Classification of Breast Density in Digital Mammograms

Keir Bovis *and* Sameer Singh

PANN Research, Department of Computer Science, University of Exeter,

Exeter EX4 4PT, UK

Abstract

In this paper we investigate a new approach to the classification of mammographic images according to breast type based on the underlying texture contained within the breast tissue. Three methods for quantifying the texture are considered and used as input in the evaluation of four different classifiers. In this study we examine two classification tasks, a three-class classification problem between dense, glandular and fatty breast types and a two-class problem, differentiating between dense and fatty breast types. We use Receiver Operating Characteristic (ROC) analysis to evaluate the performance of the two-class problem. The data set used in this study is the Mammographic Image Analysis Society (MIAS) MINIMIAS database containing Medio-Lateral Oblique (MLO) views for each breast for 161 patients. For the three-class problem using a 3-layer feed-forward artificial neural network trained with conjugate gradient descent and 10-fold cross validation, we obtain a recognition rate on test of 70.4%. For the two-class problem test using a *k*-nearest neighbour classifier and 10-fold cross validation we obtain the area under the ROC curve *Az* equal to 0.832. This study demonstrates a high sensitivity in the classification of breast types justifying the use of this prior knowledge for the detection of lesions in a proposed CAD system.

Index Terms: Mammography, computer-aided detection, breast type, texture.

I. INTRODUCTION

A report from the National Cancer Institute (NCI) estimates that about 1 in 8 women in the United States (approximately 12.6 percent) will develop breast cancer during their lifetime [4]. Government sponsored mass-screening mammography programs have been proposed as an effective method of increasing survival time for women with breast cancer [18], but the application of Computer Aided Detection (CAD) within screening programs is still to be addressed. To this end, much research has

taken place for the development of CAD techniques and systems. Previous CAD studies have focussed on two key areas; the detection of calcifications [7][8][2][24] and the detection of mammographic masses [20][19][13][10][25]. In these areas, various degrees of success have been achieved but the main challenge of accurately identifying breast cancer from digital mammograms still remains.

For a radiologist interpreting a benign mammogram, there exists an extremely wide variation in the mammographic appearance of the breast. Within the mammogram, radiographically visible density includes ducts, lobular elements, and fibrous connective tissue. The fibrous connective tissue can be of two types, intralobular or extralobular tissue, and this latter tissue type is seen as the major component of gross density variation in mammograms. Breast density is an important factor in the interpretation of a mammogram. In a breast that is considerably dense, the sensitivity of mammography for the early detection of malignancy and large cancers is reduced because of the difficulty in locating ill-defined cancers within a non-uniform background. The American College of Radiology (ACR) Breast Imaging Reporting and Data System (BIRADS), identifies four major groups for classifying breast density (Kopans [11]):

1. Predominantly fat.
2. Fat with some fibroglandular tissue.
3. Heterogeneously dense.
4. Extremely dense.

Examples of each of these breast densities are shown in figure 1. It is apparent that as the breast density increases, the associated sensitivity in detecting breast cancer will be reduced.

This paper identifies a strategy for classifying the breast according to breast type on the basis of the underlying texture of the breast that correlates well with its density. This is extremely important for two reasons. First, using texture measures we can automatically categorise a new mammogram into one of the density classes reducing any subjectivity introduced when radiologists do it manually. Second, on the basis of such categorisation, we are able to better select the optimal tools for image processing of mammograms. For example, it is quite likely that different image processing algorithms are better

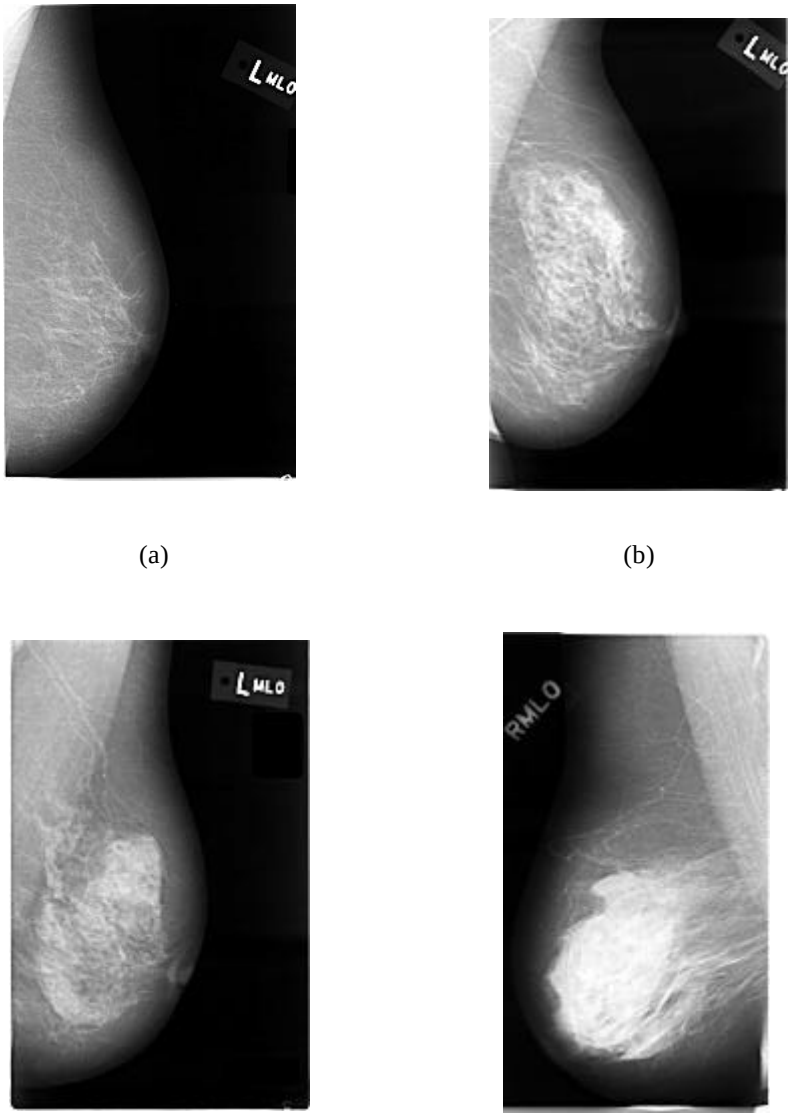
suited to certain class of mammograms in terms of their density. Also dense breasts, that are the most difficult to diagnose, can be automatically filtered out such that either a radiologist checks them or they go through double screening to ensure that correct screening has been performed. Our proposed methodology for the classification of mammograms on the basis of their texture into different density based categories has been tested on 322 oblique-view mammograms made available for researchers by the Mammographic Image Analysis Society (MIAS) [22] as part of the MIAS benchmark.

II. PREVIOUS WORK

Previous work has focussed on using the underlying texture to discriminate between breast types. In their study, Miller and Astley [17] investigated texture-based discrimination between fatty and glandular breast types, experimenting with granulometric techniques and Laws' texture masks. In the granulometric approach, the morphological operator 'opening' was applied to the original image with different sized structuring elements. The texture information was obtained by examining the reduction in the sum of pixel values against the size of the structuring element. For their comparative study, the authors applied Laws' texture masks and segmented the images on the basis of grey-level statistics obtained from the texture images. Both approaches were tested on 40 mammograms with predefined ground-truths using a Bayesian classifier tested on a leave-one-out basis. The results were evaluated with Receiver Operating Characteristic (ROC) graphs with the area under the curve indicating the performance of the evaluated technique. The granulometric approach gave a recognition accuracy of 75.7% and for Laws' texture energy, a recognition accuracy of 80.3% was obtained between the two breast types.

Taylor et al. [23] similarly investigated the classification of fatty and dense breast types. Initial experiments used a variety of texture measures extracted from a manually selected 128 x 128-pixel window for 90 images. Using the original window, re-sampling yielded various sub-images. The skewness of a grey-scale distribution in a 16 x 16 tile, fractal dimension, Laws' texture energy and the standard deviation of pixels in a 64 x 64 square, were extracted as features. An automated method of extracting the Region Of Interest (ROI) using the best texture measure resulted in 34% of the fatty breast and 5% of the dense breast samples being misclassified. In a final experiment the technique was

applied to 104 pairs of screening mammograms and resulted in the correct breast type classification of 67.35%.



(a) (b)

(c) (d)

Figure 1. Examples of different breast densities classified according to the ACR BIRADS classification (a) predominantly fat, (b) fat with some fibroglandular tissue, (c) heterogeneously dense and (d) extremely dense.

III. OVERVIEW OF SYSTEM

Our research is different from previous studies in two ways. Firstly, unlike previous studies that have attempted to classify breast tissue density on the basis of data within a predefined window, our approach considers the whole breast, in a similar manner to that perceived by a radiologist viewing the slide. Secondly, we make use of an extensive range of textural features to quantify the underlying texture for classification of breast type.

Initially the original image is sub-sampled thereby reducing the computational cost in subsequent processes and then segmented into two disjoint regions namely, breast and background. Texture measures extracted from the segmented breast region are classified using a selection of classifiers to maximise the recognition rate achieved. Finally we evaluate the performance of each of the classifiers using Receiver Operating Characteristic (ROC) analysis (Metz [16]). A block diagram identifying the major components of the proposed systems is given in figure 2.

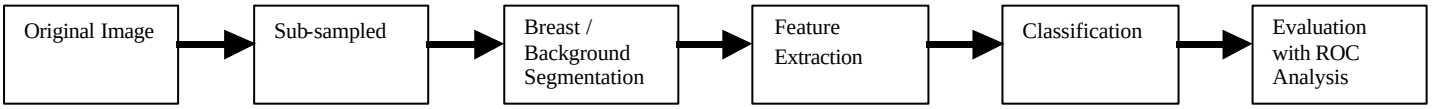


Figure 2. Overview of basic components of the proposed system

A. Sub-Sampling and Alignment

The original grey scale image is initially aligned on the *y-axis* such that the image is orientated with the nipple pointing right. It is then sub-sampled by a factor of two to reduce the computational complexity for the subsequent breast extraction component. This implementation of sub-sampling utilises averaging thus avoiding any blocking effect in the reduced image (Klette and Zamperoni [9]).

B. Breast / Background Segmentation

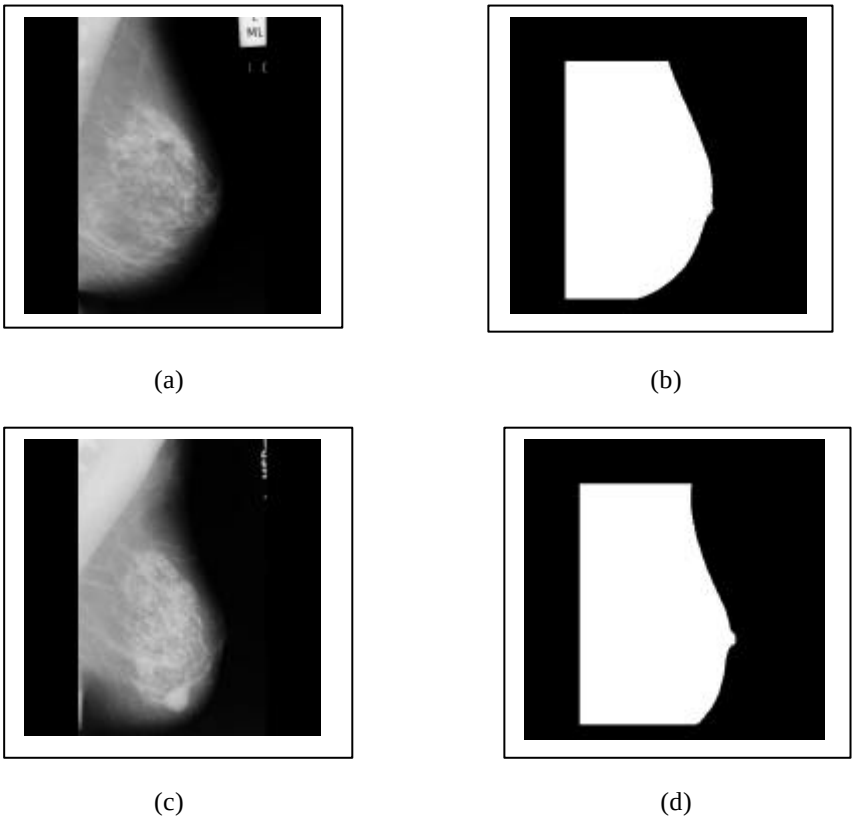
As our approach aims to utilise the whole breast for texture feature extraction, a mechanism for segmenting the breast from its background is required. Previous studies in breast/background segmentation [12][15] have difficulties due to the inherent noise within digitised mammograms. We adopt the technique proposed by Chandrasekhar and Yttikiouzel [3] with an additional step to trim the profiles, removing the top 20% and bottom 10% of the images to facilitate the removal of poor

segmentation that might still include noise. Some examples of the segmented breast profiles obtained using this method are shown in figure 3.

B. Feature Extraction

Previous approaches to classifying breast type have examined the underlying texture within the breast. In this study we employ three methods for determining texture:

- Features obtained from Spatial Grey Level Dependency (SGLD) matrices.
- Fractal dimension.
- Statistical grey-level measures.



(a) (b)

(c) (d)

Figure 3. (a) Original image mdb007 and (b) the corresponding segmented breast. (c) Original image mdb021and (d) the corresponding segmented breast.

For each texture method identified above we extract features from the original image using the segmented breast image as a map. In this way we only consider pixels in the original image within the breast region. A description of each group of texture features is given below together with parameters for their implementation.

i) Features obtained from SGLD matrices.

Second order statistics can be used to model the relationships between pixels within the breast region by constructing SGLD matrices. A SGLD matrix is the joint probability occurrence of grey levels *i* and *j* for two pixels with a defined spatial relationship in an image. The spatial relationship is defined in terms of distance *d* and angle *q*. If the texture is coarse and distance *d* is small compared to the size of the texture elements, the pairs of points at distance *d* should have similar grey levels. Conversely, for a fine texture, if distance *d* is comparable to the texture size, then the grey levels of points separated by distance *d* should often be quite different, so that the values in the SGLD matrix should be spread out relatively uniformly. Hence, a good way to analyse texture coarseness would be, for various values of distance *d*, some measure of scatter of the SGLD matrix around the main diagonal. Similarly, if the texture has some direction, i.e. is coarser in one direction than another, then the degree of spread of the values about the main diagonal in the SGLD matrix should vary with the direction *d*. Thus texture directionality can be analysed by comparing spread measures of SGLD matrices constructed at various distances *d*. From SGLD matrices, a variety of features may be extracted. The original investigation into SGLD features was pioneered by Haralick et al. [5].

Within this study each SGLD matrix is constructed at a distance of $d = 5$ and for angles $q = 0°$, $45°$, $90°$ and $135°$. A fifth SGLD matrix is constructed as the mean of all four directions. From each SGLD matrix, fifteen statistical measures are extracted. These include angular second moment (SGLD_ASM), contrast (SGLD_CONTRAST), correlation (SGLD_CORR), inverse different moment (SGLD_IDM), sum average (SGLD_SUMAVG), sum variance (SGLD_SUMVAR), sum entropy (SGLD_SUMENT), entropy (SGLD_ENTROPY), difference average (SGLD_DIFFAVG), difference variance (SGLD_DIFFVAR), difference entropy (SGLD_DIFFENT), information measure of correlation I (SGLD_MOC1), information measure of correlation II (SGLD_MOC2), inertia (SGLD_INERTIA), variance (SGLD_VARIANCE). The equations used for each feature are given in appendix A.

ii) Fractal dimension using the Hurst Coefficient

Fractal models typically relate a metric property such as line length or surface area to the scale of measurement as a basis for determining the metric property. Mandelbrot [14] illustrates this concept

using the example task of measuring a coastline with a 1km ruler laid end to end. The same procedure can be repeated with a shorter ruler and it can be seen that using a shorter ruler will result in an increased total measured length. The relationship between the ruler length and the coast length can be considered as a measure of the coastline's geometric properties, e.g. roughness. The functional relationship between the measured length L and the ruler size N can be expressed as

$$L = Nr^D \tag{1}$$

where r is a scaling constant and D is the fractal dimension. In order to estimate D, we use the Hurst coefficient described by Russ [21]. By re-writing equation (1) we can express D as:

$$D = \frac{\log N}{\log\left(\frac{1}{r}\right)} \tag{2}$$

If we plot *log(N)* against *log(1/r)*, the result should be a straight line whose slope is approximately *D*. Implementing this concept in a neighbourhood of pixels within the original image requires fitting a straight line to the log of the maximum distance (*y* co-ordinate) and the log of the distance from the central pixel (*x* co-ordinate). The slope of the line is the Hurst coefficient and replaces the pixel at the centre of the neighbour. In our experiment we use a neighbourhood of size 8 x 8 pixels. The fractal dimension *D* is calculated for every pixel in the region. The mean value over all pixels is used as the fractal dimension feature.

iii) Statistical grey-level measures

Statistical features based on the pixels making up a segmented region have been used extensively by many researchers in the field of digital mammography. In this study we use the mean (F_{MEAN}), homogeneity ($F_{HOMOGENEITY}$), standard deviation (F_{SDEV}) and skewness (F_{SKEW}) of grey-scale values of the original image, *f(x, y)*, that comprise the segmented breast region. The equations for these features are given below

$$FMEAN = \frac{1}{NM}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) \tag{3}$$

$$FHOMOGENEITY = \frac{1}{NM-1}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} (f(x,y) - FMEAN)^2 \tag{4}$$

$$FSDEV = \sqrt{FHOMOGENETY} \tag{5}$$

$$FSKEW = \frac{1}{NM} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left[\frac{f(x,y) - FMEAN}{FSD} \right]^3 \tag{6}$$

where N is the number of rows and M is the number of columns in the image $f(x, y)$.

C. Classification

For each original image we extract a set of features to classify it according to its breast type. In this study the ground truth information comprising one of three classes {Fatty, Glandular, Dense} has been decided by an expert radiologist for each image contained in the MIAS database. Using supervised learning techniques we are able to train a classifier based on a *training set* (a set of features with known classes). Subsequently we use a *test set* (a set of features with known classes) to test the performance of the classifier following the training process. The choice of classifier used is determined by the complexity of the problem, commonly dictated by the feature set. In this study we consider the following five types of classifier implemented using the SPRLIB/ANNLIB libraries [6].

- 3-Layer feed-forward artificial neural network trained with back-propagation learning ($CLASSIFIER_{BP}$).
- 3-Layer feed-forward artificial neural network trained with conjugate gradient descent ($CLASSIFIER_{CONJ}$).
- 2-Layer perceptron capable of linear classifications ($CLASSIFIER_{PERCEP}$).
- K-Nearest neighbour ($CLASSIFIER_{KNN}$).

Some of the classifiers need to be optimised according to a set of parameters. Table 1 identifies these classifiers and the parameters that should be adjusted for the purpose of optimisation. A further description of each classifier can be found in Bishop [1]. For classifiers utilising the learning rate and momentum term parameters, single optimal values are found and remain fixed throughout the training. Other parameters such as the number of neighbours and hidden nodes are varied and used to optimise the performance of the classifier detailed in section IV.

The classifiers are trained and tested using *10-fold cross-validation* to ensure exhaustive testing with all samples. Using this technique we divide the data set at random into a set of $S = 10$ distinct sets. Training is then performed on $S - 1$ of the sets and the remaining set is tested. This is then repeated for all of the possible S disjoint training and test sets. The resulting test errors are averaged over all S results. To prevent over-fitting of the training data, the classifiers $CLASSIFIER_{BP}$, $CLASSIFIER_{CONJ}$ and $CLASSIFIER_{PERCEP}$ are trained using an independent validation set is used to identify the point at which training is to stop. This is typically the point at which the validation error starts to increase.

IV. EXPERIMENTAL METHOD

A. Data Description

Within this study we use the MINIMIAS database available from the Mammographic Image Analysis Society. The database contains 322 oblique-view mammograms scanned at a pixel-edge of 50 microns and padded to a size of 1024 x 1024 pixels. All images are 8-bit grey-scale. Each image has been classed by an expert radiologist as one of three breast types namely fatty, glandular or dense.

Classifier Type	**Parameter**	**Value**
$CLASSIFIER_{KNN}$	Number of neighbours k.	k ={1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51}
$CLASSIFIER_{BP}$	Number of hidden nodes h. Momentum term in training m. Learning rate in training h.	$h = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ $m = 0.5$ $h = 0.01$
$CLASSIFIER_{PERCEP}$	None.	
$CLASSIFIER_{CONJ}$	Number of hidden nodes h. Momentum term in training m. Learning rate in training h.	$h = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ $m = 0.5$ $h = 0.01$

Table 1. Classifiers and their associated parameters.

B. Feature Sets

In this study two classification problems are investigated. In the first experiment we perform a 3-way classification of breast according to the three breast types {Fatty, Glandular, Dense}. In the second, we treat all 'glandular' breasts and 'dense' breasts as one group of dense breasts giving a two-class classification problem {Fatty, Dense}.

C. Parameter Settings

For each classification problem {2-class, 3-class}, we determine the optimal parameter settings for the classifiers listed in Table 1 through experimentation. The optimal parameters are those that maximise the recognition rate obtained using the classifier. For two of the classifiers using supervised training with gradient descent, a momentum and learning rate parameter may be set. These values were not optimised but set following initial experimentation on the data set.

D. Performance Evaluation

In our experiments we select the parameterised model that maximises the recognition rate *(r)* on test set with the least complexity, that is, a neural network model with the least number of hidden nodes or *k*-nearest neighbour model with the smallest neighbourhood. We aim to select the simplest classifier model without any loss in generalisation ability. Performance is evaluated on the basis of generalisation success measured as recognition rate (*r*) on test.

We use ROC analysis to evaluate the performance of the optimised classifiers in the two-class classification problem. ROC analysis is a statistical technique that provides information about the overlap within a two-class classification. The analysis allows for a plot of the sensitivity or True Positive Fraction (TPF) against the specificity or False Positive Fraction (FPF) at differing thresholds so that inaccuracies, which arise from assuming findings are absolutely normal or abnormal, are avoided. We evaluate the performance of each optimised classifier in the two-class classification problem. To evaluate the TPF and FPF, we define a positive case as the detection of breasts with 'glandular'/'dense' tissue. Table 2 lists the possible outcomes for different breast type classifications.

On the basis of this terminology, we can evaluate the performance of a CAD technique by calculating True Positive Fraction and False Positive Fraction and using ROC analysis. These fractions are defined as

$$TPF = \frac{TP}{TP + FN} \qquad FPF = \frac{FP}{FP + TN} \qquad (7, 8)$$

A quantitative measure of the accuracy of the classification technique is obtained by finding the area under the ROC curve termed *Az* This varies between 0.0, indicating poor classification performance, and 1.0 indicating high classification performance. This measure is sometimes preferable over the recognition rate (r), as a true indication of the classifier's performance, as it considers the correct classification of a breast as 'glandular'/'dense' against the misclassification of a 'fatty' breast as such type. This could be of significance if we are building a Computer Aided Detection (CAD) system that attempts to filter out difficult mammographic screening cases. Typically such cases comprise 'glandular'/'dense' breast types that would be best examined by an expert radiologist. The residual 'fatty' type breast is easier to screen and can be automatically processed.

Outcome	Description
True Positive (TP)	Breast classified as type 'glandular'/'dense' that prove to be type 'glandular'/'dense'.
False Positive (FP)	Breast classified as type 'glandular'/'dense' that prove to be type 'fatty'.
False Negative (FN)	Breast classified as 'fatty' that prove to be 'glandular'/'dense'.
True Negative (TN)	Breast classified as 'fatty' that prove to be 'fatty.

Table 2. Possible outcomes for different breast type classifications.

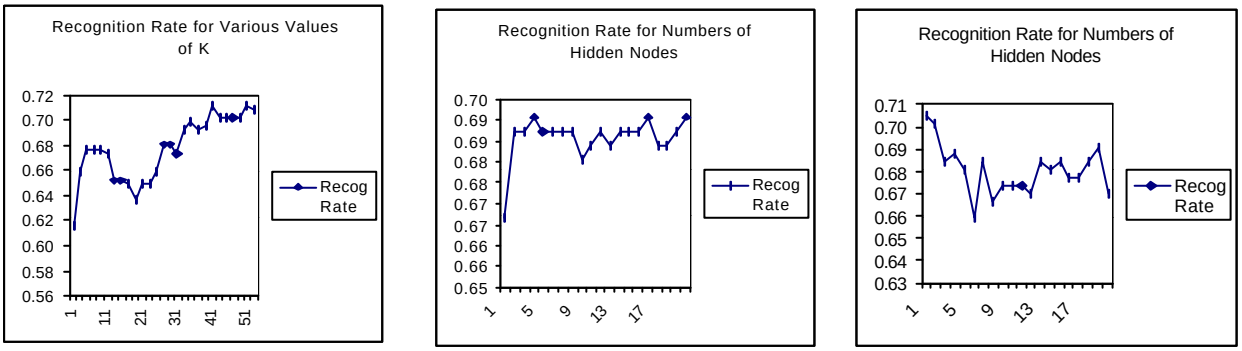
V. RESULTS AND DISCUSSION

We consider the results for the three-class and two-class problems separately.

A. Classifying Breast Types Dense, Glandular and Fatty

For $CLASSIFIER_{BP}$, $CLASSIFIER_{CONJ}$ and $CLASSIFIER_{KNN}$, we show the recognition rate (r) on test each model of classifier in figure 4. Figure 5 shows the best results obtained for each of the optimised classifiers.

For classifier $CLASSIFIER_{KNN}$, we observe that the recognition rate is maximised when $K = 41$ where $r = .71$ (equal to 71%). The classifiers $CLASSIFIER_{BP}$ and $CLASSIFIER_{CONJ}$ are optimised on the basis of the number of hidden nodes (h). Both classifiers were trained and tested varying the number of hidden nodes according to Table 1. The best recognition (r) on test for $CLASSIFIER_{BP}$ was found with $h = 4$ giving $r = 0.69$ and $CLASSIFIER_{CONJ}$ gave the best recognition rate on test at $h = 1$ such that $r = 0.70$. The classifier $CLASSIFIER_{PERCEP}$ has no optimisation parameters and the classifier gave the worst performance. The simple perceptron classifier is only capable of drawing linear boundaries in the feature space for classification and $CLASSIFIER_{PERCEP}$ gave $r = 0.68$.



(a) (b) (c)

Figure 4. Recognition rate on test (r), (a) for various values of K for $CLASSIFIER_{KNN}$; (b) for various number of hidden nodes for $CLASSIFIER_{BP}$, (c) for various number of hidden nodes for $CLASSIFIER_{CONJ.}$

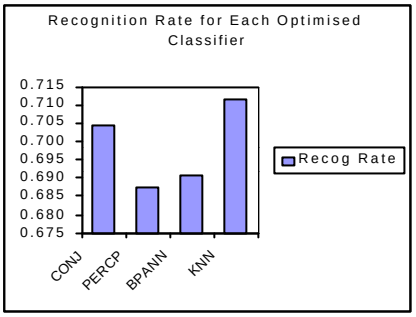


Figure 5. Recognition rate *(r)* on test for optimised classifiers.

Figure 6 shows a plot of the first two principal components extracted from all of the features. It is clear from the scatter plot that the three classes overlap extensively and as such will lead to poor Results using a linear-based method.

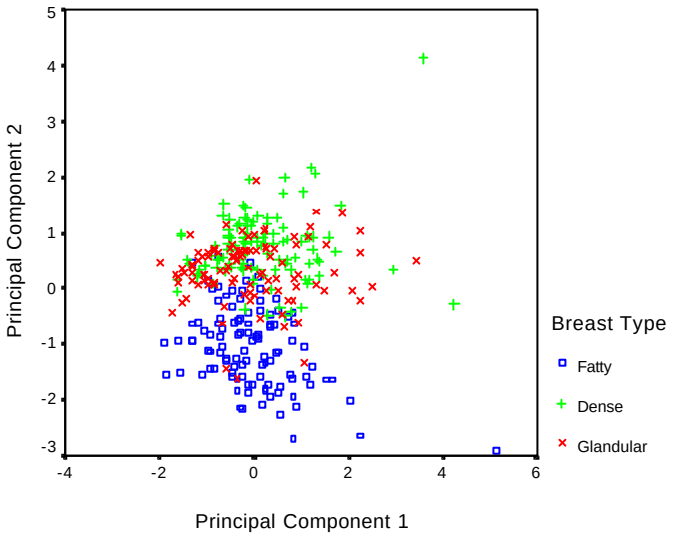


Figure 6. Scatter plot of first two principal components (PC1and PC2).

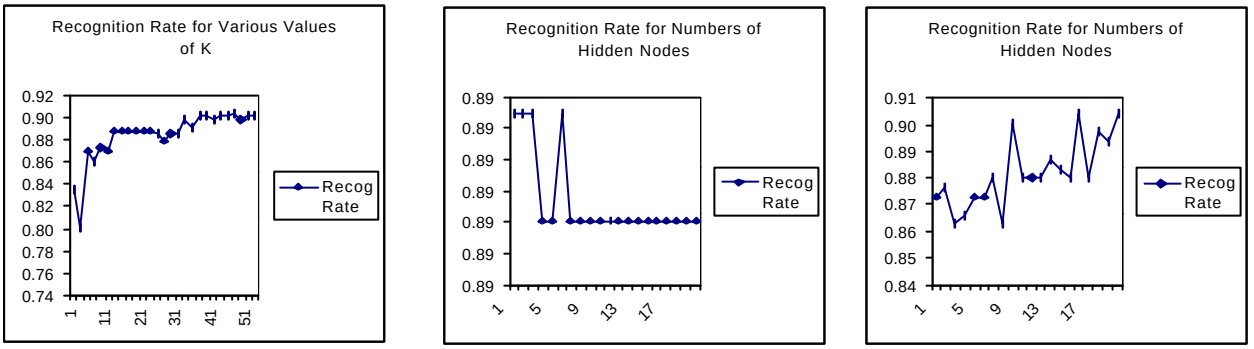
B. Classifying Breast Types Fatty and Dense

For the two-class problem we optimise classifiers where applicable. We show the recognition rate on test for each classifier in figure 7. Figure 8 shows the best results obtained for each of the optimised classifiers.

The overall performance for all classifiers improved as expected with a simplification of the problem. For classifier $CLASSIFIER_{CONJ}$, we observe that the recognition rate on test is maximised when $h = 16$

such that $r = 0.90$. The improvement in performance due to problem simplification feature set was also apparent with classifier $CLASSIFIER_{KNN}$. In optimising, the classifier gave a recognition rate on test for $k = 41$ where $r = 0.90$.

The results obtained using the classifier $CLASSIFIER_{PERCEP}$ are noticeable. The perceptron was actually comparable to the optimised $CLASSIFIER_{BP}$. The improved performance of $CLASSIFIER_{PERCEP}$ can be explained by viewing the scatter plot of the first two principal components extracted from all of the features in Fig 8. Clearly the class overlap is reduced and as such will lead to improved performance by any linear-based method. Classifier $CLASSIFIER_{PERCEP}$ recorded a value $r = .89$ and $CLASSIFIER_{BP}$ gave a similar performance with $r = .89$ when $h = 1$.



(a) (b) (c)

Figure 7. Recognition rate (r) on test (a) for various values of K for $CLASSIFIER_{KNN}$; (b) for various number of hidden nodes for $CLASSIFIER_{BP}$, (c) for various number of hidden nodes for $CLASSIFIER_{CONJ.}$

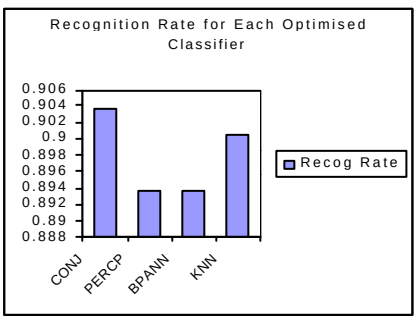


Figure 8. Recognition rate (r) on test for optimised classifiers

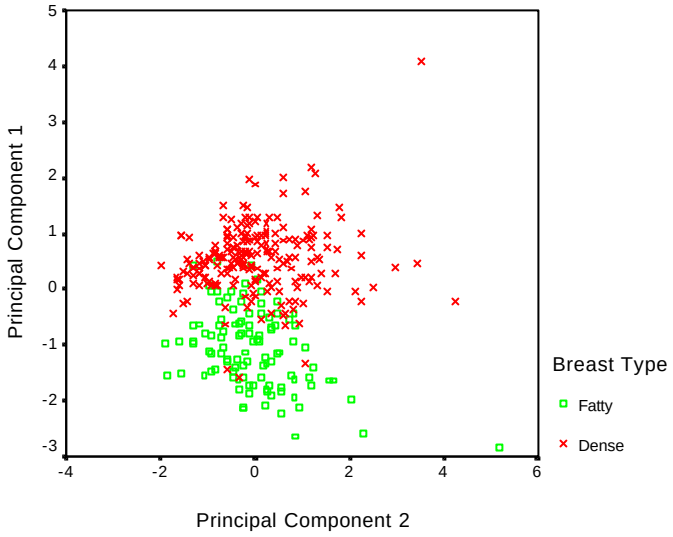


Figure 9. Scatter plot of first two principal components (PC1 & PC2).

C. Evaluation of Two-Class Classification with ROC Analysis

Figure 10 shows the constructed ROC curves for the detection of 'glandular'/'dense' breast types for each of the optimised classifiers on the data set. The diagonal line indicates the expected performance that is obtained for a 50:50 guess. Table 3 shows the values of *Az* obtained for each classifier. When evaluating the performance in terms of the sensitivity and specificity of each classifier in the detection of 'glandular'/'dense' breast types, $CLASSIFIER_{KNN}$ performs the best. This is in sharp contrast to evaluating the performance solely on recognition rate on test. This can be explained by the method used to calculate the posterior probability $P(C_k \mid x)$ for $CLASSIFIER_{KNN}$, and subsequently leads to a reduced *Az* value. In this study we use the *K-nearest-neighbour classification rule* [1] given as the ratio of C_n / K where C_n is the number of points of class C in the neighbourhood of size k. $CLASSIFIER_{KNN}$ gives an increase of approximately 32% in the TPF from 0.50 to 0.66 with a little change in the FPF as shown in figure 8(a) and (c) and accounts for the greater value of *Az* over $CLASSIFIER_{CONJ}$, the classifier with the best recognition rate on test. The smaller values obtained for the posterior probability indicate that the classifier $CLASSIFIER_{CONJ}$ appears to become more confused with the feature set (*Az* = 0.806) compared to $CLASSIFIER_{KNN}$ (*Az* = 0.832).

VI. CONCLUSIONS

We have investigated the use of co-occurrence and fractal texture features for the classification of breast tissue type and demonstrated its application by classifying segmented breast regions on mammograms. By evaluating a variety of classifiers we have been able to maximise the recognition rate on test for two classification problems. For the three-class problem of classifying 'dense', 'glandular' and 'fatty', our proposed system provides a baseline from which we aim to improve further. We have quoted the results obtained on a ten fold cross-validation. Obviously, if we use a leave-one-out strategy for validation, we get more optimistic results since we have more training data. Only kNN classifier in our analysis qualifies for a leave-one-out validation. Using this approach for the two class problem, we achieve a recognition rate of r=0.91 for k=11 with Az value of 0.873, and for the three class problem we achieve a recognition rate of r=0.70 when k=3. These results are much better than using ten fold-cross-validation.

Our results reflect the difficulty in differentiating between subtle glandular and dense tissue types. Addressing the two-class problem of classifying 'glandular/dense' and 'fatty', we have demonstrated the high sensitivity within in our results justifying the use of this technique in a proposed CAD system. Future work will investigate the use of this prior knowledge of the breast type in a CAD system at different stages including enhancement, segmentation and feature extraction. The application of this technique in this way aims to increase the sensitivity of detecting breast cancer.

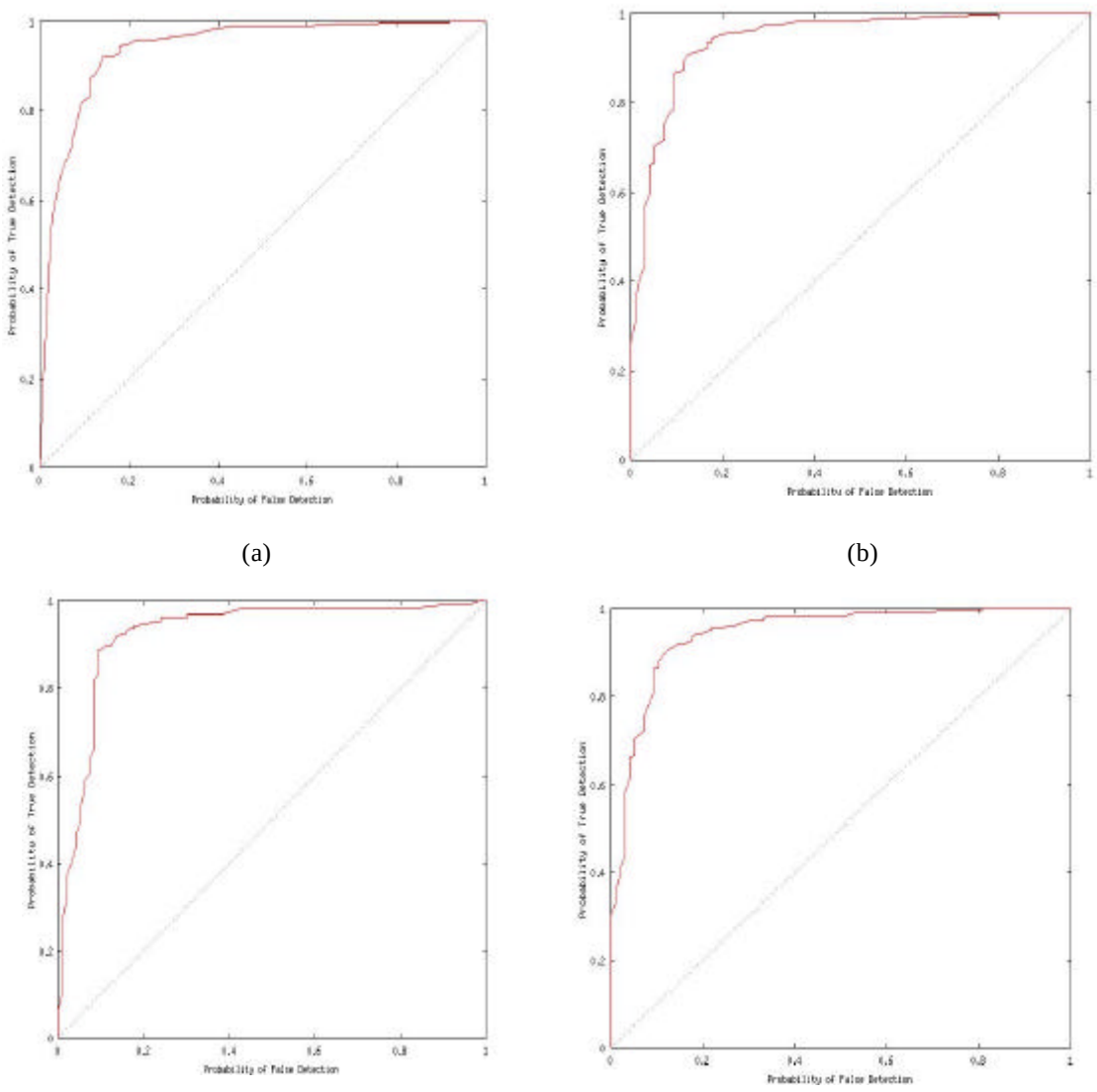


Figure 10. ROC curves for the detection of 'glandular'/'dense' breast types (a) $CLASSIFIER_{KNN}$, (b) $CLASSIFIER_{BP}$, (c) $CLASSIFIER_{CONJ}$, (d) $CLASSIFIER_{PERCEP}$.

Classifier	Data set F_{ALL} (Az)
$CLASSIFIER_{KNN}$	0.832
$CLASSIFIER_{BP}$	0.764
$CLASSIFIER_{PERCEP}$	0.765
$CLASSIFIER_{CONJ}$	0.806

Table 3. Area under the ROC curve (Az) for each classifier for all and selected feature data sets.

REFERENCES

1. C. Bishop, *Neural networks for pattern recognition*, Clarendon Press, New York, 1997.
2. H. P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. A. Goodsitt and D. D. Adler, Computerised analysis of mammographic microcalcifications in morphological and texture feature space, Medical Physics, vol. 25, issue 10, pp. 2007-2019, 1998.
3. R. Chandrasekhar, Y. Attikiouzel, A simple method for automatically locating the nipple on mammograms, IEEE Transactions on Medical Imaging, vol. 16, issue 5, pp. 483-494, 1997.
4. E. J. Feuer, L. M. Wun, DEVCAN: Probability of Developing or Dying of Cancer. Version 4.0. Bethesda MD: National Cancer Institute. 1999.
5. R. M. Haralick, K. Shanmugam and I. Dinstein, Textural features for image classification, IEEE Transactions on System, Man, and Cybernetics, vol. SMC-3, pp. 610-621, 1973.
6. A. Hoekstra, M. A. Kraaijveld, D. de Ridder, W. F. Schmidt and A. Ypma, SPRLIB – Statistical Pattern Recognition and Artificial Neural Network Library, Pattern Research Group, Department of Applied Physics, Delft University of Technology, Delft, The Netherlands, 1998.
7. N. Karssemeijer, Stochastic model for automated detection of calcifications in digital mammograms, Image and Vision Computing, vol. 10, issue 6, pp. 369-375, 1992.
8. J. K. Kim and H. W. Park, Statistical texture features for detection of microcalcifications in digitised mammograms, IEEE Transactions in Medical Imaging, vol. 18, issue 3, pp. 231-238, 1999.
9. R. Klette, P. Zamperoni, Handbook of Image Processing Operators, John Wiley & Sons, pp120 – 124.
10. H. Kobatake, M. Murakami, H. Takeo, S. Nawano, Computerised detection of malignant tumours on digital mammograms, IEEE Transactions on Medical Imaging, vol. 18, issue 5, pp. 369-378, 1999.
11. D. Kopans, *Breast imaging*, 2nd Edition, Lippincott-Raven, Philadelphia, 1998.
12. T. Lau and W. F. Bischof, Automated detection of breast tumours using the asymmetry approach, Computers and Biomedical research, vol. 24, pp. 273-295, 1991.
13. L. Li, W. Qian and L. P. Clarke, Digital mammography: computer assisted diagnosis method for mass detection with multiorientation and multiresolution wavelet transforms, Academic Radiology, vol. 4, issue 11, pp. 724-731, 1997.

14. B. B. Mandlebrot, *The fractal geometry of nature*, Freeman, New York, 1982.

15. A. J. Mendez, P.G. Tahoces, M.J. Lado, M. Souto and J.J. Vidal, Computer aided diagnosis: automatic detection of malignant masses in digitised mammograms, Medical Physics, vol. 25, issue 6, pp. 957-964, 1998.

16. C. E Metz, Basic principles of ROC analysis, Seminars in Nuclear Medicine, vol. 8, issue 4, pp. 283-298, 1978.

17. P. Miller and S. Astley, Classification of breast tissue by texture analysis, Image and Vision Computing, vol. 10, pp. 277-283, 1992.

18. NHS Breast Screening Program, Review 1999, NHS Breast Screening Program. 1999.

19. S. L. Ng and W. F. Bischof, Automated detection and classification of breast tumours, Computers and Biomedical Research, vol. 25, pp. 218-237, 1992.

20. N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie and D. D. Adler, Automated detection of breast masses using adaptive contrast enhancement and texture classification, Medical Physics, vol. 23, issue 10, pp. 1685-1696, 1996a.

21. J. C. Russ, Surface Characterisation: Fractal Dimensions, Hurst Coefficients, and Frequency Transforms, Journal of Computer Assisted Microscopy, Vol. 2, pp. 249-257, 1990.

22. 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, vol. 1069, pp. 375-378, 1994. http://www.wiau.man.ac.uk/services/MIAS/MIAScom.html.

23. P. Taylor, S. Hajnal, M-H Dilhuydy, B. Barreau, Measuring image texture to separate *difficult* from *easy* mammograms, The British Journal of Radiology, vol. 67, pp. 456-463, 1994.

24. R. Zwiggelaar, S. Astley, C. Taylor, Detecting the central mass of a spiculated lesion using scale-orientation signatures, Proc. 4th International Workshop on Digital Mammography, pp. 63-70, Nijmegen, Holland, 1998.

25. R. Zwiggelaar, T.C. Parr, J.E. Schumm, I.W. Hutt, C.J. Taylor, S.M. Astley and C.R.M. Boggis, Model-based detection of spiculated lesions in mammograms, Medical Image Analysis, vol. 3, issue 1, pp. 39-63, 1999.

APPENDIX A. – TEXTURE FEATURES EXTRACTED FROM SGLD MATRICES

Assuming the following nomenclature:

$p(i, j) = p(i, j)/R$	(i, j)th entry in a normalised SGLD matrix, $p(i, j, d,q)$ where i, j are the grey scale values of pixels at distance d pixels apart, and angle q is the angle of the line joining the centres of these pixels in the creation of the SGLD matrix.
$p_x(i) = \sum_{j=1}^{N_g} p(i, j)$	i th entry in the marginal-probability matrix obtained by summing the rows of $p(i, j)$.
N_g	Number of grey levels in the image.
$p_y(j) = \sum_{i=1}^{N_g} p(i, j)$	jth entry in the marginal-probability matrix obtained by summing the columns of $p(i, j)$.
$p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i, j); k = 2,3,...,2N_g$	The result of adding the ith and jth entries calculated above from the marginal-probability matrix.
$p_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i, j); k = 0,1,2,...,2N_g - 1$	The result of subtracting the ith and jth entries calculated above from the marginal-probability matrix.

1. Angular Second Moment

$$SGLD\_ASM = \sum_{i=1}\sum_{j=1} (p(i, j))^2$$

2. Contrast

$$SGLD\_CONTRAST = \sum_{n=0}^{N_g-1} n^2 \left(\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i, j)\right)$$

3. Correlation

$$SGLD\_CORR = \frac{\sum_i\sum_j (i, j)\, p(i, j) - m_x m_y}{s_x s_y}$$

where $m_x = \sum_{i=0}^{n-1} i p_x(i)$ $\quad s_x^2 = \sum_{i=0}^{n-1} (i - m_x)^2 p_x(i)$

$$m_y = \sum_{j=0}^{n-1} j p_y(j)$$

$$s_y^2 = \sum_{j=0}^{n-1} (j - m_y)^2 p_y(j)$$

4. Variance

$$SGLD\_VARIANCE = -\sum_i \sum_j (i-n)^2 p(i,j)$$

5. Inverse Different Moment

$$SGLD\_IDM = -\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$$

6. Sum Average

$$SGLD\_SUMAVG = \sum_{i=2}^{2N_g} i p_{x+y}(i)$$

7. Sum Variance

$$SGLD\_SUMVAR = \sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$$

8. Sum Entropy

$$SGLD\_SUMENT = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log(p_{x+y}(i))$$

9. Entropy

$$SGLD\_ENTROPY = -\sum_i \sum_j p(i,j) \log(p(i,j))$$

10. Difference Average

$$SGLD\_DIFFAVG = \sum_{i=2}^{2N_g} i p_{x-y}(i)$$

11. Difference Variance

$SGLD\_VARIANCE$ = variance of p_{x-y}

12. Difference Entropy

$$SGLD\_DIFFENT = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \log(p_{x-y}(i))$$

13. Information Measure of Correlation I

$$SGLD\_MOC1 = \frac{HXY - HXY1}{\max(HX, HY)} \; where$$

$$HXY = -\sum_i \sum_j p(i,j) \log(p(i,j)) \; and$$

$$HXY1 = -\sum_i \sum_j p(i,j) \log(p_x(i) p_y(j)) \; and$$

$$HXY2 = -\sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$$

and HX and HY are entropies of p_x, p_y.

14. Information Measure of Correlation II

$$SGLD\_MOC2 = (1 - \exp[-2.0(HXY2 - HXY)])^{\frac{1}{2}} \; where$$

$$HXY = -\sum_i \sum_j p(i,j) \log(p(i,j)) \; and$$

$$HXY1 = -\sum_i \sum_j p(i,j) \log(p_x(i) p_y(j)) \; and$$

$$HXY2 = -\sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$$

and HX and HY are entropies of p_x, p_y.

15. Inertia

$$SGLD\_INERTIA = \sum_i \sum_j (i - j)^2 p(i,j)$$

SUMMARY

The classification on mammograms on the basis of density is important for understanding the difficulty in screening and diagnosing them for breast cancer. Traditional BIRADS system is used to manually assign mammograms into one of the four chosen categories based on visible breast density. In this paper we have investigated the use of texture measures to train a classifier system to do this assignment automatically. Texture measures based on co-occurrence matrices, fractal dimension and statistical gray level values are used. We find these measures are extremely used for classification and measure the quality of our classification scheme on the basis of manually generated ground truth labelling. The classification process has been tested with a total of four different classifiers. We find that conjugate gradient based neural network model and nearest neighbour classifier give the best results. We evaluate the classifier performance on the basis of cross-validation and use of ROC statistics.