International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 03, 2020
ISSN: 1475-7192

EARLY DETECTION AND CLASSIFICATION OF BREAST TUMOR FROM MAMMOGRAM IMAGES

[1]Dr. Jayesh George Melekoodappattu, [2]Dr.V.Vijikala, [3]Dr.D.Anto Sahaya Dhas

ABSTRACT *– A low dose X-ray technique of the breast known as mammography is popular due to its advantages over other imaging techniques. Even though only 2 percentage chance of being malignant radiologist usually recommend for a biopsy test. The unwanted biopsy test not only increase the anxiety among patient but also enhance the health care cost. The existing CAD system may misinterpret the suspicious lesion as false positive or false negative. To avoid such misinterpretation it is necessary to improve the existing CAD system such that it will accurately* ***predict*** *the suspicious lesion. This paper presents a novel approach which compares several hybrid image processing techniques to enhance the accuracy. Hybrid technique is defined as the technique which combine two or more techniques together. The accuracy of the proposed system is obtained as 95 percentages.*

Key words- *CAD; Mammogram; Median filter; Preprocessing; FCM; PSO; GLCM; Genetic Algorithm*

I. INTRODUCTION

Breast Tumor is considered as a main cause of mortality in women. According to national cancer institute, Annual report to the nation 2017 incidence summary shows that between 1999 to 2013 the overall cancer incidence rate remain stable for women while the incidence rate continued to decrease among men. The slight increase in the rate of new cases of breast cancer brought up the overall incidence rate among women. This type of cancer occur almost entirely in women but in some rare cases men too gets a breast tumor. Breast cancer can begin from any part of the breast as a result there are about twenty different types of breast cancer. Most cancer occur in milk duct and some may begin in glands and also there are other cancer that are less common[1].

There are two sort of tumor and they are benign and malignant tumors. A non cancerous tumor is called benign tumor and is considered to be completely curable .A malignant tumor may invade the surrounding tissue and distribute all over the body. Real cause of breast malignancy is remaining as a big challenge but some of these are due to genetic abnormality and about 5-10% of cancers are inherited from parents[2]. An x ray imaging techniques to examine human breast is called mammography. This specialized imaging technique aids in the identification and diagnosis of breast abnormalities in female that is , it is used as a screening tool. A diagnostic mammography is recommended for the patient who has some previous abnormality and requiring some following up. A typical mammogram involves two or four views taken from different angles [3].

1Assistant Professor, Vimal Jyothi Engineering College, Kannur, Kerala, India, jayeshg1988@vjec.ac.in

2Associate Professor, Sahrdaya College of Engg and Tech., Kodakara, Kerala, India, vijikalav@sahrdaya.ac.in

3 Professor, Vimal Jyothi Engineering College, Kannur, Kerala, India, dr.anto@vjec.ac.in

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

Mammography has a missed rate at the range of 10% which is partly due to dense tissue obscuring the cancer and also large overlap of appearance of cancer with normal tissues. Interpretation of mammogram images are difficult because normal breast looks different for each women and also appearance of an image may be compromised due to several reasons. So to predict the lesions accurately, you need to have an intermediate between clinician and an input image. A computer aided detection system(CAD) highlights the abnormal areas (mass, density and microcalcification) on the images. it will incline to result in a high 'recall' rate with less effect on positive predictive value. By improving the existing CAD system it is possible to predict accurately the suspicious lesion so that rate of breast cancer recovery can be increased[4].

This paper contains following sections. Section II gives a brief idea of database used for the analysis. The proposed technique is detailed in section III. Section IV discusses the result and conclusion is given in section V.

II. IMAGE DATABASE

For the analysis and evaluation of proposed method data taken from MIAS database. It is the most easily accessed and commonly used databases. Which contain 322 mammogram images(161 pairs) that is the repository contain right and left mammograms of about 161 patients. These 322 pictures belong to 3 classes such as benign, normal and malignant. There are about 208 normal, 63 benign and 51 malignant mammograms. The MIAS database is reduced to 200 micron pixel so that every mammogram is 1024 ×1024 pixel. The repository possesses an introduction file and the information s such as type, sort, location and size.

III. PROPOSED METHOD

An improved CAD system method proposed by incorporating several hybrid image processing techniques. Preprocessing is done with median filtering and histogram equalization. A hybrid segmentation technique called PSO-FCM is performed to separate the digital images into several segments. With the help of Gray level spatial dependency matrix (GLCM), desired attributes which are informative are procured. The most desired feature is selected using Genetic Algorithm[5]. Further several classifiers are used for classification of suspicious lesion as malignant or benign. The proposed block diagram is depicted in figure 1.

A. Preprocessing

It is the basic image processing step to enhance the quality of an mammogram and preparing it for further analysis. It can be considered as the preparatory phase[6]. Unknown noise, poor contrast, inhomogeniety and weak edges are some common characteristics of medical images. All these problems can be rectified by this step. In this paper, the captured image undergoes a noise removal step and then an histogram equalization before segmenting an image[7]. The noise is removed by using median filter. One of the advantages of median filter over other filtering technique is that filtering can be repeated and it does not shift the boundaries that can happen in conventional filtering operations[8]. The median filtered image undergoes a histogram equalization process to ameliorate the contrast of mammogram and to adjust the image intensity.

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

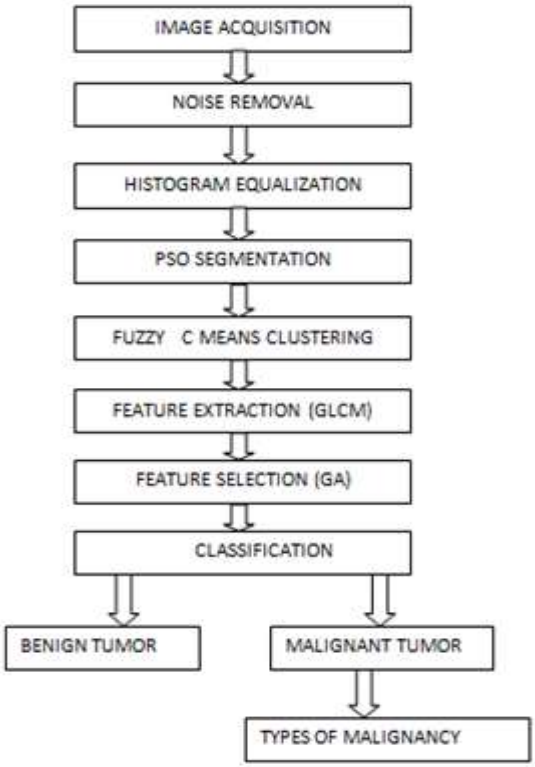

Fig 1: Block diagram of breast cancer detection system.

B. Segmentation

Segmentation can be defined as the process of altering the representation of mammogram in to something that is more easy to analyze or partitioning a digital images into different segments and each segments divulges some information. This paper uses a hybrid segmentation technique[9]. That is it combine two or more segmentation technique together. The proposed segmentation technique is PSO-FCM.

Particle swarm optimization (PSO) is a analytical method that optimizes a problem and trying to meliorate a candidate solution with regard to a given quality [10]. This optimization is a heuristic global method which is based on swarm intelligence[11].This method imitate the bird and fish floak behaviour. The main advantages of this particle swarm optimization is its simplicity and ease of implementation .It does not involve any evolutionary operators.

FCM clustering technique is a method in which the given repository data is grouped into n groups. Each point can be a part of two or more groups[12]. This technique functions by assigning certain membership function to each data point corresponding to each cluster centre and this technique gives best result for overlapped dataset

C. *Feature Extraction*

A well-accepted statistical algorithm, GLCM is a robust solution for obtaining image texture details. Several statistical parameters are grouped under the GLCM feature. The GLCM mechanisms represent the texture of an object by measuring how often pixel pairs of particular values appear in a picture and in a given spatial relation, generating a GLCM, and deriving statistical observations from this GLC matrix [13].

GLCM is pixel based technique to retrieve small details in an image rather than region based technique like Gabor filter. It is a two dimensional matrix with the size equal to the gray levels of a mammogram. From the

gray level spatial dependency matrix, 20 different features can be extracted. Each element in the matrix $P(u, v \mid d, \theta)$ is the occurrence of two intensity levels *u* and *v* at particular displacement interval d and at an angle θ. The element also computes for the frequency of brightness values at several displacement angle and interval. GLCM computes a large number of temporary data for every combination of d and θ. GLCM is very sensitive to the scale of the texture specimens because of their large dimensionality[14]. So the number of intensity levels is sometimes reduced. An example is shown in figure 2 to detail how the elements in the matrix are computed.

GLCM computed here with one pixel offset. That is a reference and immediate neighbour pixel. Use of large offset is also recommended if the window size is large enough[15]. The first cell in the given GLCM will be entered with the frequency of occurrence of the combination 0 and 0. i.e. how many times a reference pixel with intensity value 0 comes to the left of another pixel with intensity value 0 (neighbour).

D. Feature selection

The extracted features using the above discussed algorithms are large in size and not all these attributes are equally relevant for mammogram classification[16]. So the best subset of attributes must be extracted from the total features. This step is called feature selection. This process is carried out by the glowworm swarm optimisation algorithm.

The GA is a search and optimisation tool that produces solutions using natural evolution-inspired techniques [17]. This involve large classes of evolutionary techniques which involve several evolutionary operators such as mutation, cross over, selection etc. It is one of the adaptive heuristic search algorithms. A typical GA technique requires a genetic representation and a fitness function.

E. Classification

The intent of the classification process is to categorize all pixels of images into one of several classes. The proposed method uses an Naïve Bayes, SVM, KNN and ELM classifier for categorization of suspicious lesion into malignant or benign tumor.

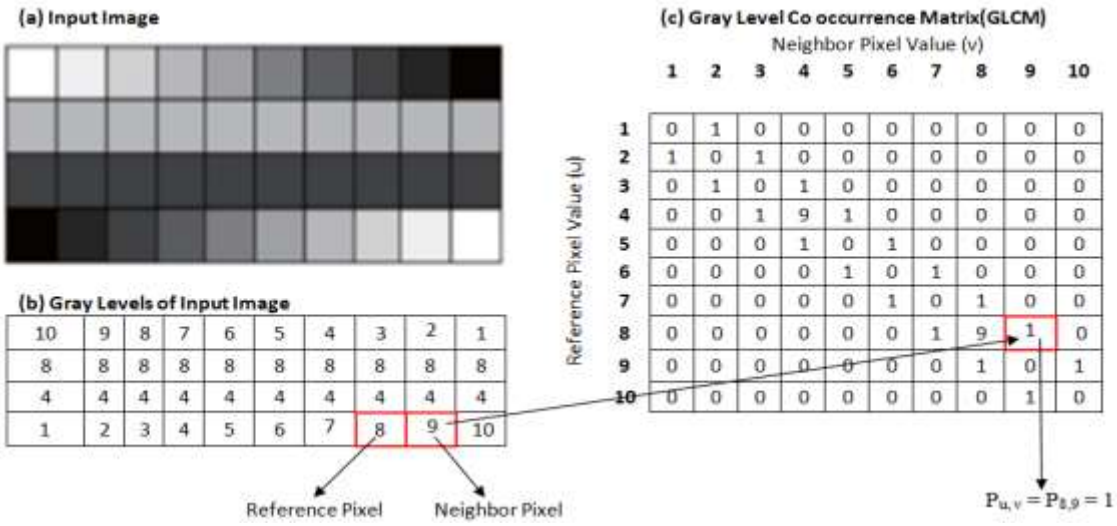

Fig 2. Computation of co-occurrence matrices of a sample image.

Support Vector Machine (SVM)

SVM is a machine learning procedure designed for regression and classification. SVM is used for linear and non-linear classification problems[18]. It separates data points into sub-sets built on the theory of the decision plane so that each entity in the sub-set has the same properties. SVM creates a hyperplane sundering the data in the training set in such a way that each class label is on either half of the hyperplane [19]. The margin is identified as the distance from the hyperplane to the nearest data point. Equation 1 could be used to construct the hyperplane.

$$g(\vec{x}) = w^{-T}\vec{x} + \omega_0 \qquad (1)$$

where ω_0 is the weight vector and if the value of this formula is higher than one for all data vectors, then it belongs to class 1. If the output of this equation for all data vectors is less than -1, then it represents class 2.

Extreme Learning Machine (ELM)

ELM is a neural based classifier that uses a single hidden layer feed-forward network. The efficacy of the SLFN can be enhanced by choosing an appropriate value for variables such as threshold value, weight and activation functions. All of these elements are updated recursively for the acceptable value in gradient-based learning methods. ELM works on the
principle of Moore-Penrose method to find out the output weight.

Naïve Bayes (NB)

NB is a probabilistic classifier which associates probability value for each class of output based on the input attribute values. It does this by calculating the conditional probabilities for each attribute values in the training dataset.

K Nearest Neighbour Classifier (KNN)

KNN is an efficient classifier employed in pattern recognition. It holds all possible instances and classifies new cases based on the distance measure. Every instance expressed with numerical values in N-dimensional feature space. The training attributes comprises of a collection of vectors with category label. Classification is accomplished by evaluating the distance to K nearest points.

IV. RESULT AND DISCUSSION

The ultimate aim of the proposed method is to classify particular suspicious mammographic lesions into benign or malignant tumor using several classifiers. The preprocessing in the designed method is carried out in two independent steps. Initially noise is removed by median filtering and then a histogram equalization is carried out to enhance the contrast of image. A hybrid segmentation technique is applied for partitioning an image into different segments. The standard result is summarized in figure 3 and 4. Attributes are procured using GLC matrix and genetic algorithm to select desired attributes. GLCM is pixel based technique to retrieve small details in an image rather than region based technique like Gabor filter.

The twenty features extracted from the GLCM are Mean, Standard deviation, Entropy, Sum Average, Sum Entropy, Difference Entropy, Sum Variance, Variance, Difference Variance, Contrast, Homogeneity, Correlation, Energy, Inverse Difference Moment, Autocorrelation, Coarseness, Cluster Shade, Inertia, Cluster Prominence and Cluster Tendency. The choosen GLC Matrix attributes are Standard Deviation, Variance, Entropy, Contrast, Correlation, Energy, Difference Entropy, Homogeneity and Difference Variance.

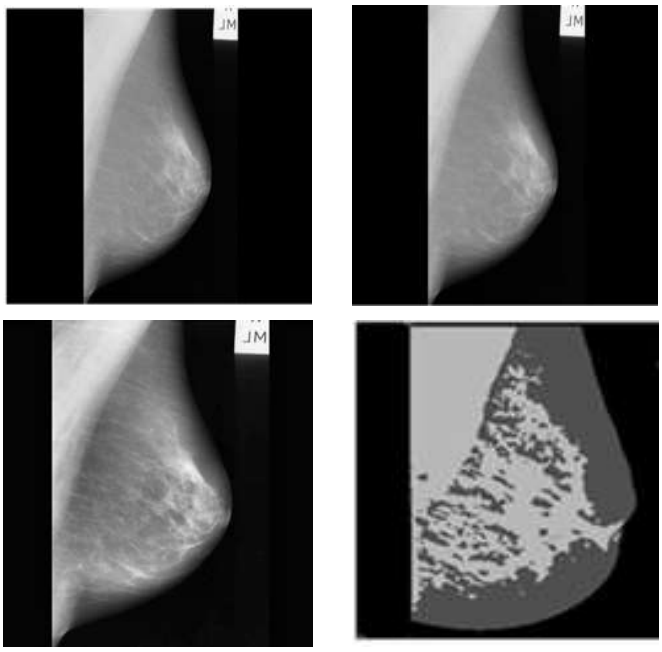

Fig 3: Input image, median filtered image, histogram equalised image and PSO segmented image.

The accuracy of a system is the degree of closeness of a quantity's measurements to the actual (true) value. The accuracy of the designed technique is computed using the formula. Accuracy is computed by,

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} * 100\% \qquad (2)$$

Sensitivity: Sensitivity or recall is detailed as the probability that a test will be positive. Sensitivity is also named as a true positive fraction. It is the ratio of true positive to the sum of true positive and false negative.

$$Sensitivity/Recall = \frac{TP}{TP+FN} * 100\% \qquad (3)$$

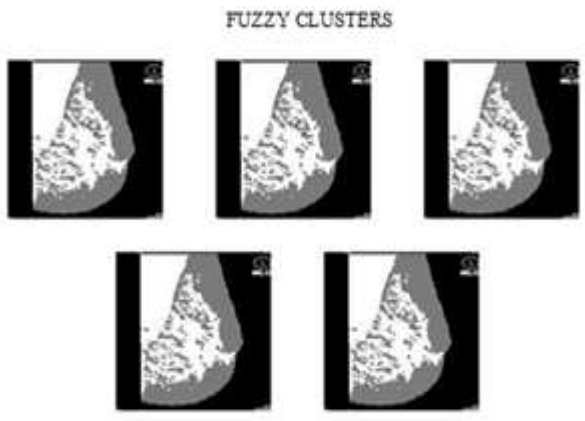

Fig 4: Different fuzzy clusters

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

Table 1. True positive rate (TPR) of various classifiers.

Classifier	**GLCM**
Naïve Bayes	0.7
KNN Classifier	0.73
SVM Classifier	0.82
ELM Classifier	0.94

Table 2. False positive rate (FPR) of various classifiers.

Classifier	**GLCM**
Naïve Bayes	0.07
KNN Classifier	0.175
SVM Classifier	0.04
ELM Classifier	0.7

Table 3. Accuracy of various classifiers.

Classifier	**GLCM**
Naïve Bayes	0.6
KNN Classifier	0.9
SVM Classifier	0.92
ELM Classifier	0.95

V. CONCLUSIONS

Medical image processing is a wide area of research. A number of researches is now carrying out on mammogram images. But still the analysis of mammogram is remaining as a big challenge. This is because the mammogram images are difficult to interpret because of its complex nature. Because of this there is a high chance of occurring false positive or false negative. To avoid such situation it is necessary to introduce an improved CAD system which can accurately predict the suspicious mammographic lesions. The proposed system uses several hybrid image processing technique to enhance the accuracy of a CAD system. The designed system found to have an accuracy of 95%.

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

REFERENCES

1. E.Malar,A.Kandaswamy,D.Chakravarthy,A.Giri Dharan "A novel approach for detection and classification of mammographic micro-calcification using wavelet analysis and extreme learning machine" Computers in Biology and Medicine 42 (2012) 898–905
2. Jayesh George Melekoodappattu and Perumal Sankar Subbian, "A Hybridized ELM for Automatic Micro Calcification Detection in Mammogram Images Based on Multi-Scale Features", Journal of Medical Systems, Volume 43, Number 7, Springer, 2019.
3. Qinghai Bai 'Analysis of Particle Swarm Optimization Algorithm' computer and information science Vol.3,No.1 feb 2010.
4. Perumal Sankar and Jayesh George Melekoodappattu, "ELM Based Detection of Microcalcification in Mammogram using GLCM Features", International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-1, pp. 1146-1151, 2019.
5. N. Gopi Raju & P A Nageswara Rao." Particle Swarm Optimization Methods for Image Segmentation Applied In Mammography". Int. Journal of Engineering Research and Application, ISSN : 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013, pp.1572-1579
6. Jayesh George Melekoodappattu and Perumal Sankar Subbian, "Efficient Preprocessing Filters and Mass Segmentation Techniques for Mammogram Images", in IEEE - International Conference on Circuits and Systems (ICCS2017), Trivandrum, Kerala, pp. 408-413, December 2017.
7. Snehal A. Mane1, Dr. K. V. Kulhalli. "Mammogram Image Features Extraction and Classification for Breast Cancer Detection". International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 07 Oct-2015
8. Nijad Al-Najdawi, Mariam Biltawi, Sara Tedmori. "Mammogram image visual enhancement, mass segmentation and classification". Applied Soft Computing 35 ,2015, 175–185.
9. Kshema and Jayesh George Melekoodappattu, "Efficient mammographic mass segmentation techniques: A review", in IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET2017), Chennai, Tamilnadu, March 2017, pp. 2253- 2256.
10. C.Abiram, R.Harikumarand, S.R.Sannasi Chakravarthy "Performance Analysis and Detection of Micro calcification in Digital Mammograms using Wavelet Features" IEEE WiSPNET 2016 conference.
11. B.Monica Jenefer and V.Cyrilraj. "An efficient image processing methods for mammogram breast cancer detection". Journal of Theoretical and Applied Information Technology 10th November 2014. Vol. 69 No.1
12. R.Harikumar, B.Vinothkumar "Performance Analysis of Neural Networks for classification of medical image with wavelet feature extraction "International Journal of Imaging Systems and Technology, Vol. 25, 33–40 (2015)
13. Kshema and Jayesh George Melekoodappattu, "Preprocessing filters for mammogram images: A review", in IEEE International Conference on Emerging Devices and Smart System (ICEDSS 2017), Tamilnadu, March 2017

14. N. Gopi Raju & P A Nageswara Rao." Particle Swarm Optimization Methods for Image Segmentation Applied In Mammography". Int. Journal of Engineering Research and Application, ISSN : 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013, pp.1572-1579
15. Zielinski, J. "Statistical Sequential Analysis for Detection of Microcalcifications in Digital Mammograms", IEEE International conference on Signal processing and Communication, July 2010, pp 1- 5.
16. Jiebo Huang, "A New Fast Fractal Coding Method for the Detection of Microcalcifications in Mammograms", IEEE International conference on Multimedia technology, July 2011, pp 4768-4771.
17. Dong Wang , "Applying Feature Selection for Effective Classification of Microcalcification Clusters in Mammograms", IEEE 10th International conference on Computer and Information technology, July 2010, pp 1384-1387.
18. Sneha John and Jayesh George Melekoodappattu, "Extreme learning machine based classification for detecting micro-calcification in mammogram using multi scale features" in IEEE International Conference on Computer Communication and Informatics (ICCCI 2019), Tamilnadu, January 2019
19. L.P .Cordella, G.Percannella, C.Sansone and M.Vento, "A graphtheoretical clustering method for detecting clusters of microcalcifications in mammographic images", IEEE on Computer-Based Medical Systems , pp . 15 – 20 ,2005 .

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020