Classification of Benign and Malignant Patterns in Digital Mammograms for the Diagnosis of Breast Cancer

Brijesh Verma¹, Peter McLeod¹ and Alan Klevansky²

¹School of Computing Sciences, Central Queensland University

Rockhampton, QLD 4701, Australia

b.verma@cqu.edu.au, peterm@practical.com.au

²Radiology Department, Gold Coast Hospital

Gold Coast, QLD 4215, Australia

Abstract

The classification of benign and malignant patterns in digital mammograms is one of most important and significant processes during the diagnosis of breast cancer as it helps detecting the disease at its early stage which saves many lives. Breast abnormalities are often embedded in and camouflaged by various breast tissue structures. It is a very challenging and difficult task for radiologists to correctly classify suspicious areas (benign and malignant patterns) in digital mammograms. In the early stage, the visual clues are subtle and varied in appearance, making diagnosis difficult; challenging even for specialists. Therefore, an intelligent classifier is required which can help radiologists in classifying suspicious areas and diagnosing breast cancer. This paper investigates a novel soft clustered based direct learning classifier which creates soft clusters within a class and learns using direct calculation of weights. The feature space for suspicious areas in digital mammograms from same class patterns can have multiple clusters and the proposed classifier uses this fact and introduces a novel idea to create soft clusters for each available class and applies them to form subclasses within benign and malignant classes. A novel learning process based on direct learning is introduced. The experiments using the proposed classifier have been conducted on a benchmark database. The results have been analysed using ANOVA test which showed that the results are statistically significant.

1. INTRODUCTION

Screening mammography has long been regarded as the gold standard in detecting breast cancer in women. Mammography has been shown to reduce the mortality rate by as much as 41% according to one South Australian study [1]. However various studies have demonstrated that an estimated 11-25% of breast cancers are missed [2] during screening mammography.

In the United States alone an estimated 240,510 new cases of breast cancer together with an estimated mortality of 40,460 during 2007 will occur [3]. Survival from breast cancer is dependant on the stage at which it is detected and the implementation of appropriate treatment.

Early stage detection and treatment results in a 98% survival rate however this plummets to 28% if metastases have spread to distant organs [4]. In the screening process, radiologists carefully search each mammogram for any visual sign of abnormality. However, in the earliest stage the visual clues are subtle and varied in appearance, making detection/diagnosis difficult; challenging even for specialists. The abnormalities are often embedded in and camouflaged by various breast tissue structures.

An abnormal growth of tissues in the breast creates lumps or tumors in the breast. Tumors perform no useful body function and grow at the expense of healthy tissues. Mammography is capable of detecting such features that may indicate a potential clinical problem, which include asymmetries between the breasts, architectural distortion, confluent densities associated with benign fibroses, calcifications and masses. Mammographic abnormalities in breast cancer can be characterized into two major classes; calcifications and masses as shown in Figure 1.

Calcifications are small mineral (calcium) deposits within breast tissues and appear as localized high-intensity regions (bright spots) on mammograms. They can be produced from cell secretion or from necrotic cellular debris. There are two types of calcifications: microcalcifications and macrocalcifications. Macrocalcifications are coarse (large) calcium deposits that are often associated with benign (non-cancerous) conditions and do not usually require a biopsy. Microcalcifications are tiny calcium specks and may be isolated, appear in cluster, or found embedded in a mass. In general, individual microcalcifications are found in size range of 0.1 – 1.0 mm with an average diameter of about 0.5 mm. A cluster is typically defined by the presence of at least three microcalcifications within a 1 cm² region. Microcalcifications are one of the mammographic hallmarks of early breast cancer. About 25% of all breast cancer is detected by the presence of microcalcification clusters. The majority of ductal carcinoma in situ (DCIS) cancers are associated with microcalcifications.

Figure 1. Breast Abnormalities

A breast mass is a localized sign of breast cancer and defined as a space-occupying lesion seen in at least two different projections. A mass may occur with or without associated calcifications. Figure 1 (b) shows circumscribed and spiculated masses. The common circumscribed mass is Fibroadenoma and usually found in younger age women. Spiculated lesions having stellate appearance on mammograms are high probability suspicious indicators of breast cancer. Generally it has a central tumor mass with spicules extending into surrounding tissues. The subtlety and camouflaged nature of masses make their detection far more challenging than calcifications in mammograms.

Masses are described by their location, size, shape, and margin characteristics as shown in Figure 2. These morphologic features are the key factors to be considered by the radiologist when judging the likelihood of cancer being present. A wide range of shapes, sizes and contrast is found in masses. Generally benign masses are well circumscribed, compact, and roughly circular or elliptical. Masses having irregular shapes and spiculated or indistinct margins suggest a higher possibility of malignancy.

Figure 2. Mass shapes and margins

Different studies have demonstrated that Computer Aided Diagnosis (CAD) of breast cancer can improve the detection rate from 4.7% to 19.5% compared to radiologists [5-9]. Various solutions have been proposed for improving the breast cancer diagnosis. In our knowledge, this is first study which investigates sub-clusters within benign and malignant masses separately and direct learning for improving overall classification accuracy. A review of existing literature in this area is presented in Section 2.

The rest of paper is organised as follows. Section 2 presents a review of existing techniques. Section 3 describes the proposed soft clustered based direct learning classifier. Section 4 presents the criteria for performance measure. Section 5 presents the experimental results obtained on a benchmark database. Section 6 presents a comparison of the results obtained with that of other researchers and section 7 presents the conclusions and future research directions.

2. RELATED WORK

The research in the area of CAD systems for breast cancer and developing intelligent techniques for improving classification accuracy have been conducted in last few decades. A comprehensive review of existing techniques for the detection and classification of masses in digital mammograms has been conducted and presented in [10-11].

Ramirez et al. [12] used seven Bayesian network classifiers for the diagnosis of breast cancer on two real-world databases. They used the breast lesion cases collected by a single observer and multiple observers and obtained an accuracy of 93.04% and 83.31% respectively.

Tourassi et al. [13] evaluated image similarity measures for detection of masses in mammograms. They used database of 1820 mammographic regions of interest and compared eight entropy-based similarity measures. They concluded that a substantial reduction in false positive while maintaining high detection rate for malignant masses was achieved.

Manrique et al. [14] utilised a genetic algorithm based radial basis function neural network for classification of masses from a Madrid hospital dataset. They obtained 83% classification accuracy (with 83% specificity and 81% sensitivity). Although their accuracy was not high, their network converged quickly.

Halkiotis et al. [15] used a MultiLayer Perceptron (MLP) type neural network on the MIAS database. They obtained a good classification rate of 94.7% with an average of 0.27 false positives per image for microcalcifications.

Georgiou et al. [16] evaluated a mass shape feature with a wide range of linear and non-linear classifiers, including linear discriminant analysis (LDA), least-squares minimum distance (LSMD), k-nearest neighbor (k-NN), radial basis function (RBF) and multi-layered perceptron (MLP) neural networks (NN), and support vector machines (SVM). They found that support vector machine produced highest accuracy. They obtained 91.54% classification accuracy on masses.

Brem [17] used the second look CAD system (version 3.4) to determine the performance of CAD systems on different sized lesions. He achieved an overall sensitivity of 89%. Brem's investigation was to try and determine if lesion size would adversely affect the performance of a CAD system.

Abdalla et al. [18] used textual features with a support vector machine classifier and they achieved a classification accuracy of 82.5% on mammograms from the Digital Database of Screening Mammography (DDSM) [19].

Panchal et al. [20] proposed an autoassociator network for feature extraction and combined it with an MLP based classifier. The network was trained using a back propagation algorithm. They achieved 91% classification accuracy on mammograms from the DDSM.

Masotti [21] used a support vector machine and obtained 90% classification accuracy on masses. Archarya [22] obtained a sensitivity of 91.67% using an artificial neural network and 95% using a Gaussian Mixture Model with 93.33% sensitivity and 96.67% specificity respectively on mammograms from the DDSM.

In a review of CAD techniques Rangayyan et al. [23] noted that several methods have achieved over 85% sensitivity for the identification of masses but also have a high false positive rate. In general mass identification is a more difficult task than microcalcifications because masses are variable in size, shape and density can exhibit poor image contrast and can be strongly intertwined with surrounding tissues making detection and classification difficult [24]. Micro calcifications tend to be of a higher density and are more readily detected.

Among existing CAD techniques, artificial neural networks have demonstrated their capabilities, but despite this, successful commercial CAD systems are not available. The main problem of developing an acceptable CAD system is inconsistent and low classification accuracy. In order to improve the training process and accuracy, this paper investigates a novel soft clustered based direct learning classifier that uses clustering to create soft clusters (sub classes) within existing classes (benign and malignant) and direct learning for finding appropriate weights during the training process.

3. SOFT CLUSTERED BASED DIRECT LEARNING (SCBDL) CLASSIFIER

3.1 Overview of SCBDL Classifier

The soft clustered based direct learning classifier (SCBDL) is a new type of neural classifier which uses clustering for creating clusters (sub-classes) within a benign/malignant class and direct calculation of weights for fast learning. The notion of soft clustered network has been proposed and investigated in this research since a Radiologist utilises many features to make a diagnostic appraisal as to the likelihood that an abnormality in digital mammograms is benign or malignant. Since any one feature is not a sufficient discriminates in its own right, and that many features are needed to allow a diagnosis to occur it means that the abnormalities can form into complex groups and subgroups within groups. This is due to the subtle nature and lack of distinctiveness between the benign and malignant abnormalities which does not readily facilitate diagnosis. By creating soft clusters within a class instead of utilising only one single binary class for benign or malignant abnormality, the training process and classification accuracy of the system can be improved. Attempting to impose a single class on benign and

malignant groups and then utilising this artificially imposed grouping in training the weights of a classifier leads to the development of sub-optimal generalisation as conflicting weights might be generated during training. However by recognising that multiple clusters exist within benign and malignant patterns; and incorporating these clusters into the output of the classifier, the resulting weights should be less conflicting in the training of the classifier with the subsequent consequence of an improvement in classification accuracy. The notion of direct learning has been combined with soft clusters in developing the proposed classifier which means that the weights of learning network have been assigned or calculated by using Gram-Schmidt orthogonalization and least squares.

The theory that soft clustered based direct learning classifier will provide statistically significant improvement in classification accuracy, we have defined null and alternate hypothesis as follows.

H0:μ The utilisation of soft clustered based direct learning classifier makes no difference to the classification of benign and malignant classes when compared to existing classifiers such as MLP.

H1:μ The utilisation of soft clustered based direct learning classifier makes an improvement in the classification of benign and malignant patterns when compared to existing classifiers such as MLP.

3.2 Mathematical Description of SCBDL Classifier

Let X (x₁, x₂, ..., x_n) be the input, Y (y₁, y₂, ... y_m) be the output, C (c₁, c₂ ..., c_k) be the number of clustered output, W be the weight matrix of hidden layer and W' be the weight matrix of the output layer.

The classifier network is defined as follows.

$$C = f(X, Y) = f_0(f_c(f_h(X * W) * W') * W'')$$

Where,

$$X = \begin{bmatrix} x_{11} & x_{12} & : & x_{1n} \\ x_{21} & x_{22} & : & x_{2n} \\ : & : & : & : \\ x_{m1} & x_{m2} & : & x_{mn} \end{bmatrix}, \quad W = \begin{bmatrix} w_{11} & w_{12} & : & w_{1n} \\ w_{21} & w_{22} & : & w_{2n} \\ : & : & : & : \\ w_{h1} & w_{h2} & : & w_{hn} \end{bmatrix}, \quad W' = \begin{bmatrix} w'_{11} & w'_{12} & : & w'_{1h} \\ w'_{21} & w'_{22} & : & w'_{2h} \\ : & : & : & : \\ w'_{c1} & w'_{c2} & : & w'_{ch} \end{bmatrix}$$

$$W'' = \begin{bmatrix} 1 & 1 & 1 & : & 0 & 0 & 0 \\ 0 & 0 & 0 & : & 1 & 1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ : \\ c_1 \\ c_2 \\ c_3 \\ : \end{bmatrix} = \begin{bmatrix} cluster_output1 - benign \\ cluster_output2 - benign \\ cluster_output3 - benign \\ : \\ cluster_output1 - malignant \\ cluster_output2 - malignant \\ cluster_output3 - malignant \\ : \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} benign \\ malignant \end{bmatrix}$$

Let X_b be the benign training patterns from X ($X_b \in X$) and X_m be the malignant training patterns from X ($X_m \in X$).

The patterns X_b and X_m are clustered into k clusters as described below.

1. Place k points into the space represented by the mammographic abnormalities that are being clustered. These points represent the initial group centroids.
2. Assign each input data (abnormality) to a group that has the closest centroid by measuring distance (d) as follows.

$$d_{benign} = \sum_{j=1}^k \sum_{i=1}^n \|x_{ibenign}^{(j)} - c_{jbenign}\|^2 \quad d_{malignant} = \sum_{j=1}^k \sum_{i=1}^n \|x_{imalignant}^{(j)} - c_{jmalignant}\|^2$$

3. When all patterns have been assigned recalculate the positions of the k centroids.
4. Repeat steps 2 and 3 until the centroids no longer move.

Using the k clusters for benign and k clusters for malignant patterns, input X for malignant and benign is divided into subgroups. They represent the following subsets.

$$C_m[c_1, c_2, \dots, c_k] = \text{cluster}(X_m)$$

$$X_{mc1} \in X_m \in X \quad X_{mc2} \in X_m \in X \quad X_{mck} \in X_m \in X$$

$$C_b[c_1, c_2, \dots, c_k] = \text{cluster}(X_b)$$

$$X_{bc1} \in X_b \in X \quad X_{bc2} \in X_b \in X \quad X_{bck} \in X_b \in X$$

The network output

$$C(c_1, c_2, \dots, c_k) = f_c(X, Y) = Y[y_1 \dots y_{n=2k}] = \{X_{mc1} \in X_m \dots X_{mck} \in X_m, X_{bc1} \in X_b, \dots X_{bck} \in X_b\} [XW]W' = Y[y_1 \dots y_{n=2k}]$$

Therefore,

$$Y = f_o \left(f_c \left(f_h \left(\begin{bmatrix} x_{11} & x_{12} & : & x_{1n} \\ x_{21} & x_{22} & : & x_{2n} \\ : & : & : & : \\ x_{m1} & x_{m2} & : & x_{mn} \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & : & w_{1n} \\ w_{21} & w_{22} & : & w_{2n} \\ : & : & : & : \\ w_{h1} & w_{h2} & : & w_{hn} \end{bmatrix} \right) * \begin{bmatrix} w'_{11} & w'_{12} & : & w'_{1h} \\ w'_{21} & w'_{22} & : & w'_{2h} \\ : & : & : & : \\ w'_{c1} & w'_{c2} & : & w'_{ch} \end{bmatrix} \right) * \begin{bmatrix} 1 & 1 & 1 & : & 0 & 0 & 0 \\ 0 & 0 & 0 & : & 1 & 1 & 1 \end{bmatrix} \right)$$

The matrix X contains input values which are known. The matrix W contains randomly generated weights and W' is calculated for each randomly generated W . The above output can be written as follows.

$$Y=f_0(X'W') \Rightarrow X'W' = \log(Y/1-Y), \text{ for } f_0\text{-sigmoidal activation function}$$

Modified Gram-Schmidt orthogonalization is used to obtain a factorization $X'=QR$ and calculation of the final weights (W').

3.3 Training and Testing Algorithms

An overview of the training and testing algorithms and research methodology are presented in Figure 1 below, together with a discussion.

Figure 3: Overview of the proposed training/testing methodology.

A. DDSM Database

The mammograms utilised for this research were obtained from the benchmark Digital Database for Screening Mammography (DDSM) [19]. The database contains over 2500 high quality images which can be downloaded from the DDSM website (located at: <http://marathon.csee.usf.edu/Mammography/Database.html>) together with case related information which is available in order to promote research into CAD systems. The use of such a database aids in comparison the results with other researchers. Both the testing and training sets are comprised of an equal number of benign and malignant classes. The total 200 mammograms have been used for training and testing.

B. Area Extraction

The method of only utilising a sub-area (region of interest) of a mammogram for classification purposes is called area extraction or image segmentation. Images obtained from the DDSM contain a chain code which allows for the extraction of the mammographic abnormality and assists in reducing system resources when performing the classification process. Areas using the supplied chain code with DDSM are extracted.

C. Feature Extraction

A group of six features have been utilised in this research which represents four BI-RADS descriptor features together with patient age and a subtlety value [19]. All six features are:

- Density
- Mass Shape
- Mass Margin
- Abnormality Assessment Rank
- Patient Age
- Subtlety Value

The feature extraction and feature selection are not the main focus of research presented in this paper. Therefore the discussion on whether or not these features are appropriate is not significant. The feature extraction process and discussion can be found in our previous papers or book chapters [11].

D. Clustering- Creating Soft Clusters

Clustering involves the partitioning of a set of data into smaller similar groups to find the natural groups based on similarity of input features. This process is achieved through evaluating the distance between a point and the cluster centroid. The output from the clustering algorithm represents a statistical description of the cluster centroids with the number of components in each cluster. K-means has been incorporated in this research. K-means uses a minimum “within-class sum of squares from the centres” criterion to select the clusters. The number of clusters that the data is to be partitioned into must be selected in advance. K-means is one of the simplest unsupervised learning algorithms [25]. The procedure follows a simple way to cluster a given data set into a certain number of clusters which are fixed a priori. In this research we separate benign and malignant classes and then apply clustering separately by creating a number of clusters (soft clusters) for each class. It starts by defining k centroids representing one for each cluster. Following this, each point belonging to a dataset with benign class is associated with the nearest centroid. When no more points are pending the first step is done and we have our initial grouping. Now we recalculate k new centroids as they vary centre of the clusters from the preceding step. After this has been done a binding is done between the data set points and the nearest new centroid position. This process iterates through until the centroids do not move. The whole process is repeated for dataset with malignant class. After soft

clusters are created, we determine subclasses for benign and malignant classes by passing all input data through clusters and identifying which cluster the input data belong to.

E. Classifier Training and Testing

Before a classifier is trained it has no knowledge or capabilities in relation to the problem domain. It is through the training process that the input data is mapped to the output vectors and through this process that the network acquires knowledge in the problem domain in question. Once trained, the classifier is then able to generalise and classify using the acquired knowledge [25]. Thus a neural network classifier maintains knowledge about the problem domain by the weighted interconnections that were used to train the network. Accordingly it follows that the capabilities of the neural network classifier are heavily dependent on its initial training. Neural networks are able to capture the complex relationship of variables better than many other models because they can capture the non-linear relationship of the training data [26]. Thus a neural network classifier is the choice where it is difficult to define the rules required for a particular outcome such as the classification of suspicious areas for detection of breast cancer.

Neural network based classifiers have demonstrated to be suitable to the dilemma of classifying medical conditions such as breast cancer [10-11, 27-28]. In fact neural networks have demonstrated their superiority in these fields as their capacity to classify abnormalities as either benign or malignant based on input features that are only weakly correlated with the diagnosis is better than other techniques [27-28]. However the literature also demonstrates that neural networks can suffer from various problems which restrict their efficacy [10-11, 14].

Traditionally MLP style classifiers would be utilised in a classification dilemma where a backpropagation of an error component (such as Root Mean Square) is passed back in order to adjust the network weights. Once the error is reduced to a threshold value the network is considered as trained. However it has been noted that a reduction in a component such as root mean square error doesn't always lead to an improvement in the classification accuracy of the network [25].

In the proposed SCBDL approach, the classifier has inputs, outputs, hidden layer and two additional layers including fixed weight output layer and a cluster layer with a variable number of outputs to represent the natural tendency of the output to group into more than a straight benign or malignant class. The weights of the classifier network define the relationship between the input features and the benign or malignant patterns. However due to the weak correlation between the input features and the class patterns this relationship is oversimplified by having only two output classes. In its simplest form it could be thought that a set of input features could represent a group such as spiculated malignant or lobulated malignant masses (the expectation would be that the groupings would be

more complicated than this) and incorporating two such separate clusters (2 sub-classes within a class) will improve network training and accuracy. The training and test algorithms are described below.

Stepwise Training Algorithm

- Step 1. Extract suspicious areas from digital mammograms
- Step 2. Extract features from suspicious areas
- Step 3. Create inputs and targets for benign class
- Step 4. Create inputs and targets for malignant class
- Step 5. Create n soft clusters within benign class
- Step 6. Create n soft clusters within malignant class
- Step 7. Assign all inputs to one of $2n$ ($n+n$) classes based on $2n$ soft clusters
- Step 8. Initialise weights (W) vector for hidden layer
- Step 9. Calculate outputs (XW) of the hidden layer
- Step 10. Calculate weights (W') using $(XW)W'=C$

Stepwise Testing Algorithm

- Step 1. Extract suspicious areas from digital mammograms
- Step 2. Extract features from suspicious areas
- Step 3. Create inputs and targets for benign and malignant classes
- Step 4. Calculate outputs
- Step 5. Check the accuracy

4. PERFORMANCE EVALUATION MEASURES

The performance evaluation measures enable appropriate evaluation and comparison of proposed SCBDL classifier with other existing techniques. The performance of the proposed approach for the classification of benign and malignant patterns is measured by classification accuracy, sensitivity, specificity, and confusion matrix. The performance evaluation methods are defined below.

4.1 Classification Accuracy

Classification accuracy is the most common evaluation method to measure the performance of the classifiers. Most researchers use it as the main parameter of criteria to evaluate the performance of their techniques. The higher the classification accuracy, the better the classifier is performing. The advantage of this measure lies in its simplicity; the disadvantage is that it can be deceptive. In this study, the classification accuracy was calculated by using the following formula.

$$\text{Classification Accuracy} = \left[\frac{\text{Correct classified patterns}}{\text{Total number of patterns}} \right]$$

4.2 Sensitivity

The sensitivity measure was first introduced in medical domain to ensure the test ability of the classifier. It was calculated in the same way as the classification accuracy. However sensitivity regards only positive cases, for instance, it can be used to find patients with observed final diagnosis. In this study, sensitivity was computed as a number of true positive decisions over a number of actual positive cases. It can be represented as follows.

$$\text{Sensitivity} = \left[\frac{TP}{TP + FN} (\%) \right]$$

Where, TP = True Positive cases and FN = False Negative cases.

4.3 Specificity

The specificity measure was also established in medical domain and computed in the same fashion as sensitivity. The difference is that it deals only with negative cases, for example patients without observed final diagnosis. The specificity can be calculated as the number of true negative decisions over number of actual negative cases. It can be represented as follows.

$$\text{Specificity} = \left[\frac{TN}{FP + TN} (\%) \right]$$

Where, TN = True Negative cases and FP = False Positive cases.

4.4 Confusion Matrix

A confusion matrix represents information about actual and classified cases produced by a classification system. Performance of such system is commonly evaluated by demonstrating the correct and incorrect patterns. The typical construction of the confusion matrix for the two class problem is presented in Table 1.

TABLE 1. CONFUSION MATRIX

Test	Actual	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

5. EXPERIMENTAL RESULTS

The proposed SCBDL approach has been implemented in C++ on the Windows platform. Many experiments were conducted using the proposed approach on DDSM benchmark dataset. To make a fair comparison, we have also conducted experiments using a standard neural network classifier based on multilayer perceptron (MLP). The experiments were conducted using the same number of hidden units and training, validation and test sets. The

average of 5 runs has been taken to find the best architecture. The best architecture is then taken to evaluate all approaches on the test set. The experimental results are presented in Tables 2-4.

TABLE 2: CLASSIFICATION ACCURACY

Technique	Hidden Units	Performance	
		Accuracy on Validation Set [%] Average of 5 runs	Accuracy on Test Set [%]
Standard MLP Classifier	10	88.1	NA
	16	89.1	88.75
	27	84	NA
Proposed SCBDL Classifier (3 Soft Clusters/Class)	10	94	NA
	16	94	
	27	95	97.5

TABLE 3: CONFUSION MATRIX FOR STANDARD MLP CLASSIFIER

Test Result	Actual (Target)		Sensitivity TP / (TP + FN)	Specificity TN / (TN + FP)
	Malignant (Positive)	Benign (Negative)		
Malignant (Positive)	34 (TP)	3 (FP)	34/40=0.85	37/40=0.925
Benign (Negative)	6 (FN)	37 (TN)		

TABLE 4: CONFUSION MATRIX FOR PROPOSED SCBDL CLASSIFIER (HU=27, 3 CLUSTERS)

Test Result	Actual (Target)		Sensitivity TP / (TP + FN)	Specificity TN / (TN + FP)
	Malignant (Positive)	Benign (Negative)		
Malignant (Positive)	39 (TP)	1 (FP)	39/40=0.975	39/40=0.975
Benign (Negative)	1 (FN)	39 (TN)		

Following the performance of these experiments, an analysis of variance was performed to determine if the change in classification accuracy was statistically significant and to determine if our alternate hypothesis was correct. A single factor ANOVA test of variance was utilised for this purpose. In accordance with standard practices a confidence level was set at 5%. Table 5 contains the summary information from our experimental results on which the ANOVA was based.

TABLE 5: ANOVA NETWORK TOPOLOGY INFORMATION

Summary of Sample Population (Network Topologies)				
Group	Count	Sum	Average	Variance
Standard MLP Classifier	15	1306	87.07	5.2166
Proposed SCBDL Classifier	15	1415	94.33	0.2381

The ANOVA analysis has been calculated and is displayed in Table 6. The P-value is less than the 5% confidence threshold indicating that the null hypothesis has been refuted. However in breast cancer diagnosis classification accuracy is more complicated as the ramifications of an incorrect diagnosis can be fatal in one case or cause undue surgery in another. Accordingly we have also calculated the sensitivity and specificity values and they were detailed above in Tables 3-4.

TABLE 6: ANOVA ANALYSIS OF THE DIFFERENT NETWORKS
ANOVA – Single Factor Analysis of Network Topology Performance

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	400.044	2	200.022	64.772	1.47E-13	3.22
Within Groups	129.7	42	3.088			
Total	529.74	44				

The sensitivity and specificity values represent a measure of the classification accuracy which take into account the variance in diagnostic consequences that a false diagnosis on a malignant abnormality can have and vice versa. True negative findings (TN) are a correct benign diagnosis, False Positives (FP) are an incorrect diagnosis about a malignant abnormality, False Negative (FN) are an incorrect diagnosis about a benign abnormality and True Positive (TP) is the correct classification of a malignant abnormality. Sensitivity is a measure of detecting cancer ($TP/(TP + FN)$) while specificity is a measure of the classification accuracy or probability that an abnormality is correctly identified ($TN/(TN+FP)$). The higher the values for both sensitivity and specificity the better the performance of the system is. However in many instances researchers have observed that a high sensitivity can often be at the cost of a low specificity value. As can be seen from Table 2, the soft clustered approach with three clusters achieved the highest percentages for sensitivity and specificity. In this case both these values were high indicating the efficacy of the classifier (since the number of identified malignant abnormalities would not come at the expense of an overt number of undue surgeries). The values for both sensitivity and specificity were higher

than for the non clustered approach (5% percentage point improvement) and much better than that obtained with a standard MLP style network.

The experiments have also been conducted using ten fold cross validation. The results are presented in Table 7. The SCBDL has achieved consistent and better accuracy than the accuracy obtained from SVM, Kmeans and SOM.

Table 7: Classification Accuracy Using Ten Fold Cross Validation

Fold#	Classification Accuracy			
	[%]			
	SCBDL	SVM	Kmeans	SOM
Fold 1	100	75	70	85
Fold 2	100	100	90	80
Fold 3	95	90	60	80
Fold 4	90	85	75	65
Fold 5	90	70	80	70
Fold 6	95	85	75	75
Fold 7	95	90	85	80
Fold 8	95	90	95	75
Fold 9	95	95	90	70
Fold 10	95	85	75	80
Average	95	86.50	84.5	76

6. COMPARATIVE ANALYSIS OF RESULTS

The results obtained by proposed SCBDL classifier have been compared with the results obtained by standard MLP classifier and other existing techniques as shown in Tables 2-4 and Tables 7-8. In many instances a comparison between one CAD system and another is not an easy task since many factors can affect the classification accuracy of the system. Many neural networks provide a decision threshold for determining the difference between a benign and malignant class and this variable is typically going to be different for different classifiers. Some networks are tested on masses while others on microcalcifications yet others still have representatives from both classes. As well as the training sample itself can lead to performance differences and even using the same benchmark database (if one is utilised) can still lead to different results due to different training samples being taken from the database.

As shown in Tables 2-4, the results obtained (over 97% on test set) from this research are much higher than the results obtained by standard MLP classifier. Table 8 details the classification accuracies obtained by different researchers employing different techniques for the classification of benign and malignant patterns in digital mammograms. The results obtained from this research are also much better than the other researchers as shown in Table 8. Overall, the proposed SCBDL classifier approach has outperformed the standard MLP and other existing techniques.

TABLE 8: CLASSIFICATION ACCURACY: PROPOSED APPROACH VS EXISTING TECHNIQUES

Authors [Reference]	Technique	Database	Micro./Masses	Accuracy[%]
This paper	SCBDL classifier	DDSM	Masses	97.5
This paper	Standard MLP classifier	DDSM	Masses	88.75
Kumar et al. [27]	Genetic algorithm neural network	DDSM	Masses	89.00
Georgiou et al. [16]	SVM	NA	Masses	91.54
Panchal et al. [20]	Auto-associator MLP	DDSM	Masses	91.00
Masotti et al. [21]	SVM	NA	Masses	90.00

7. CONCLUSIONS

This paper has investigated a novel soft clustered based direct learning (SCBDL) classifier for the classification of suspicious areas in digital mammograms. The proposed SCBDL approach has achieved over 97% classification accuracy on test set which is much higher than the standard MLP classifier and other existing techniques. This research shows that the soft clustering based direct learning classifier has a significant impact on improving overall classification accuracy. The values for both sensitivity and specificity were much better than that obtained with a standard MLP style classifier.

We defined a null and alternate hypothesis in Section 3 and proved in Section 5 that soft clustered based direct learning classifier has improved classification accuracy and the results are statistically significant. A single factor ANOVA test of variance was conducted which showed that alternate hypothesis was correct. The P-value is less than the 5% confidence threshold indicating that the null hypothesis has been refuted.

In our future research, we would like to evaluate the proposed approach on other benchmark and local databases for consistency. We would also like to extend this research and investigate the impact of larger cluster size on overall performance of the classifier. An optimal size of sub classes will also be the focus of our future research.

ACKNOWLEDGEMENT

This work was supported by Australian Research Council (ARC)'s Research Network on ISSNIP under research collaboration grant.

REFERENCES

1. Roder, D., Houssami, N., Farshid, G., Gill G., Luke., Downey, P., Beckmann, K., Iosifidis, P., Grieve, L. and Williamson, L., "Population screening and intensity of screening are associated with reduced breast cancer mortality: evidence of efficacy of mammography screening in Australia", *Breast Cancer Research and Treatment*, vol. 108, no. 3, pp. 409-416, 2008.
2. Goergen, S., Evans, J., Cohen, G. and Macmillan, J., "Characteristics of breast carcinomas missed by screening radiologists", *Radiology*, vol. 204, no. 11, pp. 131-135, 1997.
3. American Cancer Society, *Breast Cancer facts and Figures 2007-2008*, American Cancer Society, Inc, Atlanta, 2007.
4. American Cancer Society, *Global Cancer facts and Figures 2007*, American Cancer Society, Inc, Atlanta, 2007.
5. Brem, R., "Clinical versus research approach to Breast Cancer detection with CAD: Where Are We Now? *American Journal of Roentology*, vol. 188, pp. 234-235, 2007.
6. Freer, T. and Ulissey, M., "Screening Mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center", *Radiology*, vol. 220, pp. 781-786, 2001.
7. Dean, J. and Ilvento, V., "Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers", *American Journal of Roentology*, vol. 187, pp. 20-28, 2006.
8. Birdwell, R., Bandodkar, P. and Ikeda, D., "Computer-aided detection with screening mammography in a university hospital settings", *Radiology*, vol. 236, pp. 451-457, 2005.
9. Morton, M., Whaley, D., Brandt, K. and Amrami, K., "Screening mammograms: interpretation with computer-aided detection – prospective evaluation", *Radiology*, vol. 239, pp. 204-212, 2006.
10. Cheng, H., Shi, X., Min, R., Ju, L., Cai, X. and Du, H., "Approaches for automated detection and classification of masses in mammograms", *Pattern Recognition*, vol. 39, no. 4, pp. 464-668, 2006.
11. Verma, B. and Panchal, R. "Neural Networks for the Classification of Benign and Malignant Patterns in Digital Mammograms", *Advances in Applied Artificial Intelligence*, Idea Group, Inc., USA, Book Editor: John Fulcher, 2006.

12. Ramirez, N., Acosta-Mesa, H., Carillo-Calvert, H., Nava-Fernandez, L. and Barrientos-Martinez, R., "Diagnosis of breast cancer using bayesian networks: A case study", *Computers in Biology and Medicine*, vol. 37, pp. 1553-1564, 2007.
13. Tourassi, G., Haarawood, B., Singh, S., Lo, J. and Floyd, C., "Evaluation of information-theoretic similarity measures for content based retrieval and detection of masses in mammograms", *Medical Physics*, vol. 34, pp. 140-150, 2007.
14. Manrique, D., Rios, J. and Rodriguez-Paton, A., "Evolutionary system for automatically constructing and adapting radial basis function networks", *Neurocomputing*, vol. 69, pp. 2268-2283, October 2006.
15. Halkiotis, S., Botsis, T. and Rangoussi, M., "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks", *Signal Processing*, vol. 87, pp. 1559-1568, 2007.
16. Georgiou, H., Mavrofarakis, M., Dimitropoulos, N., Cavouras, D. and Theodoridis, S., "Multi-scaled morphological features for the characterization of mammographic masses using statistical classification schemes", *Artificial Intelligence in Medicine*, vol. 41, pp. 39-55, 2007.
17. Brem, R., Hoffmeister, J., Zisman, G., Simio, M. And Rogers, S., "A computer aided detection system for the evaluation of breast cancer by mammographic appearance and lesion size", *American Journal of Roentology*, vol. 184, pp. 893-896, 2004.
18. Abdalla, A., Deris, S. And Zaki, N., "Breast Cancer detection based on statistical features and support vector machine", *4th International Conference on, Innovation in Information Technology*, pp. 728-730, 2007.
19. Heath, M., Bowyer, K., Kopans, D., Moore, R. and Kegelmeyer, P., "The Digital Database for Screening Mammography", *IWDM-2000*, Medical Physics Publishing, 2001.
20. Panchal, R. and Verma, B., "Neural classification of mass abnormalities with different types of features in digital mammography", *International Journal of Computational Intelligence and Applications*, Vol. 6, No. 1, pp. 61-76, 2006.
21. Massotti, M., "A ranklet-based image representation for mass classification in digital mammograms", *Medical Physics*, vol. 33, no. 10, pp. 3951-3961, 2006.
22. Acharya, R., Ng, U., Chang, Y., Yang, J. and Kaw, G., "Computer based identification of Breast Cancer using digitized mammograms", *Journal of Medical Systems*, 2008, doi:10.1007/s10916-008-9156-6.
23. Rangayyan, R., Ayres, F. and Desautels, L., "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs", *Journal of the Franklin Institute, Special Issue: Medical Applications of Signal Processing*, Part I, Vol. 344, pp. 312-348, 2007.
24. Delogu, P., Fantacci, M., Kasae, P. and Retico, A., "Characterization of mammographic masses using a gradient based segmentation algorithm and a neural classifier", *Computers in Biology and Medicine*, vol. 37, pp. 1479-1491, 2007.

25. Jones, M., *Artificial Intelligence: A systems approach*, Infinity Science Press, Canada, 2008.
26. Russell, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, 2nd Ed, Pearson Education, USA, 2003.
27. Kumar, K., Zhang, P. and Verma, B., "Application of decision trees for mass classification in mammography", International Conference on Fuzzy Systems and Knowledge Discovery, FSKD'06, pp. 366-376, China, 2006.
28. Mazurowski, M., Habas, P., Zurada, J., Lo, J., Baker, J. and Tourassi, G., "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance", *Neural Networks*, vol. 21, pp. 427-436, 2008.