

HMPFIM-B: Hybrid Markov Penalized FCM in Mammograms for Breast Cancer

Ms. Siva Priya,
Assistant Professor
Department of Computer Science and Engg.,
MNM Jain Engineering college,
Tamil Nadu, India.

E.Ashok Kumar,
Assistant Professor
Department of Information Technology,
SMK Fomra Institute of Technology,
Tamil Nadu, India.

Abstract--Mammography is an imaging tool which uses low dose low energy x-ray for early detection of tumours in breast. Currently there are number of image based software applications to assist radiologists for better screening. Segmentation is the best way for reliable diagnosis by reducing false rate. So here, we propose novel segmentation algorithm using fuzzy logic. This new approach uses penalized fuzzy c means clustering in mammographic image to give significant improved performance while screening mammogram. The real-time implementation of this paper can be implemented using hardware and software interface with the mammography systems.

Keywords--Penalized FCM, Image processing, Markov process, Segmentation, Mammogram, MRI image, X-ray

I. INTRODUCTION

Breast cancer occurs when cells in the breast divide and grow without their normal control. Tumours in the breast tend to grow slowly. By the time a lump is large enough to feel, it may have been growing for as long as 10 years[9]. Between 50 and 75 percent of breast cancers begin in the milk ducts, 10 to 15 percent begin in the lobules and a few begin in other breast tissues. From the 1940s until the 1980s, breast cancer incidence (new cases) rates in the U.S. increased by a little over one percent each year. In the 1980s, incidence rose greatly (likely due to increased mammography screening), and then levelled off during the 1990s.

The incidence of breast cancer declined in the early 2000s. Although mammography screening rates fell somewhat over this same time period, studies show these changes were not likely related to the decline in breast cancer rates.

A mammogram is an x-ray that allows a qualified specialist to examine the breast tissue for any suspicious areas. The breast is exposed to a small dose of ionizing radiation that produces an image of the breast tissue. Mammograms can often show a breast lump before it can be felt. They also can show tiny clusters of calcium called micro calcifications. Lumps or specks can be caused by cancer, fatty cells, or other conditions like cysts. Further tests are needed to find out if abnormal cells are present. [8]

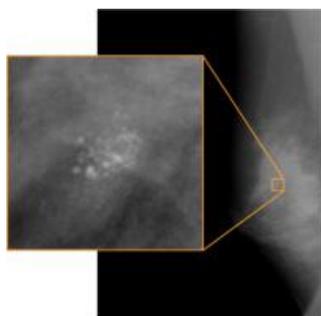


Fig1 :Example of a mammogram. The cloud of white specks visible in the magnified portion is clustered micro calcifications, which is an indicator associated with breast cancer.

Although mammography is considered the most reliable means of detecting breast cancer, between 10-30% of women subsequently diagnosed with breast cancer have false negative mammograms.[7]

They use computer-aided detection (CAD) for mammography to accurately interpret mammographic images and check for the presence of breast cancer. The CAD system puts a mammogram into digital form and then computer software searches for abnormal areas of density, mass or calcification. The system highlights suspicious areas, alerting our cancer doctors to the need for further analysis.[6]The CAD system may improve the detection of cancer in the breast by acting as a second set of eyes to find abnormal areas on a mammogram.

By performing highly-accurate computer-aided detection of shadows indicating suspected breast-cancer lesions, and making the detection results available to the radiologist interpreting the mammogram at the time, this system can help raise the degree of certainty of the radiologist's diagnosis, by, for example, preventing lesions from being missed[9]. By reducing the stressfulness of the task of interpreting mammograms, it can also make a major contribution to the early detection of breast cancer.

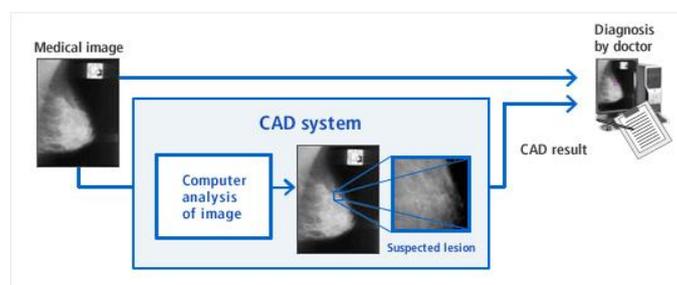
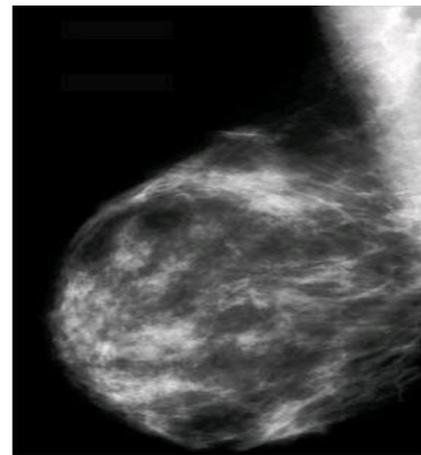
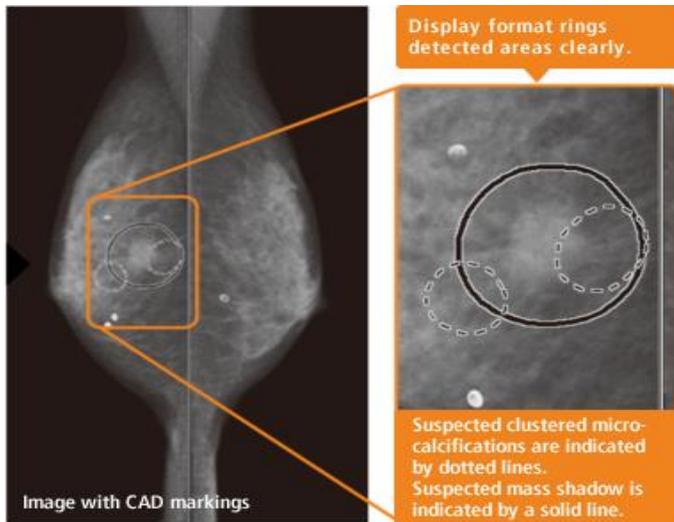


Fig2: Medical image in CAD

The precise segmentation of the breast region in mammograms is an essential pre-processing step in the computerized analysis of mammograms. It allows the search for abnormalities to be limited to the region of the breast

without undue influence from the background of the mammogram. It also facilitates enhancements to techniques such as comparative analysis, which includes the automated comparison of corresponding mammograms.[7].



(a)



(b)

Fig 4: a) Original Image b) Thresholded image

II. DIFFERENT TECHNIQUES IN MAMMOGRAM SEGMENTATION.

The image segmentation towards mammogram can be classified by four different categories: thresholding, clustering, edge detection and region extraction.

(a). Thresholding.

The simplest method of image segmentation is called the thresholding method. This method is based on a clip-level (or a threshold value) to turn a gray-scale image into a binary image. With a single threshold, it transforms a gray scale or colour image into a binary image considered as a binary region map. The binary map contains two possibly disjoint regions, one of them containing pixels with input data values smaller than a threshold and another relating to the input values that are at or above the threshold. The former and latter regions are usually labelled with zero (0) and non-zero (1) labels, respectively. The segmentation depends on image property being thresholded and on how the threshold is chosen.

General form of thresholding function is in equation

$$S=T(r)$$

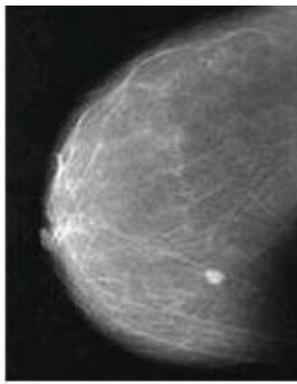
Where s & r are the input and output images respectively. [3]The effect of this transformation would be to produce an image of higher contrast than the original by darkening the levels below and brightening the levels above around particular value which is known as threshold value.

(b) Clustering method

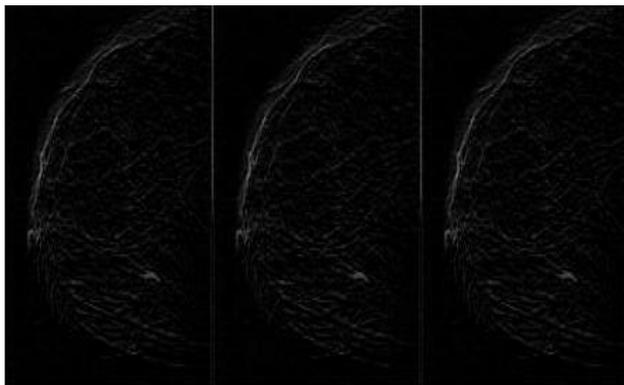
Clustering is a classification technique. Given a vector of N measurements describing each pixel or group of pixels (i.e., region) in an image, a similarity of the measurement vectors and therefore their clustering in the N-dimensional measurement space implies similarity of the corresponding pixels or pixel groups. Therefore, clustering in measurement space may be an indicator of similarity of image regions, and may be used for segmentation purposes.

(c) Edge detection

It deals with the technique of taking decision as to whether the given pixel is in the edge or not. Traditional methods of edge detection involve convolving the image through an operator/filter, which is constructed to be perceptible to large gradients in the image, although returning values of zero in uniform regions.[5]



(a)



(b)

Fig 5 a) Original Image b) Detected Image

(d) Region-based method

A region-based method usually proceeds as follows: the image is partitioned into connected regions by grouping neighbouring pixels of similar intensity levels. Adjacent regions are then merged under some criterion involving perhaps homogeneity or sharpness of region boundaries. Over stringent criteria create fragmentation; lenient ones overlook blurred boundaries and over merge. Hybrid techniques using a mix of the methods above are also popular.

Our proposed technique comes under clustering method.

III LITERATURE SURVEY ON CLUSTERING ALGORITHM.

(a) K-Mean Clustering Algorithm

K-Means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other[4]. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are clustered around centroids $\mu_i \forall i = 1 \dots k$ which are obtained by minimizing the objective

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \tag{1}$$

Where there are k clusters $S_i, i=1,2,\dots,k$ and μ_i is the centroid or mean point of all the points $\in S_i$. As a part of this project, an iterative version of the algorithm was implemented. The algorithm takes a 2 dimensional image as input. Various steps in the algorithm are as follows:

1. Compute the intensity distribution (also called the histogram) of the intensities.
2. Initialize the centroids with k random intensities.
3. Repeat the following steps until the cluster a label of the image does not change anymore.
4. Cluster the points based on distance of their intensities from the centroid intensities.

$$C_{(i)} = \arg \min_j \|x^{(i)} - \mu_j\|^2 \tag{2}$$

5. Compute the new centroid for each of the clusters.

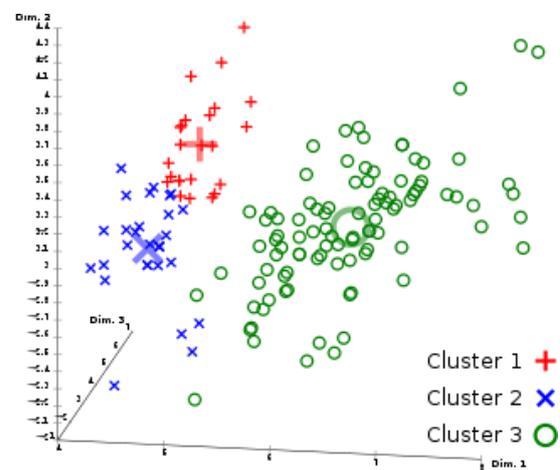


Fig 6. k-means clustering result for Iris dataset [11]

4(b). Expectation-Maximization (EM) algorithm

K-means algorithm is simple. However, it is easy to get stuck in local optimal. The EM algorithm tends to get stuck less than K-means algorithm. The idea is to assign data points partially to different clusters instead of assigning to only one cluster [10]. To do this partial assignment, we model each cluster using a probabilistic distribution. So a data point associates with a cluster with certain probability and it belongs to the cluster with the highest probability in the final assignment

We can use mixture of Gaussian distributions to model this. The mixture model is a weighted sum of K Gaussian distributions. The weights sum up to 1. Let the parameter of jth distribution be θ_j and its weight be w_j , the probability of a data point x_i given this model is

$$p(x_i | \theta) = \sum_{j=1}^k w_j p_j(x_i | \theta_j) \tag{3}$$

where $\theta = \{w_1, w_2, \dots, w_k, \theta_1, \theta_2, \dots, \theta_k\}$

(c) Fuzzy C means Algorithm

One of the problems of the k-means algorithm is that it gives a hard partitioning of the data, that is to say that each point is attributed to one and only one cluster. But points on the edge of the cluster, or near another cluster, may not be as much in the cluster as points in the center of cluster.

Therefore, in fuzzy clustering, each point does not pertain to a given cluster, but has a degree of belonging to a certain cluster, as in fuzzy logic. For each point x we have coefficient giving the degree of being in the k -th cluster $u_k(x)$. Usually, the sum of those coefficients has to be one, so that $u_k(x)$ denotes a probability of belonging to a certain cluster:

This fuzzy iterative clustering method produces an optimal c partition [2] by minimizing the weighted within group sum of squared error objective function J_{FCM} .

$$J_{FCM} = \sum_{k=1}^n \sum_{i=1}^c (U_{ik})^q d^2(x_k, v_i) \quad (4)$$

Where $x = \{x_1, x_2, \dots, x_n\} \in R^p$ is the data set in the p -dimensional vector space, n is the number of data items, c is the number of clusters with $2 \leq c < n$, u_{ik} is the degree of membership of x_k in the i th cluster, q is a weighting exponent on each fuzzy membership, v_i is the prototype of the centre of cluster i , $d^2(x_k, v_i)$ is a distance measure between object x_k and cluster centre v_i . A solution of the object function J_{FCM} can be obtained via an iterative process, which is carried out as follows:

1. Choose the values for no of clusters, weighing component.
2. Set the starting values for fuzzy partition matrix with degree of membership values $[u_{ik}]$.
3. Let iteration controller be b with initial value 0.
4. Find the c clusters centre with following formula

$$v_i^{(b)} = \frac{\sum_{k=1}^n (U_{ik})^{(b)q} x_k}{\sum_{k=1}^n (U_{ik})^{(b)q}} \quad (5)$$

5. Calculate the membership $U^{(b+1)}$. For $k = 1$ to n , calculate the following:

$$I_k = \{i \mid 1 \leq i \leq c, d_{ik} = \|x_k - v_i\| = 0\} / I(6)$$

6. if $I_k = \emptyset$, then

$$U_{ik}^{(b+1)} = 1 / \sum_{j=1}^c (d_{ik} / d_{jk})^{2/(q-1)} \quad (7)$$

else

$$U_{ik}^{(b+1)} = 0, \text{ for all } i \text{ which is not subset of } I \text{ and } \sum_i u_{ik}^{(b+1)} = 1; \text{ next } k.$$

7. If $\|U^b - U^{b+1}\| < \epsilon$, stop; otherwise,

set $b = b + 1$ and go to step 4.

(d) Penalized FCM Algorithm

The lack of classical FCM algorithm is it does not take the spatial information about neighbourhood pixels into account that is it does the clustering process only based on gray levels without considering pixels of the image in segmentation which takes the FCM to noise sensitive. So neighbourhood pixel information has been considered [2] during classification.

$$J_{PFCM} = \sum_{k=1}^n \sum_{i=1}^c (U_{ik})^q d^2(x_k, v_i) + \gamma \sum_{k=1}^n \sum_{j=1}^n \sum_{i=1}^c (U_{ik})^q (1 - U_{ij})^q w_{kj} \quad (8)$$

Where γ is the penalty factor c is the number of classes and c_{ij} is the probability that x_j belongs to class I and w_{kj} are the values of the spatial structure matrix W which can be defined as

$$W = (w_{jk}):$$

$$w_{ik} = \begin{cases} 1, & x_j \text{ and } x_k \text{ are neighbours} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

IV. PROPOSED TECHNIQUE

The classical PFCM gives best result for overlapped data set and comparatively better than k-means algorithm but still it takes number of iteration. So in this paper, we include Markov Random Theory [1] by taking the class probability of a pixel depends on class memberships of its (spatial) neighbour clusters. According to our proposed technique neighbouring pixels can be influenced on current pixel by changing the equation (4)

$$U_{ik} = P_{ik} \frac{\sum_{j=1}^c (d_{ik} / d_{jk})^{2/(q-1)}}{\sum_{j=1}^c (d_{ik} / d_{jk})^{2/(q-1)}} \quad (10)$$

$$\text{Where } P_{ik} = \frac{\# N_k^i}{\# N_k}$$

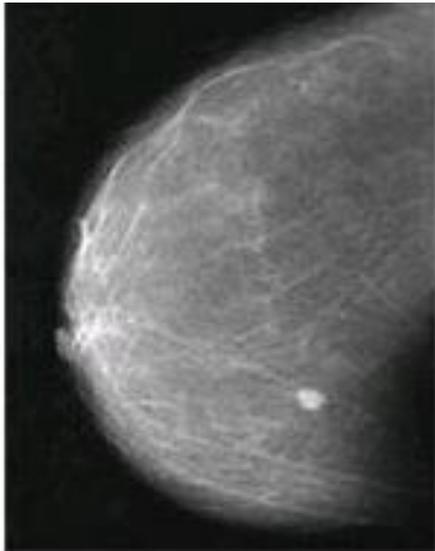
where $\# N_k$ denotes the total number of points in the neighbourhood of x_k and $\# N_k^i$ denotes the number of the neighbouring points that belongs to cluster i

By adding the penalty factor with respect to equation number (8)

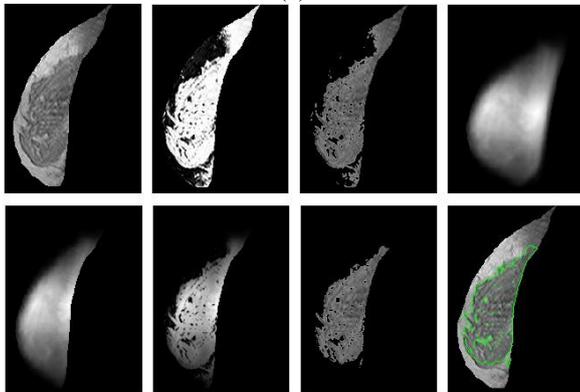
$$J_{EPFCM} = \sum_{k=1}^n \sum_{i=1}^c (P_{ik})^q d^2(x_k, v_i) + \gamma \sum_{k=1}^n \sum_{j=1}^n \sum_{i=1}^c (U_{ik})^q (1 - U_{ij})^q w_{kj} \sum_{j=1}^c (d_{ik} / d_{jk})^{2/(q-1)} \quad (11)$$

At last the procedure for finding maximum membership will assigns the object k to the class C with the highest membership:

$$C_k = \arg_i \{ \max (u_{ik}) \}, i = 1, 2, \dots, c. \quad (12)$$



(a)



(b)

Fig 7 a) Original Image b) classification result using PFCM algorithm

V. CONCLUSION AND FUTURE WORK

Breast cancer is one of the leading cancers which cause a number of deaths among women. In this paper we have presented a novel hybrid approach to identify the presence of breast cancer mass and calcification in mammograms using Markov random theory with penalized fuzzy k-means clustering. Combining these we have successfully detected the breast cancer area in mammograms images. The results shows that this system can facilitate the doctor to detect the breast cancer in the early stage of diagnosis as well as classify the total cancer affected area since it offers improved accuracy and dramatically reduced computational requirements.

VI. REFERENCES

- [1] YONG YANG Image segmentation based on fuzzy clustering with neighborhood information.. Optica Applicata, Vol. XXXIX, No. 1, 2009
- [2] Shuying Huang , Yong Yang Image segmentation by fuzzy c-means Clustering algorithm with a novel Penalty term Computing and Informatics, Vol. 26, 2007, 17–31
- [3] Nalini singh, Aambarish G Mohapatra, Breast Cancer Mass Detection in Mammograms using K-means and Fuzzy C-means Clustering, International Journal of Computer Applications (0975 – 8887) Volume 22– No.2, May 2011
- [4] Suman Tatiraju, Avi Mehta Image Segmentation using k-means clustering, EM and Normalized Cuts
- [5] Amit Chaudhary Tarun Gulati, Segmenting digital images using edge detection.
- [6] <http://www.cancercenter.com/treatments/cadmammography/>
- [7] Sara Deghani and Mashallah Abbasi Dezfooli, A Method For Improve Preprocessing Images Mammography, International Journal of Information and Education Technology, Vol. 1, No. 1, April 2011
- [8] <http://www.nationalbreastcancer.org/mammogram>
- [9] <http://www5.komen.org/BreastCancer/Statistics.html#mammography>
- [10] Daozheng Chen Expectation-Maximization Algorithm
- [11] http://en.wikipedia.org/wiki/Iris_flower_data_set
- [12] http://www.konicaminolta.com/about/research/special_healthcare/mammography.html

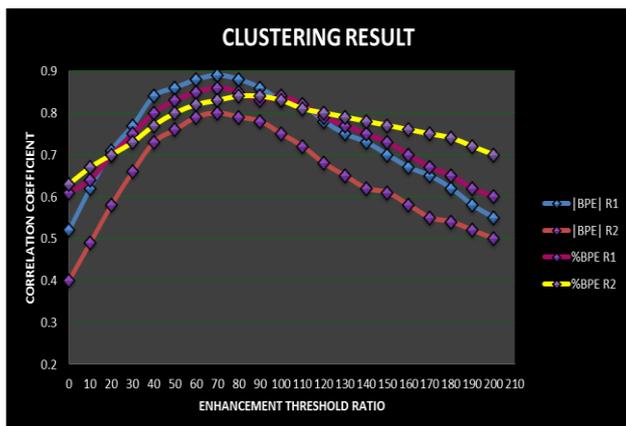


Fig 8. PFCM clustering results for enhancement ratio and correlation coefficient