Asian Journal of Engineering and Applied Technology
ISSN: 2249-068X (P) Vol. 7 No. 2, 2018, pp. 31-36
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ajeat-2018.7.2.1015

Self-Organization Map Based Segmentation of Breast Cancer

A. Arokiyamary Delphina[1], M. Kamarasan[2*] and S. Sathiamoorthy[3]
[1]Research Scholar, [2&3]Assistant Professor, [1,2&3]Division of Computer and Information Science,
Annamalai University, Annamalai Nagar, Tamil Nadu, India
[*]Corresponding Author
E-Mail: smkrasan@yahoo.com

(Received 30 May 2018; Revised 23 June 2018; Accepted 10 July 2018; Available online 18 July 2018)

***Abstract* -Breast cancer is second major leading cause of cancer fatality in women. Mammography prevails best method for initial detection of cancers of breast, capable of identifying small pieces up to two years before they grow large enough to be evident on physical testing. X-ray images of breast must be accurately evaluated to identify beginning signs of cancerous growth. Segmenting, or partitioning, Radio-graphic images into regions of similar texture is usually performed during method of image analysis and interpretation. The comparative lack of structure definition in mammographic images and implied transition from one texture to makes segmentation remarkably hard. The task of analyzing different texture areas can be considered form of exploratory report since priori awareness about number of different regions in image is not known. This paper presents a segmentation method by utilizing SOM.**

***Keywords:* Breast Cancer, Mammography, Self-Organizing Map, Euclidean Distance, Validity Measure, Double Bouldin Index**

I. INTRODUCTION

According to USA Cancer Society, breast cancer is in second place as most common type of cancer afflicting women but remains leading cause of cancer mortality in women between ages of 40 and 55. Recently, in United States, approximately 180,200 women diagnosed with invasive breast cancer. Meanwhile same year, about 44,190 women will lose fight against this life-threatening disease [1]. Although proportion of new breast cancer rose on average 4 percent between years 1982 and 2017, percentage rate has tapered off to just over one percent in years since. Much of this welcome decrease in new breast cancer diagnoses have been attributed to increased use of mammography to detect early stages of this disease. Although significant prediction technique has been made in technology of mammography, much work remains to be done to improve overall detection accuracy [2].

Segmentation is process of partitioning image into multiple fragments [3]. All pixels in region are similar to some characteristic, such as intensity, color, etc. Artificial neural networks [4] are parallel computational models, consists of densely interconnected adaptive processing units. An vital feature of these networks is that they learn by example.

The adaptive nature of artificial neural networks makes it more suitable for applications where one has small or partial understanding of problem to be resolved but where training data is readily available.

II. SELF-ORGANIZING MAP (SOM)

A Self-Organizing Map is used to project high dimensional data on to two-dimensional map [5]. The dimensional reduction could allow us to visualize important relationship among data more easily. The topology structure property which is observed in brain is also noted in SOM which is not observed in any other artificial neural network. It is said to be topology preserving since it preserves neighborhood relation of input pattern. The units that are physically located next to each other will react to classes of input vectors that are likewise located next to each other.

The basic SOM model consists of input layer and output layer. SOM network consists of neurons which are similar to neurons in brain. Each input is fully connected to all units. The number of neurons in output layer depends on number of clusters in image to be segmented, i.e., number of clusters is equal to number of output neurons. Color is one of essential features used for image segmentation. SOM is used to map patterns in three-dimensional color space to two-dimensional space. SOM learns through competition. For each input vector, only one neuron in network will respond.

This mechanism is known as competition. Once neuron is elected as winner, weights of that neuron and neurons in neighborhood of victor are updated. The neighborhood scheme for SOM may be rectangular, hexagonal or circular. The multicomponent values are given as input for training. Initially, learning rate is set to 0.1, and neighborhood size is initialized to maximum of either height or width of network divided by two. The weight vectors of neurons are initialized randomly. For every iteration, input vectors to be clustered are presented to network in random order. The neurons with weight vector that best match input vector is elected as winner or best matching unit (BMU). The winner is elected by using Euclidean distance method which is as follows.

$$\left\|x - W_l^{[k]}\right\| = \overset{min}{i}\left\|x - W_i^{[k]}\right\| \qquad (1)$$

where x is input vector, W is weight of winning unit i at each iteration k. The winning neuron and neurons within neighborhood of winning unit are updated in such way that their weights become closer to input vector being presented to network. The weights are updated as follows.

$$W_i^{k+1} = W_i^k + H_{li}^k \,(\mathrm{x} - W_i^k) \qquad (2)$$

where H is smoothing kernel defined over winning neuron. The kernel can be written concerning Gaussian function as

$$H_{li}^k = \alpha^k exp\left(-\frac{d^2(l,i)}{2(\sigma^k)^2}\right) \qquad (3)$$

where d is distance between winning neuron and neuron i and σ is neighborhood distance, and σ k is learning rate at iteration k. The learning rate and neighborhood size are updated after each iteration. As number of iterations increases, learning rate and neighborhood decreases. The learning rate is exponentially reduced as follows

$$\alpha^k = \alpha^0 \exp\left(-\frac{k}{T}\right) \qquad (4)$$

where σ 0 is initial learning rate, and T is total number of iterations which is set to 1000. The decreasing function for neighborhood is given as follows

$$\sigma^k = \sigma^0 \left(1 - \frac{k}{T}\right) \qquad (5)$$

where σ^0 is initial neighborhood size and σ^{k} is neighborhood size at iteration k. The size of neighborhood is decreased until it encompasses single unit.

Once SOM converges, input is mapped from high color space to two-dimensional map. The final result of SOM depends on initial values of weights, data used for training, and characteristics of map such as some nodes in network, learning rate, and neighborhood.

SOM suffers from drawback of over-segmentation. So, optimization method like genetic algorithm is used to identify optimal number of clusters. The data set identified from SOM is given as input to optimization method for identifying cluster centers.

III. PROPOSED FRAMEWORK FOR SEGMENTATION OF BREAST CANCER IMAGES

The following fig. 1 represents proposed framework for classification of breast cancer cell using Pre-processing step, Segmentation step. In this work, noise of image is removed by using median filtering method in pre-processing step; SOM is used in segmentation step.

A. Pre-Processing with Median Filter

Noise is unwanted signal in image. The noises in image are of three types Salt and Pepper noise, Impulse noise, Gaussian noise. Median Filter operates over window by selecting median intensity in window, The advantages of using Median filter is of its robust average, that is its unrepresentative pixel in neighborhood does not affective median value and also it has quality of preserving sharp edges. The median filter works through image by pixel by pixel and replaces it with median value of neighboring pixels and pattern of neighbors is called window which slides pixel by pixel over entire image.

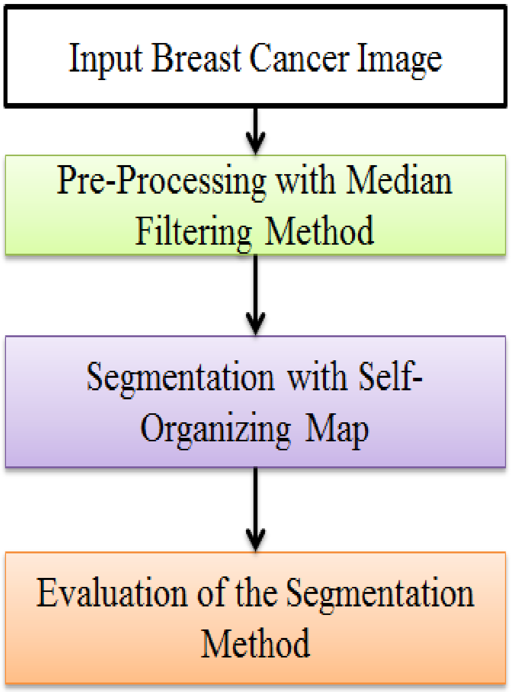

Input Breast Cancer Image
Pre-Processing with Median Filtering Method
Segmentation with Self-Organizing Map
Evaluation of the Segmentation Method

Fig. 1 Proposed Framework for SOM based Segmentation of Breast Cancer Image

B. Segmentation Using SOM

Clustering process in this study uses gray scale values of each pixel as input to SOM method. Neighborhood topology which is utilized in SOM approach in this work is linear array or also named as one dimensional (1-D) topology. Calculation of SOM approach is split into two stages, stage of learning, and stage of recognition. In this research, to determine distance does not utilize Euclidean Distance, but it utilize Normalized Euclidean Distance.

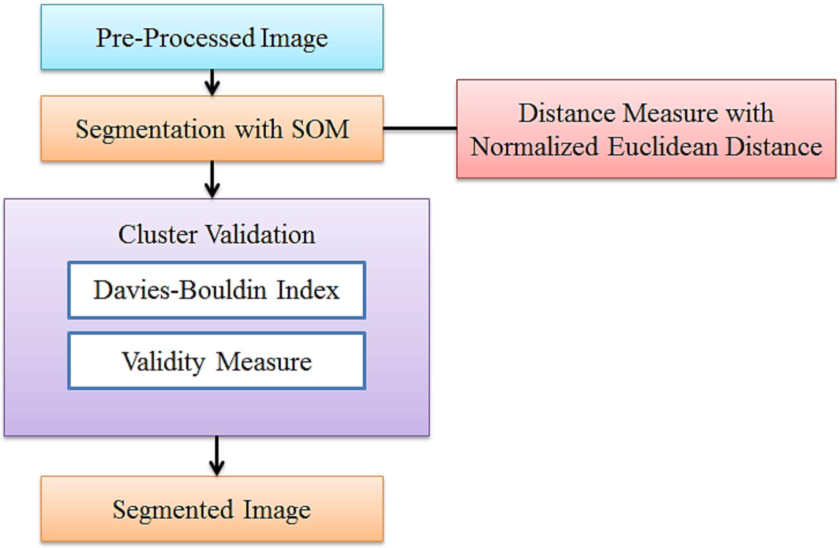


Fig. 2 Segmentation Process of Self-Organizing Map

1. *Normalized Euclidean Distance:* The computation of Normalized Euclidean Distance is modified form of Euclidean Distance [8]. Normalized Euclidean Distance of two vectors, between vector *u* and vector *v* is shown by following equation

$$d_{(uv)} = \sqrt{\sum_{k=1}^{n} (\bar{u} - \bar{v})^2}$$

Where

$$\bar{u_i} = \frac{u_i}{\|u\|}, \qquad \bar{v_i} = \frac{v_i}{\|v\|}$$

||v|| is normalized value of vector v. The normalized value is expressed in following equation

$$\|v\| = \left[\sum_{i=1}^{n} v_i^{\,2} \right]^{\frac{1}{2}}$$

2. Cluster Validation

*a) Davies-Bouldin Index (DBI):*DBI was preceded in 1979 by David L. Davies and Donald W. DBI is applied to evaluate clustering results [9]. DBI is method to measure ratio of total within-cluster scatter (spread of cluster) and between-cluster separation (distance between clusters). Below Equation is used to calculate spread of cluster value.

$$S_i = \frac{1}{T_i} \sum_{x \in C_i} \|x - z_i\|$$

Where T_i is number of member in i^{th}cluster (C_i), and z_i is i^{th} cluster center. The distance between clusters is calculated by Euclidean distance between center of i^{th} cluster and center of j^{th} cluster. Following equation is used to calculate its distance.

$$d_{ij} = \|z_i - z_j\|$$

R_{ij} is ratio value between i^{th} cluster and j^{th} cluster, which is calculated by following formula.

$$R_{ij} = \left\{ \frac{S_i + S_j}{d_{ij}} \right\}$$

Finding maximum value of ratio (D_i), it is used to find value of DBI. Following equation is used to calculate value of $D_{i.}$

$$D_i = \max_{j:j \neq i} R_{ij}$$

Then, DBI value is calculated by using Equation.

$$DBI = \frac{1}{K} \sum_{i=1}^{K} D_i$$

where k is number of clusters. DBI with most minimum value indicates most optimal clustering results and achieve well-separated cluster.

b) Validity Measure (VM): VM is one of indexes to test validity of clustering results [10]. VM is commonly used in application of image segmentation based on clustering. VM is calculated using below given equation.

$$VM = y\left(\frac{intra}{inter} \right)$$

where intra is intra-cluster distance, inter is inter-cluster distance, and y is function of number of clusters that is formed. Below equation is used to find value of intra-cluster distance.

$$intra = \frac{1}{N} \sum_{i=1}^{k} \sum_{x \in C_i} \|x - z_i\|^2$$

where N is total number of pixels in image, k is number of clusters, and zi is center of cluster Ci.

Also, to calculate VM, takes minimum value of inter-cluster distance [10]. Below Equation is used to find inter-cluster distance.

$$inter = \min\left(\left\| z_i - z_j \right\|\right)$$

Where i = 1,2 … k, and j = i+1, …, k. y is multiplied by quotient between distance of intra-cluster and inter-cluster. Below equation is utilized to calculate y.

$$y = c.N(2,1) + 1$$

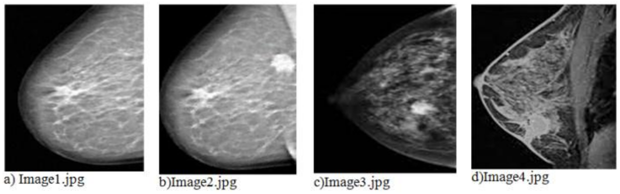
where c is constant value in range of 15 to 25, N (2,1) is Gaussian function for number of clusters (k). The Gaussian function is shown in Equation

$$N(\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{(k-\mu)^2}{2\sigma^2}\right]}$$

VM should be minimum to obtain optimal result and achieve well-separated cluster.

IV. RESULT AND DISCUSSION

Following fig. 3 presents breast cancer images considered for segmentation process whereas a) Image1.jpg, b) Image2.jpg, c) Image3.jpg, d) Image4.jpg and table I gives initialization of parameters on SOM method and spatial operations



a) Image1.jpg b)Image2.jpg c)Image3.jpg d)Image4.jpg

Fig. 3Breast Cancer Image for Segmentation process a) Image1.jpg b) Image2.jpg c) Image3.jpg d) Image4.jpg

TABLE I INITIALIZATION OF PARAMETERS ON SOM METHOD

S. No.	Parameter	Value
1	A	0.2
2	Epoch (T)	200
3	Radius of noise filter	3
4	Threshold Region (ThA)	0.003 * total pixels

Tests are performed on each image by forming 2 until 10 clusters. From each cluster which is created, then compute value of Validity Measure (VM) and Davies-Bouldin Index (DBI). Each value of VM and DBI with minimum value showed most optimal number of cluster. Results of validity measurement for each test images are shown in following table II – table V.

TABLE II VALIDITY MEASURE AND DAVIS-BOULDIN INDEX FOR GIVEN BREAST CANCER IMAGE1.JPG BASED ON NUMBER OF CLUSTERS

Number of Clusters	Image1.jpg	
	Validity Measure	Davies-Bouldin Index
2	4.98	2.710
3	4.556	3.632
4	7.851	1.662
5	4.470	2.937
6	4.954	2.380
7	4.819	3.100
8	8.661	3.789
9	8.199	3.982
10	46.11	4.150

TABLE III VALIDITY MEASURE AND DAVIS-BOULDIN INDEX FOR GIVEN BREAST CANCER IMAGE2.JPG BASED ON NUMBER OF CLUSTERS

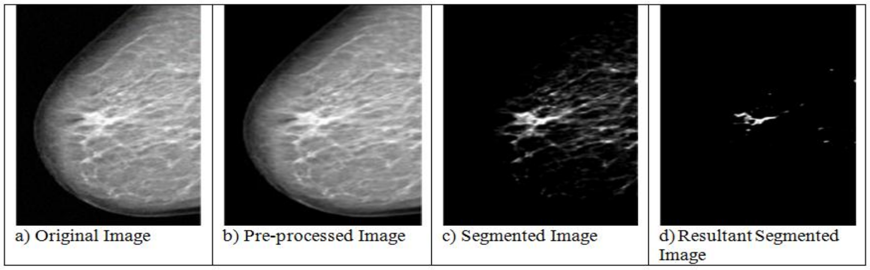
Number of Clusters	Image2.jpg	
	Validity Measure	Davies-Bouldin Index
2	3.099	1.396
3	3.752	2.730
4	2.89	1.209
5	1.946	1.897
6	3.321	1.599
7	2.652	2.477
8	8.253	2.474
9	7.577	3.751
10	7.710	3.522

TABLE IV VALIDITY MEASURE AND DAVIS-BOULDIN INDEX FOR GIVEN BREAST CANCER IMAGE3.JPG BASED ON NUMBER OF CLUSTERS

Number of Clusters	Image3.jpg	
	Validity Measure	Davies-Bouldin Index
2	4.126	2.854
3	6.686	2.323
4	3.833	1.522
5	3.487	2.533
6	1.564	2.987
7	3.122	2.495
8	14.63	2.788
9	12.51	2.779
10	30.28	3.947

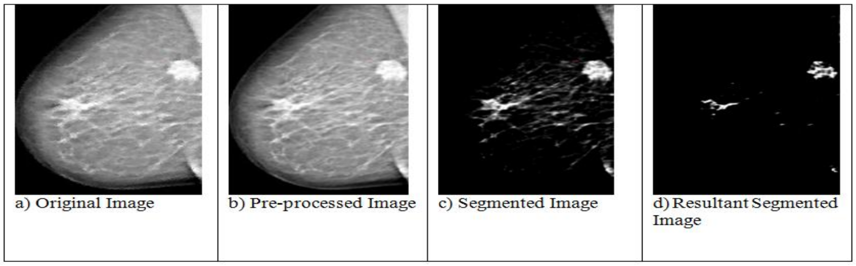
TABLE V VALIDITY MEASURE AND DAVIS-BOULDIN INDEX FOR GIVEN BREAST CANCER IMAGE4.JPG BASED ON NUMBER OF CLUSTERS

Number of Clusters	Image4.jpg	
	Validity Measure	Davies-Bouldin Index
2	2.422	4.947
3	108.2	4.896
4	15.69	2.750
5	12.86	3.221
6	18.77	4.960
7	14.68	3.331
8	13.37	4.944
9	13.79	4.667
10	16.76	4.376



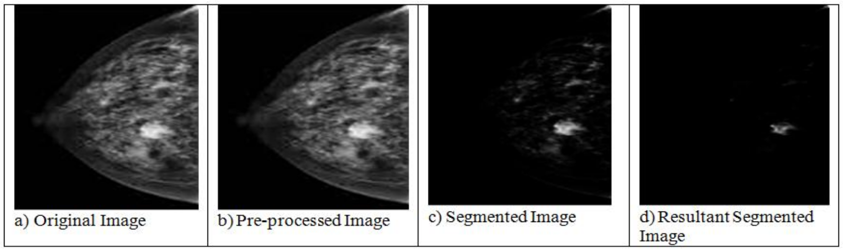
a) Original Image b) Pre-processed Image c) Segmented Image d) Resultant Segmented Image

Fig. 4 Results obtained for given Image1.jpg by Riotous Clustering and SOM Segmentation



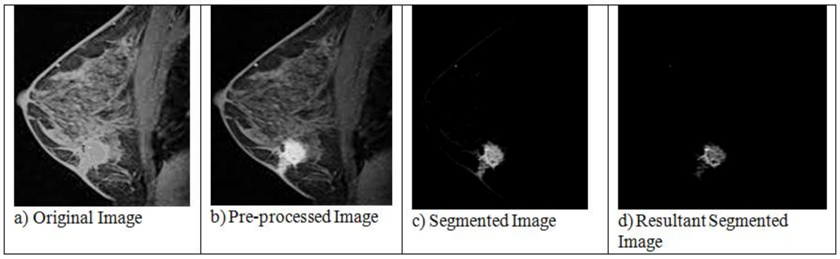
a) Original Image b) Pre-processed Image c) Segmented Image d) Resultant Segmented Image

Fig. 5 Results obtained for given Image2.jpg by Riotous Clustering and SOM Segmentation



a) Original Image b) Pre-processed Image c) Segmented Image d) Resultant Segmented Image

Fig. 6 Results obtained for given Image3.jpg by Riotous Clustering and SOM Segmentation



a) Original Image b) Pre-processed Image c) Segmented Image d) Resultant Segmented Image

Fig. 7 Results obtained for given Image4.jpg by Riotous Clustering and SOM Segmentation

Above fig.4 to fig.7 presents result obtained by proposed framework for given image1 to image 4. From segmented image, cancerous tissue can be obtained easily.

TABLE VI OPTIMAL CLUSTER FOR GIVEN BREAST CANCER IMAGES

S.No.	**Image Number**	**Optimal Number of Cluster**	
		Validity Measure	Davies-Bouldin Index
1	Image1.jpg	5	2
2	Image2.jpg	5	2
3	Image3.jpg	6	2
4	Image4.jpg	2	2

From table VI, cluster number 2 gives optimal solution by using Davies-Bouldin Index among other cluster numbers for given 4 images.

V. CONCLUSION

In this paper, SOM is employed to perform clustering and it does clustering well and gives segmentation results as in human perception, and normalized Euclidean distance is adopted for measuring the distance. The proposed work performs segmentation of breast cancer images in unsupervised and automatic manner, by utilizing measurement of cluster validity. Davies-Bouldin Index (DBI) and Validity Measurement (VM) indexes comparatively affords distinct of optimal number of clusters. For each breast cancer images, optimal numbers of clusters which are developed by DBI, on average are less than results which are obtained by VM.

REFERENCES

[1] Williams, B.Lovoria, *et al.*, “Demographic, psychosocial, and behavioral associations with cancer screening among a homeless population,” *Public Health Nursing*, 2018.

[2] Henriksen, L.Emilie, *et al.*, “The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review,” *ActaRadiologica*, 2018, 0284185118770917.

[3] Siddharth Singh Chouhan, Ajay Kaul, and UdayPratap Singh, “Image Segmentation Using Computational Intelligence Techniques: Review”, *Archives of Computational Methods in Engineering*, 2018.

[4] Ahmed, O.Isra, Banazier A. Ibraheem, and Zeinab A. Mustafa, “Detection of Eye Melanoma Using Artificial Neural Network,” *Journal of Clinical Engineering*, Vol. 43, No. 1, pp. 22-28, 2018.

[5] Shukla, and Nagesh, *et al.*, “Breast cancer data analysis for survivability studies and prediction,” *Computer Methods and Programs in Biomedicine*, Vol. 155, pp.199-208, 2018.

[6] Arora, Shaveta, MadasuHanmandlu, and Gaurav Gupta, “Filtering impulse noise in medical images using information sets,” *Pattern Recognition Letters*, 2018.

[7] Boemer, Fabian, Edward Ratner, and AmauryLendasse, “Parameter-free image segmentation with SLIC,” *Neurocomputing*, Vol. 277, pp. 228-236, 2018.

[8] Park, and Young-Seuk, *et al.*, “Multivariate Data Analysis by Means of Self-Organizing Maps,” *Ecological Informatics. Springer, Cham*, pp. 251-272, 2018.

[9] Kumar, Krishan, Deepti D. Shrimankar, and Navjot Singh, “Eratosthenes sieve based key-frame extraction technique for event summarization in videos,” *Multimedia Tools and Applications*, Vol. 77, No. 6, pp. 7383-7404, 2018.

[10] Ngo, Long Thanh, Trong Hop Dang, and WitoldPedrycz, “Towards Interval-Valued Fuzzy Set–based Collaborative Fuzzy Clustering Algorithms,” *Pattern Recognition*, 2018.