Breast Cancer Diagnosis using Artificial Neural Network

Kalpana Kaushik
Department of Computer Science & Engineering
Mtech, Scholar, Gateway Institute of Engineering & Technology (GIET), Sonepat

Anil Arora
Department of Computer Science & Engineering
Assistant Professor, Gateway Institute of Engineering & Technology (GIET), Sonepat

Abstract - Artificial neural network has been widely used in various fields as an intelligent tool in recent years, such as artificial intelligence, pattern recognition, medical diagnosis, machine learning and so on. The classification of breast cancer is a medical application that poses a great challenge for researchers and scientists. The help of technology such as data mining and machine learning can substantially improve the diagnosis accuracy. There are parameters to be set in the beginning, long time for training process, and possibility to be trapped in local minima. Databases of breast cancer (Wisconsin Breast Cancer (WBC), Wisconsin Diagnosis Breast Cancer (WDBC) and Wisconsin Prognosis Breast Cancer (WPBC) by using classification accuracy and confusion matrix based on 10-fold cross validation method. Also, we introduce a fusion at classification level between these classifiers to get the most suitable multi-classifier approach for each data set. The experimental results show that in the classification using fusion of MLP and J48 with the PCA is superior to the other classifiers using WBC data set. Results showed that Learning Machine Neural Networks (LM ANN) has better generalization classifier model than BP ANN. However, the standard Gradient-Based Back Propagation Artificial Neural Networks (BP ANN). The development of this technique is promising as intelligent component in medical decision support systems. Recently, the neural network has become a popular tool in the classification of cancer datasets. Classification is one of the most active research and application areas of neural networks.

Keywords—breast cancer; artificial neural networks; learning machine; gradient-based back propagation; medical decision support systems.

I. INTRODUCTION

The main cause of breast cancer is when a single cell or group of cells escapes from the usual controls, that regulate cellular growth and begins to multiply and spread. This activity may result in a mass, tumor or neoplasm. Many masses are benign that means the abnormal growth is mainly restricted to a circumscribed, single and expanding mass of cells (Gokhale 2009). Some tumors are malignant that means the abnormal growth invades the surrounding tissues and that may metastasize or spread to remote areas of the body (Rangayyan 2004). The benign masses may leads to complications where as Malignant tumors are serious cancer. The majority of breast tumors will have metastasized before reaching a tangible size. So far, a various numbers of imaging techniques are discovered for breast cancer detection in tissue level [9].

ANN is one of the best artificial intelligence techniques for common data mining tasks, such classification and regression problems. A lot of research showed that ANN delivered good accuracy in breast cancer diagnosis. However, this method has several limitations. First, ANN has some parameters to be tuned in the beginning of training process such as number of hidden layer and hidden nodes, learning rates, and activation function. Second, it takes long time for training process due to complex architecture and parameters update process in each iteration that need expensive computational cost. Third, it can be trapped to local minima so that the optimal performance cannot be guaranteed. Numerous efforts had been attempted to get the solutions of neural networks limitations. Huang and Babri [10] proved that Single Hidden Layer Neural Networks (SFLN) with tree steps machine learning process that could solve those problems by conjugant gradient algorithm.

The most important of these classifications are binary classification, either benign or malignant. If the cancer is in benign stage, less invasive and risk of treatments is used than for malignant stage (Rangayyan et al. 1997). The reason being the chances of survival of patient is high; it is not beneficial to increase the speed of recovery at the risk of introducing potentially life-threatening side effects caused by aggressive treatment. On the other hand a patient with malignant cancer is not so concerned about the kind of treatment or side effect of the treatment [9].

II. LITERATURE REVIEW

During the past few years, various contributions have been made in literature regarding the application of pattern recognition techniques for breast cancer diagnosis in tissue level.

Rejani and his group proposed a pattern recognition procedure to classify the breast tumor (Rejani and Selvi 2009). They used the image segmentation to segment the breast tissue corresponding to the tumor and used the discrete wavelet transform (DWT) as a feature extraction method to extract various features from the segmented images. Then they also used SVM classifier to classify the breast tissue corresponding to the features and achieved an accuracy of 88.75%.

Martin and his group proposed the technique for detection of mass on digitized mammograms (Martins et al. 2009). They used K-means clustering algorithm for image segmentation and gray level co-occurrence matrix to describe and analyze the texture of segmented structures in the image. The classification of these structures was achieved through Support Vector Machines, which separate them into two groups; using shape and texture descriptors: masses and non-masses. The classification accuracy obtained from that method was 85%.

The uses of classification systems in medical diagnosis, including breast cancer diagnosis, are growing rapidly. Evaluation and decision making process from expert medical diagnosis is key important factor. However, intelligent classification algorithm may help doctor especially in minimizing error from inexperienced practitioners [12].

Several techniques have been deployed to predict and recognize meaningful pattern for breast cancer diagnosis. Ryua [13] developed data classification method, called isotonic separation. The performances were compared against support vector machines, learning vector quantization, decision tree induction, and other methods based on two-breast cancer data set, sufficient and insufficient data. The experiment results demonstrated that isotonic separation was a practical tool for classification in the medical domain.

Sahan [14] applied hybrid machine learning method in diagnosing breast cancer. The method hybridized a fuzzy- artificial immune system with k-nearest neighbor algorithm. The hybrid method delivered good accuracy in Wisconsin Breast Cancer Dataset (WBCD). They believe it can also be tested in other breast cancer diagnosis problems. Ubeyli [15] provided comprehensive view of automated diagnostic systems implementation for breast cancer detection. It compared the performances of multilayer perceptron neural network (MLPNN), combined neural network (CNN), probabilistic neural network (PNN), recurrent neural network (RNN) and support vector machine (SVM). The aim of that works was to be a guide for a reader who wants to develop this kind of systems.

Numerous combinations and hybrid systems used neural networks as a component. However, since almost all of the employed neural networks are conventional gradient descent BP ANN, the novel or hybrid method still suffered the neural networks drawbacks.

III. ARTIFICIAL NEURAL NETWORK

Artificial neural networks (ANN) [5] have been developed as generalizations of mathematical models of biological nervous systems. A first wave of interest in neural networks (also known as connectionist models or parallel distributed processing) emerged after the introduction of simplified neurons by McCulloch and Pitts (1943). The basic components of neural networks are called artificial neurons, or simply neurons or nodes. In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function.

The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm.

A typical artificial neuron and the modeling of a multilayered neural network the signal flow from inputs x1. . .

xn is considered to be unidirectional, which are indicated by arrows, as is a neuron‟s output signal flow (O). The

neuron output signal O is given by the following relationship:

$$O = f(net) = f\left(\sum_{j=1}^{n} w_j x_j\right) \quad (1)$$

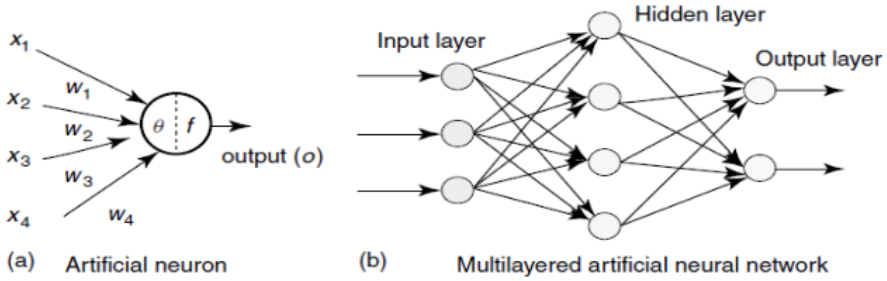

Fig 1: Architrcture of an artificial neuron and a multiple neural network

where wj is the weight vector, and the function f(net) is referred to as an activation (transfer) function. The variable net is defined as a scalar product of the weight and input vectors,

$$net = w^T x = w_1 x_1 + \cdots + w_n x_n \quad (2)$$

where T is the transpose of a matrix, and, in the simplest case, the output value O is computed as

$$O = f(net) = \begin{cases} 1 \text{ if } w^T x \geq \theta \\ 0 \text{ otherwise} \end{cases} \quad (3)$$

where θ is called the threshold level; and this type of node is called a linear threshold unit.

IV. CONJUGATE GRADIENT ALGORITHM

Back propagation (BP) process can train multilayer feed forward neural network (FFNN). With differentiable transfer functions, to perform a function approximation to continuous function f RN, pattern ass pattern classification. The term of back propagation to the process by which derivatives of network error with respect to network weights and biases, can be computed. This process can be used with a number of different optimization strategies.

V. EXPERIMENTS, RESULT & DISCUSSION

When a patient has a palpable breast lump, the first step a doctor takes is to determine whether the mass is malignant or benign. One relatively simple diagnostic procedure is a form of biopsy called fine needle aspiration (FNA). Though these tests are less invasive than others, they are historically less accurate as well. For this project, we need to create a neural network model using data published to the Machine Learning Repository by the University of Wisconsin. A neural network attempts to replicate the brain as a form of artificial intelligence through networks of computers and can be used to detect extremely complex patterns. It learns from its mistakes, so it can classify a case it hasn't seen before as malignant or cancerous based on specific criteria like clump thickness or bland chromatin.

A. Experiments

In this project we want to classify a tumor as benign or malignant, based on uniformity of cell size, clump thickness, mitosis, etc. We have 699 example cases for which we have 9 items of data and the correct classification as benign or malignant. The experiments consisted of three main steps, which are data gathering, data preprocessing, and performance evaluating. The dataset used in this experiment was Breast Cancer Wisconsin Dataset obtained from the University of Wisconsin Hospital, Madison from Dr. William H. Wolberg[25]. The data has 699 instances with 10 attributes plus the class attributes. The class distribution are 65.5% (458 instances) for benign and 34.5% (241 instances) for malignant. Each dataset consists of some classification patterns or instances with a set of numerical features or attributes. An important step in breast cancer diagnosis model is feature extraction. The optimum feature set should have effective and discriminating features, while mostly reduce the redundancy of features space to avoid "curse of dimensionality" problem. The "curse of dimensionality" suggests that the sampling density of the training data is too low to promise a meaningful estimation of a high dimensional classification function with the available finite number of training data. The attribute information can be seen in TABLE I.

TABLE I. ATTRIBUTE INFORMATION

#	**Attributes**	**Domain**
1	Sample code number	id number
2	Clump Thickness	1 - 10
3	Uniformity of Cell Size	1 - 10
4	Uniformity of Cell Shape	1 - 10
5	Marginal Adhesion	1 - 10
6	Single Epithelial Cell Size	1 - 10
7	Bare Nuclei	1 - 10
8	Bland Chromatin	1 - 10
9	Normal Nucleoli	1 - 10
10	Mitoses	1 - 10
11	Class	2 for benign, 4 for malignant

In the clump thickness, benign cells tend to be grouped in monolayers, while cancerous cells are often grouped in multilayer. While in the uniformity of cell size/shape, the cancer cells tend to vary in size and shape. That is why these parameters are valuable in determining whether the cells are cancerous or not. In the case of marginal adhesion, the normal cells tend to stick together, where cancer cells tend to lose this ability. So loss of adhesion is a sign of malignancy. In the single epithelial cell size, the size is related to the uniformity mentioned above. Epithelial cells that are significantly enlarged may be a malignant cell. The bare nuclei are a term used for nuclei that is not surrounded by cytoplasm (the rest of the cell). Those are typically seen in benign tumors. The bland chromatin describes a uniform "texture" of the nucleus seen in benign cells. In cancer cells the chromatin tends to be coarser. The normal nucleoli are small structures seen in the nucleus. In normal cells the nucleolus is usually very small if visible. In cancer cells the nucleoli become more prominent, and sometimes there are more of them. Finally, mitoses are nuclear division plus cytokines and produce two identical daughter cells during prophase. It is the process in which the cell divides and replicates. Pathologists can determine the grade of cancer by counting the number of mitoses.

In the second step, the raw dataset was preprocessed to produce well-formed data that suitable for training and testing process. The first attribute, sample code number, was removed because it was not relevant to the diagnosis. The next nine attributes were normalized into [-1, 1] and used as predictor. The last attribute was transformed to 0 (benign) and 1 (malignant) such that it can be properly fitted to the standard back propagation artificial neural network algorithm.

B. Training and Classification

The network is trained with multilayer perceptron (MLP), a feed-forward back-propagation network, most frequently used neural network technique in pattern recognition [26] [27]. Briefly, MLPs are supervised learning classifiers that consist of an input layer, an output layer, and one or more hidden layers that extract useful information during learning and assign modifiable weighting coefficients to components of the input layers. In the first (forward) pass, weights assigned to the input units and the nodes in the hidden layers and between the

nodes in the hidden layer and the output, determine the output. The output is compared with the target output. An error signal is then back propagated and the connection weights are adjusted correspondingly. During training, MLPs construct a multidimensional space, defined by the activation of the hidden nodes, so that the three classes (malignant, benign and normal tissue) are as separable as possible. The separating surface adapts to the data. Here, conjugate gradient back propagation algorithm has been used for training.
The basic back-propagation algorithm adjusts the weights in steepest descent direction (the most negative of the gradients). This is the direction in which the performance function is decreasing most rapidly. It turns out that, although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. Whereas, in the conjugate gradient algorithms a search is performed along conjugate directions, which produces generally faster convergence than steepest directions. In most of the conjugate gradient algorithms the step-size is adjusted at each iteration. A search made along the conjugate gradient direction to determine the step-size, which will minimize the performance function along that line [28].

C. Result And Discussion

Cancer Inputs is a 9x699 matrix defining nine attributes of 699 biopsies. Cancer Targets is a 2x699 matrix where each column indicates a correct category with a one in either element 1 (benign) or element 2 (malignant).With these settings, the input vectors and target vectors will be randomly divided into three sets as follows: 65% are used for training. 20% are used to validate that the network is generalizing and to stop training before over-fitting. The last 15% are used as a completely independent test of network generalization. The standard network that is used for pattern recognition is a two-layer feed-forward network, with a sigmoid transfer function in the hidden layer, and a softmax transfer function in the output layer. The number of hidden neurons is set to be 7. The number of output neurons is set to 2, which is equal to the number of elements in the target vector (the number of categories). The network is then trained. The training continues for 14 iterations.

C.1 Performance curve

Performance is measured in terms of mean squared error, it plots error vs. epoch for the training, validation, and test performances of the training record that is shown in log scale. Generally, the error reduces after more epochs of training, but might start to increase on the validation data set as the network starts over-fitting the training data. In the default setup, the training stops after six consecutive increases in validation error, and the best performance is taken from the epoch with the lowest validation error. Here in this case, the error is measured in cross entropy form. The cross-entropy measure has been used as an alternative to squared error. Cross-entropy can be used as an error measure when a network's outputs can be thought of as representing independent hypotheses (e.g. each node stands for a different concept), and the node activations can be understood as representing the probability (or confidence) that each hypothesis might be true. In that case, the output vector represents a probability distribution, and our error measure - cross-entropy - indicates the distance between what the network believes this distribution should be, and what the teacher says it should be. There is a practical reason to use cross-entropy as well. It may be more useful in problems in which the targets are 0 and 1 (thought the outputs obviously may assume values in between.) Cross-entropy tends to allow errors to change weights even when nodes saturate (which means that their derivatives are asymptotically close to 0.) Performance is shown in Fig3 for each of the training, validation and test sets. The version of the network that did best on the validation set is was after training.

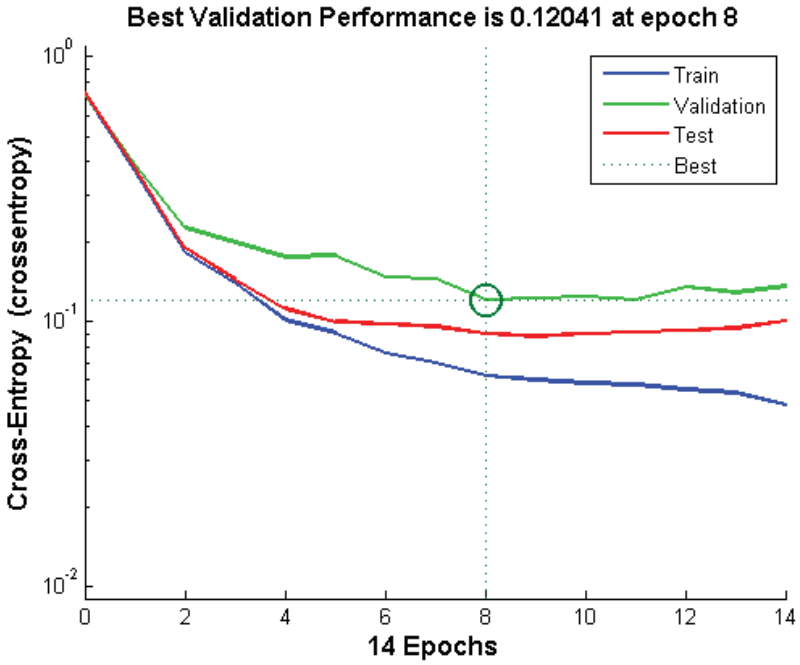

Fig.2: Performance curve for training, testing and validation

C.2 Confusion matrix curve

One measure of how well the neural network has fit the data is the confusion plot. It returns a confusion matrix plot for the target and output data in targets and outputs, respectively. On the confusion matrix plot, the rows correspond to the predicted class (Output Class), and the columns show the true class (Target Class). The diagonal cells show for how many (and what percentage) of the examples the trained network correctly estimates the classes of observations. That is, it shows what percentage of the true and predicted classes match. The off diagonal cells show where the classifier has made mistakes. The column on the far right of the plot shows the accuracy for each predicted class, while the row at the bottom of the plot shows the accuracy for each true class. The cell in the bottom right of the plot shows the overall accuracy. Here the confusion matrix is plotted across all samples as shown in Figure4. The confusion matrix shows the percentages of correct and incorrect classifications. Correct classifications are the green squares on the matrices diagonal. Incorrect classifications form the red squares. If the network has learned to classify properly, the percentages in the red squares should be very small, indicating few misclassifications. The figure below shows the confusion matrices for training, testing, and validation, and the three kinds of data combined.

Fig.3: Confusion matrix

In this figure, the first two diagonal cells show the number and percentage of correct classifications by the trained network. For example 290 biopsies are correctly classified as benign. This corresponds to 63.9% of all 699 biopsies. Similarly, 153 cases are correctly classified as malignant. This corresponds to 33.7% of all biopsies.

4 of the malignant biopsies are incorrectly classified as benign and this corresponds to 0.9% of all 699 biopsies in the data. Similarly, 7 of the benign biopsies are incorrectly classified as malignant and this corresponds to 1.9% of all data.

Out of 294 benign predictions, 98.6% are correct and 1.4% is wrong. Out of 160 malignant predictions, 95.6% are correct and 4.4% are wrong. Out of 297 benign cases, 97.6% are correctly predicted as benign and 2.4% are predicted as malignant. Out of 157 malignant cases, 97.6% are correctly classified as malignant and 2.4% are classified as benign.

Overall, 97.6% of the predictions are correct and 2.4% are wrong classifications. The network outputs are very accurate, as you can see by the high numbers of correct responses in the green squares and the low numbers of incorrect responses in the red squares. The lower right blue squares illustrate the overall accuracies.

D. Conclusion

The experimental results in WBC dataset show that the MLP with conjugate gradient algorithm has been able to classify the cancer data. This work proposes a method for refinement and categorizing the ovarian with kind, using artificial neural network technique. It gives example of detection of cancer from mass spectrometry data on protein profiles. From the results it can be seen that ovarian cancer can be diagnosed using artificial neural network.

Future Scope

For medical checking and repots will take a lots of time of patients as well doctor's time too. So, it is time consuming process proceeding by pathology specified doctor and assistants, In thesis report this time spending process is done by just few mints and report will generate in short period of time then patients treatment will do further and faster.

This software will use in pathology laboratories in short time will detect cancer in which stage after this doctor will suggest for further treatments. By more research on this topic you should improve in these field and take help by these medical hospital and institutes:-

- Use free tool- simbrain, matlab GUI,
- Use image screening software,
- You should have strong bonding with hospital medical officers, HOD, Dean, Doctors, HR departments also.
- Use basic languages with these member otherwise they are very cleaver and you just lost which important point you want to discuss with them.
- Hospitals:- BL Kapoor, Action Balaji Cancer Hospital, Maulana Azad Medical Collage and Hospital, Rajive Gandhi Cancer Hospital.

For a being future scope, we need to change the algorithm by changing its tool or software. As doing some more research work on this topic must change algorithm, for some change need "simbrain" is free tool software.

REFERENCES

[1] World Health Report 2008. Lyon: International Agency for Research on Cancer, 2008.
[2] Canadian Cancer Society NCI. Canadian Cancer Statistics. 2009.
[3] Bray F, McCarron P, Parkin DM. The changing global patterns of female breast cancer incidence and mortality. Breast Cancer Res 2004;6:229-39.
[4] Ziegler RG, Hoover RN, Pike MC et al. Migration patterns and breast cancer risk in Asian-American women. J Natl Cancer Inst 1993;85:1819-27.
[5] Stanford JL, Herrinton LJ, Schwartz SM, Weiss NS. Breast cancer incidence in Asian migrants to the United States and their descendants. Epidemiology 1995;6:181-3.
[6] Pike MC, Kolonel LN, Henderson BE et al. Breast cancer in a multiethnic cohort in Hawaii and Los Angeles: risk factor-adjusted incidence in Japanese equals and in Hawaiians exceeds that in Whites. Cancer Epidemiol Biomarkers Prev 2002;11:795- 800.
[7] Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. Int J Cancer 1975;15:617-31.
[8] Freedman LS, Clifford C, Messina M. Analysis of dietary fat, calories, body weight, and the development of mammary tumors in rats and mice: a review. Cancer Res 1990;50:5710-9.
[9] Rajesh Kumar Tripathy, An investigation of the breast cancer classification using various machine-learning techniques, department of biotechnology & medical engineering national institute of technology rourkela-769008, Orissa, India 2013
[10] G. B. Huang and H. A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions.," IEEE Trans. Neural Netw., vol. 9, no. 1, pp. 224– 9, Jan. 1998.
[11] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in Proceedings of the National Academy of Sciences, U.S.A., Volume 87, December 1990, pp. 9193–9196.
[12] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Syst. Appl., vol. 36, no. 2, pp. 3240– 3247, Mar. 2009.
[13] U. Ryu, R. Chandrasekaran, and V. S. Jacob, "Breast cancer prediction using the isotonic separation technique," Eur. J. Oper. Res., vol. 181, no. 2, pp. 842–854, Sep. 2007.
[14] S. Sahan, K. Polat, H. Kodaz, and S. Güneş, "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis.," Comput. Biol. Med., vol. 37, no. 3, pp. 415–23, Mar. 2007.
[15] E. D. Übeyli, "Implementing automated diagnostic systems for breast cancer detection," Expert Syst. Appl., vol. 33, no. 4, pp. 1054–1062, Nov. 2007.
[16] G.-B. Huang, "Real -Time Learning Capability of Neural Networks ", IEEE Transactions on Neural Networks, VOL. 17, NO. 4, JULY 2006 .
[17] W. W. Hsieh, "Machine Learning Methods in Environmental Sciences Neural Networks and Kernels", Cambridge University Press, August 31, 2008.
[18] G. Weir - Smith and C. A. Schwabe, " Spatial interpolation vs neural network propagation as a method of extrapolating from field surveys ", GIS Centre,HSRC (Human Sciences Research Council), Pretoria, 2002.
[19] J.M. Turmon ,"Assessing Generalization of Feed forward Neural Networks", phD. thesis, University of Cornell, August 1995.
[20] M.A.Ali, S.D.Gore and M. AL-Sarierah , " The Use of Neural Network to Recognize the parts of the Computer Motherboard ", Journal of Computer Sciences 1(4), pp. 477- 481, 2005.
[21] J.ILONEN, J.-K.KAMARAINEN and J.LAMPINEN, "Differential Evolution Training Algorithm for Feed-Forward Neural Networks", Neural Processing Letters 17: pp. 93 – 105, 2003.
[22] S. Breutel, "Analysing the Behaviour of Neural Networks", PhD thesis, Queensland University of Technology, Brisbane, 2004.
[23] Conjugate Gradient method For Training Feed Forward Neural Networks Luma. N. M.Tawfiq & Alaa. K. J. AL-Mosawi Department of Mathematics, College of Education-Ibn Al-Haitham, University of Baghdad .
[24] Machine Learning A Probabilistic Perspective Kevin P. Murphy The MIT Press Cambridge, Massachusetts London, England
[25] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in Proceedings of the National Academy of Sciences, U.S.A., Volume 87, December 1990, pp. 9193–9196.
[26] Duda, R.O., Hart, P.E.: "Pattern Classification and Scene Analysis", In: Wiley-Interscience Publication, New York (1973)
[27] Bishop, C.M.: "Neural Networks for Pattern Recognition". Oxford University Press,New York (1999).
[28] Adeli H., Hung S.L., "Machine learning neural networks, genetic algorithms, and fuzzy systems (John Wiley & Soons Inc., New York, NY,USA), 1995.
[29] http://www.cdc.gov/cancer/breast/basic_info/prevention.html.