Feasibility of Multi Layered Perceptron (MLP) Networks in Discriminating between Benign and Malignant Lesions in DCE-MRI of Breast

M. Sreenivas[1], P. Gibbs[1], L. W. Turnbull[1]

[1]Centre for MRI, University of Hull, Hull, East Yorkshire, United Kingdom

Introduction: Breast cancer is the second most common cause of cancer deaths in women. In the recent years mortality rates have declined owing to earlier detection and improved treatment [1,2]. Traditionally the triple assessment method (consisting of physical examination, x-ray mammography/ultrasound scan and histopathology) is used in diagnosing lesions in breast. Magnetic resonance imaging of breast is now being increasingly used as it offers very high sensitivity compared to conventional assessment. Some of the problems associated with this modality are relatively low specificity (37% to 85%)[3], the relative importance of the many parameters used for interpreting an examination is not known and nor is the complex interaction among them readily discernable, the amount of date generated is quite large and understandably the analysis will take a long time, and the scarcity of trained breast MR radiologists. In this pilot study we have tried to address these issues by demonstrating the ability of artificial neural networks such as MLP network to reliably discriminate between benign and malignant lesions at a low time cost.

Methods: In this retrospective study two hundred and twenty nine patients (mean age 53 years, range 27 to 82 years) underwent DCE-MRI of breast between January 1999 and July 2003. Patients were referred for either evaluation of difficult benign lesions (as assessed by triple assessment method) or further assessment of malignant lesions. A total of two hundred and forty eight lesions (benign = 77 and malignant = 171) were found for which either histological confirmation (n = 244) or clinical follow up information was available (n = 4, mean follow up time 29 months, range 8 to 48 months). All the imaging was performed on a 1.5T GE Signa unit with a commercially available dedicated bilateral phased-array breast coil with the patients in prone position. Following initial localisation scans a 2D T1W FSPGR (TR/TE/flip angle = 7.6 ms/4.2ms/30°) was acquired dynamically over 35 time points at nine slice locations with temporal resolution of 11.6 seconds and a post contrast fat saturated 3D T1W FSPGR (TR/TE/flip angle = 20.4 – 27.5ms/4.2ms/30°, slice thickness = 3.5 – 5.0 mm without gap, FOV 20 – 36 cm, matrix size 512x256). A single region of interest (ROI) was drawn around each lesion on the dynamic study from which pharmacokinetic parameters (n=5) were obtained by using a 2-compartment model similar to that proposed by Brix et al[4]. Similarly, from a single ROI drawn around each lesion (figure 1 and 2) on the post contrast study texture (n=14) and shape (n=10) parameters were calculated. The texture parameters were calculated with the spatial grey-level dependent matrix method (using 32 grey-levels) as suggested by Haralick et al[5]. All the analysis was performed on Sun micro systems. The parameters that we have used for constructing neural networks emulate those used by a radiologist in interpreting DCE-MRI study. Intelligent problem solver (IPS) (STATISTICA version 6, StatSoft, Inc) was used to construct one thousand three-layered feed forward MLP networks. By performing pre-processing (t-test and Mann-Whitney U test), the most significant parameters ($p < 0.05$) were first obtained in order to address the issue of curse of dimensionality[6]. Back propagation and conjugate gradient descent were used as training algorithms. Initially, the data was randomly divided into a training set (n=124), a selection set (n=62) and a test set (n=62). The single network, in terms of least error in training, selection and testing, was finally chosen.

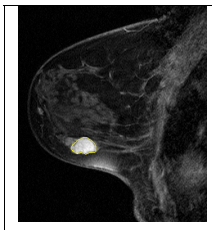

Figure 1. Benign lesion

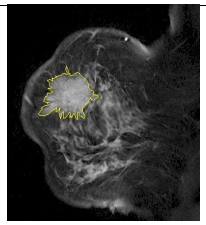

Figure 2. Malignant lesion

Results: Pre-processing resulted in reducing the input parameters from 29 to 19. The best MLP network selected 13 of the final 19 inputs resulting in an architecture consisting of 13 units in the input, 11 units in the hidden and 1 unit in the output layer. Its performance on the training, selection and test set is shown in table 1. As the network does not see the test set data during training or selection process, the real test for any network is its performance on this set. The performance of the network was further analysed using the receiver operating characteristic curve and the area under the curve was 0.896 for the test set (figure 3). Sensitivity and specificity in discriminating a malignant lesion from a benign lesion were calculated (test set data) for the MLP network and were compared with the radiologist's prediction* (table 2). The network correctly predicted 18 out of 18 benign and 35 out of 44 malignant lesions.

		True M	True B
Training set (n=124)	M	66	11
	B	19	28
Selection set (n=62)	M	32	6
	B	10	14
Test set (n=62)	M	35	0
	B	9	18

Table 1. Best MLP network's performance

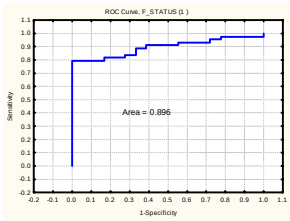

Figure 3. ROC

	MLP network	Radiologist
Sensitivity	79.5%	100%
Specificity	100%	66.6%

Table 2. MLP vs radiologist's performance on the test set

Conclusions: The feasibility of MLP networks in reliably discriminating between benign and malignant lesions is demonstrated in this pilot study. MLP networks can be trained on similar information that a radiologist would use to interpret lesions on DCE-MRI study of breast and perform as well as a trained radiologist. The time of computation is low – it took just under sixteen minutes to construct one thousand different MLP networks on a computer with a Pentium 4 processor. When given further data, interpretation using a trained network is almost instantaneous. Networks that are well trained may have a role in improving diagnostic specificity. This would have enormous implications as it could result in a reduction of unnecessary fine needle aspiration cytology and/or biopsy. In the future, this work will be extended to a larger data set and the feasibility of other neural networks such as radial basis function will be tested.

References: 1.Cancer Facts and Figures – 2004, American Cancer Society 2.Breast Cancer - October 2003, Office for National Statistics 3.CK Kuhl (2000) Eur Radiol 10:46-58 4.Brix et al (1991) J Comput Assist Tomogr 15(4):621-628 5.Haralick et al (1973) IEEE transactions on systems, man and cybernetics 3(6):610-621 6.CM Bishop (1995) Neural Networks for Pattern Recognition, Oxford: Oxford University Press, section 1.4

* Radiologist used a five point scoring system for lesion reporting where 1=normal, 2=benign, 3=probably benign, 4=suspicious of malignancy and 5= malignant. For the purpose of calculating sensitivity and specificity, lesions with score of 1-3 weere categorised as benign and those with score of 4 and 5 as malignant. It is important to note that the benign lesions in the study were difficult lesions and were referred for further assessment using MR after initial triple assessment.