International Journal of Computer Engineering and Information Technology

VOL. 8, NO. 4, April 2016, 68–70
Available online at: www.ijceit.org
E-ISSN 2412-8856 (Online)

Cancer Detection Using Gene Expression

Bhagyashri Dhayagude[1], Sarika Karanjkar[2],Snehal Kamble[3] and Kanchan Misal[4]

[1, 2, 3, 4] Computer engineering Department, SAVITRIBAI PHULE PUNE UNIVERSITY, Baramati, Maharashtra, India

[1]bhagedhri64dhaygude@gmail.com, [2]sarikakaranjkar3@gmail.com, [3]kamble.snehal23@gmail.com, [4]kanchan.misal57@gmail.com

ABSTRACT

Cancer is a major problem in medical science throughout the world. Nowadays detection of Cancer in earlier stages gives very poor results, by the lack of technologies, to get the real idea about the Cancer, reliable and precise classification of Cancer is essential for successful diagnosis and treatment of Cancer. Today DNA microarray based tumor gene expression profiles have been used for Cancer diagnosis. Problems raised this system once the testing is complete, the lab reports the results in writing to the doctor or genetic counselor. So patient gets the results during another counseling session. This may not happen until several weeks after the samples are taken. so to overcome the problem of traditional genetic testing then predicting particular Cancer and giving suggestion for that type of Cancer. Gene expression data can be clustered on both genes and samples. As a result, the clustering algorithms can be divided into two categories: gene-based clustering, sample based clustering. Supervised multi attribute clustering algorithm will be effectively work compared with others. After clustering process Best rule classification used to predict the particular type of Cancer accurately. This technique suggest the final prediction of Cancer and suggest the medicine that needs to be taken up for the Cancer diagnosis.

Keywords: *Cancer, Association rules, Classification, Clustering, Data mining, Gene expression data, Gene therapy, Epigenetics.*

1. INTRODUCTION

Cancer is a major reason for all the common mortalities and morbidities all through the world. Almost 13 percent of passing brought on are because of growth. It is a malady getting always tested by numerous famous and head analysts. Albeit, a few head ways have been reported for its clinical aversion and cure and there has been a perceptible decrease in the lives' lost, however they are not exactly satisfactory. The absence of reasonable treatment what's more, early identification is the essence of this unfriendly circumstance. It is turning out to be hard for the lasting bio-medical researchers. The development in a body is watched when the division and duplication of cells happens. At the point when the fitting division levels have been accomplished, the procedure is deactivated. In a bizarre situation be that as it may, cells proceed to imitate and structure bumps in the body, despite the fact that it begins with an irrelevant element. Growth is an irregular and wild development of cells in the body that turn threatening. This is not to be mistaken for tumors. Indeed, even a tumor is an irregular development of cells, however it can be delegated (noncancerous) favorable and dangerous, the recent one bringing about cancer. It is imperative that all growths are tumors, yet the converse is not genuine. Tumor can create in any organ on the other hand tissue, for example, the lung, colon, bosom, skin, bones, or nerve tissue. Different sorts of cancer have been distinguished to be specific, bosom tumor, colon growth, lung disease, mind cancer, cervical cancer, kidney tumor, liver growth, leukemia,Hodgkins lymphoma, non-Hodgkins lymphoma, ovarian cancer, skin cancer, thyroid cancer, uterine cancer, and testicular cancer. Cancer causes quick dissemination of cells and a cancer type can fortify and extend to another one if not treated appropriately.

2. ALGORITHMS

2.1 Support Vector Machine

The SVM learning algorithm constructs a hyperplane with maximum margin that separates the positive tuples from the negative tuples. The points that lie closest to this max-margin hyperplane are called the support vectors. The hyperplane can be defined using these points alone and the classifier only makes use of these support vectors to classify test tuples. SVMs are supervised, machine learning algorithms that classify the data into separate classes , that is by large gaps. Technically, SVMs operate by finding a hyper surface in the space of gene expression profiles, that will split the groups so that there is largest distance

between the hyper surface and the nearest of the points in the groups. More flexible implementations allow for imperfect filtering of groups.

2.2 Clustering

Algorithm for K-mean Clustering:

1. The K-Means algorithm accepts the "number of clusters" to group data into and the dataset to cluster the input values.
2. The K-Means algorithm then creates the first k initial clusters from the dataset.
3. The K-Means algorithm calculates the arithmetic mean of each cluster formed in the data set. The arithmetic mean is the mean of all the individual records in the cluster.
4. Next K-Means assigns each record in the dataset to only one of the initial clusters. Each record is assigned to the nearest cluster using proximity.
5. K-Means reassigns each record in the dataset to the most similar cluster and recalculates the arithmetic mean of the clusters in the dataset.
6. K-Means reassigns each record in the dataset to only one of the new clusters Formed.
7. The preceding steps are repeated until "stable clusters".

3. FIGURES AND EQUATIONS

3.1 Figures

In Semantic ontology the process genes of similar categories are grouped together. In mining gene expression affected and unaffected genes are classified .Then Comparative Knowledge Consolidator is used initially to extract its knowledge information .Supervised Clustering algorithm are used for the informative gene selection.By using Classiffication we further classified affected gene.final prediction technique is used to predict the particular cancer.Finally suggestion given for a particular cancer.

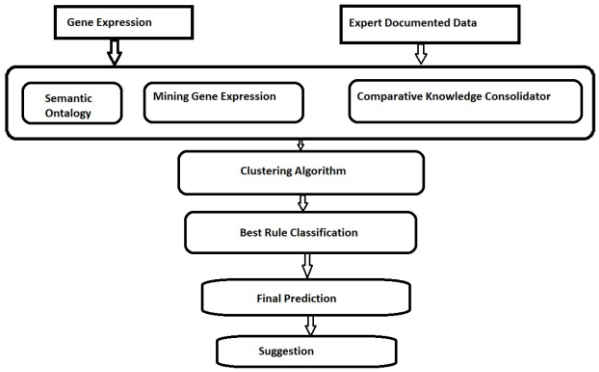

Fig. 1. Architecture diagram

3.2 Equations

1.K-mean:-Euclidean Distance

$$d=\sqrt{(x_1-y_1)^2+(x_n-y_n)^2\ldots.(x_n-y_n)^2}$$

3.3 Support Vector Machine:

Support Vector Machine used for classification.Three type of classifier used:

Linear: $x+y=c$

Nonlinear: $(x_i*y_i)^d$

RBF(Radial Basic Function): $x^2+y^2+4ac=2a$

4. CONCLUSIONS

It is watched that a dependable and exact grouping of tumors is key for effective determination and treatment of tumor. By permitting the observing of expression levels in cells for a large number of qualities all the while, microarray analyses may prompt a more finish comprehension of the sub-atomic varieties among tumors and henceforth to a better furthermore, more useful arrangement. The capacity to effectively recognize tumor classes (definitely known or yet to be found) utilizing gene expression information is an vital part of this novel way to deal with malignancy characterization. Likewise commented is that looking at the movement of qualities in a solid and carcinogenic tissue may give a few indications about the qualities that are included in malignancy.

Yet, this methodology is exceptionally constrained on the grounds that a number of the qualities serve numerous capacities and changes in gene expression can be because of components not straightforwardly concerned with the specific test. Without a doubt a micro-array information set contains various gatherings of co-communicated qualities. At that point, an average methodology for a scientist is to begin from qualities which are known not firmly identified with a natural capacity and to search a preliminary rough clustering result, to focus on a small subset of those genes which are supposed to play a role. Thus, currently biologists follow exploratory strategies by manually selecting potential groups of genes according to their knowledge.

5. ACKNOWLEDGMENTS

Its Sample acknowledgement-This project is partially supported by Grant DP123456 from the Indian Research Council and Vidya Pratishthan's College of Engineering Baramati.

REFERENCES

[1] Cluster Analysis for Gene Expression Data: A Survey Daxin Jiang Chun Tang Aidong Zhang Department of Computer Science and Engineering State University of New York at Buffalo.

[2] Accurate Cancer Classification using Expressions of Very few Genes International Journal of Computer Applications (0975 8887) Volume 14 No.4, January 2011 By N. Revathy and R. Amalraj.

[3] Comparison of discrimination methods for the classification of tumors using gene expression data. AUTHORS:S. Dudoit, J. Fridlyand, and T. P. Speed.

[4] Clustering and classification methods for gene expression data analysis AUTHORS: G.-M. Elizabeth and P. Giovanni.

[5] Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data Technical report 576, June 2000 By S. Dudoit, J. Fridlyand, and T. P. Speed

[6] Accurate cancer classification using expressions of very few genes AUTHORS: N. Revathy and R. Amalraj.