A Novel Soft Cluster Neural Network for the Classification of Suspicious Areas in Digital Mammograms

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

This paper presents a novel soft cluster neural network technique for the classification of suspicious areas in digital mammograms. The technique introduces the concept of soft clusters within a neural network layer and combines them with least squares for optimising neural network weights. The idea of soft clusters is proposed in order to increase the generalisation ability of the neural network by providing a mechanism to more aptly depict the relationship between the input features and the subsequent classification as either a benign or malignant class. Soft clusters with least squares make the training process faster and avoid iterative processes which have many problems. The proposed neural network technique has been tested on the DDSM benchmark database. The results are analysed and discussed in this paper.

1. Introduction

The analysis of medical images represents an important, yet challenging part of pattern recognition and artificial intelligent systems [1]. Computer Aided Diagnostic (CAD) systems can help a physician in diagnosing a patient. CAD systems are used in the classification task where certain features (clinical findings) are used to assign a case to a particular pattern which represents a diagnosis. In particular neural networks have demonstrated their efficacy in the clinical domain with diseases such as cancer where there is a weak relationship between the class patterns forming a benign or malignant diagnosis [2-3]. Breast cancer diagnosis is particularly challenging due to the underlying cause of the disease not being known and the

similarities that exist between benign and malignant masses. The seriousness of the disease means that breast cancer is the leading cause of cancerous deaths in women in the 20 to 59 age group [4]. This situation is then further compounded by mammograms being a fairly low contrast representation of the breast. On top of this anatomical anomalies, distortion of the breast during screening and background noise add further complications to the process. The early stages of breast cancer may only have subtle indications which can be varied in appearance, making physical examination ineffective and making diagnosis difficult even for experienced radiologists [5-6].

Various studies [7-8] have shown that the size and appearance of a lesion can affect the chance that a Radiologist may miss or incorrectly diagnose a cancer. Subtle masses and small lesions are more often missed [9] than larger ones. With an estimated 11-25% of breast cancers being missed [10] during routine screening mammography there is still room for improvement. Although different studies have found that certain lifestyle factors can increase the chance of contracting breast cancer [11], the underlying cause of the disease has not been discovered. The overall survival rate in the United States for early stage Breast Cancer is 98%. If cancer has spread to the regional lymph nodes the survival rate decreases to 84%, and if metastases have spread to distant organs the chance of survival declines to 28% [11]. This means that in the short term the likelihood of developing a cure is low indicating that early and accurate diagnosis is the surest mechanism of reducing the mortality rate.

The use of CAD can improve the performance of the radiologists in terms of reducing the number of missed diagnosis [9-10, 12-13] and reducing the time taken to reach a diagnosis. Not only can CAD lead to improvements in accuracy but CAD systems can lead to faster turnaround of results as well as improved storage and retrieval of radiographic images [14]. This can also facilitate the transmission of these images to other medical specialists. CAD systems can provide a valuable tool for the training of radiologists.

The remainder of this paper is organised into seven sections. Section 2 reviews the existing techniques. Section 3 details the theoretical underpinnings, architecture and training of the soft cluster neural network. Section 4 presents the research methodology. Section 5 details the results obtained with section 6 providing a discussion of the results together with a comparative analysis of the findings. In section 7, conclusions are drawn and future research directions are addressed.

2. Related Work

Medical diagnosis is a pattern classification dilemma where a set of input features are used to determine if a patient has a particular disorder. Breast cancer diagnosis is a classification problem in terms of distinguishing between a malignant and benign mass from a suspicious lesion or region. Such problems have lent themselves to the application of computer aided diagnostic (CAD) systems over the last few decades. A myriad of approaches such as artificial neural networks, genetic algorithms, wavelet filters and statistical transforms have been employed in order to solve this problem.

Optical Fourier transforms [15] and wavelet transforms are effective on micro calcification where the high spatial frequency in the Fourier spectrum allows for their identification, but may not be as accurate on masses where a lower spatial frequency exists similar to the surrounding tissues. Fuzzy logic [16], Bayesian Networks [17], case based systems [18] and artificial neural networks [2-3, 19-20] have been utilised to varying degrees of success. Artificial neural networks have demonstrated their capabilities in this area and have been widely adopted due to their generalisation capabilities [21]. Georgiou [21] utilised morphological features with a support vector machine to obtain 91.54% classification accuracy. Halkiotis et al. [22] used the MIAS database and a Multi Layer Perceptron (MLP) type neural network to obtain a classification rate of 94.7% with an average of 0.27 false positives per image. Brem et al. [13] used the second look CAD system (version 3.4) to determine the performance of CAD systems on different sized lesions and micro calcifications to achieve an overall sensitivity of 89%. Their investigation was to try and determine if lesion size would adversely impact upon the performance of a CAD system. Abdalla et al. [23] used textual features with a support vector machine classifier to achieve a classification accuracy of 82.5% on mammograms from the Digital Database of Screening Mammography [24]. Rangayyan et al. [25] noted that several methods have good sensitivity ($> 85\%$) for the identification of masses but also have a high false positive rate.

As can be seen from the above, neural network based techniques have been used for the classification of suspicious areas into benign and malignant classes, however there are several disadvantages with existing neural network techniques. Neural networks must undergo a training phase which can take a while to complete. The training phase needs to be repeated if new knowledge is to be added to the system making it hard to add new knowledge. The

optimisation of a neural network relies on the setting of various parameters which need to be adjusted during the training phase which can greatly affect the performance of the system if not correctly tuned. The tuning of the network parameters often requires multiple iterations to determine the optimal configuration. In other situations certain neural networks may suffer from network paralysis, or excessive use of computational resources [26] making them infeasible for use in a clinical situation. Neural networks are often tasked with handling a large number of features, many of which are only weakly correlated or are of limited diagnostic value with the class pattern that they are attempting to classify. This creates a problem for training the network and produces low accuracy on test data. In this paper, we introduce a new concept of soft clusters with least squares and propose a novel neural network which can avoid some of the above mentioned problems with improved classification accuracy.

3. Soft Cluster Neural Network

Soft Cluster Neural network (SCNN) is based on an idea that in a classification problem, each class can have more than one cluster called soft clusters and that incorporating soft clusters output values into the learning of neural network weights will improve the learning process and the overall classification accuracy. The use of soft clustered neural network for the classification of benign and malignant patterns can lead to an improvement in classification accuracy over other approaches. To validate the proposed claim, we define the null and alternative hypotheses as follows.

H_0 : μ The use of soft clustered neural networks makes no difference to the classification of benign and malignant classes when compared to traditional networks such as k-means or support vector machine.

H_1 : μ The use of soft clustered neural networks results in an improvement in the classification of benign and malignant classes when compared to traditional approaches such as k-means or support vector machine.

The theoretical underpinnings, network architecture and training methods for the soft cluster neural network are presented in the following sections.

3.1 Theoretical Underpinnings

Let $P = [p_1^{(j)}, p_2^{(j)}, \dots, p_n^{(j)}]$ be the set of j input patterns with n attributes and $O = [o_1^{(j)}, o_2^{(j)}]$ be the output classes.

Let $J = \sum_{j=1}^k \sum_{i=1}^n \|p_i^{(j)} - c_j\|^2$ be the objective function for clustering where k is the number of clusters, $\|p_i^{(j)} - c_j\|^2$ is a chosen distance measure between a data point $p_i^{(j)}$ and the cluster centre c_j , which is an indicator of the distance of the n data points from their respective cluster centres (centroids).

The clustering based on objective function mentioned above can cluster $P = [p_1^{(j)}, p_2^{(j)}, \dots, p_n^{(j)}]$ inputs into two clusters/classes $O = [o_1^{(j)}, o_2^{(j)}]$. However hard clustering with one single hard boundary creates a lot of problems. Therefore, many soft clusters are created and the output of only those clusters which strongly associate with one of the classes, are used for weight optimisation. The weak and zero clusters are removed from influencing the network parameters in the wrong directions which will reduce the number of false classifications. It will also reduce the input data for weight optimisation and provide faster training of the network.

Let $J = [j_1, j_2, \dots, j_k]$ be the output of k clusters, where $j_1, j_3 \dots j_{n-1}$ may belong to o_1 and $j_2, j_8, \dots, j_n$ may belong to o_2 . The output of the neural network can be optimised using the following output functions.

output = $f(J' * W)$, where $J' \in (j_1, j_3 \dots j_n)$ and $(j_1, j_3 \dots j_n) \in$ output of all clusters.

output = $f(J * W)$, where $J \in (j_2, j_3, j_4)$ and $(j_2, j_3, j_4) \in$ output of non zero & strong clusters.

Where the first output function utilises all the weights from the clustering process for the network training process while the second output prunes the underperforming weights from inclusion in the training process of the neural network. The actual details of these functions are expanded in Sections 3.2 and 3.3.

3.2 SCNN Architecture

An overview of the proposed SCNN architecture is shown in Figure 1. It has n inputs, m outputs and k soft clusters.

Fig. 1: SCNN Architecture

The error minimisation function for SCNN is defined as follows. SCNN is trained to minimise the following error.

$$E = \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^m (y_j^i - t_j^i)^2$$

Where p is the number of training samples, m is the number of outputs, y is the network output and t is the target (desired) output.

The output of the network is defined as follows.

$$y_j = \frac{1}{1 + e^{-net_j}},$$

$$net_j = W_o \left(\frac{1}{1 + e^{-(JW_h)}} \right), \text{ Where } J \text{ -output of the cluster layer and } W_h \text{ -weights for hidden layer.}$$

$$XW_o = \log\left(\frac{t_j}{1-t_j}\right), \text{ Where } X \text{ -output of the hidden layer and } W_o \text{ -weights for output layer.}$$

3.3 SCNN Training Methods

3.3.1 Determining Weak/Zero Clusters/Neurons

The weak and zero clusters are determined by creating a confusion matrix as shown in Table 1 below through clustering of the input data. The clustering analysis is an effective

mechanism in computational problems as it groups a set of data in d-dimensional feature space to maximise the similarity within a cluster and minimise the similarity between different clusters.

The use of clustering relies on an underlying assumption; and that is any given population will consist of n elements described by m attributes and that this population can be partitioned into k clusters. Accordingly $P_i = (p_{i1}, p_{i2}, \dots, p_{im})$ which will represent the vector of the m number of attributes of element i .

The process of clustering that searches for an optimal partition within a fixed number of clusters can be described as follows:

1. The initial seeds with the number of clusters k are selected and an initial partition is established by using the seeds as the centroids of the initial clusters.
2. Each input pattern is assigned to the centroid that is nearest, thus forming a cluster.
3. Keeping the same number of clusters, the new centroid is calculated for each cluster.
4. Steps 2 and 3 are repeated until the clusters stop changing.

$$J = \sum_{j=1}^k \sum_{i=1}^n \|p_i^{(j)} - c_j\|^2$$

This process can be explained in the above equation where J is the objective function we are trying to minimise. k is the number of clusters, $\|p_i^{(j)} - c_j\|^2$ is a chosen distance measure between a data point $p_i^{(j)}$ and the cluster centre c_j , which is an indicator of the distance of the n data points from their respective cluster centres (centroids).

A confusion matrix as shown below can be produced using the above clustering on training data. The matrix shows cluster numbers together with the number of input patterns belonging to each cluster.

Table 1- Cluster Evaluation Process

Cluster Number	1	2	3	4	5	6		k
Class 1	0	15	0	30	5	0		0
Class 2	0	3	0	2	20	0		25

The removal of zero and weak clusters is an attempt to prune the weights that have the lowest salience for the clustering of the input features, and hence provide low quality discrimination

for the relationship between the input features and subsequent benign or malignant class. Using the values in Table 1 only the output of neurons from cluster numbers 2, 4, 5 and k which are strong clusters would be fed into the MGS layer for weight calculation.

3.3.2. Determining Modified Gram-Schmidt Weights

The use of MGS in conjunction with linear system of equations to calculate/determine the output weights has been chosen since it has the potential to overcome some of the problems that could occur with traditional iterative algorithms such as backpropagation [27] where the network could suffer from local minima, paralysis and slow training process [27-28]. By using a non-iterative process such as MGS, the weights can be determined directly and the major problems as mentioned above can be avoided. Modified Gram Schmidt can also be faster at determining the output weights compared with backpropagation. The output of the neural network can be written as follows.

output = $f(J*W) = T$, where T is a target vector output for the neural network, W is a weight matrix for the output layer.

The linear system of equations $JW=T$ is solved using the MGS. The MGS's factorisation process would occur as follows:

$$J = \sum_{i=1}^n cQ_i rR_i^T$$

Where J represents a $m \times n$ matrix with $(m > n)$, and J may be decomposed into the product of $J = QR$ where Q is $(m \times n)$ orthogonal ($Q^T Q = I_n$) and R is $(n \times n)$ upper triangle. cQ_i represents the i th column of our vector Q, and rR_i refers to the i th row of R.

3.4 Advantages of SCNN over Other Existing Techniques

The major advantages of the proposed SCNN over other existing techniques such as Support Vector Machines (SVMs) are as follows (1) SCNN can have any number of inputs and outputs so it can easily deal with multi-class problems. SVM needs multiple binary SVMs to deal with multi-class problems. (2) The complexity of SCNN is lower than the other techniques such as SVMs as they need a large number of support vectors which increases complexity. (3) SCNN can be used for approximation as well as for classification. (4) SCNN

can be used with binary or real values without making any changes in architecture or training scheme. (5) SCNN uses sigmoidal functions in output and hidden layers so it maintains nonlinearity same as traditional back propagation type neural networks. (6) SCNN is faster and achieves better generalisation than other techniques including SVM. (7) SCNN avoids local minima.

4. Methodology for Experiments Using SCNN

An overview of the research methodology for using SCNN to classify benign and malignant patterns is given in Figure 2 and it involves data collection, suspicious area extraction, feature extraction, neural network training and testing.

Fig. 2: An Overview of Research Methodology

4.1 Data Collection

The mammogram images were obtained from the University of South Florida's Digital Database for Screening Mammography (DDSM) in order to evaluate the results on a benchmark database [24]. The database contains approximately 2,500 studies of malignant, benign, benign-without-callback and normal cases. The DDSM has already been interpreted by expert radiologists and the appropriate information has been provided in "ics" and "overlay" files. In this research 100 malignant and 100 benign cases are used for training and testing.

4.2 Processing And Area Extraction

The obtained digital mammograms are processed for the extraction of suspicious areas and features. The process of dividing a mammogram into distinct and discrete regions is called Area Extraction or Image Segmentation. This process is useful for subsequent phases since the mammogram is divided into Regions of Interest (ROI) which reduces the usage of system resources by discarding inappropriate regions. ROI represent regions of suspicion; however these regions contain anatomical anomalies as well as malignant and benign structures. The area extraction was based on chain code provided with the DDSM and the programs developed in our previous research [29, 30] were used to extract these areas.

4.3 Feature Extraction

A set of six features has been extracted and utilised in this research representing four BI-RADS descriptor features (density, mass margin, mass shape, abnormality assessment rank) together with patient age and a subtlety value feature [24]. The features are described below.

4.3.1 Density Feature

Breast density is measured on a scale from one to four. The values are according to the BI-RADS definitions and are classified by an expert radiologist. The classifications are reproduced below:

Table 2 - BI-RADS Breast Density Ratings

Density Ratings	
Code	Description
1	The breast is almost entirely fat.
2	There are scattered fibroglandular densities that could obscure a lesion on mammography.
3	The breast is heterogeneously dense. This may lower the sensitivity of mammography.
4	The breast tissue is extremely dense, which lowers the sensitivity of mammography.

4.3.2 Morphological Description of Breast Abnormalities

The morphological features and arrangements of abnormalities are used to determine if an abnormality is of a cancerous or benign class. Like other features morphological features are encoded into numeric values to represent the real feature types. Two sets of descriptors were

required for the Morphological description representing the Mass Margin and the Mass Shape. The mass types are as follows.

4.3.2.1 Mass Shape Feature

Table 3 - BI-RADS Mass Shape Types

Mass Shape			
Code	Description	Code	Description
1	Round	6	Tubular
2	Oval	7	Lymph Node
3	Lobulated	8	Asymmetric_Breast_Tissue
4	Irregular	9	Focal_Asymmetric_Density
5	Arcitectural_Distortion	10	Any

4.3.2.2 Mass Margin Feature

Table 4 - BI-RADS Mass Margin Types

Mass Margin			
Code	Description	Code	Description
1	Circumscribed	4	Ill_Defined
2	Microlobulated	5	Spiculated
3	Obscured	6	Any

4.3.3 Abnormality Assessment Rank Feature

The abnormality assessment rank represents a determination of the seriousness of the abnormality detected and a suggested course of action that should be taken. It is determined by an expert radiologist and is rated on a scale from 1 to 5 [24].

Table 5 - BI-RADS Abnormality Assessment Ranks

BI-RADS Abnormality Assessment Rank	
Category	Description and Course of Action
1	Negative.
2	Benign finding.
3	Probably benign finding.
4	Suspicious abnormality – Biopsy should be considered.
5	Highly suggestive of malignancy – Appropriate action should be taken.

4.3.4 Patient Age Feature

Various studies and cancer statistics have demonstrated a correlation between patient age and the risk of contracting breast cancer [6]. This makes patient age a useful feature for consideration in the diagnosis of breast cancer however cancer statistics also show a number of exceptions to the general age based correlation.

4.3.5 Subtlety Value Feature

Subtlety value is a subjective measure by a radiologist of how obvious a lesion is to detect. Cancerous abnormalities are more subtle with ill defined margins compared to non-cancerous abnormalities. It has a value between 1 and 5 with 1 being “subtle” and 5 representing “obvious”.

4.4 SCNN Training

Training a neural network is an optimisation problem of finding the set of network parameters (weights) that provide the best classification performance of the system. An inappropriate set of weights will lead to degraded performance of the system resulting in the inappropriate assignment of cases to incorrect output classes. In order to train the network the initial parameters for the number of features, number of clusters and number of hidden units are set. The training steps for SCNN are described below.

4.4.1 Stepwise Algorithm

<p>Step 1: Process mammograms and extract features from suspicious areas.</p> <p>Step 2: Use clustering algorithm to cluster data into many soft clusters.</p> <p>Step 3: Remove weak and zero clusters and keep strong clusters only.</p> <p>Step 4: Extract output of the clusters (J) for each training and test pattern.</p> <p>Step 5: Initialise hidden weights (W_h) at small random values.</p> <p>Step 6: Calculate the output (X) of hidden layer ($f(JW_h)$).</p> <p>Step 7: Use MGS to calculate weights (W_o) of the output layer
for training patterns.</p> <p>Step 8: Use test patterns (features) and calculate the accuracy.</p> |
|--|

5. Results

The SCNN has been implemented in C++ on the Windows platform and is composed of two programs. The features for the experiments were extracted from suspicious areas of digital mammograms taken from the DDSM database. The features were clustered into soft clusters. The zero and weak clusters were removed and the SCNN was trained by calculating the weights of the network. The network was then tested with the test data. A number of experiments were performed by changing the number of clusters and the results are reported in Table 6 below.

Table 6 - Classification Accuracy of SCNN

Methods/Algorithms	#of Clusters	#of Hidden Units	Classification Accuracy [%]	
			Training Set	Test Set
K-Means	16	NA	82	79
	17	NA	82	79
	19	NA	84	80
	20	NA	77	81
	21	NA	81	82
	29	NA	83	80
	41	NA	82	79
	50	NA	85	79
SCNN [All Clusters]	12	24	91	91
		29	96	91
		35	94	92
	50	45	96	91
		24	95	93
		27	93	91
		29	94	91
		45	96	91
SCNN [Strong Clusters]	12	22	91	92
		24	93	91
		29	91	89
	50->9	35	91	89
		24	91	93
		27	93	94
		29	95	94
		45	95	92

6. Analysis of Results

Once the results had been obtained a single factor ANOVA test of variance was performed on the top 15 results from each network to determine if our alternate hypothesis was correct. The alpha value was set at a 5% confidence level. Table 7 contains the summary of the sample population and table eight contains the results from the Anova analysis.

Table 7 – Summary of Network Sample Population for Anova Test

Summary of Sample Populations (Network Topologies)				
Classifier Network	Count	Sum	Average	Variance
K-means	15	1179	78.6	2.9714
SCNN – All Clusters	15	1374	91.6	0.5429
SCNN – Strong Clusters	15	1403	93.533	0.2667

Table 8 – Anova Analysis of the Different Classifier Networks

Anova – Single Factor Analysis of Network Topologies						
Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	1978.711	2	989.36	785.01	5.41E-34	3.22
Within Groups	52.93	42	1.26			
Total	2031.64	44				

The Anova analysis shown in Table 8 indicates that the P-value is less than the 5% threshold indicating that the null hypothesis (Section 3) has been disproved. However in breast cancer research due to the nature of the condition where a benign finding for a malignant tumour can have a larger impact on the patient the sensitivity and specificity values need to be calculated. A confusion matrix was calculated for each network. Essentially a confusion matrix contains information about the actual and predicted classification results. A confusion matrix is represented as follows (Table 9).

Table 9 – Representation of a Confusion Matrix

Representation of a Confusion Matrix		
Actual Result	Benign	Malignant
Benign	TN	FP
Malignant	FN	TP

Where TN represents the True Negative findings (Benign), FP represents False Positive Findings (an incorrect diagnosis about a malignant mass), FN represents False Negative findings (an incorrect diagnosis about a benign mass) and TP represents True Positive findings (or correct diagnosis of a tumour). Table 10 represents the confusion matrix for the neural networks.

Table 10 – Confusion Matrix for the Performance of Each Network

Confusion Matrix for each Classifier Network.			
Classifier Network	Actual Result	Benign	Malignant
K-means	Benign	42	10
	Malignant	8	40
SCNN – All Clusters	Benign	47	4
	Malignant	3	46
SCNN – Strong Clusters	Benign	49	5
	Malignant	1	45

A confusion matrix allows for the calculation of two ratios often used in Breast cancer research called sensitivity and specificity. Sensitivity is calculated as the True Positive classification divided by the True Positive and False Negative classifications. Specificity is calculated as the True Negative classifications divided by the False Positives plus the True Negatives. An analysis is shown in Table 11.

Table 11 – Sensitivity and Specificity Values for Each Network

Sensitivity and Specificity values for each Classifier Network.		
Classifier Network	Sensitivity %	Specificity %
K-means	83.3	80.77
SCNN – All Clusters	93.88	92.16
SCNN – Strong Clusters	97.83	90.74

The soft clustered neural networks performed better than the k-means network in terms of both their sensitivity and specificity values. Although the strong cluster neural network had a higher sensitivity value than the all clusters neural network this was at the expense of a declining specificity value. The higher sensitivity and specificity values support the efficacy of the soft clustered neural network in comparison to the k-means network.

6.1 Ten Fold Cross Validation

The analysis and comparison using tenfold cross validation has also been conducted. The classification accuracy using ten fold cross validation is shown in Table 12 below.

Table 12 - Classification Accuracy Using Ten Fold Cross Validation

Fold#	Classification Accuracy			
	[%]			
	SCNN	SVM	Kmeans	SOM
Fold 1	95	75	70	85
Fold 2	100	100	90	80
Fold 3	95	90	60	80
Fold 4	90	85	75	65
Fold 5	95	70	80	70
Fold 6	85	85	75	75
Fold 7	90	90	85	80
Fold 8	95	90	95	75
Fold 9	100	95	90	70
Fold 10	95	85	75	80
Average	94	86.50	84.5	76

6.2 Clustering Analysis

The clusters have been analysed and the clustering results are presented in Table 13. The bold numbers represent the number of patterns in correct clusters. The underlined numbers represent the number of patterns in incorrect clusters.

Table 13 – Clustering Results

Fold #		Class	Clusters										
			#1	...								#k	
Fold 1	Train	Benign				8	:	14	20	19	1	25	3
		Malignant				1	:	16	7	4	28	1	33
	Test	Benign					:	3	5		1	1	0
		Malignant					:	1	0		1	2	6
Fold 2	Train	Benign					:	15	22	20	5	24	4
		Malignant					:	17	8	4	24	4	33
	Test	Benign					:	1	3	2	0	4	0
		Malignant					:	1	1	0	2	0	6
Fold 6	Train	Benign					:	18	28	14	1	28	1
		Malignant					:	10	8	33	16	4	19
	Test	Benign					:	1	5	2	0	2	0
		Malignant					:	0	2	4	1	1	2
Fold 9	Train	Benign	0	0	7		:	27	25	5	1	23	2
		Malignant	2	2	0		:	6	9	33	16	2	20
	Test	Benign			1		:	3	2	1	0	3	0
		Malignant			0		:	0	1	3	1	0	5

6.3 Problem of Small Sample Size

The proposed SCNN has been analysed for small sample size. The experiments varying training sample size were conducted and the performance was measured. The classification accuracy is presented in Table 14 below. The increase in sample size has slightly increased the accuracy. The results also showed that the better accuracy was achieved with small number of hidden units when small sample size was used.

Table 14 - Classification Accuracy of SCNN Using Small Sample Size

Sample Size	#of Hidden Units	Classification Accuracy [%]
30	16	90
30	24	83
30	29	80
60	16	91
60	24	89
60	29	89
80	16	93
80	24	94
80	29	94

6.4 Comparative Analysis of Results

The results obtained on a DDSM benchmark database using the proposed SCNN are detailed in Table 6 and Table 12. The highest classification accuracy obtained by SCNN with strong clusters was 94%. The results obtained by SCNN with strong clusters compare favourably with other techniques and show the superiority of SCNN over other techniques.

Fig. 3: Comparison of Accuracies on DDSM's Mass Dataset

The proposed SCNN has been compared with six recently published techniques in the literature. The statistical and neural network based techniques such as Auto-Associator Neural

Network (AANN) [30], CART [29], See5 [29], Genetic Algorithm Neural Network (GANN) [29], Self Organising Map (SOM)[32-33] and Support Vector Machine (SVM) [31] were used for comparison. All the experiments for comparison purposes were conducted using same training and testing database. The weka package was used for SVM experiments. A comparison of accuracies is presented in Figure 3. The comparison showed that the proposed SCNN is very promising.

7. Conclusions

The research presented in this paper has demonstrated that the proposed approach (SCNN) with new ideas of soft clusters, removing weak clusters/neurons (keeping strong clusters) and combining clustering with MGS has been able to produce an improvement in the classification accuracy and training time. The highest classification accuracy obtained by SCNN with all clusters and without weak clusters was 93% and 94% respectively. It was found that the increased accuracy was received with a larger number of clusters. The reason for this is that the training data naturally falls into a number of groups encasing a larger variation than just benign or malignant classes (such as a round benign mass, etc.). If the number of clusters are kept low the resultant weights from the clustering process reflect this variation and effectively result in an over training of the network based on the smaller cluster size which may produce lower classification accuracy on test set. However, the accuracy with smaller cluster size will still be higher than most existing techniques.

In our future work we would like to focus on expanding this research to find the appropriate number of clusters and test it on a larger benchmark dataset. We would also like to expand it by using the value of non-conflicting clusters from the confusion matrix. This might further improve the classification accuracy and reduce the complexity.

References

- [1] Mousa, R., Munib, Q. and Moussa, A. "Breast cancer diagnosis system based on wavelet analysis and fuzzy neural", Expert Systems with Applications, Vol. 28, No. 1, pp. 713-723, 2005.
- [2] Lisboa, P. and Taktak, A. "The use of artificial neural networks in decision support in cancer: a systematic review", Neural Networks, Vol. 19, pp. 408-415, 2006.

- [3] Lisboa, P. "A review of evidence of health benefit from artificial neural networks in medical intervention", *Neural Networks*, Vol. 15, pp. 11-39, 2002.
- [4] Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T. and Thun, M., "Cancer Statistics, 2008", *CA: A Cancer Journal for Clinicians*, 2008.
- [5] Christoyianni, I., Kotras, A., Dermatas, E. and Kokkinakis, G. "Computer aided diagnosis of breast cancer in digitized mammograms", *Computerised Medical Imaging and Graphics*, Vol. 26, pp. 309-319, 2002.
- [6] Wallis, M., Walsh, M. and Lee, J. "A review of false negative mammography in a symptomatic population", *Clinical Radiology*, Vol. 44, pp. 13-15, 1991.
- [7] Birdwell, R., Ideda, D., O'Shaughnessy, K. and Sickles, E. "Mammographic Characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer aided detection", *Radiology*, Vol. 219, pp. 192-202, 2001.
- [8] Kallergi, M., Carney, G. and Gaviria, J. "Evaluating the performance of detection algorithms in digital mammography", *Medical Physics*, Vol. 26, pp. 267-275, 1999.
- [9] Bird, R., Wallance, T. and Yankaskas, B. "Analysis of cancers missed at screening mammography", *Radiology*, Vol. 184, pp.613-617, 1992.
- [10] 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.
- [11] American Cancer Society. *Global Cancer Facts and Figures 2007*. Atlanta: American Cancer Society, 2007.
- [12] Balleyguier, C., Kinkel, K., Fermanian, J., Malan, S., Djen, G., Taourel, P. and Helenon, O. "Computer-aided detection (CAD) in mammography: does it help the junior or the senior radiologist?", *European Journal of Radiology*, Vol. 54, pp. 90-96, 2005.
- [13] Brem, R., Hoffmeister, J., Zisman, G, DeSimio, 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.
- [14] Pisano, E. D., Gatsonis, C. and Hendrick, E. "Diagnostic performance of digital versus film mammography for breast cancer screening', *New England Journal of Medicine*, Vol 353, no. 17, pp. 1773-1783, 2005.

- [15] Yelleswarapu, C., Kothapalli, S. and Rao, D. "Optical fourier techniques for medical image processing and phase contrast imaging", *Optical Communications*, Vol. 28, pp. 1876-1888, 2008.
- [16] Lee, Y. and Tsai, D. "Computerised classification of microcalcifications on mammograms using fuzzy logic and genetic algorithm", In *Medical Imaging 2004: Image Processing*, Fitzpatrick J., Sonka M, editors, *Proceedings of the SPIE*, Volume 5370, pp. 952-959, 2004.
- [17] 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.
- [18] 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.
- [19] Cheng, H. D., Shi, X. J., Min, R., Ju, L. M., Cai, X. P. and Du, H. N., "Approaches for automated detection and classification of masses in mammograms", *Pattern Recognition*, Vol. 39, No. 4, pp. 646-68, 2006.
- [20] Bovis, K., Singh, S., Fieldsend, J. and Pinder, C. "Identification of masses in digital mammograms with MLP and RBF Nets", *Proc. of IEEE-IJCNN*, Vol. 1, pp. 342-347, 2000.
- [21] 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.
- [22] 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.
- [23] Abdalla, A., Deris, S. and Zaki, N. "Breast cancer detection based on statistical features and support vector machine", 4th International Conference on, *Innovations in Information Technology*, pp. 728-730, 2007.
- [24] Heath, M., Bowyer, K., Kopans, D., Moore, R. and Kegelmeyer, P. "The Digital Database for Screening Mammography", *IWDM-2000*, Medical Physics Publishing, 2001.
- [25] 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.
- [26] Mazurowski, M., Habas, P., Zurada, J. and Tourassi, G., "Impact of low class prevalence on the performance evaluation of neural network based classifiers: Experimental study in the context of computer-assisted medical diagnosis", Proceedings of International Joint Conference on Neural Networks, pp. 2005-2009, 2007.
- [27] Rumelhart, D.E., Hinton, G.E., Williams, R.J., "Learning internal representations by back-propagating errors", in: D.E. Rumelhart, J.L. McClelland, The PDP Research Group (Eds.), Parallel Distributed Processing, Vol. 1, MIT Press, Cambridge, MA, 1986, pp. 318-362.
- [28] Gori, M. and Tesi, A., "On the problem of local minima in backpropagation", IEEE Trans. Pattern Anal. Mach. Intell. Vol. 14, No. 1, pp. 76-86, 1992.
- [29] 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.
- [30] 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.
- [31] Masotti, M. "A ranklet-based image representation for mass classification in digital mammograms", Medical Physics, Vol. 33. No. 10, pp. 3951-3961, 2006.
- [32] Sawarkar, S. and Ghatol, A., "Breast cancer malignancy identification using self-organizing map", Proceedings of the 5th WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing, pp. 20-25, 2006.
- [33] Chang, P. and Teng, W., "Exploiting the self-organizing map for medical image segmentation", 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), pp.281-288, 2007.