

## COMPARISON OF HISTOGRAM FEATURE BASED THRESHOLDING WITH 3S MULTI-THRESHOLDING AND FUZZY C-MEANS

Mostafa Langarizadeh<sup>1\*</sup>, Rozi Mahmud<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran.

<sup>2</sup>Professor, Department of Medical Imaging, Faculty of Medicine and Health Sciences, University Putra Malaysia, Kuala Lumpur, Malaysia.

Article Info	ABSTRACT
<p><b>Article type:</b> Research</p> <hr/> <p><b>Article History:</b> Received: 2019-05-24 Revised: - Accepted: 2019-07-27</p> <hr/> <p><b>* Corresponding author:</b> Mostafa Langarizadeh Assistant Professor, Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran. Email: langarizadeh.m@iums.ac.ir</p>	<p><b>Introduction:</b> Thresholding is one of the most important parts of segmentation whenever we want to detect a specific part of image. There are several thresholding methods that previous researchers used them frequently as bi-level techniques such as DBT or multilevel such as 3S. New histogram feature thresholding method is implemented to detect lesion area in digital mammograms and compared with 3S (Shrinking-Search-Space) multi-thresholding and FCM method in terms of segmentation quality and segmentation time as a benchmark in thresholding.</p> <p><b>Material and Methods:</b> These algorithms have been tested on 188 digital mammograms. Digital mammogram image used after preprocessing which was including crop the unnecessary area, resize the image into 1024 by 1024 pixel and then normalize pixel values by using simple contrast stretching method.</p> <p><b>Results:</b> The results show that suggested method results are not similar with 3S and FCM methods, and it is faster than other methods. This is another superiority of suggested method with respect to others. Results of previous studies showed that FCM is not a reliable clustering algorithm and it needs several run to give us a reliable result. Results of this study also showed that this approach is correct.</p> <p><b>Conclusion:</b> The suggested method may use as a reliable thresholding method in order to detection of lesion area.</p> <p><b>Keywords:</b> FCM, Image Processing, Segmentation, Thresholding.</p>

### How to cite this paper

Langarizadeh M, Mahmud R. Comparison of Histogram Feature Based Thresholding with 3S Multi-Thresholding and Fuzzy C-Means. *Front Health Inform.* 2019; 8(1): e15. DOI: [10.30699/fhi.v8i1.195](https://doi.org/10.30699/fhi.v8i1.195)

## INTRODUCTION

Thresholding is one of the most important parts of segmentation whenever we want to detect a specific part of image. There are several thresholding methods that previous researchers used frequently as bi-level techniques such as DBT [1] or multilevel such as 3S [2]. The bi-level techniques also could be used as multilevel, since just the optimum search space should be found. The 3S multilevel thresholding method has been applied on MRI images and results showed that it is faster and more reliable in comparison with FCM.

C-means and Fuzzy C-means are two types of the earliest unsupervised learning methods in clustering. Each data will be categorized under a cluster based on the similarity criterion when c-means or fuzzy c-means are used.

According to previous researches several methods developed to produce effective multilevel thresholding techniques [3-8]. In this study a new thresholding method is suggested and compared with FCM and 3S in terms of detection of lesion area in digital mammograms. Since in digital mammograms, there are different tissues (e.g. fatty, dense, mass etc.), almost all bi-level thresholding

methods are not useful to separate different areas. Using bi-level methods leads radiologists to miss malignancy and reduce sensitivity in terms of early detection of breast lesions. Therefore, an optimum thresholding method is highly necessary to help radiologists as a second reader in order to select lesion area.

The main aim of this research was to develop a new thresholding method in order to detect lesion area in digital mammograms. The other objective of this study was to compare new suggested method with two other common methods FCM and 3s.

### C-Means and Fuzzy C-means

The C-Means method is suggested by McQueen in 1967 and it was developed during last decades. This method works based on the number of clusters that whole data divided to. A center point will be identified for each cluster in its center. Each data will be belonging to each cluster if its distance to the center point is less than other clusters [9].

At the first step the initial value of center points (there are k center points) will be defining. The number of center points defines based on the user knowledge on data distribution. At the second step all data cluster in terms of their distance to the center points.

At the next steps, calculation of center points and rearrange of clustering based on the new points frequently repeat until the objective function does not change anymore. Objective function is defined as equation (1).

$$obj\_func = \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2 \tag{1}$$

Where  $x_i^j$  is the  $i^{th}$  data point that has been recognized to belong to the  $j^{th}$  cluster with the center point  $c_j$  [2].

So the C-Means method steps could be summarized as: a) place the center points as representative of clusters b) assign data to closest cluster c) assign labels to clusters and d) repeat steps b and c until no change in objective function results. There are some limitations in using C-Means method since it needs to run several time and changing defined central point's effect on the results.

Fuzzy C-Means (FCM) is developed based on the same idea however there is a number of differences here. The most important difference between C-Means and FCM is in the clustering method. In C-Means method each data could be belong to just one cluster however in FCM each data may categorized under more than one cluster based on its membership degree [10-12]. This method which is

used by many researchers [13, 14] worked based on the following objective function:

$$obj\_func_{FCM} = \sum_{j=1}^k \sum_{i=1}^n u_{ij}^m \|x_i - c_j\|^2 \text{ for } 1 < m < \infty \tag{2}$$

Where n is the number of data points, and m is the fuzzyness degree, which is any real number greater than 1,  $x_i$  is the  $i^{th}$  data point,  $c_j$  is the center point of the  $j^{th}$  cluster, and  $u_{ij}^m$  is the membership degree of the  $i^{th}$  data point to the  $j^{th}$  cluster.

The membership and center points will be carried out based on the following equations:

$$u_{ij}^m = \frac{1}{\sum_{m=1}^k \left( \frac{\|x_i - c_j\|}{\|x_i - c_m\|} \right)^m} \tag{3}$$

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \tag{4}$$

However this method is more optimize than C-Means, it is still depending on the number of clusters and predefined centers [15, 16].

### 3S method

This method is built based on DBT method which is a bi-level thresholding technique [1]. The DBT as a bi-level thresholding method tries to compute total average information for discriminating class  $C_0$  from class  $C_1$ . Discriminating information for class  $C_1$  versus class  $C_0$  can be measured by using the logarithm of the likelihood ratios.

$$J(C_0, C_1) = \frac{1}{w_0} \ln \frac{w_1}{w_0} + \frac{1}{w_1} \ln \frac{w_0}{w_1} \tag{5}$$

Where  $w_0$  and  $w_1$  are the probability of occurrence of classes  $C_0$  and  $C_1$  respectively [1]. To get thresholding value T, to make separation between object and background, the result of the function need to be minimized. As it shown earlier, DBT as a bi-level method is useful to apply on images that there is just one object as well as text and background. However in the other situations such as digital mammograms we need to use multilevel thresholding methods. So the 3S method is resulted from extension of DBT.

3S method works based on the following steps (Fig 1):

- Identify the region of interest (ROI) that fits to whole image.
- Perform DBT to find the best thresholding value.

- Save class  $C_0$  and continue with  $C_1$  as the original image.
- Class  $C_1$  (as original image) divide into two class  $C_1$  and  $C_2$ .
- The procedure will be repeated until reach to highest value of histogram.

During above steps different optimum thresholding values will be found [2].

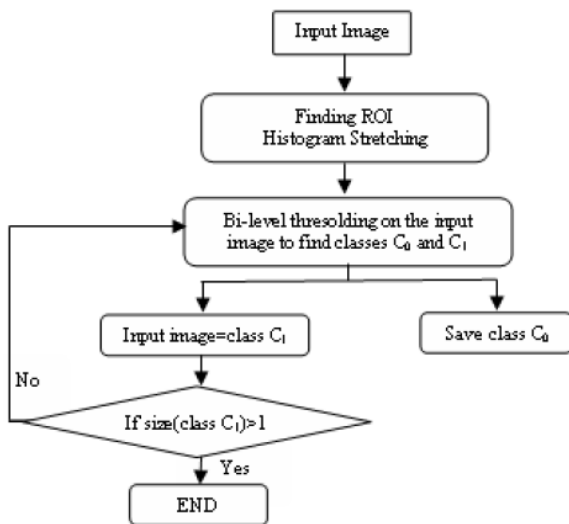


Fig 1: 3S working flowchart [2]

## MATERIALS AND METHODS

The suggested method is built based on a bi-level thresholding technique as the core of a multi-thresholding technique. Block diagram (Fig 2) shows suggested method architecture. As clearly shown in Fig 2, first of all the original images were opened. Then, one of the original images (right or left) will be rotated to make same direction and a subtraction image of right and left side will be produced. At the next step, statistical histogram features including mean and standard deviation (SD) were calculated. In normal distribution, from  $-\infty$  up to mean is including 50% of under the curve area and also mean plus one SD is including 42.5% of under the curve area. Therefore, from  $-\infty$  up to mean plus 1 SD cover 92.5% of the under the curve area. This is proven previously that the lesions are in the last 10 percentile of pixel values [13]. Because of these reasons, thresholding value to extract the lesion area in digital mammograms is equal to Mean +SD. All pixel values which are equal to or greater than this criterion will be changed to 1 and the rest changed to 0. Based on this procedure a binary image will be produced that just is containing the objects. Thresholding is highly depends on pixel values. Since preprocessing phase specially filtering make changes in pixel values, thus in this research the filtering methods are not employed and original images were used for system evaluation.

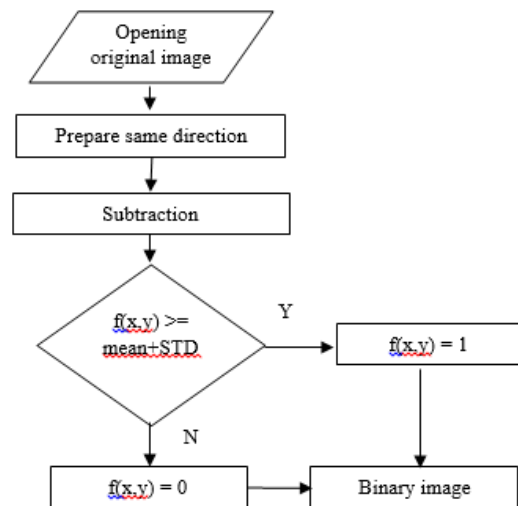
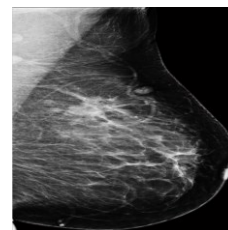


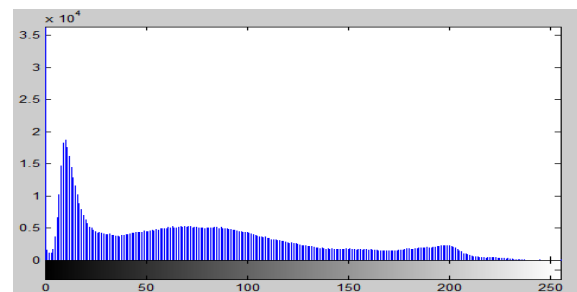
Fig 2: Architecture of the proposed method

## RESULTS

Detection processes for FCM, 3S and the suggested method was run on the same computer with Intel core 2 duo processor (1.88 Mhz) and 2 GB RAM. All algorithms have been tested using 188 digital mammograms after preprocessing, which was including crop the unnecessary area to select an ROI in size of 1024 by 1024 pixel and then normalize pixel values using simple contrast stretching method. The contrast stretching method is employed since the proposed method worked based on normal distribution (Fig 3).



(a)

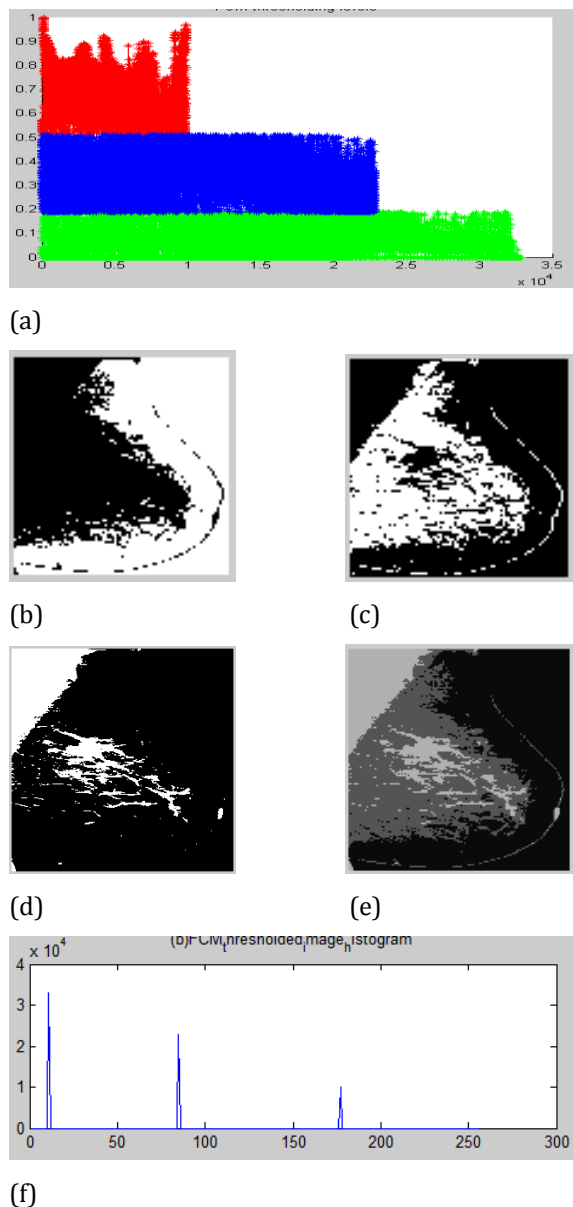


(b)

Fig 3: (a) Gray scale original image (b) Histogram of the gray scale image

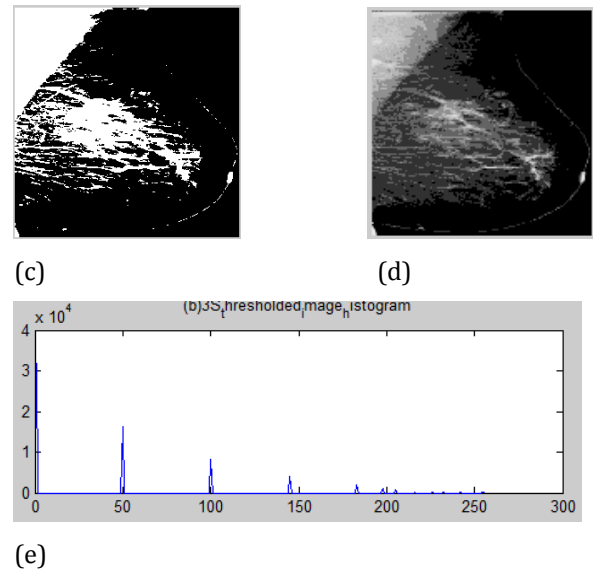
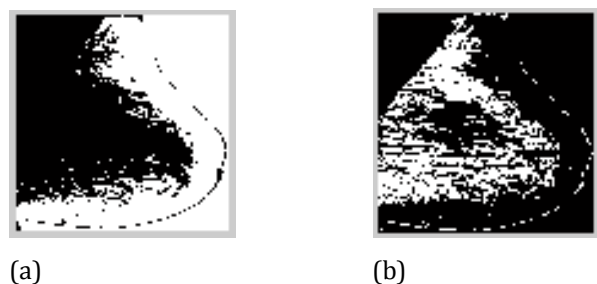
Threshold levels that FCM algorithm recognizes were shown after 235.47 seconds as average of processing time. We have defined three clusters for this algorithm, so gray levels of the input image have

been separated to three groups (Fig 4).



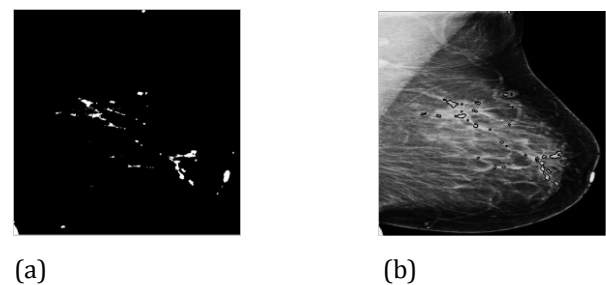
**Fig 4: (a) Clusters detected by FCM (b, c, d) Threshold images (e) All areas detected by FCM (f) Threshold values detected by FCM method**

3S multi-thresholding took an average of 33.896 seconds for processing. Comparing with the results obtained from FCM method, there was not equal quality of thresholding between FCM and 3S, and also there is different processing time (Fig 5).



**Fig 5: (a,b,c,d) Thresholded images (e) Threshold values detected by 3S method**

The result of proposed histogram feature based thresholding method was clearly shown that using proposed method is helpful since it is less complex than other methods, less time consuming (21.45 seconds) and it selected correct area including lesion (Fig 6).



**Fig 6: (a) Thresholded area (b) Selected area adjusted on original image**

Three expert radiologists with at least 5 year experience in mammogram interpretation were invited to select lesion area. The performance of three methods in terms of correct detection of lesion area compared with radiologists idea. The results show that the FCM had the lowest performance in comparison with two other methods. It means that the performance was lower and the processing time was much higher than other methods. While 3S and proposed histogram feature based methods had almost similar results however the processing time of 3S was higher than our proposed method. It was because of frequently process that 3S did (Table 1).

## DISCUSSION

We have implemented the FCM and 3S algorithm on the same data, to compare their functionality with our proposed method. The results showed that FCM

is a time consuming method since it should repeat all steps to get better result [17, 18]. In the other side 3S had better results since it shows higher correct answer in terms of detection of breast lesions. In comparison with both FCM and 3S methods, the proposed method had better performance because it takes less time for processing and give us more correct answers. This showed that our proposed method is an acceptable method in comparison with 3S and FCM in terms of processing time and accurate detection of lesions.

**Table 1: Results of different methods**

	Time (seconds)	Correct detection (%)
FCM	235.470	153 (81)
3S	33.896	174 (92.55)
Suggested system	21.450	179 (95.21)

## CONCLUSION

It could be concluded that such systems are useful in order to help radiologists in terms of detection of lesions. This research area is open yet and improvement in results is needed.

## AUTHOR'S CONTRIBUTION

All the authors approved the final version of the manuscript.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

## FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

## REFERENCES

- Chowdhury MH, Little WD, editors. Image thresholding techniques. Pacific Rim Conference on Communications, Computers, and Signal Processing. IEEE; 1995.
- Mortazavi D, Mashohor S, Mahmud R, Jantan AB, editors. Comparison of 3S multi-thresholding with fuzzy C-means method. Innovative Technologies in Intelligent Systems and Industrial Applications IEEE; 2009.
- Fan SKS, Lin Y. A multi-level thresholding approach using a hybrid optimal estimation algorithm. Pattern Recognition Letters. 2007; 28: 662-9.
- Liao PS, Chen TS, Chung PC. A fast algorithm for multi-level thresholding. Journal Information Science Engineering. 2001; 17: 713-27.
- Huang DY, Wang CH. Optimal multi-level thresholding using a two-stage Otsu optimization approach. Pattern Recognition Letters. 2009; 30: 275-84.
- Yen-Lin C, Hsin-Han C, Chuan-Yen C, Chuan-Ming L, Shyan-Ming Y, Jenq-Haur W. A vision-based driver night time assistance and surveillance system based on intelligent image sensing techniques and a heterogeneous dual-core embedded system architecture. Sensors (Basel). 2012; 12(3): 2373-99. PMID: 22736956 DOI: 10.3390/s120302373 [PubMed]
- Kockara S, Mete M, Chen B, Aydin K. Analysis of density based and fuzzy C-means clustering methods on lesion border extraction in dermoscopy images. BMC Bioinformatics. 2010; 11(6): S26. PMID: 20946610 DOI: 10.1186/1471-2105-11-S6-S26 [PubMed]
- Tay PC, Acton ST, Hossack JA. A wavelet thresholding method to reduce ultrasound artifacts. Comput Med Imaging Graph. 2011; 35(1): 42-50. PMID: 20934848 DOI: 10.1016/j.compmedimag.2010.09.003 [PubMed]
- Wu S, Pang Y, Shao S, Jiang K. Advanced fuzzy C-means algorithm based on local density and distance. Journal of Shanghai Jiaotong University (Science). 2018; 23(5): 636-42.
- Bezdek JC. Pattern recognition with fuzzy object function algorithms. Norwell, MA: Kluwer; 1981.
- Zhuge Y, Cao Y, Udupa JK, Miller RW. Parallel fuzzy connected image segmentation on GPU. Med Phys. 2011; 38 (7): 4365-71. PMID: 21859037 DOI: 10.1118/1.3599725 [PubMed]
- Yang X, Baowei F. A multiscale and multiblock fuzzy C-means classification method for brain MR images. Med Phys. 2011; 38(6): 2879-91. PMID: 21815363 DOI: 10.1118/1.3584199 [PubMed]
- Oliver A, Freixenet J, Marti R, Pont J, P'erez E, Denton ERE, et al. A novel breast tissue density classification methodology. IEEE Transactions on Information Technology in Biomedicine. 2008; 12(1): 55-65.
- Oliver A, Freixenet J, Zwiggelaar R, editors. Automatic classification of breast density. International Conference on Image Processing. IEEE; 2005.
- Keller B, Nathan D, Wang Y, Zheng Y, Gee J, Conant E, et al. Adaptive multi-cluster fuzzy C-means segmentation of breast parenchymal tissue in digital mammography. Med Image Comput Comput Assist Interv. 2011; 14(3): 562-9. PMID: 22003744 [PubMed]
- Saleck MM, ElMoutaouakkil A, Mouçouf M, editors. Tumor detection in mammography images using fuzzy C-means and GLCM texture features. 14<sup>th</sup> International Conference on Computer Graphics, Imaging and Visualization. 2017.
- Ahmadi K, Karimi A, Fouladi Nia B. New technique for automatic segmentation of blood vessels in CT scan images of liver based on optimized fuzzy C-means method. Comput Math Methods Med. 2016; 2016: 1-8.
- Singh AK, Gupta B. A novel approach for breast

cancer detection and segmentation in a mammogram. *Procedia Computer Science*. 2015; 54: 676-82.