Proceedings of International Conference on Man-Machine Systems 2006
September 15-16 2006, Langkawi, Malaysia

Classification of Breast Lesions Using Artificial Neural Network

M.Y. Mashor[1] S. Esugasini[2] N.A. Mat Isa[2] N.H. Othman[3]

[1]Electronic & Biomedical Intelligent Systems (EBItS) Research Group, School of Mechatronic Engineering
Kolej Universiti Kejuruteraan Utara Malaysia, 02600 Jejawi, Arau, Perlis, MALAYSIA
E-mail: yusoff@kukum.edu.my

[2]Control and ELectronic Intelligent System (CELIS) Research Group, School of Electrical & Electronic Engineering,
Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, MALAYSIA.

[3]Pathology Department, School of Medical Science, Universiti Sains Malaysia
Medical Campus, 16150 Kubang Kerian, Kelantan, MALAYSIA.

Abstract

This paper presents a study on classification of breast lesions using artificial neural network. Thirteen morphological features have been extracted from breast lesion cells and used as the neural network inputs for the classification. Multilayered Perceptron Network trained using recursive prediction error algorithm was used to perform the classification task. Unlike the previous studies that only classify the lesion into benign and malignant, this study extends the breast lesions classification into four categories that are malignant, fibroadenoma, fibrocystic disease and other benign cells. Based on 1300 data samples, the proposed system gives good overall diagnostic performance. The system produces 92.78% accuracy, 99.03% sensitivity and 89.58% specificity, while keeping the false negative and positive to some considerable low values.

Keywords

Breast Cancer, Neural Network, Classification, MLP Network, RPE Algorithm.

1.0 Introduction

There are various techniques that were used to interpret the cancerous cells or the lesions suspected to be cancerous in the medical field. The conventional or manual ways of interpreting the cells are microscopic diagnosis by cytopathologists or based on mammogram images by oncologists. Mammography has high sensitivity in screening of breast cancer, however, it normally produces high rate of false positive prediction that will lead to large number of biopsies of benign lesions (Starita *et al*., 2000). The application of artificial intelligence in the medical field has revealed many computer aided diagnostic (CAD) systems to assist medical experts to produce faster and more accurate diagnosis for the increasing incidence of breast cancer cases.

Giger and Huo (1999) used artificial neural network (ANN) to develop a CAD that incorporates various computer-extracted image features from mammogram images to differentiate malignant from benign masses. The performance of the computer aid with 100% sensitivity was appreciable with a positive predictive value of 83%, which was 12% higher than an experienced mammographer. Wu *et al.* (1993) used Multi-layered Perceptron (MLP) network to differentiate between malignant and benign mammographic patterns based on radiographic features extracted by radiologists. The ability of neural network classification was then compared to classification by mammographers and the results showed that the capability of the neural network to do classification is better than that of a mammographer alone.

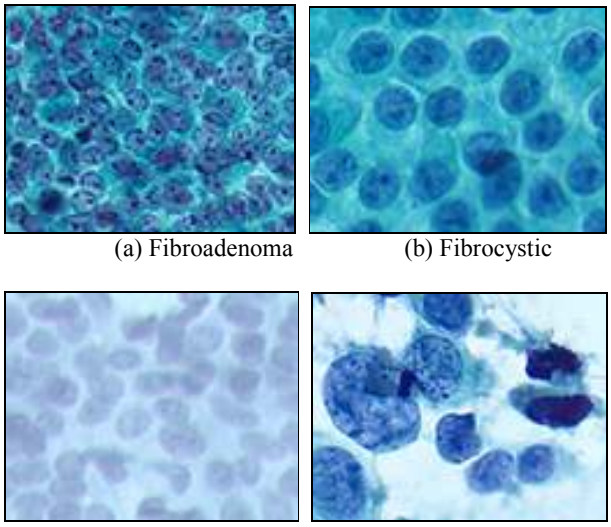
Other researches in neural networks implementation in cancer diagnosis have been done by Kok *et. al*. (1999), Mitra *et. al*. (2000) and Mashor *et. al* (2004) for cervical cancer and Yao & Liu (1999) and Kates *et. al*.(2000) for breast cancer. Yao & Liu (1999) defined two neural network approaches for breast cancer diagnosis, evolutionary and ensemble. The evolutionary approach was used to design compact neural networks automatically by evolving network architectures and weights, while the ensemble approach was aimed at tackling large problems that may not dealt with efficiently by a monolithic neural network. Kates *et. al* (2000) presented the potential contributions of neural network to a clinical decision support framework for the prediction of breast cancer therapy response.

Most of the previous studies of breast cancer diagnosis are based on the mammograms images. The limitation of these neural network based breast cancer diagnostic systems are that they only capable to classify the breast lesions into two categories which are benign and malignant tumours. The current study is based on smear slides of fine needle aspirate (FNA) cells of breast lesion. Multilayered Perceptron (MLP) Network is used to classify the breast lesion into four categories, which are malignant, fibroadenoma, fibrocystic disease and other benign cases.

2.0 Breast Lesion Classification Using Neural Network

MLP network with various training algorithms are tested to screen breast lesion cells. The inputs to the network are some features of the breast lesion cells. The images of breast lesion cells have been captured from smear slides using a computerised microscope. Thirteen features have been manually extracted from the computer images of breast lesion cells. The 13 features are cellularity, background, cohesiveness, cell in cluster, significant stromal component, clump thickness, nuclear membrane thickness, bare nuclei, normal nucleoli, mitosis, nucleus

stain, uniformity of cell and fragility. The neural networks receive these features as the inputs and classify the cells into four categories which are malignant, fibroadenoma, fibrocystic disease and other benign cells. Sample of fibroadenoma, fibrocystic disease, malignant and other benign cells are shown in Figure 1.



(a) Fibroadenoma (b) Fibrocystic

(c) Other benign (d) Malignant

Figure 1: Four categories of breast lesion cells

The features of the samples are extracted and interpreted into numerical data sets by experienced cytotechnologists with assistance and supervision by experienced pathologists. All the data were bypassed and checked thoroughly by minimum of three pathologists per sample. A total of 1300 data were collected from Hospital University Sains Malaysia and Penang General Hospital. First the captured images were revised the pathologists to determine the appropriateness of the breast lesion images. Then, thirteen features were extracted from the revised images with the pathologists' supervision, and the interpreted data were again compared with the respective medical reports of the cases. If there are any mismatches in the information, the cases were again referred and discussed with experienced pathologist and the medical report keeper before a decision is made. The data collection methodology was carefully revised and approved by the experienced pathologist before it was used to train and test the network. The 1300 data were divided into 800 training data and 500 testing data sets, where the distributions of the data are as shown in Table 1.

The MLP network has 13 input nodes that would accept the 13 features from the breast lesion cells. Then the networks will classify the lesion cells into four categories that are fibroadenoma, fibrocystic disease, other benign and malignant cells. Therefore, the network would have four output nodes to represent those four categories of breast lesion cells. In the current study, MLP network that was training using recursive prediction error (RPE) algorithm was used.

Table 1: Distribution of training and testing data sets

Category of Breast Lesion	Training data	Testing data
Fibroadenoma	240	150
Fibrocystic Disease	240	150
Other Benign	50	30
Malignant	270	170
Total	800	500

2.1 Multilayered Perceptron Network

Multilayered perceptron (MLP) network is a feed forward neural network with one or more hidden layers. Cybenko (1989) and Funahashi (1989) have proved that the MLP network is a general function approximator and one hidden layer networks will always be sufficient to approximate any continuous function up to certain accuracy. In the current study, the MLP network with a single hidden layer is used. With this simplification the network can be expressed as:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{kj}^2 F\left(\sum_{i=1}^{n_i} w_{ji}^1 v_i(t) + b_j^1 \right) \quad \text{for } 1 \le k \le n_o \qquad (1)$$

where w's, b's, n_h, n_i, n_o, $v_i(t)$ and $F(.)$ are the weights, thresholds, number of nodes in hidden layer, number of nodes in the input layer, number of nodes in output layer, inputs and an activation function respectively.

The activation function $F(.)$ is selected to be:

$$F(v(t)) = \frac{1}{1+e^{-v(t)}} \qquad (2)$$

The weights w's and threshold b's are unknown and should be selected to minimise the prediction errors defined as

$$\varepsilon_k(t) = y_k(t) - \hat{y}_k(t) \qquad (3)$$

where $y_k(t)$ represents the actual outputs and $\hat{y}_k(t)$ denotes the network outputs.

2.2 Recursive Prediction Error Algorithm

Recursive prediction error algorithm (RPE) was originally derived by Ljung and Soderstrom (1983) and modified by Chen et al. (1990) to train MLP networks. RPE algorithm is a Gauss-Newton type algorithm that will generally give better performance than a steepest descent type algorithm such as back propagation algorithm. In the present study, the RPE algorithm based on Chen et al. (1990) was used to train the MLP network to performce breast lesion diagnosis. The RPE algorithm modified by Chen et al. (1990) minimises the following cost function,

$$J(\hat{\Theta}) = \frac{1}{2N} \sum_{t=1}^{N} \varepsilon^T(t, \hat{\Theta}) \Lambda^{-1} \varepsilon(t, \hat{\Theta}) \qquad (4)$$

by updating the estimated parameter vector, $\hat{\Theta}$ (consists of w's and b's), recursively using Gauss-Newton algorithm:

$$\hat{\Theta}(t) = \hat{\Theta}(t-1) + \mathbf{P}(t)\Delta(t) \quad (5)$$

and

$$\Delta(t) = \alpha_m(t)\Delta(t-1) + \alpha_g(t)\psi(t)\varepsilon(t) \quad (6)$$

where $\varepsilon(t)$ and Λ are the prediction error and $m \times m$ symmetric positive definite matrix respectively, and m is the number of output nodes; and $\alpha_m(t)$ and $\alpha_g(t)$ are the momentum and learning rate respectively. $\alpha_m(t)$ and $\alpha_g(t)$ can be arbitrarily assigned to some values between 0 and 1 and the typical value of $\alpha_m(t)$ and $\alpha_g(t)$ are closed to 1 and 0 respectively.

$\psi(t)$ represents the gradient of the one step ahead predicted output with respect to the network parameters:

$$\psi(t,\Theta) = \left[\frac{d\hat{y}(t,\Theta)}{d\Theta}\right] \quad (7)$$

$\mathbf{P}(t)$ in equation (5) is updated recursively according to:

$$\mathbf{P}(t) = \frac{1}{\lambda(t)}\left[\mathbf{P}(t-1) - \mathbf{P}(t-1)\psi(t)\left(\lambda(t)\mathbf{I} + \psi^T(t)\mathbf{P}(t-1)\psi(t)\right)^{-1}\psi^T(t)\mathbf{P}(t-1)\right] \quad (8)$$

where $\lambda(t)$ is the forgetting factor, $0 < \lambda(t) < 1$, and normally been updated using the following scheme, Ljung and Soderstrom (1983):

$$\lambda(t) = \lambda_0\lambda(t-1) + (1-\lambda_0) \quad (9)$$

where λ_0 and the initial forgetting factor $\lambda(0)$ are the design values. Initial value of $\mathbf{P}(t)$ matrix, $\mathbf{P}(0)$ is normally set to $\alpha\mathbf{I}$ where $\mathbf{I}$ is the identity matrix and α is a constant, typically between 100 to 10000. Small value of α will cause slow learning however too large α may cause the estimated parameters do not converge properly. Hence, it should be selected to compromise between the two points, α = 1000 is adequate for most cases.

The gradient matrix $\psi(t)$ can be modified to accommodate the extra linear connections for one-hidden-layer MLP network model by differentiating equation (1) with respect to the parameters, θ_c, to yield:

$$\psi_k(t) = \begin{cases} v_j^1 & \text{if } \theta_c = w_{jk}^2 \quad 1 \le j \le n_h \\ v_j^1\left(1-v_j^1\right)w_{jk}^2 & \text{if } \theta_c = b_j^1 \quad 1 \le j \le n_h \\ v_j^1\left(1-v_j^1\right)w_{jk}^2 v_i^0 & \text{if } \theta_c = w_{ij}^1 \quad 1 \le j \le n_h,\ 1 \le i \le n \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The above gradient matrix is derived based on sigmoid function therefore, if other activation functions were used the matrix should be changed accordingly.

3.0 Results and Discussion

After some simulation analyses, the MLP network trained using RPE algorithm are found to be producing the acceptable diagnosis at hidden nodes of 20 and 300 training epochs. The number of input nodes and output nodes were fixed to 13 and 4 respectively, for the reasons mentioned in previous section. The training parameters for RPE were set as follows:

$\alpha_g = 0.05$; $\alpha_m = 0.85$; $\mathbf{P}(0) = 1000\mathbf{I}$

$\lambda_0 = 0.99$; $\lambda(0) = 0.95$

Using 800 training data and 500 testing data sets, the proposed system produce the diagnostic performance is summarised in Table 2.

Table 2: Diagnosis performance of proposed system

Analyses	**Train (%)**	**Test (%)**	**Overall (%)**
Accuracy	93.33	91.89	92.78
Sensitivity	100.00	97.47	99.03
Specificity	89.90	89.07	89.58
False Negative	0.00	2.59	1.00
False Positive	10.10	10.93	10.42

The proposed system produces high accuracy, sensitivity and specificity, with a very small false negative and a reasonable false positive. High specificity and small false negative suggesting that, the system only misclassify very small cancerous cases. The false positive rate could be reduced by changing the classification border but at the same time it would increase false negative which is more undesirable. In practice, a patient with false positive will go for further tests and normally reveals as a normal case.

4.0 Conclusion

MLP network trained using RPE algorithm has been proposed to classify breast lesion cells into fibroadenoma, fibrocystic disease, malignant and other benign cells. Thirteen morphological features were proposed to be used in this classification process. Based on the results in the previous section the system is capable to produce high accuracy, sensitivity and specificity with small false negative and considerable low false positive. Hence, MLP network trained using RPE algorithm shown to be capable to perform the breast cancer diagnosis with good accuracy.

Further study to improve the system performance by using other training algorithms and network architectures are in progress.

References

[1] Starita, A., S. La Manna, D. Majidi, D. Caramella and A. Cilotti (2000). Magnetic resonance in mammography: A quantitative evaluation tool of contrast-enhanced magnetic resonance of the breast. *Advances in Medical Signal and Information Processing, 2000*, *First International Conference on (IEE Conf. Publ. No.476),* pp. 200 – 205.

[2] Giger and Huo (1999) – missing

[3] Wu, Y., Giger, M.L., Doi, K., Vyborny, C.J., Schmidt, R.A., Metz, C.E. (1993). ANNs in mammography: Application to decision making in the diagnosis of breast cancer, *Radiology.* Pp. 81-87.

[4] Kok M.R., Schreiner-Kok P.G., Veen V.D., and Boon M.E., (1999), "Potentially difficult smears of women with squamous cell carcinoma pose fewer problems

when PAPNET is used for primary screening", *Cytopathology*, 10(5), pp 324-34.

[5] Mitra P., Mitra S. and Pal S. K., (2000): 'Staging of Cervical Cancer with Soft Computing', *IEEE Trans. on Biomedical Eng.*, **47**(7), pp. 934-940.

[6] Mashor, M. Y., Mat-Isa, N. A. & Othman, N. H. "Classification of Pap Smear Tests Using Neural Networks", *Proceedings of Kuala Lumpur International Conference on Biomedical Engineering (BioMed2004)*. pp. 21-24. Kuala Lumpur, Malaysia, 2-4 September, 2004.

[7] Yao and Liu (1999) - missing

[8] Kates, R., Harbeck, N., Schmitt, M. (2000). Prospects for clinical decision support in breast cancer based on neural network analysis of clinical survival data. *Proceedings of the 4th international Conference on Knowledge-Based Intelligent Systems & Allied Technologies*. Pp. 764-767.

[9] Cybenko, G., (1989), Approximations by superposition of a sigmoidal function, *Mathematics of Control, Signal and Systems*, **2**, 303-314.

[10] Funahashi, K., (1989), On the approximate realisation of continuous mappings by neural networks, *Neural Networks*, **2**, 183-192.

[11] Ljung, L. & Soderstrom, T. (1983). *Theory and practice of recursive identification*. Cambridge: MIT Press.

[12] Chen, S., Cowan, C.F.N., Billings, S.A., and Grant, P.M., (1990), A parallel recursive prediction error algorithm for training layered neural networks, *International Journal of Control*, **51** (6), 1215-1228.