International Journal of Electronics, Communication & Soft Computing Science and Engineering
ISSN: 2277-9477, Volume 2, Issue2

Multimedia Datamining in Medical Applications

R. R. Shelke

Abstract - **Over the past decades, data mining has proved to be a successful approach for extracting hidden knowledge from huge collections of structured digital data stored in databases. From the inception, data mining was done primarily on numerical set of data. Nowadays, large multimedia data sets such as audio, speech, text, web, image, video and combination of several types are becoming increasingly available. In the present work, an attempt has been made for software development in which slide images of patient's DNA and disease affected DNA can be processed and compared. Pixel values of these images stored in the database can be also compared for result analysis.**

Keywords - **Multimedia dataminig, DNA data**

I. INTRODUCTION

With the recent advances in electronic imaging, video devices, storage, networking and computer power, the amount of multimedia has grown anormously,and data mining has become a popular way of discovering new knowledge from such a large data sets. Multimedia data refers to data such as text, numeric, images, video, audio, graphical, temporal, relational and categorical data. Multimedia data mining refers to pattern discovery, rule extraction and knowledge acquisition from multimedia database[1] Datamining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. In recent years, there has been an explosion in the rate of acquisition of biomedical data and advances in molecular genetics technologies such as DNA microarrays [2-3].The main types of data analysis needed for biomedical applications include:

* **Gene Selection** –This is a process of attribute selection, which finds the genes most strongly related to a particular class [4-7].

* **Classification** – classifying diseases or predicting outcomes based on gene expression patterns and perhaps even identifying the best treatment for given genetic signature [8-10]

* **Clustering** – finding new biological classes or refining existing ones [11-12].

It is widely believed that thousands of genes and their products in a given living organism function in a complicated and orchestrated way that creates the mystery of life. However, traditional methods in molecular biology generally work on a "one gene in one experiment" basis which means that the throughput is very limited and the "whole picture" of gene function is hard to obtain. In the past several years, a new technology, called DNA microarray, has attracted tremendous interests among biologists. This technology promises to monitor the whole genome on a single chip so that researchers can have a better picture of the interactions among thousands of genes simultaneously.

II. OBJECTIVE OF THE PRESENT WORK

Microarrays are a revolutionary new technology with great potential to provide accurate medical diagnostics which helps to find the right treatment and cure for many diseases and provide a detailed genome-wide molecular portrait of cellular states. As it can be seen from the results they are very promising and extend the possibilities of applying computational analysis and datamining to aid research in biology and medicine. Finding new insights into the molecular basis of biological processes and searching for new drugs and treatments is a problem of high complexity and where the techniques of molecular biology has been applied for many decades. The process is analogous to a large search of a few molecular entities, connections or relationships in a large sea of possibilities now days, different diseases are arising which in turn causes an increment in death rate.

Different blood testing is available for diagnosis. Though various blood testing's or any other testing's are available, diagnosis of some diseases is very difficult. Therefore, only blood tests or any other testing (eg. spinal cord fluid testing) is not sufficient. The DNA testing is one of the advanced techniques in which affected DNA & normal DNA slides are compared for disease diagnosis. With the software development, it became a very easy and beneficial task. In software development, Slide images of patient's DNA and disease affected DNA can be processed and compared. Pixel values of these images stored in the database can be also compared for ultimate result.

In the present work, datamining of DNA microarray data is performed in which, slide images of patient's DNA sample and disease affected DNA sample are processed and compared. Database of the two samples are also compared so that exact result can be given. The result shows whether the patient is suffering from particular disease or not.

III. SYSTEM ANALYSIS AND DESIGN

Clustering and Multimedia Datamining :

Clustering and Multimedia Datamining has provided the human being to more extensively use the computing power of the computers. The utility of the clustering tools is wide. One such area is biomedical research and design of expert system. In the present work, the DNA pattern verification (as DNAs hide many characteristics of human behavior) has been chosen and observed that the none of the human being have similar DNA pattern. The system being developed in the present work is the combination of image processing and data clustering techniques. The DNA sample collected acts as the input for the system and the DNA patterns matching count is the output of the system. The system first of all accepts patient's details and the associated DNA pattern with patient's details. The DNA pattern thus selected needs to be processed for the data retrieval and clustering. The processing steps can be put in the following text.

Processing Steps:

There are three steps viz…

1) Removing the color component of the image i.e. gray scale the image by scanning the DNA pattern pixel by pixel and averaging the RGB pattern.

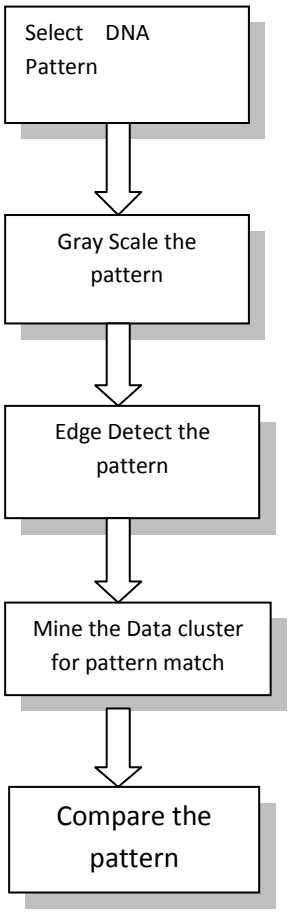
2) Detecting the edge: It is required so that one can pin point any of the defect or pattern mismatch from the ideal pattern. For the edge detection, it has been implemented the simple algorithm of post pixel mismatch which means it compares the RGB combination of the current pixel with the RGB combination of the next pixel. If the combination matches, the mark is black and if it does not match, the mark is white.

3) Real challenge of matching pattern:

Finally, it comes the real challenge of matching pattern. For this purpose, the k-mean algorithm is being used. Actually, the RGB combination of eight pixels of the ideal DNA pattern and the patient's DNA pattern is captured in the array. Then it is compared and the match count is incremented. The percentage is calculated and if only the percentage is 98% and above, the pattern match is considered.

Block diagram:

Block diagram used in the Present work is as follows:



IV. RESULT

In this work, the gray scaling, edge detection, processing, matching has been done. First, the DNA sample image of patient has been selected. The image so selected is gray scaled and edge detections have been done. After that the disease affected sample image is selected (cancer, diabetic or hepatitis). This image is then processed. In processing, gray scaling and edge detection has been done. The processed pattern is then compared with patient's image pattern. In comparison, percentage of matching is given which finally gives the result whether the patient is being suffered by particular disease or not.

V. CONCLUSIONS

Multimedia datamining and Microarrays are a revolutionary new technology with great potential to provide accurate medical diagnosis. The present work is a good sample of second generation methodologies and techniques those are being used under development today. It can be seen from the results that microarrays are very promising and extend the possibilities of applying computational analysis and multimedia datamining to aid research in biology and medicine.

VI. REFERENCES

[1] Dianhui Wang,Yong-Soo Kim,Seok Cheon,Chul Soo Lee and Yoon Kyung Han "Learning Based Neural Similarity Metrics for Multimedia Data Mining ",Soft Computing,Volume 11,Number 4,February 2007, pp.335-340

[2] Schena, M. et al,"Quantitative monitoring of gene expression patterns with a cDNA microarray." *Science 270:467-470 (1995).*

[3] DeRisi, J.L. et al.," Exploring the metabolic and genetic control of gene expression on a genomic scale." *Science 278: 680-686 (1997).*

[4] Marchal K et al.,"Comparison of different methodologies to identify differentially expressed genes in two-sample cDNA microarrays." *Journal of Biological Systems 10 (4): 409-430 Dec (2002).*

[5] Baldi P and AD Long.,"A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes." *Bioinformatics, 17: 509-519, (2001).*

[6] Li C and WH Wong.," Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application." *Genome Biology, (2001)/2/8/research/0032.*

[7] Tusher VG et al.,"Significance analysis of microarrays applied to the ionizing radiation response." *PNAS, 98:5116-5121, (2001).*

[8] Golub T. et al." Molecular classification of cancer: class discovery and class prediction by gene expression monitoring." *Science, 286:531-537, 1999. SIGKDD Explorations. Volume 5,Issue 2 - Page 4*

[9] Alizadeh L. et al.,"Identification of clinically distinct types of diffuse large B-cell lymphoma based on gene expression patterns. "*Nature 403: 503-511 (2000).*

[10] Bittner M. et al.," Molecular Classification of Cutaneous Malignant Melanoma by Gene Expression Profiling." *Nature 406: 536-540 (2000)*

[11] Eisen M. et al.," Cluster analysis and display of genome-wide expression patterns." *PNAS, 95:14863-14868 (1998).*

[12] Li H and F. Hong.,"Cluster-Rasch models for microarray gene expression data." *Genome Biology, 2(8)}: research 0031.1-0031.13, (2001).*

AUTHOR'S PROFILE

R. R. Shelke
Asstt. Professor,
Deptt. of Computer Science and Engg.
HVPM's College of Engg. Amravati
Email: rajeshrishelke@rediffmail.com