ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

Boolean Association Rule Mining on Microarray Cancer Gene Expression Data using Gene Expression Intervals

R. Vengateshkumar, S. Alagukumar,R. Lawrance,

Research Scholar, Research & Development centre, Bharathiar University, Coimbatore, Tamil Nadu, India

Assistant Professor, Department of Computer Applications, Ayya Nadar Janaki Ammal College,

Sivakasi, Tamil Nadu, India

Director, Department of Computer Applications, Ayya Nadar Janaki Ammal College,

Sivakasi, Tamil Nadu, India

ABSTRACT*:* Data Mining is one of the interdisciplinary fields on the research area. Association rule mining plays a vital role in the data mining for finding significant relations in biological data. Microarray technology is mainly used by the researchers to find the meaningful relations among gene expression data. In this research paper, the statistical t-test has been applied to select the significant genes, k-means clustering technique has been implemented for discretize the gene expression data, Boolean Association Rule Mining (BARM) generate the frequent gene expression intervals and finally, the association rules has been discovered. Association rules discover the significant relations among microarray gene expression data. It exposes the correlation among the gene expression and used to provide the significant decision for cancer diagnosis.

KEYWORDS*:* Microarray, Gene Filtering, Clustering, Frequent Pattern Mining and Association Rule Mining.

I. INTRODUCTION

Now-a-days, huge amount of data are being collected from Biological data. Analyzing and extracting information from huge amount of data is difficult. Data mining techniques have been used to get the effective knowledge from the huge amount of data. In this paper, the proposed methodology focuses on association rule mining technique to extract interesting relationships among set of genes in the field of bioinformatics.

Microarray technologies provide the opportunity to compute the expression level of tens of thousands of genes in cells simultaneously. One interesting fact about microarray data is that the behaviors of thousand of genes can be examined at different times. Gene expression is the process of transcribing DNA sequence to MRNA sequence which is later referred to as the amino acid sequence known as protein. The number of produced versions from RNA is called gene expression level. Microarray experiments contains huge amount of data. Main challenge on microarray data is high density of data. Data collected from microarray experiments is in the form of R x C matrix of expression level, where R represents Rows (experiments) and C represents Columns (genes). Microarray contains an order of magnitude more genes than experiments. In this paper, it has been focused on microarray gene expression interval association analysis from the frequent pattern mining. Frequent pattern mining is the most important task of association rule mining. Microarray gene expression interval association analysis is exploring the biological relevant association between different genes under different experimental samples. The rest of the paper is organized as given below. The related papers are reviewed in Section 2.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

The proposed methodology of Boolean association rule mining for gene expression intervals are illustrated in Section 3. The experimental results are shown in Section 4. Conclusion of the work is discussed in Section 5.

II. RELATED WORKS

In order to do the survey various algorithms have been studied. Extracting the interesting relationships among set of genes using gene intervals and association rules, the researcher must know the basic knowledge of gene filters, discretization techniques and association algorithms.

Jeanmougin, M, *et al.* have discussed the statistical approaches to select genes differentially expressed between two groups is to apply a t-test and compared with various statistical methods to find the significant genes [1].

Garcia, S., *et al.* have made a survey on discretization techniques. Discretization is an essential preprocessing technique to transform a set of continuous attributes into discrete attributes, by associating categorical values to intervals [2].

Alves, R., *et al.* have discussed frequent pattern methods for gene association analysis. Frequent pattern mining has been applied successfully in various data such as business and scientific data for discovering interesting association patterns, and is becoming a hopeful approach in microarray gene expression analysis. However, with dense datasets such as telecommunications, microarrays, etc., where there are many long frequent patterns. Hence, these methods scale very poorly and sometimes are impractical. This drawback is due to the high computational cost used by apriori algorithm. Then they pointed out that the tree based methods such as Frequent Pattern (FP-growth) may find difficulties when dealing with high dimensional datasets [3].

Zakaria, W., *et al.* have proposed a column enumeration based algorithm using high confidence association rules for up and down expressed genes. Then they explained that the generating all frequent itemsets in dense datasets requires large memory [4].

Alagukumar, S., *et al.* have discussed the microarray data analysis using association rule mining. They compared the frequent pattern mining methods using Apriori and FP-Growth on microarray gene expression data [5].

Wur, S.Y., *et.al.* have proposed effective boolean algorithm for mining association rules in large databases. The sparse matrix approach has been given better performance over the Apriori algorithm [6].

From the literature study, it has been concluded that microarray dataset typically contain high density of data. Association rules have been proved to be useful in analyzing such datasets. However, the most existing association rule mining algorithms are unable to efficiently handle normalized microarray datasets with continuous values.

The existing association rule mining algorithms requires large memory and takes exponential time for generating frequent gene expression pattern and discovering association rules. In this paper, a new algorithm called BARM is described that is specially designed to select the significant genes, generate frequent gene expressions intervals and discover association rules from microarray gene expression data using gene intervals with less memory and low computational time.

III. METHODOLOGY

Association rule mining finds frequent item-sets whose occurrences exceed a predefined threshold in the dataset. Then it generates association rules from frequent item sets with the support and confidence. Association rule mining is applied on microarray data set to extract interesting associations among set of genes.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

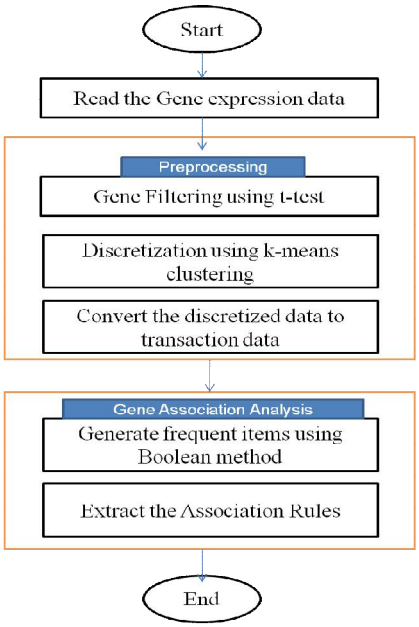

Fig.1: Block diagram of BARM

In BARM, item-sets are gene expression intervals. The aim of BARM is to extract the frequent gene expression intervals and then use them to generate association rules. Before mining, BARM selects the significant genes from microarray gene expression data, and transforms the data by converting continuous gene expression data into discretized gene expression data. Finally, the discretized data are used as transaction data for mining.

In this research paper, it has been proposed a BARM for microarray gene association analysis using frequent gene expression intervals and association rules shown in Figure 1. The BARM comprises of two phases, namely preprocessing and gene association analysis. The pseudo code for overall algorithm is illustrated in the Figure 2.

Input : Microarray gene expression data
Output : Micro array gene association rules
Begin
Step 1 : Read the gene expression data
Step 2 : Filter Significant genes using t-test
Step 3 : Discretize the gene expression data using k-means clustering
Step 4 : Convert the discretized gene intervals data to transaction data
Step 5 : Generate the frequent gene

Fig.2: BARM Algorithm

At the end of two phases the frequent patterns and significant relations among microarray gene expression intervals are extracted.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

A. PREPROCESSING

Data preprocessing is a one of the data mining technique which involves transforming unprocessed data into an understandable format. Real world data is often deficient, inconsistent. Data preprocessing is a proven method of resolving such issues. In this paper, the informative genes are selected using gene filtering and the continuous gene expression data are transformed into discrete data using discretization technique.

1. Gene Filtering using t-test

The gene filtering is the process of selecting the differentially expressed genes and statistically significant in the gene expression data using t-test method. The t-test is the most often used to analyze microarray data. The t-statistic provides a standardized estimate of differential expression based on the following formula

$$T = \frac{(\bar{A} - \bar{B})/\hat{\sigma}_p}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \qquad (1)$$

$$\hat{\sigma}_p = \frac{\sum_1^n A_i^2 - n\bar{A}^2 + \sum_1^m B_i^2 - m\bar{B}^2}{n + m - 2} \qquad (2)$$

Where $\bar{A} - \bar{B}$ represents is sample means and$\hat{\sigma}_p$is an unbiased estimator for standard deviation. First calculate the sums of squares and Correlation factor, nA^2 to subtract to give n times normal sample variance also called the sum of squared residuals, the associated probability under the null hypothesis is calculated by reference to the t-distribution with $n + m - 2$ degree of freedom. The p-value is used to determine if a number is significantly different from normal. A p-value of 0.05 or less is commonly measured statically significant. The t-statistics value will be calculated and the p-value calculated from t-distribution with n−2 degrees of freedom. Finally, the differentially expressed genes and statistically significant genes are selected or biological significance based on probability with degrees of freedom N-2 and $p < 0.05$. The uninformative genes are removed from gene expression.

B. DISCRETIZATION USING K-MEANS CLUSTERING

Data discretization is a commonly used as pre-processing method that reduces the number of distinct values for a given continuous variable by dividing its range into a finite set of disjoint intervals, and then relates these intervals with meaningful labels [7]. Subsequently, data are analyzed or reported at this higher level of knowledge representation rather than the individual values, and thus leads to the simplified data representation in data exploration and data mining process. Discretization methods can be supervised or unsupervised. Supervised methods make use of the class label when partitioning the continuous features. Unsupervised discretization methods do not require the class information to discretize continuous attributes.

The k-means clustering method is one of the popular clustering methods; k-means by MacQueen (1967) is also suitable to be used to discretize continuous valued variables because it calculates continuous distance based similarity measure to cluster data points.

K-means is a non-hierarchical partitioning clustering algorithm that operates on a set of data points and assumes that the number of clusters to be determined (k) is given. The most common distance measure used in k-means algorithm is the Euclidean distance. Initially, the algorithm assigns randomly k data points to be the centers so called centroids of the clusters. Then each data point of the given set is associated to the closest center resulting the initial distribution of the clusters. After this initial step, the next two steps are performed until the final cluster is obtained:

1. Recompute the centers of the clusters as the average of all values in each cluster.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

2. Each data point is assigned to the closest center. The clusters are formed again.

The resulting clusters or centroids are used as the states of the discretization process. The algorithm stops when there is no data point that needs to be reassigned.

$$\sum_{i=1}^{k} \sum_{y \in \text{cluster } i} d(y, C_i)^2 \qquad (3)$$

The sum of squares function over the partition of the data points into the clusters 1, 2, …, k, gets minimized where y is original data, Ci is the center of the cluster i, and d is the distance measure. After the clustering is done, the discretization cut points are defined as the minimum and maximum of the active domain of the attribute and the midpoints between the boundary points of the clusters.

B. Gene Association Analysis

The BARM algorithm finds useful patterns and rules from transaction data using boolean method. These patterns and rules are very useful for decision making. The boolean method generates the frequent microarray gene intervals without generating the candidate item sets and extracting the association rules in two steps.

Step-1: The frequent gene interval sets are identified using bitwise OR and bitwise AND operations.

Step-2: Microarray gene association rules are generated using bitwise AND and bitwise XOR operations from the frequent gene interval sets.

1. Frequent gene interval sets

Given a set of genes G= {g1, g2, g3 … gn} and a set of samples tID = {s1, s2, s3… sm}, a subset of G, S⊆G is called a frequent, if support(S) ≥ minimum support, where minimum support is a user defined threshold [8].

2. Microarray Gene Association Rules

Association Rule: Let $G = \{g_1, g_2, g_3 \dots g_n\}$ be a set of n elements called genes. A rule is defined as an implication of the form $X \rightarrow Y$, where $X, Y \subseteq G and\ X \cap Y = \emptyset$[8]. The left-hand side of the rule is named as antecedent and right-side of the rule is named as consequent.

Support: The Rule $X \rightarrow Y$ holds in the transaction set T with Support S, Where S is the percentage of samples in T that contain $X \cup Y$[8]

$$Support(X \rightarrow Y) = \frac{Support(X \cup Y)}{|T|} \qquad (4)$$

Confident: The Rule $X \rightarrow Y$ has confidence C in the transaction set T, where C is the percentage of samples in T containing X that also contain Y [8].

IV. EXPERIMENTAL RESULTS

The sample gene expression data related to breast cancer2 dataset consists of 30 samples and 16 genes are shown in table1. The sample gene expression data are filtered using statistical t-test gene filtering method. The filtered gene expression data are shown in table 2. After gene filtering, the gene expression data are transformed into gene intervals using K-Means clustering discretization method, where the data clustered into 2 distinct clusters are shown in table 3.

Finally, gene intervals are converted into transactional data where samples are represented by transactions and gene intervals are represented by item sets as shown in table 4. In microarray gene association, the frequent gene intervals sets are generated with minimum support count 50% as shown in table 5.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

From the frequent gene intervals sets, the association rules are extracted with support 50% and confidence 100% as shown in table 6. Finally the biological knowledge is extracted from the association rules. It provides gene targeting treatment decisions for cancer patients.

Table- 1 : Sample Microarray Gene Expression Data

Sample	S1	S2	S3	S4	S5	Sn
LYPD6	0.45	-1.49	-0.46	0.65	0.08	…
PTGER3	1.66	2.26	1.22	3.98	3.59	…
EST_1	-0.62	-0.68	-0.8	-0.68	-0.83	…
EST_2	-0.49	-0.38	-0.56	-0.48	-0.41	…
CHDH	0.86	1.17	0.34	1	-1.82	…
EST_3	0.64	0.48	-0.34	-0.04	-1.24	…
IL17BR	-0.7	-1.1	-2.16	-0.59	-3.56	…
SCYA4	7.13	7.11	7.05	6.51	8.58	…
IL1R2	1.37	1.65	0.53	1.44	1.1	…
ABCC11	5.96	6.55	4.35	6.82	6.02	…
HOXB13	-3.1	3.2	2.05	-3.33	1.12	…
APS	-1.45	0.68	-0.17	0.22	-0.45	…
ESTs_4	-2.57	2.3	1.11	-2.54	0.59	…
DOK2	2.04	1.69	1.77	3.09	2.24	…
EST_5	1.83	1.55	1.94	1.25	2.1	…
GUCY2D	1.99	4.25	5.49	1.56	2.59	…

Table-2: Filtered Gene Expression Data

Sample	LYPD6	EST_2	EST_3	IL17BR	IL1R2	ABCC11
S1	0.45	-0.49	0.64	-0.7	1.37	5.96
S2	-1.49	-0.38	0.48	-1.1	1.65	6.55
S3	-0.46	-0.56	-0.34	-2.16	0.53	4.35
S4	0.65	-0.48	-0.04	-0.59	1.44	6.82
S5	0.08	-0.41	-1.24	-3.56	1.1	6.02
Sn	…	…	…	…	…	…

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

Table-3: Discretized Microarray Gene Expression Data

	LYPD6	EST_2	EST_3	IL17BR	IL1R2	ABCC11
S1	[-0.29, 0.65]	-0.56, -0.45]	[-0.53, 0.65]	[-1.83, -0.59]	[0.96, 1.65]	[5.34, 6.82]
S2	[-1.49, -0.29]	[-0.45, -0.38]	[-0.53, 0.65]	[-1.83, -0.59]	[0.96, 1.65]	[5.34, 6.82]
S3	[-1.49, -0.29]	[-0.56, -0.45]	[-0.53, 0.65]	[-3.56, -1.83]	[0.53, 0.96]	[4.35, 5.34]
S4	[-0.29, 0.65]	[-0.56, -0.45]	[-0.53, 0.65]	[-1.83, -0.59]	[0.96, 1.65]	[5.34, 6.82]
S5	[-0.29, 0.65]	[-0.45, -0.38]	[-1.24 , -0.53]	[-3.56, -1.83]	[0.96, 1.65]	[5.34, 6.82]
Sn	…	…	…	…	…	…

Table-4: Transaction dataset

tID	Itemset
S1	LYPD6 [-0.29, 0.65] EST_2 [-0.56,-0.45] EST_3 [-0.53, 0.65] IL17BR [-1.83,-0.59] IL1R2 [0.96,1.65] ABCC11 [5.34,6.82]
S2	LYPD6 [-1.49,-0.29] EST_2 [-0.45,-0.38] EST_3 [-0.53, 0.65] IL17BR [-1.83,-0.59] IL1R2 [0.96,1.65] ABCC11 [5.34,6.82]
S3	LYPD6 [-1.49,-0.29] EST_2 [-0.56,-0.45] EST_3 [-0.53, 0.65] IL17BR [-3.56,-1.83] IL1R2 [0.53,0.96] ABCC11 [4.35, 5.34]
S4	LYPD6 [-0.29, 0.65] EST_2 [-0.56,-0.45] EST_3 [-0.53, 0.65] IL17BR [-1.83,-0.59] IL1R2 [0.96,1.65] ABCC11 [5.34,6.82]
S5	LYPD6 [-0.29, 0.65] EST_2 [-0.45,-0.38] EST_3 [-1.24 , -0.53] IL17BR [-3.56,-1.83] IL1R2 [0.96,1.65] ABCC11 [5.34,6.82]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

Table-5: Frequent Gene Expression Intervals set

Frequent Gene Expression Intervals set	Support Count
LYPD6[-0.291 : 0.650]	3
EST_2[-0.560 : -0.453]	3
EST_3[-0.527 : 0.640]	4
IL17BR[-1.83 : -0.59]	3
IL1R2[0.96 : 1.65]	4
ABCC11[5.34 : 6.82]	4
LYPD6[-0.291 : 0.650] IL1R2[0.96 : 1.65]	3
LYPD6[-0.291 : 0.650] ABCC11[5.34 : 6.82]	3
EST_2[-0.560 : -0.453] EST_3[-0.527 : 0.640]	3
EST_2[-0.560 : -0.453] EST_3[-0.527 : 0.640]	3
EST_3[-0.527 : 0.640] IL1R2[0.96 : 1.65]	3
EST_3[-0.527 : 0.640] ABCC11[5.34 : 6.82]	3
IL17BR[-1.83 : -0.59] IL1R2[0.96 : 1.65]	3
IL17BR[-1.83 : -0.59] ABCC11[5.34 : 6.82]	3
IL1R2[0.96 : 1.65] ABCC11[5.34 : 6.82]	4
LYPD6[-0.291 : 0.650] IL1R2[0.96 : 1.65] ABCC11[5.34 : 6.82]	3
EST_3[-0.527 : 0.640] IL17BR[-1.83 : -0.59] IL1R2[0.96 : 1.65]	3
EST_3[-0.527 : 0.640] IL17BR[-1.83 : -0.59] ABCC11[5.34 : 6.82]	3
EST_3[-0.527 : 0.640] IL1R2[0.96 : 1.65] ABCC11[5.34 : 6.82]	3
IL17BR[-1.83 : -0.59] IL1R2[0.96 : 1.65] ABCC11[5.34 : 6.82]	3
EST_3[-0.527 : 0.640] IL17BR[-1.83 : -0.59] IL1R2[0.96 : 1.65] ABCC11[5.34 : 6.82]	3
…	…

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

Table-6: Association Rules

Antecedent	Consequent	Sup.	Conf.
EST_3[-0.527 : 0.640] ABCC11[5.34 : 6.82]	IL17BR[-1.83 : -0.59] IL1R2[0.96 : 1.65]	50%	100 %
IL17BR[-1.83 : -0.59]	EST_3[-0.527 : 0.640] IL1R2[0.96 : 1.65] ABCC11[5.34 : 6.82]	50%	100 %
LYPD6[-0.291 : 0.650]	ABCC11[5.34 : 6.82]	50%	100 %
IL17BR[-1.83 : -0.59] IL1R2[0.96 : 1.65]	EST_3[-0.527 : 0.640]	50%	100 %
IL17BR[-1.83 : -0.59] IL1R2[0.96 : 1.65]	ABCC11[5.34 : 6.82]	50%	100 %
IL17BR[-1.83 : -0.59] IL1R2[0.96 : 1.65]	EST_3[-0.527 : 0.640] ABCC11[5.34 : 6.82]	50%	100 %
IL17BR[-1.83 : -0.59] ABCC11[5.34 : 6.82]	EST_3[-0.527 : 0.640]	50%	100 %
…	…	…	…

The experiments are performed on a computer with Intel Core 2 Duo CPU and 2GB of main memory. The proposed algorithm implemented in Java language with JDK1.4 version. The microarray breast cancer2 gene expression data were taken from National centre for Biotechnology Information (NCBI) [9].

A. COMPARATIVE ANALYSIS

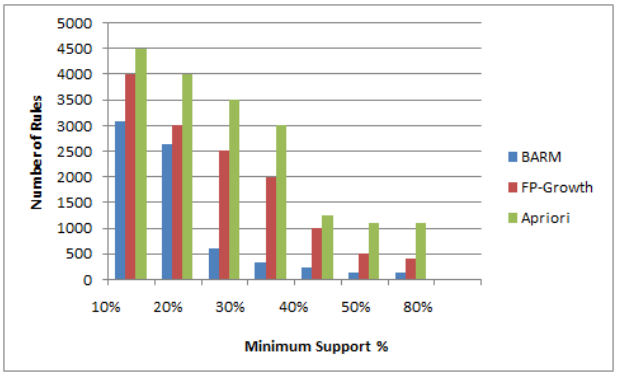

Fig.3: Comparative Analysis of Rule Generation

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 1, March 2017

The proposed association algorithm compared with classical association algorithms such as *Apriori* and *FP-Growth* algorithms. The proposed association algorithms discover the less number of rules and reduce the time complexity as well as memory to compare with bench mark algorithms. The figure 3 depicts the comparative analysis of rule generation of proposed algorithm with Apiori and FP-Growth algorithms.

V. CONCLUSION

The proposed BARM algorithm obtains frequent item set without candidate generation and scans the database only once. It reduces the time complexity and memory usage. The proposed BARM algorithm extracts significant relations among microarray genes. The experiments were carried out by using the microarray breast cancer 2 dataset. Additionally, in this paper, the algorithm has been compared with other traditional bench mark algorithms such as *Apriori*, *FP-growth*. Apriori algorithm requires large memory and takes exponential time for candidate generation.

FP-growth generates frequent item set without candidate item set generation, Hence it requires less memory and scans the database only two times. The result of the comparative analysis revealed that the BARM performed better than other methods. The result of this work can be used to reveal crucial resource for diseases and provide gene targeting treatments.

REFERENCES

[1] Jeanmougin, M, et al. "Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies." PloS one vol.5n no.9, 2010.

[2] S. Garcia, J. Luengo, J.A. Sáez, V. López, and F. Herrera, “A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning”, Knowledge and Data Engineering, IEEE Transactions, vol. 25, no.4, pp.734-750, 2013.

[3] R. Alves, B.D.S Rodriguez, and R.J.S. Aguilar, “Gene association analysis: a survey of frequent pattern mining from gene expression data”, Briefings in Bioinformatics, 2009, vol.2, no.2, pp.210-224.

[4] W. Zakaria, Y. Kotb, and F. Ghaleb, “MCR-Miner: Maximal Confident Association Rules Miner Algorithm for Up/Down-Expressed Genes”, Applied Mathematics and Information Sciences, vol.8 no.2, pp.799-809, 2014.

[5] S. Alagukumar and R. Lawrance, "A Selective Analysis of Microarray Data Using Association Rule Mining." Procedia Computer Science, no.47, pp.3-12, 2015. http://dx.doi.org/10.1016/j.procs.2015.03.177

[6] S.Y. Wur, and Y. Leu, “An Effective Boolean Algorithm for Mining Association Rules in Large Databases” Database Systems for Advanced Applications,IEEE Transactions, pp.179-186,1999.

[7] R. Dash, and R.L. Paramguru, “Comparative analysis of Supervised and Unsupervised Discretization Techniques”, International Journal of Advances in Science and Technology,vol.2,no.3,pp.29-7,2011.

[8] J. Han, and M. Kamber, “Data Mining: Concepts and Techniques”, Morgan Kaufmann Publishers, Elsevier, 2002.

[9] www.ncbi.nlm.nih.gov/geo/query/ acc.cgi?acc=GSE1379.