Breast Cancer Malignancy Identification using Self-Organizing Map

SUDHIR D. SAWARKAR
Department of Computer Engineering
Datta Meghe College of Engineering, Airoli. Mumbai University, Mumbai
(INDIA) Pin Code- 400708
DR. ASHOK A. GHATOL
Vice Chancellor, Dr. Babasaheb Ambedkar Technological University, Lonere
Dist: Raigad, M.S.
(INDIA) Pin Code- 402103

Abstract: - *An artificial neural network (ANN) is an information-processing paradigm inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information. The key element of the ANN paradigm is the novel structure of the information processing system. Learning in ANN typically occurs by example through training, or exposure to a set of input/output data where the training algorithm iteratively adjusts the connection weights. The Kohonen self-organizing map neural network performs a mapping from a continuous input space to a discrete output space, preserving the topological properties of the input. For training and testing the neural network various databases available on the Internet as well as gathered information from hospitals is used.*

Key-Words: breast cancer, neural network, efficiency, sensitivity, self-organizing map

1. Introduction

The International Agency for Research on Cancer has reported breast cancer to be by far the most frequent cancer and the leading cause of death in women. It has been estimated that 719,100 new cases (19% of all new cancers in females) occurred worldwide in 1985. With annual totals expected to be around 340,000 deaths in 1990 and 420,000 by 2000. [1]

Currently, breast imaging for the detection and characterization of suspicious breast lesions relies upon mammography and ultrasound. Mammography is the modality of choice for early diagnosis of breast cancer. Only 20% of currently biopsy cases actually reveal cancer. The remainders are all benign cases, which underwent a potentially unnecessary surgical procedure.

Preventing benign biopsies is the most important way to improve the efficacy of mammography screening, especially as screening becomes more widespread. This clearly demonstrates a need for efficient breast cancer diagnosis techniques. Some of the works done in this direction include linear programming approach (Mangasarian, 1995), machine-learning approach (Wolberg, 1994) [2,3,4]

It is proposed to develop an efficient neural network models to provide accurate diagnosis while being completely noninvasive. These models will utilize existing, available information such as mammography findings and patient history data. Artificial Neural Network (ANN) is collections of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. The key element of ANN paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to synapses. [2,3,4]

Computer based diagnostic system hold promising means the challenges of the clinical situations.

Artificial neural network (ANN) model has been developed to diagnose heart disease [5], Crutzfelt Jackob diseases [6], Acute myocardial infarction disease [7], Coronary artery disease [8], Low back pain disease [9], Dermatology disease [10], Thyroid disease [11] & Acute coronary occlusion disease [12], with encouraging results.

2. Data

This work grew out of the desire by Dr. Wolberg [2,3,4] to accurately diagnose breast masses based solely on a Fine Needle Aspiration (FNA). The feature extraction process is performed as follows:

An FNA is taken from the breast mass. This material is then mounted on a microscope slide and stained to highlight the cellular nuclei. A portion of the slide is then scanned using a digital camera and a frame-grabber board and identified nine visually assessed characteristics of an FNA sample, which he considered relevant to diagnosis. The resulting data set is well-known as the Wisconsin Breast Cancer Data.(Total attributes 9.Number of instances- 699; Missing attributes- 16;Benign- 458; Malignant-241.)

The reported sensitivity (i.e., ability to correctly diagnose cancer when the disease is present) of mammography varies from 68% to 79%, of FNA with visual interpretation from 65% to 98%, and of surgical biopsy close to 100%. Therefore, mammography lacks sensitivity, FNA sensitivity varies widely, and surgical biopsy, although accurate, is invasive, time consuming, and costly. The goal of the diagnostic aspect of this paper is to develop a relatively objective system.

3. Method

The Kohonen self-organizing map neural network performs a mapping from a continuous input space to a discrete output space, preserving the topological properties of the input.[13,14,15] This means that points close to each other in the input space are mapped to the same or neighboring PEs in the output space. The basis of the Kohonen SOM NN is soft competition among the PEs in the output space.

The Kohonen SOM NN is a fully connected, single-layer linear network (Figure 1). The output generally is organized in a one or two-dimensional arrangement of PEs, which are called neighborhoods.The training of the map is started by initializing the neurons in the grid with a small random feature vector or weight m. In every step of the learning process this weight is updated in such a way that it represents a part of the feature space as good as possible. Three steps in the learning procedure can be distinguished: i) Presentation of a randomly selected object from the input feature space ii) Evaluation of the networkiii) Updating the weight vectors of the network The first step is to randomly select an input object, although in practice the procedure is such that all objects are used at least once. This object is then mapped by the network. Each unit in the grid computes the distance of its weight vector m_i to the input vector x using the Euclidean distance.

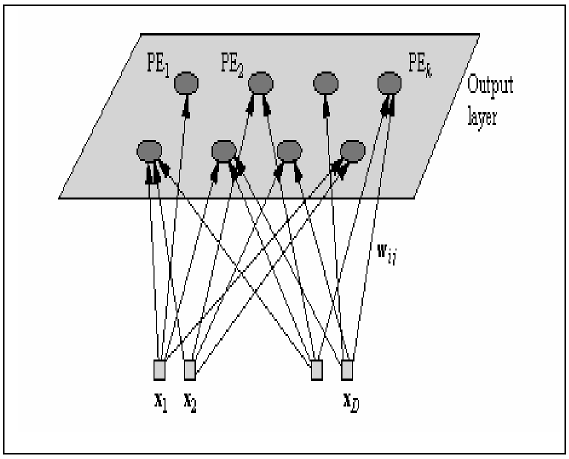

Figure 1 Architecture of the SOM with a 2D output.

The neuron having the least distance is selected as ``winner'' *s*

$$\|\overline{x}(t) - \overline{m}_s(t)\| = \min_i(\|\overline{x}(t) - \overline{m}_i(t)\|) \quad \ldots\ldots (1)$$

The iteration step is indicated with *t*. The winning neuron is used as a starting point for updating all the other neurons in the grid. All neurons within a certain neighbourhood $N_c(t)$ of the winning neuron receive an update towards the presented input object. The update rule is as follows:

$$\overline{m}_i(t+1) = \begin{cases} \overline{m}_i(t) + \alpha(t)[\overline{x}(t) - \overline{m}_i(t) & if \quad i \in N_c(t) \\ \overline{m}_i(t) & if \quad i \notin N_c(t) \end{cases} \quad \ldots.(2)$$

The neighbourhood size $N_c(t)$ decreases monotonically in time, less neurons around the winning neuron need to be updated since the objects are better represented each time an update is performed. The learning parameter $\alpha(t)$ in equation 2 linearly decreases in time to ensure that after a certain number of steps the network stops training. In the original formulation of the network [13,15] a neighbourhood parameter $h_{si}(t)$was used. This $h_{si}(t)$is in contrast to $N_c(t)$ a continuous value modelled by a Gaussian which has a kernel shrinking in time:

$$h_{is} = h_o(t)\exp\left(-\frac{\|\tau_i - \tau_0\|}{\sigma(t)^2}\right)^2 \quad \ldots\ldots\ldots(3)$$

where τ_i and τ_0 are the co-ordinates of the neurons in the grid and $\|\tau_i - \tau_0\|$ is a suitable distance measure between those neurons. For instance, if the co-ordinates would be numbered as in figure 1, then the distance between *i* and *s* would be: $\max(|r_{i1} - \delta_{i1}|, |r_{i2} - \delta_{i2}|)$ The size of the Gaussian function, determined by $\sigma(t)^2$, decreases as the network reaches convergence. The parameter $h_o(t)$has the same role as the learning parameter α in equation 2. The update rule stated in equation 2 changes into;

$$\overline{m}_i(t+1) = h_{oi}(t)[\overline{x}(t) - \overline{m}_i(t)] \quad \ldots\ldots\ldots(4)$$

When the network reaches convergence [16, 17] it can be used for classification. The final values of the neuron weights are used to compute the inter neuron distances. This will enable to visualize the scatter of the data set.

4. Result

Self-organizing Map method was used on the set of 683 samples of actual data. The accuracy or efficiency of the diagnosis of breast cancer by ANN is evaluated by using the Magnitude of Relative Error which is calculated using formula:

MRE = Abs ((AD - DD) /AD)

Where

AD is actual detection, DD is desired detection

Pred (0.25) gives % of input that were predicted with an MRE is less than 0.25.

Measure of Average efficiency is calculated using:

Pred(p) = if (MRE < 0.25,1,0); Pred(p) = K/N

Where N total no of historical data and K is number of cases output with MRE less than or equal to p. [18]

This experiment is aimed at finding the efficiency using self organizing map neural network. For this total 480 experiments are performed by selecting different sets of training and testing samples out of 683 samples of data set.

Training parameters used to train self organizing map neural network are as follows.

Activation functions selected:

TanhAxon, SigmoidAxon, Linear TanhAxon, Linear SigmoidAxon, SoftMaxAxon, BiasAxon, LinearAxon, and Axon

Momentum = 0.7,

Step size: Hidden layer =1.0, output layer =0.1

Number of Hidden layer = 1, 2, 3

Number of neurons in hidden layer = 4

Number of epochs = 5000

The artificial neural network is trained with proper training parameters. Training returns new values of weights and biases. The errors are plotted with respect to epochs. The example with activation function Linear TanhAxon and hidden layer 1 training error versus number of epoch is plotted as shown in figure 2.

For 20 readings with activation function Linear TanhAxon and hidden layer 1, percentage

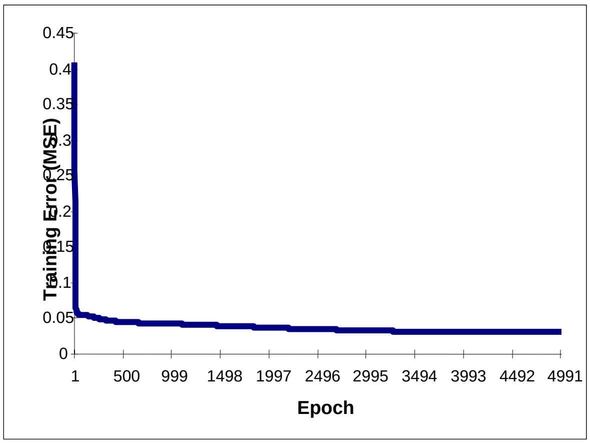

Figure 2 MSE verses number of epoch

efficiency versus number of testing samples is shown in figure 3, for the same testing error versus number of training samples is shown in figure 4.

Table-1 gives 20 experiments results out of 480.

Number of Samples		**Activation Function TanhAxon and hidden layer-1**			
Training	**Testing**	**Efficiency**	**Min. Train. error**	**Final Train. error**	**Testing error**
459	224	98.66	0.01	0.01	0.04
469	214	98.13	0.02	0.02	0.06
479	204	98.53	0.01	0.01	0.04
489	194	98.45	0.01	0.01	0.03
499	184	98.91	0.02	0.02	0.03
509	174	98.85	0.01	0.01	0.04
519	164	98.78	0.02	0.02	0.03
529	154	98.05	0.01	0.01	0.05
539	144	99.31	0.01	0.01	0.01
549	134	98.51	0.02	0.02	0.03
559	124	99.19	0.02	0.02	0.01
569	114	98.25	0.02	0.02	0.03
579	104	99.04	0.01	0.01	0.04
589	94	97.87	0.02	0.02	0.04
599	84	98.81	0.01	0.01	0.01
609	74	98.65	0.02	0.02	0.01
619	64	98.44	0.02	0.02	0.02
629	54	98.15	0.02	0.02	0.05
639	44	97.73	0.01	0.01	0.02
649	34	100	0.01	0.01	0.00
Average		**98.62**	**0.01**	**0.01**	**0.03**

Table-1 Sample reading for 20 experiments with activation unction tanhAxon and hidden layer one.

The average efficiency for 20 experiments each is calculated using different activation functions, varying number of hidden layers (1, 2, 3) and varying number of training and testing samples. The neural network is trained for 5000 epochs. Figure 5 shows keeping hidden layer as 1 for activation function Linear TanhAxon 98.88% average efficiency is obtained which is good as compared to other activation functions.

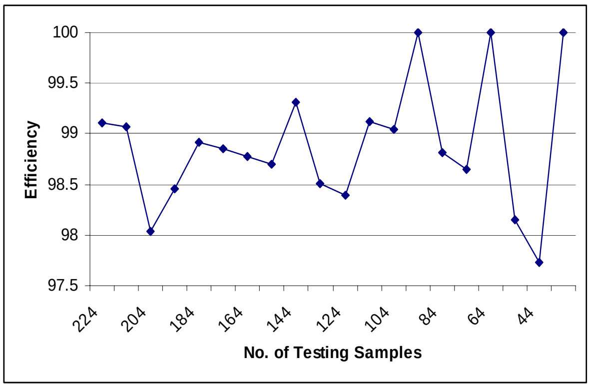

Figure 3 Efficiency verses testing samples

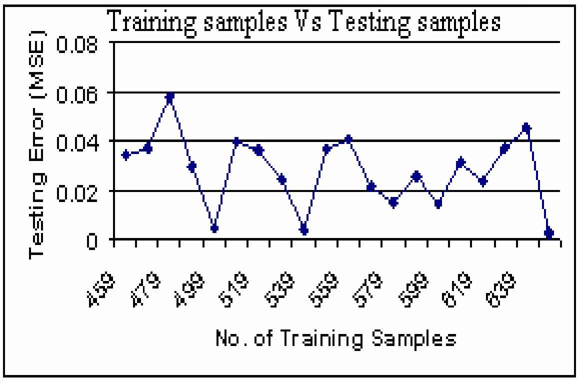

Figure 4 Number of training samples verses

Figure 6 shows overall training error i.e. Minimum MSE is good for TanhAxon

Figure 7 shows overall training error i.e. Final MSE is good for TanhAxon

Figure 8 shows overall testing error (MSE) is good for TanhAxon.

From above, it is observed that the self organizing map neural network is trained for transfer function TanhAxon. It is also observed that for Linear TanhAxon function, the efficiency (in percentage) decreases as number of hidden layer increases and for TanhAxon, the efficiency (in percentage) increases as number of hidden layer increases.

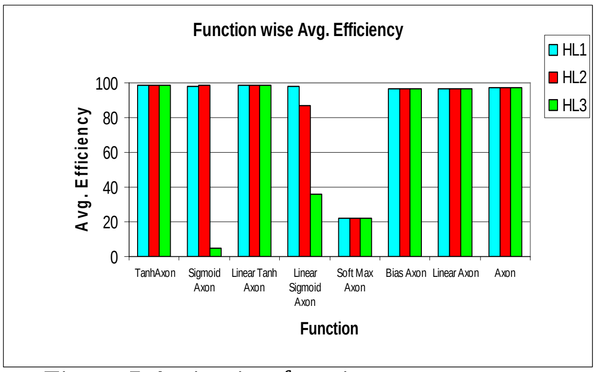

Figure 5 Activation function versus average efficiency in % (SOM NN)

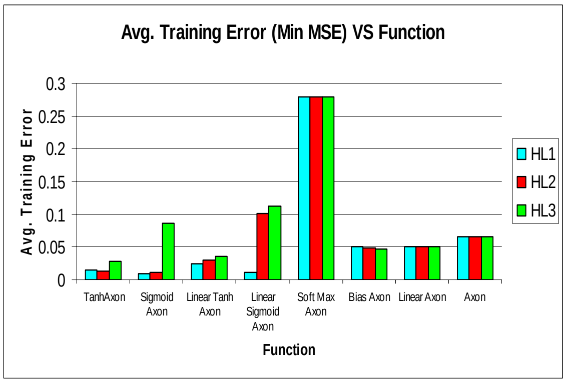

Figure 6 Activation function versus average training error (SOM NN)

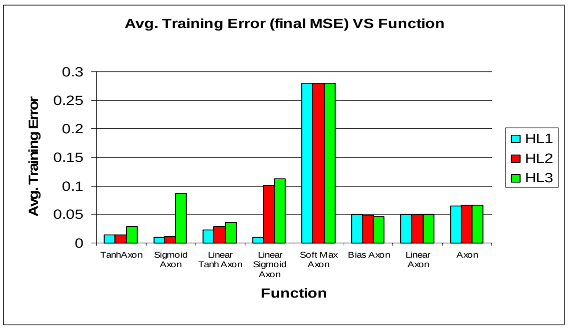

Figure 7 Activation function versus average final training error (SOM NN)

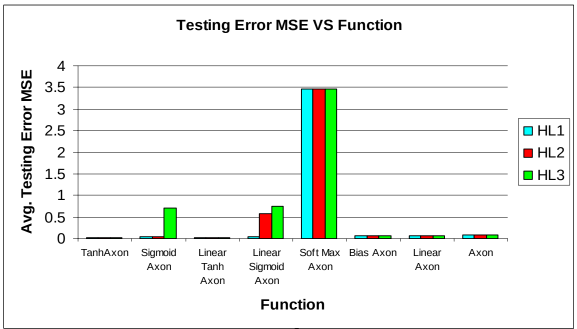

Figure 8 Activation function versus average testing error (SOM NN)

Receiver Operating Characteristic, Area Under Curve, Specificity and Sensitivity:

An ROC curve is a two-dimensional depiction of classifier performance, on which *TP* (*true positive rate*) is plotted on the *Y* axis and *FP* (*false positive rate*) is plotted on the *X* axis.[19] AUC (area under the ROC curve) has been recently used as an alternative measure for machine learning algorithms. There are many advantages of AUC, such as its independency to the decision sensitivity in Analysis of Variance. Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1. [20]

The typical ROC curve for SOM NN is shown in figure 9. From ROC curve of self organizing map neural network the area under curve (AUC) is calculated using trapezoidal rule. The area under curve is found to be 1.Which clearly reveals an ideal classification. This result is obtained for neural network trained with 549 samples and testes for 134 samples of which 35 are malignant and 99 are benignant.

Classification efficiency has been widely used as the main criterion for comparing the classification quality of classifiers.

$$sensitivity = \frac{truePositiveExamples}{totalPositiveExamples}$$ and

specificity measures the ability of a test to be negative when the condition is actually not present, or how many of the negative test examples are excluded:

$$specificity = \frac{trueNegativeExamples}{totalNegativeExamples}$$

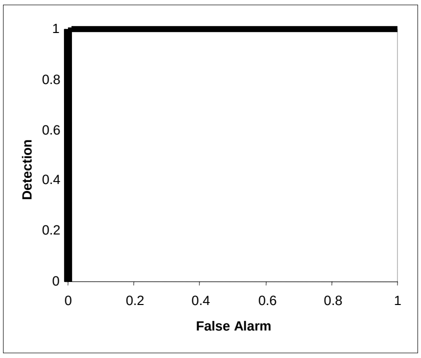

Figure 9 ROC curve for SOM NN model on testing dataset

Table 2 shows the selectivity, sensitivity and efficiency for SVM and SOM for the experiment performed to calculate ROC. The two different parameters used are pred (0.25) and default values of neural network.(same experiment as that of for ROC)

Parameter	With Pred(0.25)		With default value	
	SVM [21]	SOM	SVM [21]	SOM
Sensitivity (in %)	97	100	97	100
Selectivity (in %)	95.95	98.99	98.9	98.99
Efficiency (in %)	96.26	99.25	98.5	99.25

Table 2 Sensitivity, selectivity, efficiency

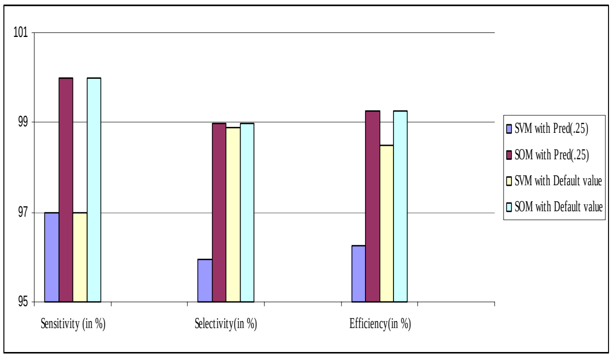

Figure10. Sensitivity / specificity / efficiency using default MSE and Pred.(0.25) Support Vector Machine and Self Organizing Map

4. Conclusion

Neural network aided breast cancer diagnosis gives promising results. It can be a good supplement to the conventional clinical diagnosis system. This study shows the decision taking ability of artificial neural network model.

For diagnosis the efficiency of self organizing map neural network shows that it can support the doctors or physicians to consider it as a second opinion of the learning machine to prevent biopsy. In addition, this neural network based clinical Decision Support System avoid unnecessary excision and expenses.

Using self organizing map ANN, the breast cancer diagnosis is comparably accurate than the human being. The average efficiency of the Self organizing map ANN is nearly 98.88%. The pred (p) analysis gives low efficiency than the default value.

The optimal classifier activation function of neurons for the self organizing map neural network is seen to be lineartanh Axon. With this activation function the estimated classifiers are able to discriminate between the malignant and the benign patients. The proposed classifiers are seen to be a good supplementary aid for the physicians to make precise clinical decisions pertaining to diagnosis.

5. **Future Scope**

It may be possible to improve efficiency of module that has currently been implemented. Some of the areas in which future work can be done to make the present scheme more effective are:

1. The neural network can be realized in hardware with the programmable weights and biases so that the real time capability of the scheme can be effectively analyzed.
2. The network has to be trained with more input patterns, so that the generalization ability of network will be enhanced.
3. The present study uses Euclidean norm (L_2 norms) to compute the MSE between the output of the neural network model and the desired one. Possibility of other L_p norms (p>2) for the calculation of the MSE may be investigated to suggest an optimal cost function.

References:

[1] 2003: Imag(NBCC) National Breast Cancer www.nbcc.org.au/pages/info/resource/nbccpubs-/bc21-94/intro.htm

[2] O.L. Mangasarian, W.N. Street and W.H. Wolberg. "*Breast cancer diagnosis and prognosis via linear programming*". Operations Research, 43(4), pages 570-577, July-August 1995.

[3] W.H. Wolberg, W.N. Street, and O.L. Mangasarian. "Image analysis and machine learning applied to breast cancer diagnosis and prognosis. Analytical and Quantitative Cytology and Histology, Vol. 17 No. 2, pages 77-87, April 1995.

[4] W.H. Wolberg, W.N. Street, and O.L. Mangasarian. Machine learning techniques to diagnose breast cancer from fine-needle aspirates. Cancer Letters 77 (1994) 163-171.

[5]N. Kumaravel, K. S. Sridhar, N. Nithiyanandan. '*Automatic diagnosis of Heart diseases using neural network.*' IEEE, Southern Biomedical Engineering conference, Piscataway, USA, pp 319-322,1996.

[6]K. Sutherland, R. Desilva, R. G. Will. '*Clinical diagnosis of Crentzfeldt – Jakob disease using multi layer perceptron neural network classifier*.' Journal of Intelligent systems, Vol. 7, No. 2, U.K.,pp 1-18,1997.

[7]R. L. Kennedy, R. F. Harrison, A. M. Burton. '*Artificial neural network system for the diagnosis of acute myocardial infraction (AMI) in the accident & emergency department*.' Computer Methods and Programs in Biomedicine, Vol. 52, No.2, Studerland, U. K., pp 93-103,1997

[8]Tang K. Wendy, Pingle. '*Artificial neural network for the diagnosis of Coronary Artery disease* ' Journal of Intelligent system, Vol – 07, No, USA, pp. 307-338, 1997.

[9] D. L. Bounds, P. J. Lioyd, B. Mathew & G. G. Wadell.' *A multilayer perceptron network for the diagnosis of low back pain.*' Proceeding of the IEEE International conference on neural networks, san diego, Vol. II, IEEE, New York, pp. 481-489,1988.

[10] Y. O. Yoon, R. W. Brobst, P. R. Bergstresser & L. L. Peterson. ' *A desktop neural network for dermatology diagnosis.*' Journal of neural network computing, summer, pp. 43-52,1989.

[11] P. N. Chatur & A. A. Ghatol, ' *Artificial neural network for the diagnosis of Thyroid disease,*' International Conference of IEEE on Evolutionary Computing for computer, Communication, Control and Power, ECCAP - 2000, Chenni. January 4th – 7th, pp. 21-26, 2000.

[12] W. G. Baxt. '*Use of artificial neural network for data analysis in clinical decision-making. The diagnosis of acute coronary occlusion.*' Neural computation, pp. 480-489, April 1990.

[13]T. Kohonen, Self Organisation and Associative Memory, Third Edition, Springer-Verlag, New York, 1989.

[14]Kohonen, T. The self-organizing map. In Proceedings IEEE, pages 1464-1480, September 1990. vol 78, no 9.

[15]Ritter, H., Schulten, K., and Martinetz, T. Neural computation and self-organizing maps: an introduction. Addison-Wesley, 1991.

[16]Barnas, A. and La Vigna, A. Convergence of kohonen's learning vector quantization. In IJCNN, International Joint Conference on Neural Networks,

[17]Ritter, H. and Schulten, K. Kohonen's self-organizing maps: Exploring their computational capabilities. In IEEE International Conference on Neural Networks. IEEE, 1988. vol 1, section : Self-organization.

[18]Ali Idri, Taghi M Khoshgoftaar, "*Can Neural Network be easily interpreated in software cost estimation*", 2002, world congress computational intelligence, Honolulu, Hawaii

[19]Fawcett, T., "ROC Graphs: Note and Practical Considerations for Data Mining Researchers", HP Labs Technical Report (HPL-2003-4), 2003.

[20]Hanley, J. A., McNeil, B. J., "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve", Radiology, Vol.143, pp.29-36, 1982.

[21]Sudhir D. Sawarkar, Ashok A. Ghatol "Breast cancer malignancy identification Using Support Vector Machine" International Journal, WSEAS transaction on computers, Issue 8, Volume 5,Augest 2006, pp1707-1712 (ISSN-1109-2750)