Identification of masses in digital mammograms with MLP and RBF Nets

Keir Bovis[1], Sameer Singh[1], Jonathan Fieldsend[1], Chris Pinder[2]
[1]*PANN* RESEARCH, DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF EXETER, EXETER, UK
[2]BREAST CARE UNIT, ROYAL DEVON & EXETER HOSPITAL, EXETER, UK
{k.j.bovis, s.singh, j.e.fielsend, @exeter.ac.uk; pind@pind.screaming.net}

ABSTRACT
In this paper we study the identification of masses in digital mammograms using texture analysis. A number of texture measures are calculated for bilateral difference images showing regions of interest. The measurements are made on co-occurrence matrices in four different direction giving a total of seventy features. These features include the ones proposed by Haralick et. al., (1973) and (Chan et al., 1997). We study a total of 144 breast images from the MIAS database. The dimensionality of the dataset is reduced using principal components analysis (PCA). PCA components are classified using both multilayer perceptron networks using backpropagation (MLP) and radial basis functions based on Gaussian kernels (RBF). The two methods are compared on the same data across a ten fold cross-validation. The results are generated on the average recognition rate over these folds on correctly recognising masses and normal regions. Further analysis is based on the Receiver Operating Characteristic (ROC) plots. The best results show recognition rates of 77% correct recognition and an area under the ROC curve value Az of 0.74.

1. INTRODUCTION

X-ray mammography is the most common technique used by radiologists in the screening and diagnosis of breast cancer in women. Although it is seen as the best examination technique for the early detection of breast cancer reducing mortality rates by up to 25%, their interpretation requires skill and experience by a trained radiologist. The aim of this study is to analyse digitised mammograms by applying computer image processing techniques to enhance x-ray images and then subsequently extract features from suspicious regions characterising the underlying texture of the breast regions. These features can then be passed to a classifier for discrimination for different regions of interest to test whether they are masses or non-masses.

In order to identify suspicious regions, the clinical observation of focal asymmetric densities within the internal structure of the breast is implemented (Kopans, 1997). Implementing this method requires left and right breast images to be aligned around a common reference point, then subtracted bilaterally (see Figure 1). On the basis of several past studies, we follow the fundamental assumption that asymmetries represent a region of interest that could be a mass. Several previous studies have applied the bilateral subtraction technique to mammographic image pairs. Yin et al. (1993) investigated mammographic asymmetries for the identification of mass lesions. In their research the authors compared the performance of a non-linear bilateral subtraction technique with grey level thresholding. For bilateral subtraction, left and right breast images were aligned, thresholded and following subtraction subsequently thresholded again. For the comparative method, the authors implemented a method of local grey scale thresholding a single image divided into 100 x 100 pixel blocks. For bilateral subtraction using ROC analysis the performance of the technique, measured by the area under the ROC curve Az, was found at Az=0.530 and for local grey level thresholding the Az=0.385. Ideally, the area under the curve Az should be as close to 1 as possible for an accurate technique. Mendez et al.(1998), investigated a computerised method to automatically detect malignant masses on digital mammograms based on bilateral subtraction to identify asymmetries between left and right images. The nipple was detected to align the images, and alignment of image pairs achieved by translation using the nipple location and rotation against a correlation coefficient. To reduce false positives, texture tests were applied to each suspicious region, coarseness and contrast. Using ROC analysis, the value of Az=0.667 was achieved. In their study, Chan et al. (1995), studied the effectiveness of using texture features derived from Spatial Grey Level Dependency (SGLD) matrices for classification of masses and normal breast tissue on mammograms. All images were digitised at 4,096 grey levels and a region of interest confirmed by biopsy comprising 256 x 256 pixels. The following texture features were extracted from the SGLD matrices: correlation, entropy, angle of second moment, inertia, inverse difference moment, sum average, sum entropy, difference entropy. The accuracy was evaluated using ROC analysis and the maximum value of the area under the ROC curve obtained was Az = 0.823.

2. METHODOLOGY

The MiniMIAS database (Suckling et al., 1994) contains left and right breast images for a total of 161 subjects. In this study we study 72 subjects. All images containing spiculated/circumscribed masses and a selection of normal types are

0-7695-0619-4/00 $10.00 © 2000 IEEE

considered. Supplied ground truth data for each image includes tissue type and regions of interest encompassing the abnormalities given in terms of a centroid and region radius. All images are digitised at a resolution of 1024 X 1024 pixels and at 8-bit grey scale level. Forty images contain abnormalities (circumscribed masses n=21, spiculated masses n=19) and 104 images are classed as normal. In the analysis of results within the study, we use the following definitions:

True Positive (TP): lesions called cancer and prove to be cancer
False Positive (FP): lesions called cancer that prove to be benign
False Nagative (FN): lesions that are called negative or benign and prove to be cancer
True Negative (TN): lesions that are called negative and prove to be negative

On the basis of this terminology, we can evaluate the performance of our technique by calculating True Positive Fraction (TPF) and False Positive Fraction (FPF). These are defined as:

$$TPF = \frac{TP}{TP + FN} \qquad FPF = \frac{FP}{FP + TN}$$

The study involved five different phases:

1. *Location of a Common Reference Point*
In order to align left and right breast image pairs prior to bilateral subtraction a common reference, the spatial position of the nipple, is located.

2. *Alignment and Bilateral Subtraction of Left and Right Breast Images*
Once the nipple has been located in each breast image of a matching pair, the observed image is translated such that the nipple locations of both breast images are aligned. Taking different rotations of the observed image and determining a correlation measure against the reference image, the left original breast image, the best alignment can be determined. Initially the observed image is subjected to large degrees of rotation, following which the best range is located and the process repeated with smaller incremental rotations until the best correlation is found. Using aligned left and right breast image pairs, two images are generated by bilaterally subtracting one image from the other. One is a positive image and details differences that occur in the left breast image and not in the right, and a negative image that details differences that occur in the right breast image and not in the left.

3. *Reduction of False Positives*
Each difference image generated from bilateral subtraction, following contrast enhancement using a histogram stretching technique (Umbaugh, 1998), will contain regions that are true positives, and those that are false positives. Difference images are segmented using a region splitting technique, the complete set of regions being modelled using a quad-tree data structure representation (Jain et al., 1995). Once regions have been identified, the aim is to subsequently remove as many false positive regions on the basis of region characteristics size, shape (Petrick et al., 1996), difference in homogeneity (Umbaugh, 1998) and entropy (Gonzalez et al., 1993).

4. *Feature Extraction*
Using the quad-tree region model generated in the previous phase for all remaining suspicious regions within a difference image, five co-occurrence matrices are constructed in four different spatial orientations, horizontal, left diagonal, vertical and right diagonal, (0°, 45°, 90° and 135°, respectively). A fifth matrix is also constructed as the mean of the first four. Each co-occurrence matrix reflects the joint probability of a pixel pair at a given orientation and distance. From each of the five normalised co-occurrence matrix, the 14 textures features are extracted giving a total of 70 texture features per region. Each region is labelled mass or non-mass using the supplied MIAS ground-truth data. These 70 features are extracted from the co-occurrence matrices constructed at four pixel distances (d=1, 3, 6, 9) giving four separate feature vectors.

5. *Classification*
The classification of masses from normal regions requires high quality classification system as most of these differences can be at times subtle. In this paper, we develop two different types of networks for recognition of masses. The first network is a three layered multi-layer perceptron (input, hidden and output layers). The second network is a Gaussian kernel based Radial Basis Function network.

In order to determine the discriminating effectiveness of texture features extracted from co-occurrence matrices constructed at different distances (each distance yields a different feature set), classification is initially performed using each normalised feature set and linear discriminant analysis. The feature set giving the best true positive fraction

was selected for subsequent classification using the two networks described above. Both networks are developed using the Stuttgart Neural Network Software (SNNS) package. For classification using an ANN, the selected feature set is normalised and principal component analysis performed to reduce the dimensionality of the data. We find a total of seven principal components for our data. Subsequently, samples for classes are interleaved and training/test files can be generated using a 10-fold cross validation method (Bishop, 1998). In each fold, 90% of data is used for training the classifier and the remaining 10% is used for testing. The process continues 10 times taking different training and test partitions, each time with a different training and test set.

In the MLP network, each of the seven principal components is mapped onto an individual input node and linked by a varying number of hidden nodes to two output nodes (mass, non-mass). For training, a back-propagation with momentum learning function is used (learning rate η= 0.01, momentum μ= 0.9). In order to prevent over-generalisation by controlling the complexity of the ANN, the technique of structural stabilisation (Bishop, 1998) is used. Using this technique, the TPF for a range of ANNs with differing number of hidden nodes is compared. The ANN model giving the highest TPF is selected. ROC analysis is finally performed (McNicol, 1972) on the test results from each fold and the value of Az computed.

Similarly, the RBF network is trained to a maximum of 500 epochs using a learning rate for the centres of .0002. The same learning rate value was used for the weights and bias for the output. The maximum error allowed for each neuron was set to 0.8 and the momentum used was equal to 0.8. It should be noted that these values have been optimised through experimentation.

Two separate trials are conducted. In the first trial the complete data available initially is used. In the second trial, we eliminate 10% of outliers from each class data. Outliers are selected using a nearest neighbour method that computes the mean pattern for each class. The Euclidean distance between this mean pattern and all other samples for a given class are computed and arranged in their order of increasing distance. The highest 10% values are removed from the data.

3. RESULTS

From the 144 input images, the dataset comprised 820 non-masses and 48 actual masses. Tables 1 and 3 show results on this data. Tables 2 and 4 show the results after outliers have been removed which leaves 738 non-masses and 43 masses. The co-occurrence matrices constructed using a distance *(d=3)* gave the largest TPF value thus being the most sensitive to detecting true positives. This distance represents a trade-off between the construction of the complete joint probability histogram for each region, needed to describe its texture, limited by a region's size, coupled with a minimum value of d required to characterise the underlying texture associated with pixel pairs. For each test set tested on a trained ANN using 10-fold cross-validation, the associated feature vector (where d=3) is subsequently used for classification.

In Table 1, we show the results of classification performance based on the RBF network. The recognition rates vary over the ten trials; on average we get a value of 65%. The TPF and FPF values listed for the network performance shows that this configuration of an RBF network run on original data (without any outlier removal) gives an unacceptably high FPF and a poor TPF related to the overall lower figure of Az. The value of Az (area under the ROC curve) is found to be 0.70. As mentioned earlier, an ideal system should have a value close to 1.0. When the outliers are removed from training data, test performances improve as shown in Table 2. A recognition rate of 78% is obtained with better TPF and FPF scores. Also the value of Az improves to 0.73. In both experiments, 5 hidden nodes give the optimal results. Increasing them leads to over-generalisation such that all masses are classified as non-masses.

The same analysis is performed using a multi-layer perceptron with backpropagation learning. In this, the number of optimised hidden nodes used in different trials are different. The recognition rate obtained equals 72% and the average Az equals 0.72. The TPF and FPF values listed for the network performance shows that this type of ANN gives a lower FPF and an improved sensitivity in detecting masses. Finally Table 4 shows the results of classifying outlier removed data with backpropagation net. This shows the best performance on recognition rate (77% correct), highest TPF (.73), lowest FPF (.25) and highest Az value (.74).

Finally we plot the ROC curves for the best performance of outlier removed data classified using backpropagation network. These are shown in the Appendix. The curves plot TPF vs. FPF. Ideally, the curves should be as close to the y-axis and the upper edge of the graph as possible. As seen from the graphs in the appendix, most graphs have this trend.

4. CONCLUSIONS

A key aim of this study was the implementation of a technique that could automatically detect suspicious regions. The

bilateral subtraction technique of left and right breast image pairs is seen as an effective method of identifying such suspicious regions based on focal asymmetric densities. The reduction of false-positive regions based on size, shape and uniformity measures ensured that the ultimate FPF could be kept low without affecting the overall sensitivity. The use of statistical measures of texture as features, their subsequent extraction and classification, can be seen from the results in Tables 1-4 to be highly discriminant in differentiating between masses and non-masses with a very high TPF with a low TPF. In comparison with previous studies the overall performance of the technique as indicated by the value of Az=0.74, which improves on similar previous studies where the detection masses has been the objective. Future work will attempt to improve on these results and include different texture and shape measures applied to other mammography databases.

REFERENCES

[1] C. Bishop, 1998, Neural Networks for Pattern Recognition, Clarendon Press.

[2] H. Chan, B. Sahiner, N. Petrick, M. Helvie, K.L. Lam, D. Adler, M. Goodsitt, 1997, Computerised Classification of Malignant and Benign Microcalcifications on Mammograms: Texture Analysis Using and Artificial Neural Network, *Physical Medical Biology*, vol. 42, pp. 549 - 567.

[3] H. Chan, D. Wei, N. Helvie, B. Sahiner, D. Adler, M. Goodsitt, N. Petrick, 1995, Computer-aided Classification of Mammographic Masses and Normal Tissue: Linear Discriminant Analysis in Texture Feature Space, *Physical Medical Biology,* vol. 40, pp. 857 - 876.

[4] R. Gonzalez, R. Woods, 1993, Digital Image Processing, Addison Wesley.

[5] R. Haralick, K. Shanmugam, I. Dinstein, 1973, Textural Features for Image Classification, *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610 - 621.

[6] A. Jain, 1995, Fundamentals of Digital Image Processing, Prentice-Hall of India.

[7] D. Kopans, 1998, Breast Imaging, Lippincott-Raven.

[8] D. McNicol, 1972, A Primer of Signal Detection Theory, Australasian Publishing Company.

[9] A. Mendez, P. Tahoces, M. Lado, M. Souto, J. Vidal, 1998, Computer-aided Diagnosis: Automatic Detection of Malignant Masses in Digitised Mammograms, *Medical Physics,* vol. 25, no. 6, pp 957-964.

[10] N. Petrick, H. Chan, D. Wei, B. Sahiner, M. Helvie, D. Adler, 1996, Automated Detection of Breast Masses on Mammograms Using Adaptive Contrast Enhancement and Texture Classification, *Medical Physics*, vol. 23, no. 10, pp. 1685-1696.

[11] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal and J. Savage. "The mammographic images analysis society digital mammogram database." Exerpta Medica. International Congress Series 1069: 375-378 (1994). (mias@sv1.smb.man.ac.uk).

[12] S. Umbaugh, 1998, Computer Vision and Image Processing: A Practical Approach Using CVIPtools, Prentice Hall.

[13] F. Yin, M. Giger, C. Vyborny, K. Doi, R. Schmidt, 1993, Comparison of Bilateral-Subtraction and Single-Image Processing Techniques in the Computerised Detection of Mammographic Masses, *Investigative Radiology,* vol. 28, no. 6, pp. 473 - 481.

Fold No.	Hidden Nodes	Recog. Rate %	TPF	FPF	Az
1	5	74	0.40	0.24	0.837
2	5	62	0.60	0.38	0.739
3	5	62	0.60	0.38	0.600
4	5	67	1.00	0.35	0.829
5	5	77	0.80	0.23	0.734
6	5	65	0.80	0.35	0.723
7	5	52	0.60	0.49	0.609
8	5	57	0.40	0.41	0.474
9	5	76	0.80	0.24	0.790
10	5	61	0.40	0.39	0.711
Mean	-	**65**	**0.64**	**0.35**	**0.70**

Table1 Results of classification using an RBF net

Fold No.	Hidden Nodes	Recog. Rate %	TPF	FPF	Az
1	5	74	0.60	0.24	0.683
2	5	66	0.80	0.35	0.722
3	5	67	0.60	0.32	0.672
4	5	73	0.60	0.26	0.660
5	5	80	0.80	0.20	0.898
6	5	85	0.40	0.12	0.650
7	5	84	0.20	0.49	0.664
8	5	68	0.40	0.12	0.656
9	5	83	1.00	0.30	0.983
10	5	100	-	-	-
Mean	-	**78**	**0.60**	**0.27**	**0.73**

Table2 Results of Classification using an RBF net (10% outliers have been removed)

Fold No.	Hidden Nodes	Recog. Rate %	TPF	FPF	Az
1	3	74	0.80	0.27	0.717
2	1	62	0.80	0.39	0.700
3	4	74	0.20	0.23	0.486
4	1	77	0.80	0.23	0.781
5	4	78	0.80	0.22	0.776
6	1	64	0.80	0.37	0.721
7	3	57	0.80	0.44	0.589
8	4	70	0.60	0.29	0.704
9	5	89	0.80	0.11	0.832
10	4	76	1.00	0.24	0.896
Mean	-	**72**	**0.74**	**0.28**	**0.72**

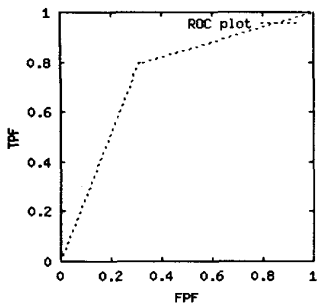
Table3 Results of classification using an MLP net

Fold No.	Hidden Nodes	Recog. Rate %	TPF	FPF	Az
1	1	70	0.80	0.31	0.743
2	5	78	0.40	0.19	0.617
3	1	70	0.60	0.30	0.601
4	3	77	0.80	0.23	0.789
5	1	75	1.00	0.27	0.836
6	2	68	0.60	0.31	0.674
7	4	67	0.80	0.34	0.736
8	4	82	0.60	0.16	0.737
9	5	83	1.00	0.18	0.898
10	5	100	-	-	-
Mean	-	**77**	**0.73**	**0.25**	**0.74**

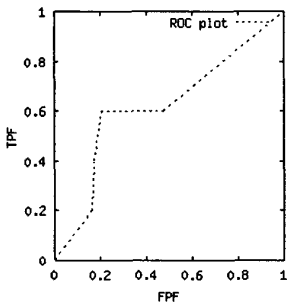
Table 4 Results of Classification using an MLP net (10% outliers have been removed)

Appendix: ROC plots for folds 1 to 9

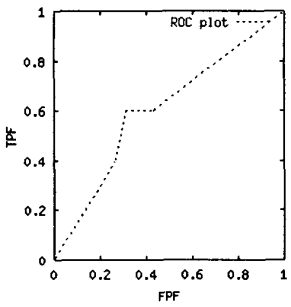
(there are no samples of masses in the final fold so we do not plot the ROC curve for the last fold)



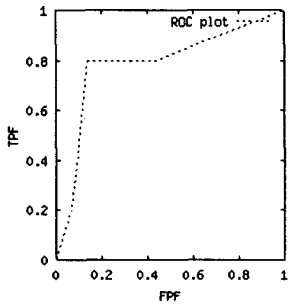
ROC curve for fold_1



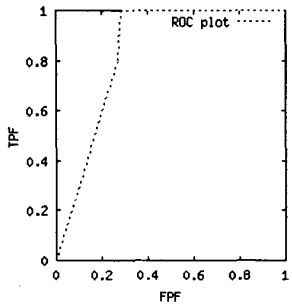
ROC curve for fold_2



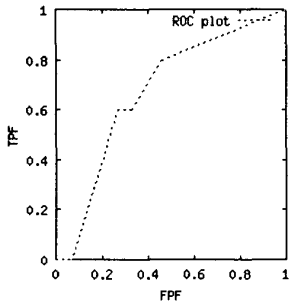
ROC curve for fold_3



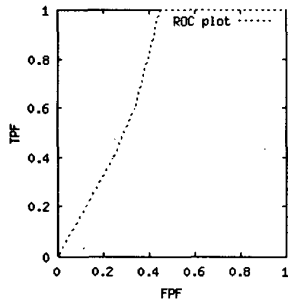
ROC curve for fold_4



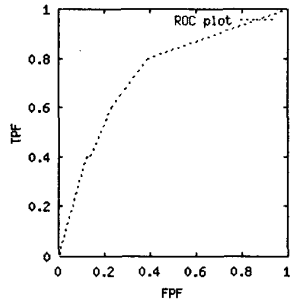
ROC curve for fold_5



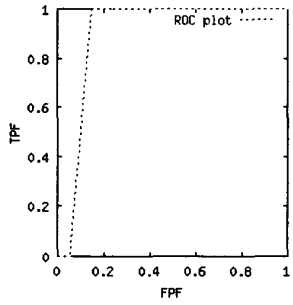
ROC curve for fold_6



ROC curve for fold_7



ROC curve for fold_8



ROC curve for fold_9