Detection and Comparison of Breast Cancer Using Clustering Algorithm and Bayesian Networks

[1]Nancy.P, [2]Leo Sahaya Dharshini.A
[1]Assistant Professor, [2]Assistant Professor
[1]Biomedical Department, [2]Biomedical Department
[1] Sri Sakthi Institute of Engineering and Technology, Coimbatore, Tamilnadu, India, [2] RVS Educational Trust's and Group of Institutions, Dindigul, Tamilnadu, India

Abstract : Breast cancer is one of the leading causes of women's death. In most of the cases the curing is very difficult due to the inability in early detection of the cancer symptoms. One of the main reasons behind this is the insufficient number of experienced doctors to examine the huge data and other types of diagnostic facilities. Using computational techniques we can identify micro calcification and gradiation masses in different groups of benignancy and malignancy stages of breast cancer. The mammogram image considered for the analysis is detected using clustering segmentation techniques and the results of this segmentation is compared with other techniques.The Bayesian network is a powerful tool to describe the uncertainty and complexity of many problems in the real world. Hence the grades of cancer is found using Bayesian network. It determines the stage of cancer whether it is beningn or malignant as grade 0, grade 1, grade 2 and grade 4.

***Keywords*-Tumor, Preprocessing, Enhancement, Segmentation, Clustering Algorithm and Bayesian Network, Grades or Stages.**

I. INTRODUCTION

Breast cancer is the second leading cause of death in women. Early diagnosis of the breast cancer is required where the survival rate is increased(1). One in Eight women's are affected by means of breast cancer leading cause of death in the world (2). The best method for earlier detection is done by mammography (3). In order to diagnose the severity of the disease now a day's Image processing techniques are proposed. Mammogram image are bit complex to detect cancer, due to noise which affects the overall diagnosis of the patient (4). One of the strongest known risk factors for cancer is mammographic density (5,6) which is caused due to gender, age, gene mutation and family history. Mammography is not the only non-invasive diagnostic technique allow to diagnose the breast cancer in early stage there is a clustering technique used to detect a micro calcification (7). In this the images is divided into segments or parts based on few important characteristics after a preprocessing (8,9).

The work which was survived to detect breast cancer in early stages consist of Preprocessing which improves the visual quality by removing the noise in the mammogram image (10). The mass density in a breast was identified using feature extraction (11). Whereas the Morphological techniques was another methodology used where the structural element id detected in breast (12). Another most commonly used technique is Computer Aided Detection (CAD) and diagnosis system which have gained more and more space in modern medicine (13) in identifying the disease and breast cancer. Frame work of mammogram will provide an advantage for managing information and image technique (14). Feature of digital mammography can be extracted with soft computing methods such as neural network and fuzzy logics(15), Spectrum Vector Machine for early detection of breast cancer. Above mentioned methods uses segmentation to process mammographic image (18,19). K- Mean, Fuzzy Logic algorithm are two techniques used to find the suspicious area in mammogram image (20).

The proposed work here gives an early detection and comparison of breast cancer and its grades, were grades indicate stages of the breast cancer as grade 0, grade 1, grade 2 & grade 3 with help of feature extraction in a mammography image.

The pre-processing and Image enhancement system helps to improve the image quality by suppressing unwanted distortions in the mammogram image captured. The image is then segmented as parts and is grouped according to the characteristic in order locate the place where the tumor is. Breast cancer is one of the leading causes for women death. The proposed work here is to detect and compare the severity of cancer in breast using feature extraction by digital image processing technique in which the accurate data can help the doctors to analyze the location and severity of cancer. Feature is extracted in the suspicious area from mammography image with the help of clustering and Bayesian network which gives us a result in which stage the cancer is.

II. METHODS & ALGORITHMS

The proposed method is to detect the abnormalities present in the breast. MATLAB software is used to process the different methods. The image processing technique is used here were the features is extracted from the mammogram image using 5 different steps shown in Fig.1

Proposed Block diagram

Fig 1 Block Diagram of different stage of Processing

2.1Input Image - Mammogram Image

The first and foremost step is to input an image in M X N size. The input image used here is the mammogram image which is converted it into Gray scale image. It is the high quality imaging modality to detect the breast cancer that is why mammogram image is taken as the input image. There will be distortion in mammogram image and the noise in the image should be removed. In order to remove the noise in the mammogram image we proceed with the next step image preprocessing where the image acquisition is done.

2.2 Preprocessing – Image Acquisition

Once getting the input image the mammogram image is preprocessed using image processing technique called Image Acquisition. In this method the image will be first scaled and then if any noise present in the image will be filtered using a median filter. In this the noise usually present in the images are removed The noise usually found in these kind of imaging are guassion noise, speckle noise, impulsive noise and etc. By using a median filter these noises can be removed, it is very important to remove the noise since due to the noise there will be false result obtained. Median filter is used to smoothen the non repulsive noise in the absence of edge blurring and also retain edge information whenever there is a high density impulse noise. The pixels are differentiated by comparing each pixel with its neighborhood, after setting size of the neighborhood and threshold for the comparison. A pixel whose intensity not matched with majority of its neighboring pixels is considered to be an impulse noise and these noise pixel get replaced with the median pixel values in that neighborhood. Using this method the image quality will be improved in order to analyze for suitable processing this is done and as a next step the image should be enhanced.

2.3 Image Enhancement

Image enhancement is to enhance the contrast of the images. Here the histogram equalization is used to enhance the preprocessed image, this adjust the intensity values and make the image to enhance as next step the image is processed to next step image segmentation.

2.4 Image Segmentation

The main goal of segmentation is to partition an image into regions. Segmentation algorithms generally are based on one of two basic properties of intensity values: discontinuity and similarity. In the first category, the approach is to partition an image based on abrupt changes in intensity, such as edges in an image. The principal approaches in the second category are based on partitioning an image into regions that are similar according to a set of predefined criteria. Thresholding region growing and region splitting and merging are examples of methods in this category.

2.5 Clustering Algorithm

Clustering refers to classification of objects into groups according to certain characteristic. In this the pixels which have a similar intensity are grouped to determine the region of interest. Three different clustering algorithms are used to determine the part where the tumor is and it is compared shown in Fig 2.

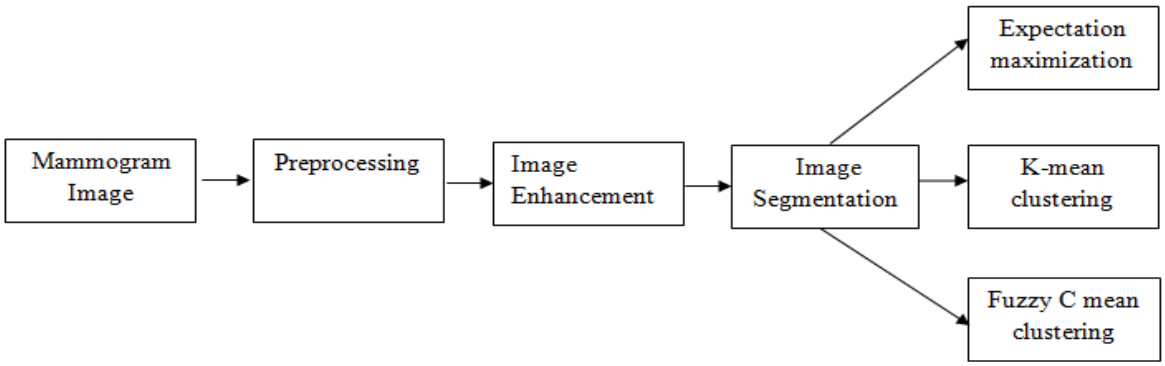

Fig 2 Comparison between the clustering Algorithms

2.6Expectation Maximization

- ✓ First, initialize the parameters to some random values.
- ✓ Compute the best value for given parameter values.
- ✓ Then, use the computed values to compute a better estimate for the parameters. Parameters associated with a particular value will use only those data points associated latent variable.
- ✓ Finally, iterate until convergence.

$$L(\boldsymbol{\theta}; \mathbf{X}) = p(\mathbf{X}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

2.7 K-mean Algorithm

The K-mean clustering algorithm groups a sample set of feature vectors into K clusters via an appropriate similarity (or dissimilarity) criterion (such as distance from the center of the cluster).

$$J = \sum_{j=1}^{K} \sum_{n \in S_j} |x_n - \mu_j|^2,$$

The k-means algorithm assigns feature vectors to clusters by the minimum distance assignment principle, which assigns a new feature vector to the cluster, such that the distance from the center. The basic k-means algorithm is as follows:

- ➢ Put the first K feature vectors as initial centers.
- ➢ Assign each sample vector to the cluster with minimum distance assignment principle.
- ➢ Compute new average as new center for each cluster.
- ➢ If any center has changed, then go to step 2, else terminate.

2.8 Fuzzy C Mean Algorithm

1. One of the pixel is placed as constant from group of clusters.
2. Identify the distance among the pixels and calculate given dimension of input image.
3. Start the iterations and if probable iteration is reached, then stops the process and get segmenting image otherwise iteration process is continued.

$$\mathbf{Z}_n^{(k)} = \sum_{\{q:\, q \in k\}} W_{qk} \mathbf{X}^{(q)}_n$$

Fuzzy clustering plays an important role in solving problems in the areas of pattern recognition and fuzzy model identification. A variety of fuzzy clustering methods have been proposed and most of them are based upon distance criteria. One widely used algorithm is the fuzzy c-means (FCM) algorithm. It uses reciprocal distance to compute fuzzy weights.

2.8.1Advantages of Fcm

- ➢ Fuzzy is the most natural way for clustering.
- ➢ It not only gives the accurate outline, but also the central and densely located cancerous area.
- ➢ Does not depend upon the user to identify the cluster values.

2.9Bayesian networking

Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. It calculates the probability that a disease is present in a finding, *P*(*disease finding*) , using the probability of finding a given disease, *P*(*finding disease*) , the probability of disease, *P*(*disease*) , and the probability of a finding, *P*(*finding*) is given as:

$$P(\text{diseases}\backslash\text{finding}) = \frac{P(\text{finding}\backslash\text{disease})\, P(\text{diseases})}{P(\text{finding})}$$

X is a Bayesian network with respect to G if its joint probability density function (with respect to a product measure) can be written as a product of the individual density functions, conditional on their parent variables

$$p(x) = \prod_{v \in V} p\left(x_v \mid x_{\mathrm{pa}(v)}\right)$$

where pa(v) is the set of parents of v (i.e. those vertices pointing directly to v via a single edge)

With the help of pixel intensity the Bayesian network determines the stages of cancer and whether it is a normal tumor curable or it is cancerous. Also determine the severity of disease as grades.

III. RESULT AND DISCUSSION

3.1 Mammogram image of breast- Filtering and enhancement of mammogram image

Initially patient's mammogram image were taken (Fig.3.A) and then fed as an input image for image acquisition. In image acquisition the images were scaled and filtered using a median filter to filter the image where the background noises were removed (Fig.3.B).In the next step image enhancement were done where the image is enhanced using a histogram equalization, contrast of the image is enhanced using the pixel intensity values and the output is shown in (Fig.3.D).

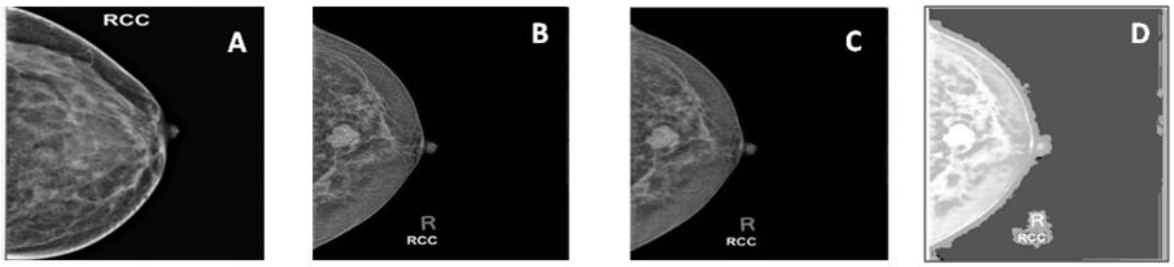

Fig 3.Mammographic image of normal and cancer affected breast A. Normal Breast B. Tumor breast
C. preprocessed image of B obtained by removing the noise using median filter and
D. enhanced image of C

3.2 Comparison of clustering algorithm

Once the image is enhanced using the image enhancement then it is segmented using the image segmentation here it is proposed to segment using clustering algorithm. Enhanced image shown in Fig 3 D is then further processed using different clustering algorithm such as K-mean clustering, Expectation Maxima and Fuzzy C Mean clustering algorithm.

TABLE 1
COMPARISON TABLE

Algorithms	Accuracy %
Expectation Maximization	80.24%
K-Mean Algorithm	85.98%
Wavelet Techniques	87%
Watershed Segmentation	88%
Fuzzy C Mean	92.47%

Fig 4 Shows the Comparison result of the Clustering Algorithm. The enhance image of Fig 4 A tumor affected breast will be further processed using image segmentation algorithm. By grouping the pixel intensity of the same size the tumor can be determined. The exact location of the tumor in breast is located using three algorithms. Fig 4B shows the segmented image using K-mean algorithm which shows a location and size of tumor, Fig 4C is the segmented image using expected maxima where the result is not up to the mark. Fig 4D segmented image using Fuzzy C Mean algorithm here the accuracy of the location is prompt and very clear compared to other two results. All over these algorithms give a result of tumor location in the breast using the clustering. These three algorithms are compared and the best result is then processed with Bayesian Network. Table 1 shows the accuracy of output obtained from each algorithm.

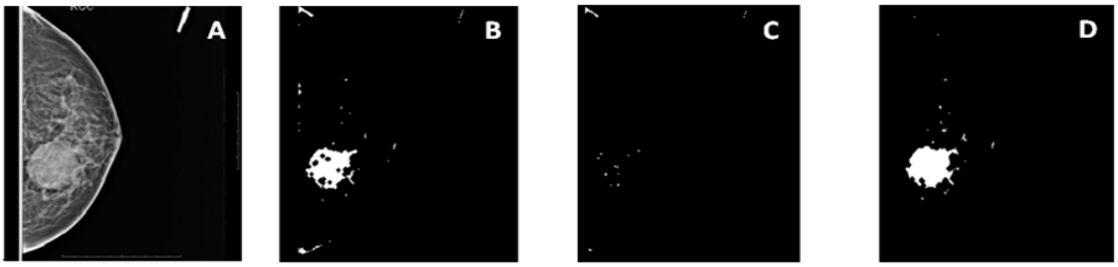

Fig 4.Comparison of Clustering Algorithm A. Tumor Affected Breast B. Segmented image of A using K-mean clustering Algorithm C. Clustered image of A using Expectation Maximization Algorithm D. Clustered image of A using Fuzzy C Mean Algorithm

The FCM which generates best location of tumor and the output is fed to ANN and trained using BAYESIAN NETWORK. This is done in order to find grade of the breast cancer whether it is grade 0, grade 1, grade 2 or grade 3 which is shown in Fig 5 and Fig 6. According to threshold, volume, density the result is given by the ANN training method.

3.3Grades / Stages

Different stages of cancer from the benignancy to malignancy are determined using Bayesian classifier. Here the input fed is the output of FCM clustering algorithm since the FCM gives the best result of the location of breast cancer, the grades are generated using the pixel intensity in the image.

Fig 5.Stage of Cancer A. Tumor Affected Breast B. Segmented image of A and Grade determined using Bayesian Network

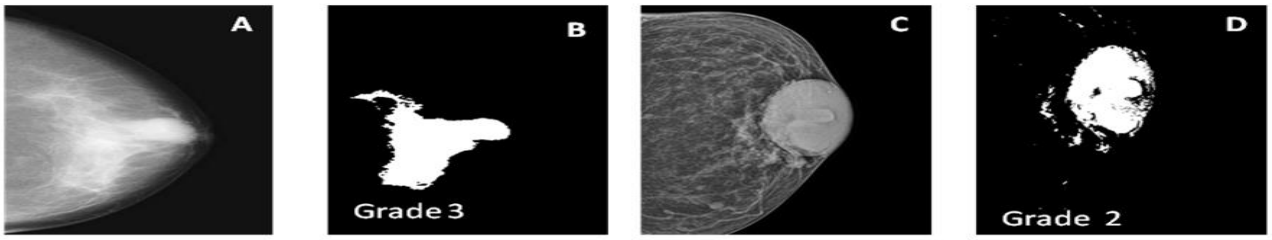

Fig 6.Grades of Cancer A. Tumor Affected Breast B. Segmented image of A and Grade 1 determined using Bayesian Network C. Tumor Affected Breast D. Segmented image of C and Grade3. determined using Bayesian Network

3.3.1 Grade0

Pixel intensity 1-25: Indicate whether the tumor present is cancerous or non cancerous.

3.3.2 Grade1

Pixel intensity 26-50: Fig 5 B Shows the beginning stage of the cancer.

3.3.3 Grade2

Pixel intensity 51-100: Fig 6D Shows the semi-normal stage of the cancer.

3.3.4 Grade3

Pixel intensity above 100: Fig 6 B Shows the abnormal stage of the cancer.

The proposed work is to find the malignancy and benignancy of the tumor from mammogram images which is painless when compared with biopsy test, here we have performed only with the images which are processed and trained to get the desired output. .The software used is MAT lab, which is used to detect the cancerous area and the different grades of the cancer.

IV. CONCLUSION

In this work, mammogram breast cancer is detected by clustering methods and the results of segmentation is compared. Different Grades are detected by ARTIFICIAL NEURAL NETWORK, based on BAYESIAN NETWORK.

The mammogram breast cancer image is detected using clustering segmentation techniques and the results of this segmentation are compared. The Bayesian network is a powerful tool to describe the uncertainty and complexity of many problems in the real world. Hence the grades of cancer is found using Bayesian network.

Bayesian networks have an unparallel advantage of being able to exploit the explicit structure of the domain model to derive a graphical representation for learning.

The encoding of independencies in the network topology admits the design of efficient procedures for performing computations over the network. The Bayesian network is a powerful tool to describe the uncertainty and complexity of many problems in the real world. Within the idea of graphical model, a complex system consists of simple parts, which are bound together with probability theory.

For the application of computer-aided detection in mammography, the researchers intend to design an interface between the project's Bayesian network learning algorithm and the radiologists, so that the radiologists can have interaction with the system by labeling only a small number of informative images presented by the active learning algorithm. In this way, the project's system only requires very little labeling effort from radiologists, while significant improvements of classifiers are achieved

REFERENCES

[1] Breast Contour Detection on Digital Mammogram H. Mirzaalian Isfahan University of Technology, Iran ir.ac.iut.Mirzaalian@ec M. R. Ahmadzadeh Isfahan University of Technology, Iran Ahmadzadeh@cc.iut.ac.ir F. Kolahdoozan Isfahan University of Medical Sciences, Iran fkolahdoozan@yahoo.com.
[2] Breast Cancer Detection in Mammogram Using FuzzyC-Means And Random Forest Classifier by Aleena Johny1, Jincy J Fernandez2 © 2018 IJSRSET | Volume 4 | Issue 8 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099 Themed Section: Engineering and Technology.
[3] Accurate Breast Contour Detection Algorithms in Digital Mammogram Indra Kanta Maitra Research Fellow Department of Computer Science & Engineering, University of Calcutta, India Sanjay Nag Research Scholar Department of Computer Science & Engineering, University of Calcutta, India Samir K Bandyopadhyay Professor Department of Computer Science & Engineering, University of Calcutta, India International Journal of Computer Applications (0975 – 8887) Volume 25– No.5, July 2011.
[4] Mahua Bhattacharya, Arpita Das, IDENTIFICATION OF MICROCALCIFICATIONS AND GRADING OF MASSES USING DIGITAL MAMMOGRAM, International Journal of medical engineering and informatics 2010-vol.2, no2, pp.122-138.
[5] Detection of Microcalcifications Clusters in Mammograms through TS-MRF Segmentation and SVM-based Classification C. D'Elia°, C. Marrocco°, M. Molinara°, G. Poggi*, G. Scarpa*, F.Tortorella.
[6] Alison J. Noble and Djmal Boukerroui (2006) 'Ultrasound image segmentation', IEEE Transaction on Medical Imaging, Vol. 25, No. 8, pp. 987-1010.
[7] Kandaswamy A. and Sheshadri H.S. (2006) 'A Detection of breast cancer by mammogram image segmentation', Journal of Cancer Resther, Vol. 1, No. 4, pp. 232-234.
[8] Nóra Székely and Norbert Tóth (2008), 'Hybrid System for Detecting Masses in Mammographic Images', IEEE Transactions on Instrumentation and Measurement, Vol. 55, No. 3, pp. 944-952.
[9] Rafael C. Gonzalez (2002), 'Digital image processing', Second Edition, Prentice Hall, New Jersey, USA, p. 674.
[10] Saheb Basha S. and Sathya Prasad K. (2009**), '**Automatic detection of breast cancer mass in mammogram using morphological operators and fuzzy c-means clustering', Journal of Theoretical and Applied Information Technology, pp. 704-706.
[11] Thangavel K. and Karnan M. (2005), 'Automatic Detection of Microcalcification in Mammograms', ICGST-GVIP Journal, Vol. 5, No. 5, pp. 31-61.
[12] Varsha Patil and Vaishali Pawar S. (2006), 'An Automated computer Aided Breast Cancer Detection Technique', GVIP Journal, Vol. 6, No. 1, pp. 69-70.
[13] Detection of masses in mammogram images using CNN, geostatistic functions and SVM Wener Borges Sampaioa, Edgar Moraes Diniza, Aristo ´fanes Corrˆea Silvaa,, Anselmo Cardoso de Paivaa, Marcelo Gattass Computers in Biology and Medicine 41 (2011) 653–664.
[14] Y.IreaneusAnna Rejani et al, EARLY DETECTION OF BREAST CANCER USING SVM CLASSIFIER TECHNIQUE, International Journal on Computer Science and Engineering 2009, Vol.1(3), pp 127-130.
[15] A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram Mohamed Meselhy Eltoukhya,, Ibrahima Fayeb, Brahim Belhaouari Samirb Computers in Biology and Medicine 40 (2010).
[16] Jyotirmay Gadewadikar1*, Ognjen Kuljaca1, Exploring Bayesian networks for medical decision support in breast cancer detection African Journal of Mathematics and Computer Science Research October 2010, Vol. 3(10), pp. 225-231.
[17] Diagnosis of Breast Cancer using Clustering Data Mining Approach ,Jahanvi Joshi Research Scholar KSV University Gujarat, India Rinal Doshi Director, Mahavir Computer Acadmy, Kalol,Gujarat, India Jigar Patel, Ph.D Research Guide Gujarat Technological University International Journal of Computer Applications (0975 – 8887) Volume 101– No.10,September2014.
[18] A comprehensive study on data clustering for breastcancer prognosis and risk exposure NIHARIKASAXENA1,Prof.MEETAKUMAR2 International Journal of Pure and Applied Mathematics Volume 118 No. 24 2018.

[19] Survey on Breast Cancer Detection Using Weka Tool S.Padmapriya1, M.Devika2,V.Meena3, S.B.Dheebikaa4 & R.Vinodhini5 Imperial Journal of Interdisciplinary Research (IJIR) Vol-2, Issue-4, 2016.
[20] Comparative study on different classification techniques for breast cancer dataset Ahamed Lebbe Sayeth Saabith1, Elankovan Sundararajan2, Azuraliza Abu Bakar3 International Journal of Computer Science and Mobile Computing, Vol.3 Issue.10, October-2014.
[21] Performance and Evaluation of Data Mining Techniques in Cancer Diagnosis R.M. Chandrasekar Ph.D, V. Palaniammal M.C.A., M.Phil. IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 15, Issue 5 (Nov. - Dec. 2013.
[22] Prediction of benign and malignant breast cancer using data mining techniques Vikas Chaurasia1, Saurabh Pal1 and BB Tiwari2 Journal of Algorithms & Computational Technology 2018, Vol. 12.