# Prediction of Malignant Breast Lesions from MRI Features: A Comparison of Artificial Neural Network and Logistic Regression Techniques

**K. Nie[1], C. E. McLaren[2], W-P. Chen[2], J-H. Chen[1,3], O. Nalcioglu[1], and M-Y. L. Su[1]**

[1]Tu & Yuen Center for Functional Onco-Imaging, University of California, Irvine, Irvine, CA, United States, [2]Department of Epidemiology, University of California, Irvine, Irvine, CA, United States, [3]Department of Radiology, China Medical University, Taichung, Taiwan

***Introduction:***

Breast MRI has demonstrated a high sensitivity of > 95%, with specificity of approximately 67% [1]. To address the challenge of accuracy and efficiency in interpretation of breast MRI, a computer-aided diagnosis (CAD) system that can automatically analyze lesion features to differentiate between malignant and benign lesions would be very useful. The general approach to breast CAD involves applying computer algorithms for tumor characterization and then developing statistical models using logistic regression analysis (LRA) or artificial neural networks (ANN) for classifying a lesion as malignant or benign. As there is no universal approach to select models for data classification; evaluation of tasks on a case-by-case basis is necessary. To the best of our knowledge, there is no study directly comparing the diagnostic performance of ANN and LRA techniques based on breast MRI features. We have developed a CAD that incorporates a clustering-based algorithm for lesion segmentation and derives a full panel of quantitative morphological and texture descriptors for lesion characterization using ANN [2]. In this study we investigated and compared the diagnostic performance using the ANN and logistic regression in the development of breast MRI CAD.

***Methods:***

The study included 43 malignant and 28 benign histological-proven lesions. 8 morphological parameters, 10 gray level co-occurrence matrices (GLCM) texture features, and 14 Laws' texture features were obtained using automated lesion segmentation. Firstly, the artificial neural network (ANN) with three-layer back-propagation, known as multi-layer perceptrons (MLP), was utilized to obtain optimal classifiers. While ANN can select features robustly, the non-linear diagnostic model makes the weighting of each individual feature not readily interpretable. For the feature set selected by ANN, although ANN already gave the performance (Area under ROC, sensitivity, specificity), we also applied logistic regression to analyze the diagnostic performance to gain more insightful understanding for how each selected feature can be interpreted. Secondly, we applied logistic regression for feature selection. The following statistical procedures were applied sequentially: 1) transformation of feature to induce normality, 2) model selection and validation, 3) assessment of correlation between features to reduce collinearity, 4) reduction of the number of variables in models to avoid overfitting. The overall dataset was separately and randomly assigned into four sub-cohorts. Three sets of any four sub-cohorts were combined as a training set, with the remaining one used as the corresponding validating cohort. The diagnostic performance of each model was evaluated, and compared to that achieved by ANN.

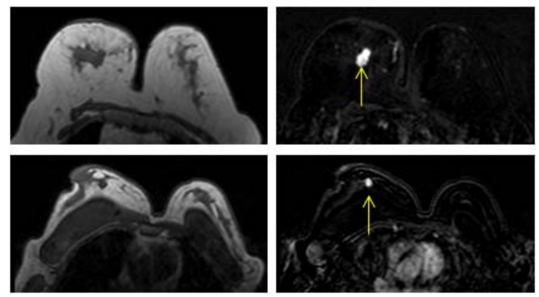

**Fig.1 Two cases with high/low compactness index (CI). The top-malignant (top) lesion shows irregular shape and margin with CI=63; the benign (bottom) case shows a round and smooth small lesion with CI=2.**

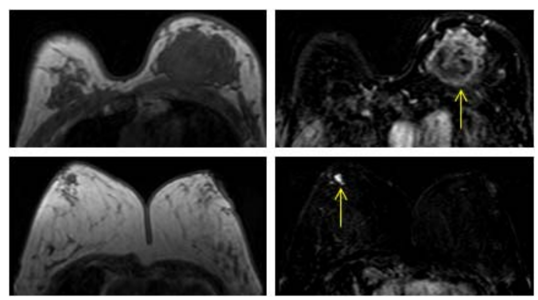

**Fig.2 Two cases with low/high GLCM energy index (EI). The top-malignant with EI=0.08 indicating lower homogeneity; the bottom-benign case with EI=0.4, indicating higher homogeneity.**

***Results:***

Using ANN, the final selected features were compactness, energy, homogeneity, and Law_LS, with AUC= 0.82, and accuracy= 0.76. The diagnostic performance of these 4-features computed on the basis of logistic regression yielded AUC= 0.80, similar as that of ANN. For the outcome of presence of a malignant lesion, the logistic regression equation was: logit ($p$) = 0.69 + 1.69 Compactness - 0.61 Energy + 0.24 Homogeneity - 0.73 Law_LS. Using logistic regression, 3 models were selected from training cohort 1-4 (model D and E are identical). AUC values varied from 0.75 to 0.80 as shown in **Table 1**. However, the 95% confidence intervals for the AUC overlap, indicating no significant difference between AUC values. Model C comprised of compactness, NRL entropy, and gray level sum average had the highest accuracy of 0.75, with AUC= 0.77. The logistic regression equation was logit($p$)=0.66 + 1.28 Compactness - 0.66 NRL Entropy + 0.72 Gray Level Sum Average.

For all these models, compactness has been consistently chosen. The compactness is defined as the ratio of the square of surface area to the volume; therefore it is sensitive to spherical (low compactness index) vs. spiculation (high compactness index) lesion morphology. An example is given in **Fig 1**. The distribution of enhancements within the lesion was described by the GLCM and Laws texture features. GLCM Energy was selected by both ANN and LRA (Models A & B). Benign lesions were found to show more homogeneous enhancements (high energy index) than malignant tumors (low energy index). An example is given in **Fig 2**.

**Table 1. Diagnostic evaluation of models selected using the artificial neural network (ANN) and logistic regression analysis (LRA) techniques**.

| Classifier by ANN | Image descriptors | Methods | Accuracy | Sensitivity | Specificity | AUC |
|---|---|---|---|---|---|---|
| A | Compactness, Energy, Homogeneity, LAW_LS | ANN | 0.76 | 0.84 | 0.64 | 0.82 |
| | | LRA | 0.72 | 0.86 | 0.50 | 0.80 |
| Classifier by LRA | | | | | | |
| B: Cohort-1 | Compactness, NRL Entropy, Energy | LRA | 0.73 | 0.95 | 0.39 | 0.75 |
| C: Cohort-2 | Compactness, NRL Entropy, Gray level sum average | LRA | 0.75 | 0.91 | 0.50 | 0.77 |
| D: Cohort-3 | Compactness, NRL Entropy, LAW_LW | LRA | 0.72 | 0.86 | 0.50 | 0.80 |
| E: Cohort-4 | Compactness, NRL Entropy, LAW_LW | LRA | 0.72 | 0.86 | 0.50 | 0.80 |

***Discussion:***

In summary, we have shown that the diagnostic performance of models selected by ANN and logistic regression was similar when small number of variables was chosen. The ANN methodology is more robust which does not require a high level of operator judgment. On the other hand, logistic regression may generate many sets of models that yield similar diagnostic performance, and have to rely on the operator to make intellectual judgments to select the best model(s). However, the poor interpretability is the limitation of ANN. In developing the breast MRI CAD, the direct association between quantitative predictors and the outcome should be established to give a better understanding of the effect of the selected predictor on outcome variables. As such, logistic regression may be still more favorable to ANN due to improved interpretation of individual predictors. During the development of the breast CAD system, ANN may be used to first identify the optimal classifier and logistic regression technique may be used complimentarily to provide more insight information about the diagnostic role of the selected predictors.

***Reference:*** [1]. Bluemke et al. JAMA. 2004;292(22):2735-42. [2]. Nie et al. 2008 ISMRM proceedings, program # 3752 or Acad Radiol. 2008 (in press).

***Acknowledgement:*** This work was supported in part by NIH/NCI CA90437, CA121568 and CBCRP 9WB-0020, 14GB-0148.