Iran. J. Radiat. Res., 2004; 1(4): 217-228

Comparison of logistic regression and neural network models in predicting the outcome of biopsy in breast cancer from MRI findings

P. Abdolmaleki[1*], M. Yarmohammadi[2], M. Gity[3]

[1]*Dept. of Biophysics, Tarbiat Modarres University, Tehran, Iran*
[2]*Dept. of Statistic, Payam Noor University, Tehran, Iran*
[3]*Dept. of Radiology, Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran*

ABSTRACT

Background: We designed an algorithmic model based on the logistic regression analysis and a non-algorithmic model based on the Artificial Neural Network (ANN).
Materials and methods: The ability of these models was compared together in clinical application to differentiate malignant from benign breast tumors in a study group of 161 patients' records. Each patient's record consisted of 6 subjective features extracted from MRI appearance. These findings were encoded as features for an ANN as well as a logistic regression model (LRM) to predict biopsy outcome. After both models had been trained perfectly on samples (n=100), the validation samples (n=61) were presented to the trained network as well as the established LRMs. Finally, the diagnostic performance of models were compared to that of the radiologist in terms of sensitivity, specificity and accuracy, using receiver operating characteristic curve (ROC) analysis.
Results: The average output of the ANN yielded a perfect sensitivity (98%) and high accuracy (90%) similar to that one of an expert radiologist (96% and 92%) while specificity was smaller than that (67% verses 80%). The output of the LRM using significant features showed improvement in specificity from 60% for the LRM using all features to 93% for the reduced logistic regression model, keeping the accuracy around 90%.
Conclusion: Results show that ANN and LRM prove the relationship between extracted morphological features and biopsy results. Using statistically significant variables reduced LRM outperformed of ANN with remarkable specificity while keeping high sensitivity is achieved.
Iran. J. Radiat. Res., 2004; 1(4): 217-228

Key words: *Breast cancer, neural networks, logistic regression model, ROC curves.*

INTRODUCTION

There is an ongoing effort by radiologists to predict the biopsy results by using Artificial Neural Network (ANN). ANN, as a well-established computer aided in diagnosis (CAD) system, is a computer algorithm capable of learning important relationship from a set of data and applying this knowledge to evaluate new cases. This method has been intensively used in breast evaluation, using different high sensitive algorithm (Vomweg *et al.* 2003, Biganzoli *et al.* 2003). ANN has two basic elements: processing elements and weighted connections. Collection of processing elements are defined as different layers including an input, one or more hidden layers, and an output layer. The connection weights store the information in form of weight matrices (Wasserman 1989). The neural network learning procedure determines, in turn, the value of the connection weights.

The outcome of biopsy commonly confirms the presence or absence of the malignancy and therefore is a binary outcome. A commonly used

[*] **Corresponding author:**
Dr. P. Abdolmaleki, Dept. of Biophysics, Faculty of Science, Tarbiat Modarres University, Tehran, Iran.
P.O. Box 14115-175, Fax: +98 21 8009730
E-mail: parviz@modares.ac.ir

statistical logistic regression model can evaluate this binary outcome, which is a useful method for discrimination. Logistic regression analysis is a technique with sufficient capability for separating distinct sets, when the dependent variable shows dichotomy, and the independent variables are continuous and/or discrete. The distinction is performed through establishing the discrimination rules. The rules will be estimated during the training procedure and can be used to allocate the new cases into the previously defined classes (Hosmer *et al.* 1989).

Both methods have been individually applied in breast cancer diagnosis using subjective impression of different features based on defined criteria (Tzacheva *et al.* 2003, Degenhard *et al.* 2002, Abdolmaleki *et al.* 2001). However, despite impressive results for each of them, a few works focused on the comparison between the advantages and/or disadvantages of the two models. In the present study, we established an ANN and a logistic regression model to take time intensity curve patterns and morphological findings from MR imaging to predict the outcome of biopsy or surgery.

Our objectives in this study were: (1) To compare the diagnostic performance of both methods in distinction between malignance and benign patterns, (2) To reduce the number of benign cases sent for biopsy using the best model as a supportive tool, and (3) To validate the capability of each model to recognize new cases as an expert system.

MATERIALS AND METHODS

Our goal was to compare the logistic discriminant analysis with ANN using the data collected in a study designed to predict the malignancy of breast cancer on the basis of radiological features that had been extracted from MRI appearance. Our study group consists of 161 consecutive patients (age 15-79 years; mean age, 51.2 years) with histopathologically proof. The patient group included 126 malignant lesions and 35 benign entities. Most of the malignant cases were invasive carcinoma (n=114), with the majority ductal carcinoma (n=107), while most of the benign lesions were fibroadenoma (n=19). Table 1 summaries the distribution of lesions at histopathologic analysis.

Table 1 : Demonstrating the distribution of lesions at histopathologic analysis.

Histopathologic diagnosis	No. of lesions
Malignant (n=126)	
Invasive carcinoma (n=118)	
Ductal	105
Mucinous	4
Lobular	2
Medullary	2
Squamous	5
Lymphoma	3
Malignant phyllodes tumor	5
Benign (n=35)	
Fibroadenoma	19
Fibrocytic disease	7
Fat necrosis	1
Benign phyllodes tumor	3
Intraductal papilloma	3
Granulomatous mastitis	2

Data acquisition

MR Imaging

For taking the images we used a Signa 1.5 Tesla unit (GE Medical Systems, Milwaukee, WI) in the first 123 patients. Patients underwent MR Imaging in the prone position using a single 5-inch circular general-purpose surface coil. Initial sagittal or axial T1-weighted spin-echo images (T1W) were performed at 400/16 (repetition time msec/ echo time msec), and axial or sagittal T2-weighted fast spin-echo images (T2W) with or without fat suppression for tumor localization were performed at 3,000/108. Other MR parameters used were a 20 cm field of view, 5 mm section thickness, and 256×192 (T1-weighted) or 256 × 256 (T2-weighted) matrix. Following the rapid administration of gadopentetate dimeglumine (Magnevist; Schering, Berlin, Germany) at a dose of 0.1 ml/kg body weight over 10-15 seconds the dynamic study was performed using a fast radio-frequency spoiled gradient-recalled-echo

(SPGR) sequence (11.4/3.3; flip angle, 35°; matrix, 256×192; section thickness, 5 mm; gap, 1.0-2.5 mm).

In the other 38 patients MR imaging was done on a Siemens Magnetom 1.5 Tesla system equipped with a dedicated double breast coil. A modified protocol was used with initial gadolinium enhanced coronal three-dimensional fast low angle shot (3D-FLASH) images, TR/TE 20/6.0 for localization, followed by 2D-FLASH TR/TE 60/5.0 with administration of gadolinium by automatic injector at a rate of 1 ml/second for dynamic study was used. Images were taken at 15-second intervals up to 5 minutes. In most cases maximum intensity projection (MIP) images, TR/TE, 40/6, and postprocessed subtraction images were obtained to improve the detection of contrast-enhancing lesions. All lesions were histologically confirmed after biopsy or surgical excision. We followed similar criteria for either imaging or feature categorization. An expert radiologist read the MR images and graded his finding on the following features: size, shape, lesion margin, enhancement homogeneity, time-intensity curve type, as well as other associated features like internal septations, duct-like enhancement, peripheral enhancement, breach of prepectoral fat plane, satellite nodules, and enhancement and/or retraction of overlying skin. The morphological features including lesion margin, enhancement homogeneity, peripheral enhancement, and shape were ranked using a five-scale categorization with increasing likelihood of malignancy. The presence of associated features was ranked on a scale of 0-8 with increasing likelihood of malignancy. In the case of more than one associated feature, the one with the highest rank was considered. Data acquisition of the dynamic study was done in the time of the injection. Then the images were called back one by one and a free size ROI was drawn in the most enhancing part of the lesion (figure 1). The obtained time-intensity values were used to generate the time intensity curve. Figure 2 shows a schematic diagram of four different types of the time intensity curves, which has been observed in this study. Type A and B had similarity with malignant cases while type C and D had similarity with benign cases. However there was some overlap between these patterns in some cases. The classification for the curve type has been previously reported (Buadu *et al.* 1996).

Figure 1. Free size regions of interest ROI were drawn in the most enhancing part of the lesion.

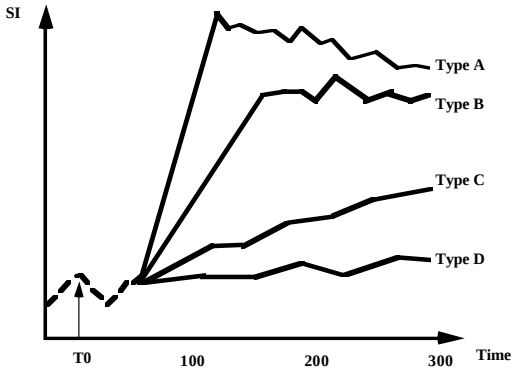

Figure 2. Schematic diagram showing time intensity curve classification.

In order to evaluate the capability of the established models to perform as an expert system and to learn the particular benign and malignant patterns presented in the training samples, we initially used the all-available data to train the ANN and to extract the estimation function for logistic regression model. The same database was then presented to both models to validate them after models had been trained. The neural network's output yielded a perfect accuracy (100%), demonstrating that the neural network learned perfectly all of the presented

patterns and was capable of recognizing all of them correctly. The LRM output showed a comparable accuracy (97%) demonstrating that the LRM was also capable to recognize most of the original cases.

To determine the performance of the established models in practical usage we divided the database into two separate database (a) the training samples comprising 100 patient records (20 benign, 80 malignant) and (b) the validation samples comprising 61 patient records (15 benign, 46 malignant).

Initially, using the patient's records in training sample the models trained by adjusting the weight values for interconnection links for the ANN and estimating the parameters needed to establish the classification rules for logistic regression model. Table 2 summarizes the radiologic features used as input into the models during the training and validation procedures.

Table 2. Coding of the evaluated parameters of MR images of 161 patients, which used as input into the models during the training and validation procedures.

Radiological Features	Findings	Code
Mass Size	No mass	0
	Mass	Size(mm)
Mass Shape	No mass	0
	Round	1
	Oval	2
	Lobulated	3
	Irregular	4
Mass Margins	No mass	0
	Well-defined	1
	Microlobulated	2
	Ill-defined	3
	Spiculated	4
Homogeneity	No mass	0
	No mass	0
	Homogeneous	1
	Slightly inhomogeneous	2
	Inhomogeneous	3
	Markedly inhomogeneous	4
Associated features	None	0
	Internal septations	1
	Intracystic mass	2
	Cystic spaces	3
	Skin and/or nipple enhance	4
	Satellite nodules	5
	Ductlike enhancement	6
	Nipple retraction	7
	Peripheral enhancement	8
	Axillary adenopathy	9
Time intensity curve type	Type D	1
	Type C	2
	Type B	3
	Type A	4

Then, the patients record in validation sample (n=61) was used to evaluate the generalizing ability of the established models separately. The best performance of the established models was compared with the reader in terms of accuracy, sensitivity, specificity, false positive, false negative, and the misclassification rate (where misclassification rate is defined as the ratio of the number of misclassified cases including those without a definite diagnosis to the total number of cases expressed as a percentage). In summary, the following stages were performed and compared:

1- Using data extracted by radiologist with considerable experience in breast MR image interpretation, both models established on features derived from 100 patients.
2- The established models were tested on 61 new cases.
3- To evaluate the performances of both models for clinical assessment, the best obtained outputs for each model compared with the participated radiologist in terms of accuracy, sensitivity, and specificity using ROC analysis.

Neural network structure

The neural network, which was employed in this study, had three layers. The first layer consisted of 6 input elements, each of which corresponded to the subjective data extracted from MR images as well as time-intensity curves type; the second layer, the hidden one, had 5 nodes and finally the output layer with 1 elements, which represented 1 for malignant and 0 for benign lesions. In order to determine the best optimized structure for the neural network, we simulated a large number of neural networks by varying the number of hidden nodes, iterations and learning rates. In all the simulations the Sum Square Error (SSE) was used as an index of the learning efficiency of the network during the training process. The details of the ANN simulations have been already reported (Abdolmaleki *et al.* 1997).

Finally, after the network had been trained perfectly the testing set was presented to the trained network giving a diagnostic output vector in the range of (0-1). Our network was trained perfectly over 100,000 iterations in each learning process within one hour on an IBM compatible personal computer (Pentium III 800 MHz). The software used to construct the neural network was written locally in MATLAB programming language.

Logistic regression models

We used logistic regression model as a classifier to predict the outcome of biopsy in breast cancer. The training and validation samples were used to build and validate the logistic regression model, respectively. Briefly, the logistic regression analysis was a statistical technique through which to examine the relationship between a dependent variable (result of biopsy) and a set of independent variables (radiological features). Then the independent variables, which could provide the best prediction, will be selected. This approach is commonly applied to predict membership in two groups using a set of predictors. Suppose we have two populations with different top probabilities. Using the cases presented in the training samples as well as the top probability the posterior probabilities for each group was obtained. Then, the cases presented in the validation sample were separated based on the obtained posterior probability associated with variables. The simplest optimizing method of discrimination was to maximize the posterior probability of correct allocation. To obtain the posterior probability the logit coefficients could be estimated using the Maximum Likelihood Estimation (Hosmer and Lemeshow 1989). Allocation of new cases can be performed using logit function, which could be obtained using the natural logarithm of the ratio of the calculated posterior probabilities. If the outcome of the logit function is positive (with the assumption of equal prior probabilities) the individual is allocated to class one (benign group). On the other hand, if the outcome is negative, the case is allocated to class two (malignant group). In the

present study, we established two logistic regression models. The first model, named full model, was using 6 variables including: size, shape, lesion margin, enhancement homogeneity, time-intensity curve type and other associated features. The second model, named reduced model, was using 3 variables from which lesion margin and time-intensity curve types were significant at the level of 0.05 using wald statistic. The last variable, enhancement homogeneity, was not statistically significant (p-value 0.08); however, its p-value was noticeable. The wald statistic is the square of the ratio of the unstandardized logit coefficients to its standard error, which has a chi-square distribution (Hosmer and Lemeshow 1989). We used the stepwise procedure of PROC LOGISTIC in SAS statistical package (SAS Institute, Inc., Cary, North Carolina) to establish the logistic regression models.

Performance evaluation

The commonly used ROC analysis was chosen to evaluate the predictive accuracy of neural network approach, logistic regression model and radiologist (Metz 1989). After the network had been trained perfectly the testing set (n=51) was presented to the trained network giving a diagnostic output vector in the range of 0-1. In the same way after the establishment of the logistic regression models the testing set (n=51) was presented to the models giving two posterior probabilities; one posterior probability for class one (benign group) and the other posterior probability for class two (malignant group). Each of these obtained probabilities, which are higher; the case will be allocated to the related classes. The outputs of the testing set were then analyzed to determine the true-positive and the false-positive fractions for each models including the ANN, full model as well as the reduced model. These data were then used for plotting the ROC curves. The area under the ROC curve (A_z) was used to compare the performance of ANN, full model, reduced model as well as the radiologist participating in the testing procedure (Metz 1986, 1989). In this regard the higher ROC areas indicating the better performance of the models.

To evaluate the performance of the observer, an expert radiologist was asked to read the MR images and grade his overall impression into one of the five categories with increasing likelihood of malignancy; 1= benign, 2= probably benign, 3= indeterminate, 4= probably malignant, 5= malignant. Similarly, to evaluate the performance of the neural network, the network output was classified into five categories; output in range of (0-0.2)= benign, (0.2- 0.4)= probably benign, (0.4-0.6)=equivocal, (0.6-0.8)= probably malignant and output in range of (0.8-1)= malignant. In the same way, to evaluate the performance of the established logistic regression models (Full model and reduced model) the obtained posterior probability for class two (malignant group) was considered and its value was then classified into five categories; posterior probability in range of (0-0.2)= benign, (0.2 - 0.4) = probably benign, (0.4 - 0.6) = equivocal, (0.6-0.8)= probably malignant and posterior probability in range of (0.8-1)= malignant.

RESULTS

Radiologist performance

An experienced radiologist read the images and classified them into benign and malignant groups using a five-scale category with increasing likelihood of malignancy. The statistical results of sensitivity, specificity and accuracy obtained were 96%, 80% and 92% respectively.

Neural network performance

The output of neural network on validation samples (n=61) showed a correct classification (45 of 46 of the patients with malignant breast cancers and 10 of 15 with benign entity). The average results of sensitivity, specificity and accuracy of 98%, 67%, and 90% obtained for the ANN were comparable to the results obtained for the participating expert radiologist: 96%,

80%, and 92%. However the results show that the radiologists with high level of experience are more specific than ANN in determining the benign cases.

Logistic regression model performance

The estimated logistic regression parameters were obtained from the training sample. Table 3 shows the maximum likelihood estimates of the parameters, standard errors, wald statistic and p-values of the logistic regression model. Taking into consideration all available variables, a logistic regression model, named as full model, established. Using the following allocation rule, a new case will be allocated to the malignant class if:

$$Rule1 = (-13.9360 - 0.0425 * Size + 0.6047 * Shape + 1.4392 * Margin + 1.0300 * Homogeneity + 2.2556 * Curve\ types + 0.3122 * Associated\ features) \rangle 0$$

$$Rule2 = -13.9360 + 1.4392 * Margin + 1.0300 * Homogeneity + 2.2556 * Curve\ types$$

Table 3. Indicating the maximum likelihood estimates of the parameters, standard errors, wald statistic and p-values of the logistic regression models fitted to the training sample.

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCPT	-13.9360	3.8230	13.2885	0.0003
Size	-0.0425	0.1668	0.0648	0.7990
Shape	0.6047	0.4360	1.9231	0.1655
Margin	1.4392	0.6839	4.4286	0.0353*
Homogeneity	1.0300	0.5919	3.0279	0.0818
Curve types	2.2556	0.6966	10.4852	0.0012*
Associated features	0.3122	0.1903	2.6926	0.1008

*significant at level of 0.05

The small p-values obtained for lesion margin and time-intensity curve types indicated that they are most significant predictor of malignancy in the model and the remained parameters including size and shape of tumor as well as associated features are not significant at level of 0.05. Also, the p-value obtained for homogeneity is not statistically significant but it was noticeable (p=0.0818). Therefore, the reduced model has been established by a minor modification of the full model. Using the significant variables and homogeneity, the allocation rule is modified into the following form.

The performance of the logistic regression models using rule 1 and rule 2 were evaluated. The best performance of the established models was then compared with the reader in terms of accuracy, sensitivity, specificity, false positive fraction, false negative fraction, misclassification rate and correlation with pathology (table 4, figure 3).

We also applied ROC analysis as a measure of the discriminating ability of a model, with higher areas indicating better predictive ability to compare the performance of the established models. Using the best results obtained for the ANN, full model, reduced model as well as the radiologist ROC analysis were performed (figure 4). The obtained areas under the receiver operating characteristic curves (Az) were presented in table 4.

Table 4. Comparative performance of the participating radiologist, neural network, logistic regression full model and logistic regression reduced model on validation sample (n=61).

Parameter	Radiologist	ANN	Logistic Discriminant Analysis	
			Full Model	Reduced Model
Sensitivity (%)	96	98	96	89
Specificity (%)	80	67	60	93
Accuracy (%)	92	90	87	90
Misclassified rate (%)	8	10	13	8
Correlation* with pathology	0.81	0.72	0.62	0.82
P_value	0.001	0.001	0.001	0.001
Area under the ROC curve(A_z)	0.9521±0.0294	0.9225±0.0561	0.9243±0.0393	0.9448±0.0357

*The measures of association of model's output and radiologist confirmed relations.

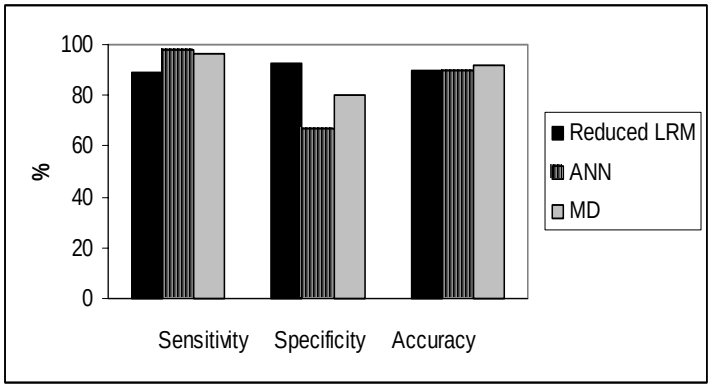

Figure 3. Comparative histogram of the average sensitivity, specificity and accuracy obtained for the participating radiologist, artificial neural network and reduced logistic regression model.

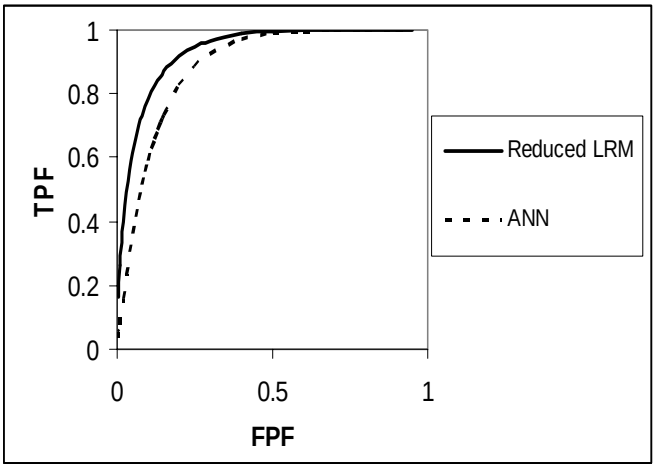

Figure 4. Resulting ROC curve comparing the diagnostic performance of the best results obtained for Artificial Neural Network as well as the reduced LRM.

DISCUSSION

In this study, we designed two algorithmic models based on the logistic regression analysis and a non-algorithmic model based on the artificial neural network. The ability of these models to differentiate malignant from benign tumors were compared among a group of 161 patients with approved breast lesions. Our main goal was to investigate which model obtains more reasonable specificity while keeping high sensitivity. By doing so, we hope to decrease the number of cases sent to the biopsy; especially in a significant fraction of patients who are going under the biopsy procedure for apparently benign lesions.

Using the guidelines for features selection from the previous literatures, the parameters were evaluated by a participating radiologist with a high level of experience. The extracted data was later presented to the established neural network. The average output of the ANN yielded a perfect sensitivity (98%) and high accuracy (90%) similar to the one obtained by the radiologist (96% and 92%). In contrast, the specificity obtained by ANN is clearly smaller than the specificity reported by our radiologist (67% versus 80%). This finding demonstrates a consistent high sensitivity with a moderate specificity for the ANN in differentiating between benign and malignant breast tumors. The moderate to low specificity values obtained for the ANN may appear as a limitation of the ANN. However, this might be related to the existence of a considerable degree of overlap between the enhancement patterns of malignant and benign lesions in MR imaging (Heywang Kobrunner 1994, Weinreb and Newstead 1995). On the other hand, ANN cannot eliminate the existing overlap due to the lack of a comprehensive and balanced database. Similarly, the specificity decreased by 60% for the logistic regression model when we used all extracted features from MR imaging (full model). This means that when the input is exactly similar for both models, the ANN performed better than the logistic regression model, as far as the specificity is concerned. The higher level of performance of the ANN may be related to the unique ability of neural network in making associations among too many nonlinear and dependent parameters by addressing them as proportional weights. For example, if an input parameter, such as mass shape, has a high correlation (0.60, p=0.001) with the output (result of pathology used as target for training), its neural network connection weights will be set higher than the others (26.12 ± 10.42, mean ± standard deviation). Similarly, if an input parameter such as associated features has a medium correlation (0.34, p=0.003), its weights will be set to lower values (0.63±0.38). However, even in the case of very complex input parameters, like mass size with no detectable correlation (-0.01, p=0.878), the independent relationship with the remaining parameters will be represented by a non-zero connection weight (0.08±0.04). These weights, which are adjusted by training procedure, are important for neural network because it is addressing the importance of each input elements for internal calculation on testing procedure. This provides little help for radiologists who want to clarify the relative prognostic importance of each feature. So, although the ANN may work as an excellent predictor of malignancy, it may not be able to explain which findings are more relevant in reaching the diagnosis. This can be pointed out as another limitation for the ANN. In contrast, the logit coefficient obtained from wald test in logistic regression model is somehow signifying the importance of any feature in making differentiation between benign and malignant breast tumor. Results obtained from wald test (table 3) supported the previous publication which has reported the high correlation of curve type, tumor margin and homogeneity with the results of biopsy (Buadu *et al.* 1996).

Finally, using a stepwise logistic regression procedure, we removed the variables with the largest wald test p-value from the full model. The large p-value obtained for the size (0.79) and the shape (0.16) indicated that these variables were the least significant predictor of malignancy. Therefore, the size and shape values

were then excluded from further proceeding. In contrast, the curve types had the smallest p-value indicating that it was the most significant indicator of malignancy. The tumor margin followed the curve types with a p-value of (0.03). Although the p-value obtained for the homogeneity was not statistically significant at the level of 0.05, it seemed to be noticeable (0.08). Therefore, the homogeneity was also considered in the model, as well. The output of the reduced model showed a very sharp improvement (93%) in the specificity in comparison with one obtained from the full logistic regression model (60%); while the accuracy remained about 90% for both. However, the remarkable specificity of the reduced model was obtained in the cost of a decrease in the sensitivity (from 96% for the full logistic regression model to 89% for the reduced one). Nevertheless, reduced model could still be considered as highly sensitive. In addition, the correct prediction of 14 out of 15 cases with benign entity as well as correct prediction of 41 out of 46 cases with proved malignancy demonstrated the high performance of the reduced logistic regression model. These results showed that if we use the most significant features in logistic regression model, its performance would be even better than the established back propagation neural network.

Another limitation of this study was that we just made a comparison between one training paradigm (back propagation training algorithm, which is the most popular training algorithm in medical assessments) with the logistic regression analysis in a specific clinical task. Since there is many artificial neural network training paradigms which we did not check and would do better discrimination, It needs more research to make comprehensive conclusion with greater performance in clinical applications.

Previous reports suggested that the accuracy, sensitivity and specificity of each diagnostic procedure was strongly dependent on the prevalence of patients' population. Therefore, the obtained data by ANN, logistic regression models and participated radiologist may not show the exact performance of them (Abdolmaleki *et al.* 1997). To justify this point the participating radiologist, we used ROC analysis to evaluate the performance of all models. By introducing a relative ROC area (Az) of 0.94 for the reduced model compared to 0.92 and 0.84 obtained by radiologist and full model respectively, the ROC analysis supported and enforced our results.

In conclusion, we established a non-algorithmic model based on the back propagation neural network and two algorithmic models based on the logistic regression analysis to differentiate malignant from benign breast tumors. Our results showed that our network and logistic regression models learned similar relationships between extracted morphological features and biopsy results. However, the results of this study suggested that the diagnostic performance of ANN is better than the logistic regression model (full model) when all the input and/or variables are similar. On the other hand, using statistically significant variables (reduced model), the logistic regression model had the best performance by preparing a remarkable specificity while keeping high sensitivity.

Appendix:

Theory of Logistic regression analysis:

Logistic regression is a statistical model for analysis of the relationship between an observed proportion (binary outcome) y and a vector $\mathbf{X}' = [X_1, X_2, ..., X_p]$ of regressor variables which are continuous, categorical or both for each of N individuals. The logistic regression model relates y to $\mathbf{X}$ in assuming that

$$\Pr(y = 1|\mathbf{X}) = \Lambda(\beta_0 + \sum_{i=1}^{p} \beta_i x_i) \quad (1)$$

where $\Lambda(u) = (1+\exp(-u))^{-1}$ denotes the logistic function. The logistic model as a non-linear regression model is a special case of a generalized linear model, i.e. $E(y|\mathbf{X}) = \Pi(\mathbf{X}, \boldsymbol{\beta})$

and $\mathrm{Var}(y|X) = \Pi(\mathbf{X},\boldsymbol{\beta})(1-\Pi(\mathbf{X},\boldsymbol{\beta}))$, using the logit link

$$\Pi(\mathbf{X},\boldsymbol{\beta}) = \Lambda(\beta_0 + \sum_{i=1}^{p}\beta_i x_i) = \Lambda(\boldsymbol{\beta}'\mathbf{X})$$

setting $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)$ and adding $X_0 \equiv 1$ to the vector $\mathbf{X}$. Estimation of $\boldsymbol{\beta}$ is usually based on maximizing the log-likelihood function numerically by the Newton-Raphson method (Thisted 1988). For an individual with covariate vector $\mathbf{X}' = [X_1, X_2, ..., X_p]$ the probability $\Pr(y=1|\mathbf{X})$ can be predicted by

$$\hat{\Pi} = \hat{\Pi}(\mathbf{X},\hat{\boldsymbol{\beta}}) = \Lambda(\hat{\beta}_0 + \sum_{i=1}^{p}\hat{\beta}_i x_i)$$

These terms are often simply referred to as "prediction" for given $\mathbf{X}$.

The major purpose of logistic regression is to correctly predict the category of outcome for individual cases using the most parsimonious model. To accomplish this goal, a model is created that includes all predictor variables that are useful in predicting the response variable. Variables can be entered into the model in the order specified by the researches or logistic can test fit of the model after each coefficient is added or deleted, called stepwise regression.

Day and Kerridge (1967) both suggested the logistic regression model for posterior probabilities as a basis for discrimination two populations Π_1 and Π_2 with prior probabilities p_1 and p_2 respectively. The objects are ordinarily separated or classified on the basis of measurements on p associated random variables $\mathbf{X}' = [X_1, X_2, ..., X_p]$. The simplest optimizing method of discrimination is to maximize the probability of correct allocation. This is achieved by allocating the sample point $\mathbf{X}$ to Π_1 (i.e. the response variable y=1) if

$$\Pr(y=1|\mathbf{X}) = \Pr(\Pi_1|\mathbf{X}) \geq \Pr(\Pi_2|\mathbf{X}) = \Pr(y=0|\mathbf{X})$$

otherwise to Π_2. Where, $p = \Pr(y=1|\mathbf{X}) =$ pr($\Pi_1|\mathbf{X}$) is given at (1) and $\Pr(\Pi_1|\mathbf{X}) + \Pr(\Pi_2|\mathbf{X}) = 1$.

The allocation of new individuals can be performed on the basis of scores given by the logit function i.e.

*Logit(***p***)* = *ln(***p** */1-* **p***)* $= (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ,..., + \beta_p X_p)$

If it is positive (with assumption of equal prior probabilities) the individual is allocated to Π_1 otherwise to Π_2. The logit coefficients $\boldsymbol{\beta}$ are estimated by the maximum likelihood estimation (MLE) using the iterative equations. To test the null hypothesis that a particular logit coefficient is zero the Wald's statistic is used. This is the square of the ratio of the estimated logit coefficient to its standard error and has a chi-square distribution (Hosmer and Lemeshow 1989).

ACKNOWLEDGEMENT

This work was supported financially by providing technical support from the Center of Modeling, Department of Biophysics, Faculty of science, Tarbiat Modarres University. Also the authors are indebted to Dr. Lawrence Danso Buadu for his invaluable assistance.

REFERENCES

Abdolmaleki P., Buadu L.D., Murayaama S., Murakami J., Hashiguchi N., Yabuuchi H., Masuda K. (1997). Neural Network Analysis of Breast Cancer from MRI Findings. *Radiation Medicine., **15:** 283-293.*

Abdolmaleki P., Buadu L.D., Naderimansh H. (2001). Feature extraction and classification of breast cancer on dynamic magnetic resonance imaging using artificial neural network. *Cancer Lett., **10:** 171-183.*

Biganzoli E., Boracchi P., Coradini D., Grazia Daidone M., Marubini E. (2003) Prognosis in node-negative primary breast cancer: a neural network analysis of risk profiles using routinely assessed factors. *Ann. Oncol., **14:** 1484-93.*

Buadu L.D, Murakami J., Murayama S., Hashiguchi N., Sakai S., Masuda K., Toyoshima S, Kuroki S., Ohno S. (1996). Breast lesions: correlation of contrast medium enhancement patterns on MR images with histopathologic findings and tumor angiogenesis. *Radiology*, ***200:*** *639-649.*

Day N.E. and Kerridge D.F. (1967). A general maximum likelihood discriminant. *Biometrics*, ***23:*** *313-23.*

Degenhard A., Tanner C., Hayes C., Hawkes D.J., Leach M.O. (2002). UK MRI Breast Screening Study. Comparison between radiological and artificial neural network diagnosis in clinical screening. *Physiol. Meas.*, ***23:*** *727-39.*

Heywang Kobrunner S.H. (1994). Contrast-enhanced magnetic resonance imaging of the breast. *Invest. Radiol.*, ***29:*** *94-104.*

Hosmer D.W. and Lemeshow S. (1989). Applied logistic regression. New York; Wiley

Metz C.E. (1986). ROC methodology in radiologic imaging. *Invest. Radiol.*, ***21:*** *720-733.*

Metz C.E. (1989). Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest. Radiol.*, ***24:*** *234-245.*

Thisted R.A. (1988). *Elements of statistical computing: Numerical computation*. New York, Chapman and Hall.

Tzacheva A.A., Najarian K., Brockway J.P. (2003). Breast cancer detection in gadolinium-enhanced MR images by static region descriptors and neural networks. *J. Magn. Reson. Imaging.*, ***17:*** *337-42.*

Vomweg T.W., Buscema M., Kauczor H.U., Teifke A., Intraligi M., Terzi S., Heussel C.P., Achenbach T., Rieker O., Mayer D., Thelen M. (2003). Improved artificial neural networks in prediction of malignancy of lesions in contrast-enhanced MR-mammography. *Med. Phys.*, ***30:*** *2350-9.*

Wasserman P.D. (1989). Neural computing: theory and practice. Vannortrand Reinhold Press.

Weinreb J.C., Newstead G. (1995). MR imaging of the breast. *Radiology*, ***196:*** *593-610.*