

Novel Network Architecture and Learning Algorithm for the Classification of Mass Abnormalities in Digitized Mammograms

Brijesh Verma

School of Computing Sciences, Central Queensland University

Bruce Highway, North Rockhampton, Queensland 4702, Australia

Phone: +61749309058, Fax: +61749309027, E-mail: b.verma@cqu.edu.au

Summary

Objective

The main objective of this paper is to present a novel learning algorithm for the classification of mass abnormalities in digitized mammograms.

Methods and Material

The proposed approach consists of new network architecture and a new learning algorithm. The original idea is based on the introduction of an additional neuron in the hidden layer for each output class. The additional neurons for benign and malignant classes help in improving memorization ability without destroying the generalization ability of the network. The training is conducted by combining minimal distance based similarity/random weights and direct calculation of output weights.

Results

The proposed approach can memorize training patterns with 100% retrieval accuracy as well as achieve high generalization accuracy for patterns which it has never seen before. The grey-level and breast imaging reporting and data system based features from digitized mammograms are extracted and used to train the network with the proposed architecture and learning algorithm. The best results achieved by using the proposed approach are 100% on training set and 94% on test set.

Conclusion

The proposed approach produced very promising results. It has outperformed existing classification approaches in terms of classification accuracy, generalization and memorization abilities, number of iterations, and guaranteed training on a benchmark database.

1. Introduction

Breast cancer continues to be the most common cause of cancer deaths in women. Every year more than 1 million women develop breast cancer world-wide [1, 2]. A report by a national cancer institute [3] estimates that 1 in 11 women develop breast cancer in Australia [3], 1 in 8 in US, 1 in 9 in UK and Canada. In 2005, an estimated 1,150,000 women worldwide were diagnosed with breast cancer and 411,000 women died from the disease [1].

As there is currently no means of preventing breast cancer, the focus in reducing deaths from the disease has been on finding breast cancer as early as possible. Early detection of cancer saves patients from the more aggressive radical treatments and increases the overall survival rate. The introduction of screening mammography in 1963 brought major revolution in breast cancer detection and diagnosis. It is widely adopted in many countries including Australia as a nation wide public health care program. The decline in the number of breast cancer deaths [4] corresponds directly to an increase in routine mammography screening.

At present digital mammography is one of the most reliable procedures for early detection of breast cancer [1-3, 5, 6]. Digital mammography screening has advantages compared with film mammography, according to the results of a recent study reported [6] in September 2005 in the new England journal of medicine. Current image processing techniques for digitized mammography make the primitive breast abnormalities detection easier; however their classification as malignant or benign remains one of the most difficult tasks for radiologists and researchers. The key reason is a lack of individuality in benign and malignant class patterns. In many cases, both classes exhibit similar characteristics, i.e. size, shape and distribution of microcalcification. Radiologists' interpretation of such cases often produce screening errors; either to miss malignant cases or more benign biopsies. Research has shown that the computer-aided diagnosis (CAD) can significantly reduce misdiagnosis. In a 2005 study published in the European journal of radiology, researchers found that CAD was able to correctly point out cancers that junior radiologists tended to miss more often [6]. When applied to the United States population, researchers have estimated that for every 100,000 cancers currently detected with screening mammograms (in women with no obvious signs of breast cancer, such as a lump), the use of CAD technology could result in the detection of an additional 20,500 breast cancers.

The ability of neural classifiers to learn from the attributes of given class patterns and to classify the

unknown patterns of given classes into appropriate classes using the acquired knowledge has shown its potential [7-22] in the field of mammography. Advancement in computational intelligence and pattern recognition techniques [7-44] has improved the performance of computer-aided classification of the breast abnormalities into malignant and benign patterns in digitized mammograms.

In this paper, we propose a neural architecture and a training algorithm which can learn and classify mass abnormalities in digitized mammograms. The approach is based on introduction of additional neurons in hidden layer for benign and malignant classes and a weight adjustment technique for the calculation of weights.

The remainder of this paper is broken down into five sections. Section 2 reviews existing techniques for the classification of benign and malignant patterns. Section 3 discusses the proposed technique. Section 4 presents the experimental results obtained using the proposed architecture and learning algorithm. Section 5 presents a discussion and analysis of the obtained results. Finally, Section 6 concludes the paper and describes the future research directions.

2. Review of Existing Techniques

The detection and classification of breast abnormalities in digitized mammograms has been studied for the last few decades and many scientific papers have been published in the literature. Recent surveys by Cheng et al. [7] and Verma et al. [15] present a comparative analysis of various algorithms and techniques for the diagnosis of abnormalities in digitized mammograms. The techniques such as artificial neural networks [7-20], fuzzy logic [8, 23, 24, 37], and wavelet transforms [25, 26] are the most commonly used for the classification of malignant and benign patterns in digitized mammograms. Verma et al. [8] used a back-propagation neural network for the classification of the suspicious lesions extracted using a fuzzy rule based detection system. They obtained 88.9% classification rate using manual combination of features. Zhang et al. [9] used a genetic algorithm for neural network learning in their study of breast abnormality classification. Two types of features; grey level based statistical features and radiologist's interpretation features including patient age are extracted from the DDSM database to test their proposed technique. They have attained a good (90.5%) accuracy rate on test set at the cost of low accuracy rate on training set. Chitre et al. [12] compared the artificial neural networks and the statistical methods for microcalcification patterns classification. They obtained a classification rate of 60%, which was better than the statistical classifiers. A comparative study of a radial basis

function (RBF) and a multi-layer perceptron (MLP) based neural networks for the classification of breast abnormalities using the texture features were performed by Christoyianni et al. [13] and Bovis et al. [14] in their research work. They concluded that MLP obtained 4% higher accuracy than RBF. Wroblewska et al. [27] proposed a new segmentation and feature extraction technique for reliable classification of microcalcifications which achieved low classification rate (78%) on DDSM database. Yu et al. [31] used a multilayer feed forward neural network to classify the potential pixel as a true or false microcalcification object. They obtained good true positive accuracy at the cost of very low false positive rate.

As can be seen in previous paragraphs and literature [1-44] that there has been a lot of research in last few decades and neural network based techniques have shown promising future in breast cancer diagnosis. There is no doubt that neural networks have achieved better performance [7, 8, 15, 43] in terms of classification accuracy than other traditional techniques. Neural networks have improved accuracy rate for the classification of benign and malignant patterns in digitized mammography. Recent reviews [7, 15] have also reported the superiority of neural networks over other techniques. However, as mentioned earlier there are many drawbacks with neural network based techniques for the diagnosis of breast cancer. First such drawback is that current neural network based techniques do not provide any assurance to radiologists that they are not going to misclassify patterns which are already in classified database (known patterns). Second drawback is that the current neural network approaches are very slow in adapting and learning new knowledge from new acquired mammograms. Third drawback is that its generalisation ability and consistency, if we try to get good accuracy on unseen mammograms (test data), the accuracy rate on seen mammograms (training data) may drop and vice versa. Fourth and final drawback is that there are too many parameters such as learning rate, momentum, hidden units, hidden layer which should be adjusted/optimised during the training. The research work presented in this paper focuses on solving some of the major drawbacks by developing new architecture and a learning algorithm.

3. Research Methodology

The research methodology consists of 4 parts as follows: (1) acquisition of digitized mammograms from a benchmark database, (2) extraction of features, (3) selection of features (optional), and (4) classification of features (classifier). An overview of the research methodology is presented in Figures 1 and 2 and details are described in the following sections.

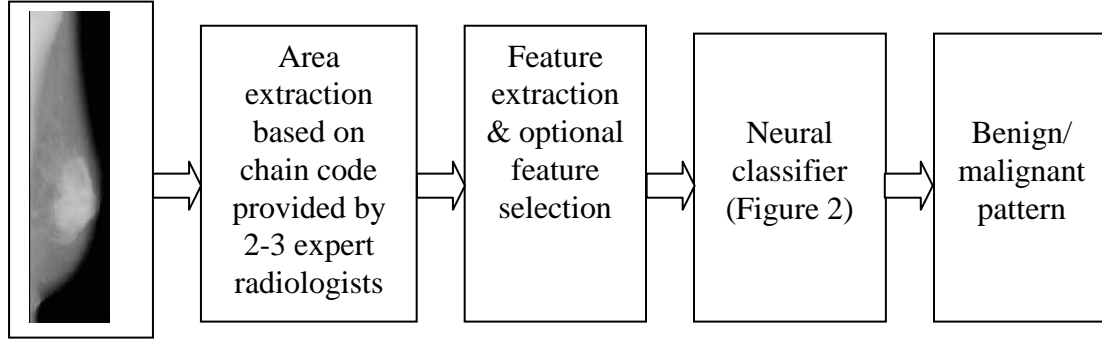


Figure 1. An overview of the proposed approach.

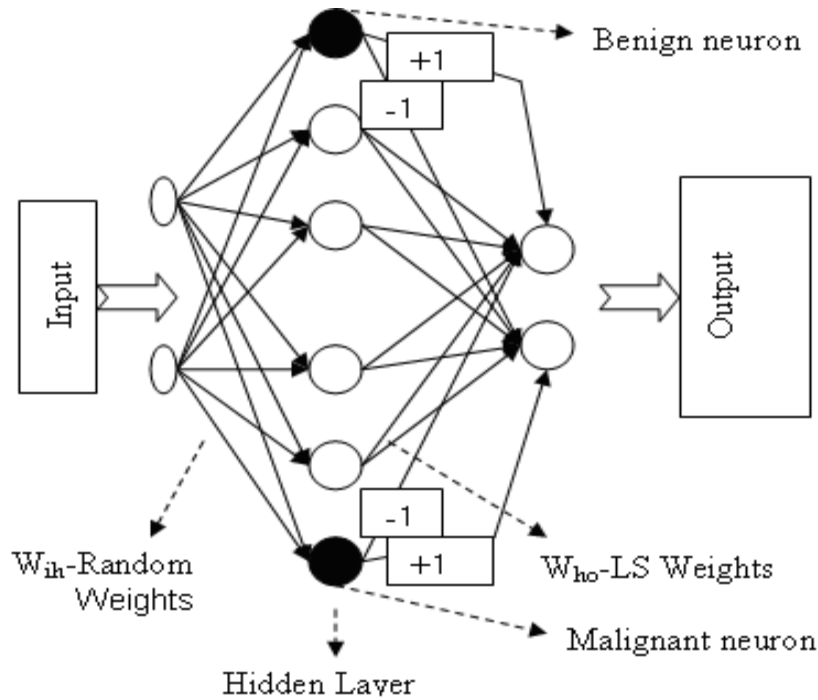


Figure 2. Proposed network architecture.

3.1. Benchmark Database of Digitized Mammograms

The digitized mammograms from university of South Florida's digitized database for screening mammography (DDSM) are used in this research. The reason for using this database is to compare the results of our study with other national and international researchers. DDSM is a benchmark database and widely used by researchers to carry out and evaluate their research work with other researchers in the area of CAD of breast cancer [32]. We have used this database in our previous research so we have already developed software to decode the database and extract features. The database contains approximately 2,500 studies of malignant, benign, benign-without-callback and normal cases. Digitized

mammograms of DDSM have 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. The 100 training and 100 test sets were taken as recommended by DDSM benchmark so that the results can be compared with other existing and future research. How training and testing sets were created by DDSM is described below. The description presented below is taken from DDSM's website.

Sampling a set of cases from the DDSM database to use for evaluating a spiculated mass detection algorithm required making some choices. A set of cases from the DDSM that had at least one, malignant, spiculated mass in it were selected. For simplicity, a set of cases that were all scanned on the same scanner and that all had the ground truth marked by the same radiologist were selected. The resulting set of cases were split into a training set and a test set using while attempting to balance the lesion subtlety and ACR breast density in the two datasets. The resulting list of cases can be found at <http://marathon.csee.usf.edu/Mammography/Database.html>. Each case contains four mammograms from a screening exam. The images were scanned on a HOWTEK 960 digitizer with a sample rate of 43.5 microns at 12 bits per pixel. The images were preprocessed to crop out much of the image that did not contain imaged breast tissue and to darken regions of the image that contained patient information or technician identifiers by setting pixels in those regions to the value zero. Each image was then compressed using a truly lossless compression algorithm.

3.2. Feature Extraction

Once all suspicious areas are extracted from digitized mammograms, for each suspicious area of digitized mammogram, grey level features are then extracted employing statistical formulas on grey level values of suspicious area and, BI-RADS lesion descriptor features, patient age feature and the subtlety value feature are extracted from associate case information files. The selected features represent the textural properties and morphological structure of an abnormality. These measures are very important considerations for effective classification of breast abnormality patterns. Extracted values of each feature are normalised to be consistent between 0 and 1.

3.2.1. Grey Level-based Features

Grey level-based features are calculated using statistical formulas on the grey level pixel values of each suspicious area and respective surrounding boundary area. A total of 14 features are considered as grey

level features in this study. All 14 features/formulas are briefly described below. Refer to [8] for detail description.

Number of Pixels: it is a count of the pixels in the extracted suspicious area.

Average Boundary Grey: it is the average grey level value around the outside of the suspicious area.

$$\text{Average Histogram} = \frac{1}{k} \sum_{j=0}^{k-1} N(j) / T \quad (1)$$

$$\text{Average Grey} = \frac{1}{T} \sum_{g=0}^{T-1} I(g) \quad (2)$$

$$\text{Difference} = \text{Average Grey} - \text{Average Boundary Grey} \quad (3)$$

$$\text{Contrast} = \frac{\text{Difference}}{\text{AvgGrey} + \text{AvgBoundryGrey}} \quad (4)$$

$$\text{Energy} = \sum_{j=0}^{k-1} [P(j)]^2 \quad (5)$$

$$\text{Modified Energy} = \sum_{g=0}^{T-1} [P(I(g))]^2 \quad (6)$$

$$\text{Entropy} = - \sum_{j=0}^{k-1} P(j) \log_2 [P(j)] \quad (7)$$

$$\text{Modified Entropy} = - \sum_{g=0}^{T-1} P(g) \log_2 [P(I(g))] \quad (8)$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\sum_{j=0}^{T-1} (j - \text{AvgGrey})^2 P(j)} \quad (9)$$

$$\text{Modified Std. Deviation } (\sigma) = \sqrt{\sum_{g=0}^{T-1} (I(g) - \text{AvgGrey})^2 P(I(g))} \quad (10)$$

$$\text{Skew} = \frac{1}{\sigma^3} \sum_{j=0}^{k-1} (j - \text{AvgGrey})^3 P(j) \quad (11)$$

$$\text{Modified Skew} = \frac{1}{\sigma^3} \sum_{g=0}^{T-1} (I(g) - \text{AvgGrey})^3 P(I(g)) \quad (12)$$

Where,

T is the total number of pixels,

g is an index value of image I,

K is the total number of grey levels (i.e. 4,096),

j is the grey level value (i.e. 0-4,096),

$I(g)$ is the grey level value of pixel g in image I ,

$N(j)$ is the number of pixels with grey level j in image I ,

$P(I(g))$ is the probability of grey level value $I(g)$ occurring in image I , $P(g)=N(I(g))/T$,

$P(j)$ is the probability of grey level value j occurring in image I ,

$P(j) = N(j)/T$.

3.2.2. BI-RADS Lesion Descriptor Features

Each case in the DDSM contains information such as breast density, an abnormality description and abnormality assessment rank that were specified by an expert mammography radiologist using the BI-RADS lexicon [32]. This information is stored in '.ics' file and '.OVERLAY' file of each case. The density of breast tissues and abnormality shape/type and its distribution/outline/margin inside breast tissues are the key factors radiologists consider when judging the likelihood of cancer being present. Abnormality assessment rank suggests the severity of abnormality. Considering their importance in the human interpretation process BI-RADS features have been used to evaluate their significance in computer-aided classification of breast abnormalities in digital mammograms.

The morphological descriptors for mass lesions are mass shape and mass margin. So the total four BI-RADS lesion descriptor features are density, mass shape, mass margin, and abnormality assessment rank.

3.2.2.1. Breast Tissue Density

This feature gives information about the density of breast tissues rated in 1 to 4 according to BI-RADS standards by an expert radiologist [32].

3.2.2.2. Morphological Description of Breast Abnormalities

Radiologists consider the arrangement and structural formation of abnormalities within the breast tissue to distinguish the abnormality into cancerous (malignant) and non-cancerous (benign) classes. Thus morphological descriptions of breast abnormality are considered as features. Like other features, morphological descriptions of breast abnormality are encoded into numeric values to get real feature values. The encoded numeric feature values for mass shapes (round, oval, lobulated, irregular,

architectural_distortion, tubular, lymph node, asymmetric_breast_tissue, focal_asymmetric_density) and margins (circumscribed, microlobulated, obscured, ill_defined, spiculated) features found in DDSM cases [32] are 1 to 9 and 1 to 5 respectively.

3.2.2.3. Abnormality Assessment Rank

It categorises the seriousness of the abnormality found and suggests the course of the action should be taken accordingly. It is rated in 1 to 5 by an expert radiologist [32].

3.2.2.4. Patient Age Feature

The DDSM also provides information such as the patient age at the time the mammogram was taken for each case study [32]. Various medical findings and past breast cancer statistics show that the risk of breast cancer increases with age. This makes patient age a very useful feature for medical practitioners to consider in diagnosis of breast cancer. However the current breast cancer statistics reveal many exceptions to this correlation. In recognition of the correlation and exceptions, patient age feature is used in the proposed research work for neural classification of breast abnormalities.

3.2.2.5. Subtlety Value Feature

This is a subjective impression of the subtlety of a lesion by an expert radiologist. It is rated in 1 to 5 by an expert radiologist where 1 is "subtle" and 5 is "obvious". A higher subtlety rating indicates a more obvious lesion. The subtlety value for a lesion may indicate how difficult it is to find [32]. Subtlety of lesion plays an important role in their interpretation as subtle mammographic lesions are associated with significantly different search parameters than obvious lesions. Cancerous abnormalities are more subtle with ill-defined edges compared with non-cancerous abnormalities. This is one of the key factors which affect the overall interpretation accuracy.

3.3. Classifier

As mentioned in Section 2, many classifiers including neural classifiers and traditional classifiers [7-44] have been used to classify suspicious areas in digitized mammograms into benign and malignant patterns. It has been shown [7, 15, 43] that neural classifiers outperformed traditional classifiers. However, there are many problems (1) current classifiers can misclassify even known cancer cases (e.g. 80% accuracy on training set means that 20% known cases will be misclassified), (2) difficult to train (may take many days to learn), (3) do not provide consistency (accuracy rates are different every time

you train on same or on different database), and (4) no guarantee for solution (sometime there are no results after a few days of training). In this study, novel network architecture and a learning algorithm to overcome these problems are introduced and investigated. An overview of the network architecture is presented in Figure 2. As seen in Figure 2, two additional neurons for benign and malignant patterns are introduced which are connected to the output of the network. The main idea behind this is to improve the memorization/association abilities of the network without destroying/degrading generalization capabilities. It is well known fact in traditional classifier learning that we aim to achieve a high recognition/classification rate on test data, but we forget that the classification on training data may suffer. With current learning algorithms, it is impossible to get 100% classification rate on training data and a high classification rate on test data which means that the network may incorrectly classify already known cases. In situation such as diagnosis of breast cancer, radiologists or doctors are most unlikely to use the network which misclassifies obvious cases (true cases) from the database. The new network with 100% memorisation capabilities doesn't misclassify any benign or malignant pattern from training data or very close to training data. The weights (input-hidden-output) of additional neurons are trained differently (refer to learning algorithm) than the rest of the network.

A new learning algorithm for training of the network in Figure 2 is investigated in this study. The main idea behind the new learning algorithm is that it is possible to minimize output error based on calculation of weights between hidden layer and output layer using least square methods. The nonlinear sigmoidal function at the output can be converted into a linear function by using a log function. A linear system can be established for each output class and the output weights can be calculated by solving the linear system. The weights between input and the hidden layers are randomly selected. By doing this we are removing traditional gradient based neural networks' problems such as local minima, paralysis, long training time, uncertainty, etc. We will also have a 100% guarantee that there will always be a solution [34]. There will be no such situation that after many hours we do not have any solution. The use of this idea also makes the training very fast which solves the problem of slow adaptation of new knowledge from new mammograms. The weights for two additional hidden nodes/neurons (black top and bottom circles in Figure 2) are trained separately using the following idea.

Minimal distance based calculation of weights for additional neurons

The basic idea is to find a distance between two patterns which is minimal. This principle has been used in many well known statistical and intelligent techniques [43] such as nearest neighbour, hamming

network and distance based clustering, to find the similarity in patterns. The idea here in the proposed approach is to maximise the output value of the class which the input pattern belongs to. The process is independent to the rest of the training and the output neuron for only input class is fired (e.g. if input test pattern is from benign database then the output value for benign class will be higher than the output value for malignant class). The weights between input and benign neuron are set using the benign training patterns and the weights between input and malignant neuron are set using malignant training patterns. The following function is formulated and investigated in this study.

$$W_{cMal} = \log\left(\frac{\text{target}}{1 - \text{target}}\right) * e^{-\min(\|x - X_{allMalPatterns}\|) * \log\left(\frac{\text{target}}{1 - \text{target}}\right)} \quad (13)$$

Where target is the desired output for malignant class, x is an input test pattern and $X_{allMalPatterns}$ is a set of all malignant patterns from training set.

The function maximizes the output value for the test input which has exact match in training data. The value decreases according to closeness between the test input and the training data. The aim of the *min* function is to find a minimal value between the input and X (weights). The output from additional neuron is passed to the following activation function.

$$f(out) = \begin{cases} out & \text{if } out \geq \log\left(\frac{\text{target}}{1 - \text{target}}\right) \\ out * gen & \text{if } out < \log\left(\frac{\text{target}}{1 - \text{target}}\right) \end{cases} \quad (14)$$

Where *gen* should be less than 0.5 for 100% classification on training data. The output value is added to network's output before activation function. The aim of the above activation function is to produce the maximum value for the output of the network only if the input pattern is from training data or similar to training data.

The following steps are used to train the classifier network.

Step 1: Set network's initial parameters

n - inputs (number of features),

h - hidden units (number of hidden units, for example, start with h=2)

m - outputs (number of classes, for example m=2 as we have malignant and benign)

p- training pairs (total number of benign and malignant samples for training).

Step 2: Set the weights between inputs and hidden units.

The weights between inputs and hidden units are set using the following two steps.

Step 2.1 Set weights (W_{ih}) between inputs (X_n) and hidden units (H_h)

The weights are set to small random values. Use rand function in C++.

Step 2.2 Set Weights between inputs (X_n) and 2 additional hidden neurons/units (H_{ben}, H_{mal})

The weights for additional hidden neurons are calculated using minimal distance approach as described above in Section C3. Each weight is a vector and the length of vector is equal to p (eg.

$W_{iben[n][p]} = X$ (all benign patterns)).

Step 3: Set the weights between hidden units and outputs

The weights ($W_{ho}, W_{beno}, W_{malo}$) between hidden units and outputs are set by calculating them as shown in steps below.

Step 3.1 Feed each input pattern to the network and calculate output of the hidden layer (H_h).

Step 3.2 Calculate the output value (O_{baf}) before the activation function as follows.

$$O_{baf} = \log\left(\frac{\text{target}}{1 - \text{target}}\right) \quad (15)$$

Step 3.3 Set a linear system of equations by using H_h , O_{baf} and W_{ho}

$H_h W_{ho} = O_{baf}$ - use least square methods (e.g. Gram-Schmidt) to calculate W_{ho}

Step 3.4 Repeat Steps 3.2 and 3.3 for each output.

Step 3.5 Set weights W_{beno}, W_{malo} to +1 and -1 as shown in Figure 2.

Step 4: Increment #HU (hidden unit), repeat Steps 2-3, stop when the network starts memorizing instead of generalizing (find a number in terms of training patterns (e.g. p/4, p/2)).

Step 5: Select #HU (hidden unit) with the best accuracy results on test set of training data.

The following steps are used to test the network.

Step 1: Feed test input (X) and output to the network.

Step 2: Calculate the output of the hidden layer.

The output (except benign and malignant neurons) is calculated as follows:

$$\text{net} = \text{sum}(X.W) \quad (16)$$

$$\text{output} = f(\text{net}) = \frac{1}{1 + e^{-\text{net}}} \quad (17)$$

The output for benign and malignant neurons is calculated as follows:

$$\text{net} = \log\left(\frac{\text{target}}{1 - \text{target}}\right) * e^{-\min(\|x - X_{allMalPatterns}\|) * \log\left(\frac{\text{target}}{1 - \text{target}}\right)} \quad (18)$$

$$f(net) = \begin{cases} net & \text{if } net \geq \log\left(\frac{target}{1-target}\right) \\ net*0.5 & \text{if } net < \log\left(\frac{target}{1-target}\right) \end{cases} \quad (19)$$

Step 3: Calculate the output of the network.

The output of the network is calculated as follows.

$$net = \text{sum}(H.W) \quad (20)$$

$$output = f(net) = \frac{1}{1 + e^{-net}} \quad (21)$$

3.4. Theoretical Underpinning

Let $X (x_1, x_2, \dots, x_n)$ be the input, $Y (y_1, y_2, \dots, y_n)$ be the output, W be the weight matrix of hidden layer and W' the weight matrix of the output layer.

Let x be the training pattern from $X (x \in X)$, M be the set of all malignant patterns and B be the set of all benign patterns. The output of the proposed network A and A' can be calculated as follows.

$$A \sim (xW)W' + \log\left(\frac{target}{1-target}\right) * e^{-\min(\|x-M\|) * \log\left(\frac{target}{1-target}\right)} \quad (22)$$

$$A' \sim (xW)W' + \log\left(\frac{target}{1-target}\right) * e^{-\min(\|x-B\|) * \log\left(\frac{target}{1-target}\right)} \quad (23)$$

To prove that the proposed approach will always work, it is shown below that $A > A'$ for all $x \in M$ and $A' > A$ for all $x \in B$.

For all $x \in M$,

$$e^{-\min(\|x-M\|) * \log\left(\frac{target}{1-target}\right)} > e^{-\min(\|x-B\|) * \log\left(\frac{target}{1-target}\right)} \quad (24)$$

$$\Rightarrow \log\left(\frac{target}{1-target}\right) * e^{-\min(\|x-M\|) * \log\left(\frac{target}{1-target}\right)} > \log\left(\frac{target}{1-target}\right) * e^{-\min(\|x-B\|) * \log\left(\frac{target}{1-target}\right)} \quad (25)$$

$$\Rightarrow A > A'$$

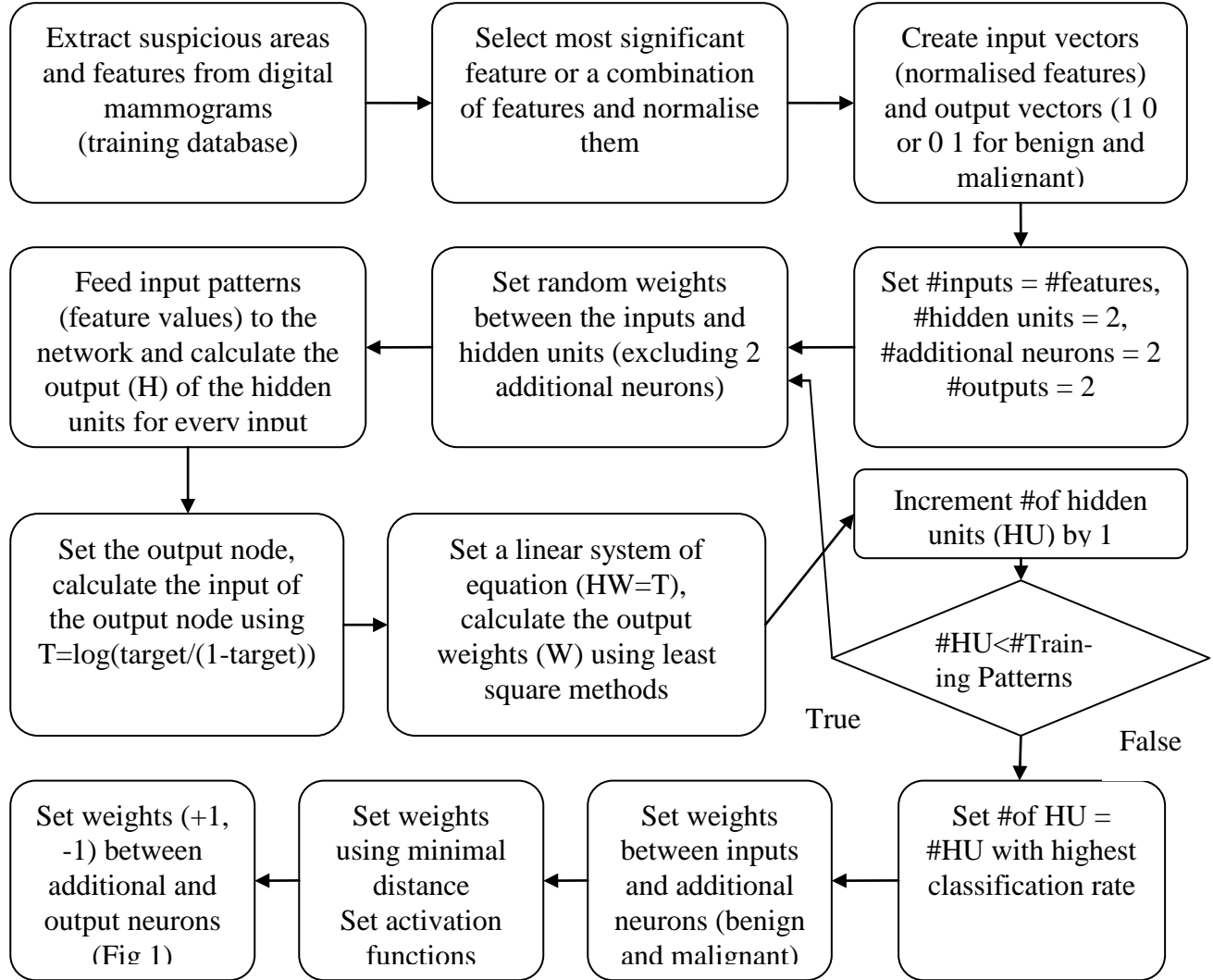
For all $x \in B$

$$e^{-\min(\|x-B\|) * \log\left(\frac{target}{1-target}\right)} > e^{-\min(\|x-M\|) * \log\left(\frac{target}{1-target}\right)} \quad (26)$$

$$\Rightarrow \log\left(\frac{\text{target}}{1-\text{target}}\right) * e^{-\min(\|x-B\|) * \log\left(\frac{\text{target}}{1-\text{target}}\right)} > \log\left(\frac{\text{target}}{1-\text{target}}\right) * e^{-\min(\|x-M\|) * \log\left(\frac{\text{target}}{1-\text{target}}\right)} \quad (27)$$

$\Rightarrow A' > A$

Flow Chart-Training Process



Flow Chart-Test Process

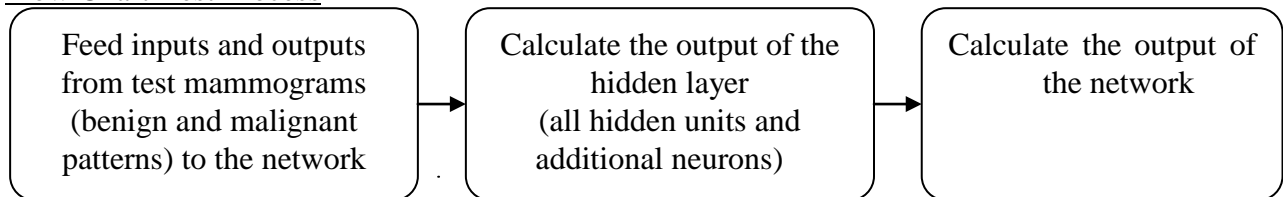


Figure 3. Flow-chart for training and test processes

4. Implementation and Experimental Results

The proposed approach has been implemented in C++ on Windows XP platform and UNIX platform. The experiments included in this paper were conducted on Windows XP (Pentium 4 machine) using the approach presented in this paper and other approaches such as MLP-BP (MLP trained using back-propagation) MLP-GA (MLP trained using GA), DA (discriminatory analysis) and LR (logistic regression). The experiments were conducted using the same database and computing environment. Table 1 presents results with 6 BI-RADS features (radiologist interpretation) and varying parameters such as iterations and hidden units. The best results are recorded here. Table 2 presents results with 14 grey level features and all 6 BI-RADS features. Table 3 presents the results with 14 grey level features and only 4 BI-RADS features.

Table 1. Results using 4 BI-RADS features, patient age feature and subtlety value feature

	Performance			
	# of hidden units	# of iterations	Classification rates	
			Training set	Test set
Proposed approach	10	1	100	84
	10	947	100	93
	20	1	100	84
	20	1560	100	94
	32	1	100	90
	32	1190	100	94
MLP-BP	10	70000	99	91
	10	80000	100	83
	20	10000	95	91
	20	60000	100	88
MLP-GA	18	NA	96	87
DA	NA	NA	85	88
LR	NA	NA	93	90

Table 2. Results using 14 grey level based features, 4 BI-RADS features, patient age feature and subtlety value feature

	Performance			
	# of hidden units	# of iterations	Classification rates [%]	
			Training set	Test set
Proposed approach	15	1	100	90
	15	1937	100	94
	16	1	100	88
	16	1043	100	93
	20	1	100	85
	20	2010	100	91
MLP-BP	16	30000	100	84
	16	40000	100	90
	20	10000	100	91
MLP-GA	18	NA	81	89
DA	NA	NA	90	88
LR	NA	NA	97	89

Table 3 Results using 14 grey level based features and 4 BI-RADS features

	Performance			
	# of hidden units	# of iterations	Classification rates [%]	
			Training set	Test set
Proposed approach	10	1	100	75
	10	2649	100	94
	12	1	100	91
	12	1808	100	94
	14	1828	100	93
MLP-BP	10	30000	100	90
	14	20000	99	91
MLP-GA	18	NA	81	89
DA	NA	NA	90	88
LR	NA	NA	97	89

4. Discussion and Analysis of Results

The results using the proposed approach on three types of feature combinations have been presented in Tables 1-3. The ROC curves using results from experiments have been displayed in Figures 4-6. The proposed approach obtained highest classification rate (94%) for all three feature combinations. It achieved the highest classification rate with just a few iterations. It is good to notice that classification rate on training set is always 100% which means that the proposed approach has a good memorizing ability with an excellent generalization ability. Training in just a few iterations means that the new data can be adapted just in few minutes.

The obtained results with highest classification rates were analyzed for finding benign and malignant class errors, classes with consistent errors and impact of hidden units. As listed in the Tables 1-3, the best results produced 6% misdiagnosed regardless of feature combinations used. Table 4 presents the analysis of results with two different hidden units, same input features and the same classification rates. In both cases (20 and 32 hidden units), the malignant class has only 1 misclassification error and the benign class has 5 misclassification errors. However, all misclassified classes were not the same. The class pattern numbers (mammograms) 14, 38 and 94 were consistent and misclassified in both cases.

Table 4. Different hidden units, same features (bi-rads, patient age and subtlety value features) and same classification rate

Hidden units: 20, Classification rates: 94%				Hidden units: 32, Classification rates: 94%			
Case #	Correct outputs	Network outputs	Network class	Case #	Correct outputs	Network outputs	Network class
14	0.1 0.9 (benign)	0.782 0.218	malignant	14	0.1 0.9 (benign)	0.720 0.280	malignant
38	0.1 0.9 (benign)	0.912 0.088	malignant	38	0.1 0.9 (benign)	0.931 0.069	malignant
64	0.1 0.9 (benign)	0.552 0.448	malignant	64	0.1 0.9 (benign)	0.327 0.673	benign
81	0.9 0.1 (malignant)	0.462 0.538	benign	81	0.9 0.1 (malignant)	0.504 0.496	malignant
92	0.1 0.9 (benign)	0.634 0.366	malignant	92	0.1 0.9 (benign)	0.403 0.597	benign
94	0.1 0.9 (benign)	0.964 0.036	malignant	94	0.1 0.9 (benign)	0.997 0.003	malignant
				50	0.1 0.9 (benign)	0.529 0.471	malignant
				62	0.1 0.9 (benign)	0.806 0.194	malignant
				45	0.9 0.1 (malignant)	0.391 0.609	benign
Benign errors ¹ : 5, Malignant errors ² : 1 Sensitivity: 98% Specificity: 90%				Benign errors ¹ : 5, Malignant errors ² : 1 Sensitivity: 98% Specificity: 90%			
Patterns 14, 38, 94 are misclassified in both cases.							
¹ Benign error – Benign classified as malignant, ² Malignant error – Malignant classified as benign							

Table 5 presents the analysis of results with two different hidden units, different classification rates and same features (14 grey level based features, 4 BI-RADS features, patient age feature and subtlety value feature). It is interesting to see that the change of features resulted in more misclassifications in malignant classes. The misclassification errors for benign and malignant are same (3 each for 15 hidden units). The consistency of misclassified patterns is better as 5 misclassified classes (14, 38, 45, 75, 81) were same in both cases (hidden units 15 and 16). The grey-level based features provided more misclassification malignant cases than BI-RADS features alone.

Table 5. Different hidden units, same features and different classification rates

Hidden units: 15, Classification rates: 94%				Hidden units: 16, Classification rates: 93%			
Case #	Correct outputs	Network outputs	Network class	Case #	Correct outputs	Network outputs	Network class
2	0.1 0.9 (benign)	0.503 0.497	malignant				
14	0.1 0.9 (benign)	0.709 0.291	malignant	14	0.1 0.9 (benign)	0.789 0.211	malignant
38	0.1 0.9 (benign)	0.834 0.166	malignant	38	0.1 0.9 (benign)	0.885 0.115	malignant
45	0.9 0.1 (malignant)	0.338 0.662	benign	45	0.9 0.1 (malignant)	0.343 0.657	benign
75	0.9 0.1 (malignant)	0.318 0.682	benign	75	0.9 0.1 (malignant)	0.240 0.760	benign
81	0.9 0.1 (malignant)	0.268 0.732	benign	81	0.9 0.1 (malignant)	0.351 0.649	benign
				63	0.9 0.1 (malignant)	0.414 0.586	benign
				94	0.1 0.9 (benign)	0.820 0.180	malignant
Benign errors ¹ : 3, Malignant errors ² : 3				Benign errors ¹ : 3, Malignant errors ² : 4			
Sensitivity: 94% Specificity: 94%				Sensitivity: 92% Specificity: 94%			
Patterns 14, 38, 45, 75, 81 are misclassified in both cases.							
¹ Benign error – Benign classified as malignant, ² Malignant error – Malignant classified as benign							

The proposed network can automatically find the number of hidden units required for the best solution (highest classification rate on test data). It gives you the best number of hidden units in the first run. It was found that the number of hidden units with highest classification rates in the first run can achieve the best solution. This was true for all the experiments.

The results obtained by the proposed approach are better than MLP-BP, MLP-GA, DA and LR. The results are also much better than the most of the recently published results in the literature. Sometime it is very difficult to compare results from various techniques because researchers use different database. We have selected some recent papers with results using DDSM database for comparison purposes.

Zhang et al. [9] used same DDSM benchmark database and reported the highest 90.5% classification rate for the calcification cases and 87.2% classification rate for the mass cases. The highest results obtained (94%) in this research are for mass cases so nearly 7% improvement has been achieved. Wroblewska et al. [27] have used the same DDSM benchmark database and reported 76% classification rate which is much lower (18%) than the obtained classification rate in this research. Masotti et al. [40] have used DDSM benchmark database and support vector machine (SVM), they achieved 90% accuracy which is 4% less than obtained accuracy in this research. The Figures 4-6 and Tables 6 and 7 show the ROC curves and areas under the curves for various parameters such as hidden units, classification accuracy, etc. Overall, the proposed technique has outperformed all the existing neural network based and statistical techniques.

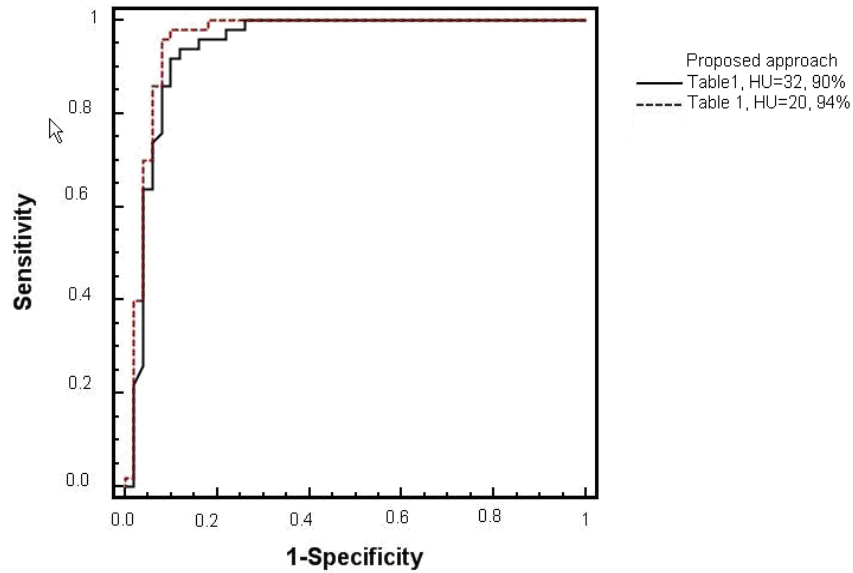


Figure 4. ROC curve for results in table 1

Table 6. Area under the curve

	Area	Std. error
Table 1 (Proposed approach, #HU=20, accuracy=94%)	0.957	0.024
Table 1 (Proposed approach, #HU=32, accuracy=90%)	0.943	0.027

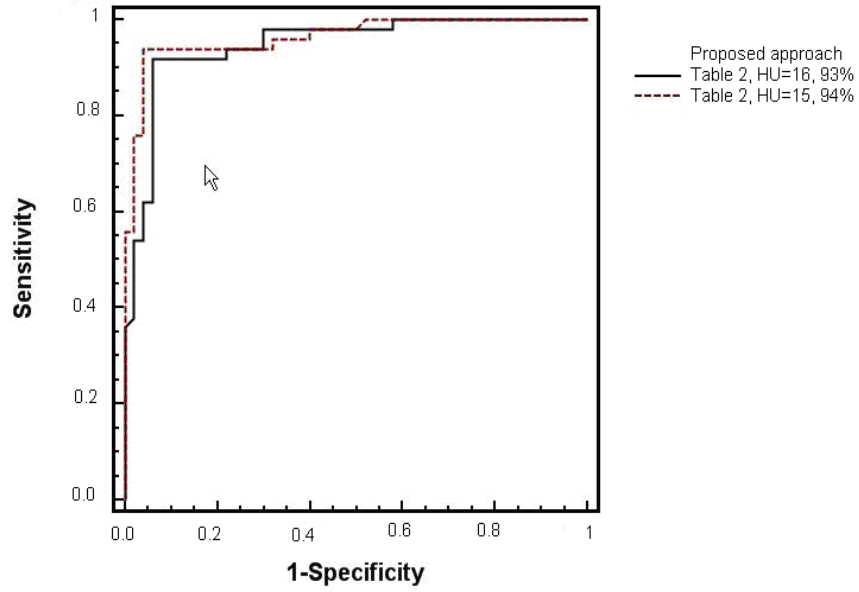


Figure 5. ROC curve for results in table 2

Table 7. Area under the curve

	Area	Std. error
Table 2 (Proposed approach, #HU=15, accuracy=94%)	0.964	0.017
Table 2 (Proposed approach, #HU=16, accuracy=93%)	0.947	0.022

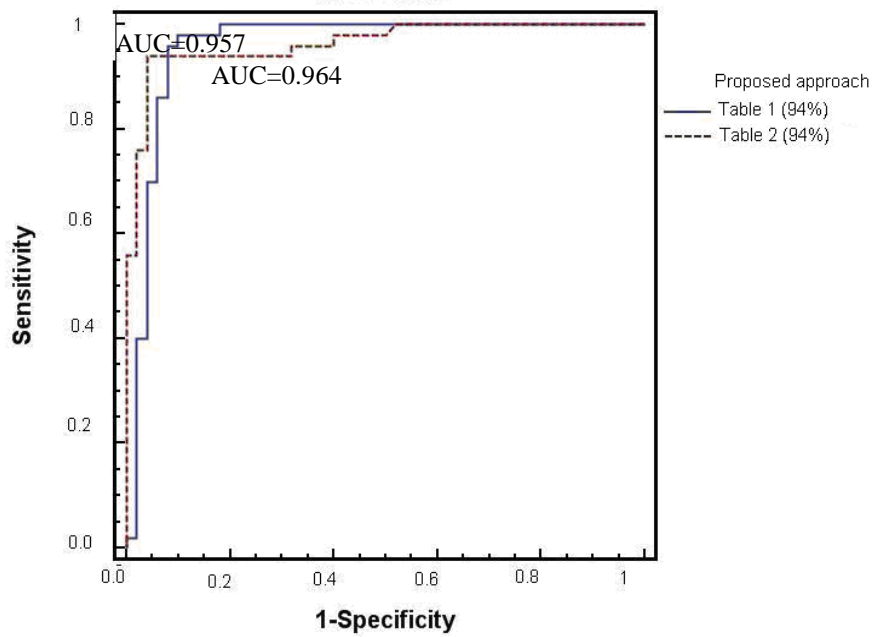


Figure 6. ROC curve and area under curve obtained for different features with the same 94% accuracy.

6. Conclusions

We have presented a novel neural architecture and a learning algorithm for the classification of mass abnormalities in digitized mammograms. The proposed methodology has been implemented and various experiments on a digital mammography benchmark database have been conducted. The experiments using the proposed approach on a benchmark database produced 94% classification accuracy on test set and 100% classification accuracy on training set. The results are very promising. The proposed approach has significantly improved the results in terms of classification rates, number of iterations, memorization, generalization and fast guaranteed training.

There are four major advantages of the proposed approach over existing approaches for breast cancer diagnosis. First advantage is that the medical community in particular radiologists can have 100% assurance that the system based on the proposed architecture and learning algorithm is not going to misclassify patterns from already classified/diagnosed database. Second advantage is that the new approach can adapt and learn new knowledge from new training patterns very quickly. Third advantage is that it can generalize much better than multilayer perceptron with standard backpropagation algorithm. Fourth and final advantage is that there are not many parameters such as learning rate, momentum, etc. which should be adjusted during the training.

The various sets of statistical and breast imaging reporting and data system based features were investigated. The different features provided different classification errors. A feature selection process can be implemented and used which may further improve the classification accuracy.

In our future research, we would like to conduct more experiments on DDSM and other benchmark databases. The various new activation functions for additional and output neurons will be investigated. We would like to further test the automatic finding of hidden units for large number of training data. A feature selection process will also be implemented and incorporated with the proposed approach.

Acknowledgments

Author would like to thank Rinku Panchal, a PhD student at Central Queensland University and Ping Zhang, a PhD student at Bond University for helping with MLP-BP, MLP-GA, DA, and LA experiments.

References

- [1] Parkin, M., Freddie Bray, F, Ferlay, J. and Pisani, P. “Global Cancer Statistics”, A Cancer Journal for Clinicians, Vol. 55, pp. 74-108, 2005, American Cancer Society.
- [2] Breast Cancer Facts, 2005. www.breastcancerfund.org (Accessed: 12 September 2007).
- [3] National Breast Cancer Centre, <http://www.nbcc.org.au/bestpractice/statistics/> (Accessed: 12 September 2007).
- [4] American College of Radiology. www.acr.org (Accessed: 12 September 2007).
- [5] NSW Breast Cancer Institute, <http://www.bci.org.au> (Accessed: 12 September 2007).
- [6] Pisano, E. D., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, K., Acharyya, S., Conant. F., Fajardo, L., Bassett, L., D'Orsi, C., Jong, R. and Rebner, M., “Diagnostic Performance of Digital Versus Film Mammography for Breast Cancer Screening”, New England Journal of Medicine, Vol. 353, pp. 1773-83, 2005.
- [7] Cheng, H. D., Cai, X., Chen, X., Hu, L., and Lou, X., “Computer-aided Detection and Classification of Microcalcification in Mammograms: a survey”, Pattern Recognition, Vol. 36, pp. 2967-2991, 2003.
- [8] Verma, B. and Zakos, J., “A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques”, IEEE Transactions on Information Technology in Biomedicine, Vol. 5, pp.46-54, 2001.
- [9] Zhang, P., Verma, B., and Kumar, K., “A Neural-Genetic Algorithm for Feature Selection and Breast Abnormality Classification in Digital Mammography”, Proceedings of IEEE International Joint Conference on Neural Networks, IEEE, Piscataway, NJ, Vol. 3, pp. 2303-09, 2004.
- [10] Zhang, P., Verma, B. and Kumar, K. Neural Vs. Statistical Classifier in Conjunction with Genetic Algorithm Feature Selection in Digital Mammography, Pattern Recognition Letters, 26(7), pp. 909-919, 2005.
- [11] Wu, Y., He J., Man Y., and Arribas J. I., “Neural Network Fusion Strategies for Identifying Breast Masses”, Proceedings of IEEE International Joint Conference on Neural Networks, IEEE, Piscataway, NJ, Vol. 3, pp. 2437-2442, 2004.
- [12] Chitre Y., Dhawan A. P., and Moskowitz M., “Artificial Neural Network Based Classification of Mammographic Microcalcifications using Image Structure Features”, State of The Art in Digital Mammographic Image Analysis, Editors: Kevin W. Bowyer, Sue Astley, World Scientific, Singapore, Vol. 9, pp. 167-197, 1994.

- [13] Christoyianni L., Dermatas E., and Kokkinakis G., “Neural Classification of Abnormal Tissues in Digital Mammography using Statistical Features of the Texture”, Proceedings of The 6th IEEE International Conference on Electronics, Circuits and Systems, IEEE, Piscataway, NJ, Vol. 1, pp. 117-120, 1999.
- [14] Bovis K., Singh S., Fieldsend J., and Pinder C., “Identification of Masses in Digital Mammograms with MLP and RBF Nets”, Proceedings of IEEE International Joint Conference on Neural Networks, IEEE, Piscataway, NJ, Vol. 1, pp. 342 – 347, 2000.
- [15] 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.
- [16] Markey, M., Lo, J., Tourassi, G., and Floyd, C., “Self-organizing Map for Cluster Analysis of a Breast Cancer Database”, Artificial Intelligence in Medicine, 27, pp. 113–127, 2003.
- [17] Wu, Y., “Application of Neural Networks in Mammography”, Radiology, vol. 187, pp. 81-87, 1993.
- [18] Guran, M., Chan, H., Sahiner, B., Hadjiiski, L., Petrick, N., and Helvie, M., “Optimal Neural Network Architecture Selection: Improvement in Computerized Detection of Microcalcifications”, Academic Radiology, 9 (4), pp. 420–29, 2002.
- [19] Guran, M., Sahiner, B., Chan, H., Hadjiiski, L., and Petrick, N., “Selection of an Optimal Neural Network Architecture for Computer-Aided Detection of Microcalcifications—Comparison of Automated Optimization Techniques”, Medical Physics, 28 (9), pp. 1937–1948, 2001.
- [20] Papadopoulos A., Fotiadis D. I., and Likas A., “An Automatic Microcalcification Detection System Based on a Hybrid Neural Network Classifier”, Artificial Intelligence in Medicine, Vol. 25, pp. 149-167, 2002.
- [21] Lo, J. Y., Land, W. H., Morrison, C. T., “Application of Evolutionary Programming and Probabilistic Neural Networks to Breast Cancer Diagnosis”, Proceedings of IEEE International Joint Conference on Neural Networks, IEEE, Piscataway, NJ, Vol. 5, pp. 3712-3716, 1999.
- [22] Lo, J. Y., Land W. H., Morrison, C. T., “Evolutionary Programming Technique for Reducing Complexity of Artificial Neural Networks for Breast Cancer Diagnosis”, Proceedings of SPIE (Society of Photo-Optical Instrumentation Engineers), SPIE, Bellingham, WA, USA, Kenneth M. Hanson (editor), Vol. 3979, pp. 153-158, 2000.
- [23] Pena-Reyes, C. A., and Sipper, M., “A Fuzzy-Genetic Approach to Breast Cancer Diagnosis”, Artificial Intelligence in Medicine, Vol. 17, pp. 131-155, 1999.

- [24] Cheng, H., Lui, Y. M., and Freimanis, R. I., "A Novel Approach to Microcalcification Detection using Fuzzy Logic Technique", *IEEE Trans. on Medical Imaging*, Vol. 17, pp. 442-450, 1998.
- [25] Yoshida, H., Nishikawa, R. N., Geiger, M.L., and Doi, K., "Signal/Background Separation By Wavelet Packets for Detection of Microcalcifications in Mammograms", *Proceedings of SPIE on Wavelet Applications in Signal and Image Processing*, SPIE, Bellingham, WA, USA, Michael A. Unser, Akram Aldroubi, Andrew F. Laine (editors), Vol. 2825, pp 805-811, 1996.
- [26] Wang, T. C., and Karayiannis, N. B., "Detection of Microcalcification in Digital Mammograms using Wavelets", *IEEE Transactions on Medical Imaging*, Vol. 17, pp. 498-509, 1998.
- [27] Wroblewska, A., Boninski, P., Przelaskowski, A., and Kazubek, M., "Segmentation and feature Extraction for Reliable Classification of Microcalcification in Digital Mammograms", *Opto-Electronics Review*, Vol. 11, pp. 227-235, 2003.
- [28] Lo, J. Y., Gavrielides, M., Markey, M. K., and Jesneck, J. L., "Computer-aided Classification of Breast Microcalcification Clusters: Merging of Features from Image Processing and Radiologists", *Proceedings of SPIE on Medical Imaging: Image Processing*, SPIE, Bellingham, WA, USA, Milan Sonka, J. Michael Fitzpatrick (editors), Vol. 5032, pp. 882-889, 2003.
- [29] Bovis, K., and Singh, S., "Detection of Masses in Mammograms using Texture Measures", *Proceedings of IEEE International Conference on Pattern Recognition*, IEEE, Piscataway, NJ, Vol. 2, pp. 267-270, 2000.
- [30] Markey, M. K., Lo, J. Y., and Floyd, C. E., "Differences between Computer-aided Diagnosis of Breast Masses and That of Calcifications", *Radiology*, Vol. 223, pp. 489-493, 2002.
- [31] Yu, S., and Guan, L., "A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films", *IEEE Transactions on Medical Imaging*, Vol. 19, pp. 115-126, 2000.
- [32] Heath, M., Bowyer, K., Kopans, D., Moore, R., and Kegelmeyer, Jr. P., "The Digital Database for Screening Mammography", *Proceedings of International Workshop on Digital Mammography*, Medical Physics Publishing, Yaffe, M.J. (editor), Madison, USA, pp. 212-218, 2000.
- [33] Rangayyan, R. M., Mudigonda, N. R., and Leo, J. E., "Boundary Modeling and Shape Analysis Methods for Classification of Mammographic Masses", *Medical and Biological Eng. and Computing*, Vol. 38, pp. 487-496, 2000.
- [34] Verma, B. "Fast Training of Multilayer Perceptrons (MLPs)", *IEEE Transactions on Neural Networks*, vol. 8, no. 6, pp. 1314-1321, 1997.

- [35] Sahiner, B., Chan, H., Petrick, N., Helvie, M. and Goodsitt, M., “Design of a High-Sensitivity Classifier Based on a Genetic Algorithm: Application to Computer-Aided Diagnosis”, *Physics in Medicine and Biology*, 43, pp. 2853–2871, 1998.
- [36] Anastasio, M., Yoshida, H., Nagel, R., Nishikawa, R., and Doi, K., “A Genetic Algorithm-Based Method for Optimizing the Performance of a Computer-Aided Diagnosis Scheme for Detection of Clustered Microcalcifications in Mammograms”, *Medical Physics*, 25, pp. 1613-1620, 1998.
- [37] Lee, Y., and Tsai, D., “Computerized Classification of Microcalcifications on Mammograms using Fuzzy Logic and Genetic Algorithm”, *Proceedings of SPIE on Medical Imaging*, SPIE, Bellingham, WA, USA, J. Michael Fitzpatrick, Milan Sonka (editors), Vol. 5370, pp 952-959, 2004.
- [38] Khuwaja, G. A., and Abu-Rezq, A. N., “Bi-Model Breast Cancer Classification System”, *Pattern Analysis and Applications*, Vol. 7, pp. 235-242, 2004.
- [39] Shen, L., Rangayyan R. M., and Desautels, J. E. L., “Detection and Classification of Mammographic Calcifications”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 7, pp. 1403-1416, 1993.
- [40] Masotti, M., “A Ranklet-based Image Representation for Mass Classification in Digital Mammograms”, *Medical Physics*, vol. 33, no. 10, pp. 3951-3961, 2006.
- [41] Sahiner, B., Petrick, N., Chan, H-P., Hadjiiski, L., Paramagul, C., Helvie, M., and Gurcan, M., Computer-Aided Characterization of Mammographic Masses: Accuracy of Mass Segmentation and Its Effects on Characterization, *IEEE Transactions on Medical Imaging*, Vol. 20, No. 12, pp. 1275-1284, 2001.
- [42] Sahiner, B., Chan, H-P., Petrick, N., Helvie, M., and Hadjiiski, L., Improvement of Mammographic Mass Characterization using Spiculation Measures and Morphological Features, *Medical Physics*, Vol. 28, No. 7, pp. 1455-1465, 2001.
- [43] Looney, C., “Pattern Recognition using Neural Networks”, Oxford University Press, New York, USA, 1997.
- [44] Kallergi, M, Carney, M and Gaviria, J. “Evaluating the Performance of Detection Algorithms in Digital Mammography”, *Medical Physics*, 26(2):267-75, 1999.