

Segmentation of Fuzzy Enhanced Mammogram Mass Images by using K-Mean Clustering and Region Growing

Nidhi Singh¹, S. Veenadhari²
Dept. Name of CSE
Rabindranath Tagore University
Bhopal, India

Abstract—Providing intention to encourage radiologist's appraisal for distinguishing proof or order of mammogram images, different methods were suggested by specialists since past two decades. By means of this technical paper, we propose segmentation on advanced mammogram imaging with k-means clustering and locale developing systems tending to support specialists or radiologists to figure out cancerous areas with computer-aided techniques. The suggested task is further classified within two stages: Applied/implemented pre-processing, at primary stage. With the pre-processing stage, we carried a median filter to expel undesirable salt and pepper clamor. Further, we apply fuzzy intensification operator (INT) to upgrade the distinction of intake images. During subsequent stage, improved fuzzy imaging conduces as input for k-mean clustering. Secondly, the locale developing technique is employed with previously generated clustered imagery to partition mammogram into homogeneous areas indicated through force from pixels. With the end goal of the experiment, we utilized the smaller than normal MAIS dataset. The experiment's end result shows that proposed strategy accomplishes higher precision.

Keywords—INT operator; feature extraction; k-mean clustering; mammogram; median filter; segmentation

I. INTRODUCTION

In today's time, Cancer (tumor) is major fatal diseases precisely made of widely different related ailments. In every single category about cancer expansion, body somatic (cell) starts to separate repeatedly additionally disseminate amongst encompassing tissue. Similarly, a breast carcinoma emerges from breast organ tissues. Breast comprises of billions of infinitesimal cells. These somatic would begin expanding compulsively which causes cancer of breasts. Cancer of breast growth may be partitioned into two sorts: Ductal Carcinomas and Lobular Carcinomas. Ductal carcinomas are highly prominent cancers that commences in mammary duct whereas lobular carcinomas are particularly ailments that grows around lobes [1]. Prior, the analysis and treatment end up with being ruinous without productive methods. At every stage of carcinoma, the demise rate including dynamism from this disease raises. To decrease the deaths rate and limit the dynamism elicits need for earlier breast cancer identification strategies. So, automated computerized detection is unavoidable. There exists none absolute reasons available for breast carcinoma, and we might see these as causative factors only. They were going to be hereditary or natural. Hereditary

factors incorporate family ancestry, individual wellbeing history, menstrual and reproductive history, dense breast tissues, certain genome changes, age, sex etc. The ecological factors incorporate corpulence, under stellar eating ruts, liquor utilization, radiation, less physical activity, etc. [2]. The elementary factors of breast cancers are forming lumps. Reason for this is tiny sedimentation of calcium called micro calcification and tumors called circumscribed mass such tumors are often benign and not malignant. The benign tumors are generally non-aggressive and non-harmful. It does not disseminate to another body parts [3]. There are several distinctive imaging systems for earlier prediction of breast cancer. These incorporate MRI, X-Ray imaging, ultrasound imaging, computerized mammography, screening etc. The computerized mammograms are widely utilizes currently on account including favorable circumstances over others. X-Ray imaging is typically used for the benefit of discovering indications of carcinoma while utilizing mammograms that generally examines the problem. A mammogram uses X-rays to make breasts imaging [4]. Earlier there exist film mammograms with which images being stored on films, today computerized mammography are generally used that captures and stores straightforwardly on digital computer and every single corners and niches are visible for simple detection. In breasts ultrasound [5], the sound waves are used to form the images. However, currently this is not used since that's done with handheld devices. It would facilitate false positives as well as false negatives, if in this specific instance the person operating is not good grasp or skilled and thereby the quality about the images will differ. Point wise identification of the images gets done by applying feature extracting and texture extraction methods [6]. This alone open other research area, as broad assortment of strategies employed for segmentations, feature extraction, enhancements, done mostly by wavelet techniques [7], clustering using GLCM matrix [8] etc. specifically depicted plainly during relevant works. Therefore a definitive point from this survey is to furnish distinctive improvements, detection and classification approaches for quick recognizing of breast carcinoma.

Several authors have published research papers for breast regions segmentation based on difference in density. For instance, Saidin et al. [9] proposed technique to perform breast segments into four regions: backgrounds, skin-air boundaries, fatty boundaries and pectoral muscles boundaries.

Karssemeijer [10] used approach for mammograms subdivisions towards three separate regions: breast tissues, pectoral muscles and backgrounds. Adel et al. [11] discuss way for breast regions segmentation with three distinct areas: pectoral muscles, fatty and fibro glandular regions by using Bayesian techniques with the adaption of Markov random field for region detection of various tissues on mammograms. El-Zaart et al. [12] provides mammograms segmented images with three regions, which are fibro glandular disc, breast regions and backgrounds. Camilus et al. [13] intend a technique for graph cuts computations seeking the pectoral muscles automatically.

The intended work is divided within two phases. During its primary phase, we applied pre-processing. To be a part of pre-processing step, we had utilized median filter to expel undesirable salt and pepper commotion. We apply fuzzy intensification operator (INT) to strengthen the differentiation of information picture. In the subsequent level, enhanced effective fuzzy images used as input for k-means clustering. Besides, the algorithms for region growing should apply to earlier produce clustered images to divide mammograms within homogeneous parts (regions) in accordance with the relevant intensity of the pixels. Intending to the experiment we used the mini MAIS datasets. The trial result indicates this intended strategy accomplishes higher precision.

The setup of this proposed paper is arranged as follows: Introduction is provided in Section I; Section II discusses substances/material and methods; Section III portrays the proposing (suggested) method; Section IV include outcomes (results) and discussion; conclusion and future work are provided with Section V; we end by references in Section VI.

II. MATERIAL AND METHOD

A. Database

To test out proposed methodology we have use mini-MIAS dataset [14]. This dataset consist of 322 mammogram image. These images are kind of three separate classes. Among 322 images, 106 images belong to Fatty (F) class, another 104 images belongs to Fatty-Glandular (G) class and rest 112 images belong to Dense-Glandular (D) class images. All images have size of 1024 × 1024 pixels in PGM Format. Each pixel in the images is corresponded to the 8-bit word, where the images are in grayscale format with a pixel intensity of range [0, 255] [15].

B. Median Filter

Median filter [16] recurrently used to overcome salt and paper noise while trimming down noise. It also preserves edges in image.

For applying the median filter we have to select a square window of size 2*k+1, where k lies between 1 and N, around the considered pixel. After then arrange and sort all pixel values belongs into the square window and then calculate median value and replace median value with value considered pixel value. This process is repeated for every pixel in the image from left to right and top to bottom manner.

Fig. 1 illustrates an example calculation. In this example we have select k=1; so window size will be 3*3.

105	100	106	97	98	100	102
108	107	98	97	99	98	100
100	104	99	120	101	102	101
100	101	99	99	99	99	97
97	99	100	101	103	98	101

Fig. 1. Matrix of 5×7 where Every Element Represents Gray Value of Pixels.

Neighborhood pixel values:

98,97,99,99,120,101,99,99,99

Sorted pixel values:

97,98,99,99,99,99,99,101,120

Median value: 99

It can be seen that the central pixel value of 120 is rather unrepresentative of the surrounding pixels and is replaced with the median value: 99. A 3×3 square neighborhood is used here and larger neighborhoods will produce more severe smoothing. Image Enhancement using fuzzy intensification operator.

C. Image Enhancement using Fuzzy Intensification Operator

Image enhancement using fuzzy logic [17] is done by using Fuzzification which converts image form spatial domain to fuzzy domain, enhancement if fuzzy membership value and Defuzzification which convert back original image from fuzzy domain to spatial domain.

1) *Fuzzification*: An image X of size M×N can be viewed as an array of fuzzy singleton by converting into fuzzy set notation using following formula:

$$\mu_{ij} = (X_{ij} - X_{\min}) / (X_{\max} - X_{\min}) \quad (1)$$

Here, where $\mu_{i,j}$ is a membership value that represents the amount of brightness acquired by the pixel intensity value $x_{i,j}$ at i^{th} row and j^{th} column in image [18].

2) *Modification of the membership function*: The goal of our proposed method is to take care of the fuzzy nature of an image and formulate the contrast improvement more adjustable and valuable and to prevent from over-enhancement/under-enhancement. So, we employ Adjustment of memberships function $\mu_{i,j} \rightarrow \mu_{i,j}$ by applying following PAL and KING transformation [19] (or the intensification operator (INT)).

$$\mu_{ij} = T(\mu_{ij})$$

$$\mu_{ij} = \begin{cases} 2 * [\mu_{ij}]^2 & 0 \leq \mu_{ij} \leq \mu_c \\ 1 - 2 * [1 - \mu_{ij}]^2 & \mu_c < \mu_{ij} \leq 1 \end{cases} \quad (2)$$

It transforms the membership values that are above threshold value to much larger values and membership values that are below than threshold to much smaller values in a nonlinear manner to obtain an enhanced image. Otherwise, display the unenhanced image.

3) *Defuzzification*: Defuzzification procedure is performed by applying reverse operation of fuzzification.

$$X'_{ij} = G^{-1}(\mu'_{ij}) \quad (3)$$

$$X'_{ij} = X_{\min} + \mu'_{ij} * (X_{\max} - X_{\min}) \quad (4)$$

Thus, the ultimate image will be contrast improved image [18].

D. Histogram based Self Initializing K-Means Clustering

K-Mean clustering [20] is among the most applicable unsupervised learning for classification that needs manual initialization clusters count and primary centroids of each cluster.

The automatic initialization of initial centroids can be achieved with uniform distribution of gray values of the normalized histogram.

1) Algorithm 1: K-Means clustering

a) Input the image X and Initiate the number of clusters K.

b) Calculate the histogram of X. Find out the smallest and largest histogram gray level respectively, say p and q. Based on the value of p and q, apply discrete uniform distribution to initiate the initial centroids c_j where $j=1, 2 \dots K$.

The general formula for the probability density function (pdf) for the uniform distribution is:

$$f(x) = 1/b - a \quad \forall a \leq x \leq b$$

c) Allocate each pixel to the nearest class. This process is completed by minimizing the function J as given below.

$$J = \sum_{i=1}^N \sum_{j=1}^K \|x_i^j - c_j\|^2$$

Where, $\|x_i^j - c_j\|^2$ is an absolute distance from a data point x_i^j to the cluster center C_j .

d) Evaluate mean value for each cluster and substitute this mean value as new centroids.

e) Do again steps and 4 until no more new centroids are created.

E. Region-based Segmentation

RBS [21-22] is a method for finding out region on the basis of some region membership criteria directly. The basic steps for RBS are:

a) Pick a set of seed points. Number of seed points depends on number of segments. Seed point selection is based on user choice or any random process.

b) The process starts from the selected seed points and then regions are expanded from these points to its neighboring points depending on region relationship norms. The norms may be pixel intensity, pixel color or texture.

c) Repeat step 2 for each of the recently added pixels; stop if no more pixels can be added.

III. PROPOSED METHOD

In this Article ornate (elaborate) the intended method working sensible to easily identify and segment the boundary of breast tissues area in mammograms image. Our suggested method deals via a pair of phases:

- 1) Preprocessing.
- 2) K-means Clustering based Segmentation.

The first step, preprocessing comprises of two sub section i.e. noise Removal and improvement in image contradictions. Median filter utilizes in removal of salt and paper noise and contrast of images strengthened from assistance of Fuzzy Pal and King Method.

Second phase, the image splits in multiple clusters by applying K-mean clustering and appertaining to Seed Based Region Growing (SBRG) Technique, required appropriate cluster is segmented by selecting the seed/start point.

a) Proposed Algorithm

Step 1: Input the mammogram images.

Step 2: Median filter is imposed thus removes various noises from imaging.

Step 3: Change images (getting from step 2) from spatial domain to fuzzy area and subsequently applying the pal king function, as demanded INT operator on fuzzy domain that modifies fuzzy area for low contrast area to upper contrast area and moreover lastly(modified) fuzzy area transformed into intensified image using Defuzzification process.

Step 4: Perform the Histogram based Self Initializing K-Means Clustering algorithm that partition the images in different number of clusters. The count of clusters is mentioned before processing.

Step 5: Apply SBRG methodology which segments the appropriate region by performing region growing from seed/start points to adjoining points in concordance with region membership criterion.

Fig. 2 show the Block diagram of proposed work.

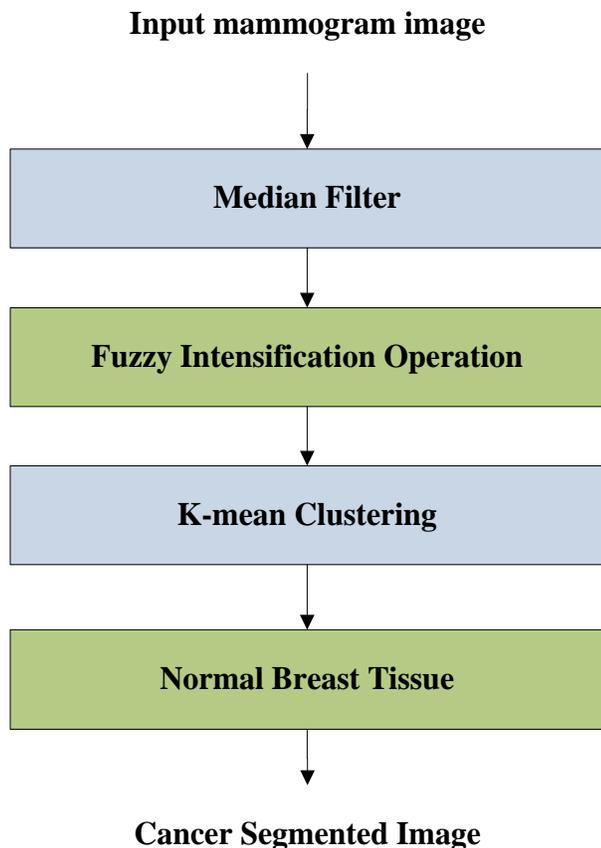


Fig. 2. Flow Chart of the Proposed Method.

IV. EXPERIMENTS AND RESULT

Thirty-one MIAS images brought into play the experiment. 15 together were cancerous images and 16 were normal images. Regarding this the proposed algorithm, referring to 15 cancerous images, 12 patients detected correctly and remaining 3 images, the cancerous region including tumor found incorrect. When the proposed methodology applied to remaining 16 patients on normal images, the significant cancer tissues was determined in 2 images, and in case of 14 images, the algorithm determines that these images does not contain carcinoma.

Fig. 2 shows the result of proposed method against the various mammogram images. Fig. 3(a) depicts Original image, Fig. 3(b) shows Mammography image after preprocessing step, Enhanced Mammography image after fuzzy INT operation is shown in Fig. 3(c). Fig. 3(d) depicts segmented image after K-mean clustering and finally Fig. 3(e) shows segment extracted by using region growing technique.

Table I depicts confusion matrix which describes relation between actual and predicted class by the performed classifier in our case we have two classes i.e. benign and malignant.

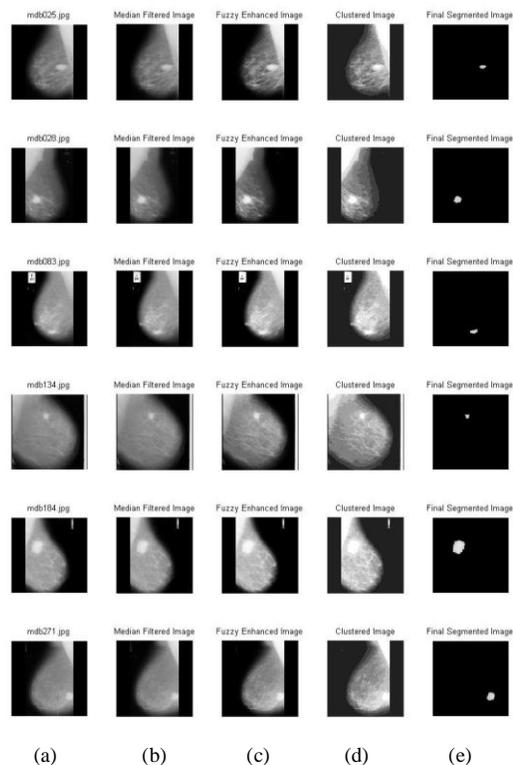


Fig. 3. Segmentation of Images mdb025, mdb028, mdb083, mdb134, mdb184 and mdb271:(a) Original Image, (b) Mammography Image after Preprocessing Step, (c) Enhanced Mammography Image after Fuzzy INT Operation (d) Segmented Image by K-mean Clustering (e) Segment Extraction by using Region Growing.

Table II shows the performance measure calculated from the Table I. From the Table II, it is clear that our proposed method has good precision value while error is low.

Two vital evaluation parameters for calculating effectiveness of our suggested technique are True Positive Rate (TPR) and False Positive Rate (FPR). TPR determines the correctly classified cancerous ROIs among all cancerous ROIs available whereas FPR determines incorrectly classified cancerous ROIs among all non-cancerous ROIs.

TABLE I. CONFUSION MATRIX

Actual Class	Predicted class	
	Cancerous image	Normal Image
Cancerous image	12	03
Normal Image	02	14

TABLE II. PERFORMANCE MEASURE

Measure	Definition
Error	0.1612
FPR	0.8235
precision	0.8235
ACC	83.87
F1_Score	0.8484
MCC	0.6778

Depending of evaluating the binary classification, F-measure and MCC also contribute a very significant role. F-measure examines harmonic mean of precision and recall and expounds as:

$$F - measure = \frac{2 \times recall \times precision}{recall + precision} \quad (5)$$

The MCC anatomizes correlation coefficient within observed and predicted classification and described as:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{(TP + FN)(TN + FP)(TP + FP)(TN + FN)} \quad (6)$$

The F-measure value separate from 0 to + 1 and MCC value lies from -1 to +1. Higher value measurement of both F-measure and MCC specify higher classification quality.

V. CONCLUSION

For radiologist's consideration to easily identification and categorization of concerning mammography images, different methods by researchers sustained bearing in mind since past decades. By means of this research report, we aims mammography image segmentation with k-means clustering and region growing techniques benefit to experts or radiologists to detect cancerous regions with computer-aids. The intended work carve up within two stages: At start stage, pre-processing employed, we retained a median filter that expel undesirable salt and pepper noise. Further, taken pal king approach that reinforce input images contrast. In subsequent stage, more effective image is exerted as an input for k-mean clustering. Image is divided in order to get multiple clusters and with Seed Based Region Growing (SBRG) Technique, required appropriate cluster is segmented by selecting the seed point. For achieving rationale behind this experiment, we are using mini MAIS dataset. The end outcome of experiments demonstrates that proposed strategy accomplishes higher precision.

REFERENCES

- [1] Saurabh PrasadLori Mann BruceJohn E Ball," A Multi-classifier and Decision Fusion Framework for Robust Classification of Mammographic Masses" International Conference of the IEEE Engineering in Medicine and Biology , 2008.
- [2] K.Subashini, K.Jeyanthi, 'Masses detection and classification in ultrasound images', IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS), Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51.
- [3] Ojo J. A., Adepoju T. M., Omdiora E. O., Olabiyisi O. S. and Bello O. T, 'Pre-Processing Method for Extraction of Pectoral Muscle and Removal of Artefacts in Mammogram,' IOSR Journal of Computer Engineering (IOSR-JCE) e-Volume 16, Issue 3, Ver. V (May-Jun. 2014).
- [4] Huanping Zhao,Weidong Xu,Lihua Li,Juan Zhang "Classification of Breast Masses Based on Multi-View Information Fusion Using Multi-Agent Method", 5th International Conference on Bioinformatics and Biomedical Engineering, (iCBBE) 2011.
- [5] K.Subashini, K.Jeyanthi, 'Masses detection and classification in ultrasound images', IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS),Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51.
- [6] S. Deepa, Dr.V.Subbiah Bharathi, 'Textural Feature Extraction and Classification of Mammogram Images using CCCM and PNN', IOSR Journal of Computer Engineering (IOSR-JCE) ,Volume 10, Issue 6 (May. - Jun. 2013).
- [7] Mencamttini, A.et al., "Mammographies images enhancement and denoising for breast cancer detection using dyadic wavelet processing", IEEE Trans. Instrumentation Measure., 57: 1422-1430. DOI: 10.1109/TIM.2007.915470.
- [8] Abdelali ELMOUFIDI et al.,"Detection of Regions of Interests in Mammograms by Using Local Binary Pattern, Dynamic K-Means Algorithm and Gray Level Co-occurrence Matrix", 2014 Fifth International Conference on Next Generation Networks and Services (NGNS'14) 28- 30 May 2014, Casablanca, Morocco.
- [9] N. Saidin et al., "Density based breast segmentation for mammograms using graph cut and seed based region growing techniques", 2nd International Conference on Computer Research and Development (ICCRD '10), pp. 246-250, Kuala Lumpur, Malaysia, May 2010.
- [10] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms", Phys. Med. Biol., vol. 43, 1998, pp. 365-378.
- [11] A.El-Zaart et al., "Expectation Cmaximization technique for fibroglandular discs detection in mammography images", Comput. Biol. Med. 40(4), 392-401 (2010).
- [12] A.El-Zaart et al., "Expectation Cmaximization technique for fibroglandular discs detection in mammography images", Comput. Biol. Med. 40(4), 392-401 (2010).
- [13] K. S. Camilus et al., "Computer- Aided Identification of the Pectoral Muscle in Digitized Mammograms", Journal of digital imaging, October 2009, pp. 1-19.
- [14] [Online] Available: <http://www.wiau.man.ac.uk/services/MIAS/MIASweb.htm>.
- [15] Suckling, J. et al., "The Mammographic Image Analysis Society digital mammogram database", Exerpta Medica, International Congress Series 1069 pp. 375-378., 1994.
- [16] Makandar, Aziz and Bhagirathi Halalli. "Breast Cancer Image Enhancement using Median Filter and CLAHE." (2015).
- [17] M. Hanmandlu and D. Jha, "An Optimal Fuzzy System for Color Image Enhancement", Image Processing, IEEE Transactions on, Vol. 15, pp. 2956-2966, 2006.
- [18] Khairunnisa Hasikin and Nor Ashidi Mat Isa, "Enhancement of the low contrast image using fuzzy set theory"14th International Conference on Modelling and Simulation, IEEE,2012.
- [19] Pal S K, King R A. Image enhancement using smoothing with fuzzy sets. IEEE Trans. Systems, Man & Cybernetics, 1981, 11(7): 494-501.
- [20] S.Julian Savari Antony ,Dr.S.Ravi, 'A New Approach to Determine the Classification of Mammographic Image Using K-Means Clustering Algorithm', International Journal of Advancements in Research & Technology, Volume 4, Issue 2, February -2015.
- [21] R. Adams and L. Bischof. "Seeded region growing", IEEE Trans, Pattern Analysis Machine Intelligence, 16(6):641-647, Jun 1994.
- [22] Harikrishna Rai G.N et al., "Gradient Based Seeded Region Grow method for CT Angiographic Image Segmentation", InterJRI Computer Science and Networking, Vol. 1, Issue 1, July 2009.