

Optimistic Diagnosis of Acute Leukemia Based On Human Blood Sample Using Feed Forward Back Propagation Neural Network

Arunkumar Sivaraman^{#1}, S.Arun Rajesh^{#2}, Dr.M.Lakshmi^{*3}

^{#1} Department Of Computer Science and Engineering, Manonmanium Sundaranar University, Trinaveli, India

^{#2} Department Of Computer Science and Engineering, Manonmanium Sundaranar University, Trinaveli, India

^{*3} Professor & Head, Department of Computer Science and Engineering Sathyabama University, Chennai, India

ABSTRACT— Blood cancer disease is one of the leading causes of death among men in developed and developing countries. Its cure rate and prognosis depends mainly on the early detection and diagnosis of the disease. In order to conserve the life of the individuals who are endured by the Blood cancer disease, it should be pre-diagnosed. So there is a demand of pre-diagnosis method for Blood cancer disease which should provide superior results. In this manuscript we illustrate a process to classify the microarray gene expression data based on their blood sample types using data mining and image processing techniques. The proposed Blood cancer pre-diagnosis system is a combination of Feed Forward Back Propagation Neural Network grouping with Statistical Approach and Fuzzy Inference System. The ultimate objective is to solve the drawbacks in dimensionality reduction as they have a direct impact on the robustness of the generated fuzzy rules. Consequently, the goal is to generate fuzzy rules based on dimensionality reduced data. Then the risk factors and the indications from the dimensional concentrated dataset are given to the Feed Forward back Propagation Neural Network to accomplish the training process. In the testing practice, more data are given to the trained fuzzy system to finalize whether the given testing data envisage the Blood disease perfectly or not.

KEYWORDS— Data Mining, Feed Forward Back Propagation Neural Network, DNA Microarrays, Cancer Diagnosis

I. INTRODUCTION

Blood is a connective tissue consisting of cells suspended in plasma. From the identification of blood disorders, it

can lead to classification of certain diseases related to blood. This paper describes a preliminary study of developing a detection of leukemia types using microscopic blood sample images. Analyzing through images is very important as from images, diseases can be detected and diagnosed at earlier stage. From there, further actions like controlling, monitoring and prevention of diseases can be done. Blood's major functions are to transport various agents such as oxygen, carbon dioxide, nutrients, wastes, and hormones. Blood cells are composed of erythrocytes (red blood cells, RBCs), leukocytes (white blood cells, WBCs) and thrombocytes (platelets). The most abundant small reddish cells are erythrocytes and called red blood cell. An erythrocyte is a discoid cell with a thick rim and a thin sunken center [1]. RBCs' two principal functions are to move oxygen from lung to tissues elsewhere and transport carbon dioxide from tissues to the lung. Whereas, the Leukocytes or white blood cells are part of the immune system. The conventional device used to count blood cells is the hemocytometer. It consists of a thick glass microscope slide with a rectangular indentation creating a chamber of certain dimensions. This chamber is etched with a grid of perpendicular lines. It is possible to count the chamber of cells in a specific volume of fluid, and calculate the concentration of cells in the fluid [2,3]. To count blood cell, physician must view hemocytometer through a microscope and count blood cells using hand tally counter. The organized in order and are of equal size. Cancer cells are different than normal cells. They are in dispersed order, their sizes are different and they are not structured well. Issue in this method is that a doctor with his naked eye and a microscope cannot remember a large number of patterns

Optimistic Diagnosis of Acute Leukemia Based on Human blood Sample Using Feed Forward Back Propagation Neural Network

of the disease. The need is to automate this process to make the cancer diagnosis efficient and fast with the use of state of the art technology.

A. Genes and their importance in Cancer Diagnosis

Genes provide very valuable information which can be used to study any disease in depth. Study of genes from a cancer patient helps us diagnose cancer and differentiate between types of cancer. It also helps in separating the healthy people from the patients. Genes contains infinite patterns that cannot be recorded manually using a microscope. DNA Micro Arrays are used to study the information obtained from Genes.

B. DNA Micro Arrays

DNA microarrays are the latest form of biotechnology. These allow the measurement of genes expression values simultaneously from hundreds of genes. Some of the application areas of DNA microarrays are obtaining the genes values from yeast in various ecological conditions and studying the gene expression values in cancer patients for different cancer types. DNA Microarrays have huge potential scientifically as they can be useful in the study of genes interactions and genes regulations. Other application areas of DNA microarrays are clinical research and pharmaceutical industry [1].

C. Data Retrieval from DNA Micro Arrays

Gene expression data is retrieved from DNA microarray through Image processing techniques. Data for a single gene consists of two intensity values of fluorescence i.e. Red and Green. These intensities represent expression level of gene in Red and Green labeled mRNA samples. Image of a microarray is scanned. This image is then processed through image processing techniques [2].

II. RELATED WORKS

Kiran et al. [4], have surveyed the various Neural Networks that have been used in successful classification of medical data for various disorders. Examples include - Feed Forward Neural Network, Radial Basis Function (RBF) Network, Kohonen self-organizing network, Fuzzy Neural Network, Probabilistic Neural Network. For the successful detection of Blood Cancers, various methods have been surveyed. They are as follows: To perceive the Blood cancer in its premature stages Hopfield Neural Network (HNN) and Fuzzy C-Mean (FCM) clustering algorithm were used for segmenting sputum color images. HNN showed better classification result than FCM, and it flourished in extorting the nuclei and cytoplasm regions. Nevertheless FCM unsuccessful in extorting the nuclei, as a substitute it detected only part of it; and was not

sensitive to intensity variations. [5] With the development of medical technology, the medical images play a more and more important role in diagnosis. X-ray, CT and MRI images have widely been applied to diagnosis. [6]

ANN has the potential to improve the diagnostic accuracy. Biochemical Diagnosis, Imaging Diagnosis and Cytology Histology diagnosis are the three main methods for Blood cancer diagnosis. Imaging diagnosis includes X-ray imaging, CT, MRI, angiography and interventional radiology. CT is an important medical imaging method employing tomography, which is one of the most widely used for the diagnosis of Blood cancer. The result of ANN was evaluated with those of logistic regression by ROC curve study. The analytical accuracy of ANN and logistic regression with all samples of the test group and training group were 84.6% and 96.6%. [7] In [8] a Radial Basis Neural Network (RBFN) was used for Blood cancer screening.

Because of its learning characteristics it was selected to train the samples and then extract the internal relation between the pathogenic factors and inducing Blood cancer, and eventually it generates empirical function and forecasts the new samples. The training function adopted Linear Least Square method (LLS) and the Gradient Descent hybrid learning algorithm to optimize the training process and the screening results. The accuracy of Blood cancer identification was 95.32%. [9] In order to recognize features of patient fragments where tolerable survival is significantly higher/lower than middling endurance across the intact dataset, Association Rule Mining techniques were used for the identification of hotspots from Blood cancer data. Automated association rule mining practices answered in hundreds of rules, from which many outmoded rules were physically removed based on domain understanding. The ensuing rules conformed to existing biomedical knowledge and provided interesting insights into Blood cancer survival. The Hotspot algorithm is an association rule mining algorithm which is directed by a target attribute, which means that the consequent is fixed to the target attribute. It can be used for segmentation with both nominal and numeric targets. [10] Computer Aided Diagnosis (CAD) is procedures in medicine that assist doctors in the interpretation of medical images. Imaging techniques in X-ray, MRI, and Ultrasound diagnostics yield a great deal of information, which the radiologist has to analyze and evaluate comprehensively in a short time. A relatively young interdisciplinary technology, CAD combines the elements of Artificial Intelligence and Digital Image Processing with radiological image processing.[11] and [12] use CAD for Blood cancer. The design and development of a two stage CAD system that can automatically detect and diagnose histological images such as CT scan of Blood with a nodule into cancerous or noncancerous nodule was done in [10]. In the first stage the input image is pre-processed and the cancerous

Optimistic Diagnosis of Acute Leukemia Based on Human blood Sample Using Feed Forward Back Propagation Neural Network

nodule region is segmented. Second stage involves diagnosis of the nodes based on fuzzy system and the grey level of the nodule region. While maintaining a high degree of true-positive diagnosis, this proposed method attained an accuracy of 90% and also high detection sensitivity and specificity, which meets the basic requirements of clinical diagnosis [10]. Machine Learning techniques were utilized to develop a CAD system, which consisted of feature mining phase, feature collection phase and cataloging phase. Different wavelets functions have been used in Feature Extraction/Selection, to find the one that produces the uppermost precision. Clustering K-nearest-neighbor algorithm has been consumed for classification. Testing was done using Japanese Society of Radiological Technology's standard dataset of Blood cancer. Of the 154 nodule regions (abnormal) and 92 non-nodule regions (normal), a precision level of over 96% was achieved for classification. [11]

III. PROPOSED METHODOLOGY

Our proposed architecture for Blood cancer pre-diagnosis system is shown in Fig1. At the early stage, the dimensionality of the given dataset is reduced using Statistical Analysis with Fuzzy Inference System (FIS). After the dimensionality reduction, the reduced dataset are given as the input to the pre-diagnosis stage. While exploiting a supervised training process, the network must be afforded with both sample inputs and anticipated outputs. The anticipated outputs are compared against the actual outputs for given input.

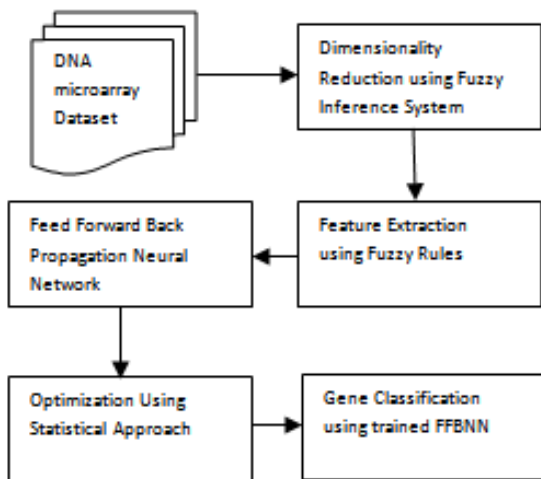


Fig. 1 Proposed Architecture

Using the predictable outputs, the back propagation training algorithm acquires an intended error and fine-tuned the weights of the different layers rearward from the output layer to the input layer.

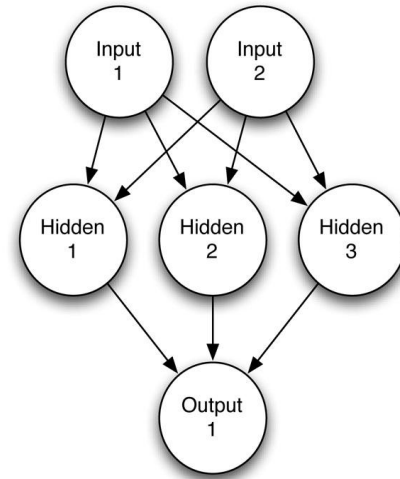


Fig. 2 Structure of a Feed-Forward Neural Network

The following steps are followed to build and train a network [19];

1. Create an initial neural network with number of hidden unit $h = 1$. Set all the initial weights of the network randomly within a certain range.
2. Train the network on training set by using a training algorithm for a certain number of epochs that minimizes the error function.
3. If the error functions $av \xi$ on validation set is acceptable and, at this position, the network classifies desired number of patterns on test set that leads the efficiency E to be acceptable then *stop*.
4. Add one hidden unit to hidden layer. Randomly initialize the weights of the arcs connecting this new hidden unit with input nodes and output unit(s). Set $h = h + 1$ and go to *step 2*.

For back propagation algorithm the weight adjustment is:
For the output-layer weights:

$$w_{kj_{new}}^o = w_{kj_{old}}^o + \eta \delta_{pk}^o j_{pj} \text{-----(1)}$$

Where, $\delta_{pk}^o = \delta_{pk} f_k'(net_{pk}^o)$

For the hidden-layer weights:

$$w_{ji_{new}}^h = w_{ji_{old}}^h + \eta \delta_{pj}^h x_{pi} \text{-----(2)}$$

Where, $\delta_{pj}^h = f_j'(net_{pj}^h) \sum_k \delta_{pk}^o w_{kj}^o$

Where, k indicates the k_{th} output unit, j indicates the j_{th} hidden unit; i indicates the i_{th} input node, p is the input vector, is the learning rate is the error term, x_{pi} is the input

Optimistic Diagnosis of Acute Leukemia Based on Human blood Sample Using Feed Forward Back Propagation Neural Network

value to the i , is the output function of j connected to k
The error function is usually defined as the mean squared-errors.

$$e_k = d_k(n) - y_k(n) \text{-----(3)}$$

$$\xi(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) \text{-----(4)}$$

$$\xi_{av} = \frac{1}{N} \sum_{n=1}^N \xi(n) \text{-----(5)}$$

Where, k denotes k_{th} output unit, n denotes the n_{th} iteration, C is the number of output units, N is the total number of patterns, d_k denotes the desired output from k , y_k denotes the actual output of neuron k , e_k denotes the error term for k_{th} output unit.

IV. CONCLUSION

In this manuscript, we have proposed a Blood cancer pre-diagnosis system with the aid of Statistical Approach with Fizzy Inference System and Feed Forward Back Propagation Neural Network. The proposed system was implemented and a huge set of test data's were utilized to analyze the outcomes of the proposed Blood cancer pre-diagnosis system. Thus the proposed Blood cancer pre-diagnosis system offers a significant tempo of accuracy, sensitivity and specificity. We can say that proposed method more precisely diagnosis the Blood cancer from the given test data by seeing the elevated rate of measurements.

REFERENCES

[1] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. "Gene selection for cancer classification using support vector machines", Machine Learning, 2000. [2] Duerinckx AJ, Pisa EJ. Filmless Picture Archiving and Communication System (PACS) in Diagnostic Radiology. Proc SPIE 1982;318:9-18. Reprinted in IEEE Computer Society Proceedings of PACS'82, order No 388
[3] Principles of Data Mining. Max Bramer, BSc, PhD, CEng, FBCS, FIEE, FRSA, Digital Professor of Information Technology, University of Portsmouth, UK. ISBN-10: 1-84628-765-0.
[4] Kiran S. Darne & Suja S. Panicker, "Use of Artificial Neural Networks in Medical Data Classification: A Review", International Conference on Advanced Computer Sciences, Communication and Information Technologies (ICACSIT), 2012.
[5] Fatma Taher and Rachid Sammouda, "Blood Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Methods", IEEE GCC conference and exhibition, 2011..
[6] Thessa T.J.P. Kockelkorn, Eva M. van Rikxoort, Jan C. Grutters and Bram van Ginneken, "Interactive Blood Segmentation in CT Scans with Severe Abnormalities" 2010.
[7] Yongjun WU, Na Wang, Hongsheng ZHANG, Lijuan Qin, Zhen YAN and Yiming WU, "Application of Artificial Neural Networks in

the Diagnosis of Blood Cancer by Computed Tomography", IEEE conference on Natural Computation, 2010.
[8] Wang Tao, Lv Jianping and Liu Bingxin, "Research of Blood Cancer Screening Algorithm Based On RBF Neural Network", International Conference on Computer and Management (CAMAN), 2011.
[9] Ankit Agrawal and Alok Choudhary, "Identifying HotSpots in Blood Cancer Data Using Association Rule Mining", 11th IEEE International Conference on Data Mining Workshops, 2011.
[10] S.Aravind Kumar, Dr.J.Ramesh, Dr.P.T.Vanathi, Dr.K.Gunavathi, "Robust and Automated Blood Nodule Diagnosis from CT Images Based On Fuzzy Systems" International Conference on Process Automation, Control and Computing (PACC), 2011.
[11] Ankit Hamada R. H. AI-Abs, Brahim Belhaouari Samir, Khaled Bashir Shaban, and Suziah Sulaiman, "Computer Aided Diagnosis System based on Machine Learning Techniques for Blood Cancer" International Conference on Computer & Information Science (ICCIS), 2012.
[12] Ankit Jia Tong, Wei Ying, Wu Cheng Dong, "A Blood Cancer Lesions Detection Scheme Based on CT Image", 2nd International Conference on Signal Processing Systems (ICSPS), 2010.
[13] Ankit Xiaozhou Li, Rong Wang and Ming Lei, "Analysis on data fordetection of Blood cancer using serum auto-fluorescence", International Symposium on IT in Medicine and Education (ITME), 2011.
[14] Ankit Aminmohammad Roozgard, Samuel Cheng, and Hong Liu, "Malignant Nodule Detection on Blood CT Scan Images with Kernel RX -algorithm", Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI 2012), 2012.
[15] Ankit Jung Hun Oh, Jeffrey Craft, Rawan Al-Lozi, Manushka Vaidya, "Predicting local failure in Blood cancer using Bayesian networks" Ninth International Conference on Machine Learning and Applications, 2010.
[16] Ankit Tadashi Kondo, Junji Ueno and Shoichiro Takao, "Medical Image Diagnosis of Blood Cancer by Hybrid Multi-layered GMDH-type Neural Network Using Knowledge Base", Proceedings of ICME International Conference on Complex Medical Engineering, 2012.
[17] Ankit Thessa T.J.P. Kockelkorn, Eva M. van Rikxoort, Jan C. Grutters and Bram van Ginneken, "Interactive Blood Segmentation in CT Scans with Severe Abnormalities" 2010.
[18] Ankit PENG Gang, YANG Xiong, LIU Li , "Parallel Immune Algorithm for Blood Cancer Detection in X-Ray Images Based on Object Shared Space" 12th International Conference on Parallel and Distributed Computing, Applications and Technologies, 2011.
[19] Ankit Rudy Setiono and Huan Liu, "Improving Backpropagation Learning with Feature selection", Appears in Applied Intelligence, Vol. 6, No. 2, 1996, pp. 129-140.