidence rate rose from 5.3 per 100,000 people in 1993 to 6.1 in 1998. The number of patients diagnosed with breast cancer is also increasing every year. In the Samarkand region, 154 patients with the disease were initially registered in 2000, and in 2019, their number reached 300. The fact that the incidence of breast cancer has increased dramatically in recent years, with most patients seeking medical attention in stages III-IV with a much later onset of the disease, and therefore a 5-year survival rate for patients, is even more relevant. According to the Samarkand Regional Oncology Service, the number of patients registered as primary "D" in the region in 2020 was 1863, of which 315 patients with malignant breast cancer, In 2021, the number of registered patients was 2295, and 380 were diagnosed with this dangerous disease. Breast cancer has been the number one cancer type in the region in recent years.

One of the most effective ways of skimming the structure of the breast is mammography. It classifies the mammogram images into three classes: benign, malignant, and normal. Mammogram’s image area unit is diagrammatical as terribly correct and sophisticated pictures to be taken [4-5]. Moreover, around 25% of breast cancer cases could not be detected on-screen, hence there is a need for a system that may assist radiologists in diagnosing cancer [4]. In the field of medical imaging, computer-aided design (CAD) tools are utilized to interpret images. To comprehend and interpret mammography pictures, these systems are required [4]. CAD software is applied to help radiologists by validating their findings during mammography screening [5]. The CAD system's evaluation begins with the radiologist analyzing the mammogram for abnormal patterns, and then the CAD system scans the mammogram for suspicious areas. Finally, the radiologist examines the outputs of the CAD system [6]. A CAD system for processing mammogram images includes four main steps: preprocessing, segmentation of the region of interest, feature extraction, and classification [5]:

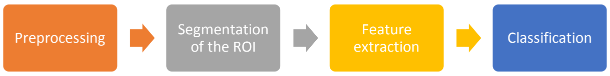

Fig. 1. A CAD system for processing of mammograms

Preprocessing of mammogram images includes noise removal, suppression of artifacts, and contrast enhancement. Image segmentation means dividing the image into categories or sub-regions that are related to different objects [15]. One of these categories corresponds to each pixel in the image. The feature extraction stage generates high-level features that are required for object categorization and anomaly detection. They can also tell the difference between healthy and unhealthy lesions [5].

SIGNIFICANCE OF THE SYSTEM

The paper mainly focuses on preprocessing and segmentation of digital mammogram images for early detection of breast cancer. Otsu's thresholding method is analyzed and reviewed. The study of literature survey is presented in section III. The methodology is explained in section IV. Section V covers the experimental results of the study, and section VI discusses the future study and conclusion.

LITERATURE SURVEY

In the literature, segmentation based on thresholding has been used to classify images, detect tumors, identify, and represent features [2,3,8,9]. One of the most important application areas is medical imaging. In [4] for breast cancer detection, they observed the effects of various thresholding methods like Otsu, Niblack, Bernsen, Thepade's Sorted Block Truncation Coding (TSBTC), and feature-level fusions of them in the identification of breast cancer at an early stage. All thresholding methods work well for normal images, but the emphasis method has given the best results for malignant images in the neighborhood valley. In [7], segmentation was carried out on 15 dermoscopy of melanoma images that were subjected to grayscaling, histogram, segmentation with Otsu Thresholding, binarization, image negation, and testing. Jelena Bozek and others [16] gave a survey of image processing algorithms that have been developed for the detection of masses and calcifications. An overview of algorithms in each step (segmentation step, feature extraction step, feature selection step, classification step) of the mass detection algorithms was given. The Wavelet detection methods and other recently proposed methods for calcification detection were presented. An overview of contrast enhancement and noise equalization methods was given, as well as an overview of calcification classification algorithms. Ayman A. and others [17] introduced a preprocessing technique for reducing the size and enhancing the quality of USF and MIAS mammogram images. The algorithm analyzed the mammogram image to determine if a 16-bit to 8-bit conversion process was required. Enhancement was applied later, followed by a scaling process to reduce the mammogram size. The performance of the algorithms was evaluated objectively and subjectively. On average, there was an 87% reduction in size with no data loss in the breast region. R.Kalyani and others [2] proposed an efficient metaheuristic algorithm, the exchange market algorithm (EMA), for multilevel thresholding (MLT) of distinct medical images. The optimal threshold is effectively obtained through the most promising objective functions such as Kapur, Otsu, and minimum cross-entropy (MCE) aided with EMA. The EMA involves the exchange of shares among investors in stable and unstable market situations to achieve a profit. Exploration and exploitation are achieved by the second and third groups of stable and unstable modes of EMA. [10] studied the workings of image thresholding based on a genetic algorithm. The proposed variant in this paper was tested on a set of images and the results are compared with the original Otsu method. A. Gurung and others [11] demonstrated that their approach outperforms the popular Otsu's method in terms of CPU computational time. They observed a maximum speed-up of 35.58 and a minimum speed-up of 10.21 on popular image processing benchmarks. The proposed algorithm in [13] employs Median filtering, Optimal Global Thresholding using Otsu's method, and morphological processing to: (1) improve the quality of mammography images; (2) segment masses based on the resulting Region of Interest (ROI); and (3) extract segmented masses from images. In order to promote the real-timeliness of image segmentation, [3] introduces the fruitfly optimization algorithm (FOA) to OTSU segmentation, creating an FOA-OTSU segmentation algorithm. In the proposed algorithm, the optimal threshold for segmentation is searched for by the FOA.

Therosholding Categories

Image segmentation is a very important phase of image processing and separates images into similar components and partitions them into ROIs (regions of interest). This process is performed before feature extraction and classification. Much of the research work done in early breast cancer identification using segmentation techniques was discussed in the literature review section. The threshold method, edge-based segmentation, region-based segmentation, and others are examples of image segmentation methods. Thresholding or image binarization separates pixels into two groups: pixel values of black pixels (background) and white pixels (foreground). In the thresholding process, a gray image is converted into a binary image.

Binarization of grayscale images is categorized into two groups: global thresholding binarization and local thresholding binarization. Global thresholding methods are much faster and produce good results for traditional grayscale images.

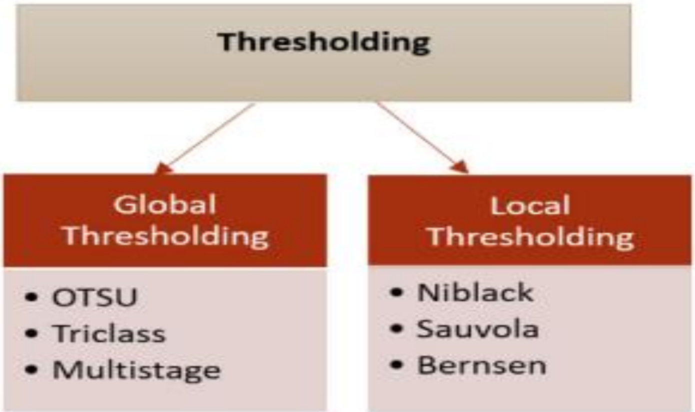

Fig. 2. Thresholding methods

Otsu, Triclass, Multistage global thresholding techniques are examples for global thresholding.

Otsu's thresholding method

The Otsu method is a global adaptive binarization threshold image segmentation algorithm. This method was proposed by Japanese scholar Nobuyuki Otsu in 1979 [14]. The image is divided into two groups; the target and the background, by searching for an optimal threshold in the range of the image's gray level. The between-class variance of the two classes is used as an evaluation criterion. The larger the variance, the greater the difference between the target and the background, and the

WEB OF SCIENTIST: INTERNATIONAL SCIENTIFIC RESEARCH JOURNAL
ISSN: 2776-0979, Volume 3, Issue 8, Aug., 2022

better the segmentation results. The smaller the variance, the greater the probability the two classes will be divided and the worse the segmentation results [10]. The technique of image thresholding, which can be thought of as an extreme type of gray-level quantization, is the most basic example of such an abstraction.

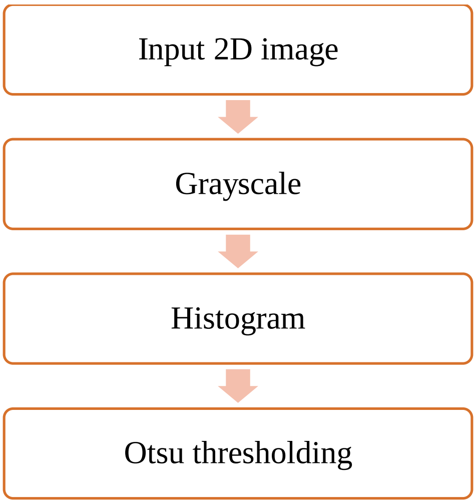

Fig. 3. Flow of Otsu's thresholding method in image processing

Suppose that the gray level is (used to be 256), and the number of pixels of gray level at i is n_i then the number of total pixels of an image is $N = \sum_{i=0}^{L-1} n_i$, the probability that each gray level appears is $p_i = \frac{n_i}{N}$. The Otsu method selects gray level k as the segmentation threshold, and divides the image into two classes: the background class named w_1 which gray level ranges from 0 to k, the target class named w_2 which gray level ranges from $k+1$ to $L-1$ [10]. The gray level probability distributions for the two classes are given as:

$$P_{w1} = \sum_{i=0}^{k} p_i ; \qquad (1)$$

$$P_{w2} = \sum_{i=k+1}^{L-1} p_i = 1 - P_{w1} . \qquad (2)$$

The mean gray level of w_1 and w_2 are:

$$\mu_{w1} = \sum_{i=0}^{k} \frac{i \cdot p_i}{P_{w1}} ; \qquad (3)$$

$$\mu_{w2} = \sum_{i=k+1}^{L-1} \frac{i \cdot p_i}{P_{w2}} . \qquad (4)$$

The mean of total gray levels is denoted by μ:

$$\mu = P w_1 \mu_{w1} + P_{w2} \mu_{w2} = \sum_{i=0}^{L-1} i \cdot p_i . \qquad (5)$$

The between-class variance is:

$\sigma^2(k) = P_{w1}(\mu_{w1} - \mu)^2 + P_{w2}(\mu_{w2} - \mu)^2$.   (6)

Otsu's method chooses the optimal threshold k by maximizing the between-class variance $\sigma^2(k)$. The larger the variance, the better the image segmentation. The optimal threshold T^* for the perfect image segmentation is given as:

$T^* = \arg\max(\sigma^T(T)); \quad 0 \le T \le L-1$.   (7)

Methodology

The technique of image thresholding, which can be thought of as an extreme type of gray-level quantization, is the most basic example of such an abstraction. Assume that a gray-level image f has g different gray levels to choose from: $0,1,2,\dots,k-1$. Define an integer threshold, T, that falls inside the gray-scale range of $(0,1,2,\dots,k-1)$. Thresholding is a simple comparison process in which each pixel value in f is compared to the threshold, T. A binary decision is produced based on this comparison, which determines the value of the relevant pixel in an output binary picture g. Iterating over all conceivable threshold values and calculating a measure of spread for the pixel levels on each side of the threshold, i.e., the pixels that either fall in the foreground or background, is how Otsu's thresholding approach works. The goal is to achieve the smallest total threshold value for foreground and background spreads. The Otsu thresholding method is used in image processing for automatic binarization level determination depending on the structure of the histogram. The algorithm considers the image to be divided into two main classes: foreground and background. It then calculates an ideal threshold value that minimizes these two classes' weighted within-class variances. According to mathematics [6], minimizing within-class variance is the same as maximizing between-class variance.

If $g(x,y)$ is a thresholded version of $f(x,y)$ at some global threshold T:

$$g(x,y) = \begin{cases} 1, & if\ f(x,y) \ge T; \\ 0, & otherwise. \end{cases} \qquad (1)$$

Here we explained the algorithm of Otsu's thresholding method.

Step 1. Computed histogram for a two-dimensional(2D) image.

Step 2. For a single threshold, calculate the foreground and background variances (measure of spread).

i) Calculate weight of background pixels and foreground pixels.

ii) Calculate mean of background pixels and foreground pixels.

iii) Calculate variance of background pixels and foreground pixels.

Step 3. Calculate "within class variance".

WEB OF SCIENTIST: INTERNATIONAL SCIENTIFIC RESEARCH JOURNAL
ISSN: 2776-0979, Volume 3, Issue 8, Aug., 2022

EXPERIMENTAL RESULTS

This section presents the results obtained for the Otsu thresholding algorithm. The result of the segmented masses for three mammogram images is shown in fig. 4.

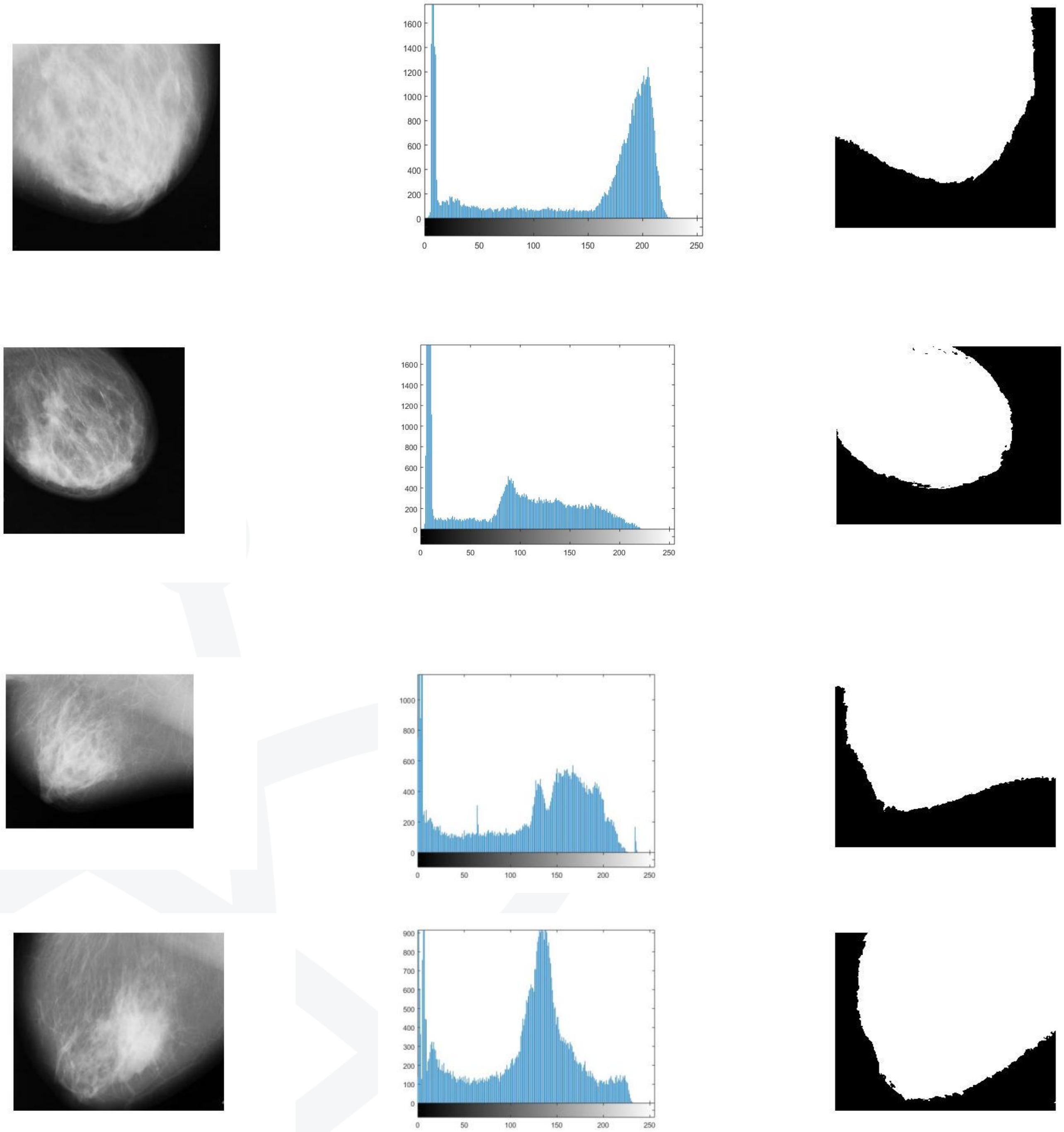

Fig. 4. Results of mammogram images after Otsu thresholding

CONCLUSION AND FUTURE WORK

The literature review will be valuable to anyone doing research in this field. From the perspective of discriminant analysis, a method for automatically selecting a threshold from a gray-level histogram has been developed. This immediately addresses the issue of determining the usefulness of thresholds. The discriminant criteria select an appropriate threshold (or combination of thresholds) by maximizing the discriminant measure q. (is the gray-level measure of the resultant classes' separability). Taking these considerations into account, the approach proposed in this correspondence may be recommended as the most basic and standard way for automatic threshold selection that can be used for a wide range of actual issues. In conclusion, the method's results suggest that it is a reliable strategy, albeit it could be improved in terms of accuracy. Even so, we analyzed this method because it provides useful regions (there is no meaningful loss of information)

REFERENCES

1. World Health Organization. www.who.org
2. R.Kalyani, P.D.Sathya,V.P.Sakthivel. “Medical image segmentation using exchange market algorithm” 2021. Alexandria Engineering Journal 60, 5039-5063.
3. Ch.Huang, X.Li, Y.Wen. “An Otsu image segmentation based on fruitfly optimization algorithm”2021. Alexandria Engineering Journal 60,183-188.
4. S.Khairnar,S.D.Thepade,Sh.Gite. “Effect of image binarization thresholds on breast cancer identification in mammography images using Otsu,Niblack, Burnsen, Thepade’s SBTC” 2021. Intelligent Systems with Applications. 10-11.
5. N.F.Soliman,N.S.Ali, M.I.Aly, A.D.Algarni, W.El-Shafai, F.E.Abd El-Samie. “An efficient breast cancer detection framework for medical diagnosis applications”CMC,2022. vol.70, no.1,
6. S.L.Bangare,S.Patil. “Reviewing Otsu’s method for image thresholding”.International Journal of Applied Engineering Research.2015.
7. P.A.Pitoy, P.G.H.Suputra. “Dermoscopy image segmentation in melanoma skin cancer using Otsu thresholding method”.Journal of Elektronik Ilmu Komputer Udayana. Vol.9, no.3, 2021.
8. Fazilov Sh.Kh and O. R. Yusupov, "Algorithm for Extraction of the Iris Region in an Eye Image," IEEE East-West Design & Test Symposium (EWDTS), Batumi, Georgia, 2019, pp. 1-5, doi: 10.1109/EWDTS.2019.8884441.
9. Fazilov Sh.Kh and O. R. Yusupov, " Improvement of the Daugman Method for Nonreference Assessment of Image Quality in Iris Biometric Technology," IEEE

East-West Design & Test Symposium (EWDTS), Batumi, Georgia, 2021, pp. 1-5, doi: 10.1109/EWDTS52692.2021.9581040.

10. W.Shang, Y.Cheng. "An improved Otsu method based on Genetic algorithm",Advances in Computer Science Research,vol.71, 2016.
11. A.Gurung, S.L.Tamang. "Image Segmentation using multi-thresholding thechnique by histogram sampling ", 2019.
12. A.Kumar,A.Tiwari. "A comparative study of Otsu thresholding and k-means algorithm of image segmentation ".International Journal of Engineering and Technical Research(IJETR), vol.9, issue-5,2019.
13. Sh.Thawkar, Dr.R.Ingdikar. "Segmentation of masses in digital mammograms using optimal thresholding with Otsu's method".IJCST, vol.5, issue 3, 2014.
14. N.Otsu. "A threshold selection method from gray level histogram". IEEE Trans.system.Man. and Cybernetics. 9:62-66,1979.
15. O.R.Yusupov, Kh.S.Abdiyeva, A.Primov. "Preprocessing and segmentation of digital mammogram images for early detection of breast cancer".IJARSET, vol.8, issue 9, 2021.
16. Ayman A. AbuBaker, R.S.Qahwaji, Musbah J. Aqel, Hussam Al-Osta and Mohmmad H. Saleh.2006. Efficient Pre-processing of USF and MIAS Mammogram Images Journal of Computer Science 3 (2): 67-75
17. Sara Dehghani, Mashallah Abbasi Dezfooli. A Method For Improve Preprocessing Images Mammography International Journal of Information and Education Technology, Vol. 1, No. 1, April 2011.