International Journal of Emerging Technologies in Engineering Research (IJETER)
Volume 5, Issue 4, April (2017) www.ijeter.everscience.org

Comparing the Performance of Texture and Cluster Based Segmentation Algorithms in SVM Classifier for Mammogram Analysis

S.K.Sajan[1], Dr.M.Germanus Alex[2]

[1]Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, India.

[2]Assistant Professor and Head, Department of Computer Science, Kamarajar Government Arts College, Surandai, Tirunelveli(Dt), Tamilnadu, India.

Abstract –

PURPOSE:

An approximate 12-15% reduction in breast cancer mortality is associated with mammography screening for women aged 40-69 years reported by the public health foundation of india[18]. Breast cancer is predicted to be a great threat to the society in the coming decade. Some women wonder about the risks of radiation exposure due to mammography. Modern-day mammography only involves a tiny amount of radiation — even less than a standard chest X-ray. The main risk of mammograms is that they aren't perfect. Normal breast tissue can hide a breast cancer so that it doesn't show up on the mammogram[20]. The problems with analysis of the mammogram image by Radiologists during breast cancer screening is reduced with our proposed segmentation algorithm and computer assisted techniques.

METHOD:

The enhanced mammogram image is given as input for the two segmentation algorithms Fuzzy C Means clustering and Texture segmentation using texture filters and the segmented image is classified using SVM classifier (non-linear) method.

Index Terms – Radiologists. Mammography, segmentation, screening, mortality, clusters

1. INTRODUCTION

Breast cancer screening includes three methods of diagnosis: 1) Breast self-exams. 2) Clinical breast exams. 3) Mammographic screening. Among the three methods of diagnosis breast self-exams and clinical breast exams is required for a women at the age of 20 and above, but these two methods are not efficient in reducing the mortality. Mammographic screening is needed annually at the age of 40 and above. A mammogram is an x-ray of the breast that uses very low levels of radiation (0.1-0.2 rads per picture). It captures the abnormal masses in the breast, and biopsy is needed to confirm the breast cancer [18]. In India the government is conducting cancer awareness programmes to detect breast cancer at its early stage before the symptoms were felt. Mammographic screening programs produce large number of mammograms to be frequently evaluated by radiologists. Computer assisted methods of diagnosis supports radiologists to evaluate the mammogram image and it helps to increase the accuracy in diagnosis of breast cancer.

Computer assisted techniques explores the advances in the digital image processing analysis and recognition technology. We present image segmentation techniques that partition the image into meaningful regions. This partitioning is done either by region extraction or by edge detection.

Breast Cancer Statistic Analysis in India, the average age of developing a breast cancer has undergone a significant shift over last few decades [5]. Consider the graph below (This is only a rough representation of the data):

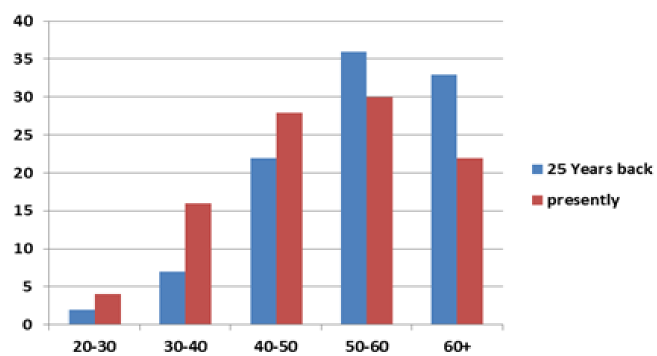


Fig 2 Breast Cancer in India

The horizontal line lower down represents the age groups: 20 to 30 years, 30 to 40 years and so on. And the vertical line represents the percentage of cases. The blue colour represents the incidence 25 years back, and maroon colour represents the situation today. 25 years back, out of every 100 breast cancer patients, 2% were in 20 to 30 years age group, 7% were in 30 to 40 and so on. 69% of the patients were above 50 years of age. Presently, 4% are in 20 to 30 yrs age group, 16% are in 30 to 40, 28% are in 40 to 50 age group. So, almost 48% patients are below 50. An increasing numbers of patients are in the 25 to 40 years of age, and this definitely is a very disturbing trend.

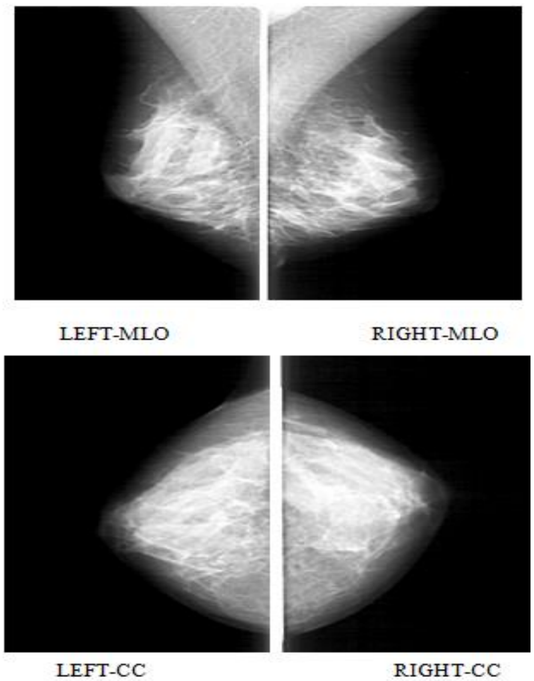


Fig3 Normal breast tissue

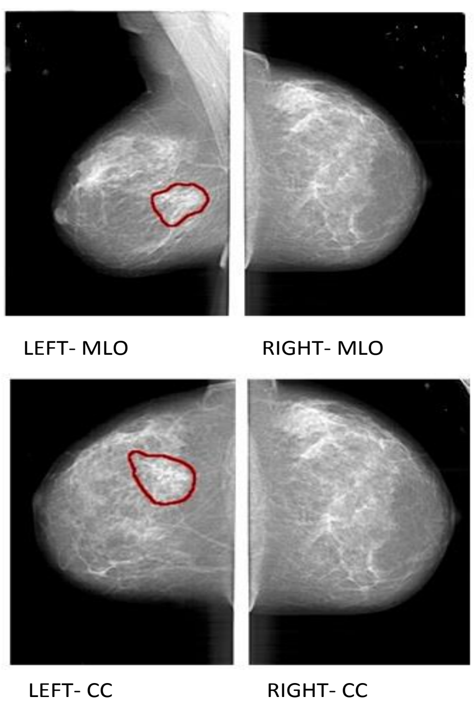


Fig4 Malignant breast tissue

Fig5 shows the normal breast tissue in a mammogram and fig6 shows the malignant breast tissue in a mammogram therefore the suspicious region was marked by the redlines. The dataset of mammogram images for analysis was taken from the DDSM(Digital Database for Screening Mammogram) . The primary purpose of the database is to facilitate sound research in the development of computer algorithms to aid in screening. Secondary purposes of the database may include the development of algorithms to aid in the diagnosis and the development of teaching or training aids.

2. ALGORITHMS AND TECHNIQUES IN COMPUTER ASSISTED DIAGNOSIS

Image processing techniques provide sufficient assessment strategies to categorize the mammogram image into normal and abnormal. MATLAB is the computational tool of choice for research, development and analysis [4]. It integrates computation, visualization and programming in a user friendly environment where problems and solutions are expressed in mathematical notation. MATLAB incorporates state of the art numerical computation software that is highly optimized for modern processors and memory architectures. MATLAB is an interactive system whose basic data element is a matrix [4]. Since MATLAB is a matrix-oriented language, basic knowledge of matrix analysis is mandatory.

In computer assisted diagnosis of breast cancer there are four stages in the evaluation process, they are image quality enhancement, segmentation, feature extraction and classification. Among the four stages segmentation plays an important role in identifying the regions of the breast in the mammogram image for analysis. Segmentation accuracy determines the eventual success or failure of computerized analysis procedures, for this reason considerable care should be taken to improve the probability of accurate segmentation. Segmentation technique basically divides the spatial domain, on which the image is defined into meaningful parts or regions. The segmentation algorithms are based on one of the following approaches

- Satisfying homogeneity property in image feature(s) over a large region.
- Detecting abrupt change in image feature(s) within a small neighbourhood

The first approach extracts the regions as a whole over which some measure shows the presence of homogeneity in feature value, while the second one detects the border between two regions and is commonly known as edge detection.

2.1 Segmentation Algorithms in diagnosis

Image segmentation is the process of partitioning an image into parts or regions. This division into parts is often based on the characteristics of the pixels in the image. For example, one way to find regions in an image is to look for abrupt discontinuities in pixel values, which typically indicate edges. These edges can define regions. Another method is to divide the image into regions based on color values.

Cluster analysis, also called segmentation analysis or taxonomy analysis, creates groups, or clusters, of data. Clusters are formed in such a way that objects in the same cluster are very similar and objects in different clusters are very distinct. Measures of similarity depend on the application.

Hierarchical Clustering groups data over a variety of scales by creating a cluster tree or dendrogram. The tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next level. This allows you to decide the level or scale of clustering that is most appropriate for your application. The Statistics Toolbox™ function clusterdata performs all of the necessary steps for you. It incorporates the pdist, linkage, and cluster functions, which may be used separately for more detailed analysis. The dendrogram function plots the cluster tree.

k-Means Clustering is a partitioning method. The function kmeans partitions data into k mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. Unlike hierarchical clustering, k-means clustering operates on actual observations (rather than the larger set of dissimilarity measures), and creates a single level of clusters. The distinctions mean that k-means clustering is often more suitable than hierarchical clustering for large amounts of data.

Gaussian Mixture Models form clusters by representing the probability density function of observed variables as a mixture of multivariate normal densities. Mixture models of the gmdistribution class use an expectation maximization (EM) algorithm to fit data, which assigns posterior probabilities to each component density with respect to each observation. Clusters are assigned by selecting the component that maximizes the posterior probability. Clustering using Gaussian mixture models is sometimes considered a soft clustering method. The posterior probabilities for each point indicate that each data point has some probability of belonging to each cluster. Like k-means clustering, Gaussian mixture modeling uses an iterative algorithm that converges to a local optimum. Gaussian mixture modeling may be more appropriate than k-means clustering when clusters have different sizes and correlation within them.

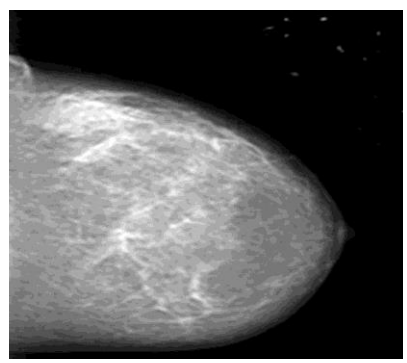
2.1.1 Texture segmentation using texture filters

Texture segmentation is a method to identify the regions based on their texture using filters. The use of image texture can be used as a description for regions into segments. There are two main types of segmentation based on image texture, region based and boundary based. Though image texture is not a perfect measure for segmentation it is used along with other measures, such as color, that helps solve segmenting in image. texture segmentation is concerned with automatically determining the boundaries between various textured regions in an image. Although quantitative texture measures, once determined, are useful in segmentation, most of the statistical methods for determining the texture features do not provide accurate measures unless the computations are limited to a single texture region. Both region-based methods and boundary-based methods have been attempted to segment textured images. These methods are analogous to those used for object-background separation methods. Texture segmentation is still an active area of research, and numerous methods, each designed for a particular application, have been proposed in the literature. However, there are no general methods which are useful in a wide variety of situations.

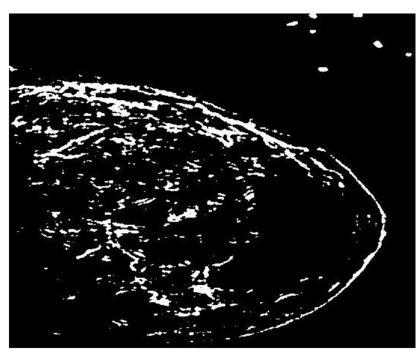
Extraction of texture features can be done with multiple methods like fourier spectra, edge detection methods, autocorrelation and decorrelation methods, dependency matrix method, microstructure methods, singular-value decomposition methods and so on. Among the methods edge detection method is broadly used in texture based segmentation. In edge detection method an edge map array E(j,k) is produced by some edge detector, such that E(j,k)=1 for a detected edge and E(j,k)=0 otherwise. Texture measure is thus defined as

$$T(j,k) = \frac{1}{w^2} \sum_{m=-w}^{w} \sum_{n=-w}^{w} E(j+m, k+n)$$

Where w is the dimension of the observation window.



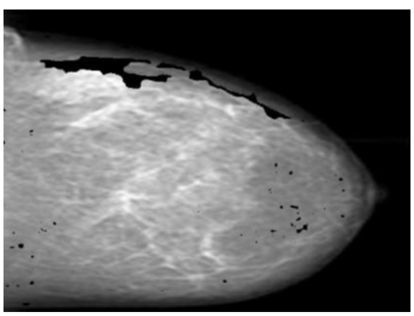
(a)



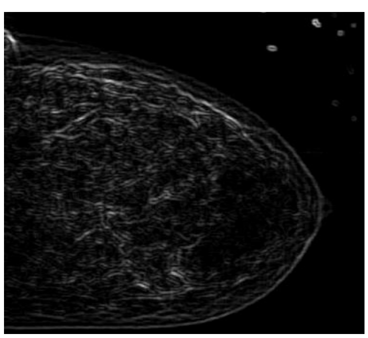
(b)

(c)

(d)



(e)



(f)

Fig 5 – Output of texture segmentation (a) original image (b) Rescaled image to segment the textures (c) compare the binary image (rough mask) with the original image. (d) Entropy filter to calculate the texture image (e) Outline boundary between two textures (f) segmented image

Algorithm:

i. Read input image
ii. Use Entropy filter to create a texture image
iii. Identify the intensity values of pixels along the boundary between the textures.
iv. Create rough mask for the bottom texture
v. Use rough mask to segment the top texture
vi. Compare the binary image (rough mask) to the original image
vii. Display the segmentation result.

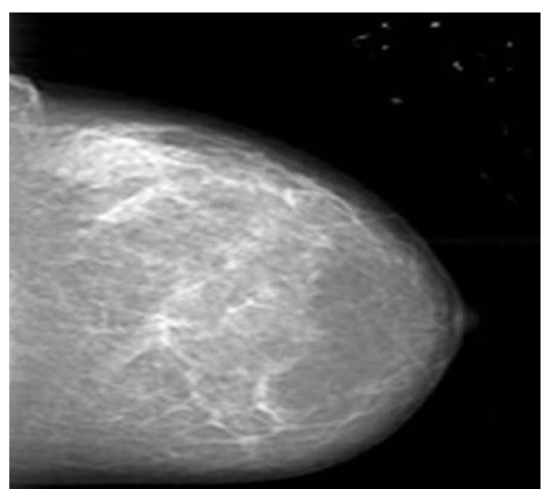
2.1.2 Fuzzy C Means Algorithm

Clustering analysis plays an important role in the medical field, it is a method of clustering objects into different groups. It attempts to organize unlabeled input objects into clusters, so that data points within a cluster are more similar to each other than those belonging to different clusters. To maximize the intra-cluster similarity while minimizing the inter-cluster similarity. In the field of clustering analysis, a number of methods have been put forward and many successful applications have been reported.

The FCM algorithm partitions the collection of n elements P={P_1,P_2…..P_n} into clusters based on certain criteria. Given a finite set of data, the algorithm returns a list of n cluster centres C={c_1,c_2…..c_n}and a partition matrix

$$W = w_{i,j} \in [0,1], i = 1,.......n \quad \text{and} \quad j=1,........n$$

Where each element $w_{i,j}$ tells the degree to which element p_i belongs to cluster c_j



(a)

(b)

Fig6 – Output of Fuzzy C Means algorithm - (a) input mammogram image (b) Segmented image

3. PROPOSED WORK

The proposed method deals with acquisition of the X-ray images of the breast through the mammogram. The mammogram image undergoes enhancement. Enhancement is the process of manipulating an image so that the result is more suitable than the original for a specific application [6]. Spatial filters are used for enhancing the mammogram image, the term spatial domain refers to the image plane itself, and image processing methods in this category are based on the direct manipulation of pixels in the image. The enhanced mammogram image is given as input for the two segmentation algorithms Fuzzy C Means clustering and Texture segmentation using texture filters and the segmented image is classified using SVM classifier using non-linear method.

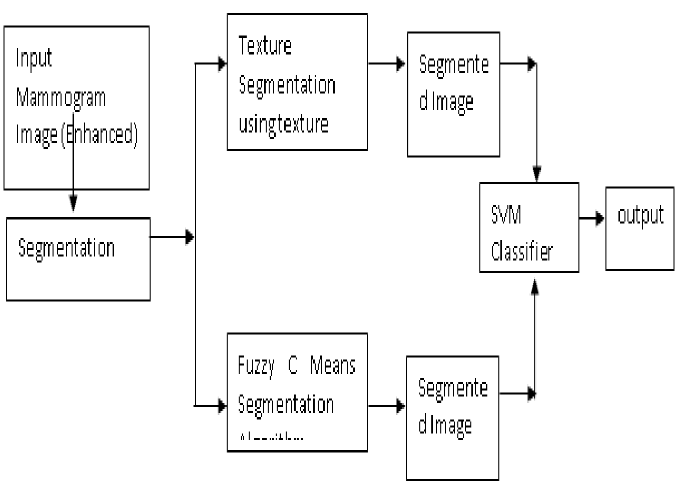


Fig7 ProposedApproach

4. EXPERIMENTAL RESULTS

Table1: Performance of the segmentation algorithms with SVM classifier

Mammogram Image (DDSM)	Algorithm	No of cases Misclassified	Accuracy in Classification (%)
Dataset1 (25 images)	Texture segmentation using texture filters	5	80
	Fuzzy C Means	3	88
Dataset2 (25 images)	Texture segmentation using texture filters	4	84
	Fuzzy C Means	3	88

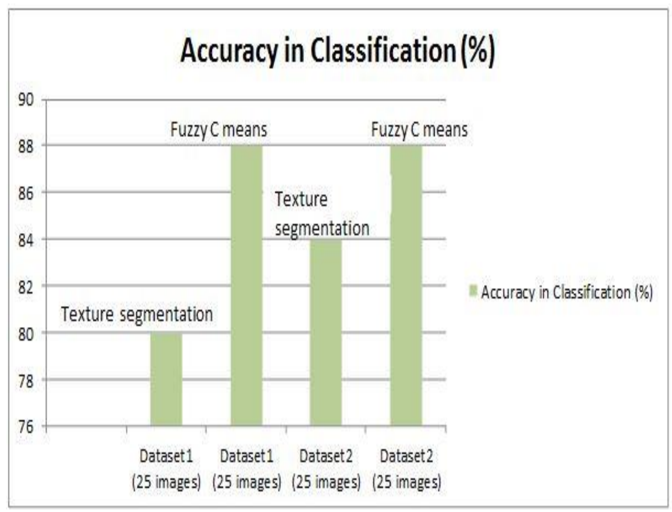


Fig8 Chart representation of the experimental result

5. CONCLUSION AND FUTURE WORKS

The Fuzzy C Means segmentation algorithm with SVM classifier yields good results compared to texture segmentation using texture filters. This method for diagnosis of breast cancer works efficiently with X-ray images of mammograms possessing denser breast tissues. In this context the mammogram images of DDSM dataset is used for evaluating the performance of the two segmentation algorithms with SVM classifier. The Support Vector Machine is trained with limited feature vectors based on the features attained by feature extraction that can be capable of diagnosing denser breast tissues. The proposed method can be added with more features so that the percentage of accuracy during classification can be improved in screening mammogram.

REFERENCES

[1] www.cancer.org/breast-cancer-early-detection
[2] www.cancer.org/cancer/breastcancer/.../breastcancerearlydetection/breast
[3] www.cancer.gov/cancertopics
[4] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, "Digital Image Processing UsingMATLAB",2nd Edition, McGraw Hill Education (India) Private Limited, 2013.
[5] http://www.breastcancerindia.net/bc/statistics
[6] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", Third Edition, Pearson Education, Inc, 2009.
[7] C.Cortes and V.Vapnik, "Support-vector networks," Machine learning, vol.20, no.3, pp.273–297, 1995.
[8] Milan Sonka, Vaclav Hlavac, Roger Boyle,"Digital Image Processing and Computer Vision", Cengage Learning, 2011.
[9] B. Chanda, D.Dutta Majumder, "Digital Image Processing and Analysis", Prentice Hall of India Private Limited,2007.
[10] Rajjan Shinghal, "Pattern Recognition Techniques and Applications", Oxford University Press, 2008.
[11] Madhuri A.Joshi, "Digital Image Processing An Algorithmic Approach", PHI Learning Private Limited,2009.
[12] www.mias.org
[13] www.birads.at
[14] F. Eddaoudi, F. Regragui, A. Mahmoudi, N. Lamouri, Masses Detection Using SVM Classifier Based on Textures Analysis, Applied Mathematical Sciences, Vol. 5, no. 8, 367 - 379, 2011.
[15] T. Ema, K. Doi, R.M. Nishikawa, Y. Jiang, J.Papaioannou, Image feature analysis and computer-aided diagnosis in mammography: reduction of false-positive clustered microcalcifications using local edge-gradient analysis, Med. Phys. 22 (2) (1995) 161–169, 1995.
[16] C.Y. Enderwich, E.M. Tzanakou, Classification of Mammographic tissue using shape and texture features, Proceedings of the 19th International Conference-IEEE/EMBS, 1997, pp.810–813, 1997.
[17] M.M. Eltoukhy, I. Faye, B.B. Samir, A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram, Computers in Biology and Medicine,40 (2010) 384–391, 2010.
[18] Preet K Dhillon, Breast cancer factsheet, Public Health foundation of india, 2011.
[19] http://naturalsociety.com/why-mammograms-waste-of-time-money
[20] http://www.breastcancer.org/symptoms/testing/types/mammograms/benefits_risks
[21] https://www.mathworks.com