2006 International Joint Conference on Neural Networks
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

A Neural Learning Algorithm for the Diagnosis of Breast Cancer

Brijesh Verma

Abstract—**This paper presents a new learning algorithm for the diagnosis of breast cancer. The proposed algorithm with novel network architecture 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 BI-RADS features (radiologists' interpretation) from digital mammograms are extracted and used to train the network with the proposed learning algorithm. The new learning algorithm has been implemented and tested on a DDSM Benchmark database. The proposed approach has outperformed other existing approaches in terms of classification rate, generalization and memorization abilities, number of iterations, fast and guaranteed training. Some promising results and a comparative analysis of obtained results are included in this paper.**

I. Introduction

BREAST cancer continues to be the most common cause of cancer deaths in women. Every year more than a million women develop breast cancer world-wide [1-5]. In 2005, an estimated 1,150,000 women worldwide were diagnosed with breast cancer, and 411,000 women died from the disease [1]. A recent study [2] on breast cancer shows that one of every three cancer diagnosis in women is a breast cancer. A report by cancer institute estimates that 1 in 8 women develops breast cancer in US [1-2], 1 in 9 in UK and Canada [3-4], and 1 in 11 in Australia [5]. It also reports that nearly 3% women die from breast cancer worldwide,with risk increases with age particularly after 50.

At present digital mammography is considered to be one of the most reliable methods for early detection of breast cancer. 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 corresponds directly to an increase in routine mammography screening [6].

The main objective of this research work is to investigate a new architecture and a new learning algorithm for the classification of benign and malignant patterns. The research aims to achieve. (1) 100% classification accuracy on known benign and malignant patterns in database (2) high generalization accuracy (3) fast and guaranteed training, and (4) fast adaptation of new knowledge.

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.

B. Verma is with Central Queensland University, Australia (phone: +61749309058; fax: +61749309027; e-mail: b.verma@cqu.edu.au).

II. Background

A recent survey by Cheng et al. [7] presents a comparative analysis of various techniques for computer-aided detection and classification of microcalcification patterns in digital mammography. The techniques such as artificial neural networks [13, 15-25], fuzzy logic [8-10], and wavelet transforms [11, 12] are the most commonly used for detection and classification of malignant and benign patterns in digital mammograms. Performance of any pattern classification system strongly depends on the characteristics (features) of the input patterns which effectively discriminate each pattern. Selection and extraction of significant type(s) of features which characterize each pattern uniquely are very important for reliable classification. Features found in the literature to classify malignant and benign patterns are shape features [14, 17, 26], image structure features [13, 15, 16, 18, 19], texture features [17, 20-22], wavelet features [11, 12] and BI-RADS ((breast imaging- reporting and data systems) lesion descriptor (radiologists' interpretation) features [18, 19, 23]. Lo et al. [18] and Zhang et al. [19] have researched the significance of different types of features for the classification of different types of microcalcification patterns in digital mammograms. Selection of the significant individual feature(s) or the combination of features, which can effectively discriminate the microcalcification patterns, from the extracted features is another important issue for researchers. Genetic algorithms [16, 19] and sequential forward/backward selection [24] have been efficiently used for optimal feature(s) subset selection for breast abnormality classification.

Chitre et al. [15] compared the statistical methods and the artificial neural networks for microcalcification patterns classification. They obtained a classification rate of 60%,

0-7803-9490-9/06/$20.00/©2006 IEEE

which was better than the statistical classifiers. A comparative study of a radial basis function (RBF) and a multi layer perceptron (MLP) neural networks for the classification of breast abnormalities using the texture features was performed by Christoyianni et al. [20] and Bovis et al. [21] in their research work. They concluded that MLP obtained 4% higher accuracy than RBF. Yu et al. [24] 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 with very low false positive rate. 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. Their proposed technique could classify 88.9% of the 40 cases from the Nijmegen database. Zhang et al. [19] used a genetic algorithm for neural network learning in their study of microcalcification classification in digital mammograms. Two types of features; grey level based statistical features and radiologists' interpretation features including patient age are extracted from the USF-DDSM database to test the proposed technique. With selective features they have attained 90.5% accuracy rate for the calcification type breast abnormalities and 87.2% for the mass type breast abnormalities. Wroblewska et al. [17] evaluated their proposed automated detection and classification technique of microcalcifications in digital mammograms with expert mammographers. They reported that their proposed technique is better in the classification of microcalcifications and the same in detection of the microcalcifications compared with experts.

III. PROPOSED APPROACH

The proposed approach consists of 4 parts as follows: (1) acquisition of digital mammograms from a benchmark database, (2) extraction of features, (3) feature selection, (4) neural classifier (proposed architecture and learning algorithm). An overview of the proposed approach is presented in Figure 1 and details are described in the following sections.

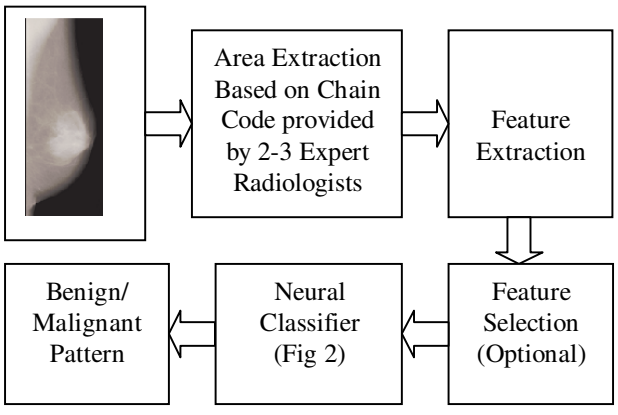

Fig. 1. An overview of the proposed technique.

A.. Benchmark Database of Digital Mammograms

The digital mammograms used in this research are from University of South Florida's Digital Database for Screening Mammography (DDSM). 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 computer aided detection and diagnosis of breast cancer [15]. The database contains approximately 2500 studies of malignant, benign, benign-without-callback and normal cases. Digital mammograms of DDSM are already interpreted by expert radiologists.

The DDSM database includes both calcification and mass types of abnormalities. Mammograms containing mass type of abnormalities have been used in this research work. The experimental dataset contains a total of 200 suspicious areas, 100 (50 malignant, and 50 benign) for training and the same for testing.

B. Feature Extraction

Features discussed below are used in this research for reliable classification of mass type abnormalities in digital mammograms into malignant and benign.

Grey Level based Features

Suspicious areas are already marked in all digital mammograms of the DDSM by 2-3 expert radiologists. The DDSM also provides pixel-level 'ground truth' information about the locations of suspicious areas in mammograms of each case. Grey level based features are calculated using statistical formulas on the grey level pixel values of suspicious areas of digital mammograms. The feature extraction process is accomplished in two steps: extract already marked suspicious areas from mammograms, and extract grey level based features from extracted suspicious areas employing statistical formulas. For area extraction, first the boundary of each suspicious area of the mammogram is defined by solving chain code values available in the '*.OVERLAY*' file of the DDSM case. Using the boundary information each suspicious area is defined on the mammogram. Grey level values for each suspicious area and the respective surrounding boundary area are extracted to calculate the feature values using statistical formulas.

Number of pixels, histogram, average grey, average boundary grey, contrast, difference, energy, modified energy, entropy, modified entropy, standard deviation, modified standard deviation, skew, and modified skew, totaling 14 features are considered as grey level based features.

BI-RADS Features

Each case in the DDSM contains information such as the patient age at time of first study, ACR breast density, a subjective impression of the subtlety of an abnormality, and an abnormality description and abnormality assessment rank that were specified by an expert mammography radiologist using the BI-RADS™ (ACR 1998) lexicon [15].

The morphological descriptors of mass abnormalities are the shape of abnormality and its margin. Hence the total four BI-RADS features are density, mass shape, mass margin, and abnormality assessment rank. Morphological descriptions of breast abnormality are encoded into numeric values to get real feature values.

Patient Age Feature

As mentioned earlier the DDSM provides all information about the case including patient age at the time of the mammogram was taken.

Subtlety Value Feature

Subtlety value is a subjective impression of the subtlety of an abnormality by an expert radiologist. It is rated 1 to 5 by an expert radiologist.

C. Proposed Learning Algorithm

This section presents the proposed neural architecture and learning algorithm for the classification of benign and malignant patterns. An overview of the network architecture is presented in Figure 2. As seen in Figure 2, two additional neurons for benign and malignant patterns with a class weight 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 neural 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 very obvious cases. In situation such as diagnosis of breast cancer, radiologists or doctors are most unlikely use the network which misclassifies obvious cases (true cases) from the database. By introducing two additional neurons which can maximize the outputs of benign and malignant classes for inputs from training patterns, the network will never 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 has also been presented below. The main idea behind this 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 is converted into a linear function by using a log function. A linear system is established for each output class and the output weights are 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 also have a 100% guarantee that there will be a solution. There is no such situation that after many hours we do not have any solution. The use of this idea is making the training very fast, it takes just a few seconds to train the network which solves the problem of slow adaptation of new knowledge from new mammograms. The weights for additional neurons are trained separately. The idea is to maximize the weighting for the input class independent to the rest of the training so that the output neuron for this class is fired (eg. if input test pattern is in benign database then weight for benign class will be higher than the weight for malignant class) and test pattern will be correctly classified. The weights between input and benign additional neuron are initialized with benign training patterns and the weights between input and malignant additional neuron are initialized with malignant training patterns. The following function $W_{cMal}=\log(999999)*(\exp(-\min(\|x-X_{allMalPat}\|)*\log(999999)))$, is used which produces maximum output value for the test input which has exact match in training data and the value decreases according to closeness between test input and training data. The final output from additional neurons is passed to the following function

$$f(out) = \begin{cases} out & out >= \log(999999) \\ out*0.49 & out < \log(9999999) \end{cases}$$

The new learning algorithm consists of training and testing phases as described below. The following steps are used to train the network.

Step 1: Set inputs (#of input features (n), #of hidden units (h), start with h=2, #of outputs (m)-malignant & benign, and #of training pairs (p)).

Step 2: Set weights (W_{ih}) between Inputs (X_n) and Hidden Units (H_h, H_b-benign, H_m-malignant).

Step 2.1 Weights between X_n and H_h

Weights are initialised using small random values.

Step 2.2 Weights between X_n and H_b, H_m (Additional Hidden Neurons)

Weights are initialised with training patterns. Each weight is a vector and the length of vector is equal to p. The weight matrix can be initialised with X ($W_{iben[n][p]} = X$).

Step 3: Calculate weights (W_{ho}, W_{bo}, W_{mo}) between Hidden Units and Outputs

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

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

O_{baf} = log(Target)-log(1-Target);

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

$H_hW_{ho} = O_{baf}$ - Use Least Square Method such as Modified 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_{bo}, W_{mo} to +1 and -1 as shown in Figure 2.

Step 4: Increment #HU & Repeat Steps 2-3 for p/4 times.

Step 5: Select #HU with best results & Repeat Steps 2-3.

The following steps are used to test the network.
Step 1: Feed new input to the network.
Step 2: Calculate the output of the hidden layer.
The output (except benign and malignant neurons) is calculated as follows:
net= sum(X.W)
output=f(net)=1/1+exp(-net);
The output for benign and malignant neurons is calculated as follows:
net = log(999999)*(exp(-min(||x-$X_{allMalPat}$||)*log(999999))),
f(net) = net net >= log(999999)
net*0.49 net < log(9999999)
Step 3: Calculate the output of the network.
The output of the network is calculated as follows.
net= sum(H.W)
output=f(net)=1/1+exp(-net);

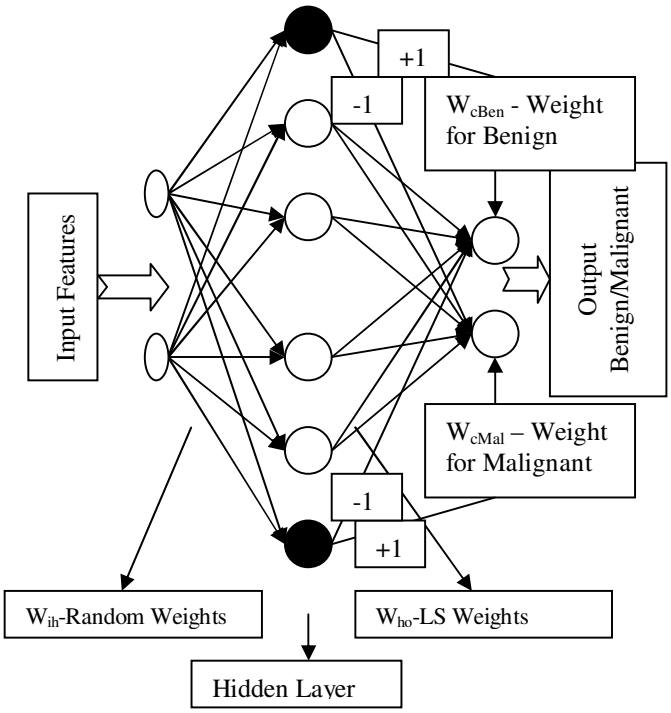

Fig 2. A neural classifier for the proposed technique.

IV. EXPERIMENTAL RESULTS

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) have been implemented and the experiments were conducted using the same database and computing environment. Table I 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 II presents results with grey level features and all 6 BI-RADS features. Table III presents the results with grey level features and only 4 BI-RADS features.

TABLE I
RESULTS USING 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 II
RESULTS USING GREY LEVEL BASED FEATURES, BI-RADS FEATURES, PATIENT AGE FEATURE AND SUBTLETY VALUE FEATURE

	Performance			
	# of Hidden Units	# of Iterations	Classification Rates	
			Training Set	Test Set
Proposed Approach	5	1	100	71
	5	246	100	93
	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
	20	20000	100	84
MLP-GA	18	NA	81	89
DA	NA	NA	90	88
LR	NA	NA	97	89

TABLE III
RESULTS USING GREY LEVEL BASED FEATURES AND BI-RADS FEATURES

	Performance			
	# of Hidden Units	# of Iterations	Classification Rates	
			Training Set	Test Set
Proposed Approach	10	1	100	75
	10	89	100	93
	10	2649	100	94
	12	1	100	91
	12	1808	100	94
	14	1	100	88
	14	1828	100	93
MLP-BP	10	30000	100	90
	10	40000	100	84
	14	10000	100	91
	14	20000	99	91
MLP-GA	18	NA	81	89
DA	NA	NA	90	88
LR	NA	NA	97	89

V. Discussion and Analysis of Results

The results using the proposed approach on three types of feature combinations have been presented in Tables I-III. The best results from experiments have been displayed in Fig. 3. 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 it has a good memorizing capability with an excellent generalization ability. Training in just a few iterations means that the new data can be adapted just in few minutes in comparison to other approaches where we need many hours and days.

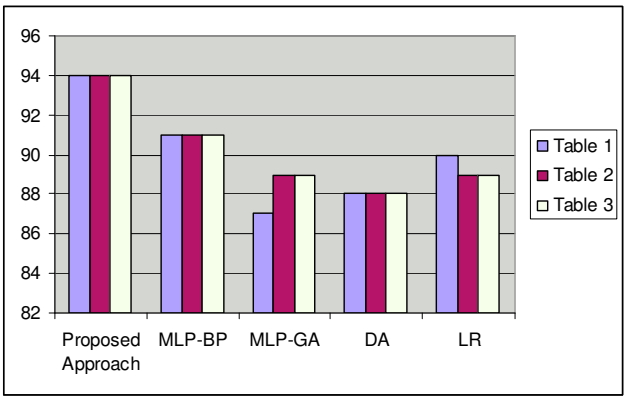

Fig. 3. The best results obtained using the proposed approach and 4 other approaches (comparison purposes)

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 three recent papers to compare the results. Wu et al [25] used neural network ensembles for identifying breast masses. They tested their technique on database taken from the China Society for Industrial and Applied Mathematics. They reported 87.77% classification rate using weighted average (WA) fusion algorithm on balanced input patterns and 88.27% classification rate with perceptron average fusion algorithm on imbalanced input patterns. The highest classification rate obtained in this research (94%) is approximate 5% higher than Wu's results. Zhang et al [26] 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. Again the results in this research are for mass cases so nearly 7% improvement has been achieved. Wroblewska et al [17] have also used the same DDSM benchmark database and reported 76% classification rate which is much lower (18%) than the obtained classification rate in this research. Overall, the proposed technique has outperformed all the existing neural network based and statistical techniques.

VI. Conclusions

We have presented a novel neural architecture and a learning algorithm for the classification of benign and malignant patterns. The experiments using proposed technique on a benchmark database produced 94% classification rate on test set and 100% on training set which are among the best published results. 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 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 other approaches. Fourth and final advantage is that there are not many parameters such as learning rate, momentum, etc. which should be adjusted during the training.

In our future research, we are planning to conduct more experiments and test the consistency and adaptability of the proposed approach on a large benchmark database.

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
3. Canadian Breast Cancer Foundation, http://www.cbcf.org/
4. Cancer Screening Programs-Breast Cancer, http://www.cancerscreening.nhs.uk/ breastscreen/breastcancer.html
5. Australian National Breast Cancer Centre, http://www.nbcc.org.au/.
6. American College of Radiology. www.acr.org
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. K. and Zakos J., "A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques", IEEE Trans. on Info. Tech. in Biomedicine, Vol. 5, pp.46-54, 2001.
9. 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.
10. 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.
11. Yoshida H., Nishikawa R. N., Geiger M.L., and Doi. K., "Signal/Background Separation By Wavelet Packets for Detection of Microcalcifications in Mammograms", Proc. of SPIE on Wavelet Applications in Signal and Image Processing IV, Vol. 2825, pp 805-811, 1996.

12. 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.
13. Woods K. S., Solks J. L., Priebe C. E., Kegelmeyer W. Phillip, Doss C. C., and Bowyer K. W., "Comparative Evaluation of Pattern Recognition Techniques for Detection of Microcalcifications in Mammography" , State of The Art in Digital Mammographic Image Analysis, World Scientific Publication, Vol. 9, pp. 213-231, 1994.
14. Rangayyan R. M., Mudigonda N. R., and Leo Desautels J. E., "Boundary modeling and shape analysis methods for classification of mammographic masses", Medical and Biological Engineering and Computing, Vol. 38, pp. 487-496, 2000.
15. 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, World Scientific Publication, Vol. 9, pp. 167-197, 1994.
16. Zhang, P., Verma, B., and Kumar, K., "Neural vs. Statistical Classifier in Conjunction with Genetic Algorithm Feature Selection in Digital Mammography", Proc. of IEEE-CEC, Vol. 2, pp. 1206-1213, 2003.
17. 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.
18. 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", Proc. of SPIE on Medical Imaging: Image Processing, Vol. 5032, pp. 882-889, 2003.
19. Zhang P., Verma, B., and Kumar K., "A Neural-Genetic Algorithm for Feature Selection and Breast Abnormality Classification in Digital Mammography", Proc. of IEEE-IJCNN, Vol. 3, pp. 2303-2309, 2004.
20. Christoyianni L., Dermatas E., and Kokkinakis G., "Neural Classification of Abnormal Tissues in Digital Mammography using Statistical Features of the Texture", Proc. of IEEE-ICECS, Vol. 1, pp. 117-120, 1999.
21. 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.
22. Bovis K., and Singh S., "Detection of Masses in Mammograms using Texture Measures", Proc. of IEEE-ICPR, Vol. 2, pp. 267-270, 2000.
23. 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.
24. Yu S., and Guan L., "A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films", IEEE Trans. on Medical Imaging, Vol. 19, pp. 115-126, 2000.
25. Wu Y., He J., Man Y., and Arribas J. I., "Neural Network Fusion Strategies for Identifying Breast Masses", Proc. of IEEE-IJCNN, Vol. 3, pp. 2437-2442, 2004.
26. Zhang P., Verma, B., and Kumar K., "A Neural-Genetic Algorithm for Feature Selection and Breast Abnormality Classification in Digital Mammography", Proc. of IEEE-IJCNN, Vol. 3, pp. 2303-2309, 2004.