Gene Expression Analysis

R.Radha

Assistant Professor, Department of Computer Science, S.D.N.B.Vaishnav College for Women, Chromepet, Chennai-600 044
radhasundar1993@gmail.com

Abstract

In the past decade rapid advances of microarray technologies have made it possible to monitor the expression profiles of thousands of genes under various experimental conditions. The requirements for methods to handle such amounts of data have arisen. These massive source of information extracted from the genome project contain the keys to address fundamental problems relating to the prevention and cure of diseases, biological evolution mechanisms and the understanding of particular functional elements in the human organism. The knowledge of the coding sequences of virtually every gene in an organism is an exciting opportunity to develop methods to study the role of a gene in a specific organism or biological function. One of such methods consists of the monitoring of the level of expression of a gene. It has been shown that specific patterns of gene expression occur during different biological states such as cell development and during normal physiological responses in tissues and cells. There are many data mining techniques which help to analyze the gene expression data. This paper discusses some of these methods adopted by different researchers.

***Keywords* :** *Clustering , Classification, Neural Networks, FNN, Gene Analysis, Microarray*

1. Introduction

Machine learning is a branch of artificial intelligence that employs a variety of statistical, probabilistic and optimization techniques that allows computers to “learn” from past examples and to detect hard-to-discern patterns from large, noisy or complex data sets. This capability is particularly well-suited to medical applications, especially those that depend on complex proteomic and genomic measurements. As a result, machine learning is frequently used in cancer diagnosis and detection. More recently machine learning has been applied to cancer prognosis and prediction. Most approaches to the computational analysis of gene expression data are functionally significant classification of genes in unsupervised fashion and the discrimination of high risk patients from low risk ones. On the other hand, supervised learning techniques use training set to optimize the discrimination model. Artificial Neural Network (ANN) is one of supervised methods and a powerful tool for accurately detecting causal relationships. Fuzzy Neural Network (FNN) is one of the advanced ANN models. Microarray technologies are capable of simultaneously measuring the signals for thousands of messenger RNAs and large numbers of proteins from single samples. Arrays are now widely used in basic biomedical research for mRNA expression profiling and are increasingly being used to explore patterns of gene expression in clinical research. Most research has focused on the interpretation of the meaning of the microarray data which are transformed into gene expression matrices where usually the rows represent genes, the columns represent various samples. Clustering samples can be done by analyzing and eliminating of irrelevant genes. However, majority methods are supervised (or assisted by domain knowledge), less attention has been paid on unsupervised approaches which are important when little domain knowledge is available.

2. Clustering of Genes

Clustering algorithm are currently being applied to search for meaningful patterns in microarray datasets [10,12,32]. Consider a data set $D = \{x_1, x_2, ..., x_n\}$, obtained from a microarray experiment, where each x_i is a vector that contains d measurements of a particular gene and of different time points. The problem of clustering such data is to cluster the genes into groups which posses' similar biological functionality. Genes in the same clusters or groups are expected to have strong similarity of activity patterns, while those in different clusters have weak similarity to each other.

Currently, one of the main methods to analyze the gene expression data rely on clustering algorithms. The fundamental premise for applying such methods is that genes with similar functions under the same conditions are co-expressed in the cycle of development of the cell. Thus, one of the purposes for clustering such gene expression data is to predict the function of unknown genes by grouping them in terms of functional similarity.

The automated interpretation of data originating from the human genome may play a crucial role in cancer treatment. The completion of the genome of several model organisms represents a fascinating opportunity to explore important normal and abnormal biological phenomena. DNA microarrays provide a systematic and comprehensive framework to achieve this goal. This method generates data that reflects the logic and structure of the genetic program. Thus, the automation of the genome expression analysis process should significantly contribute to the understanding of gene function and the development of better cancer diagnosis strategies [4].. The clustering method of data mining is used for automatic unsupervised identification of genes with similar characteristics.

A number of authors have applied hierarchical clustering to organize genes into dendrograms based on their expression patterns [17,14,12] for instance, implement a clustering method based on pairwise average-linkage cluster analysis. One of the main disadvantages of this type of clustering techniques is that the identification of categories and informational associations is left to the observer. Additionally, the complexity of the cluster visualization task can be directly proportional to the number of elements to be grouped. Recently, a number of expression analysis techniques have been based on more advanced data mining approaches.

[30] for instance, describe a method for monitoring gene expression, in which differential expression is demonstrated by a simultaneous two-colour hybridisation scheme . Microarrays prepared by high-speed robotic printing of complementary DNAs on glass were used for quantitative expression measurements of the corresponding genes. Because of the small format and high density of the arrays, hybridization volumes of 2 micro liters could be used that enabled detection of rare transcripts in probe mixtures derived from 2 micrograms of total cellular messenger RNA. Differential expression measurements of 45 Arabidopsis genes were made by means of simultaneous, two-color fluorescence hybridization.

[33] have presented a framework for the unsupervised analysis of gene expression data. They developed an interrelated two-way clustering method which they applied on the gene expression matrices transformed from the raw microarray data. This approach detects significant patterns within samples while dynamically selecting significant genes which

manifest the conditions of actual empirical interest. Through iterative clustering the number of genes are reduced which improves the accuracy of sample class discovery. The method was proved effective by conducting experiment with two multiple-sclerosis data sets and a leukemia data set. These experiments indicate that this appears to be a promising approach for unsupervised sample clustering on gene array data sets.

[12] described a system of cluster analysis for genome-wide expression data from DNA microarray hybridization that uses standard statistical algorithms to arrange genes according to similarity in pattern of gene expression. They displayed the graphical output, conveying the clustering and the underlying expression data simultaneously in a form intuitive for biologists.They have found in the budding yeast Saccharomyces cerevisiae that clustering gene expression data groups together efficiently genes of known similar function, and found a similar tendency in human data. Thus patterns seen in genome-wide expression experiments can be interpreted as indications of the status of cellular processes. Also, coexpression of genes of known function with poorly characterized or novel genes may provide a simple means of gaining leads to the functions of many genes for which information is not available currently.

3. Classification of Genes

The systematic classification of types of tumor is crucial to achieve advances in cancer treatment and research. It has been suggested that the specification of therapies according to tumor types differentiated by pathogenetic patterns may maximize the efficiency of the treatment and minimize toxicity on the patients [17,1]. Several limitations about the conventional classification techniques based on morphological features of the tumor have been reported in the literature [5]. Moreover, by analyzing complex patterns defined by molecular markers, it has been demonstrated that there are subtypes of *acute leukaemia*, prostate cancer and *non-Hodgkin's lymphomas* [17]. There are two useful tasks in cancer classification: prediction of classes and discovery of classes. The prediction task consists of the assignment of particular tumor samples to known types of cancer. The discovery task refers to the unsupervised identification of relevant groups of samples and the characterization of subtypes of cancer. Their research aims to implement a discovery task based on a global expression analysis approach.

One important application of gene expression analysis is to classify tissue samples according to their gene expression levels. Gene expression data are typically characterized by high dimensionality and small sample size, which makes the classification task quite challenging. [18] present a data-dependent kernel for microarray data classification. This kernel function is engineered so that the class separability of the training data is maximized. A bootstrapping-based resampling scheme is introduced to reduce the possible training bias. The effectiveness of this adaptive kernel for microarray data classification is illustrated with a k-Nearest Neighbor (KNN) classifier. Their experimental study shows that the data-dependent kernel leads to a significant improvement in the accuracy of KNN classifiers. Furthermore, this kernel-based KNN scheme has been demonstrated to be competitive to, if not better than, more sophisticated classifiers such as Support Vector Machines (SVMs) and the Uncorrelated Linear Discriminant Analysis (ULDA) for classifying gene expression data.

Classification of biomedical data faces a special challenge because of the characteristics of the data: too few data examples with too many features. How to improve the classification performance or the generalization ability of a classifier in the biomedical domain becomes one of the active research areas. One approach is to build a fusion model to combine multiple classifiers together and result in a combined classifier which can achieve a better performance than any of its composing individual classifiers. [35] proposed a SVM classifier fusion model to combine multiple SVMs by applying the knowledge of fuzzy logic and genetic algorithms. The fuzzy logic system (FLS) is constructed based on SVM accuracies and distances of data examples to SVM hyperplanes in SVM feature spaces. A genetic algorithm (GA) is used to tune the fuzzy membership functions (MFs) in the FLS and determine the optimal fuzzy fusion model. They applied the proposed model to two biomedical data: colon tumor data and ovarian cancer data. Our experiment shows that multiple SVM classifiers complement each other well in the proposed fusion model and the ensemble achieves a better, more robust and more reliable performance than individual composing SVMs.

4. Neural Networks

Machine learning techniques such as neural networks are adequate for this type of analysis for their well-known pattern recognition and data organization capabilities [28,6]. Advanced neural learning algorithms have not only improved the accuracy, reliability and efficiency of many medical pattern recognition systems, but they also show several advantages for the implementation of decision support systems in physiological genomics [4].

Tamayo et al. have illustrated the value of Kohonen's Self-Organising Feature Maps (SOFM) [20] to interpret gene expression patterns during yeast growth cycle and *haematopoietic* differentiation [32]. They identify predominant gene expression patterns in those biological processes that suggested, for instance, novel hypotheses about haematopoietic differentiation useful for the treatment of acute *promyelocytic leukaemia.* Although cancer classification has improved over the past 30 years, there has been no general approach for identifying new cancer classes (class discovery) or for assigning tumors to known classes (class prediction). [17] adopted a generic approach to cancer classification based on gene expression monitoring by DNA microarrays and applied to human acute leukemias as a test case. A class discovery procedure automatically discovered the distinction between acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) without previous knowledge of these classes. An automatically derived class predictor was able to determine the class of new leukemia cases. They applied Self-Organizing Feature Maps (SOFM), approach to the problem of molecular classification of cancer. They propose a procedure that automatically discovers the distinction between *acute myeloid leukaemia* and *acute lymphoblasic leukaemia* based on the clusters obtained after training the network with a small set of cases. They applied a two-cluster SOM to automatically group the 38 initial leukemia samples into two classes on the basis of the expression pattern of all 6817 genes The results demonstrate the feasibility of cancer classification based solely on gene expression monitoring and suggest a general strategy for discovering and predicting cancer classes for other types of cancer, independent of previous biological knowledge.

A recent effort to understand how genes contribute to disease approaches the discovery of sub-classes of diffuse large B-cell lymphoma (DLBCL) by using expression analysis [1]. It has been shown that the discovery of sub-classes in DLBCL has not been successful by relying exclusively on morphological features [1] demonstrate that the molecular profile of a tumor obtained from cDNA microarrays can indeed be interpreted as a robust and clearer picture of the tumor's biology. GCS neural networks discussed in [15] are a variation of the Kohonen's SOFM (Self-Organizing Feature Maps) [20]. GCS (Growing Cell Structures) offers several advantages for the implementation of gene expression pattern interpretation systems over both non-self organizing neural networks and the Kohonen SOFM [15,16] . The GCS model and programs applied have demonstrated to be successful in a number of medical decision support domains. The basic learning process in a GCS network consists of topology modification and weight vector adaptations [16].

[9] did a comprehensive study to investigate the capability of the probabilistic neural networks(PNN) associated with a feature selection method, a so-called signal-to-noise statistic, in the application of cancer classification. The signal-to-noise statistic, which represents the correlation with the class distinction, is used to select the marker genes and trim the dimension of data samples for the PNN. Their experimental results show that the association of the probabilistic neural network with the signal-to-noise statistic can achieve superior classification results for two types of acute leukemias and five categories of embryonal tumors of central nervous system with satisfactory computation speed. Furthermore, the signal-to-noise statistic analysis provides candidate genes for future study in understanding the disease process and the identification of potential targets for therapeutic intervention.

[4] presents a method that automatically constructs a case retrieval model from existing data based on growing cell structure artificial neural networks. Within the case-based reasoning (CBR) framework, the method is evaluated for two medical prognosis tasks, namely, colorectal cancer survival and coronary heart disease risk prognosis. The results of the experiments suggest that the proposed method is effective and robust.

[8] present a new self-organizing neural network model that has two variants. The first variant performs unsupervised learning and can be used for data visualization, clustering, and vector quantization. The main advantage over existing approaches (e.g., the Kohonen feature map) is the ability of the model to automatically find a suitable network structure and size. This is achieved through a controlled growth process that also includes occasional removal of units. The second variant of the model is a supervised learning method that results from the combination of the above-mentioned self-organizing network with the radial basis function (RBF) approach. In this model it is possible—in contrast to earlier approaches—to perform the positioning of the RBF units and the supervised training of the weights in parallel. Therefore, the current classification error can be used to determine where to insert new RBF units. This leads to small networks that generalize very well. Results on the two-spirals benchmark and a vowel classification problem are presented that are better than any results previously published.

5. Microarray Data

Microarray technology permits monitoring of the expression levels of thousands of genes simultaneously. A few previous studies have shown promising results for outcome prediction using gene expression profiles for certain diseases [29,7,34,21,13]. This kind of analysis provides techniques to predict disease progression and clinical outcome at the molecular level. It also identifies genes which are responsible for non- survival of patients. Carefully verifying and understanding these genes would lead to innovative therapies and may also generate opportunities for drug discovery. Various approaches have recently been used on outcome prediction using gene expression profiles. It has been shown that specific patterns of gene expression occur during different biological states such as cell development and during normal physiological responses in tissues and cells.

Generally speaking the expression of genes provides a measure of how active a specific gene is under certain biochemical conditions. This level of expression is related to the relative concentration of messenger RNA(mRNA) which encodes the gene under considerations. The generation of quantitative expression patterns of thousands of genes can be achieved by using techniques based on complementary DNA (cDNA) microarrays [1].

Various approaches have recently been used on outcome prediction using gene expression profiles. In the Cox proportional hazard regression method [11,23] genes most related to survival are first identified by a univariate Cox analysis, and a risk score is then defined as a linear weighted combination of the expression values of the identified genes [7,29]. In [2], gene expression profiles are fed to a Fuzzy Neural Network (FNN) system to predict survival of patients. They first predict the outcome of each patient using one gene at one time. Then they rank each gene by their accuracy. Next, one by one, they use the ten highest ranked genes and the selected partner genes for prediction. Finally, the formed ten FNN models using combinatorial genes are optimized by the back-propagation method.

In [26], gene expression data are linked to patient survival times using the partial least squares regression technique which is a compromise between principal component analysis and ordinary least squares regression. In [31], the weighted voting algorithm is used to identify cured vs fatal for outcome of diffuse large B-cell Lymphoma. The algorithm calculates the weighted combination of selected informative marker genes to make a class distinction. In [21] they develop a gene index method to investigate genes that jointly relate to patient outcome and to a specific “reference gene” of interest. The paper written by [22] presented a methodology for the identification and filtration of genes.

Microarray technologies have recently shown that expression signals of genes can divide heterogeneous DLBCL in terms of chemotherapy response [1]. These microarray data were also useful in disease diagnosis and prognosis. Most approaches to the computational analysis of gene expression data are functionally significant classification of genes in unsupervised fashion and the discrimination of high risk patients from low risk ones. On the other hand, supervised learning techniques use training set to optimize the discrimination model. Artificial Neural Network (ANN) is one of supervised methods and a powerful tool for accurately detecting causal relationships [19]. Fuzzy Neural Network (FNN) is one of the advanced ANN models. They have used the FNN model for microarray analysis, and proved that the model was precise and simple prediction method of survival of patients. In previous study, the FNN model identified only four from 5,857 genes for prognostication of DLBCL patients. The accuracy of this predictor was 93% [3]. However, the number of patients were quite small. In [2] they constructed the various FNN models for outcome prediction of the larger number of DLBCL patients. By using these models, FNN was advanced for more precise and optimal and rigidly. The best FNN model showed the high accuracy of 72%. The prediction accuracy increased with increase of number of FNN model using for final decision. Thirteen combinations achieved the highest prediction accuracy of 90%. Kaplan-Meier survival analyses indicated that the patients predicted alive by these FNN models showed significant longer survival than these predicted dead. ($p<0.0001$) These selected combinations consisted of 63 genes including HLA-DRa and AA805575. HLA-DRa and AA805575 are overlapping with the sixteen genes selected by [29]. This result indicated that FNN extracted these genes as significant biological markers affecting prognosis. These combinations might implicate crucial pathways of DLBCL tumor genesis. FNN is the powerful tool for not only predicting precise outcome but also selecting genes from a huge number of genes.

[36] investigate the use of transformation models in microarray survival data . The transformation model, which can be viewed as a generalization of proportional hazards model and the proportional odds model, is more robust than the proportional hazards model, because it is not susceptible to erroneous results for cases when the assumption of proportional hazards is violated. In this they have analyzed a gene expression dataset from [7] and shown that the transformation model provides higher prediction precision than the proportional hazards model.

[25] introduce a new method of functionally classifying genes using gene expression data from DNA microarray hybridization experiments. They used support vector machines in their methods. They tested several SVMs that use different similarity metrics, as well as some other supervised learning methods, and found that the SVMs best identify sets of genes with a common function using expression data. Finally, they used SVMs to predict functional roles for uncharacterized yeast ORFs based on their expression data.

Histopathology is insufficient to predict disease progression and clinical outcome in lung adenocarcinoma. [7] show that gene-expression profiles based on microarray analysis can be used to predict patient survival in early-stage lung adenocarcinomas. Genes most related to survival were identified with univariate Cox analysis. Using either two equivalent but independent training and testing sets, or 'leave-one-out' cross-validation analysis with all tumors, a risk index based on the top 50 genes identified low-risk and high-risk stage I lung adenocarcinomas, which differed significantly with respect to survival. This risk index was then validated using an independent sample of lung adenocarcinomas that predicted high- and low-risk groups. This index included genes not previously associated with survival. The identification of a set of genes that predict survival in early-stage lung adenocarcinoma allows delineation of a high-risk group that may benefit from adjuvant therapy.

Advances in techniques for high throughput data gathering such as microarrays and DNA sequencing machines have opened up new research avenues in genomics. Large-scale biological research such as genome projects are now producing enormous quantities of genomic data using these rapidly growing technologies. Transforming the massive data to useful biological knowledge is the present challenge. Different analysis tools are being developed in order to detect and understand the phenomena of gene regulation and physiological functions and assessing the quality of a genomic sequence. Fuzzy systems are suitable for uncertain or approximate reasoning when systems are difficult to describe with a mathematical model. They allow problem solving and decision making with incomplete or uncertain information. This unique feature makes them an ideal tool for analyzing complex genomic data. [27] presents application of fuzzy systems in (1) developing a confidence measure to assess the accuracy of bases called by a DNA basecalling algorithm, and (2) building a gene interaction model that identifies triplets of activators, repressors, and targets in gene expression data. It is shown that applying appropriate fuzzy conjunction and aggregation rule increases the resilience of the fuzzy gene interaction model to noise.

Broad availability of molecular sequence data allows construction of phylogenetic trees with 1000s or even 10 000s of taxa. [24] reviews methodological, technological and empirical issues raised in phylogenetic inference at this scale. Before phylogenetic analysis, data must be generated de novo or extracted from existing databases, compiled into blocks of homologous data with controlled properties, aligned, examined for the presence of gene duplications or other kinds of complicating factors, and finally, combined with other evidence via supermatrix or supertree approaches. After phylogenetic analysis, confidence assessments are usually reported, along with other kinds of annotations, such as clade names, or annotations requiring additional inference procedures, such as trait evolution or divergence time estimates. Prospects for partial automation of large-tree construction are also discussed, as well as risks associated with outsourcing phylogenetic inference beyond the systematics community.

References

[1] A.A.Alizadeh et al., ‘Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling’, *Nature*, (February, 2000),vol. 403, pp.503-511

[2] Ando, T., & Katayama, M., 'Selection of causal gene sets from transcriptional profiling by FNN modeling and prediction of lymphoma outcome', *In 13th Intl.Conf. Genome Informatics*, 2002, pp. 278–279.

[3] T.Ando, M.Suguro, T.Hanai, T.Kobayashi, H Honda, Seto, "Fuzzy Neural Network applied to gene expression profiling for prognosis of diffuse large B-cell lymphoma", *Jpn.J.Cancer*, 2002.

[4] Azuaje, F.; Dubitzky, W.; Black, N.; Adamson, K. ,' Discovering relevance knowledge in data: a growing cell structures approach' , *Systems, Man, and Cybernetics, Part B: Cybernetics,2000, IEEE Transactions* on Volume:30 Issue:3 On page(s): 448 - 460

[5] F.Azuaje. 'Interpretation of genome expression patterns: computational challenges and opportunities', *IEEE Engineering in Medicine and Biology*, November 2000b.

[6] F.Azuaje, W.Dubitzky, P.Lopes, N.Black, K.Adamson, X.Wu, and J.White., 'Predicting Coronary Disease risk Based on Short-term RR Intervals Measurements: A Neural Network Approach', *Artificial Intelligence in Medicine*, 1999, 15, 275-298.

[7] Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, Taylor JM, Iannettoni MD, Orringer MB, Hanash S.' Gene-expression profiles predict survival of patients with lung adenocarcinoma' *Nature Medicine* 8, 816 - 824 (2002)

[8] Bernd Fritzke, "Growing cell structures—A self-organizing network for unsupervised and supervised learning", *International Computer Science of Institute*,1993, TR 93 026

[9] Chenn-Jung Huang and Wei-Chen Liao, 'Application of Probabilistic Neural Networks to the Class Prediction of Leukemia and Embryonal Tumor of Central Nervous System ', *Neural Processing Letters* , 2004,Volume 19, Number 3, 211-226,

[10] R.J.Cho, M.J. Campbell, E.Winzeler, L.Steinmetz, A.Conway, L.Wodicka, T.G.Wolfsberg, A.E.Gabrielian, D.Landsman, D.J.Lockhart, and R.W.Davis., 'A genome-wide transcriptional analysis of the mitotic cell cycle', *Molecular Cell*, 1998, vol. 12, pp. 65-73

[11] Cox, D.R., 'Regression models and life-tables (with discussion)', *J. R. Stat. Soc*., 1972 B34:187–220

[12] M.B.Eisen, P.T.Spellman, P.O.Brown, and D.Botstein., 'Cluster analysis and display of genome-wide expression patterns,' *Proceeding of the National Academy of Sciences*, 1998, vol.95, pp. 14863-14868

[13] Fayyad, U. & Irani, K., 'Multi-interval discretization of continuous-valued attributes for classification learning', *In 13th Intl. Joint Conf. Artificial Intelligence*,1993, pp. 1022–1029.

[14] W.M.Fitch and E.Margoliash., 'Construction of phylogenetic trees', *Science*, 1967, 155, 279-284.

[15] B.Fritzke., 'Growing Self-Organizing Networks- Why?', in ESANN' 96: *European Symposium on Artificial Neural Networks*, 1996, pp. 61-72.

[16] B.Fritzke., 'Growing Cell Structures- A Self-Organizing Networks for Unsupervised and Supervised Learning', *Neural Networks*,1994, vol. 7, pp. 1441-1460.

[17] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, et al., 'Molecular classification of cancer: class discovery and class prediction by gene expression monitoring', *Science*, 1999, 286, pp. 531-537

[18] Huilin Xiong, Ya Zhang, Xue-Wen Chen, "Data-Dependent Kernel Machines for Microarray Data Classification", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Volume 4 Issue 4, October 2007, *IEEE Computer Society Press Los Alamitos*

[19] Khan, J. et al., 'Classification and diagnostic prediction of cancers using gene expression profiling and artificial networks', *Nat. Med*., 2001, 7:673–679.

[20] T.Kohonen., 'Self-Organizing Maps', *Heidelberg, Springer*,1995.

[21] LeBlanc, M. et al., (2003), 'Directed indices for exploring gene expression data', *Bioinformatics*,2003, 19(6):686–693.

[22] H. Liu, Jinyan Li, LimsoonWong., 'Selection of patient samples and genes for outcome prediction', *IEEE Computational Systems Bioinformatics Conference* (CSB'04), 2004, pp. 382-392.

[23] Lunn, M., & McNeil, D.R., 'Applying Cox Regression to Competing Risks', *Biometrics*, 1995, 51:pp. 524–532.

[24] Michael J Sanderson ,'Construction and annotation of large phylogenetic trees', *Australian Systematic Botany* (2007),Volume: 20, Issue: 5, Pages: 287-301

[25] Michael P. S. Brown, William Noble Grundy, David Lin, Nello Cristianini, Charles Sugnet, Terrence S. Furey, Manuel Ares, Jr., David Haussler, 'Knowledge-based Analysis of Microarray Gene Expression Data Using Support Vector Machines', *Proceedings of the National Academy of Sciences*,2000, 97(1):262-267

[26] Park, P.J., Tian, L., & Kohane, S., 'Linking gene expression data with patient survial times using partial least squares', *Bioinformatics*, 2002, 18(Suppl 1):S120–S127.

[27] H. Ressom, P. Natarajan, R.S. Varghese and M.T. Musavi, "Applications of fuzzy logic in genomics ', *Fuzzy Sets and Systems*, 16 May 2005, Volume 152, Issue 1, Pages 125-138

[28] B.Ripley., 'Pattern Recognition and Neural Networks', Cambridge, England, *Cambridge University Press*, 1996

[29] Rosenwald, A. et al., 'The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma', *NEJM*, 2002, 346(25):1937–1947

[30] Schena M, Shalon D, Davis RW, Brown PO., 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray', *Science*., 1995, Oct 20;270(5235):467-70.

[31] Shipp, M.A. et al., 'Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning', *Nat. Med*., 2002, 8(1):68–74.

[32] P.Tamayo, D.Slonim, J.Mesirov, Q.Zhu, S.Kitareewan, E.Dmitrovsky, E.S.Lander, and T.R.Golub., 'Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation', *Proc Natl Acad Sci U S A*, 1999 vol. 96, 2907-2912.

[33] C. Tang and A. Zhang, 'Interrelated Two-Way Clustering and its Application on Gene Expression Data', presented at *International Journal on Artificial Intelligence Tools*, 2005, pp.577-598.

[34] Van de Vijver, M.J. et al., 'A gene-expression signature as a predictor of survival in breast cancer', NEJM, 2002, 347(25):1999–2009

[35] Xiujuan Chen, Yong Li, Robert Harrison, Yan-Qing Zhang, **'**Genetic fuzzy classification fusion of multiple SVMs for biomedical data', *Journal of Intelligent & Fuzzy Systems*. Volume 18, issue 6, December 2007, IOS Press Amsterdam.

[36] Xu J, Yang Y, Ott J, 'Survival analysis of microarray expression data by transformation models', *Comput Biol Chem*. 2005 Apr;29(2):91-4.

Author

Dr.R.Radha, aged 42, working as assistant Professor in Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, Chromepet, Chennai, Tamil Nadu, India. She has 18 years of teaching experience. She did her research in *Fuzzy Based Data Mining For Effective Decision Support In Bio-Medical Applications.* Her research interest is in bio informatics. She has published 5 international and 1 national publications. She has presented papers in 1 national and 1 State level conferences.