Automatic Breast Tumor detection and classification of Asymmetries in Mammograms using Neural Network Classifier and Hybrid GA

S.Thilagavathi[1], M.S. Irfan Ahmed[2]

Abstract

Breast Cancer has become a common mortality factor in India. Despite the fact, not all general hospitals have the mammogram facilities. The long waiting for diagnosing a breast cancer may increase the possibility of fatality and the cancer spreading. Therefore a computerized breast cancer may increase diagnosis prototype has been developed to reduce the time taken and indirectly reducing the probability of death. Micro calcification on X-ray mammogram is a significant mark for early detection of breast cancer. In this paper , A hybrid Genetic algorithm and Neural Network is proposed to automatically detect the suspicious regions on digital mammograms based on irregularity between left and right breast image. The basic idea of Irregularity approach is corresponding left and right images are subtracted to extract the suspicious region. The proposed system consists two steps: First, the mammogram images are enhanced using the median filter, pectoral muscle region is removed and the border of the mammogram is detected for the both left and right images of the binary image.

Keywords:

Breast Cancer, Neural Network, Genetic Algorithms, Breast Tumor Classifications, Image Segmentation Algorithm.

1. Assiatant Professor, Department of Computer Application, Hindusthan College of Engineering & Technology,

2.Director, Department of Computer Application, Hindusthan College of Engineering & Technology, Coimbatore-32, TamilNadu, India.

E-mail: thilaga.hicet@gmail.com,msirfan@gmail.com

1.Introduction

Breast Cancer is the one of the leading cancers. About 10% of all women develop breast cancer and about 25% of all cancers diagnosed in women are breast cancers[1]. Although effective prevention is not possible, early detection can at least reduce the chance of breast cancers from becoming incurable[2]. Mammography has been shown to be the most effective and reliable method for early cancer detection. Mammogram interpretation is both time-consuming and difficult, requiring the expertise of trained radiologists.

Breast Cancer as a Domain

Breast tumor may be benign or malignant lesions. Benign is a non cancerous lesion and malignant is cancerous lesion[3] .Benign lesions in the breast tend to be well defined and circumscribed, while malignant lesions appeared ill defined with speculated margins. Classification of breast lesions can be performed based on border characteristics of lesions, by using the fact that benign lesions have smooth borders, while malignant lesions have rougher border.

Breast cancer is a malignant tumor that has developed from cells of the breast. A malignant tumor is a group of cancer cells that may involve surrounding tissue or spread to distant areas of the body.

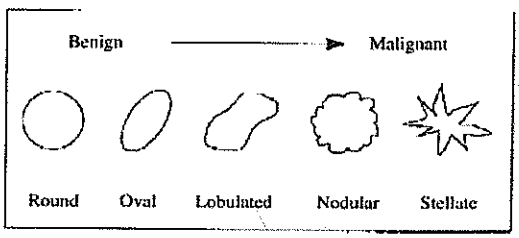

Fig 1: Morphologic Spectrum of Masses

The etiologies of breast cancer remain unclear and no single dominant cause has emerged. Preventive way is still a mystery and the only way to help patients to survive is by early detection. If the cancerous cells are detected before spreading to other organs, the survival rate for patient are more than 97%. Therefore it has been our motivation factor to develop such system.

Some early signs of having high possibility of cancerous cells, like microcalsifications, mass, architectural distortion and breast asymmetries; might be detected from mammogram images. Mammogram[4], Fig 2, introduced in 1969 is the first digital steps in detecting cancerous cell in breast tissues, and it has been a useful tool in diagnosing breast cancer ever since.

A closer inspection of the mammograms reveal several difficulties for the asymmetry approach. First, the global appearance (brightness, contrast, etc.,) of the two breasts may differ, usually due to variations in the recoding procedure. This can be solved in the enhancement phase. Second, due to natural asymmetry, and due to the mammography recording procedure, the shapes of the left and right breast do not match. Defining corresponding positions in both breasts becomes therefore a nontrivial task. Third, asymmetries exist not only at the tumor position but also in the appearance of healthy breast tissue at corresponding locations in the two breasts. The asymmetry method thus has to discriminate tumor related asymmetries from naturally occurring asymmetries.

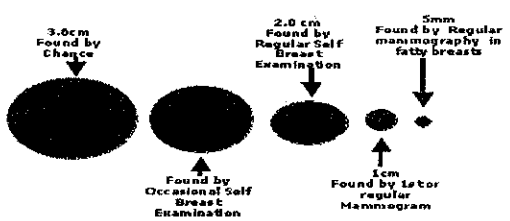

Fig 2: Sizes of mass that may be detected by mammogram or hand

As some of cancerous tissues can be very aggressive[4], it is very important to identify them as early as possible. But the waiting time from capturing mammogram images to biopsy result is 2 weeks to a month, on average. However, 10-30% visible abnormalities are usually detected, due to technical or human error. The whole procedure in diagnosing breast cancer in woman is shown in Figure 3 below.

The effectiveness of early detection has been proven to reduce a lot of mortality among breast cancer patients[5]. As a proof, 80% of American society detected cases are still in early stage, but the mortality among them is only 30%in 2006, believed as a result of early detection and improved treatment[6].

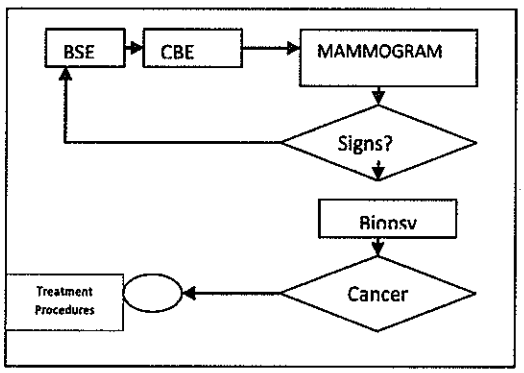

Fig.3 : Procedure in diagnosing breast cancer

II. Artificial Neural Network as a Classifier

To date many of the organizations, groups and educational researchers have put an interest in computerizing breast cancer diagnosis and detection. However, India is still lacking of Artificial Neural Network(ANN) application in Medical field. Therefore, some research in computerizing breast cancer diagnosis, as well as implementation of artificial intelligence method in the system.

In ANN(Artificial Neural Network), the assigned weights for each connector of node resemble the long term memory. They contain information of the input's importance and ANN learns by repeated adjustments of these weights. The weight adjustments are carried out according to the mathematical functions known as learning or activation function, which will be compared to the threshold value of the network.

To develop the prototype model, 15 ANN models in computerizing breast cancer diagnosis are computed and analyzed. Among these, six have used BPNN[8][9][10][11][12][7] and 7 have included image processing in their application. These researches has included many ANN architecture including Convolution Neural Network, radial Basis Function in Neural Network, hybrid and Fuzzy logic as well as various network architecture and attributes.

A closer inspection of the mammograms reveals several difficulties for the asymmetry approach. First, the global appearing (brightness, contrast, etc.) of the two breasts may differ, usually due to variations in the recording procedure. This can be solved in the enhancement phase. Second, due to natural asymmetry, and due to the mammography recoding procedure, the shapes of the left and right breast do not match.

III System Approach and Development

From papers discussed, many ANN model including convolution, Radial-basis-function, Resilient Back propagation and a lot more with different network architecture and parameters, or a hybrid with fuzzy logic as well as GA has been discussed and tested. It has been proven that sigmoid activation function give better result for predictions and this is inline with Haykin[20]. As for number of hidden layers and neurons, there are still debates as it really depends on the NN architecture and number of input.

Apart from that, Multilayer perceptrons with back propagation has provides good result too in predicting and classifying breast cancer. The only matter is to find out the effect of including other features as input, best NN architecture, number of hidden layers as well as the neurons and the best weight distribution functions for an almost perfect result. Initial weights are important as it may affect the NN ability and performance.

Therefore , a hybrid function using GA and BPNN will be tested in this research to solve some of the stated problems. Optimum initial weight for each layer in the NN architecture is calculated with GA approach, following the method suggested by Khairuddin Omar, while for mining process, multilayer perceptrons with back propagation will be used. Ideally, setting the initial weight for NN model will reduce the time needed for mining process as it has a good staring point to start the learning process. The overall processes flowchart in GAwNN prototype model are as shown in figure 4.

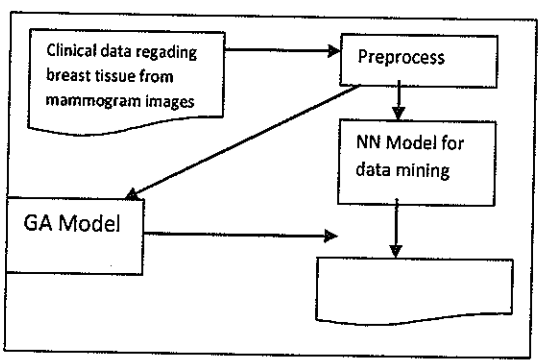

Fig. 4. GAwNN prototype model

IV. The Neural Network Classifier

Several Architecture on the input-hidden-output nodes were tested to avoid overfitting problem. Overfitting is a phenomena where the network will simply memories

all training examples, thus prevent it from generalizing, or producing correct output when presented with test data. This happens if the number of hidden neurons is too big.

However , simulation or testing is very important as number of hidden neurons as well as the layer affects both the accuracy of pattern recognition and the speed of training. Complex pattern might not be detected by small number of hidden neurons, but too many of it might increase the computational burden. Tsoi[24] and Hush & Horne[25] have suggested that simulating training process with increasing number of hidden nodes to identify the best solution.

V. Classification Alogrithm

In the present work, the neural networks are used for cancer affected micro object classification purposes. The neural networks derive their power due to their massively parallel structure, and an ability to learn from experience. Neural network can be used for fairly accurate classification of input data into categories, provided they are previously trained to do so. The accuracy of the classification depends on the efficiency of training. The knowledge gained by the learning experience is stored in the form of connection weights, which three issues used to be settled in designing an ANN for a specific application.

- Tolpology of the network:
- Training algorithm;
- Neuron activation functions.

In our topology, the number of neurons in the input layer is 8 neurons for the ANN classifier. The output layer was determined by the number of classes desired. The outputs are type4, type3, type2, type1 micro object therefore; the output layer consists of 4 neurons. The hidden layer consists of 30 neurons. Before the training process is started, all the weights are initialized to small random numbers. This ensured that the classifier network was not saturated by large values of the weights. In this experiment, the training set was formed by choosing 420 data sets for the testing process.

Cancer object (nucleus) is classified into one of four type object(nucleus) using the feed forward back propagation neural network classifier. After the classification of each cancer objects(nucleus), score and grade are computed. The classifier was trained and tested on the images created by one of the experts. The main steps of the classification algorithm are summarized as follows.

Extra Features for Each Nucleus (Micro Object)

The following local and global features are extracted: Area, Diameter, solidity, Eccentricity, Extent, perimeter, Roundness, and Compactness.

5.1 Classify each Cancer Objects:

A cancer objects(Nucleus) is classified into one of four type cancer objects using the feed forward back propagation neural network classifier.

VI. Training and Testing

The proposed network was trained with all 75 Micro Objects (Tumor) data cases. Thses 75 are fed to the FNN with 5 input neurons, one hidden layer of 10 neurons and four outputs neuron. MATLAB software package version 8 is used to implement the software in the current

work. When the training process is completed for the training data (75 cases), the last weights of the network were saved to be ready for the testing procedure. The time needed to train the training datasets was approximately .40 second.

The testing process is done for 50 cases. These 50 cases are fed in to the proposed network and their output is recorded for calculation of the sensitivity, specificity and accuracy of prediction.

VII. Biospy Scoring and Grading

Scoring according to the positive cells

The scoring of the cancer detected biopsy image was evaluated by computing the percentages of positive cancer cells, the respective definitions are as follows.

Table I

Scoring based on the micro objects:

% of cell positive	Score
0	0
1-25%	1
26-50%	2
51-75%	3
>=76%	4

Scoring based on the micro objects:

All the detected cells (micro object) in the high resolution images are classifies into one of the four type's micro objects, Based on the type of object scoring perform.

Overall grading

The overall grade of the patients defined in terms of the total. Then, the overall is assigned as follows:

- Grade I(low grade) : G=3,4,5
- Grade II (intermediate Grade):G=6,7,8

Grade III (high Grade) G=9,10,11

TABLE II
GRADE P: PATHOLOGIST, S: S YSTEM

Patient Image		Grade
Patient Image 1	P	3
Patient Image 1	S	3
Patient Image 2	P	2
Patient Image 2	S	2
Patient Image 3	P	3
Patient Image 3	S	2
Patient Image 4	P	2
Patient Image 4	S	1
Patient Image 5	P	1
Patient Image 5	S	1

VIII. Results and Discussion

Figure 1 shows a sample high – resolution original image. Figure 2 illustrates detected cancer cells for nuclear scoring. Nevertheless, scoring is still reliable as long as enough type-2 and type-3 cells are detected. This is analogous to clinical most 5 image frames to pick up enough type-2 and type-2 cells to make an assessment. Table II compares the grading results of the algorithm and a pathologist. It can be seen the system's scores are very close to the pathologist's scores. The system 's scores tend to be slightly lower than the pathologist's scores. This could be due to slightly more stringent criteria taken by the system. Given these encouraging results, we are confident that an automatic breast cancer detection and grading system can be developed to assist the pathologists by providing second opinions and alerting them to case that require attention. Nevertheless, more comprehensive tests are needed to provide better evaluation of the algorithm's performance. The performance of the classification algorithm was evaluated by computing the percentages of Sensitivity (SE), Specificity (SP) and Accuracy (AC), the respective definitions are as follows.

SE=TP/(TP+FN)*100 (1)

SP=TN/(TN+TP)*100 (2)

AC=(TP+TN)/(TN+TP+FN+FP)*100 (3)

Where TP is the number of true positives. TN is the number of true negatives; FN is the number of False negatives, and FP is the number of false positives. Since it is interesting to estimate the performance of classifier based on the classification of benign and malignant breast cell nuclei, the true positives (TP), false positives(FP), true negatives (TN), and false negatives(FN) are defined appropriately as shown below:

FP: Predicts benign as malignant
TP: Predicts malignant as malignant
FN: predicts malignant as benign
TN: Predicts benign as begin.

Sensitivity, specificity and accuracy of prediction have been calculated according to the above formals for all of the testing data (50 micro object cases). Table 1 shows the resulted SE and SP and AC for testing data of the proposed networks.

No of Cases	Sensitivity	Specificity	Accuracy
50	99.10%	95.70%	98.60%

9. Conclusion :

This paper presented a multi- segmentation method for automatic breast cancer detection and grading of histopathological images. The individual cells are detected and classified in the high- resolution image frame. They are then scored according to the three criteria of the Nottingham system. Given the encouraging test results, we are confident that an automatic grading system can be developed to assist the pathologists by providing second opinions and alerting them to cases that require further attention. FNN has been implemented for classification of micro object of breast cancer tumor. Twenty six hundred sets of cell nuclei characteristics obtained by applying image analysis techniques to microscopic slides of H& E stained samples of breast biopsy have been used in the current work. MATLAB software package version 8 is used to implement the software in the current work. These feature vectors which consist of eight image analysis features each were carried out to generate training and testing of the proposed NN. The accuracy is calculated to evaluate its effectiveness of the proposed network.

The obtained accuracy of the network was 98.60% whereas the sensitivity and specificity were found to be equal 99.10% and 95.70% respectively. We conclude that the proposed system gives fast and accurate classification of breast tumors.

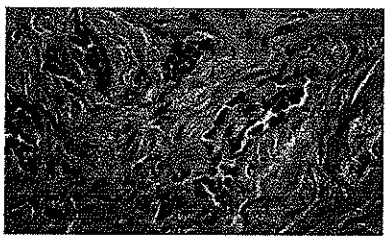

Fig. 1 Biospy Image

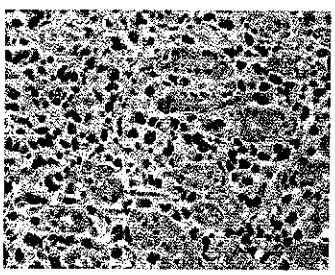

Fig. II Cancer detected Image

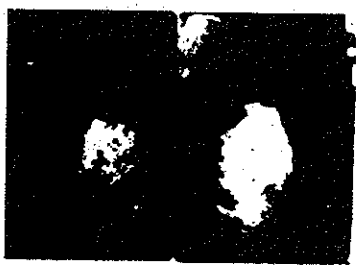

Fig. III Classified Cancer Detected Image.

REFERENCES

1. Ajay Nagesh BAravanhalty, Shridar Gancean, SHANNAN Agner, James Peter Monaco, " Computerised Image based detection and grading of tymphocytic infiltration in HER2 breast cancer histopathology ", IEEE Transactions on biomedical Engieering vol 57, no-3 march-2010.

2. Yu –Len Huang and Dar-Ren Chen , 2004 " Watershed Segmentation for Breast tumor in 2-D Sonography" , Ultrasound in Med & Biol.,vol.30, No.5 , pp 625 – 632.

3. H.S.Sheshadri and A.Kandaswamy , 2004 , " Detection of Breast Cancer tumor based on Morphologicalwatershed Algorithm",www.icgst.com

4. Chen DR , Chang RF , H unag YL. 2000,"Breast cancer diagnosis using self organizing map for Sonography". Ultrasound Med Biol ., Vol.26, No.3 , pp 405-411

5.Chen DR , Chang RF , Hunag YL. 2000, " Texture analysis of breast tumors on sonograms", Semin ultrasound CT MR, Vol.24,No.4, pp 308-316

6. Rafel C. Gonzalez and Richard E.Woods, 2002, " Digital image processing" second edition, by Pearson Education Inc.

7.Haris K, Efstrafiadis SN,1998, "Hybrid image segmentation using watershed and fast region merging" , IEEE Trans Image Processing Vol. 8. No.12, pp 1684-1698.7. Horsch K, Giger ML,Venta LA, Vyborny CJ, 2001, "Automatic segmentation of breast lesions on ultrasound" Med Phys 2001, Vol 28. No 8, pp 1652 – 1659.

9. Sokol Petushi, Fernando U Garcia, Marian MHaber, Constantine katsinis and Aydin Toaeren, "Large –Scale Computations on histology images reveal grade-differentiating parameters for breast cancer", BMC Medical Imaging 2006, 6:14 doi:10.1186/1471-2342-14.

10. Tao Shi, david Seligson, Arie S Belldegrun, Aarno Palotie and Steve Horvath, " Tumor Classification by tissue microarray profiling: random forest clustering applied to reneal cell carcinoma", Modern Pathology (2005) 18,547-557.

11. Hussian Fatakdawala, jun Xu, Ajay Basavanhally, Gyan Bhanot, Shridar Ganesan , Michael Feldman, John E, Tomaszewski, and Anant Madabhushi, " Expectation Maximization driven Geodesic Active Contour with overlap Resolution (EMaGACOR): Application to Lymphocyte Segmentation on breast Cancer Histopathology", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL.-ISSUE.-, DECEMBER 2009.

12. Scarff RW, Torloni H: Histological typing of breast tumors. International histological classification of tumors. World health Organization 1968, 2(2):13-20.

13. C.Demir and B.Yener, "Automated cancer diagnosis based on histopathological images: a syatematic survey," Rensselaer polytechnic institute, tech. rep., 2005.

14. H.Soltanian-Zadeh and K.Jafari-Khouzani, "Multiwavelet grading of prostate pathological images", IEEE Trans. On Biomedical Engineering, vol 50, PP.607-704, 2003.

15. H.Jenog , T-Y Kim,H.-G Hwang, and H.-J Choi, "Comparison of thresholding methods for breast tumor cell segmentation, "in proc.of 7th Int. Workshop on Enterprise networking and computing in Healthcare Industry, 2005, PP 392-395.

16.T.K.Lau, and W.F.Bischof. Automated detection of breast tumors using the asymmetry approach. Comp. Biomed. Res.,24:273-295, 1991.

17.A.J.Mendez.,P.G.Tahocesb,M.J.Lado, M.Souto, J.L.Correa, and J.J. Vidal. Automatic detection of Breast Border and Nipple in Digital Mammograms. Computer Methods and Programs in Biomedicine, 49:253-262, 1996.

18. R.J.Ferrari, R.M.Rangayyan, J.E.L.Desautels, and A.F.Frere. Analysis of Asymmetry in Mammograms via Directional Filtering with Gabor Wavelets. IEEE Trans on Medical Imaging, 20(9):953-964, 2001.

19. M.Gudmundsson, E.A.El-Kwae, and M.R. Kabuka. Edge detection in Medical Images using a Genetic Algorithm. IEEE Transactions on Medical Imaging, 17(3):469-472,1998.

Author's Biography

Mrs. S. Thilagavathi received her Professional degree MCA., from IGNOU, New Delhi, and M.Phil degree from Vinayaga Mission University, salem in Computer Application Her area of Research includes Medical Image Processing and Genetic algorithm, Neural Network. She has presented 5 papers in National and InternationalConferences. She has 5 years of experience in teaching and currently working as Assistant Professor in MCA department at Hindusthan College of Engineering & Technology. Coimbatore.

Dr.M.S.Irfan Ahmed received his Professional degree MCA.,and M.Phil degree in Computer Application and he received his Professional Doctoral Degree Ph.D in Computer Science and Engineering from the Algappa University. He is currently working as a Director , Department of MCA in Hindusthan College of Engineering and Technology, Coimbatore. He is a research Guide of Anna University, Bharathiar University as well as Karpagam University in the field of Computer Application. He has published several papers in International Journals and Conferences. His research interests are in Trustworthy Computing , Networking and Datamining etc. He has 17 Years of teaching experience.