

# A Novel DBSCAN Approach to Identify Microcalcifications in Cancer Images with Noise

Sandhya G  
JNTU, Hyderabad,  
AndhraPradesh, India

D Vasumathi  
Dept., of CSE  
JNTU, Hyderabad,  
AndhraPradesh, India

G T Raju  
Dept., of CSE, RNSIT  
Bangalore, Karnataka  
India

---

**Abstract:** Cancer is the most deadly disease among the human life. Breast Cancer is one of the most common cancers in this industrialized world and it is the most common cause of cancer related death among worldwide. Many segmentation technologies and clustering technologies like K-Means, K-Mediod, CLARANS etc have been proposed to identify the microcalcifications but this paper presents our new approach of identification of cancer cells in the images containing with noise and the performance analysis.

**Key Words:** Cancer, Breast Cancer, Segmentation, Clustering, K-Means, K-Medoid, CLARANS.

---

## 1. INTRODUCTION:

The most frequently diagnosed cancer in females is breast cancer, the variation may be due to the racial and genetic differences, cultural differences and the environmental factors that are varied throughout the world. There are two main types of breast cancer non-invasive (in situ) and invasive breast cancer. In non-invasive breast cancer the cancer cells remains within their place of origin and they have not spread to the breast tissue around the duct or lobule. In this we have two subtypes Ductal Carcinoma InSitu (DCIS) which is precancerous lesion and Lobular Carcinoma In Situ (LCIS) which is not precancerous and may increase the risk of cancer in both breasts. In invasive breast cancer it spreads outside the membrane that lines a duct or lobule, invading the surrounding tissues. Cancer stages with I, II, III, IV (will be discussed later) are invasive breast cancer.

### 1.1 Parts of Breast cancer:

- Milk Ducts: DCIS is the most common type of breast cancer which forms in the lining of a milk duct within the breast.
- Milk producing lobules: LCIS starts in the lobules of the breast, where breast milk is

produced. The lobules are connected to the ducts, which carry breast milk to nipple.

- Connective tissues: the connective tissue made up of muscles, fat and blood vessels will be rarely affected by the cancer cells.

### 1.2 Symptoms of breast cancer:

For women:

- Breast lump in the armpit which will be hard and having uneven edges which usually doesn't hurt.
- Change in the size, shape or feel of the breast or nipple.
- Fluid coming from nipple may be bloody, clear to yellow, green and look like pus.

For men:

- Breast lump and breast pain and tenderness, advanced stage may include bone pain, skin ulcers, swelling of arms, weight loss.

### 1.3 Issues in breast cancer:

- Mammography is one of the tool which is used to identify the breast lump screening with mammography is suggested by many doctors for the detection of breast cancer but the mammography is done once the age of women reaches 40 which fails for the early detection of breast cancer. Mammograms [12-13]are most often used in women over 40, unless they are at high risk, like carrying a mutation of the BRCA1 or BRCA2 gene. Having such a

mutation increases the risk of developing cancer five-fold. Even though many methods[1] are involved for the detection the radiation and the mammography techniques may increase the risk of breast cancer for younger age women

- Which types of tumors benefit most from early detection? The major problem is breast cancer at the early stages is not seen any case in the sense the symptoms will not give a clear picture of presence of the cancer cells.
- Monitoring and evaluation of individualized screening has to be taken care which means not only a single screening technique and different therapy can be suggested based on the risk of breast cancer[7-9].
- Many modern techniques have to be found for the faster evaluation of the results of the screening technologies [4]so that the accuracy of the results can be upgraded.
- Some studies have suggested women with the genetic mutations could be more sensitive to radiation because the genes are involved in fixing DNA problems. If those genes are damaged by radiation, they may not be able to repair DNA properly, raising the cancer risk. Researchers found women with a history of chest radiation in their 20s had a 43 per cent increased relative risk of breast cancer compared to women who had no chest radiation at that age. Any exposure before age 20 seemed to raise the risk by 62 per cent. Radiation after age 30 did not seem to affect breast cancer risk.
- Psychological issues also should be taken care in the sense proper knowledge about the disease, role and care of the family, sexual functioning, type and degree of disruption in life cycle tasks such as proper menstrual cycle, child bearing etc., personality and ability to cope up with stress, anxiety, prior psychiatric history, availability of psychological and social support and other factors.

## 2. METHODOLOGIES USED:

### 2.1 Improved k-Means clustering.

In the proposed methodology[5-6] there are mainly three important steps the first is the image which will be input to the machine which will be provided by the image database and filtering of the image and next important step is the implementation of k-means algorithm and lastly the classification of the image and producing the output image. The block diagram below shows the steps

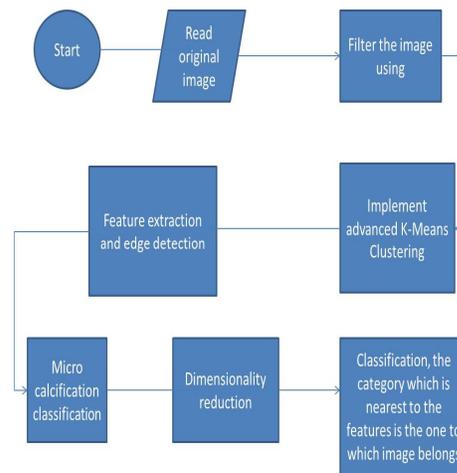


Figure 1: block diagram of the proposed adaptive k-means algorithm.

Clustering algorithms[10-12] can be applied to solve the segmentation problem. They consist in choosing an initial pixel or region that belongs to one object of interest, followed by an interactive process of neighborhoods analysis, deciding if whether each neighboring pixel belongs or not to the same object. In this work we use the K-Means[2-3] to resolve the mass detection task on mammograms using texture information obtained from Haralick's descriptors. The K-means algorithm is one of the simplest non-supervised learning algorithms class that solves the clustering segmentation problem . The method follows the usual steps to satisfy the primary objective: clustering all the image objects into K distinct groups. First, K centroids are defined, one for each group, being their initial position very important to the result. After that, it is determined a property region for each centroid, which groups a set of similar objects. The interactive stage of the algorithm is started, in which the centroid of each

group is recalculated in order to minimize the objective function. This function, for K-means, is the minimum square method, calculated by

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

$\|x_i^{(j)} - c_j\|^2$  → Distance metric from any point from any point  $x_i^{(j)}$  to the group  $c_j$

Thus, the J (objective function) represents the similarity measure of the n objects contained in their respective groups.

The proposed idea comes from the fact that the k-means algorithm discovers spherical shaped cluster, whose center is the gravity center of points in that cluster, this center moves as new points are added to or removed from it. This motion makes the center closer to some points and far apart from the other points, the points that become closer to the center will stay in that cluster, so there is no need to find its distances to other cluster centers. The points far apart from the center may change the cluster, so only for these points their distances to other cluster centers will be calculated, and assigned to the nearest center. In the proposed method, we write two functions. The first function is the basic function of the k-means algorithm, that finds the nearest center for each data point, by computing the distances to the k centers, and for each data point keeps its distance to the nearest center.

## 2.2 Improved K-Medoid

Input: K: The number of segments D: An images

Output: A segmented image that minimizes the sum of the dissimilarities of all the pixels to their nearest medoid. Method: Convert image into gray scale; Equalize histogram; Store the equalized intensities into an array; Select randomly K medoids from array; Remove the selected medoids from array; Segment image using this medoids; Calculate the total cost T and store medoids and cost;

Repeat: Randomly select a non medoid Orandom from array and remove it from array ; Assign each remaining

pixel to the segment with the nearest medoid; Compute the new total cost Tnew of swap point Oj with Orandom if Tnew < T then swap Oj with Orandom to form the new set of k medoid Until array is not empty;

In the place of advanced K-Means we implement Improved K-Medoid along with the two phase approach to increase the performance and to obtain accurate results.

### Phase 1

Step 1: Threshold the image using Otsu's method.

Step2: Label the connected components of the binary image as L1, L2..., Ln where n is the number of connected areas.

Step3: Find the reference area  $A=(L1.area+L2.area+...+Ln.area)/n*10$ .

Step 4 : Copy all connected regions having area less than A into the output image.

### Phase 2

Step 1: The process begins in the region of interest.

Step 2: a) For each pixel in the identified region calculate the intensity difference between that pixel and eight neighboring pixels

b) If there is any pixel having intensity difference greater than a predefined value, put it into the new image of the same size as the input image.

c) Repeat the above two steps for two times(Done in order to create ample gap between the cancer nodules and noncancerous tissues).

Step 3: Erode each of the resulting images with their

Step 4: Combine all the resulting images to form the final output image

## 2.3 CLARANS:

CLARANS (A Clustering Algorithm based on Randomized Search) . CLARANS draws sample of neighbors dynamically. The clustering process can be presented as searching a graph where every node is a potential solution, that is, a set of k medoids. If the local optimum is found, CLARANS starts with new randomly selected node in search for a new local optimum. It is more efficient and scalable than both PAM and CLARA . A node is represented by the set of k objects {Om1, ..., Omk}. Two nodes are neighbors if their sets differ by

only one object  $S1 = \{Om1, \dots, Omk\}$ ,  $S2 = \{Ow1, \dots, Owk\}$

$$|S1 \cap S2| = k - 1$$

each node has  $k(n-k)$  neighbors. each node represent a collection of  $k$  medoids, each node corresponds to a clustering (dynamic). draws a sample of neighbors in each step of a search. if a better neighbor is found, moves to the neighbor's node

### 3 DBSCAN:

Relies on a density-based notion of cluster: A cluster is defined as a maximal set of density-connected points. Discovers clusters of arbitrary shape in spatial databases [9]with noise. Arbitrary select a point  $p$ . Retrieve all points density-reachable from  $p$  with respect to  $Eps$  and  $MinPts$ . If  $p$  is a core point, a cluster is formed. If  $p$  is a border point, no points are density-reachable from  $p$  and DBSCAN visits the next point of the database. Continue the process until all of the points have been processed.

### 4 PERFORMANCE EVALUATION:

We have implemented the same methodology for identifying the masses which is specified in the figure above. Compared to other technologies DBSCAN can give a better performance even when noise is included oin the image. Based on our study conducted we have the possible performance evaluated based on various parameters.

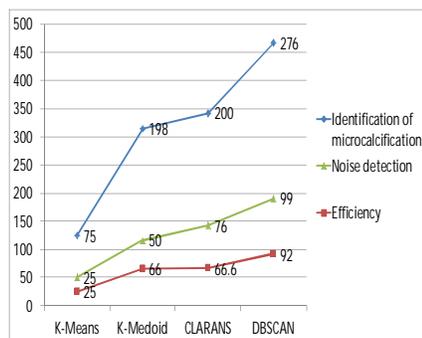


Figure.2: Graph representing the results

The above graph specifies the performance evaluation considering the parameters like identification of microcalcification, noise detection and efficiency for the methodologies discussed.

### 5. CONCLUSION:

Developing countries need a concentration on the diseases like cancer especially the cervical cancer and breast cancer which accounts for major death in women. This method provides an efficient way to identify the tumor and treatment can be taken. And awareness program has to be conducted in the country. This method provides an accurate result and which says that DBSCAN provides efficient way t identify noise in cancer cells and increase the performance.

### 6. REFERENCES:

- [1] R. Rajesh N. Senthilkumar, "A Note on Image Segmentation Techniques," *International J. of Recent Trends in Engineering and Technology*, vol. 3, no. 2, May 2010.
- [2] T. Velmurugan and T. Santhanam, "Computational Complexity between K-Means and K-Medoids Clustering Algorithms," *Journal of Computer Science*, vol. 6, no. 3, 2010.
- [3] Amit Yerpude and Dr. Sipi Dubey, "Colour image segmentation using K – Medoids Clustering," *Int.J.Computer Techology & Applications*, vol. Vol 3 (1), pp. 152-154, January 2012
- [4]S.Pradeesh Hosea, S. Ranichandra, and T.K.P.Rajagopal, "Color Image Segmentation – An Approach," *Color Image Segmentation – An Approach*, vol. 2, no. 3, March 2011
- [5]N. Senthilkumar and R. Rajesh, "A Note on Image Segmentation Techniques," *International J. of Recent Trends in Engineering and Technology*, vol. 3, no. 2, May 2010
- [5] Krishna Kant Singh and Akansha Singh, "A Study Of Image Segmentation Algorithms For Different Types of Images," *IJCSI International Journal of Computer Science Issues*, vol. 7, no. 5, September 2010.

[6] Catherine a. Sugar and Gareth m. James, "Finding the number of clusters in a data set :An information theoretic approach".

[7] G. Braz Jr, E. C. Silva, A. C. Paiva, A. C. Silva, Breast tissues classification based on the application of geostatistical features and wavelet transform, IEEE Computer Society Press, Tokio, 2007, pp. 227–230.

[8] J. Tang, R. Rangayyan, J. Xu, I. El Naqa, Y. Yang, Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances, Information Technology in Biomedicine, IEEE Transactions on 13 (2) (2009) 236–251.

[9] Wikipedia. Medical imaging. 2010.  
[http://en.wikipedia.org/wiki/Medical\\_image\\_processing](http://en.wikipedia.org/wiki/Medical_image_processing)  
ult  
rasound

[10] INCa, Internet site address: <http://www.inca.gov.br>  
accessed in 04/12/2010

[11] Domínguez, A. Rojas. Nandi, A. K. "Detection of masses in mammograms using enhanced multilevel thresholding segmentation and region selection based on rank", In: Proceedings of the fifth conference on Proceeding of the Fifth IASTED International Conference: biomedical engineering. 2007

[12] Campos, L. F. A. ;Costa, D. D.; Barros, A. K. "Segmentation on Breast Cancer Using Texture Features and Independent Component Analysis", In: Bioinspired Cognitive Systems, BICS 2008.

[13] Sheikholeslami, G., Chatterjee, S., Zhang, A., 1998. Wave- Cluster: A Multi-Resolution Clustering Approach for very Large Spatial Databases. Proc. 24th Int. Conf. on Very Large Data Bases. New York, p.428-439.

[14] D. D. Costa, A. K. Barros, A. C. Silva, Independent component analysis in breast tissues mammograms images classification using lda and svm., Information Technology Applications in Biomedicine - ITAB2007 - Tokyo. Conference on 6th International Special Topic (2007) 231–2