INTELLIGENT CANCER CLASSIFICATION AND PREDICTION IN MICRO ARRAY GENE ANALYSIS

Dr. M. Thangamani[1], M s. J. Malar vizhi[2], Ms. M. Indhumathi[3]
[1]*Assistant Professor, Kongu Engineering College, Perundurai (India)*
[2,3]*Students, Engineering, Kongu Engineering College, Perundurai (India)*

ABSTRACT

The main task of this work is to intelligent prediction system for Cancer treatment by micro array gene analysis and making a timely decision regarding the treatment procedure. Detection of cancer using micro array analysis, study will be useful for medical and biological students in the area of micro array analysis to predict the cancer and its size easily. This research increased understanding of biological relationship with cancer diseases which may lead to the identification of additional biomarkers or therapeutic targets. Interpretation of more features or characteristics which help to identify and monitor the target diseases or functions. The system will be helpful to find all type of cancer exactly. It improves the quality of healthcare in the modern era.

Keywords: Data Mining, Prediction, Classification, Micro Array Gene Analysis, Cancer Detection.

I. INTRODUCTION

Over the last few years, microarray gene analysis has emerged as a tool for managing thousands of gene expression levels in parallel. This technology has been successfully used to determine disease states and response to stimuli between cells. Experimental techniques such as oligonucleotide allow comparison of thousands of genes simultaneously. Oligonucleotide arrays provide abundant data which gives an insight into the cells under consideration. The analysis and extraction of useful information from similar cells are challenging and they can help in identifying cancer diseases.

Data mining has been effectively used for clustering and classification problems of normal and abnormal samples of microarray data. The challenges faced in microarray gene expression data involve the high dimensionality of the data. Though gene expression data involve the high dimensionality of the data. Though gene intensity is extensively used to select the genes for modeling the class under study, selecting the number of genes required for classification still remains a challenge. Cancer prediction using microarray data poses a major problem because the number of genes greatly exceeds the number of tissue samples. This is an important machine learning problem, which is referred to as feature selection.

The key implication of this fact is that biological spending will tend to be higher than under a system where individuals bear the costs of their decisions. This has two implications. First, the gains to society from micro array analysis research, take account of the effect of increased knowledge on medical spending. Since individuals do not bear the costs of medical choices, it is possible that the induced increase in health expenditures could offset the direct gains from the biological knowledge. The most practical solution to this problem is to calculate the increased value of improved health net of the increase in medical and biological

spending for cancer diseases. This eliminates the need to separate the contributions to health of increases in biological and medical knowledge and the associated increases in biological and medical spending. Second, if increases in medical knowledge increase medical spending, any divergence between the cost and value of these expenditures will be accounted for in the calculations. Thus it would appear that while important, the impact of third party payers for evaluating the returns to biological relationship with the target diseases research is something that can be dealt with.

II. NATIONAL AND INTERNATIONAL STATUS

Develop centers of excellence for education, training, clinical care and prevention and treatment of cancer disease, located in several world regions and funded by international, national, and nongovernmental agencies. These centers should be closely linked to international centers for cancer disease. The research on the physical, chemical and physiologic mechanisms of crystal nucleation, growth, and aggregation of crystals that form within the nephron and in related vascular tissue is very useful in the view of people affected with cancer in nationalized countries. This national data system collects, analyzes, and distributes information about the use of renal calculi research. The study characterizes the total population of people being treated for cancer; reports on incidence, prevalence, mortality rates, and trends over time; and develops data on the effects of various treatment modalities. The report also helps identify barriers to the delivery of quality health care and opportunities for more focused studies of cancer research issues. Cancer diseases are rising globally, particularly in developing countries where the major underlying causes, diabetes and hypertension, are also on the rise. The renal calculi diseases place long-term demands on health care systems. Patients with advanced stages of the disease need some form of renal replacement therapy, but the increased demand is unlikely to be met in developing countries because of the complexity and high cost of such therapy. Thus, detecting cancer disease and managing the underlying causes are key to saving lives.

III. ORIGIN OF THE PROPOSED RESEARCH

The Cancer prediction tool is known for various features and functions. This type of detection system When a small number of genes are selected, their biological relationship with the target diseases is more easily identified. System reducing the dimension to reduced the computational cost. Assists clinicians in making effective decisions on diagnosis and treatment planning. This system can be used as a second opinion to the doctors. Supports biological students in verifying the concepts. It makes the information process quicker and efficient by accommodating with the knowledge base systems. Provides consistency in clinical decision making, in reliability, in accuracy and in enhanced problem solving capability.

IV. METHODOLOGY

The sequence of a gene does not provide information. Gene expression analysis is obtained from transcribing the gene (DNA) into mRNA and the measure of corresponding mRNA from a particular gene gives the level of activity. The major challenge faced in cancer classification using microarray data is that the number of genes greatly exceeds the number of samples. Each sample has anywhere between a couple of thousands to tens of

thousands of features in it. This problem is addressed by feature selection. Feature selection selects a small set of genes from the sample which has strong correlation with the targeted phenotypes.

The proposed method steps as follows.

- **Preprocessing Data:** Datasets are preprocessed using weka tool. Stemming is performed on dataset to reduce the cube size.
- **Feature selection:** Feature selection to reduce the number of genes used by the classifier. Thus the relationship with the target disease is easily identified. Feature selection also performed using weka tool.
- **Implementation of fuzzy**: Fuzzy logic to be applied in feature selection for faster convergence along with the input vector.
- **Buliding Classifier**: Building classifier is the next stage in this system. In this work used Support Vector Machine (SVM) algorithm using weka SVM classifier. Given a set of binary labeled training data (e.g normal and cancer subjects gene expression profiles), the SVM finds a hyperplane that best separates the two classes of training data. Filters and wrappers to select useful biomarkers.

Next step of our system is classification stage. Classification is used either to categorize different types of cancerous tissues or to distinguish a cancerous tissue from a normal tissue. Many classification techniques are used to analyze microarray gene expression data. Recurrent Neural Network (RNN) are dynamic neural networks which mean that the network uses not only the current inputs but also of the previous operations of the network. In RNN, neuron outputs are fed back into the network as additional inputs with time delay elements. RNNs have one-to-one recurrent connections where neuron outputs are fed back into the network as the input of one neuron and not to all neurons. The recurrent connections have a time-delay and the rest of the forward connections are instantaneous. Context layers are layers that use recurrent connections in its computations.

V. RESULT AND DISCUSSION

The parameters to be used in the proposed neural network design are Number of neurons in input layer is 133, number of neurons in output layer is 1, number of hidden layer is 1, Context unit time constant is 0.8 second, Transfer function of context unit is Integrator, Number of neurons in hidden layer is 4, learning rules are back propagation and number of epochs for termination is 500. In this research for purpose of accuracy improvement, genetic algorithm to be used. Genetic algorithm can be investigated to evolve the alpha and momentum in the hidden layer of the neural network architecture. This type of prediction and classification technique is crucial in the development of cancer detection of risk analysis systems. The system is designed to provide timely expertise regarding the presence of cancer.

Dataset: Prediction of cancer is very useful in the view of people affected with cancer in nationalized countries. In this research colon cancer datasets obtained from Kent Ridge Biomedical Data Repository are used. The dataset consist of the microarray profiling of two thousand genes. These genes are used to differentiate between the expressions patterns of different diseases. The Kentridge dataset consists of 62 gene profiles of different patients in which 40 are normal profiles and the remaining 22 profiles pertain to colon cancer patients.. The study characterizes the total population of people being treated for cancer failure; reports on incidence, prevalence, mortality rates, and trends over time; and develops data on the effects of various treatment modalities. The report also helps identify barriers to the delivery of quality health care and opportunities for more focused studies of renal research issues.

VI. CONCLUSION

In this research reduce the time to discover and develop new cures and treatment options for cancer diseases. □ □ □ □ It reduce the economic burden of diseasby reducing hospitalization and other costs. Hence decrease levels of disability and improves quality of life by reducing pain and suffering. Also increase the substantial support for university research, education, and tomorrow's leading scientists.

REFERENCES

[1] E. Bair and R. Tibshirani, Machine Learning Methods Applied to DNA Microarray Data Can Improve the Diagnosis of Cancer, SIGKDD Explorations, 5(2), 2003, 48-55.

[2] D. Jiang, C. Tang, and A. Zhang, "Cluster Analysis for Gene Expression Data: A Survey," IEEE Trans. Knowledge and Data Eng., 16(11), 2004, 1370- 1386.Chinchalkar, S., "Determination of Crack Location in Beams using Natural Frequencies", Journal of Sound and Vibration Volume 247 (3), 2001, 417-429.

[3] L. Li, C. R. Weinberg, T. A. Darden, and L. G. Pedersen, Gene Selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the ga/knn method, Bioinformatics, 17, 2001, 1131-1142.

[4] Rui Xue, Jianying Li and Denni J.Streveler, Microarray Gene Expression Profile Data Mining Model for clinical cancer research in Proceedings of the 37th Hawaii International Conference on System Sciences – 2004.

[5] Yuchun Tang, Yan-Qing Zhang, and Zhen Huang, Development of Two- Stage SVM-RFE Gene Selection Strategy for Microarray Expression Data Analysis, IEEE/ACM Transactions On Computational Biology And Bioinformatics, 4(3), 2007, 365-381.

[6] Hala Alshamlan, Ghada Bads and Yousef abhali, A Comparative Study of Cancer Classification Methods Using, Lecture Notes in Electrical Engineering 285, *2014*, 389-398.

Biographical Notes

Dr. M. Thangamani is working as Assistant Professor in Kongu Engineering College, Tamil Nadu, India. She has published 23 articles in International journals and presented papers in 53 National and International conferences. She has published 11 books for polytechnic colleges and also guided many UG projects. She is also the editorial member for many International Journals and organizing chair and keynote speaker in International conferences in India and other countries. She received best paper award in many International conferences.

Ms. J. Malar vizhi is presently pursuing B.Sc. final year in Kongu Engineering College, India.

Ms. M. Indhumathi is presently pursuing B.Sc. final year in Kongu Engineering College, India.