ANALYSIS OF OVARIAN CANCER USING BACK-PROPAGATION ALGORITHM WITH GRADIENT-DESCENT MODELS

M. Malathi [1]

ABSTRACT

Most Women is at danger ovarian cancer; about 90 percent of women who get ovarian cancer are above 40 years of age, with the greatest number of ovarian cancers taking place at the age of 6o years and above. Ovarian cancer arises from the ovary indicates growth of cancer. Each year, almost 20,000 women in the United States get ovarian cancer and also, ovarian cancer is the eighth most common cancer and the fifth leading cause of cancer death. More than 91% of ovarian cancers are known as "epithelial tumors" which start from the outer surface epithelium, the thin tissue forming the outer layer of a body's surface and lining the alimentary canal and other hollow structures of the ovary. The intention of this study is to observe the performance of the Back propagation techniques using Gradient-Descent Model over Genetic Algorithm on the analysis of ovarian cancer using proven ovarian dataset fig 2. Ovarian cancer [2] accounts for the most caused cancer diagnoses among women. We propose a comparison between Genetic Algorithm and BPA in ANN [1] for preoperative guess of enmity in ovarian tumors. Most of the present methods do not meet the requirements which deal with the drawbacks like accuracy and noise. Gene ranking methods like T-Score, ANOVA, went for wrong prediction the rank when large database is applied. The typical BPA is proposed to form part of a predictable tool to distinguish between kind and unkind ovarian tambours. This may help doctors to fix on the applicable treatment for the patients.

Keywords : Gene Ranking, predictable tool, ovarian tambours, Gradient-descent model.

I. INTRODUCTION

Cancer precedents to almost 27% of all mortalities, making it the leading cause of death in America and also around the world. Timely and exact detection of cancer is life-threatening to the comfort of patients.

A major application of microarrays has been to the study of cancer. Recognition of the signals that are symptoms for the disease phenotype and its progression requires the use of hardy techniques

Cancer can be identified through the analysis of genetic data. The human genome contains almost 10 million single nucleotide polymorphisms which will be in charge for the difference that lies between human beings. The microarray technology is used to achieve gene expression levels and SNPs of an individual. . In this paper, two methods are used, namely Artificial Neural Network using Back propagation algorithm and Genetic algorithm, and use techniques such as dimensionality reduction to improve the accuracy rate of the classifier.

---

[1] Lecturer, L.R.G. Govt Arts College For Women, Tirupur -4, E-mail : malathi145@gmail.com

Gene expression profiling by micro array techniques are playing a vital role. Various researchers have proposed a number of machine learning and data mining approaches for identifying cancerous nodule using gene expression data. But, these existing techniques have certain limitations that do not handle the particular needs of gene micro array examination. Initially, micro array data is featured by a high-dimensional feature space repeatedly exceeding the sample space dimensionality by a factor of 100 or higher. Moreover, micro array data consists of a high degree of noise. Most of the conventional approaches do not adequately handle with the limitations like dimensionality and noise. Gene ranking techniques are later proposed to overcome those problems. Some of the widely used Gene ranking techniques are T-Score, ANOVA, etc. But those approaches will sometimes wrongly predict the rank when large database is used. To overcome these issues, this paper proposes an efficient feature selection technique

Artificial Neural Networks (ANN) "direct information processing tactics, built from linked elementary processing devices called neurons stimulated by the biological nervous systems" [3]. Artificial neural networks are extensively used with appliances in science and technology

An ANN using Back propagation algorithm is a mathematical illustration of the human neural design, representing its "learning" and "generalization" capability. For this cause, ANNs have its place in the area of artificial intelligence. ANNs are extensively applied in research for the reason that they can model highly non-direct systems in which the relationship among the variables is undetermined or very complex.

II. Related Work

There are different techniques proposed by different authors for the prediction of cancer regions. Every technique has its own advantages and disadvantages. Some of the existing techniques are presented in this section.

Rui *et al.*, [12] proposed a multiclass cancer classification using semisupervised ellipsoid ARTMAP and particle swarm optimization with gene expression data [11]. It is critical for cancer prediction and treatment to perfectly categorize the site Of origin of a cancer. With huge progress of DNA microarray Techniques, creating gene expression profiles [8] for various cancer kinds has previously turn out to be a capable way for cancer classification [10]. In addition to research on binary classification like normal versus tumor samples that focuses on various issues from a mixture of disciplines, the discrimination of multiple tumor kinds is also essential. In the meantime, the choosing of genes that are appropriate to definite cancer kinds not only enhances the performance of the classifiers, but also offers molecular insights for treatment. Here, the author utilizes the semisupervised ellipsoid ARTMAP (ssEAM) for multiclass cancer discrimination and particle swarm optimization for informative gene selection. SsEAM is a neural network technique [14] embedded in adaptive resonance theory and applicable for classification purpose.

III. BACKGROUND STUDY

Classification of cancer by gene expression data is known to hold the keys for indicating the basic troubles relevant to cancer analysis and drug identification. The latest advent of DNA microarray method has made simultaneous care over of thousands of gene expressions probable. With the lavishness of gene expression data researchers have started to discover the chances of cancer recognition using gene expression data. Recently, many methods have been proposed with promising results.

In order to achieve in-depth knowledge into the cancer classification problem, it is must to have a close observance which look at the problem the proposed method and the relevant problem all together. In this paper present a widespread indication of numerous proposed cancer classification methods and estimate them based on their calculated time, recognition accuracy and ability to reveal biologically useful gene information.

Accuracy is the major factor in cancer classification, where it alone only the objective that wants to achieve. Biological significance is another issue, since any biological information obtained during the process can support in further gene function exploration and other research studies. Some useful information can be gained from the classification process is the purpose of the genes that work as a group in resolving the cancerous tissues or cells or the genes that are under-expressed or over-expressed in certain tissues or cells. Dealing with a huge amount of unrelated attributes is a challenging factor. Though unrelated attributes are exists in almost all kind of data sets, researchers have with earlier dealings, but the ratio of unrelated attributes to the related attributes is not as bulk as that in the gene expression data.

IV. METHODOLOGY

In the medical research, Cancer research is one of the foremost research areas. Exact predictions of various tumor types has great value in granting better treatment and reduce harmfulness on the patients. In earlier days, cancer identification has always been morphological and clinical based. These typical cancer classification methods are boomed to have several demerits in their diagnostic capability. The cancer training gene data set is divided into several subsets with almost 1000 genes in each subset as shown in Fig.1. The division of the data sets can be taken place unintentionally.

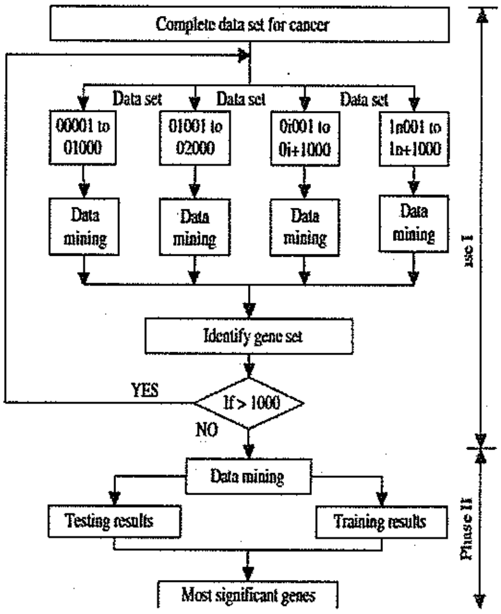


Figure 1. Integrated gene-search algorithm

4.1 Ovarian cancer

Ovarian cancer is a cancerous growth rise from the ovary where the indications are often very faint at initial stage on and may have bloating, pelvic pain, difficulty eating and frequent urination, and other disease. More than 90% of ovarian cancers are categorized as "epithelial" and are supposed to arise from the base (epithelium) of the ovary. In some case, suggests that the fallopian tube might be the cause for ovarian cancers. Meanwhile, the ovaries and tubes are tightly connected to each other; it is assumed that these fallopian cancer cells can simulate ovarian cancer. The Genetic algorithm was autonomously implied to the data set to extent the involvement of each individual gene. When the number of expressive genes was more than 1000, the Genetic algorithm was applied one more time for experiment, which reduces the number of significant genes by **81.12%.**

Types of Ovarian cancer :

The ovaries consist of 3 main types of cells; each of these cells can develop into different types of tumor.

Epithelial tumors start from the outer surface cells that cover the ovary. Most ovarian tumors are epithelial cell tumors.

Germ cell tumors start from the cells that produce the eggs (ova) .

Stromal tumors start from structural tissue cells that hold the ovary together and produce the female hormonesestrogen.

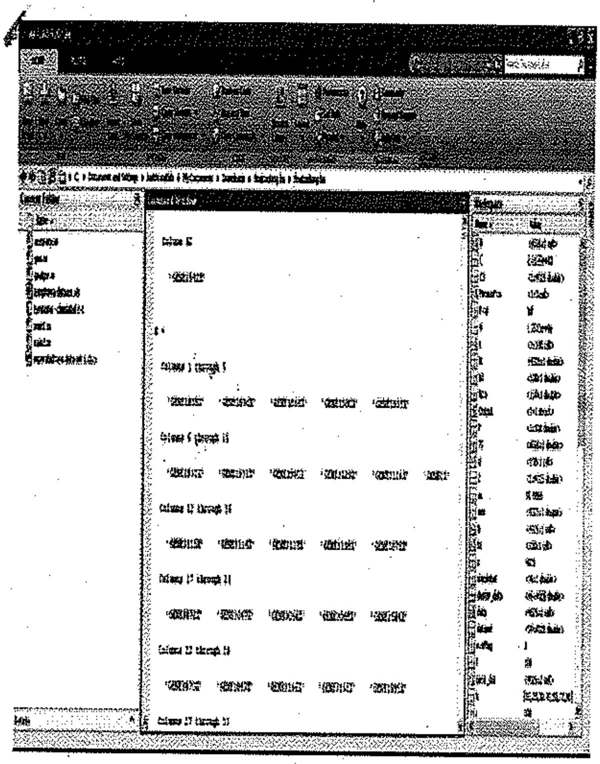


Figure 2. Dataset

4.2 Back Propagation Algorithm

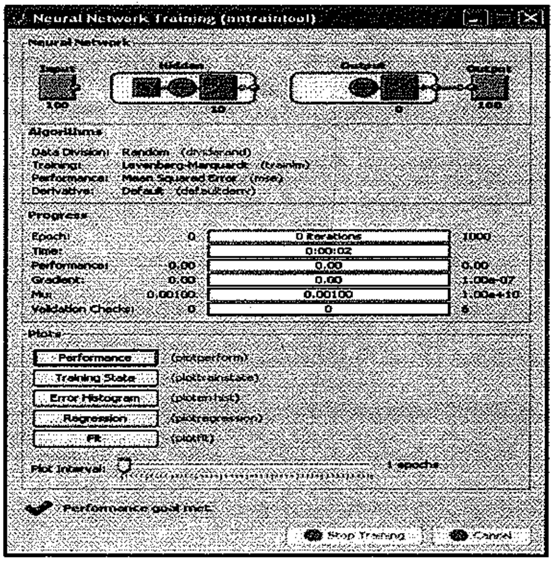
Back propagation is the proposed algorithm, in which there will neural network for each stage. The code takes the preferred picture and implements it through the neural networks one by one, each time outcomes in a score. For each network we can review the most important methods. Artificial Neural Networks using BPA represent a powerful method to support clinical researchers to perform diagnosis and other enforcements. Several neural network learning algorithms exist, but this section will describe what perhaps the most commonly used algorithm is, back-propagation learning. The weights and biases are initially assigned to a random continuous (real) value between -1 and 1. The algorithm trains the neural network incrementally, meaning that each instance is considered independently, and after considering a training instance, the weights and biases are updated before considering another one. The update is done in a way that performs a gradient descent model minimization of the error with respect to the weights and biases. Methods of summarizing and detailed information and intelligent data are continuously improving and can provide greatly to successful, complete and swift medical analysis as results in fig 3.

Training and Testing stages: The method and its parameters below are used to experimented our neural network: **net = train (net, inputs, targets);** The function parameters are :

1. **net :** the neural network which created previously.

2. **Inputs :** inputs of the created neural network as defined before.

3. **Targets:** stated target the neural network.



Algorithms

Data Division: Random (dividerand)
Training: Levenberg-Marquardt (trainlm)
Performance: Mean Squared Error (mse)
Derivative: Default (defaultderiv)

Progress

Epoch: 0 | 0 iterations | 1000
Time: 0:00:02
Performance: 0.00 | 0.00 | 0.00
Gradient: 0.00 | 0.00 | 1.00e-07
Mu: 0.00100 | 0.00100 | 1.00e+10
Validation Checks: 0 | 0 | 6

Plots

(plotperform)
(plottrainstate)
(ploterrhist)
(plotregression)
(plotfit)

Plot Interval: 1 epochs

✓ Performance goal met.

Figure 3. Neural network model using Training Tool

4.3 Gradient-Descent Model

The standard back propagation algorithm we use an approximate steepest descent rule model . The performance index is approximated by value.Where the total sum of squares is replaced by the squared errors for a single input/output pair. The approximate steepest (gradient) descent algorithm is then used the learning rate. The update is done in a way that performs a gradient descent minimization of the error with respect to the weights and biases.

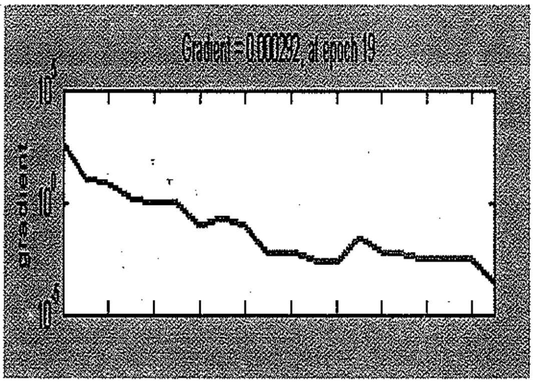


Figure 3.3 (a) The approximate steepest (gradient) descent algorithm using train the data

V EXPERIMENTAL RESULTS

The Ovarian data sets were used in this research to identify affected that to differentiate the kind and malicious of ovarian cancer cases. The data sets consisted of 100 samples, 100 normal tissues samples, and 14 kind samples. Only the cancerous and noncancerous samples are included in this paper. Each sample consisted of 14.094 intensities with intensities of features. The proposed system results that Genetic algorithm and Neural Networks using

Back propagation will be useful tool to distinguish affected or unkind, kind and normal cells. Furthermore, the ANN using BPA obtained the best performance, classifying those tissues, with a good success rate, accuracy and compassion. It may minimize the number of unwanted tissues and delayed cancer identification. Based on this performance, it is experimented that such features provide considerable assistance to a medical research, and the results were boosting when tissues were classified with Neural Networks. Genetic algorithm is used to hold those genes in the dataset that provide most to its conflict. The monetary and social values of Ovarian Cancer are very high and its outcome has the problem which impressed many researchers in the computational intelligence in recent times. Because of the significance of achieving exact classification, Artificial Neural Networks (ANNs) are the best then genetic algorithm.

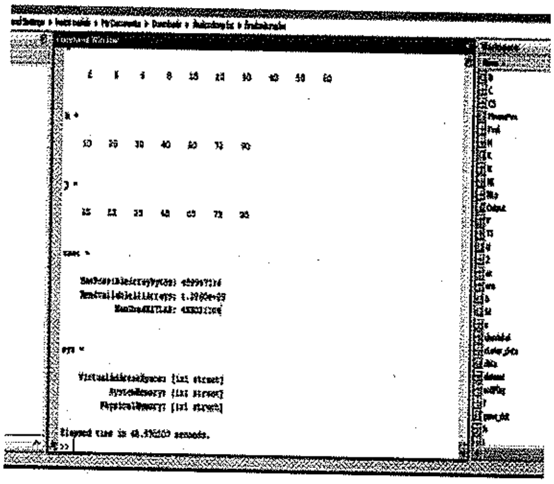


Figure 4. Memory and speed calculation

The experimental results show **in Fig 5(a) & (b)** that when Genetic algorithm is compared with Artificial Neural Networks using Back propagation, later is the best algorithm for classification of medical data. It is also observed that Artificial Neural Network performs well for classification on medical data sets of increased size. Although the research described in this proposed paper has given an extensive comparison of models and techniques, and extended the work done previously it has not exhausted the possibilities of classification models.

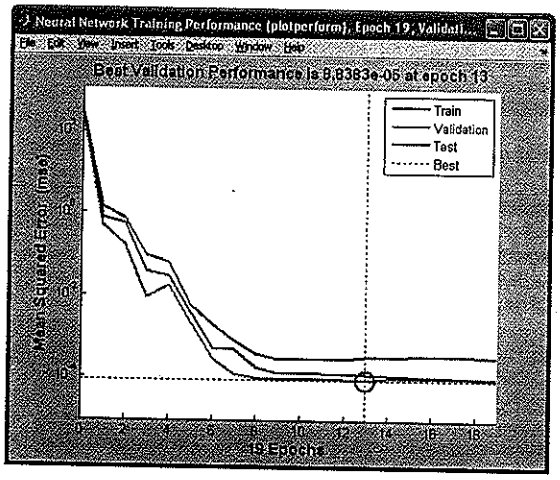


Figure 5.(a). Performance of the Neural Network Training data of Mean Squared Error (mse) using Best validation at Epoch 13

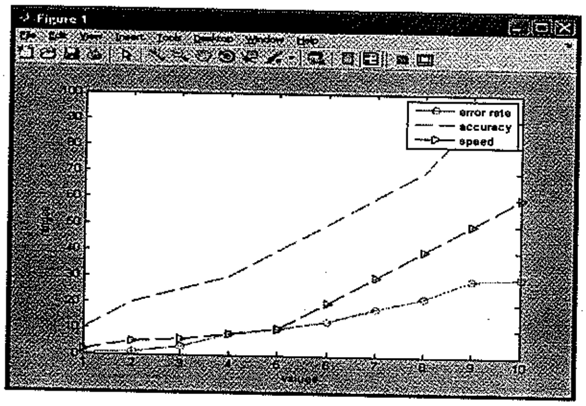


Figure 5. (b). Performance of the Artifical neural network using Ranges and Values

VI. CONCLUSION

In the past cancer categorization is generally depends on morphological and clinical analysis. To overcome those drawbacks in cancer classification, efficient technique in accordance with the global gene examination have been evolved. The microarray gene data must be preprocessed for classification with good accuracy using the classifier. This paper proposes a method for refinement and Categorizing the ovarian with kind, spreading and normal tissue using Genetic Algorithm and Back propagation algorithm of ANN. This method was experimented for a ovarian dataset and radial basis operates neural networks. The best result was acquired with probabilistic back propagation using neural networks resulting with high accuracy and good success results rate with 98% of performance for classification when compared to the conventional techniques.'

VII. REFERENCES

[1] Basheer I, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Meth. 43: 3–31, 2000.

[2] J.K.Yu, "An integrated approach utilizing proteomics and bioinformatics to detect ovarian cancer," *Jornal of Zhejiang University SCIENCE,* vol. 6, no.4, pp. 227-231, July 2005.

[3] Deepa S and Aruna Devi B (2011) A survey on artificial intelligence approaches for medical image classification. *Indian.J. Sci. Technol.* 4(11), 1583- 1595

[4] D. Lawrence, Handbook of Genetic Algorithms, Van Nostrand Reinhold, New York, 1991.

[5] Dayhoff J, Deleo J. Artificial Neural Networks: Opening the Black Box. Cancer. 91: 1615–1635, 2001.

[6] Runxuan Zhang, Huang, G.B., Sundararajan, N. and Saratchandran, P., "Multicategory Classification Using An Extreme Learning Machine for Microarray Gene Expression Cancer Diagnosis", IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 4,No.3, Pp. 485 – 495, 2007

[7] Xiao gang Ruan, Jillian Wang, Hui Li and Xiaoping Li, "A Method for Cancer Classification Using Ensemble Neural Networks with Gene Expression Profile", The 2nd International Conference on Bioinformatics and Biomedical Engineering, Pp. 342-346, 2008.

[8] Khan, J., Wei, J., Ringner, M. and Saal, L.,"Classification and Diagnostic Prediction of Cancers using Gene Expression Profiling and Artificial Neural Networks", Nature Medicine, 2001.

[9] Zhang, H., Yu, C., Singer, B. and Xiong, M., "Recursive Partitioning for Tumor Classification with Gene Expression Microarray Data". PNAS, Pp. 6730–6735, 2001.

[10] Guyon, I., Weston, J., Barnhill, S. and Vapnik, V.,"Gene Selection for Cancer Classification using Support Vector Machines", Machine Learning, 2000.

[11] Berns, A., "Cancer: Gene Expression in Diagnosis", Nature, Pp. 491–492, 2000

[12] Rui Xu, Anagnostopoulos, G.C. and Wunsch, D.C.I.I. ,"Multiclass Cancer Classification Using Semisupervised Ellipsoid ARTMAP and Particle Swarm Optimization with Gene Expression Data", IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol.4,No.1, Pp. 65-77, 2007

AUTHOR'S BIOGRAPHY

M.Malathi MCA., M.Phil., working as a Lecturer in the Department of Computer Science, LRG Govt. Arts College for Women, Tirupur. She has 9years of teaching experience. Her area of interest is Data Mining. She has presented 1 National Conference, 1 International Journal and 1 National Workshop