ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Breast Cancer Classification using Feed Forward Back Propagation Neural Network (BPNN)Classifier

Rose Merlin Jose[1], Sangharsh Syal[2]

B.E., Dept. of Biomedical Engineering, W.I.E.E.C.T, Mumbai University, Worli, Maharashtra, India

Assistant Professor, Dept. of Biomedical Engineering, W.I.E.E.C.T , Mumbai University, Worli, Maharashtra, India

ABSTRACT: The project proposes Breast Cancer Classification using Feed Forward Back Propagation Neural Network (BPNN) classifier. Breast cancer is the leading cause of non-preventable cancer death among women. The detection of the breast cancer is a challenging problem, due to the structure of the cancer cells. A typical mammogram is an intensity X-ray image with gray levels showing levels of contrast inside the breast that which characterize normal tissue and different classifications and masses. Therefore, it poses inaccuracy in identifying the presence of breast cancer. Now a days, detection of classifications in mammograms has received much attention from researchers and public health practitioners. Our experimental result of Breast Cancer Classification using Feed Forward Back Propagation Neural Network classifier has achieved excellent classification accuracy compared with the other technique like support vector machine.

KEYWORDS: Breast cancer detection techniques, Mammogram,Feed Forward Back Propagation Neural Network (MLP)Multi-layer Perceptron.

I. INTRODUCTION

A.Breast Anatomy

Histologically, the breast is divided in 15 to 20 lobes or segments. Working backwards from the nipple, each lobe begins with a major duct that transfers the milk to the nipple during lactation. This major duct branches several times, forming minor (or sub segmental) ducts with correspondingly smaller diameters. Finally, the branching ducts end up in Terminal Ductal Lobular Unit's (TDLU's). A TDLU consists of a lobule and its extra lobular terminal duct. The lobule is responsible for milk production during the period of lactation and it is build up from 10 to 100 sac-like units called acini. As shown in figure 1, the ductal and lobular system as a whole is surrounded by an uninterrupted basement membrane on the outside. With age the breast tissue will change. In young women the breast tissue is dense and rich of glandular tissue. On aging, the glandular tissue is gradually replaced by fat. This increased fat content of the breast in older women makes their mammograms relatively easier to diagnose.

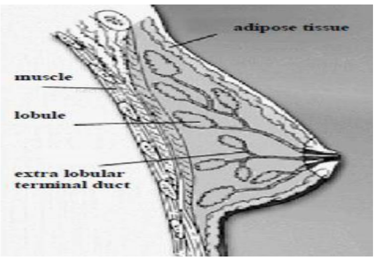

Fig 1 Anatomy of breast

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

B. Breast Cancer

Breast cancer is the uncontrolled growth of abnormal cells in the breast. As with other forms of cancer, breast cancer is considered to be a result of malfunctioning DNA due to damage or inherited mutation. Breast cancer is a disease that typically develops in women; however, it is also possible, although rare, for breast cancer to develop in men. These cancers start in the lobules or ducts of the breast but have broken through the duct or glandular walls to invade the surrounding tissue of the breast. The seriousness of invasive breast cancer is strongly influenced by the stage of the disease; that is, the extent or spread of the cancer when it is first diagnosed. Many risk factors for the disease have been identified, such as family history of breast cancer, late age of first birth, early onset of first menstruation and late age at menopause. As far as current knowledge goes, these risk factors cannot explain the major part of the incidence. Moreover, these factors are not a basis for prevention. Therefore, at this point in time, the most important strategy to reduce breast cancer mortality seems early detection through organized breast cancer screening programs. Early diagnosis enables effective treatment and thus increases the survival chance.

C.Mammography

Mammography is a diagnostic breast imaging method using X-rays. It is widely used to detect and characterize breast cancer and because of its high performance and low costs it is by far the most suited imaging technique for screening programs. A typical mammogram is a low intensity x-ray image with gray levels showing levels of contrast inside the breast which characterize normal tissue and different calcification and masses. The contrast level of a typical mammogram image is proportional to the difference in x-ray attenuation between different tissues. In general, a clear separation between normal functioning tissue and abnormal cancerous tissues is difficult to identify. Since their attenuation is very similar.

There are two general types of mammography: screening and diagnostic. Screening mammography aims at detecting clinically occult breast cancer, i.e., cancer that is still too small to be felt by a physician or during self-examination by the patient. It is the regular low-dose x-ray examination of the breast that is performed on non-symptomatic women for early detection and prevention purposes. Diagnostic mammography is an x-ray examination of symptomatic women that is performed to pinpoint the exact size and location of the palpable or non-palpable abnormalities detected during physical examination or routine screening mammography. It is also aims to image the surrounding tissue and lymph nodes for possible spread of the disease. Even though it is usually more time-consuming and costly than screening mammography, it sheds more light on the likelihood of malignancy for the abnormality and may avoid the need for a biopsy of the lesion.

II. LITERATURE SURVEY

Prediction of breast cancer in mammogram image using SVM and Fuzzy C-means (GulShairaBanu, AmjathFareeth, Taif University, Saudi Arabia)

In this work they have used a novel technique that uses continuous wavelet transform as feature selection technique and SVM as classifier. The Support Vector Machine classifier exhibits superior efficacy of pattern recognition, in comparison to f-c mean .The true positive rate is raised up to 100% and the false positive is Zero, if CWT and SVM are both employed. Finally, SVM shows the no MSE in Classification and Prediction

Evaluation of Breast Cancer risk by using Fuzzy logic (Victor BALANICĂ, Ioan DUMITRACHE, Mihai CARAMIHAI, William RAE, Charles HERBST)

The aim of this work is to propose a fuzzy logic technique for the prediction of the risk of breast cancer based on a set of judiciously chosen fuzzy rules utilizing patient age and automatically extracted tumor features. For future research, this method can be extended to include parameters like the number of invaded axillary nodes, calcifications, disease extent, etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Breast Cancer Classification System by Means of SVM and Extensive Feature Set(Suman Mishra, RGCE, Chennai)

In the proposed technique in this paper, initially the suspicious regions containing micro calcifications in the digital mammograms are extracted and then they are classified into two categories that is, whether they contain benign or malignant cancer cells. Here, the learning of extracted features as well as the classification is done by using the SVM Classifier. As a result in this paper SVM based mammogram classification system was proposed for early detection of breast cancer.

III. FEED FORWARD BACK PROPOGATION NEURAL NETWORK

Classification is the categorization of entities, objects or patterns in to specific classes or groups depending on their similarities and characteristics. Neural Network is basically a statistical computational model of the human biological nervous system. Neural Networks are considered one of the most competent techniques in the applications regarding Decision Making, Pattern Recognition, and Target Detection etc.

Neural networks are mathematical and computational equivalent model of biological neural network. The model is rough, as the human brain is a parallel computational device, which achieves great power due to connectivity of billions of simple neurons.

The general process responsible for training the network is mainly composed of three steps:

1. Feed forward the input signals
2. Back propagate the error
3. Adjust the weights

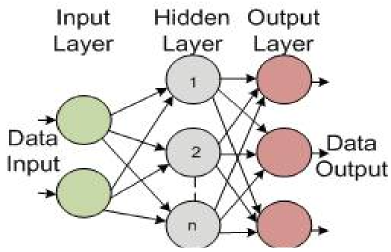

Fig 2 Basic structure of a NN

The input data being fed up at the input layer and the output data being collected at the output layer. Training the NN is divided into four steps:

1. Data selection- Data selection step restricts subsets of data from larger databases and different kinds of data sources. This phase involves sampling techniques, and database queries.

2. Data pre-processing– Data pre-processing represents data coding, enrichment and clearing which involves accounting for noise and dealing with missing information.

3. Data transformation- Data transformationhas the purpose to convert data into a form suitable to feed the NN. For example categorical data like (YES/NO) cannot be used to train the NN. Therefore at this step YES/NO inputs are transformed into numerical values +1/-1.

4. NN selection and training - FFNN (MLP) is suitable for causal forecasting, while TDNN, RNN and NARX NN are more suitable for time series forecasting. Therefore, according to these specifications, a NN model is selected with respect to the available data. Note that TDNN, RNN and NARX NN can also be used for causal forecasting.

ALGORITHM

Step 1 :The dataset instances are mixed to provide most generalized results.

Step 2: The dataset is divided in to two parts; the training and the testing datasets.

Step 3: The training dataset comprises of about 70% of the original dataset and the testing dataset occupies about 30%.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Step 4 :Now the neural network is trained on known target value (in our case binary).

Step 5: After the BPNN has been trained, the testing dataset that the neural network has never seen is applied to check the accuracy of the classification.

IV. PROPOSED METHODOLOGY

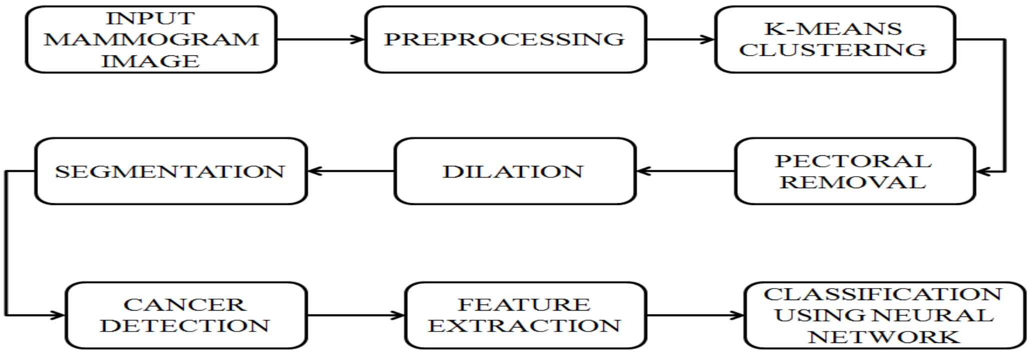

Fig3 Block Diagram of System Design

Pre-processing

It is indispensable to perform preprocessing on the mammography image for increasing the quality ofthe image as well as for making the feature extraction more reliable. In this stage, the noises in theimages are removed, and then the black background and the existing artifacts such as the written labeletc. are also removed by using the cropping technique. Cropping removes the unnecessary parts of theimage particularly peripheral to the area of interest. The cropping operation was made automatically bysweeping through the image and cutting the superfluous regions of the image horizontally andvertically that have the mean below a certain predefined threshold value.

K-meansClustering

This nonhierarchical method initially takes the number of components of the population equal to the final required number of clusters. In this step itself the final required number of clusters is chosen such that the points are mutually farthest apart.Next,it examines each component in the population and assigns it to one of the clusters depending on the minimum distance. The centroid’s position is recalculated every time a component is added to the cluster and this continues until all the components are grouped into the final required number of clusters.

Pectoral Removal

Pectoral muscles are the regions in mammograms that contain brightest pixels. These regions must be removed before detecting the tumor cells so that mass detection can be done efficiently. Pectoral muscles lie on the left or right top corner depending on the view of the image and must be detected before removing it. For this, searching for nonzero pixels are simultaneously done from the left and right top corner. Width of the image in which the non-zero pixel detected from both the corner are counted and compared. If the left width is smaller than the right width then it is assumed that pectoral is on the left side of the image else it is on the right side. From the detected corner pixel the intensity discontinuity is detected on each and every column of the same row. Coordinates of the pixel in which the intensity change is encountered is considered as width of the pectoral region. All the pixels, which lie inside pectoral width and half of the height of the whole image is segmented from the original image. This rectangle shaped image contains the entire pectoral muscles. To extract the pectoral muscles from this image binary image should be obtained by simple Thresholding. This binary image contains pectoral muscles and other tissues. To segment the pectoral muscles alone from the binary image raster scanning is done from the right or left side of the image to detect the intensity discontinuities. The resulting image contains pectoral muscles alone and this region is completely removed from the original image.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Dilation

Dilation is one of the basic operations in mathematical morphology. Originally developed for binary images, it has been expanded first to grayscale images, and then to complete lattices. The dilation operation usually uses a structuring element as shown in figure 3for probing and expanding the shapes contained in the input image.

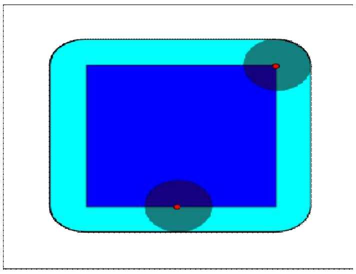

Fig 4 Dilation image

Segmentation

After preprocessing, the image contains dark backdrop with low intensity level and relatively small brighter fragments. The strong tissue structure with lower intensity level is considered as background and the brighter regions are micro calcifications. In this stage, the calcification objects are extracted based on the threshold value which should make a compromise between the elimination of the background and detection of the accurate micro calcification with good descriptions of the shape which is crucial for the object feature extraction. The intensity values of the filtered image that are greater than the threshold value are belongs to the micro calcification. However it is impossible to achieve 100% separatibility between noise and micro calcification for more complicated situations. The threshold value is extracted based on the experiments as follows.

Let $a=\{a_{ij} \mid 1\le i\le r; 1\le j\le c\}$ be the mammogram image after preprocessing, then threshold $T=\mu+k\sigma$

Where $\mu = \frac{1}{i*j}\cdot\sum_{i=1}^{I*j} a_{ij}$

$$\sigma=\sqrt{\frac{1}{i*j}\cdot\sum_{i=1}^{I*j}(a_{ij}-\mu)^2}$$

Feature Extraction

After segmentation, the features relevant to the classification are extracted from the segmented images. The selected features for the classification process should return an accurate specificity for micro calcifications mammogram and also it should emphasize the dissimilarity between benign and malignant. Using some image processing techniques, the visual content features such as texture, shape, gray level intensity, edge strength, area and bounding region of the micro calcification are extracted from the segmented images. Then, a feature that is already available in the original database that is type of tissue (dense, fatty and fatty-glandular) is combined with these extracted features for improving the effectiveness of classification.

Feed Forward Back Propagation Neural Network

Backpropagation Neural Network is a network that based on Backpropagation learning technique and that works on the principle of supervised learning. In general it is called the Feed Forward Backpropagation neural network. With regard to architecture it is basically a Multi-layer Perceptron.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

The Backpropagation neural network is the gemstone that enchanted and mesmerized researchers and showed the true power of neural networks. It opened research doors with endless opportunities in various fields of engineering, sciences and statistics while being computationally economical. But on the darker side the BPNN has also been called the 'black box' as it has a fixed algorithmic operation and that's about it, there is no fixed topology (number of nodes and neurons used) for it and exhibits different results with different data subsets of the same dataset that is it is very hard to sway it to global minima. Regardless of all these factors, overall the BPNN is quite accurate and easy to manipulate with respect to other neural networks.

V. EXPERIMENTAL RESULTS

The proposed technique is implemented in the working platform of MATLAB (version 7.10) and its performance is evaluated using the MIAS (MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY) Mini Mammography Database. The size of all the images is 1024 x 1024 pixels. The images have been centered in the matrix.

In this work, a mammogram image is taken as the input as in Fig 5; which is pre-processed by resizing and cropping unwanted portions so as to improve the quality of the image Fig 6 & 7. Fig 8 shows the result image attains the final required number of clusters by k- means clustering. As in Fig 9 Pectoral removal must be done before detecting the tumor cells so that mass detection is done efficiently; for extracting the pectoral muscles from the image, a binary image must be obtained by simple thresholding as in Fig 10.Fig 11 shows the dilated mammogram image which is done to expand the input image. Fig 12 shows segmented image which is essential for feature extraction. Table 1 shows the various parameters and certain values taken to determine whether the tumor is Malignant or Benign. Fig 13 (a) shows the tumor is malignant when the parameter values as in Table 1 were considered. Fig 13 (b) shows the tumor is benign when the parameter values as in Table 1 were considered.

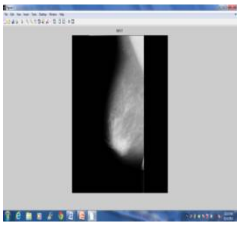

Fig 5 Input Mammogram Image

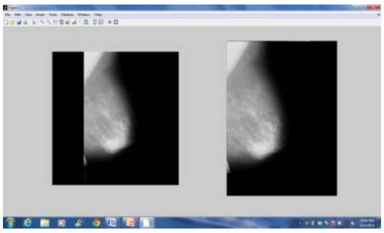

Fig 6 Resizing of input mammogram image

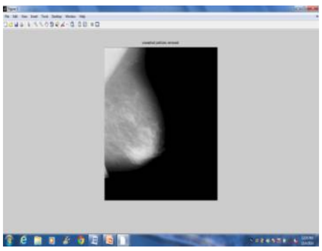

Fig 7 Unwanted portions removal

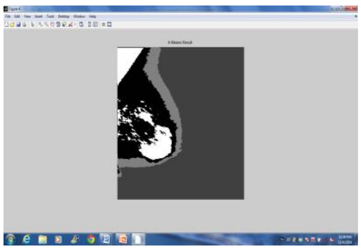

Fig 8 K-means result image

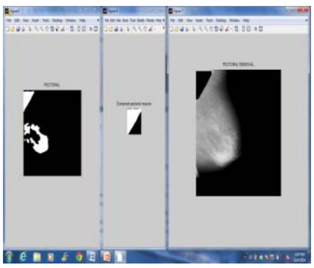

Fig 9 Pectoral removal

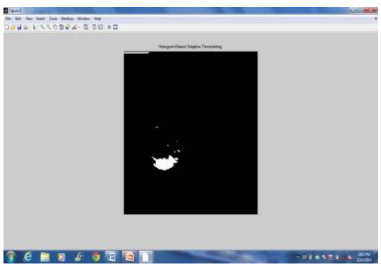

Fig 10Histogram-based adaptive thresholding

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

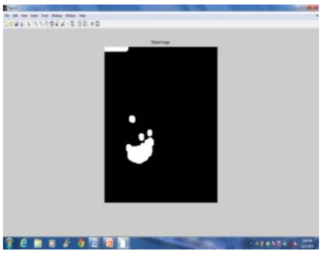

Fig 11 Dilated mammogram image

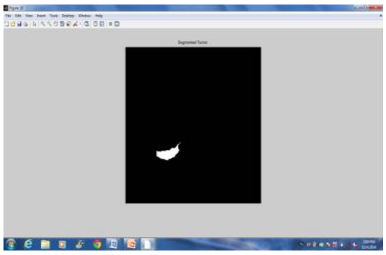

Fig 12 Segmented mammogram image

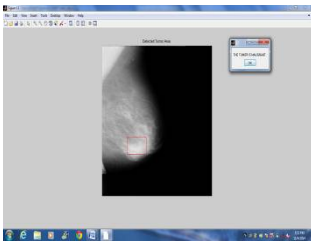

Fig 13 (a) Malignant Tumor

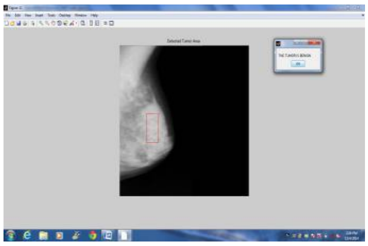

Fig 13 (b) Benign Tumor

Table 1 CLASSIFICATION RESULTS FOR CANCER DETECTION

PARAMETERS	MALIGNANT	BENIGN
ENTROPY	0.0715	0.0452
STANDARED DEVIATION	0.0827	0.0649
MEAN	0.0086	0.0050
VARIANCE	0.0085	0.0049
COVARIANCE	2.4351e-004	4.5540e-005
TUMOR AREA IN Sq mm	565	294
DATA	5.6517	2.9412

VI. CONCLUSION AND FUTURE WORK

In this work, a Feed Forward Back Propagation Neural Network based mammogram classification system was proposed for detection of breast cancer. The biological features as well as the visual content features were used for the classification of mammograms.K-Means clustering process was applied on the extracted features and the extracted features are combined to obtain better classification performance. The proposed system with K-Means clustering has given higher performance and confirmed that the performance of our proposed technique is better than the other

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

classification systems.Further the future works can be done in this one by using other artificial neural networks (ANN), K-nearest neighbour (KNN) etc.

REFERENCES

1. Gul Shaira Banu,Amjath Fareeth, NisarHundewale , ,College of Computers and information Technology, Taif University ,'Prediction of Breast Cancer in Mammagram image using Support Vector Machine and Fuzzy C-means', IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI 2012) Hong Kong and Shenzhen, China, 2-7 Jan 2012.
2. Shekhar Singh1, Dr P. R. Gupta2,1C-DAC, NOIDA,'Breast Cancer detection and Classification using Neural Network'.(IJAEST) International Journal of Advanced Engineering Sciences and technologies, Vol No. 6, Issue No. 1, 004-009.
3. http://www.mammoimage.org/databases: Database-Mammographic Image Analysis Society(MIAS).
4. Rafael .C.Gonzalez and Richard.E.Woods.,Digital image processing (3rd edition).

BIOGRAPHY

Rose Merlin Jose completed B.E. in Biomedical Engineering from Watumull Institute of Electronics Engineering and Computer Technology, Mumbai University. Her research interests are in Medical devices, Image Processing, bio robotics, biomedical instrumentation, biomaterials, and cardiovascular mechanics.

Sangharsh Syal is an Assistant Professor in the Biomedical Department,Watumull Institute of Electronics Engineering and Computer Technology, Mumbai University in India. Her areas of research are on Biomedical Instrumentation, Modelling and Simulation ,Image Processing and Nuclear Medicine.