Proceedings of the 7th WSEAS International Conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006 (pp158-163)

# Neural Network Aided Breast Cancer Detection and Diagnosis Using Support Vector Machine

SUDHIR D. SAWARKAR
Department of Computer Engineering
Datta Meghe College of Engineering, Airoli. Mumbai University, Mumbai
(INDIA) Pin:400708

ASHOK A. GHATOL
Vice Chancellor, Dr. Baba saheb Technological University, Lonere
Dist: Raigad, M.S.
*(INDIA) Pin : 402103*

AMOL P. PANDE
Department of Computer Engineering
Datta Meghe College of Engineering, Airoli. Mumbai University, Mumbai
(INDIA)Pin:400708

***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 (synapses). These connection weights store the knowledge necessary to solve specific problems. In this work, we have used neural networks Support Vector Machine method for diagnosis of breast cancer. SVMs can only be used for classification, not for function approximation. The Support Vector Machine (SVM) is implemented using the kernel Adatron algorithm. The kernel Adatron maps inputs to a high-dimensional feature space, and then optimally separates data into their respective classes by isolating those inputs, which fall close to the data boundaries.The proposed neural network model hold promise for radiologists, surgeons, and patients with information, which was previously available only through biopsy, thus substantially reducing the number of unnecessary surgical procedures. For training and testing the neural network various databases available on the Internet as well as gathered information from hospitals is used.*

*Key-Words: Neural network, support vector machine, efficiency, breast cancer, diagnosis*

## 1. Introduction

Breast cancer is the most commonly diagnosed cancer and the most common cause of death in women all over the world. In 1992, deaths from breast cancer accounted for around 27,000 woman before 75 years of age. The International Agency for Research on Cancer has reported breast cancer to be by far the most frequent cancer, apart from non-melanocytic skin 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. In developed countries such as the USA and other Western countries, it is the most commonly diagnosed cancer in women (excluding non-melanocytic skin cancers), causing about 16% of all deaths due to cancer in women. Worldwide, 308,000 women died of breast cancer in 1985, with annual totals expected to be around 340,000 deaths in 1990 and 420,000 by 2000. The most recent statistics by Asian Project Report 2002 depicted in the (fig.1) clearly shows the severity of the problem. [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 detection of breast cancer. Although mammography is very sensitive at finding cancer, it results in many false positives. Only 20% of currently biopsy cases actually reveal cancer. The remainders are all benign cases, which underwent a potentially unnecessary surgical procedure.

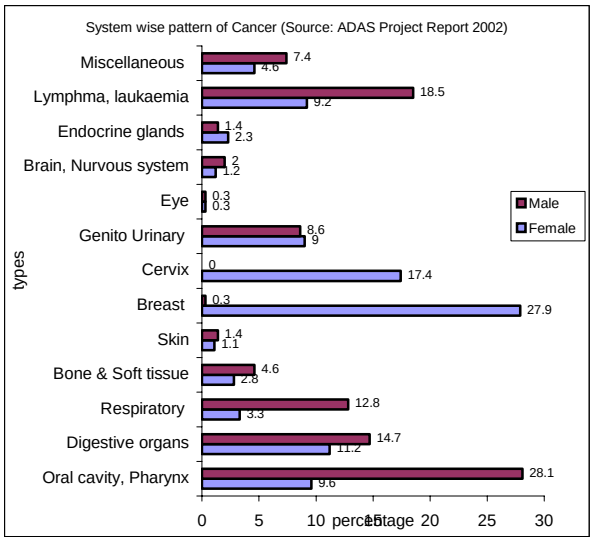


**Figure 1**

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 detection and 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. The application of Artificial Intelligence, Expert system, Knowledge system finds its use to supplement the decision making of the clinician. Artificial neural network captures the basic knowledge that allows the clinician to act as expert while dealing with such complicated problem.

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 in which the cells are well differentiated 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.

| # Attribute | Domain |
|---|---|
| 1. Clump Thickness | 1-10 |
| 2. Uniformity of Cell size | 1-10 |
| 3. Uniformity of Cell shape | 1-10 |
| 4. Marginal Adhesion | 1-10 |
| 5. Single Epithelial Cell Size | 1-10 |
| 6. Bare Nuclei | 1-10 |
| 7. Bland Chromatin | 1-10 |
| 8. Normal Nucleoli | 1-10 |
| 9. Mitoses | 1-10 |
| 10. Class (2 for benign, 4 for malignant) | |
| Number of instances | : 699 |
| Missing attributes | : 16 |
| Benign | : 458 |
| Malignant | : 241 |

Most breast cancers are detected by the patient as a lump in the breast. The majority of breast lumps are benign so it is the physician's responsibility to diagnose breast cancer, that is, to distinguish benign lumps from malignant ones. There are three available methods for diagnosing breast cancer: mammography, FNA with visual interpretation and surgical biopsy. 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 research is to develop a relatively objective system.

## 3. Method

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression.

**Linear classification**When used for classification, the SVM algorithm creates a hyperplane that separates the data into two classes with the maximum-margin. Given training examples labeled either "yes" or "no", a maximum-margin hyperplane is identified which splits the "yes" from the "no" training examples, such that the distance between the hyperplane and the closest examples (the margin) is maximized.

The use of the maximum-margin hyperplane is motivated by Vapnik Chervonenkis theory, which provides a probabilistic test error bound that is minimized when the margin is maximized. However the utility of this theoretical analysis is sometimes questioned given the large slack associated with these bounds: the bounds often predict more than 100% error rates.[ 13]

The parameters of the maximum-margin hyperplane are derived by solving a quadratic programming (QP) optimization problem. There exist several specialized algorithms for quickly solving the QP problem that arises from SVMs. The most common method for solving the QP problem is Platt's SMO algorithm.

Non-linear classification with the "kernel trick"

The original optimal hyperplane algorithm proposed by Vladimir Vapnik in 1963 was a linear classifier. However, in 1992, Bernhard Boser, Isabelle Guyon and Vapnik [13,14] suggested a way to create non-linear classifiers by applying the kernel trick (originally proposed by Aizerman) to maximum-margin hyperplanes. The resulting algorithm is formally similar, except that every dot product is replaced by a non-linear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in the transformed feature space. The transformation may be non-linear and the transformed space high dimensional; thus though the classifier is a hyperplane in the high-dimensional feature space it may be non-linear in the original input space.

If the kernel used is a radial basis function, the corresponding feature space is a Hilbert space of infinite dimension. Maximum margin classifiers are well regularized, so the infinite dimension does not spoil the results. Some common kernels include,

- Polynomial (homogeneous):
$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}')^d$
- Polynomial (inhomogeneous):
$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + 1)^d$
- Radial Basis RBF:
$k(\mathbf{x}, \mathbf{x}') = \exp(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{2\sigma^2})$
- Sigmoid:
$k(\mathbf{x}, \mathbf{x}') = \tanh(\kappa \mathbf{x} \cdot \mathbf{x}' + c)$, for $\kappa > 0$ and $c < 0$

Soft margin: In 1995, Corinna Cortes and Vapnik [15] suggested a modified maximum margin idea that allows for mislabeled examples. If there exists no hyperplane that can split the "yes" and "no" examples, the Soft Margin method will choose a hyperplane that splits the examples as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples. This work popularized the expression Support Vector Machine or SVM.
Regression: A version of a SVM for regression was proposed in 1997 by Vapnik, Steven Golowich, and Alex Smola[16]. This method is called support vector regression (SVR). The model produced by support vector classification (as described above) only depends on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin. Analogously, the model produced by SVR only depends on a subset of the training data, because the cost function for building the model ignores any training data that is close (within a threshold ε) to the model prediction.

## 4. Result

Support Vector Machine method was used on the set of 683 samples of actual data. Additional set of

data of 117 samples is generated using Neural Network.

The Accuracy or Efficiency of the detection 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 [17]

where N total no of Historical Data and K is number of cases output with MRE less than or equal to p.

Various groups of training and testing data were formed and mean square error was found out as shown in Table-1. The graphs of MSE verses Epochs obtained for various experiments of training samples are shown in figure 2a, 3a, 4a, 5a. The graphs of Actual Output (2 for Benignant and 4 for Malignant) verses Desired Output for the same experiments of training and testing samples are shown in figure 2b, 3b, 4b, 5b.

| Sr. No | No. of Training samples | No. of Testing samples | Efficiency | Min. Training error (MSE) | Refer Fig No |
|---|---|---|---|---|---|
| 1 | 450 | 250 | 95 | 0.025 | Fig2.a,b |
| 2 | 500 | 200 | 96 | 0.024 | Fig3.a,b |
| 3 | 500 | 250 | 96 | 0.024 | Fig4.a,b |
| 4 | 550 | 250 | 97 | 0.022 | Fig5.a,b |

**Table-1**

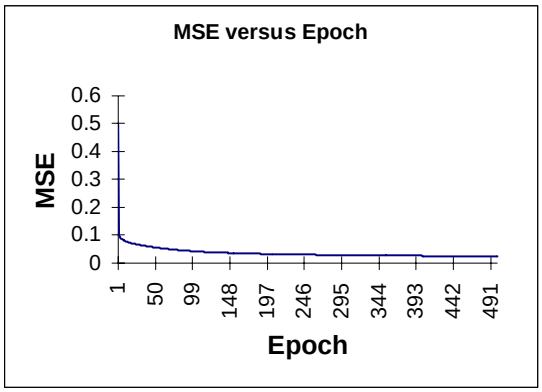


**Figure 2a**

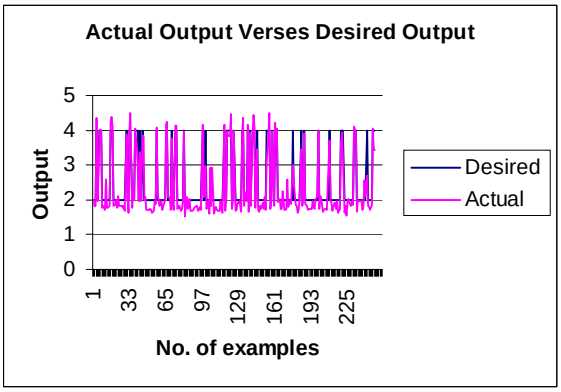


**Figure 2b**

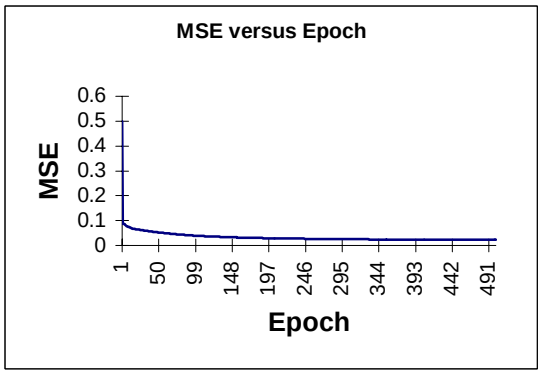


**Figure 3a**

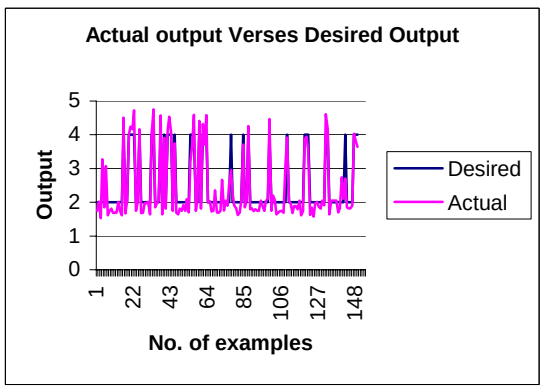


**Figure 3b**

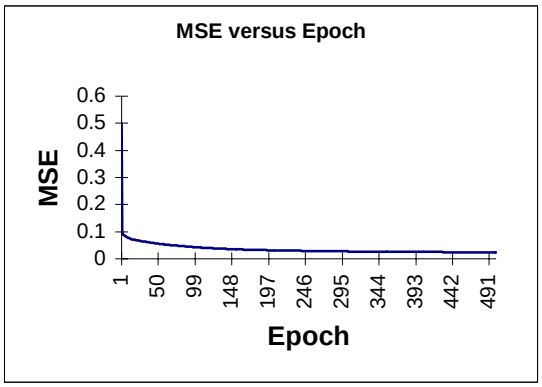


**Figure 4a**

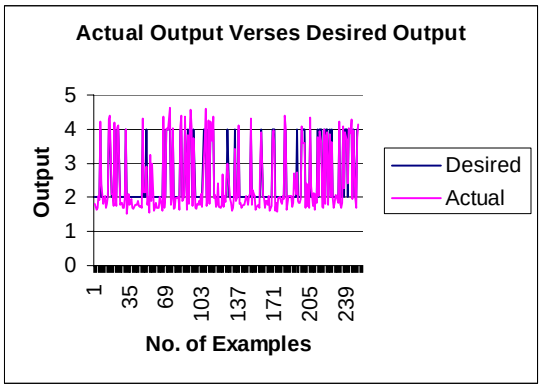


**Figure 4b**

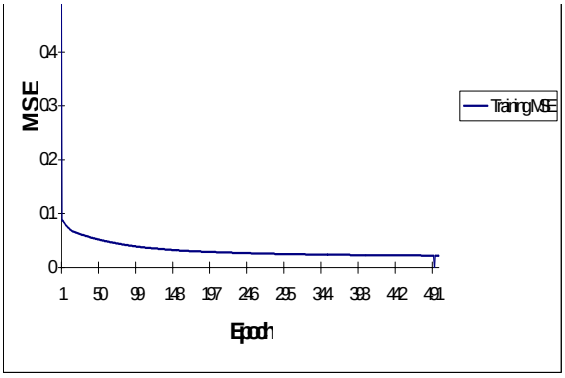


**Figure 5a**

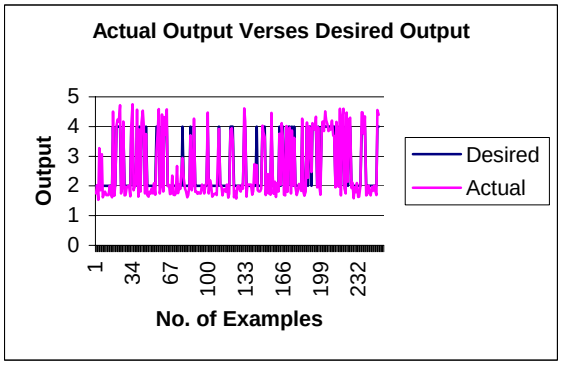


**Figure 5b**

From the above figures, it is observed that as the number of epochs increases upto 500, the MSE decreases but thereafter MSE remains nearly constant.

| Sr. no | Num. of Training sample | Num. of Testing sample | Experiment performed for different step sizes keeping epochs same (500) Efficiency in % | |
|---|---|---|---|---|
| | | | For Step size = 0.01 Efficiency in % | For Step size = 0.02 Efficiency in % |
| 1 | 459 | 224 | 96.42 | 96.43 |
| 2 | 469 | 214 | 96.26 | 96.26 |
| 3 | 479 | 204 | 96.07 | 96.08 |
| 4 | 489 | 194 | 96.39 | 96.39 |
| 5 | 499 | 184 | 96.19 | 96.20 |
| 6 | 509 | 174 | 95.97 | 95.98 |
| 7 | 519 | 164 | 95.73 | 95.73 |
| 8 | 529 | 154 | 95.45 | 95.45 |
| 9 | 539 | 144 | 95.83 | 95.83 |
| 10 | 549 | 134 | 96.26 | 96.27 |
| 11 | 559 | 124 | 95.96 | 95.97 |
| 12 | 569 | 114 | 95.61 | 95.61 |
| 13 | 579 | 104 | 95.19 | 95.19 |
| 14 | 589 | 94 | 94.68 | 94.68 |
| 15 | 599 | 84 | 94.04 | 94.05 |
| 16 | 609 | 74 | 17.56 | 94.59 |
| 17 | 619 | 64 | 18.75 | 95.31 |
| 18 | 629 | 54 | 94.44 | 94.44 |
| 19 | 639 | 44 | 93.18 | 93.18 |
| 20 | 649 | 34 | 94.11 | 94.12 |
| **Average Efficiency** | | | **87.70** | **95.39** |

**Table –2**

Using actual data of 683 samples and varying the step size, the efficiencies are obtained with 20 experiments, shown in table-2. As the step size is increased, the efficiency increases which is depicted by the graph shown in figure-6.

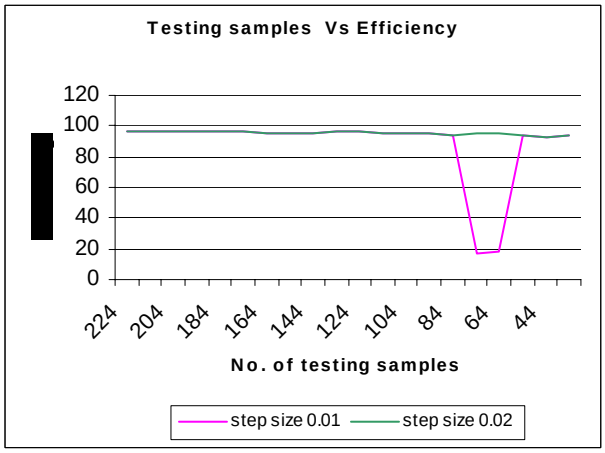


**Figure-6 (For Table-2)**

## 5. Conclusion

Using Support Vector Machine ANN, the prediction of diagnosis and detection of breast cancer is comparably accurate than the human being. The efficiency of manual detection of breast cancer is 85% and the efficiency of the Support Vector machine recognition obtained is nearly 97%. This high rate of accuracy can be utilized to support the Doctor's decision to avoid Biopsy

## 6. Future Scope

The network has to be trained with more input patterns, so that the generalization ability of network will be enhanced. This method can be used for the diagnosis of other types of diseases.

## *References:*

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

[2](Mangasarian, 1995) 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](Wolberg, 1995) 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](Wolberg, 1994) 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] E. Boser, I. M. Guyon, and V. N. Vapnik.' *A training algorithm for optimal margin classifiers'*. In D. Haussler, editor, 5th Annual ACM Workshop on COLT, pages 144-152, Pittsburgh, PA, 1992. ACM Press

[14]Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1999. ISBN 0-387-98780-0

[15] Corinna Cortes and V. Vapnik, "Support-Vector Networks*, Machine Learning, 20, 1995.*

[16] Bernhard Schölkopf, Christopher J.C. Burges, and Alexander J. Smola (editors). "*Advances in Kernel Methods: Support Vector Learning*". MIT Press, Cambridge, MA, 1999. ISBN 0-262-19416-3.

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