ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 3, May 2016

Analysis and ingenious approach for segmentation of Breast Cancer cells in dual modality images for Computer Aided Diagnosis

Manjunatha V.G, Meharunnisa S.P

Abstract— ***This paper defines a unified approach for achieving ultrasound feature analysis through detection of tumor, in terms of their shape and size with experimental work for early identification of breast tumor. The main focus is to predict the abnormality in the breast tissues using three stages: Preprocessing, enhancement and Tumor segmentation. Preprocessing includes noise reduction and pectoral mass removal. Advanced techniques are adapted in image enhancement process, such as, CLAHE. And image segmentation is analyzed through Kmeans clustering. Size and stage of tumor is also detected in these steps. The evaluation of the proposed concept is applied to ultrasound images, a set of statistical features are calculated for every image. These features are then classified in normal and cancerous stages with few machine learning methods and compared the results with existing methods***

Index Terms— **Breast Cancer, CAD, K-means Clustering, Mammography.**

I. INTRODUCTION

Breast Cancer existed in ancient times and reference to this disease can be discovered going back as 3000 BC, in an Egyptian papyrus. Breast cancer is the second-most driving and normal explanation behind death in view of tumor among women. It has turned into a major health issue in the world over the period of last 20 years, and has increased over the recent years. As indicated by ICMR (Indian Council of Medical Research it is accounted for that One in 22 women in India is prone to experience the ill effects of breast cancer in her lifetime while in the US with One in 8 being a victim of this deadly cancer. In metropolitan cities such as Mumbai, Delhi and Kolkata, breast cancer accounts for 30 %, 26.9 % & 27.2 % respectively. Among women in Bangalore, it is almost 26.9 %, while the occurrences in men have been slightly less.

Prime prevention seems difficult because the causes of this disease still remain unfamiliar. So Early detection is the way to enhance breast cancer prognosis. Ultrasound mammography is the most widely recognized procedure utilized by radiologists as a part of the screening and determination of breast cancer, with ultrasound images supplying complementary information. The problems in mammography images such as high brightness value, dense tissues, noise and inefficient contrast level make analysis of these images a time consuming and hard task for physicians for identification of diagnostic signs. Although this technique is considered as the most reliable technique for early identification of breast cancer cells, its interpretation is very difficult. About 10 to 30 % of breast lesions are missed during routine screening due to over sightedness.

Mass classification is most important signatures of breast cancer. Masses which may be hidden or be similar to normal breast parenchyma is the primary objective of this study. Reading these digital mammograms is a challenging job for radiologists; hence, a Computer Aided Detection (CAD) System can provide a consistent second opinion to a radiologist and greatly improve the detection accuracy. This early detection can be reached by subjecting women at risk to a mammography once in every two years, since it takes around five years for a breast tumor to reach 1 mm in size, two years longer to reach 5mm and 1 to 2 years to measure 2 cm, quite big enough to detect by palpation. Now a days, magnetic resonance imaging (MRI), X-Ray mammography and ultrasound are imaging modalities routinely used to screen for breast cancer.

Cancer is detected by identifying either of four signatures of breast cancer:-

- Micro calcifications
- Masses

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 3, May 2016

- Architectural distortion.
- Bilateral asymmetry

It is well known that there is no approach at present, which is capable of curing cancer disease. But, it is well known that early identification of tumor can help in recovery and prolong patient life. The main reason for these errors is due to the way that radiologists rely on upon visual inspection. During diagnosing of an extensive number of mammograms, radiologists may get effectively exhausted, forgetting essential hints while studying the scans. For the hundreds of images scanned, only a few of them are cancerous. While diagnosing, some of them may be skipped, as the detection of suspicious and abnormal images is a recurrent mission that causes strain and tiredness to human eye. Using enhanced images and segment the suspicious area, extract features, select more accurate features and then classify them into appropriate category are the most important steps that computer aided detection systems should follow. However, due to more noise and low contrast, digital mammograms are the most challenging images to examine.

This project explores different Image Enhancement and segmentation techniques, Machine Learning algorithms and Classifying methods, which are used to increase the accuracy of predicting cancer regions. Also, a new CAD model is proposed to detect abnormal masses and statistically analyze the results of the findings in the Digital Mammogram. The project proposes algorithms based on image processing techniques for detection of breast cancer by keeping following points as main considerations

- Identification of masses in the given mammogram & ultrasound image.
- Adopting new techniques for detection of tumor cells
- Improve detection rate with existing systems.
- Quantify enhancement methods by statistical modeling.
- Develop a Graphical User Interface to view details of tested patients.

II. LITERATURE SURVEY

Masses are defined as space inhabiting lesions that are described by their margin properties and shapes. A benign neoplasm is smoothly marinated, whereas a malignancy is described by an unclear border that becomes more speculated with time [2]. If there is a slight variation in the X-ray image between glandular tissue and masses, they seems to appear as low contrast region with very often blurred.

The mammogram images are enhanced and the suspicious areas segmented, lesion features are extracted and are more accurately processed. For example, In MENCATTINI et al. [3], we see a technique where the denoising phase is based on a local iterative noise variance estimation. A new segmentation method is developed combining dyadic wavelet transformation with mathematical morphology. This innovative approach comprises of using the same algorithm for processing images to detect both masses and micro calcifications. Furthermore, the design of the algorithm goes toward the hardware implementation of the heavy core of the wavelet computation, permitting the realization of a fast real-time processing of the image.

In HU et al [4], a novel algorithm is used to identify suspicious lesions in mammograms. This algorithm uses both adaptive global and adaptive local thresholding segmentation on a mammogram images. The wavelet transform on the original mammograms removes the singularities and generated the lesion grayscale information. Subsequently, the wavelet transform on the histograms (PDF curves) removed the fluctuations. Hence, the global local minima can be found as the adaptive global threshold to implement the coarse segmentation. In the meantime, the morphological filter on the transforms images not only removes the unwanted components and the structure noise inside the suspected mass pattern but also enhances the gray-level feature and shape feature of lesions. Lastly, after a convolution between the coarse segmentation and the morphological enhancement filters the gray-level image to perform the fine segmentation.

In Sundarmai et al [5], fuses both histogram changes as an improvement method and CLAHE. It is interesting to note that from the subjective and quantitative measures, this proposed technique implements better contrast

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 3, May 2016
enhancement with maintaining the local information of the mammogram images as the conventional histogram equalization generally results in too much of contrast enhancement because of lack of control on the level of enhancement.

In Marija et al [6], digital mammogram images were explored by using basic point operators to highlight actual features and to extract quantitative information. The main focus was on enhancing contrast between the suspicious breast structures and contiguous tissues. Few basic algorithms like normalization, equalization and thresholding were applied and their actions have been displayed in both: the image and its histogram. Basic point operators could be used successfully to assist radiologists by increasing probability in early stage detection of breast cancer, even if the original image was not optimally taken.

Feature extraction is considered as the most effective step in mammogram classification and can be distinguished in the three following stages [7]:

- Statistical methods:
- Model-based methods:
- Signal processing methods

After feature extraction, selecting the appropriate classifier makes an important role for obtaining good result, there are many classifier have been used such as k-nearest neighbors (KNN), Neural Network (NN), support vector machines (SVM), Euclidean Distance and Bayesian classifier. In this paper two techniques are introduced for characterization of irregularity in digital mammograms. The Statistical and LBP features as a good tool for features extraction.

III. COMPUTER AIDED DETECTION – DIAGNOSIS

Growth of malignant tumor inside the breast tissue causes Breast cancer. Depending on the size of the tumor and spread to lymph nodes stages of breast cancer will be decided. Detection and diagnosis of breast cancer can be done by various imaging techniques and biopsy, the procedure in which small amount of tissue removed and looked at under a microscope to confirm whether cancer is present. Digital Mammography is a specific type of imaging technique that adapts a low-dose X-ray system to examine the specimen and is currently the most effective method for cancer detection before it becomes clinically palpable. It offers high quality images at a low radiation dose and is currently the only widely accepted method used for routine breast cancer screening. As double reading of mammograms and diagnosis by imaging techniques other than mammography, such as MRI, are expensive, the costs incurred is very high for the patients, the project proposes the development of Computer - Aided Detection (CAD) and diagnosis method, by developing efficient algorithms for early identification of breast masses with low cost mammography images. The Process flow of this approach is given in Figure 1.

We propose to preprocess the mammogram image, segment the suspicious mass region and extract features before analyzing it. We use Contrast Limited Adaptive Histogram Equalization (CLAHE) method for preprocessing the mammogram image before extracting the lesion.

A. CLAHE Enhancement Algorithm

In this approach the given images are converted to grayscale and then equalized in spatial domain. The algorithm for enhancing the relative image regions by adapting CLAHE technique and it is described below;

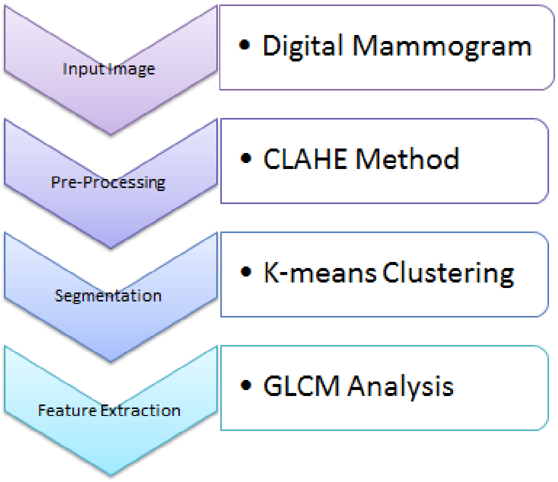

Fig 1: Process Flow

Step 1: First the given color image is converted to grayscale image using rgb2gray function. The transformed image is a 2D representation of intensity values.

Step 2: Grayscale image is divided into small bins of 8*8 matrices called as tiles.

Step 3: In equalization process the clip limit is selected.

Step 4: Histogram is calculated for each bin separately.

Step 5: For increasing the contrast enhancement, histogram representation of each bin is transformed in such a way that its height did not exceeds the selected clip limit.

Ultimately segmentation requires the smooth histogram thus uniform level of distribution is used for enhancement in this paper. The mathematical expression for changed grayscale levels for CLAHE technique with uniform distribution can be given as

$$g = gmax - gmin * P(f) + g \qquad (1)$$

Where gmax is maximum pixel value, gmin is the minimum pixel value. Where, g is computed pixel value and P(f) is the Cumulative probability distribution. Flow chart of the method is given in Figure 2.

The difference gmax-gmin represents the actual contrast of the image. Since CLAHE technique is essentially intended for maximum entropy of image therefore, it is more often used for image enhancement applications. Figure 3 describes output image of CLAHE method.

The main idea of the image segmentation is to assemble comparative pixels in homogeneous locales and the typical way to deal with do this is by common feature. In segmentation, clustering algorithms are most standard as they are intuitive and are anything but difficult to execute. The K -means clustering algorithm is one of the most widely used methods. It is an iterative technique used to divide an image into k clusters. K-means the easiest and simplest clustering techniques.

Mathematical expressions for k-means:

For a given image, compute the cluster means M:

$$M = \sum_{i=ci}^{k} \sum Xi \,/ Nk \qquad (2)$$

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 3, May 2016

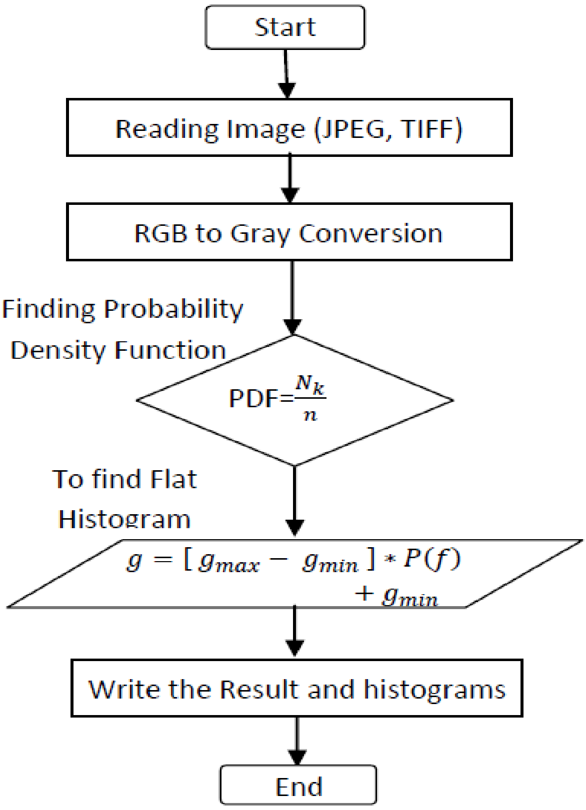

Fig 2: Data Flow of CLAHE Method

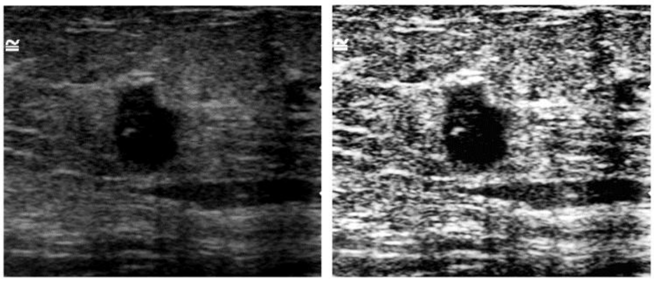

Fig 3 : CLAHE method preprocessing output

Now, compute the distance between the cluster centers to each pixel:

D(i)=arg min ||xi - Mk||2 , i=1,...N (2)

Repeat the above steps until mean value converges.

Algorithm for k means:

Step 1 : Give the number of cluster k.

Step 2 : Randomly assingn the cluster centers.

Step 3 : Estimate the center of cluster.

Step 4 : Find euclidean distance between each pixel to each cluster center.

Step 5 : If the distance is near to the center then move to that cluster.

Step 6 : Otherwise move to next cluster.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 3, May 2016

Step 7 : Re-assign the center.

Step 8 : Repeat the above process until the center doesn't move.

B. Feature Extraction

After the segmentation, the features can be extracted from each of the segmented images for further classification. Feature extraction is done using GLCM. GLCM is an arrangement of how frequently pixel intensity values occur in a pixel pair in an image. Each element (i, j) in GLCM specifies the number of intervals that the pixel with value i occurred horizontally adjacent to a pixel with value j. The resulting matrix will be analyzed and based on the existing information, the feature vectors are formed. Contrast, correlation, energy, entropy, homogeneity are prime features used in this approach. Contrast is used to measure the local variations in the GLCM matirx Homogeneity measures the closeness distribution among elements. Its extent is from 0 to 1. Homogeneity is 1 for a diagonal GLCM. Entropy gives the measure of randomness. The segments with hotspots should be having lesser entropy.

$$\text{Contrast} = \sum_{n=0}^{G-1} n^2 \{\sum_{i=1}^{G} \sum_{j=1}^{G} P(i,j)\}, \qquad |i-j| = n \tag{3}$$

$$\text{Homogeneity} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} P(i,j) \tag{4}$$

$$\text{Entropy} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) \times log(P(i,j)) \tag{5}$$

$$\text{Variance} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-u)^2 P(i,j) \tag{6}$$

These values provide a further insight into the type of tumor or the class of tumors which they are likely to fall into.

IV. CONCLUSION

The low contrast digital mammogram images are enhanced by using CLAHE. The Contrast enhanced image is segmented using K-means clustering, which is one of the simplest unsupervised learning algorithm for image segmentation. The mammogram images were collected from HCG hospital, Bangalore and the features extracted have been compared with publicly available mini-MIAS database. K-means clustering takes less computational time compared to other clustering techniques such as FCM. It provides information about the pixels from which the result of classification will be improved. To characterize these segmented images, a set of features based on texture are extracted using Gray Level Co-Occurrence Matrix (GLCM). Classification of masses into benign and malignant can be predicted and CAD performance can be compared in the future work using the same method.
.

REFERENCES

[1] www.icmr.nic.in.

[2] http://www.radiologyinfo.org/en/info.cfm?pg=mammo.

[3] "Mammographic Images Enhancement and Denoising for Breast Cancer Detection Using Dyadic Wavelet Processing", Mencattini et al., 0018-9456, 2008 IEEE.

[4] "Detection of Suspicious Lesions by Adaptive Thresholding Based on Multiresolution Analysis in Mammograms", Kai Hu, Xieping Gao, and Fei Li: IEEE Transactions On Instrumentation And Measurement, Vol. 60, No. 2, February 2011, IEEE.

[5] "Histogram Based Contrast Enhancement For Mammogram Images", M.Sundarami, K.Ramar, N.Arumugami, G.Prabini, Proceedings of 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN 2011), 2011 IEEE.

[6] "Basic Feature Extractions from Mammograms", Marija Dakovic and Slavoljub Mijovic, Mediterranean Conference on Embedded Computing MECO - 2012.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 3, May 2016
[7] "Breast Mass Classification using Statistical and Local Binary Pattern Features", Mohamed A. Berbar, Yaser. A. Reyad, Mohamed Hussain, 16th International Conference on Information Visualization, 2012 IEEE.

[8] "Automated Abnormal Mass Detection in the Mammogram Images Using Chebyshev Moments", Alireza Talebpour, Doman Arefan and Hamid Mohamadlou, Res. J. Appl. Sci. Eng. Technol., 5(2): 513-518, 2013.

[9] "Suspicious Lesion Detection in Mammograms using Undecimated Wavelet Transform and Adaptive Thresholding", Abhijit Nayak, Dipak Kumar Ghosh, Samit Ari, 978-1-4673-2818-0/13/$31.00 ©2013 IEEE.

[10] "A Computer-Aided Diagnosis System For Breast Cancer Detection By Using A Curvelet Transform", Nebi Gedik, Ayten Atasoy, Turk J Elec Eng & Comp Sci (2013) 21: 1002 { 1014 TUBITAK.

[11] "Malignancy Detection in Mammogram using Gray Level Gradient Buffering Method", Meenalosini S, Kannan E, International Journal of Cancer Research, ISSN:2051-784X, Vol.47, Issue.1, March 2013.

[12] "Particle Swarm Optimization Based Contrast Limited Enhancement for Mammogram Images", Shelda Mohan and T.R. Mahesh, 978-1-4673-4601-6/13 © 2013 IEEE.

[13] "A Computer Aided System for Breast Cancer Detection and Diagnosis", Hamada R. H. Al-Absi, Brahim Belhaouari Samir & Suziah Sulaiman.

[14] "Mass Detection in Digital Mammograms System Based on PSO Algorithm", Ying-Che Kuo, Wei-Chen Lin, Shih-Chang Hsu & An-Chun Cheng, 978-1-4799-5277-9/14.

[15] "Detection of Mammograms Using Honey Bees Mating Optimization Algorithm (M-HBMO)", R.Durgadevi, B.Hemalatha & K.Vishnu Kumar Kaliappan, 978-1-4799-2876-7/13, 2014.

[16] "Isolation of Breast Cancer Accumulation in Mammograms for Improving Radiologists Analysis", Prapti V. Patil, Rode Yogesh & K. V. Kale, ©2014 Engineering and Technology Publishing, International Journal of Electrical Energy, Vol. 2, No. 2, June 2011.