

# Multi Attribute Gene Manipulation for Cancer

<sup>1</sup>D.Priyadharshini, <sup>2</sup>Mr.J. MannarMannan

<sup>1,2</sup>Information Technology, Anna University Regional Campus Coimbatore, India

---

**Abstract:** In the present decade, the growth of medical fields with the help of computers opened the new beginning for medical research. In medical research, data mining techniques plays a vital role in predict disease earlier and diagnose. Cancer research is one of the research areas in the medical field. Accurate prediction of different cancer types has great value in providing better treatment and cure the disease all well as possible. The systematic approach is to analyze global gene expression which provides a accurate solution for the identified problem area. In this paper, the input from multiple sources, create a store. Cluster the data with attribute match association rule and followed by classification with the knowledge acquired.

**Keywords:** Gene manipulation, gene expression, cDNA.

---

## 1. INTRODUCTION

Cancer research-One of the major research areas in the field of medical. Pointed out to the exact tumour types provides a best solution for the better treatment and toxicity minimization due to medicines on the patients.

Cancer tissues can develop in any organ in our human body, such as the lung, colon, breast, skin, bones, or nerve tissue. Various types of cancer have been identified such as, breast cancer, colon cancer, lung cancer, brain cancer, cervical cancer, kidney cancer, liver cancer, leukemia, Hodgkin's lymphoma, non-Hodgkin's lymphoma, ovarian cancer, skin cancer and thyroid cancer To analyse clear picture of cancerous cells, a cancer classification analysis system needs to analyse global gene expression which provides a best solution for the identified problem area to cure the cancer.

Molecular diagnostics provided a promising option of cancer classification, but these tests are not applied because characteristic of molecular markers for solid tumours have yet to be identified. Recently, DNA based gene expression profiles and various methods have been used for cancer diagnosis. Here we follow the data mining techniques for cancer diagnosis.

This paper is correlated as follows: Section 2 gives the existing techniques (survey) of predicting the cancer and overview. Section 3 explains proposed model predicting cancer using XXX. Section 4 explains about the implementation and result. Section 5 gives conclusion of the paper including future work.

## 2. RELATEDWORK

Cancer prediction from gene sequence analysis is complicated task. Shauryajauhari[1] et al., reviews from traditional approaches to current advancement in analysis of gene expression data from cancer datasets. In their work, they suggest several machine learning techniques and data mining approaches for identifying the gene expression data to predict cancer.

Similarly Sandrine Dudoit [2] et al., classified the tumours for diagnosis and treatment of cancer. In their work, they used cDNA microarrays for monitoring the gene expression level in cells. It monitors the complete variation in tumours and classes in gene expression data.

Alternate method for gene expression analysis is proposed by G.M.Elizabeth [3] et al., used the method for converting the

large expression into small which requires computerized data analysis approaches. EShally [4] et al., also used the microarrays for monitoring the gene expression. It includes many stages and reviews microarray technology with greater emphasis on cluster analysis methods.

Yoshinor Murata [5] et al. used the effects of taking advantage of DNA microarrays and bioinformatics for clustering the gene. They suggested that DMSO induces glucose depletion by new energy synthesis via metabolic pathway.

Sared Tawazoie[6] et al. used the recapitulate the structure from networks without the knowledge of their biological molecules. It identifies the new component of networks and characterized their structure.

Nicolas Pasquier [7] et al., focus on analysis the gene activity in a cell at given time. It particularly interested the way that knowledge can be combined using various phases of the datamining process.

N.Revathy [8] et al., deals with microarray data that can predict the rank when large database is used. They used the classifying technique called Support Vector machine (SVM) which can shows the better accuracy of classification.

Yang su [9] et al., ranked the gene for integrate into one system on popular ranking criteria. He ranked the gene by expression analysis and its feature selection.

CTang[10] et al., used DNA microarray technique for monitoring thousands of gene expression. It deals with the outliers whether it's robust. In this they use the random index and sum within the cluster.

Lipo Wang [11] et al., describes the accurate cancer classification using the expressions only for few genes using SVM. In this they use both SVM and FNN.

### 3. MULTI ATTRIBUTE GENE MANIPULATION

To analysis gene and converting the expression for the gene to predict cancer is the proposed concept of the paper. To Control tumour cells from Gene to Functional Protein & Phenotype has also been analyzed in order to identify the cancer cells. In this proposed method, the expert's documental DNA data methylation is analyzed using data mining techniques. It is (Gene expression segments) a kind of binding site for proteins which make DNA inaccessible in state. Semantic based Mining - gene expression analysis tends to compare the gene expression values by using the comparative knowledge consolidator. Multi Attribute Clustering Algorithm has been used along with Semantic characteristic of sequences to find the Best Rule for Classification in the gene expression to find the Final Prediction of cancer disease. The semantic analysis is done using Ontology developed by experts to distinguish cancer cells with normal cells.

#### A. System Architecture:

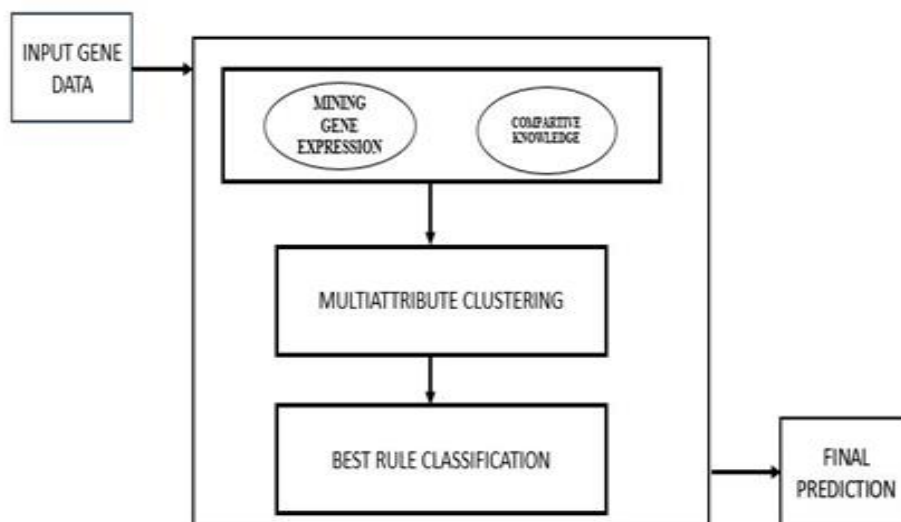


Figure 1: System architecture

The figure 1 shows the system architecture diagram of the proposed system describes that the input gene data are collected from various source and form a store its used for Compare the gene expression values by using the comparative Knowledge Consolidator Multi Attribute Clustering has been used to find the Best Rule classification in the gene expression to find the Final Prediction of cancer disease.

#### 4. IMPLEMENTATION AND RESULTS

The modules are:

- A. Gene knowledge extraction.
- B. Comparative mapping.
- C. Rule classification.
- D. Building association.
- E. Disease prediction.

##### A. Gene knowledge extraction:

Based on the analysis from the data set input comparison. They will analyse the extracted gene to predict its characteristics and other features. Gene knowledge extraction mainly focuses towards the performance of the individual gene expression data. In this they can analysis the performance of each gene in the particular dataset.

##### B. Comparative mapping:

In this it analysis the gene expressions from the data analysis and comparing the gene among them. It also result the in multi attribute clustering. If the cancerous data gene belong to the same category. It compares the attributes of different genes and forms a separate clusters. In this we follows the EM Clustering algorithm which can be represented as:

$$L(\theta;x) = p(X|\theta) = \sum p(X,Z|\theta)$$

##### C. Rule classification:

Classification is the process of finding the accuracy in the dataset. In this it classifies the gene data by follows the best rule in the particular system. It can follow the classification algorithm to find the cancerous data. In this we use decision table algorithm for classification y which can be define as

$$p(c|x) = \frac{p(c)p(x|c)}{p(x)}$$

##### D. Building association

It is a method of discovering relations among the variables in the system. For example:

T1: {genedata 7, genedata 34, genedata 55}

T2: {genedata 8, genedata 6} .....

... Tn: {genedata 4, genedata 9}

Finally,

{gene data 4, gene data 8} => {lung cancer}.

##### E. Disease prediction

Based on the analysis of the Comparative Knowledge and semantic clustering rules. They have been predicting the cancerous genes under the best rule classification constrain followed by association rule. They can suggest the final prediction of cancer as well as the stages of the cancer.

#### 4.1 PERFORMANCE EVALUATION:



**Figure 2: Visualizing the gene data**

The Figure 2 shows the visualizing the gene data which results in difference between the cancerous data as well as its gene type. In which they can found difference between the genes. Each individual gene should contain its own specification by which the system should be interrelated. If the two or more genes combine to form an association rule then we can easily predict the cancerous data.

## 5. CONCLUSION

In this predicting Cancer by analyzing gene and converting the gene expression is the proposed concept, which leads to identifying and analyzing the cancer result set. Mining Gene Expression analysis tends to compare the gene expression values by using the comparative Knowledge of the cancerous data. Multi Attribute Clustering Algorithm has been used to find the best rule for Classify the gene expression to find the Final Prediction of cancer disease.

## REFERENCES

- [1] Shaurya Jauhari and S.A.M. Rizvi "Mining Gene Expression Data *Focusing Cancer Therapeutics: A Digest*" Ieee/ Acm transactions On Computational Biology And Bioinformatics, Vol. 11, No. 3, May/June 2014.
- [2] S. Dudoit, J. Fridlyand, and T. P. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," J. Amer. Statist. Assoc., vol. 97, no. 457, pp. 77–87, Mar.2002.
- [3] G.M. Elizabeth and P. Giovanni, (2004, Dec.). "Clustering and classification methods for gene expression data analysis." Johns Hopkins Univ., Dept. of Biostatist. Working Papers. Working Paper 70.[Online].Available: <http://biostats.bepress.com/jhubiostat/paper70>.
- [4] E. Shay, (2003, Jan.). "Microarray cluster analysis and applications"[Online]. Available:<http://www.science.co.il/enuka/Essays/Microarray-Review.pdf>.

- [5] M. B. Eisen, T. P. Spellman, P. O. Brown, and D. Botstein, "Clusteranalysis and display of genome-wide expression patterns," Proc.Nat. Acad. Sci. USA, vol. 95, no. 25, pp. 14863–14868, Dec. 1998.
- [6] S. Tavazoie, D. Hughes, M. J. Campbell, R. J. Cho, and G. M.Church, "Systematic determination of genetic network architecture"nature genetics,vol.22,pp281-285,1997
- [7] N. Pasquier, C. Pasquier, L. Brisson, and M. Collard, (2008)."Mining gene expression data using domain knowledge," Softw.Informat, vol. 2, no. 2, pp. 215–231, [Online]Available:<http://www.ijsi.org/1673-7288/2/215.pdf>.
- [8] N. Revathy and R. Amalraj, "Accurate cancer classification usingexpressions of very few genes," Int. J. Comput. Appl., vol. 14, no. 4,pp. 19–22, Jan. 2011.
- [9] Y. Su, T. M. Murali, V. Pavlovic, M. Schaffer, and S. Kasif, (2003)"RankGene: Identification of diagnostic genes based on expression data," Bioinformatics, vol. 19, no. 12, pp. 1578–1579, [Online] Avaialble: <http://bioinformatics.oxfordjournals.org/content/19/12/1578.full.pdf>.
- [10] D. Jiang, C. Tang, and A. Zhang, "Cluster analysis for gene expression data:A survey," IEEE Trans. Knowl. Data Eng., vol. 16,no. 11, pp. 1370–1386, Nov. 2004.
- [11] D. R Rhodes, T. R. Barrette, M. A. Rubin, D. Ghosh, and A.M. Chinnaiyan, "Meta-analysis of microarrays: Interstudy validation of gene expression profiles reveals pathway sysregulation in prostate cancer"Cancer Res.,vol.62,pp.4427-4433. Data and Statistics. World Health Organization, Geneva, Switzerland,2006.