Birads Score For Mammographic Images

Parveen Jaseela Regina .O. M
Regional Centre Of Anna University, Tirunelveli

Abstract
Breast imaging and screening has evolved for accurate diagnosis of breast cancer at earlier stages of development. 20% of malignant cases which has been proved as cancerous have been misinterpreted as non-detected cancers owing to technical problems in imaging procedure. These cancers is generally referred to as missed cancers (MC). Various image segmentation algorithms like K-Means Clustering , Expectation Maximization has proved for segmenting cancerous part from tumour part and statistically proved approaches for extracting features such as Area, Entropy, Clustering image, Uniformity, Mean evolved for classifying features of abnormal tissue. Evidences prove that about 65-80% of breast biopsies result in benign diagnosis and many false negative errors in retrospect considered as false positive biopsies. Hence clearer vision of classifying image based on textures provide wider view for radiologists in assessing the categories.

Keyword-malignant, missedcancers, K-Means Clustering,biopsies

1. INTRODUCTION

Breast Cancer is a malignant tumour which is a collection of cancer cells arising from the cells of the breast.BC is the one of the most leading cancer which exhibits an exceptionally heterogenous phenotype in histopathology[7].The availability of a computerized image analysis for automated quantification will enable development of an inexpensive image-based system for predicting disease survival and outcome. Definitive diagnosis of BC is performed by a pathologists via examination of tissue histopathology typically obtained via a needle biopsy. Hence, a clinician's ability to predict survival and disease outcome may be affected by inter and intra observer variability. introduction of a novel method that enables professionals to efficiently produce medical reports that are less error-prone and can be used in decision support systems without extensive post-processing methodology[3]. Filtering algorithm begins by storing the data points in a kd-tree Recall that, in each stage of Lloyd's algorithm, the nearest center to each data point is computed and each center is moved to the centroid of the associated neighbors[10]. Accordingly,K-Means Clustering algorithm results in a partitioning of the data space into Voronoi cells where each observation is a d-dimensional real vector. Staging is the process of determining the extent of the cancer and its spread in the body. Together with the type of cancer, staging is used to determine the appropriate therapy and to predict chances for survival. Staging system is used by the health care team to summarize in a standard way the extent and spread of the cancer. This staging can then be used to determine the treatment most appropriate type of cancer. Staging is the process of determining the extent of the cancer and its spread in the body. Together with the type of cancer, staging is used to determine the appropriate therapy and to predict chances for survival. Staging system is used by the health care team to summarize in a standard way the extent and spread of the cancer. This staging can then be used to determine the treatment most appropriate type of cancer.

The most widely used system in the U.S. is the American Joint Committee on Cancer TNM system. Besides the information gained from the imaging tests, this system also uses the results from surgical procedures. After surgery, a pathologist looks at the breast cancer and associated lymph nodes under the microscope. This information gained is incorporated into the staging as it tends to be more accurate than the physical exam and X-ray findings alone. . The system, called BI-RADS, includes seven standardized categories, or levels. Each BI-RADS category has a follow-up plan associated with it to help radiologists and other physicians appropriately manage a patient's care.BI-RADS is a quality assurance tool designed to standardize mammography reporting, reduce confusion in breast imaging interpretations, and facilitate outcome monitoring. It is a lexicon of standardized terminology, a reporting organization and assessment structure, a coding system and a data collection structure Results are communicated to the referring physician in a clear fashion with a final assessment that indicates a specific course of action. Results are compiled in a standardized manner that permits the maintenance and collection analysis of mammographic and outcome data. It is important for CAD to assess not only the computer performance, but also the performance by physicians. It is thus necessary to evaluate quantitatively and accurately by use of receiver operating characteristic (ROC) analysis whether the performance by physicians can be improved by use of the computer results. In fact, even if the ROC curve for computer results in detecting clustered micro calcifications on mammograms is substantially lower than that by radiologists, the ROC curve obtained by radiologists using the computer results can be improved[4]. The extent

of this improvement due to CAD was confirmed to be statistically significant. The higher the performance of the computer, the better the overall effect on the final diagnosis. The rest of the paper is organized as follows. ection II describes about lesion segmentation from normal tissue. Section III describes the experimental results and discussions. Section IV describes the conclusion and the further work.

2. PROPOSED METHOD

The Fig.1 describes the distinctions between detection and computer applications. The work presented here may be considered as the groundwork for an overall automated classification system for use in digital mammography (DM).

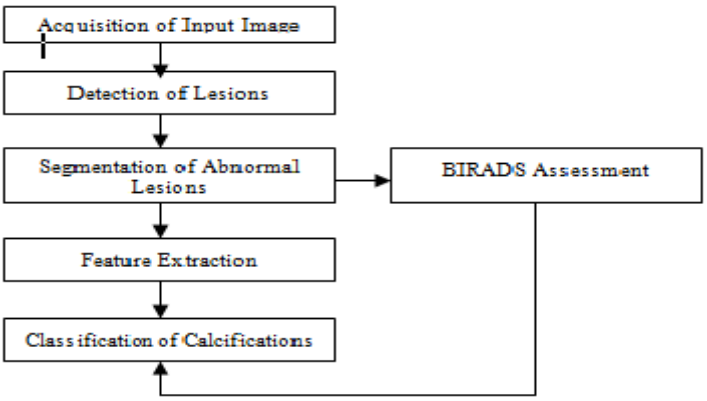

Fig.1.Block Diagram for Automated Mass classification System

Once the system is cued to the abnormality location, the classification consists of three main processing steps:
• Separate the abnormality from normal tissue
• Feature analysis.
• Classify the degree of malignancy.

3. Acquisition Of Input Image

The most common sign of breast cancer is a new lump or mass in the breast The doctor will also look for lump and calcifications. Lump or mass where the size, shape and edges of a lump sometimes can give doctors information about whether or not it may be cancer. On a mammogram a growth that is benign often looks smooth and round with a clear, defined edge. A digital mammogram also uses x-rays to produce an image of the breast, but instead of storing the image directly on film, the image is stored directly on a computer. This allows the recorded image to be magnified for the doctor to take a closer look. Even though mammography can detect tumors that cannot be felt, finding a small tumor does not always mean that a woman's life will be saved. Mammography may not help a woman with a fast growing cancer that has already spread to other parts of her body before being found. False negatives can happen**.** This means everything may look normal, but cancer is actually present. False negatives don't happen often. Younger women are more likely to have a false negative mammogram than are older women. The dense breasts of younger women make breast cancers harder to find in mammograms. False positives can happen. This is when the mammogram results look like cancer is present, even though it is not. The indifferences exists between the selection of the processes and the acquisition of the image helps to classify the calcifications and provide ease for the radiologists in assessing the category of BIRADS formulated by ACR.The acquisition is followed by the detection of lesions that leads to confusion.

3.1 Detection of Lesions

Lesions may be comprised of calcifications of type micro and macro which means tiny specks of calcium and large deposits often caused by aging. If calcifications are grouped together ,in a certain way, it may be a sign of cancer. Depending on how many calcium specks present, how big they are, and what they look like, doctors may suggest for further tests. Calcium in the diet does not create calcium deposits, or calcifications, in the breast. For detection the margins have to be sharply demarcated with an abrupt transition between the lesion and the surrounding tissue. Without additional modifiers there is nothing to suggest infiltration.

3.2 Segmentation of Abnormal Lesions

The most widely used segmentation algorithm preferred so far include Expectaion Maximization algorithm. The EM algorithm is a general method of finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values. There are two main applications of the EM algorithm. The first occurs when the data indeed has missing values, due to problems with or limitations of the observation process. The second occurs when optimizing the likelihood function is analytically intractable but when the likelihood function can be simplified by assuming the existence of and values for additional but missing (or hidden) parameters. The latter application is more common in the computational pattern recognition community .

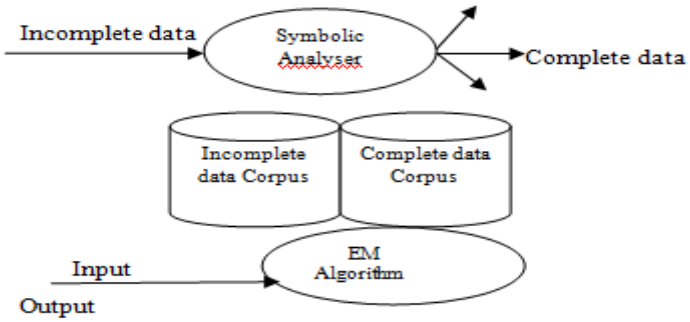

Fig. 2. Input and Output of EM Algorithm

The EM algorithm is more probabilistic in nature and more widely characterized for segmentation processes. The more appropriate algorithm that best suites for the spot detections is K-Means Clustering Algorithm since it can be refined by filtering processes. Given a set of observations (x_1, x_2, …, x_n), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets ($k \leq n$) $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS).

(1)

where μ_i is the mean of points in S_i.

k-initial "means" (in this case k=3) are randomly generated within the data domain . k clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means. The centroid of each of the k clusters becomes the new mean. The result of k-means can also be seen as the Voronoi cells of the cluster means. Since data is split halfway between cluster means, this can lead to suboptimal splits as can be seen in the "mouse. The Gaussian models used by the Expectation-maximization algorithm (which can be seen as a generalization of k-means) are more flexible here by having both variances and covariances. Attach label to each observation or data points in a set.You can say this "unsupervised classification"Clustering is alternatively called as "grouping".

Intuitively, if you would want to assign same label to a data points that are "close" to each other Thus, clustering algorithms rely on a distance metric between data points Sometimes, it is said that the for clustering, the distance metric is more important than the clustering algorithm. The EM result is thus able to accommodate clusters of variable size much better than k-means as well as correlated clusters .The parameters of the Gaussian mixture model were calculated for each texture in the image using EM algorithm. A likelihood function is calculated which gives the probability of a pixel as belonging to a particular class which forms the basis of labeling of the pixel. The segmentation step is the crucial stage addressed here; if it fails, the entire classification analysis fails. The goal of this work is to develop a robust method of segmenting breast masses from the normal background breast tissue. The success of automated classification requires knowledge of the mass, ambient normal-tissue, background border region, and the tumor area. The refining algorithms prevent the initial selection seed cluster points and the filtering processes makes ease of selecting the centroid that becomes a seed point to be a new cluster. The EM algorithm is applied to estimate the mean and variance of features for every texture in the image. At last a Bayesian classification rule is applied to attribute a label for each pixel by defining a likelihood function, which computes the probability for a given pixel as belonging to a given class. The Clustering algorithm will there provide simpler context in estimating the threshold value from which mean and variance need not be a probabilistic value. Algorithmically, very simple to implement .K-means converges, but it finds a local minimum of the cost function. Works only for numerical observations .K is a user input;

alternatively BIC (Bayesian information criterion) or MDL (minimum description length) can be used to estimate K. Outliers can considerable trouble to K-means.

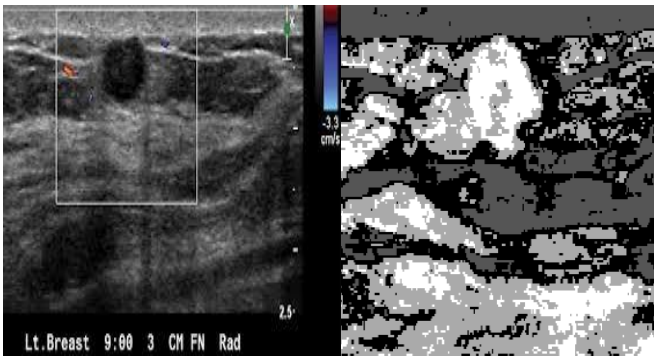

Fig. 3. Output of the Segmentation process

Where the cluster count K=3 and more the cluster count more the perfect segmentation and careful observation has to be performed for reducing the computational time.

4. D. Feature Extraction

To reduce the computational time required for extracting features for 256 pixel values the input image is quantized before applying the feature extraction process. Quantization can be done either to the pixel values or to the spatial coordinates. Operation on pixel values is referred to as gray-level reduction and operating on the spatial coordinates is called spatial reduction. Texture measures are derived using the gray-level co occurrence matrices. The EM algorithm is applied to estimate the mean and variance of features for every texture in the image. The feature extraction algorithms analyze the spatial distribution of pixels in grey scale images. The different methods capture how coarse or fine a texture is. The textural character of an image depends on the spatial size of texture primitives. Large primitives give rise to coarse texture and small primitives fine texture. To model these characteristics, spatial methods are found to be superior to spectral methods. Texture segmentation has long been an important topic in image processing. Basically, it aims at segmenting a textured image into several regions with the same texture features. An effective and efficient texture segmentation method will be very useful in applications like the analysis of aerial images, biomedical images and seismic images as well as the automation of industrial applications. Like the other segmentation problems ,the segmentation of textures requires the identification of proper texture-specific features with good discriminative power. Generally speaking, texture feature extraction methods can be classified into three major categories, namely, statistical, structural and spectral. In statistical approaches, texture statistics such as the moments of the gray-level histogram, or statistics based on gray-level co-occurrence matrix are computed to discriminate different textures. For structural approaches, "texture primitive", the basic element of texture, is used to form more complex texture pattern by grammar rules which specify the generation of texture pattern. Finally, in spectral approaches, the textured image is transformed into frequency domain To circumvent the above-mentioned issue, in this study, based on the fact that each interior point in a texture region must posses similar properties with its neighbors, a new statistical method is proposed. The key idea is that if the pixels of the input image can be classified into interior pixels and boundary ones, the interior pixels stand for the interior parts of texture regions, then the segmentation can be achieved by applying region growing on the interior pixels.

4.1 Statistical approaches

A frequently used approach for texture analysis is based on statistical properties of intensity histogram. One such measure is based on statistical moments. Various features are as follows,

(2)

which gives a measure of average intensity.

Variance, *(z)* (3)

which gives a measure of average contrast.

(4)

which gives a measures the uniformity of intensity in the histogram.

$$Ent \qquad (5)$$

which gives a measure of randomness.

Once the features are extracted from the input image set the threshold value thereby Area ,Entropy ,Clusterized image, Mean ,Variance are found using matlab code and the work will be made ease for the radiologists to assess the abnormalities. Statistical approaches like uniformity also specify the texture of the lesion and the probability of occurrence will be visually observed through the above mentioned features.The unit of specification is not mentioned as if predictions are based on the count of pixels and the image boundary.

5. RESULTS AND DISCUSSIONS

After number of simultations performed on the images varies statistical textures were obtained .

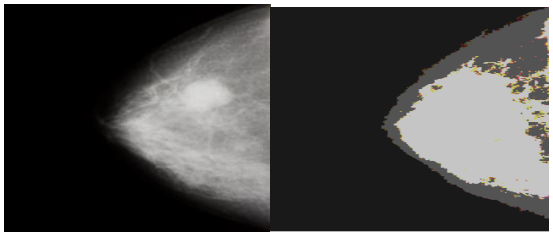

Fig .4 (a) Input Image Fig .4 (b)Clusterized Image

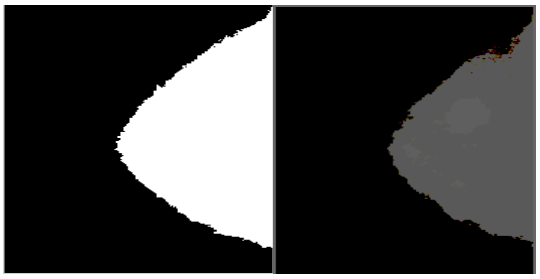

Fig .4 (c) Binary Image Fig .4 (d) Gray Image

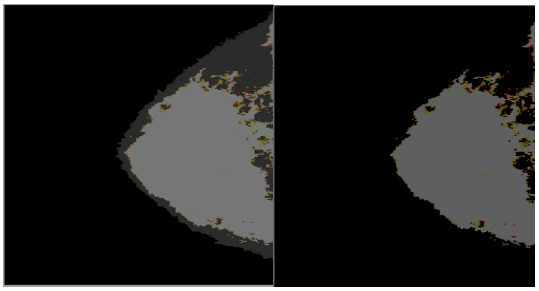

Fig .4(e) Area Fig .4(f) Mean

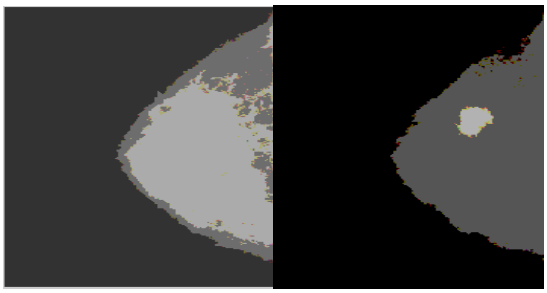

Fig .4(g) Variance Fig .4(h)Resultant Image

Figure. 4(a) represents the calcification of size of the input <400*267*3uint8>. (b) represents the cluster count that has been taken here as K=3. (c) represents the Binary image which provides the differentiation of the object and background.(d) represents the gray image that provides the threshold range of 0.3137.(e) gives the spread

of metastasis state where the area of the image is clearly shown as 3449 and (f) represents the mean of about 4 that provides the average smooth of the spread and makes it easier for assessing the image. The assessment will be assisted by the use of CAD tool with automated tool for maintaining the patient archiving ,Medical data with CAD assess that leads to better diagnosis. (g) represents the average contrast that is a measure of uniformity and it varies depends upon the cluster count and (h) provides the final resultant image of the clusterized image from which the analysis of the category of the image begins. The further processing has been made for large number of samples for classifying the abnormalities such as Compressed tissue, fatty and Non-fatty that leads to complexion and the categories for biopsies are verified using these samples.In this paper three samples have been taken and the output is verified using the resultant image,Variance analyzing the contrast.

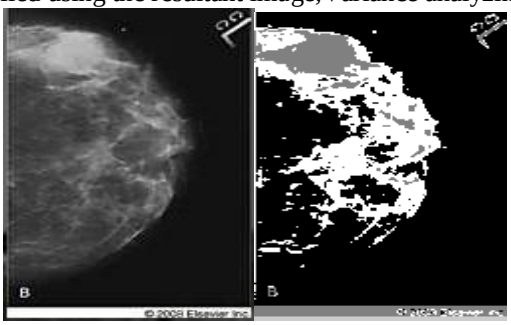

Fig. 5 (a) Input Image Fig. 5 (b) Clusterized Image

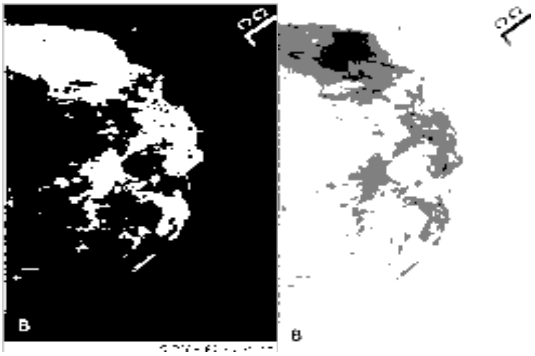

Fig. 5(c) Binary Image Fig. 5(d) Gray Image .

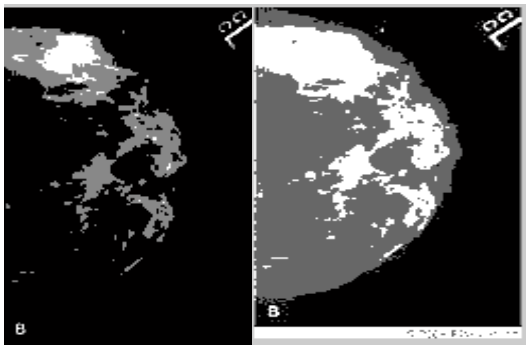

Fig. 5 (e) Area Fig. 5 (f) Mean

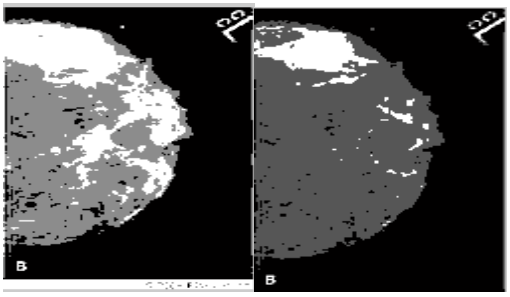

Fig. 5(g) Variance Fig. 5(h) Resultant Image

Figure. 5(a) represents the calcification of size of the input <200*126*3uint8>. (b) represents the cluster count that has been taken here as K=3. (c) represents the Binary image which provides the differentiation of the object and background.(d) represents the gray image that provides the threshold range of 0.3608(e) gives the spread of metastasis state where the area of the image is clearly shown as 2001 and (f) represents the mean the provides the average smooth of the spread and makes it easier for assessing the image. The assessment will be assisted by the use of CAD tool with automated tool for maintaining the patient archiving ,Medical data with CAD assess that leads to better diagnosis. (g) represents the average contrast that is a measure of uniformity and it varies depends upon the cluster count and (h) provides the final resultant image of the clusterized image from which the analysis of the category of the image begins.

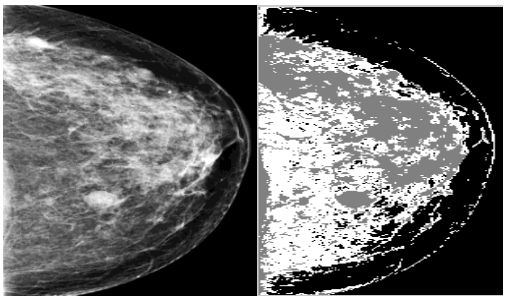

Fig. 6(a) Input Image Fig. 6 (b) Clusterized Image

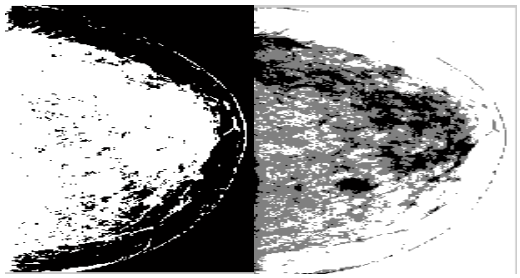

Fig. 6 (c) Binary Image Fig. 6 (d) Gray Image

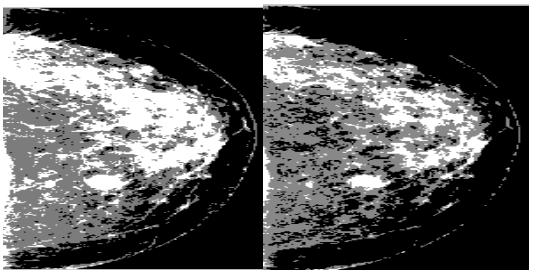

Fig. 6 (e) Area Fig. 6 (f) Mean

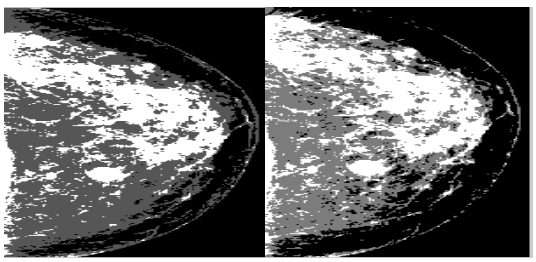

Fig. 6 (g)Variance Fig. 6 (h) Resultant Image

Figure. 6(a) represents the calcification of size of the input <440*227*3uint8>. (b) represents the cluster count that has been taken here as K=3. (c) represents the Binary image which provides the differentiation of the object and background.(d) represents the gray image that provides the threshold range of 0.3176.(e) gives the spread of metastasis state where the area of the image is clearly shown as 49813 and (f) represents the mean the provides the average smooth of the spread and makes it easier for assessing the image. The assessment will be assisted by the use of CAD tool with automated tool for maintaining the patient archiving ,Medical data with

CAD assess that leads to better diagnosis. (g) represents the average contrast that is a measure of uniformity and it varies depends upon the cluster count and (h) provides the final resultant image of the clusterized image from which the analysis of the category of the image begins.

Images	Cluster	Area	Threshold	Mean	Variance
Image a	[185,161,17]	3449	0.3137	4	170
Image b	[25,83,197]	2001	0.3608	50	63
Image c	[50,171,109]	49813	0.3176	63	228

Table 1. Performance Analysis of Image Types

The above experimental results were performed for inputs that comes under the category of uncompressed fatty tissue since the benign and malignant stages were absolute for cluster count K=3.For more beneficiary outputs the computation time of CAD processing becomes more tedious and further analysis have to be made for detecting compressed and fatty tissues.

6. Conclusion

Cluster value chosen as K=3 and more the cluster count more the efficient image segmentation. My future work includes the Classification of images by a better algorithm that aids CAD system to help Radiologists even more better**.**For better performance analysis cluster count can raised and segmentation along with computer aids can be used for better radiological assessment. Raising the cluster count may take more time consuming and using refining filtering algorithm procedures differentiation from tumour ,the cancerous part can be detected. The Radiologists will be able to provide report for biopsy proven cancer tissue and be able to categorize benign and malignant tissue.

7. Future Scope

The future scope includes the Classification of images by a comparison made with breast tomosynthesis and evaluated using thermographic analysis that aids CAD system to help Radiologists even more better for differentiating between masses and density. The Breast tomosynthesis method employ reconstruction method for improved detectability of microcalcifications (MCs).

References

[1] Marwa A. Shaaban , Abo El-Ata K. Aly , "Real-time ultrasound elastography: Does it improve B mode ultrasound characterization of solid lesions ." The Egyptian Journal of Radiology and Nuclear Medicine (2012) 43, 301-309.

[2] BasakOztan, Hui Kong, Metin N. G¨urcab, and B¨ulentYener,"Follicular Lymphoma. Grading using Cell-Graphs and Multi-Scale Feature Analysis."Medical Imaging 2012: Computer-Aided Diagnosis,Proc. Of SPIE Vol .8315,831516.

[3] Kuru, S. Girgin, K. Arda, U. Bozlar ,"A novel report generation approach for medical applications: The SISDS methodology and its applications." International Journal of Medical Informatics (2012).

[4] Arnau Oliver, Jordi Freixenet ,Joan Marti ,Elsa Perez , Josep Pont , Erika R.E. , "A review of automatic mass detection and segmentation in mammographic images." Medical Image Analysis 14 (2010) 87 -110.

[5] Georgia Giannakopoulou , George M.Spyrou , ArgyroAntaraki , IoannisAndreadi, " Downgrading BIRADS 3 to BIRADS 2 category using a computer -aided microcalcification analysis and risk assessment system for early breast cancer", Computers in Biology and Medicine 40 (2010) 853-859.

[6] Ajay NageshBasavanhallyand James Peter Monaco, "Computerized Image-Based Detection and Grading of Lymphocytic Infiltration in HER2+Breast Cancer Histopathology".IEEE Transactions on Biomedical Engineering Vol 57,No 3 March (2010).

[7] Hussian Fatakdawala, Jun Xu,Ajay Basavanhally,Gyan Bhanot ,Shridhar Ganesan, "Expectation-Maximization-Driven Geodesic Active Contour with Overlap Resolution (EMaGACOR): Application to Lymphocyte Segmentation on Breast Cancer Histopatholgy." IEEE Transactions on Biomedical Engineering ,Vol 57 No 7 July 2010.

[8] H.S. Sheshadri and A. Kandaswamy , "Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms".Computerized Medical Imaging and Graphics 31 (2007) 46–48."

[9] Dr.K Revathy,Roshni V. S , "Applying EM Algorithm for Segmentation of Textured Images." Proceedingsa of the World Congress on Engineering 2007 Vol I.

[10] Catarious, D.M., Baydush, A.H., Floyd, C.E., (2006). "Characterization of Difference of Gaussian filters in the detection of mammographic regions."Med.Phys.33(11)4104-4114.