American Journal of Engineering Research (AJER) 2015

American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-4, Issue-4, pp-153-158
www.ajer.org
Research Paper Open Access

Mammogram Image Segmentation Quality EnhancementUsing Clustering Techniques

[1]Mrs. Sandhya G, [2]Dr. D Vasumathi, [3]Dr. G T Raju,
[1]Assoc. Prof, Dept. of CS&E ICEAS, Bangalore, Karnataka
[2]Prof, Department of CS&E JNTU-H, Kukatpally, Hyderabad
[3]Professor, Department of CS&E, RNSIT, Bangalore, Karnataka

Abstract: *Breast cancer is the most commonly observed cancer in women both in the developing and the developed countries of the world .Cancer refers to the uncontrolled multiplication of a group of cells in a particular location of the body. A group of rapidly growing or dividing cells may form lump or mass of extra tissue. These masses are referred to as tumors. Cancer cells are termed as malignant tumors. Any form of malignant tumor developed from breast cells is nothing but breast cancer. Breast cancer detection is the standard diagnosis and prognosis. Mammogram Image segmentation isbest method used for detection breast cancer by using various clustering techniques such as K-Means modified K-Means (KM), Fuzzy C-Means. The 14 Haralick features are extracted from mammogram image using Gray Level Co- occurrence Matrix (GLCM) for different angles.*

Keywords: *Mammogram, Breast cancer detection, K-Means, K-Mediods, Fuzzy c-means*

I. INTRODUCTION

The mammography is the most effective procedure to diagnosis the breast cancer at an early stage. This paperproposes mammogram image segmentation quality enhancement using various clustering techniques such as K-Means, modified K-Means (KM), Fuzzy C-Means. The 14 Haralick features are extracted from mammogram image using Gray Level Co- occurrence Matrix (GLCM) for different angles. The features are clustered by K-Means, Fuzzy C-Means (FCM) and modified K-Means algorithms to segment the region of interests (ROIs) for classification. The results of these clustering techniques compared and analyzed using Mean Square Error (MSE) and Root Means Square Error (RMSE). It is observed that the modified K-Means method gives better results compared to all the other methods clustering is defined as the optimal partitioning of a given set of *n* data points into specified number of subgroups, such data points belonging to the same group are as similar to each other [5]. The data points from two different groups share the different group. Image segmentation is considered as a clustering problem where each pixel corresponds to a pattern, and each image pattern region corresponds to a cluster. Some of hard clustering approaches do not consider overlapping of classes which occur in many practical image segmentation problems.

The main objective in cluster analysis is to group objects that are similar each other and separate other objects that are dissimilar by assigning them to different clusters. One of the most popular clustering methods is K-Means clustering algorithm. It classifies object to a pre-defined number of clusters, which is given by the user (assume K clusters). The idea is to choose random cluster centers, one for each cluster. These centers are preferred to be as far as possible from each other. In this algorithm mostly Euclidean distance is used to find distance between data points and centroids [7]. The Euclidean distance between two multidimensional data points are

$X = (x_1, x_2, x_3, \ldots, x_m)$ and

$Y = (y_1, y_2, y_3, \ldots, y_m)$ is described as follows:

$$D(X,Y) = \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2} \quad (1)$$

The K-Means method helps to minimize the sum of squared distances between all points and the cluster center. This procedure consists of the following steps, as described below.

K-Means Algorithm:

Require: $D = \{d_1, d_2, d_3, \ldots, d_n\}$ // Set of n data points.

K - Number of desired clusters

Ensure: *A* set of *K* clusters.
Steps-1: Arbitrarily choose *k* data points from *D* as initial centroids;
Steps-2: **Repeat:** Assign each point $d_i$to the cluster which has the closest centroid;
Calculate the new mean for each cluster;
Steps-3: **Until** convergence criteria is met.
Though the K-Means algorithm is simple, it has some drawbacks in final clustering, since it highly depends on the arbitrary selection of the initial centroids. Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar to each other, and items in different classes are as dissimilar as possible. Depending on the nature of the data and the purpose for which clustering is being used, different measures of similarity may be used to place items into classes, where the similarity measure controls how the clusters are formed. Some examples of measures that can be used as in clustering include distance, connectivity, and intensity.
In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels in cluster. These indicate the strength of the association between data element and a particular cluster. Fuzzy clustering method is a process of assigning membership levels, and then using them to assign data elements to one or more clusters.
Fuzzy C-Means Algorithm
Input: Dataset *X* of *n* objects with *d* features, value of *K* and fuzzy value *m* >1
Output: Membership matrix *Uij*for *n* objects and *K* clusters
Procedure:
Step-1: Declare a membership matrix *U* of size *n* X *K* .
Step-2: Generate *K* cluster centroids randomly within the range of the data or select *K* objects randomly as initial cluster centroids. Let the centroids be $c_1, c_2, \ldots, c_K$.
Step-3: Calculate the distance measure d_{ij}using Euclidean distance, for all cluster centroids C_i,j= 1, 2,...., K, and data objects x_i i = 1, 2,…, K .
Step-4: Compute the Fuzzy membership matrix U_{ij}
Step-5: Compute new cluster centroids c_j
Step-6: Repeat steps 3 to 5 until convergence.

II. Modified K-Means Clustering

The modified K-Means algorithm uses three basic steps
1. A data object can be a member of one lower approximation cluster.
2. A data object that is a member of the lower approximation of a cluster is also, amember of the upper approximation of the same cluster.
3. A data object that does not belong to any lower approximation is a member of at least two upper approximations.

According to the above steps, the lower approximation is a subset of the upper approximation. The difference between upper and lower approximation is called boundary region, which contains objects in multiple clusters. The membership of each objects in lower and upper approximation is determined by three parameters W_l,W_u and ∈the parameters W_l
and W_u correspond to the relative importance of lower and upper bounds, and W_l and W_u=1.
The ∈ is a threshold parameter used to control the size of boundary region.

Input: Dataset of n objects with d features, number of clusters k and values of parameters Wlower,Wupper and epsilon.
Output: Estimate Lower asV(K) and Upper as V'(k) of k clusters.
Procedure:
1. Randomly assign each data object as one Lower V(k) by step 2, the data object also belong to Upper V'(k) of the same cluster
2. Compute cluster centroids C_j.

If V (k) ≠ Ø and V'(k) - V(k) = Ø

$$C_j = \frac{\sum_{x \in V(k)} xj}{V(k)}$$

Else

V (k) ≠ Ø and V'(k) - V(k) = Ø

$$C_j = \frac{\sum_{x \in (V'(k) - V(k))} xj}{|V'(k) - V(k)|}$$

Else

$$C_j = W_l \, X \frac{\sum_{x \in V'(k)} xj}{V(k)} + W_u \, X \frac{\sum_{x \in (V'(k) - V(k))} xj}{|V'(k) - V(k)|}$$

3. Assign each object of the Lower V(k) or Upper approximation V'(k)of cluster i cluster respectively, for each object vector **x** , let d(x, C_j) is the distance between itself and the centroid **d** of cluster C_j , Let d(x, C_j) is min
   $1 \leq i , j = K$,
   Then ratio
   d(x, C_i) / d(x, C_j)
   $i \leq j, j \leq K$ is used to determine the member ship of **x** as follows.
4. Repeat the steps 2 and 3 until Convergence

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this paper the image samples are taken from the benchmark MIAS database for analyzing the proposed method. 14 Haralick features were extracted using Gray level Co-occurrence Matrix (GLCM). The sub-matrices of size 5 x 5 is used for constructing GLCM at different angle with distance d = 1 and then feature are extracted. Further feature are clustered into five groups by modified KM algorithm, each groups is partition into one segment, the each segmented image show in Figure 1. The same features are used to cluster using K-Means and FCM algorithms with five groups each groups is partition into one segment. The quality of segmentation result are measured using MSE and RMSE if the error value becomes low means that the better results. The MSE and RMSE values for the modified KM segmentation, FCM segmentation and K-Means segmentation are tabulated in tables1,2,3 and 4 respectively. According to the segmentation errors: means square error (MSE) and root mean square error (RMSE), the GLCM at distance 1 and angle 450 gives the best result for all tested image as shown in figures 1,2,3,4 and 5 K-means and FCM are helpful in early stage of clustering in medical diagnosis [7]. The cancerous mode can easily be separated from a fatty breast region as well as from dense region. As the number of cluster increases more and more information is obtained about the tissue which can't be identified by the pathologists

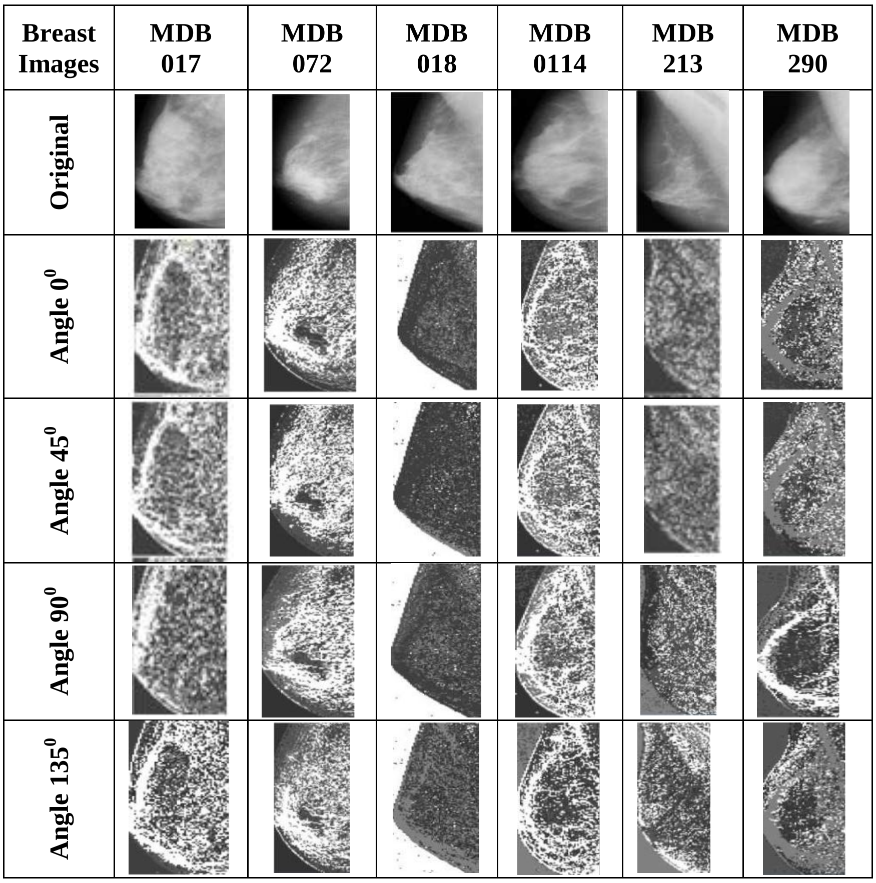

Figure.1 Results of Segmentation using modified K-MeansAlgorithm

Table.1 MSE values for modified K-Means Segmentation

Sample Image	Mdb 17	Mdb 72	Mdb 18	Mdb 114	Mdb 213	mdb290
Angle 0^0	9.75E+03	7.65E+03	6.27E+03	8.23E+03	5.63E+03	7.38E+03
Angle 45^0	8.05E+03	9.17E+03	6.34E+03	8.26E+03	5.77E+03	7.31E+03
Angle 90^0	9.82E+03	8.09E+03	6.02E+03	8.06E+03	5.79E+03	8.06E+03
Angle 135^0	9.11E+03	7.15E+03	5.74E+03	1.10E+04	6.18E+03	6.91E+03

Table.2 RMSE values for modified K-Means Segmentation

Sample Image	mdb17	mdb72	mdb18	mdb114	mdb213	mdb290
Angle 0^0	98.76	87.51	79.19	90.73	75.04	85.91
Angle 45^0	89.17	95.77	79.63	90.91	75.97	85.54
Angle 90^0	99.15	89.97	77.6	92.75	76.13	89.79
Angle 135^0	100.91	84.59	75.77	104.96	78.65	83.13

Table 3 MSE values for FCM segmentation

Sample Image	mdb17	mdb72	mdb18	mdb114	mdb213	mdb290
Angle 0^0	1.08E+04	1.18E+04	1.41E+04	8.77E+03	8.84E+03	1.10E+04
Angle 45^0	8.11E+03	1.06E+04	1.01E+04	8.41E+03	7.94E+03	9.43E+03
Angle 90^0	1.11E+04	1.30E+04	1.19E+04	9.97E+03	9.86E+03	1.07E+04
Angle 135^0	1.16E+04	1.29E+04	1.10E+04	1.17E+04	1.01E+04	1.09E+04

Table.4 RMSE values for K-Means segmentation

Sample Image	Mdb 17	Mdb 72	Mdb 18	Mdb 114	Mdb 213	Mdb 290
Angle 0^0	111.61	127.26	119.82	114.64	107.27	108.91
Angle 45^0	108.97	111.41	109.81	102.23	101.85	104.26
Angle 90^0	109.34	135.16	112.85	103.77	103.29	108.93
Angle 135^0	111.99	136.69	111.66	113.55	107.38	111.58

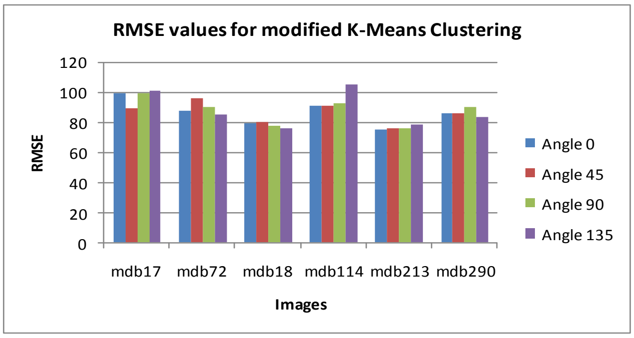

Figure.2 RMSE values for Segmentation using modified K-MeansClustering

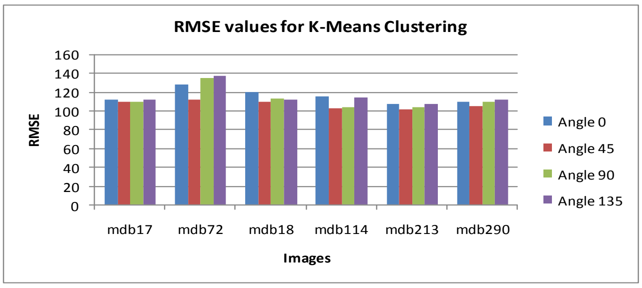

Figure.3 RMSE values for Segmentation usingK-MeansClustering

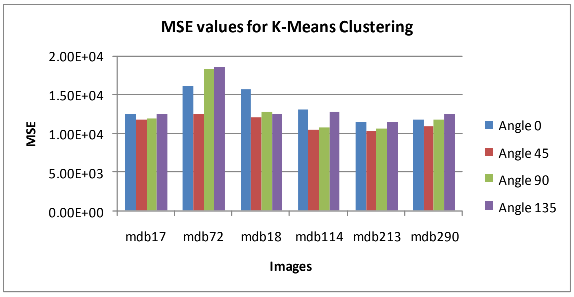

Figure.4 MSE values for Segmentation using K-MeansClustering

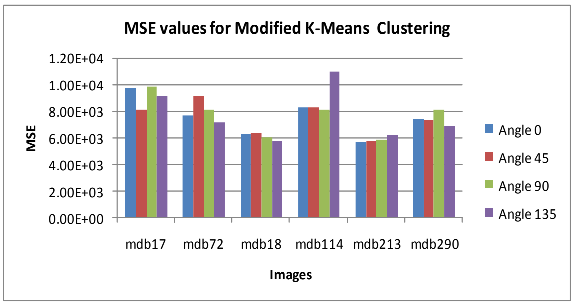

Figure 5 MSE values for Segmentation using modified K-MeansClustering

IV. CONCLUSION

In this paper, modified K-Means algorithm is proposed for mammogram image segmentation. The 14 Haralick features are extracted from mammogram image using Gray Level Co-occurrence Matrix (GLCM) for different angles. The features are clustered by K-Means, Fuzzy C-Means (FCM) and modified KM algorithms inorder to segment the region of interests for further classification. The performance of the modified KM segmentation is evaluated using MSE and RMSE measures. The proposed segmentation algorithm is compared with K-Means algorithm and FCM algorithm. It was observed that modified KM segmentationalgorithm out performs the benchmark K-Means algorithm and FCM algorithm. Further the resultant mammogram can be used for the detection of abnormalities in human breast like calcification, circumscribed lesions etc. This is the direction for further research.

REFERENCES

[1] Sampat, M.P., Markey, M.K., Bovik, A.C.: Computer-Aided Detection and Diagnosis in Mammography. In: Bovik, A.C.(ed.) Handbook of Image and Video Processing. ElsevierAcademic Press, Amsterdam (2005)

[2] Cheng, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P., Du,H.N.: Approaches for Automated Detection and Classification of Masses in Mammograms. Pattern Recognition 39(4), 646–668 (2006)

[3] Brzakovic, D., Luo, X.M., Brzakovic, P.: An approach to automated detection of tumors in mammograms. IEEE Transactions on Medical Imaging 9(3), 233–241 (1990)

[4] Li, H.D., Kallergi, M., Clarke, L.P., Jain, V.K., Clark, R.A.: Markov Random Field for Tumor Detection in Digital Mammography. IEEE Transactions on Medical Imaging 14(3), 565–576 (1995)

[5] Li, L.H., Qian, W., Clarke, L.P., Clark, R.A., Thomas, J.: Improving Mass Detection by Adaptive and Multi-Scale Processing in Digitized Mammograms. Proceedings of SPIE—The International Society for Optical Engineering 3661 1, 490–498 (1999)

[6] Székely, N., Tóth, N., Pataki, B.: A Hybrid System for Detecting Masses in Mammographic Images. IEEE Transactions on Instrumentation and Measurement 55(3), 944–951 (2006)

[7] Zheng, B., Mello-Thoms, C., Wang, X.H., Gur, D.: Improvement of Visual Similarity of Similar Breast Masses Selected by Computer-Aided Diagnosis Schemes. In: 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2007, April 12-15, pp. 516–519 (2007)

[8] Pappas, T.N.: An Adaptive Clustering Algorithm for Image Segmentation. IEEE Transactions on Signal Processing 40(4), 901–914 (1992)

[9] Sahiner, B., Hadjiiski, L.M., Chan, H.P., Paramagul, C.,Nees, A., Helvie, M., Shi, J.: Concordance of Computer-Extracted Image Features with BI-RADS Descriptors for Mammographic Mass Margin. In: Giger, M.L., Karssemeijer, N. (eds.) Proc. of SPIE Medical Imaging 2008: Computer-Aided Diagnosis, vol. 6915 (2008)

[10] Rangayyan, R.M.: Biomedical Image Analysis. CRC Press LLC, Boca Raton (2005)

[11] Fauci, F., Bagnasco, S., Bellotti, R., Cascio, D., Cheran, S.C., De Carlo, F., De Nunzio, G., Fantacci, M.E., Forni, G., Lauria, A., Torres, E.L., Magro, R., Masala, G.L., Oliva, P., Quarta, M., Raso, G., Retico, A., Tangaro, S.: Mammogram Segmentation by Contour Searching and Massive Lesion Classification with Neural Network. In:2004 IEEE Nuclear Science Symposium Conference Record, Rome, Italy, October 16–22, vol. 5, pp. 2695–2699 (2004)

[12] Petrick, N., Chan, H.P., Sahiner, B., Wei, D.: An Adaptive Density Weighted Contrast Enhancement Filter for Mammographic Breast Mass Detection. IEEE Transactions on Medical Imaging 15(1), 59–67 (1996)

[13] Zou, F., Zheng, Y., Zhou, Z., Agyepong, K.: Gradient Vector Flow Field and Mass Region Extraction in Digital Mammograms. In: 21st IEEE International Symposium on Computer-Based Medical Systems, CMBS 2008, Jyvaskyla, June 17-19, pp. 41– 43 (2008)

[14] Ferreira, A.A., Nascimento Jr., F., Tsang, I.R., Cavalcanti, G.D.C., Ludermir, T.B., de Aquino, R.R.B.: Analysis of Mammogram Using Self-Organizing Neural Networks Based on Spatial Isomorphism. In: Proceedings of International Joint Conference on Neural Networks, IJCNN 2007, Orlando, Florida, USA, August 12-17, pp. 1796–1801 (2007)

[15] Yuan, Y., Giger, M.L., Li, H., Sennett, C.: Correlative Feature Analysis of FFDM Images. In: Giger, M.L., Karssemeijer, N. (eds.) Proc. of SPIE Medical Imaging 2008: Computer-Aided Diagnosis, vol. 6915 (2008)