IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.1, January 2021 E-ISSN: 2347-2693

Multi-Model Analysis of Mammograms

Vijaylaxmi K. Kochari

Bharatesh College of Computer Applications, Belagavi, Karnataka, India

Corresponding Author: vijaylaxmirrao@gmail.com Tel.: +919008311396

DOI: https://doi.org/10.26438/ijcse/v9i1.3035 | Available online at: www.ijcseonline.org

Received: 16/Jan/2021, Accepted: 20/Jan/2021, Published: 31/Jan/2021

***Abstract*—**In this paper mammogram classification is introduced. The system takes mammogram, pre-processes it by applying Adaptive Histogram Equalization. The enhanced image is segmented using K Means Clustering algorithm. Statistical features such as standard deviation and mean of a segmented mammogram are extracted. SVM takes these features as input. DCT is applied on the segmented mammogram, these extracted features are fed as input to FFBPNN. These classify the mammogram as normal or abnormal. The training time of both the classifiers are compared to know which classifier takes less training time. The accuracy of the classifiers are determined by analyzing the results.

Keywords— Mammograms, pre-process, SVM, FFBPNN

I. INTRODUCTION

In the human body all the organs and tissues are made up of cells. These cells proceed to partition and increment. In order to keep the organs in the correct working condition and to maintain the concerns, balance is obtained between increase and death of the cells. The creation and death of cells will be imbalanced if there is going to be any DNA changes or inherited hereditary imperfections. New changes rise and accumulate with a snowball impact as cells create and duplicate themselves; these wild cells eventually turn into a tumor. Disease cells grab supplements from ordinary cells, and after that intrude on the nearby tissues. From a tumor, malignancy cells can even disconnect and metastasize to different parts of the human body through the lymphatic or the vascular frameworks.

As technology progresses, lifestyles have turn out to be fast, it causes many changes in the living style of the people. To mention few of those changes are unhealthy meals adoptions, absence of physical workout, and changed every day practices. These causes increase in the rate of infection in the human body. One of those diseases may be a breast cancer. For females all over the world breast cancer disease is the most widely recognized reason for death, it is also recognized as fifth common cancer overall, with about 522,000 deaths from breast cancer in 2012 [1]. Now a days breast cancer disease suffers rate has been increased a lot and it can be found in smaller age people also. Breast cancer occurrence rate is different for different places for example it's occurrence is maximum for the females in the advanced countries as compared to Asia or Africa. When the survey was made by the American Cancer Society, they came to a conclusion that the breast cancer happens to ladies who has overweight, post-partum ladies and postmenopausal ladies. These ladies have to take care of their health properly. In the earlier stage it is not possible to find any indications of breast cancer, but as cancer progresses masses would start developing. These masses are painless. So we should find the way to detect the breast cancer in the earlier stage, through which we can save many lives. One way to detect it in the initial stage is through self-assessment, but is not that effective. One more way is through mammography, and it has been recognized as the most effective method. For the early identification and evaluation of breast alterations screening method known as mammogram can be used. Mammogram is nothing but the x-ray of the breast. Through this mammogram some other breast diseases such as nipple discharge, lump or pain can also be identified and analyzed. For the breast cancer screening presently mammography is the best imaging methodology. The presence of cancer can be identified by examination of digitized mammographic image for calcifications, masses or areas of irregular density. Through this automated techniques mammograms can be easily examined and categorized as normal or abnormal. This helps doctors to take proper decision and give proper treatments to the patients. Identifying the cancer in the earlier stage has saved lives of many women. In our proposed system we are using K Means Algorithm for segmentation of mammograms. Different steps which are used from starting are image pre-processing, segmentation, extracting features, training, testing and comparing the accuracy and training time of both the classifiers. After comparison of results we get the best classifier which classifies the mammogram as either normal or abnormal.

II. SYSTEM DESIGN

In this section, we present an approach to find the best classifier out of support vector machine and feed forward back propagation neural network. The system takes mammogram, pre-processes it by applying Adaptive Histogram Equalization. The enhanced image is segmented using K Means Clustering algorithm. Statistical features such as standard deviation and mean of a segmented mammogram are extracted. These extracted features act as input to the classifier Support Vector Machine and DCT is applied on the segmented mammogram; these extracted features are given as input to feed-forward back propagation neural network. These classify the mammogram as either benign or malignant. The training time and accuracy of both the classifiers are compared to known which is the best classifier.

Algorithm for the proposed approach
Input: Sample mammogram.
Output: Classification of mammogram and comparing accuracy of classifiers.
Step 1: Select the mammogram.
Step 2: Pre-process the mammogram to enhance it.
Step 3: Segment the mammogram using K means algorithm.
Step 4: Extract the features of segmented mammogram, store the extracted features for training.
Step 5: Build the Support Vector Machine and Feed Forward Back Propagation Neural Network for training and classification of mammogram.
Step 6: Once the mammogram is classified, the system shows the type of mammogram.
Step 7: Compare the results of both the classifiers and get the best classifier.
Step 8: End.

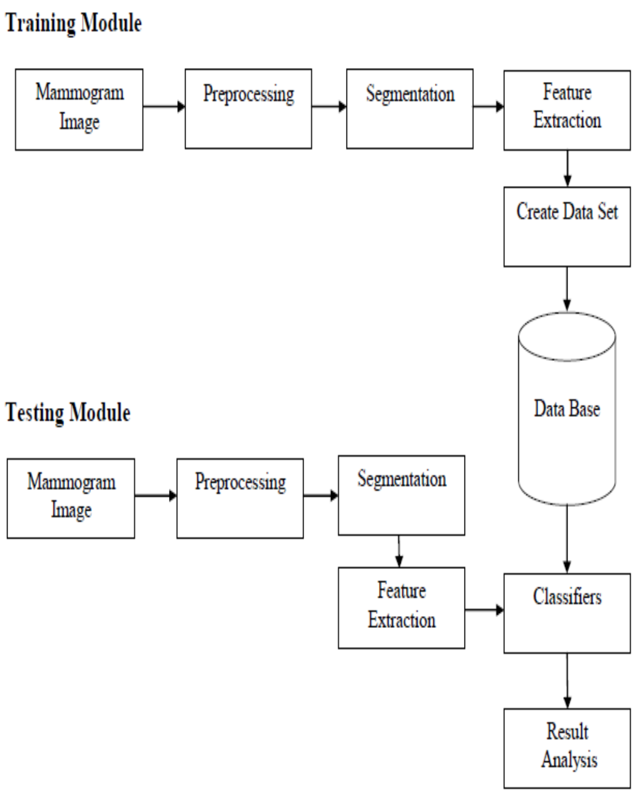

Figure 1 Architecture of the proposed system

III. IMPLEMENTATION OF MODULES

Modules are sub parts of the system each designed for a specific purpose. Over all system can be divided into to two main modules Training module and Testing modules. Details of all the modules used in the system are explained below:

A. Mammograms Acquisition

Different types of mammograms are collected from standard MIAS (Mammographic Image Analysis Society) database for training as well as testing purpose [2]. MIAS database of digital mammograms is generated by UK research group. The database contains 161 patients' right and left breast images. So totally the database contains 322 images. To make all the images of same size i.e. 1024 x 1024 pixel the database has been reduced to 200-micron pixel.

B. Preprocessing

To enhance the quality of the image it is essential to do image pre-processing. Once the image is pre-processed it becomes easy to extract the features of the image and it also helps for trustworthy extraction of features [3].

Image Enhancement: Translation of an image into a superior quality image and more understandable format is called as image enhancement. Adaptive histogram equalization is an image enhancement technique which is applied on mammogram to enhance contrast in image by transforming values in the intensity image[5]. Image enhancement provides better contrast to the regions which have very low contrast. Images which have a foregrounds and backgrounds either dark or bright, this method can be applied.

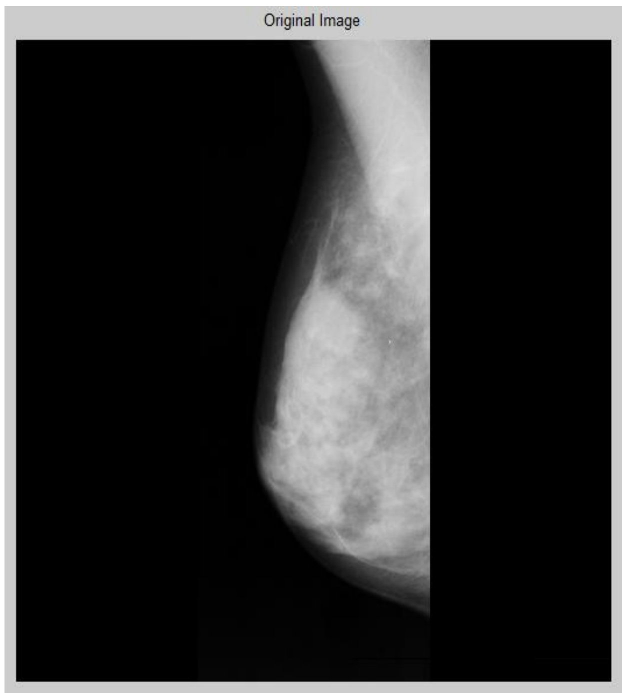

Figure 2 Original Mammogram

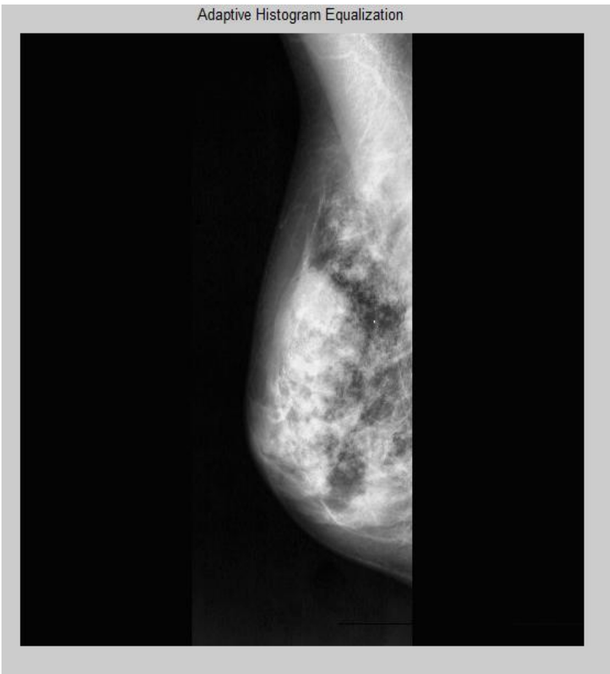

Figure 3 Mammogram after applying Adaptive Histogram Equalization

Figure 2 shows the original mammogram selected from standard MIAS database. After application of Adaptive Histogram Equalization mammogram is improved. The improved mammogram is displayed in the figure 3. In figure 2 the regions which had low contrast are improved and shown in figure 3. Once the original mammogram is enhanced enhancement factor is calculated and it is displayed on the screen [2].

A. Segmentation

Dividing an image into distinct regions where the pixels of each region have similar characteristics is called segmentation [4]. The success of image analysis depends on an effective image segmentation process. K-Means clustering algorithm is used to device the segmentation of mammogram. The algorithm works as follows,

K Means clustering algorithm:

Step 1: Number of clusters must be known previously to be K

Step 2: Select K number of cluster centers such that they are farthest apart from each other

μ_i = some value , i=1,...,k

Step 3: Consider each pixel and assign it to the cluster which is closest

$c_i = \{j : d(x_j, \mu_i) \leq d(x_j, \mu_l), l \neq i, j=1,...,n\}$

Step 4: Recalculate cluster centers by finding mean of pixels belonging to the same cluster

$$\mu i = 1/|ci| \sum j \in ci \; xj, \forall i$$

$$|c| = \text{number of elements in c}$$

Step 5: Repeat step 3 and step 4 till shifting of cluster centers are observed [4].

Figure 4 Mammogram after applying K Means Clustering

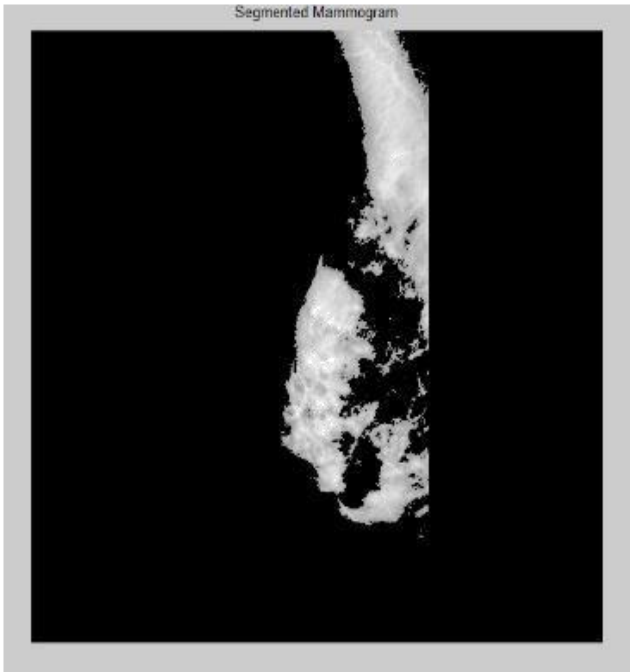

Figure 5 Segmented Mammogram

The K Means Clustering algorithm is enforced on the enhanced image. Here the value of K is three. Total three clusters are formed. The clustered image is shown in figure 4. The black cluster shows the area which is of not at all useful. The gray cluster is also not so important. The brightest cluster shows the area which is of very important [4]. This brightest cluster is selected and segmented. The segmented mammogram is shown in figure 5.

C. Feature Extraction

Converting a mammogram image into set of features is called feature extraction of mammogram [6].

Very much care should be taken while selecting the features. Because depending on these features only the mass present in the mammogram can be detected. So the features provide complete information about the mammogram. To assign the images to different classes it is necessary to select proper features for each and every image. Once the features are extracted these are provided as input to the classifiers, then depending on these features the classifier classify the mammogram.

To extract features first all the images are resized to same size for training and testing. Then fft2 i.e. two dimensional discrete Fourier transform is applied. After applying DFT on the image, it creates a set of samples which can completely represent the original image. So it does not have all the frequencies of image but only set of frequencies through which image can be represented. The number of pixels in the image are equal to the number of frequencies in the image. It represents that the size of the image in the Fourier domain and image in the spatial domain are equal. The two-dimensional DFT for the image of size N x N can be represented as

$$F(k,l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j)\ e^{-\iota 2\pi \left(\frac{ki}{N} + \frac{lj}{N}\right)}$$

Here in the above equitation bias function is represented by the exponential term, which is nothing but representation of Fourier space's each point *F(k,l)* and image which is present in the spatial domain is represented by *f(a,b).* The interpretation of equation can be done as: by obtaining the product of spatial image with the respective base function and adding the output, the value of each point *F(k,l)* can be obtained [4].

- Mean – Average of a set of two or more numbers is called as Mean. Along various dimension of an array average values of elements is calculated and it is returned as mean value. The following equation is used to calculate value of mean.

$$\text{Mean} = \frac{\sum_{i=1}^{N} Fi}{N}$$

- Standard deviation– In statistics and probability theory, the standard deviation (SD) is the quantity of difference from the mean. If the data points are very much inclined towards the mean value then we have very low standard deviation value or else if the standard deviation value is very high then it shows that over a huge variety of values the data points are spread. Standard deviation is calculated using the following formula

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{N} (Fi-\mu)2}{N}}$$

Where µ is the mean of *Fi*, N is the number of elements in the sample. These features are fed as input to SVM.

A discrete cosine transform (DCT) represents a limited sequence of data points in the form of addition of cosine functions fluctuating at various frequencies. DCT is same as DFT and deals with Fourier related transformation, and it only makes use of real numbers. After applying DCT on a mammogram thirty-five features are extracted for each mammogram and stored in the database. These features are used by the FFBPNN classifier.

D. Training and Testing

In this work two classifiers are used namely SVM and FFBPNN. How these two classifiers are trained and used for testing is explained below.

SVM

The extracted features are then used to train the SVM [7]. The inbuilt function provided by MATLAB "svmtrain" is used for training of svm.

To test the mammogram follow the same procedure till feature extraction to extract the features. The inbuilt function "svmclassify" compares the extracted features of mammogram under test to all the features of mammograms in the database. Depending on the feature it classifies the mammogram as either benign or malignant.

FFBPNN

The extracted features after applying DCT on mammogram are supplied as input to the feed-forward back propagation neural network at the input layer. Random weight and bias values are assigned and output is calculated by summation unit and transfer function sends output to the next hidden layer. For all the hidden layers this process is continued. At the output layer there is only one neuron which can classify the mammogram as benign or malignant. Then the output is compared with actual output if there is a difference between them it is calculated and treated as error. If any error is there it is calculated and it is propagated back to the last hidden layer. This NN is called Backward Propagation Neural Network, because there is propagation of delta values or error values backward. Depending on the delta value the weight and bias values are changed and process is repeated till there is no error. Once this training process is over, the next step is testing. In testing process the mammogram which needs to be tested is preprocessed, segmented and then DCT is applied and features of the segmented mammogram are extracted. These features are sent as input to the feed-forward back propagation neural network and it will classify the mammogram as either benign or malignant [2].

To create FFBPNN MATLAB provides a function "newff" and "sim" simulate a Simulink model. Inbuilt function "sim" compares the extracted features of the mammogram under test conditions to all the mammograms in the database, if the features are matched it displays the appropriate result to the user.

IV. RESULT ANALYSIS

In this training time of the both the classifiers are obtained and compared. Also accuracy of both the classifiers are obtained and compared.
Following table represents the accuracy of SVM

Table 1 Performance analysis after testing different mammograms using SVM

		Mammogram Classified As		
	Total = 50	Abnormal	Normal	
Actual Mammogram	Abnormal	TP = 22	FN = 03	25
	Normal	FP = 04	TN = 21	25
		26	24	Total = 50

Accuracy= (TP+TN) / (TP+TN+FP+FN)

= (22+21)/(22+21+04+03))

= 43/50

= 0.86

The accuracy of the SVM is 86.00%

Following table represents the accuracy of FFBPNN

Table 2 Performance analysis after testing different mammograms using FFBPNN

		Mammogram Classified As		
	Total = 50	Abnormal	Normal	
Actual Mammograms	Abnormal	TP = 24	FN = 01	25
	Normal	FP = 01	TN = 24	25
		25	25	Total = 50

Accuracy= (TP+TN) / (TP+TN+FP+FN)

= (24+24)/(24+24+01+01)

= 48/50

= .96

The accuracy of the FFBPNN is 96.00%
From the above tables it is clear that accuracy of FFBPNN is 10% more than SVM

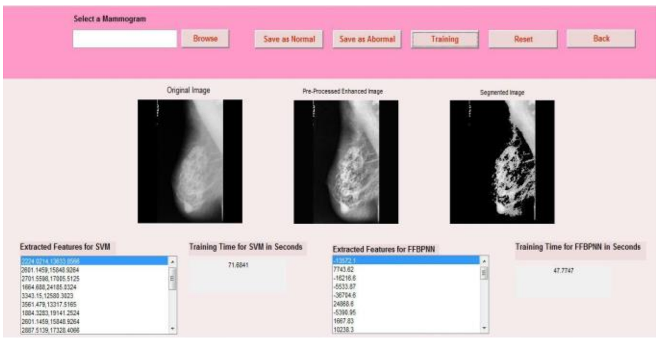

Figure 6 Feature extraction and training time of SVM and FFBPNN.

From the figure 6 it is clear that the training time for SVM is 71.6841 seconds and for FFBPNN it is 47.7747 seconds. From the figure 6 it is clear that training time for FFBPNN is less compare to SVM.

Depending on these values it is clear that out of these two classifiers FFBPNN is the best classifier. Because it takes less training time and more accurately classify the mammogram as either normal or abnormal.

V. CONCLUSION

This system takes mammogram, and applies Adaptive Histogram Equalization to enhance the image. On this pre-processed image K Means Clustering algorithm is applied to get segmented mammogram. From the segmented mammogram Statistical features such as standard deviation and mean are extracted. SVM takes these features as input. DCT is applied on the segmented mammogram, these extracted features are fed as input to FFBPNN. These classify the mammogram as normal or abnormal. The training time and accuracy of both the classifiers are compared. FFBPNN training time is less and accuracy is 10% more than SVM. The approach which is based on FFBPNN got a better results compared to SVM.

In future we can increase accuracy by using other classification methods or by extracting other features.

ACKNOWLEDGMENT

The author would like to thank the Principal of BCCA for support and guidance for doing the research work and Bharatesh College Of Computer Applications, Belagavi for providing the support throughout the Project.

REFERENCES

[1] S. Shanthi, V. Murali Bhaskaran, "Computer aided detection and classification of mammogram using self-adaptive resource allocation network classifier", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering , **March 21-23, 2012**

[2] Vijaylaxmi Kochari "Feed Forward Back Propagation Neural Network for Detection of Breast Cancer" *International Journal of Computer Science Trends and Technology (IJCST)* – **Volume 6 Issue 6, Nov-Dec 2018,** *ISSN :* 2347-8578.

[3] K Menaka, S Karpagavalli , "Mammogram Classification using Extreme Learning Machine and Genetic Programming," International Conference on Computer Communication and Informatics (*ICCCI* -2014), **Jan. 03 – 05, 2014**

[4] Digambar A Kulkarni, Vijaylaxmi K Kochari "Detection of Breast Cancer Using K Means Algorithm" International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, **Volume 6, Issue 4, April 2016.**

[5] S .Bhadra "Analysis of Tumor Detection Methods for Mammogram Images" International Journal Of Computer Science and Engineering **Vol.- 7, Issue- 5, May 2019** E-ISSN : 2347-2693

[6] M Arfan Jaffar, Nawazish Naveed, Sultan Zia, Bilal Ahmed and Tae-Sun Choi, "DCT Features Based Malignancy and Abnormality Type Detection Method For Mammograms"

International Journal of Innovative Computing, Information and Control **Volume 7, No.9, September 2011.**

[7] Leonardo de Oliveira Martins, Geraldo Braz Junior, Aristofanes Correa Silva,Anselmo Cardoso de Paiva, and Marcelo Gattass_, “Detection of Masses in Digital Mammograms using K-means and Support Vector Machine”, Electronic Letters on Computer Vision and Image Analysis **8(2):39-50, 2009**

AUTHORS PROFILE

Mrs. Vijaylaxmi Kochari pursed Bachelor of Engineering in Computer Science and Master of Technology in Computer Science from Gogte Institute Of Technology, affiliated to Visvesvaraya Technological University Belagavi, Karnataka, INDIA in the year 2016. She is currently working as Assistant Professor in Bharatesh College of Computer Applications, Belagavi affiliated to Rani Channamma University, Belagavi, Karnataka, INDIA since 2012. She is a member of CSI. She has published research papers in reputed international journals related to cancer identification. Her main research work focuses on image processing, Machine Learning. She has 10 years of teaching experience and 2 years of Research Experience.