International Journal of Advanced Trends in Engineering, Science and Technology(IJATEST)Volume.4, Special Issue.1Dec.2016

# AN EFFICIENT HYBRID MODEL FOR EXPOSURE OF BREAST CANCER IN MAMMOGRAMS BY IMAGING TECHNIQUES

Mr. T. Lakshmi Narayana<sup>1</sup>, Dr. T. Sreenivasulu Reddy<sup>2</sup>, Mr. C. Venkatesh<sup>3</sup> Research Scholar, Dept of E.C.E., S.V.U.C.E, Tirupati, India<sup>1</sup> Professor, Dept of E.C.E., S.V.U.C.E, Tirupati, India<sup>2</sup> Assistant Professor, Dept of E.C.E., A.I.T.S, Rajampet, India<sup>3</sup>

# career.lak@gmail.com

Abstract: Now a day Image processing is the best solution for various problems especially in the medical field. At present, humankind throughout the world is affected with one or other types of cancer. In all over the world breast cancer is one of the major causes of death among women. Early detection and screening of cancer which greatly increases the chances for successful treatment. Mammography is most widely used test by radiologists for screening, detecting and diagnosis of breast cancer. The available detection procedures and models are unable to forecast the abnormalities within time because of attenuation in acquiring device, low contrast and blurreness present in breast cancer related images.

Hence a dynamic model is proposed for diagnosing and detecting the abnormalities which lie on the breast cancer related images. This model combines various State-Of-Art clustering algorithms in combination with image segmentation techniques and neural network algorithm to make images easier to detect the abnormalities accurately and precisely. The first step should be a segmentation procedure able to distinguish various neoplasms from background and second step should be a morphological procedure for classification and feature extraction of tumors accurately. The main objective of the model is to detect and forecast the abnormalities in the breast cancer related images.

Keywords: Mammography, Computer Aided Detection (CAD), Breast Cancer, microcalcification, mass detection, ROI, Image segmentation.

## I. INTRODUCTION

At present, breast cancer is the most common cancer among women, after cancers of the skin, and the second leading cause of cancer death in women after lung cancer [1-3]. The malignancy cell that grows in the breast tissue is termed as breast cancer. The risks of the breast cancer increases with the factors such as female gender, obesity, lack of physical exercise, having children late or not at all etc .It is to be found that the 80% of women who are above age of 50. The most healthier and cost effective way of screening the breast cancer is through mammograms. A mammogram is one of the best radiographic methods to detect the breast cancer earlier. It even detects the tumors which are too small to be felt and gives us the X-ray image as an output. Image Processing techniques provide a sufficient assessment category to the

abnormalities.

Classification of these images are done with the three main stages, they are preprocessing, segmentation and classification. Pre-processing: Prior processing is very necessary to direct the algorithm for classification. Segmentation: Extraction of region of interest and analyzing the same can be done in his step. One of the essential step to analyze, object representation and visualization. The different kinds of segmentation are threshold segmentation, edge Region based segmentation, based segmentation, clustering techniques and mapping segmentation. These categories are contextual or noncontextual. Classification: The segmented images are taken where which classification techniques are applied to identify the results accordingly. The intent of this process is to find the severity of tissues and

## National Conference on Emerging Trends in Information, Digital & Embedded Systems(NC'e-TIDES-2016)

## International Journal of Advanced Trends in Engineering, Science and Technology(IJATEST)Volume.4, Special Issue.1Dec.2016

categorize the data sets into several classes. It will be classified according to the categories such as density, texture etc in the feature space. Hence one of the important steps is classification and one has to concentrate on choosing the technique wisely. **II.PROPOSED WORK** 

# **Segmentation of Suspicious MC Regions**

In this section, the suspicious MC regions are identified and for this purpose, the proposed algorithm is given below. As seen in Fig. 6., the upper region identifies the pectoral muscle region and the limit value is used to control if the pectoral muscle is seen in mammography image or not. Because there is no pectoral muscle in some mammograms or it is just a bit visible. Therefore, if the pectoral muscle is visible in the mammography image it can be removed because it is an unnecessary area for MC regions.



#### International Journal of Advanced Trends in Engineering, Science and Technology(IJATEST)Volume.4, Special Issue.1Dec.2016

### **Otsu's N=3 Thresholding**

Otsu's N thresholding is an automatic threshold selection method which separates the image into classes [9]. The steps of Otsu's

N=3 thresholding algorithm are given below. The

grey levels of the image are identified as [0,1,2,...,L-1] where L is the number of grey levels.

Step 1. The normalized histogram of the input image is

computed and the histogram elements are shown with  $P_i$  where i=0, 1, 2, ..., L-1.

Step 2. The threshold values k1, k2, k3 separates the image into 4 classes and the cumulative sums P1, P2, P3 and P4 are computed for 4 classes with the following formulas.

$$P_{1} = \sum_{i=0}^{k_{1}} P_{i}$$

$$P_{2} = \sum_{i=k_{1}+1}^{k_{2}} P_{i}$$

$$P_{3} = \sum_{i=k_{2}+1}^{k_{3}} P_{i}$$

$$P_{4} = \sum_{i=k_{3}+1}^{k_{4}} P_{i}$$

Step 3. The global intensity average,  $m_G$ , is computed by the formula below.

$$m_G = \sum_{i=0}^{L-1} i P_i$$

Step 4. The average intensity values for the classes are computed with the following formulas.

$$m_{1} = \frac{1}{P_{1}} \sum_{i=0}^{k_{1}} i P_{i} \qquad m_{2} = \frac{1}{P_{2}} \sum_{i=k_{1}+1}^{k_{2}} i P_{i} m_{3} = \frac{1}{P_{3}} \sum_{i=k_{2}+1}^{k_{3}} i P_{i} \qquad m_{4} = \frac{1}{P_{4}} \sum_{i=k_{3}+1}^{k_{4}} i P_{i}$$

Step 5. The variance between classes can be computed as below.

$$\sigma_{\beta}^{2}(k_{1},k_{2},k_{3}) = P_{1}(m_{1}-m_{G})^{2} + P_{2}(m_{2}-m_{G})^{2} + P_{3}(m_{3}-m_{G})^{2} + P_{4}$$

$$\sigma_{\beta}^{2}(k_{1},k_{2},k_{3}), \begin{cases} k_{1} = 1, \dots, k_{2} - 1 \\ k_{2} = k_{1} + 1, \dots, k_{3} - 1 \\ k_{3} = k_{2} + 1, \dots, L - 2 \end{cases}$$

Step 6. The optimum threshold values correspond to the values which maximizes the variance between classes.

#### **K-Mean Clustering Algorithm**

In statistics and machine learning, k-means clustering is a

method of cluster analysis which aims to partition "n" observations in to "k" clusters in which each observation belongs to the cluster with the nearest [7]. For a given set of observation mean  $(x_1, x_2, \dots, x_n)$ , where each observation is a ddimensional real vector, then k-means clustering ,,n" aims to partition the observations in to (k<n),  $\{S = S_1, S_2, \dots, S_n\}_{so as to}$ ..k" sets minimize the within cluster sum of squares (WCSS) in equation (1).

$$\arg\min_{s} \sum_{i=1}^{k} \sum_{x_{j} \in s_{i}} \|x_{j} - \mu_{i}\|^{2}$$
(1)

Where,  $\mu_i$  is the mean  $\delta_i$  .The number of cluster k is assumed to be fixed in k-means clustering.

#### Standard algorithm

Given an initial set of k-means in which may be specified randomly or by some heuristic, the algorithm produces by alternating between two steps.

## I. Assignment Step

Assign each observation to the cluster with the closest mean (i.e. partition the observation according to the voronoi diagram generated by the means) in equation (2).

$$s_{i} = \{x_{j} : ||x_{j} - m_{i}^{t}|| \leq ||x_{j} - m_{i^{*}}^{t}|| \forall i^{*} = 1.....k\}$$
*II.* Update Step
(2)

Calculate the new means to be the centroid of the observations in the cluster in equation (3)

$$m_{i}^{t+1} = \frac{1}{\left|s_{i}^{t}\right|} \sum_{x_{i} \in S_{i}^{t}} x_{j}$$
(3)

The algorithm is usually very fast, it is common to run it multiple times with different starting conditions. Theoretically it has seen that there exit certain point sets on which k-means takes superpolynomial time, but practically it is not so far. National Conference on Emerging Trends in Information, Digital & Embedded Systems(NC'e-TID -2016)

International Journal of Advanced Trends in Engineering, Science and Technology(IJATEST)Volume.4,Special Issue

### **3.2 Fuzzy C-Mean Algorithm**

The Fuzzy C-means algorithm, also known as fuzzy ISODATA, is one of the most frequently used methods in pattern recognition. Fuzzy C-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters

piece of data to belong to two or more clusters [7]. It is based on the minimization of objective function to achieve a good classification. "J" is a squared error clustering criterion, and solutions of minimization are least squared error stationary point of "J" in equation (4).

$$j_{m} = \sum_{i=1}^{k} \sum_{j=1}^{c} u_{jj} || x_{i} - c_{j} ||^{2}$$

$$1 \le m \le \infty$$
(4)

Where,

Where "m" is any real number greater than 1, is the degree of membership of in the cluster "j", is the d-dimensional measured data, is the dimension center of the cluster and is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above,

with the update of member ship  $u_{ij}$  in equation (5) and the cluster centers by equation (6)

$$u_{jj} = \frac{1}{\sum_{k=1}^{c} \left[ \frac{\|x_{i} - c_{j}\|}{\|x_{i} - c_{k}\|} \right]^{\frac{2}{m-1}}}$$
(5)

$$c_j = \frac{\sum_{i=1}^{k} u_{ij} \cdot x_i}{\sum_{i=1}^{k} u_{ij}} \tag{6}$$

## **Hierarchical Clustering**

Hierarchical clustering taking successively by either merging smaller clusters into larger ones or by dividing larger clusters. The result of the algorithm is a hierarchy of clusters, called dendrogram, which shows how the clusters are related. By wounding the dendrogram at a desired level, a clustering of the data items into disjoint groups is acquired. It constructs a hierarchy of clusters or, in other words, a tree of clusters, also known as a dendrogram[6]. There are two type of hierarchical clustering such as agglomerative and Divisive clustering. In agglomerative clustering start w ith one point and recursively add two or more appropriate clusters. Also it stop with k number of clusters achieved.

# **III COMPARISION ANALYSIS**

The experimental results of basic clustering algorithms are discussed in this section using the data mining tool Breast cancer data contains tumors which represent the severity of the disease. The two kinds of tumors are benign and malignant. To cluster them correctly from the training data set accuracy using clusters are evaluated. The accuracy of the various clustering technique for diagnosis of breast cancer either benign or malignant are compared is shown in Table 1

Table.2	Comparisions	of	clustering
Algorith	ms		

Clusterin g	Beni gn	Malign ant	Accura cv	Tim e
Fuzzy	552	168	78%	0.01
K-Means	592	152	76%	0.02
Hierarchy	458	241	66%	1.69

# IV CONCLUSION

Breast cancer is one of the major causes of death among women. In this paper we have investigated a novel approach to detect the presence breast cancer mass and calcification in of mammograms using image processing functions, Kmeans and Fuzzy C-Means clustering for clear identification of clusters. Combining these we have successfully detected the breast cancer area in raw mammograms images. In the paper there was a comparison between Fuzzy C-Means and K-Means Algorithm.

## REFERENCES.

[1]. Shah, "comparison of data mining classification algorithms for breastcancer prediction", IEEE fourth international conference on computing, communication and networking technologies 2013

[2]. R.Delshi Howsalya Devi , P.Deepika, "Performance Comparison of Various Clustering Techniques for Diagnosis of Breast Cancer", IEEE 2015.

[3]A Sathish, J Mohana Sundaram "A Comparative Study on K-Means and Fuzzy C-Means Algorithm for Breast Cancer Analysis" International Journal of Computational Intelligence and Informatics, Vol. 4: No. 1, April - June 2014

## National Conference on Emerging Trends in Information, Digital & Embedded Systems(NC'e-TIDES-2016)

#### International Journal of Advanced Trends in Engineering, Science and Technology(IJATEST)Volume.4,Special Issue.1Dec.2016

[4].Nalini Singh,Ambarish G Mohapatra, Gurukalyan Kanungo" Breast Cancer Mass Detection in Mammograms using K-means and Fuzzy Cmeans Clustering"International Journal of Computer Applications (0975 – 8887) Volume 22–No.2, May 2011

[5] T.Hieken, J.Harrison, J.Herreros, Velasco, "Correlating sonography, mammography, and pathology in the assessment of breast cancer size," American Journal of Surgery 182(4), pp. 351-354, .2001.

[6] I.Saarenmaa, T.Salminen, U.Geiger, P.Heikkinen, S.Hyvrinen, J.Isola, V.Kataja, M.Kokko, R.Kokko, E.Kumpulainen, "The effect of age and density of the breast on the sensitivity of breast cancer diagnostic by mammography and ultasonography," Breast Cancer Research and Treatment, 67(2), pp. 117-123, 2001.

 [7] W.Pedrycz, J. Waletzky, "Fuzzy clustering with partial supervision," IEEE Transactions on Systems, Man and Cybernetics," 27, pp. 787–795, May 1997.

[8] D.T.C. Lai, J.M. Garibaldi, "A comparison of distance-based semisupervised fuzzy c-means clustering algorithms," in 2011 IEEE International Conference on Fuzzy Systems, pp. 1580–1586, June 2011.

[9]. Aijuan Dong and Baoying Wang "Feature selection and analysis on mammogram classification", Communications, Computers and Signal Processing, 2009. Pac Rim 2009. IEEE Pacific Rim Conference on 23-26 Aug. 2009.

[10]. Thangavel, K.Mohideen, A.K. Dept. of Comp.Sci., Periyar Univ., Salem, India, "Semi-supervised k-means clustering for outlier detection in mammogram classification" appears in Trandz in information sciences & computing, (2010).

[11]. Cahoon, T.C. Sutton, M.A. Bezdek "Brest cancer detection using image processing techniques", J.C.Dept.of Comp.Sci.Univ.of West Florida, Pensacola, FL Fuzzy IEEE 2000.The ninth IEEE conference.

[9]. "Microcalcification detection in digital mammograms based on wavelet analysis and neural networks" by Jasmine, J.S.L. Govardhan, A. Baskaran, S. Dept. of CSE, Velammal Eng. Coll., Chennai, India on Control