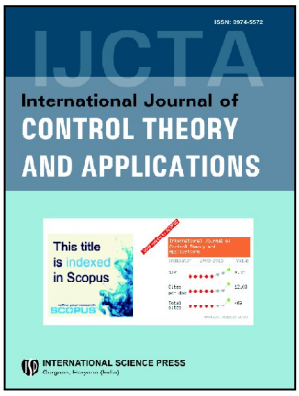


International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 9 • Number 50 • 2016

Classification of Breast Cancer in Digital Mammography using Hybrid Model

K. Preetha[1]

[1] *Assistant Professor, Vellalar College for Women, Erode, India, E-mail: kpreetha.msc@gmail.com*

Abstract: Mammography is a specific type of breast imaging that uses low-dose x-rays to detect cancer early before women experience symptoms. The goal of this research is to classify the breast cancer into Normal or Abnormal patterns in Digital Mammography which helps radiologists in reducing false positive predictions. A robust approach to increase the accuracy of the classification is based on the hybrid model which combines Artificial Neural Network with Genetic Algorithm. Mammogram image is pre-processed using Adaptive Median Filtering. The extracted features of the pre-processed image play a vital role in classification process for identifying normal and abnormal which will help the doctor to diagnose and make decisions about the treatment.

Keywords: Breast Cancer, Mammogram, Neural Network, Genetic Algorithm

I. INTRODUCTION

Breast cancer is the most common type of cancer in women, aged between 35 and above, that can even cause death. Mammograms can often detect breast cancer before it can be felt. A mammogram can show small deposits of calcium in the breast. Although most calcium deposits are benign, a cluster of very tiny specks of calcium (called *micro calcifications*) may be an early sign of cancer. Radiologists basically look for two types of patterns in mammography. They are micro-calcification and masses. The diagnosis result of tissue is classified into three categories: Normal represents the mammogram without any cancerous cell, benign represents the mammogram showing tumor, but not formed by cancerous cells and malign represents the mammogram showing tumor with cancerous cells. About 20% of women having breast cancer and undergoing mammography has negative mammograms, i.e. are misdiagnosed. Furthermore only 20 - 40% of the women who undergo biopsy have cancer.

Now-a-days, the mammograms are analysed in a much better way than the traditional methods. The quick growth of computer technology, pattern recognition, digital image processing, and artificial intelligence has provided a new way in mammogram analysis and breast cancer diagnosis in modern times. By considering the knowledge of radiologists, the computer-based systems give a second opinion in finding abnormalities and make diagnostic decisions.

The objective of this paper is to improve the detection of cancer in mammogram images, a method to classify the mammogram image with the aid of Artificial Neural network based genetic algorithm. Various image processing steps like Pre-processing, feature extraction is carried out along with neural network based Genetic Algorithm.

This paper is organized as follows: Section 2 briefly reviews some existing techniques for classification. Section 3 describes the proposed methodology. Section 4 discusses the experimental result analysis. Performance analysis of our method is explained in section 5 and finally conclusion and future work is presented in section 6.

II. LITERATURE REVIEW

Currently, breast cancer detection is a challenging issue for women. Breast cancer is curable, if it is detected in an initial stage. Number of researchers has tried to arrive an exact solution for this work by proposing different classification techniques.

B. Monica Jenefer et al [10] proposed a method for mammogram breast cancer detection using SVM classifier and GLCM feature extraction and achieved 97% sensitivity, 98% accuracy and 100% Specificity for classifying 80 mammogram images randomly selected from the MIAS database. Sonal Naranje [11] proposes an automated technique using artificial neural network for early detection of breast cancer gives maximum accuracy at high speed.

R. Nithya and B. Santhi [3] proposed a method for Classification of normal and abnormal patterns in Digital mammograms for the breast cancer diagnosis using ANN and GLCM features. The work shows that the sensitivity and specificity more than 90% for a sample set of 50 digital mammogram images from the DDSM Database.

The overall literature survey says that various methods and classification techniques are applied for classifying the images into normal or abnormal . The existing methods are taken, limited number of mammogram images from the MIAS and DDSM database and that are classified using less number of extracted features. Based on this study, we propose better and efficient classification methods are applied for improving the accuracy.

III. PROPOSED METHODOLOGY

The proposed system consists of three phases for breast cancer detection from the mammogram image. They are Preprocessing, Feature Extraction and Classification. The anticipated research uses the data set obtained from Mammographic Image Analysis Society (MIAS) [7]. The set consists of 322 images that fall into one of the following classes: 67 benign, 54 malignant and 201 Normal images. The overview of proposed methodology is depicted in Fig. 1.

(A) Pre-Processing

Pre-processing is to improve the image quality to make it suitable for further processing by removing or reducing the noise in an image. The image quality is enhanced by using adaptive median filter. Adaptive Median filter solves the dual purpose of removing the impulse noise from the image and reducing distortion in the image. It also smoothens out other types of noise, thus giving mush better output image than the Median filter [9]. The output images are processed for feature extraction.

(B) Feature Extraction

Feature Extraction is very important for the detection and classification of mammogram images. It involves simplifying the amount of resources required to describe a large set of data accurately. Intensity based features are the first order statistics depends only on individual pixels values. The intensities and its variations inside the

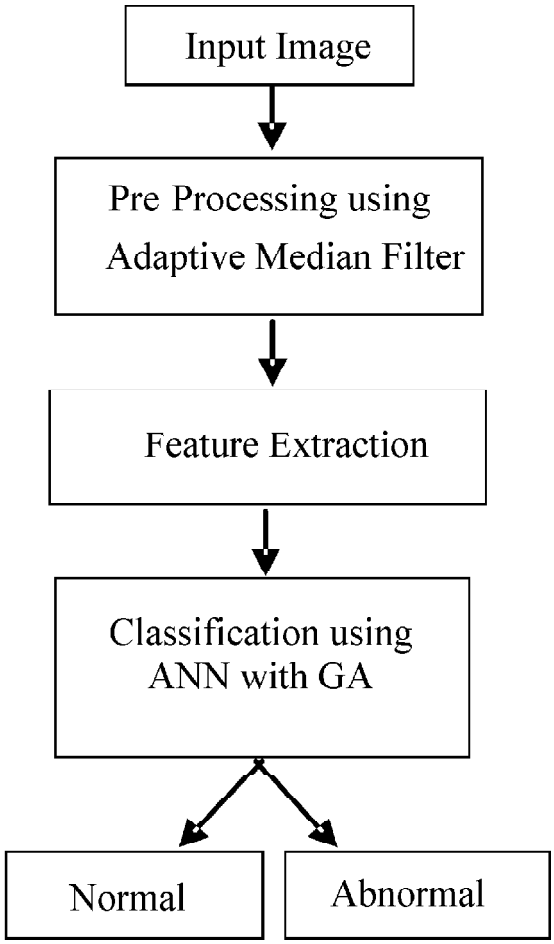


Figure 1: Diagrammatic Representation of Proposed Method

mammograms can be measured by features like mean, standard deviation and skewness. Gray-level co-occurrence matrix (GLCM) is the second order statistical method of examining the textures of an image that considers the spatial relationship of the pixels. The GLCM functions characterize the textures of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM and then extracting statistical measures from this matrix. A GLCM features like contrast, correlation, Homogeneity, Area. Table 1. Shows the explanations and formula for feature extraction.

(C) Classification

The extracted feature is used as an input to the classifier. Classification task usually involves with training and testing data which consists of some data instances. The neural networks are trained such that the input has to send a particular output. The neural network has superior compatibility with the classification procedure. In this method, the Feed Forward Neural Network is used for training. The problem with neural network is that the number of parameters have to set before any training can begin. however, there is no clear rules how to set these parameters. By combining genetic algorithm and Neural networks (GANN), the genetic algorithm is used to find these parameters.

1) Artificial Neural Network: An Artificial Neural Network (ANN) is inspired by the way biological nervous systems process information. An ANN consists of a collection of processing elements that are highly interconnected and transform a set of inputs to a set of desired outputs in which each connection has a weight

Table I
Explanation and formula for feature extraction

Features	*Explanation*	*Formula*
Contrast	Measures of intensity contrast between a pixel and its neighbour.	$\Sigma\Sigma\,(i-j)^2\,p(i,j)$
Correlation	It indicates the size of the linear relationship of the neighborhood pixel gray level.	$\frac{\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x\sigma_y}$
Homogeneity	Measures the distribution of elements in the GLCM with respect to the diagonal.	$H=\sum_{i,j}^{N-1}\frac{P(i,j)}{1+(i-j)^2}$
Area	Area of a particular region of interest from images.	Total number of ON pixels in an image.
Mean	Mean reveals the general brightness of an image.	$\mu=\sum_{x=1}^{N_S}x,p(x)$
Standard Deviation	Standard deviation reveals the contrast of an image.	$\sigma=\sqrt{\sigma^2}$
Skewness	Skew measures asymmetry (unbalance) the distribution of the gray level.	$S(z)=\sum_{i=0}^{L-1}(z_i-m)^3\,p(z_i)$

associated with it. The advantage of ANN is their capability of self-learning, and often suitable to solve the problems that are too complex to use the conventional techniques, or hard to find algorithmic solutions. The neural network trained by adjusting the weights so as to be able to predict the correct class. In this work, a method that combines ANN with GA algorithm was proposed to optimize the weights of target value.

Neural Network Algorithm

Step 1: Extract features from mammogram images.

Step 2: Create input and target for normal and abnormal class.

Step 3: The initial weights are chosen randomly.

Step 4: Calculate the predicted output.

An issue in neural network is difficult to train: the training outcome can be nondeterministic and depend crucially on the choice of initial parameters. To address the issue hybrid approach of neural network with Genetic Algorithm is used for cancer prediction.

2) Neural Network with Genetic Algorithm: Genetic algorithm is an optimization technique inspired by natural selection and natural genetics. Unlike many search algorithms, GA is a stochastic general search method, capable of effectively exploring large search space. Genetic algorithm is mainly composed of three operators: selection operator, crossover and Mutation. In GA, randomly initialize population of individuals. The individuals

in the genetic space are called chromosomes. It is the collection of genes where genes can generally represent as the database images. Gene is the basic building block of the chromosomes. In each generation, the population is evaluated using fitness function. Next comes the selection process, where high fitness chromosomes are used to eliminate low fitness chromosomes. But selection alone does not produce any new individuals into the population. Crossover is the process by which two-selected chromosome with high fitness value exchange part of the genes to generate new pair of chromosomes. The crossover tends to facilitate the evolutionary process to progress towards potential regions of the solution space. Mutation is the random change of the value of a gene, which is used to prevent premature convergence to local optimum solution.

The Genetic algorithm has been used to initialize and optimize the weights of neural network to improve the detection performance. The genetic algorithm based optimization of weight factor is explained in the below section.

Algorithm

Step 1: Extract the features from the mammogram images.

Step 2: Create Input and target for normal and abnormal class.

Step 3: Initializing and weight optimization of ANN is done by Genetic Algorithm is as follows.

Step 4: 1. Generate random population of n chromosomes.

2. Evaluate the fitness f(x) of each chromosome x in the population.

3. Create a new population by repeating following steps

Selection: Select two parent chromosomes from a population according to their fitness.

Crossover: Cross Over the parents to form new offspring.

Mutation: With a mutation probability mutates new offspring.

4. Use new generated population for a further run of the algorithm.

5. If the end condition is satisfied, stop and return the best solution in current population.

6. Go to fitness function evaluation step again.

Step 5: Output node predicts the correct class (i.e.) Normal or abnormal

Using the above algorithm the weightage can be optimized on target values for the efficient breast cancer detection using ANN with GA. The result shows that one can easily and effectively detect breast cancer at an early stage.

IV. EXPERIMENTAL RESULT ANALYSIS AND DISCUSSION

The proposed system for grading the severity level in abnormal mammogram images using hybrid neural network with Genetic Algorithm was implemented in the working platform of MATLAB. Table II shows the Feature extraction values for Normal and abnormal mammogram images.

Table II
Feature values for normal and abnormal images

Features	*Normal Mammogram Image*		*Abnormal Mammogram Image*	
	Image 1	*Image 2*	*Image 3*	*Image 4*
Contrast	6.1604	6.3690	5.5066	5.1064
Correlation	-0.2241	-0.0697	0.0665	0.1158
Homogeneity	0.0525	0.0440	0.0474	0.0488
Standard Deviation	31.1734	32.4499	32.2522	26.3692
Skewness	0.9381	1.5417	1.3147	0.6561
Mean	4.5938	4.5977	15.1484	20.3320

The extracted feature values are passed to train the neural network based genetic algorithm to classify the normal or abnormal mammogram images. Figure 2 shows the Classification of Normal and Abnormal Mammogram Images.

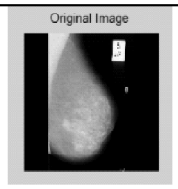
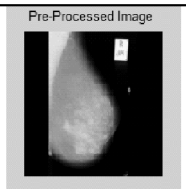
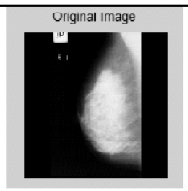
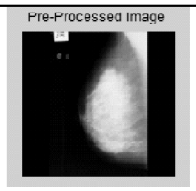
Original Image	Pre-Processed Image	Classification
		Abnormal
		Normal

Figure 2: Classification of Normal or Abnormal Mammogram Images

V. PERFORMANCE EVALUATION

The various assessment metrics are used to calculate and analyze that our proposed technique is the efficient mammogram image segmentation and classification of breast cancer. The metric values like Sensitivity (SE), Specificity (SP) and Accuracy (AC) are used to evaluate the performance of the classifier. The formulas are given in Table III. Sensitivity is a proportion of positive cases that are well detected by the test and the specificity is proportion of negative cases that are well detected by the test. Classification accuracy depends on the number of samples correctly classified.

Table III
Formula for measures

Measures	*Formula*
Sensitivity	SE=TP/(TP+FN)
Specificity	SP=TN/(TN+FP)
Accuracy	AC=(TP+TN)/(TP+FP+TN+FN)

Where, TP is the number of true positives; FP is the number of false positives; TN is the number of true negatives; FN is the number of false negatives. Confusion matrix is shown in Table IV.

Table IV
Confusion matrix

Actual	*Predicted*	
	Positive	*Negative*
Positive	TP	FP
Negative	FN	TN

TP - Predicts abnormal as abnormal.
FP - Predicts abnormal as normal.
TN - Predicts normal as normal.
FN - Predicts normal as abnormal.

Table V
Confusion matrix for testing ann-pso

Method	*Actual*	*Predicted*	
		Cancer (Positive)	*Normal (Negative)*
ANN-PSO	Cancer (Positive)	41(TP)	7(FP)
	Normal (Negative)	6(FN)	246(TN)

Table VI
Confusion matrix for testing ann-ga

Method	*Actual*	*Predicted*	
		Cancer (Positive)	*Normal (Negative)*
ANN-GA	Cancer (Positive)	44(TP)	4(FP)
	Normal (Negative)	2(FN)	250(TN)

Table VII
Performance comparison between ann-ga and ann-pso

Test Image	*Methods*	*Sensitivity*	*Specificity*	*Accuracy*
48 Malign	ANN-PSO	87.23%	97.23%	95.66%
252 Normal	ANN-GA	95.65%	98.03%	98%

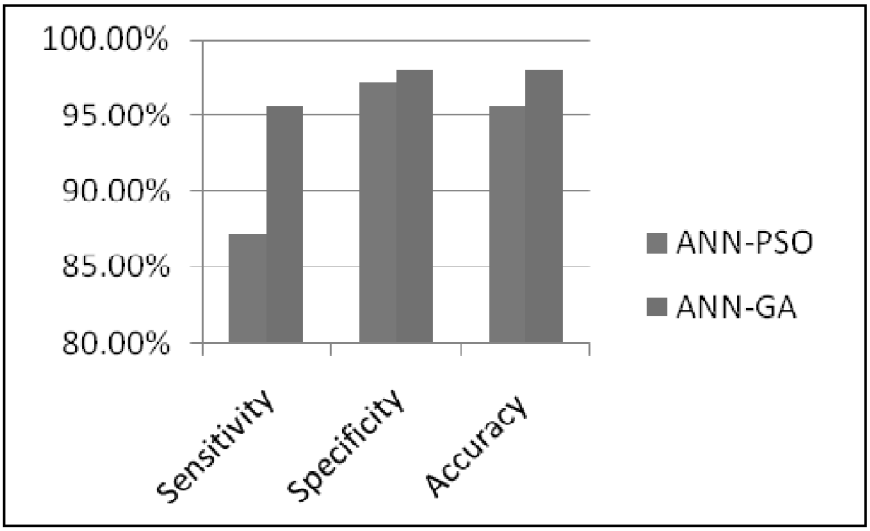


Figure 3: Comparison between ANN-GA and ANN-PSO

To evaluate this work, the proposed method trained with 322 mammogram images and tested with 300 mammograms (48 malignant and 252 normal) images. Confusion Matrix for testing ANN-PSO and ANN with GA is shown in Table V and Table VI. The result in Table VII shows the performance comparison between ANN-GA and ANN-PSO. Figure 3 show the computed sensitivity, Specificity and accuracy for testing data of proposed and existing method. ANN-GA reveals a better classification rate in sensitivity, specificity and accuracy.

VI. CONCLUSION AND FUTURE SCOPE

Breast cancer is one of the major causes of death among women. Due to the wide range of features associated to breast abnormalities, some abnormalities may be missed or interpreted. There is a number of false positive findings and therefore a lot of unnecessary biopsies. Breast cancer is curable when detected in early stages. The classification of mammogram images is emphasized in this paper for cancer diagnosis using hybrid model. It provides a faster diagnosis of cancer into normal or abnormal with accurate results. It will help the doctor and radiologists to analyse the patient is in and according to which he/she can take necessary and appropriate treatment steps and also reduce the number of false positives.

REFERENCES

[1] Satishsaini, RituVijay,"Performance Analysis of Artificial Neural Network Based Breast Cancer Detection System", *International Journal of Soft Computing and Engineering (IJSCE) ISSN:2231-2307,* Volume-4, September 2014.

[2] Shekharsingh, Dr.P.R Gupta, "Breast Cancer detection and Classification using Neural Network", *International Journal of Advanced Engineering Sciences and Technologies* Vol No:6, Issue No:1,004-009,2011

[3] R.Nithya, B.Santhi, "Comparative Study on Feature Extraction Method for Breast Cancer Classification", *Journal of Theoretical and Applied Information Technology", Vol.33 No.2,* November 2011.

[4] Al NutazM.Abdalla, Safaai Dress, NazarZaki, "Detection of Masses in Digital Mammogram using Second order Statistics and Artificial Neural Network," *International Journal of Computer Science and Information Technology (IJCSIT),* Vol 3, No 3, June 2011.

[5] Anu Alias, B.Paulchamy, "Detection of Breast cancer using Artificial Neural Networks", *International Journal of Innovative Research in Science, Engineering and Technology,* Vol.3 Issue 3, March 2014

[6] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., et al., "The mammographic images analysis society digital mammogram database", *Exerpta Medical International Congress Series,* vol. 1069, 1994, pp. 375–378. URL: http://peipa.essex.ac.uk/ipa/pix/mias/

[7] Monika Sharma, R. B. Dubey, Sujata, S. K. Gupta, "Feature Extraction of Mammograms", *International Journal of Advanced Computer Research,(IJACSA), Vol. 2 No.3 September 2012*.

[8] Anuradha C. Phadke and Preeti P. Rege, "Classification of Architectural Distortion from other Abnormalities in Mammograms", *International Journal of Application and Innovation in Engineering and Management (IJAIEM), Vol.2, No.2, 2013.*

[9] K.Preetha,, Dr.S.K.Jayanthi "Breast Cancer Detection and Classification using Artificial Neural Network with Particle Swarm Optimization" *International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST),* Vol.2, Oct 2016.

[10] B.Monica Jenefer, V. Cyrilraj, "An Efficient Image processing Methods for Mammogram Breast Cancer Detection", *Journal of theoretical and Applied Information Technology*, Vol. 69 No.1 November 2014.

[11] Sonal Naranje, " Early Detection of Breast cancer using ANN", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol.4 Issue 7, July 2016.