International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 8th Jan 2021. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=ISSUE-01

DOI: 10.48047/IJIEMR/V10/I01/05

Title: **FINDING THE GENE SIGNIFICANCE USING ANALYTIC TOOL WITH DATAMINING ALGORITHMS**

Volume 10, Issue 01, Pages: 28-31

Paper Authors

Dr. K. Mohan Kumar, S.Devi

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

FINDING THE GENE SIGNIFICANCE USING ANALYTIC TOOL WITH DATAMINING ALGORITHMS

Dr. K. Mohan Kumar[1], S.Devi[2]

[1]Research Supervisor & Head, PG & Research Department of Computer Science, Rajah Serfoji Government College (Autonomous), Thanjavur

[2]Research Scholar, Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

For Correspondence: #tnjmohankumar@gmail.com

Abstract:

Data mining algorithms applied in healthcare industry play a significant role in prediction and diagnosis of the diseases. This study deals with schizophrenia datasets to which data mining techniques are applied to segregate the co genes associated with the differentially expressed genes.Though, there are many studies about schizophrenia reporting the candidate genes, they are not fully accurate because of the techniques used to find out the highly responsible genes. To encounter this problem, the significant genes that play a major role in schizophrenia are to be identified.Significant gene finding is necessary to identify the disease and associated gene finding helps in determining the co-diseases, therefore enabling drug target identification.This study attempts to develop a gene predictor tool to simplify the process of significant gene finding. The Gene predictor tool involves several process to overcome the procedure for identifying the significant and associated genes.This tool is used to automate the manual process involved in finding of significant genes and associated genes. This tool can be used for any microarray data set from array express.

Key Words: Gene predictor tool, Significant gene, Associated gene

INTRODUCTION

Data mining is one of the newest analytical methods that have been used to serve medical science research and has been shown to be a valid, sensitive, and reliable method to discover patterns and relationships. The purpose of data mining is to extract useful information from large databases or data warehouses. Data mining algorithms applied in healthcare industry play a significant role in prediction and diagnosis of the diseases. There are a large number of data mining applications are found in the medical related areas such as Medical device industry, Pharmaceutical Industry and Hospital Management. To find the useful and hidden knowledge from the database is the purpose behind the application of data mining (1). Data mining techniques like classification, clustering and association rule mining can be used to analyse data and extract meaningful information. Some of the important current applications of data mining in health care includes predicting the future outcomes of diseases based on previous data collected from similar diseases, predicting the genes causing the diseases using microarray technology, predicting frequently occurring set of genes causing diseases along with another similar diseases, diagnosis of disease based on patient data, analysing treatment costs and demand of resources, pre-processing of noisy, missing data and minimizing the time to wait for the disease diagnosis (2). Depending on the nature of the data as well as desired knowledge, there are a large number of algorithms for each task. This study deals with schizophrenia datasets to which data mining techniques are applied to segregate the co genes associated with the differentially

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

expressed genes.People with a diagnosis of schizophrenia are at higher risk of suicide (Approximately 10% of general population). It can't be cured completely (3). But if treated, it may help to reduce its aggressiveness. Genes are responsible for causing any kind of disease. Though, there are many studies about schizophrenia reporting the candidate genes, they are not fully accurate because of the techniques used to find out the highly responsible genes. To encounter this problem, the significant genes that play a major role in schizophrenia are to be identified. The gene predictor tool is developed by using PHP scripting language. The reason for taking PHP is it is an open source scripting language (4).

USE OF GENE PREDICTOR TOOL

Gene predictor tool is helpful to easily identify the significant gene and associated gene from the considered dataset. Significant gene finding is necessary to identify the disease and associated gene finding helps in determining the co-diseases, therefore enabling drug target identification.The Gene predictor tool involves several process to overcome the procedure for identifying the significant and associated genes (5).This tool can be used for any microarray data set from array express.

For the case study, the Schizophrenia Dataset is taken from Array Express. Array Express contains Micro array datasets. Micro array data set contains values of gene expression which resides in a gene chip (6). The processed data is downloaded and saved as '.csv' file which is further loaded and processed.

CONCEPTUAL FRAMEWORK

The gene predictor tool model consists of the following modules as in the Figure 1. Quality checking, Data pre-processing, Significant gene finding and Associated genes finding.

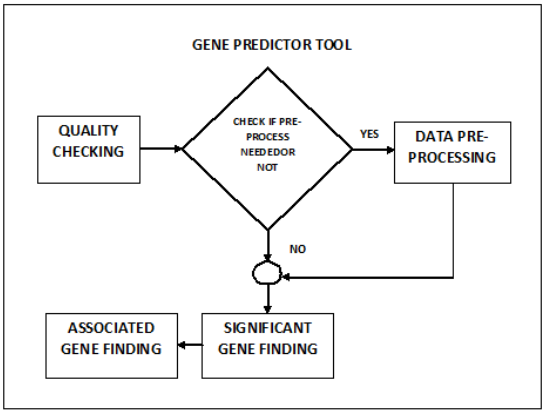


Figure 1: GENE PREDICTOR TOOL MODEL

GENE PREDICTOR TOOL

This tool is used to automate the manual process involved in finding of significant genes and associated genes. The manual process consumes more time since it involves lot of manual intervention. The following Figure 2 to 6 are the screen shots of the developed software tool at various levels of execution.

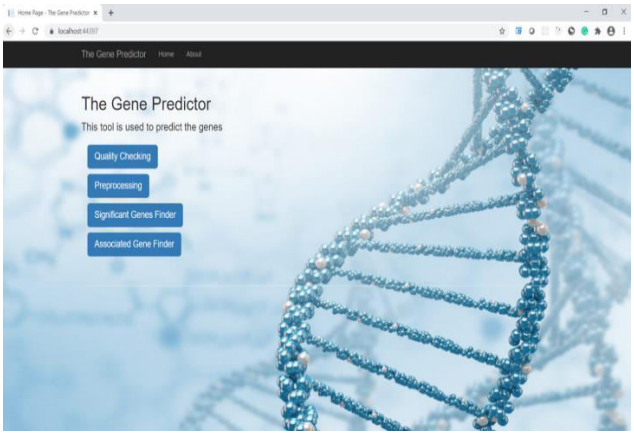


Figure 2: Home Page for Gene Predictor Tool

MODULE 1:

In this module of gene predictor tool, Quality check of the retrieved data set is performed. Figure 3 displays the functioning of Quality checking using two statistical methods namely Box Plot and Density Plot. The processing is done in PHP embedded in R.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

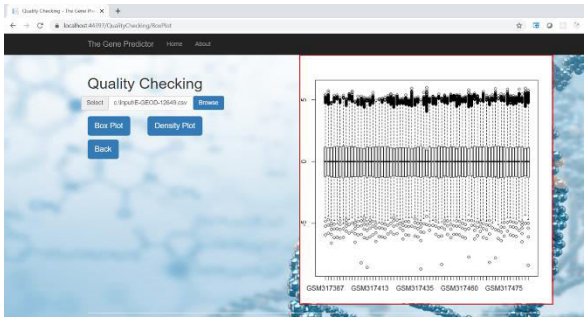


Figure 3: Quality Checking in Gene Predictor Tool

MODULE 2:

Module 2 enables the data pre-processing for clearing the irrelevant data from the given input data set and skewing the larger data set into smaller data set involving Log Transformation and Z-Score Normalization. The following figure 4 depicts this procedure.

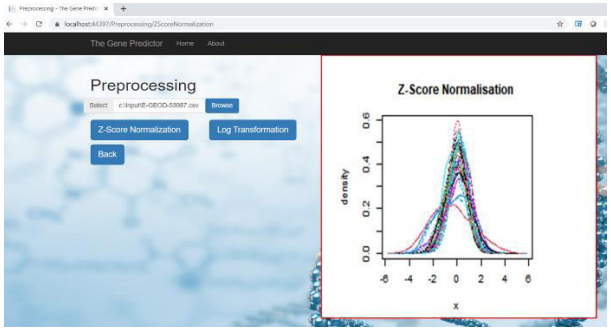


Figure 4: Pre-Processing in Gene Predictor Tool

MODULE 3:

This module represents the Significant Gene Finding. Here any microarray data sets can be feeded for analysing the significance of the genes which is done by utilizing T-test . T-test gives the significance genes as output as shown in the figure 5.

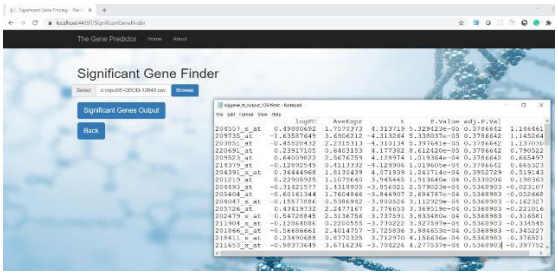


Figure 5: Significant Gene Finding in Gene Predictor Tool

MODULE 4:

Module 4, the associated genes or co-genes are found by applying 'Apriori' algorithm implied in PHP with R. The sample result of Association Rule Mining is displayed in the below Figure 6.

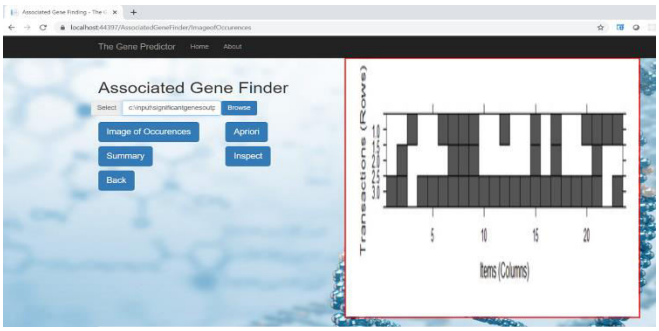


Figure 6: Associated Gene Finding in Gene Predictor Tool

COMPARISION OF EXISTING AND DEVELOPED TOOL

In this field of Data Mining in pharmacy and medical, many researchers have developed tools for gene prediction which involves each and every process performed individually. For finding of significant genes and associated genes several steps are involved. In the existing tools these steps are done separately thereby it takes more time to complete the process. The developed tool is coded by assembling these several steps at a time using PHP scripting with R. Therefore it takes only few minutes to complete the process.The following Table 1 clearly depicts the performance of the developed gene predictor software tool with existing tools. The average time taken for processing the gene finding procedure with the help of existing tool and the developed tool is compared and exhibited in the below table.

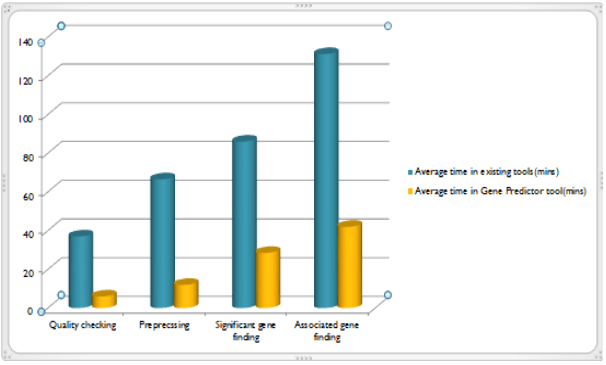
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

TABLE 1: TIME CONSUMPTION FOR EXISTING TOOLS VS GENE PREDICTOR TOOL

S.NO.	PROCESS	EXISTING GENE PREDICTION TOOL	TIME TAKING FOR PROCESSING IN MINS	AVERAGE TIME FOR EXISTING TOOLS IN MINS	AVERAGE TIME FOR GENE PREDICTOR TOOL IN MINS
1	QUALITY CHECKING	FATIGO	27	37	6
		EASE ONLINE	23		
		GO TOOL BOX	33		
		GO START	29		
2	PRE-PROCESSING	FATIGO	73	66.5	12
		EASE ONLINE	60		
		GO TOOL BOX	69		
		GO START	64		
3	SIGNIFICANT GENES FINDING PER DATASET	FATIGO	88	86	28.5
		EASE ONLINE	92		
		GO TOOL BOX	78		
		GO START	86		
4	ASSOCIATED GENES FINDING	FATIGO	121	131.25	42
		EASE ONLINE	112		
		GO TOOL BOX	144		
		GO START	148		

The following Figure 7 shows the graphical representation of the above Table 1.

Figure 7: TIME CONSUMPTION FOR EXISTING TOOLS VS GENE PREDICTOR TOOL



CONCLUSION

This study mainly focuses on finding significant genes and co-genes causing schizophrenia by using datamining techniques with the help of gene predictor tool. The overall architecture is framed and coded using 'R' and embedded in scripting language 'PHP'.The main purpose of the study is to help the society by find the effective drugs for any diseases through significant genes and co-genes by using the frame work. The gene predictor tool can be further improved by getting the dataset automatically from any other biological databases.

REFERENCES

1. Jin Yang, Yuanjie li, Qingquingliu, Li Li and Aozi Feng, "Brief introduction of medical database and data mining technology in big data era", *Journal of Evidance based Medicine*;13:57–69, 2020.

2. R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines", *International Journal of Medical Informatics*, vol. 77, pp. 81-97, 2008.

3. Julio Licinio, "Advances in Schizophrenia Research: First Special Issue", Molecular Psychiatry, 25, pp.699-700,2020.

4. EleftheriaDroosopoulou, EleftheriaKiourtzoglou andGeorge Palaiologopoulos, " PHP Programming CookBook", *Web Code Seeks – Web Developers Research Center,*1-72, 2016.

5. Vignesh M and Balaji R, "Data analysis using Box and Whisker Plot for Lung Cancer", *International Conference on Innovations in Power and Advanced Computing Technologies*, March 2017.

6. RussulAlanni, JingyuHou, HasseebAzzawi and Yong Xiang. "Deep Gene Selection Method to Select Genes From Microarray datasets for Cancer Classification", *BMC Bioinformatics*, 20-608,November 2019.