Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 3, Issue, 8, pp.054-057, August, 2011

INTERNATIONAL JOURNAL
OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

AUTOMATIC MASS CLASSIFICATION OF BREAST CANCER USING ANN

Prof.Dr. P.K. Srimani*[1] and Smt. Parvati N Angadi[2]

[1]Director, R&D(CS),DSMIT,Bangalore
[2]Research Scholar, Rayalaseema University, Kurnool

ARTICLE INFO

Article History:
Received 27th May, 2011
Received in revised form
9th June, 2011
Accepted 28th June, 2011
Published online 5th August, 2011

Key words:
Artificial Neural Networks,
Multilayer Perceptron,
Backpropogation,
Levenberg Marquardt learning,
Breast Cancer.

ABSTRACT

Breast cancer is one of the leading causes of mortality among women, and the early diagnosis is of significant clinical importance. To perform classification of medical data, the neural network was trained and tested with breast cancer database by using feed forward neural network multilayer perceptron model and back propogation learning algorithm with Levenberg Marquardt learning and variable learning rate. The performance of the network was evaluated and the experimental results showed that by applying this neural network model, stable and accurate results could be achieved. The results obtained by using the neural network model predicted a higher degree of accuracy when compared with the existing results.

INTRODUCTION

Over the past 50 years, breast cancer has become the most commonly diagnosed cancer and it is the leading cause of cancer-related death among women. Experts say early detection is the best way to prevent the disease from spreading and improve the odds of survival. If detected in time, the cure rate is an astonishing 97 percent. Whilst the causes of breast cancer are not completely understood, modern research is making rapid progress in controlling the disease, through easier and accurate diagnosis and improved methods of treatment. X-ray mammography is currently considered as standard procedure for breast cancer diagnosis. However, retrospective studies have shown that radiologists can miss the detection of a significant proportion of abnormalities in addition to having high rates of false positives. The estimated sensitivity of radiologists in breast cancer screening is only about 75% (Bird, 1990) Double reading has been suggested to be an effective approach to improve the sensitivity. But it becomes costly because it requires twice as many radiologists' reading time. This cost will be quite problematic while considering the ongoing efforts to reduce the costs of the health care system. Cost effectiveness is one of the major requirements for a mass screening program to be successful. The ultimate diagnosis of all types of breast disease depends on a biopsy. In most cases the decision for a biopsy is based on mammography findings. Only 20% of current biopsy cases actually reveal cancer. The remainder is all benign cases, which underwent a potentially unnecessary surgical procedure. Preventing benign biopsies is the most important way to improve the efficacy of mammography screening, especially as screening becomes more widespread. This clearly demonstrates a need for efficient breast cancer detection and diagnosis techniques. Therefore, the present investigation is undertaken. In this paper, automatic mass classification into benign and malignant is presented based on the statistical and textural features extracted from mass from the breast region using ANN.

***Corresponding author:** profsrimanipk@gmail.com

SOME RELATED WORKS

A thorough survey of the literature pertaining to the topic revealed that, most of the works dealt with intelligent diagnostic systems specifically to provide 'second opinion' for pathologists in making diagnosis (Hatanaka *et al*., 2001, Santos Andre *et al.,*2002, Tsujii *et al*.,1999). Artificial neural network (ANN) was employed to classify between benign and malignant cases. Those approaches required breast features taken from mammogram images, as the input data for the ANN. It has been shown in (Burke, H. B *et al*., 1997, Seker *et al*., 2002, Land Jr *et al*., 2001) that in general, feedforward ANNs could produce the breast cancer diagnosis results almost favorably in comparision with those from human experts. The applicability of ANNs combined with image processing techniques to predict the stages of breast cancer has been carried out in (Polakowski *et al*.,1997, Schnorrenberg *et al*.,1997, Verma *et al.,*2001). The system proposed by Polakowski *et al*., managed to achieve 92% of sensitivity out of 272 cases. A detailed study of breast cancer classification based on morphological features of breast cells had been done by many researchers (Demir and Yener 2005, Lo *et al.,* 2003, Tozaki *et al.,* 2005, Wedegartner *et al.,* 2001). In another research, Abbas introduced evolutionary multi-objective approach to artificial neural networks (ANNs) for breast cancer diagnosis. This survey showed that some literature was available with regard to neural networks approach to cancer detection problem but, no literature is available with regard to the present approach.

ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANNs) are powerful tools for classification problems. An ANN can learn the classification task from a set of examples known as training set. Multi Layer Perceptron (MLP) is one of the most popular supervised learning ANN models which is frequently used for classification problems. Artificial Neural Networks consist of the large number of processing elements with their interconnections. ANN's are basically parallel computing systems similar to biological neural networks. They can be characterized by three components: nodes, weights (connection strength), an activation (transfer) function. ANN modeling is a nonlinear statistical technique; it can be used to solve problems that are not amenable to conventional statistical and mathematical methods. In the past few years, there has been constantly increasing interest in neural networks modeling in different fields of engineering. The basic unit in the artificial neural network is the node. Nodes are connected to each other by links known as synapses, associated with each synapse there is a weight factor. Usually neural networks are trained so that a particular set of inputs produces, as nearly as possible, a specific set of target outputs.

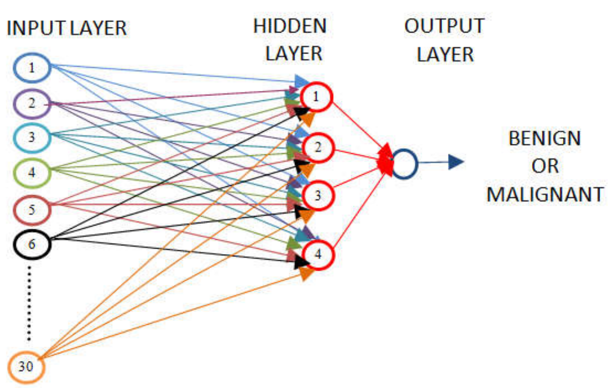

Fig. 1. General Architecture MLP

Multilayer perceptron (MLP) network models are one of the most popular ANN architectures used in most of the research applications in medicine, engineering and other applications. In MLP, the weighted sum of the inputs and bias term are passed to activation level through a transfer function to produce the output, and the units are arranged in a layered feed-forward topology called Feed forward neural Network (FFNN). Supervised learning is achieved by an iterative adjustment of the network connection weights in order to minimize an error function, computed over the training cases. The schematic representation of MLP with 30 input nodes, 4 hidden nodes and one output layer is given in Fig I. An ANN has 3 layers: input layer, hidden layer and output layer. The hidden layer vastly increases the learning power of the MLP. The output layer had one output node representing the classification of each cancer type as benign or malignant. The transfer or activation function of the network modifies the input to give the desired output. The choice of the number of hidden layers, hidden nodes and type of activation function plays an important role in model construction (White *et al.*, 1992, Bishop,1995). Back propagation error correction method is generally used for breast cancer survival prediction.

TRAINING THE MODEL

Once a network has been constructed for a particular application, that network is ready to be trained. To start this process the initial weights are chosen randomly. Then, the training, or learning, begins. The ANN has been trained by exposing it to sets of existing data (based on the follow up history of cancer patients) where the outcome is known. Multi-layer networks use a variety of learning techniques; the most popular is back – propagation algorithm. It is one of the most effective approaches to machine learning algorithm developed by David Rummelhart and Robert McLelland (1994). Information flows from the direction of the input layer towards the output layer. A network is trained rather than programmed. Learning in ANN's is typically accomplished with the help of examples. This is also called 'training' in ANNs because the learning is achieved by adjusting the connection weights in ANNs iteratively. The number of iterations of the training algorithm and the convergence time will vary depending on the weight initialization. After repeating this process for a sufficiently large number of training cycles the network will usually converge to some state where the error of the calculations is small. In this case one says that the network has learned a certain target function. Learning techniques are often divided into supervised, unsupervised and reinforcement learning. Nominal variables are used to represent the input values in the nodes of the input layer. Nominal variables may be two-state or many-state. A two-state nominal variable is easily represented as a numeric value by a transformation. Neural networks have facilities to convert both two-state and many-state nominal variables. It is well known that the performance of learning systems on the training data often does not reflect the performance on unknown data. This is due to the fact that the system often adapts well with training data to the observed particularities. Therefore, the training data should be randomly chosen from all available data. It should represent the typical data properties. If there is an initial bias in the training data, then one encounters performance problems for the test data later. In other words, the test data should not contain samples coming from the training data of the patients. It is a fact that patient data is individual based and hence it is difficult to generalize. By ignoring this fact, better results may be predicted but, may not be practically feasible.

MATERIAL AND METHODS

The material that was used in the present work was derived from the internet site of University of California at Irvine (UCI) Machine Learning Data Repository (Tsujii *et al*., 1999). The file downloaded contains medical data concerning breast cancer classification cases that were categorized by medical experts to malignant or benign. The file wdbc.data contains features that describe characteristics of the cell nuclei of a fine needle aspirate (FNA) of a breast mass (Burke *et al*.,1997). For each cell nuclei, 30 real-valued features (Seker, *et al*., 2002) are provided, which refer to three different values of each of the following diagnostic parameters:

1. Radius (mean of distances from centre to points on the perimeter)
2. Texture (standard deviation of grey-scale values)
3. Perimeter
4. Area
5. Smoothness (local variation in radius lengths)
6. Compactness (perimeter^2 / area - 1.0)
7. Concavity (severity of concave portions of the contour)
8. Concave points (number of concave portions of the contour)
9. Symmetry
10. Fractal dimension ("coastline approximation" - 1)

Three sets (mean, standard error and "worst") of values are provided for each parameter. Therefore, the file contains in total the measured values of 30 different parameters, for 569 cases of breast cancer. These 569 cases were classified by medical experts as 357 (62.75%) benign and 212 (37.25%) malignant. The result of medical classification to benign or malignant is provided as an extra feature in the downloaded data file with an ID code particularly for each evaluated case. In this experiment, the neural network is trained with Breast Cancer database by using feed forward neural network MLP model and back propagation algorithm with Levenberg Marquardt learning and variable learning rate. The input layer of the network consists of 30 neurons to represent each attribute as the cancer database consists of 30 attributes. The numbers of classes are 2; one Benign and another is Malignant. So, one neuron in the output layer is sufficient to represent these two classes. Neural networks with different training and testing samples were constructed and trained with the cancer dataset.

PERFORMANCE OF THE NETWORK

The various phases in the classification problems solved by neural network techniques were construction, training and testing. Construction and training of the neural network were explained in the previous section. The classification of the test data and the performance of the network are discussed in this section. The test data was given as the input to the trained network and the output of the network was calculated with the adjusted weights. Since we know the target output, the output of the network was compared with the target output to study the learning ability of the network for classifying the cancer data. The classification performance was assessed in terms of the Sensitivity, Specificity and Accuracy of the system. Sensitivity (SN) is the proportion of actual positives which were correctly identified and is mathematically expressed in equation 1. Specificity (SP) is the proportion of negatives which were correctly identified and is mathematically expressed in equation 2. Accuracy (AC) is the degree of closeness of the actual output to the theoretical output and is mathematically expressed in equation 3. Various groups of training and testing data were formed and mean square error (MSE) was computed as shown in Table-I. The Sensitivity, Specificity and Accuracy of the built and tested multi layer perceptron is shown in Table II .

$$SN=TP/TP+FN \qquad \dots (1)$$

$$SP=TN/TN+FP \qquad \dots (2)$$

$$AC=TN+TP/TN+TP+FN+FP \qquad \dots (3)$$

Where TP- True positive, TN- true negative, FP-false positive and FN- false negative.

Table I. Experimental Results for MSE.

Trial No.	Number of training samples	Number of testing samples	MSE Diagnosis (Benign)	MSE Diagnosis (Malignant)
1	57	57	0.0600883	0.06008825
2	114	114	0.060038	0.06003799
3	171	171	0.062967	0.06296697
4	171	228	0.0210627	0.02115003
5	228	285	0.0600259	0.06002592
6	228	341	0.0423115	0.0441581
7	285	398	0.0293238	0.02932381
8	285	455	0.0226639	0.02218522
9	341	512	0.0248956	0.02489558
10	341	28	0.0225391	0.02282713
11	398	57	0.0548452	0.05484516
12	398	85	0.0172735	0.01727351
13	455	142	0.0879255	0.08792547
14	455	142	0.0014129	0.00141136
15	512	142	0.0470604	0.04706038
16	512	142	0.0689827	0.06898265
Average values			0.04096222	0.041078325

Table 2. Experimental Results for Sensitivity, Specificity and Accuracy

Trial No.	Number of training samples	Number of testing samples	Sensitivity SN (In %)	Specificity SP (in %)	Accuracy AC (in %)
1	512	57	95	95	95
2	455	114	89	97	91
3	398	171	95	94	95
4	341	228	99	96	97
5	285	285	96	94	95
6	228	341	96	90	94
7	171	398	95	92	94
8	114	455	94	93	94
9	57	512	94	92	93
10	512	28	95	89	93
11	455	57	100	100	100
12	398	85	98	96	98
13	341	142	98	96	97
14	285	142	96	96	96
15	228	142	97	89	94
16	171	142	99	94	97
Average values			96	93.9	95.2

Table3. Comparison of existing and present experimental results

Parameters	Existing results (In %)	Present Experimental Results (In %)
Sensitivity	92	96
Specificity	90	93.9
Accuracy	88.9	95.2

Table II clearly shows that accurate and stable results could be obtained only for large number of testing samples. It is also evident from Table III that the present experimental results are more accurate when compared to the existing results.

CONCLUSION

To classify the medical data set, a neural network approach was adopted. The experiment was conducted with the cancer dataset by considering the MLP. Backpropogation algorithm with Levenberg Marquardt learning and variable learning rate was used to train the network. To analyze the performance of the network, various test data were given as input to the network. The results showed that the MLP neural network's classification efficiency is cent percent for 455 training samples and 57 testing samples. Our experimental results using neural networks technique proved to be more efficient and accurate for the classification task when compared to the existing results (Table III).

FUTURE SCOPE

This paper forms the basis for further implementation and computation of different network classifiers and the other training algorithms. The work in this direction is in progress. This method can be used for the diagnosis of other diseases also.

ACKNOWLEDGEMENT

We thank Dr. William .H Wolberg at the University of Wisconsin for supporting us with the breast cancer dataset which we have used in our experiments.

REFERENCES

Bird R.E,1990. Professional quality assurance for mammography screening programs, Journal of Radiology, Vol. 175, pp. 587-605.

Bishop C. M, 1995, *Neural Networks for pattern recognition*, Oxford University Press, New York.

Burke H. B., Goodman P. H, Rosen D. B, Henson D. E, Weinstein J. N, Harral Jr. F. E, Marks J. R, Winchester D. P. and Bostwick D. G. 1997, Artificial Neural networks Improve the Accuracy of Cancer Survival Prediction. *Cancer*. pp. 857 – 862.

Hatanaka Y, Hara T, Fujita H, Kasai S, Endo T and Iwase T, 2001, Development of An Automated Method for Detecting Mammographic Masses with A Partial Loss of Region, IEEE Transaction on Medical Imaging. Vol. 20. No. 12. pp. 1209 – 1214.

Land W. H. Jr, Master T, Lo. J. Y. and McKee D. W. 2001, Application of Evolutionary Computation and Neural Network Hybrids for Breast Cancer Classification Using Mammogram and History Data, Proceedings of IEEE Conference on Evolutionary Computation, pp. 1147 – 1154.

Polakowski W. E, Cournoyer D. A, Rogers S. K, DeSimio M. P, Ruck D.W, Hoffmeister J. W and Raines R.A. 1997, Computer-Aided Breast Cancer Detection and Diagnosis of Masses Using Difference of Gaussians and Derivative-Based Feature Saliency, IEEE Transaction on Medical Imaging. Vol. 5. pp. 811 – 819.

Santos Andre T. C. S. and Da Silva A. C. R. 2002, Segmentation of Digitized Mammograms Using Self-Organizing Maps in A Breast Cancer Computer Aided Diagnosis System, Neural Network. Vol. 5. pp. 256 – 266.

Schnorrenberg F, Pattichis C. S, Kyriacou K. C. and Schizas C. N. 1997, Computer Aided Detection of Breast Cancer Nuclei, IEEE Transactions on Information Technology in Biomedicine, Vol. 1. No. 2, pp. 128 – 140.

Seker H., Odetayu M. O., Petrovic D., Naguib R. N. G, Bartoli C., Alasio L, Lakshmi M. S and Sherbet G. V. 2002, *A* Hybrid System for Nodal Involvement Assessment in Breast Cancer Patients, Proceedings of the Second Joint on 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference. Vol. 2. pp. 1049 – 1050.

Simon Haykin 2001, Neural Networks – A Comprehensive Foundation. Pearson Education

Tsujii O, Freedman M.T and Mun S. K. 1999, Classification of Microcalcifications in Digital Mammograms Using Trend-Oriented Radial Basis Function neural Network, Pattern Recognition. Vol. 4. pp. 891 – 903.

Verma B and Zakos J, 2001, A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques, IEEE Transactions on Information Technology in Biomedicine. pp. 46 – 54.

White H and Gallant A.R, 1992, On learning the derivatives of an unknown mapping with multiplayer feed forward networks, Neural Networks, vol.5, pp.129-138.

White H, 1992, Artificial Neural Networks. Approximation and learning theory, Blackwell, Cambridge, MA, http://en.wikipedia.org/wiki/ Artificial neural network# Multi-layer_perceptron

****