Future Computing and Informatics Journal

Volume 2
Issue 2 (2020) *2017, Volume 2, Issue 2*

Article 4

2017

Informative gene selection using Adaptive Analytic Hierarchy Process (A2HP)

Abhishek Bhola
Department of Computer Science and Engineering, PEC University of Technology, Chandigarh 160012, India, abhishek_bhola@hotmail.com

Shafa Mahajan
Department of Computer Science and Engineering, PEC University of Technology, Chandigarh 160012, India, 6march.shafa@gmail.com

Available online at www.sciencedirect.com

ScienceDirect

Future Computing and Informatics Journal 2 (2017) 94–102
http://www.journals.elsevier.com/future-computing-and-informatics-journal/

Informative gene selection using Adaptive Analytic Hierarchy Process (A2HP)

Abhishek Bhola, Shafa Mahajan, Shailendra Singh*

Department of Computer Science and Engineering, PEC University of Technology, Chandigarh 160012, India

Received 24 May 2017; accepted 29 July 2017
Available online 31 August 2017

Abstract

Gene expression dataset derived from microarray experiments are marked by large number of genes, which contains the gene expression values at different sample conditions/time-points. Selection of informative genes from these large datasets is an issue of major concern for various researchers and biologists. In this study, we propose a gene selection and dimensionality reduction method called Adaptive Analytic Hierarchy Process (A2HP). Traditional analytic hierarchy process is a multiple-criteria based decision analysis method whose result depends upon the expert knowledge or decision makers. It is mainly used to solve the decision problems in different fields. On the other hand, A2HP is a fused method that combines the outcomes of five individual gene selection ranking methods t-test, chi-square variance test, z-test, wilcoxon test and signal-to-noise ratio (SNR). At first, the preprocessing of gene expression dataset is done and then the reduced number of genes obtained, will be fed as input for A2HP. A2HP utilizes both quantitative and qualitative factors to select the informative genes. Results demonstrate that A2HP selects efficient number of genes as compared to the individual gene selection methods. The percentage of deduction in number of genes and time complexity are taken as the performance measure for the proposed method. And it is shown that A2HP outperforms individual gene selection methods.

Keywords: Gene selection; Microarray experiments; High dimensionality; Gene expression dataset; Adaptive Analytic Hierarchy Process (A2HP)

1. Introduction

Bioinformatics is a versatile discipline for the study of biological data which integrates computer science, engineering, mathematics and statistics. Bioinformatics is capable of transforming all faces of society. The main purpose of bioinformatics is to explore the genome of different organisms to identify genes and their encoded products that govern the biological reactions that provide fuels, food and different significant materials for health [1,2].

Human genome consists of tens of thousands of genes, genes which encodes information for the synthesis of biological molecules. They are very useful in the early diagnosis of various deadly diseases like cancer, etc. and for the distinction of various types of tumor. Study and analysis of such a large number of genes of an organism manually by scientists is a cumbersome task [3–5].

Here microarray technology came into picture. It is an imperative tool which is used by various scientists, researchers and biologists to monitor the expression level of genes of an organism [6]. A microarray gene expression dataset can be represented in the form of matrix where each row represents a specific gene whereas each column represents a sample or a time point. Each entry in the table respective to a particular gene and sample or time point is the measured expression level

* Corresponding author.
E-mail addresses: abhishek_bhola@hotmail.com (A. Bhola), 6march.shafa@gmail.com (S. Mahajan), shailendra_sing@yahoo.com (S. Singh).
Peer review under responsibility of the Faculty of Computers and Information, Future University in Egypt

http://dx.doi.org/10.1016/j.fcij.2017.07.004

of a particular gene respectively in a sample or time point [7–9].

One of the major applications of microarray experiments is gene expression profiling. In this, the expression levels of thousands of genes are analyzed and monitored simultaneously to study consequence of various treatments, diseases and development stages on gene expression [10].

Microarray experiments allow to study and analyze different gene expression datasets simultaneously. But due to the problem of high dimensional datasets (large number of genes as compared to small sample size or time point), managing these huge dataset and classification of genes becomes difficult [11].

Apart from this curse of dimensionality, other problems like missing values in gene expression datasets, noisy and redundant data exists. And the major challenge is to search an optimal way to retrieve the relevant and informative genes from these huge gene expression datasets. For this purpose, the concept of informative gene selection is highlighted. Many dimensionality reduction and gene selection methods have been applied to gene expression datasets in the past [12].

The rest of the paper is organized into following sections: Section 2 describes the background which includes the literature of various gene selection techniques, Section 3 describes the proposed methodology for identifying informative genes using A2HP, Section 4 presents the results and discussion and Section 5 gives the conclusion and future scope.

2. Background

Gene selection is the procedure which involves selection of subset of relevant genes and removal of genes with less or no predictive information [13]. The presence of large number of genes rises the dimensionality problem, causing increase in computational costs and noise [14]. Gene selection methods provide better understanding of data for the study of gene expression data for diagnosis of diseases. A lot of work has been done in this area for reducing the dimensionality and deriving a subset of good marker genes.

Gene selection methods are classified as supervised, unsupervised and semi-supervised on the basis of prior knowledge and distinguish into filter, wrapper, embedded, hybrid and ensemble depending upon the selection algorithm used along with model building [15]. Filter methods are the open loop methods which use ranking techniques for variable selection. Wrapper and embedded methods are classifier dependent methods and models feature dependencies [16]. Hybrid method combines the strengths of both filter and wrapper methods and gives higher performance results than filter and better computational complexity than wrapper method. Ensemble is the robust method which overcomes the instability of other gene selection methods [17]. It is stated that filter methods are preferred even when the subset of genes is not optimal due to their computational and statistical scalability [17].

Supervised feature selection is most commonly used method which utilizes labeled data for feature selection. Unsupervised feature selection method is used for scrutiny of biological data and it is helpful in finding the insights for classification of disease types. The disadvantage of using unsupervised feature selection method is that it relies on mathematical principles and neglects the correlation between different features whereas semi-supervised feature selection method utilizes both labeled and unlabeled data and shows better results for gene expression microarray data [18].

Many methods have been proposed for gene selection in past. Deepthi and Thampi [19] integrated population based search technique Particle Swarm Optimization (PSO) with k-means for gene selection. Alshamlan et al. [20] proposed a hybrid gene selection method which integrates the advantages of Genetic Algorithm (GA) and Artificial Bee Colony (ABC) algorithm. Application of Principal component analysis (PCA) has been used for selection of informative and highly dominating genes [21].

Predictive and informative gene selection is still a problem for researchers working on high dimensional gene expression datasets. The high dimensionality problem, noise, incomplete and uncertain genes makes the gene expression dataset complex. Various researchers worked on integration of two or more individual gene selection methods to get promising results.

Rathore et al. proposes a feed forward gene selection technique where two gene selection methods are used one after another [12]. Nguyen and Nahavandi [14] presented a modification to Analytic Hierarchy Process (AHP) by considering the quantitative factors of individual gene selection and ranking methods. These ranking methods generate stable subset of genes which are further used for classification. Nguyen et al. [22] worked on modified AHP and concluded that the results obtained gives superior performance in terms of high classification accuracy, stability across classifiers and decreased computational cost.

3. Proposed methodology

3.1. Analytic hierarchy process

In this paper, a method for gene selection using enhancements in AHP is proposed. The AHP developed by Saaty is called Saaty's AHP, which is a technique for system analysis and is used for solving the decision problems by dividing the whole problem into small problems [23]. AHP is a powerful multiple criteria decision making technique which has been used to solve the decision problems in various fields. The problem in AHP is decomposed into a hierarchy of criteria and alternatives. Fig. 1 shows the tree structure of AHP. The AHP commonly deals with qualitative criteria and depends upon the assumptions taken by decision-makers/experts. It is believed that the accuracy of the comparisons of the criteria is affected depending upon the knowledge of the decision-makers, their previous perceptions and understandings [24].

Knowledge to the decision makers are often limited when working on such a huge dataset of genes (in thousands) and a wide number of criteria referring to different areas. AHP

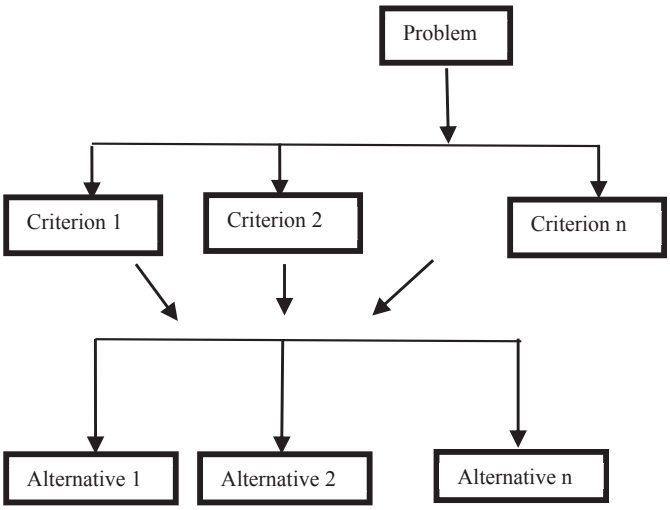

Fig. 1. The tree structure of AHP including criteria and alternatives.

involves these basic three steps: i) State the objective. ii) Defining the criteria. iii) Selecting the alternatives.

The next subsection includes the details of individual gene selection ranking methods employed, which are succeeded by our proposed method: Adaptive AHP.

3.2. Individual gene selection methods

The details of each gene selection methods with their mathematical expression that are employed in the proposed method is given below:

A. T-test

It is widely used in the study of gene expression datasets due to its simplicity and interpretability. For gene selection, t-test is applied on each gene by considering each expression level of given gene at different sample conditions. The test statistics is explained as:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{1}$$

where $\bar{x}$ is mean of the sample, μ is hypothesized mean of population, s is the standard deviation of sample and n is the size of the sample.

Ranking of the genes based on t-test is done by the absolute value of t. Genes with higher absolute values are more likely to be selected [25].

B. Chi-square variance test

Chi-square variance test is an important method for gene selection. It assigns chi-square score to each gene based on their chi-square statistics. For the application of chi-square variance test for gene selection, the test is applied on individual gene and the respective values obtained from the test are responsible for ranking of genes. The test statistics is given as:

$$T = (n-1)\left(\frac{s}{\sigma_0}\right)^2 \tag{2}$$

where n is the size of sample, s is the standard deviation of sample, σ_0 is the hypothesized standard deviation of population [26].

C. Z-test

The z-test is a parametric hypothesis test which is used to determine whether sample data set comes from a population with a particular mean. The test statistics is:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{3}$$

where $\bar{x}$ is the mean of sample, μ is the mean of population, σ is the standard deviation of population and n is the size of sample [27].

D. Wilcoxon method

The Wilcoxon signed rank test is a nonparametric test. When used for one sample, W is the sum of the ranks of positive differences between the observations and the hypothesized mean value. For large sample size, the p-value using the z-statistics is calculated as:

$$z = \frac{\left(W - \frac{n(n+1)}{4}\right)}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \tag{4}$$

where n is the size of sample [28].

E. Signal to noise ratio

Each gene is represented by an expression vector $V(g) = (e_1, e_2, e_3, \ldots..e_n)$ where e_i represents the expression level of each gene g in the ith sample. SNR infers that the separation between the means for two classes is a measure for partition. Moreover, the slight standard deviation supports the partition between classes. The separation between mean values is then standardized by the standard deviation of the classes [29]. The higher the value of SNR, the stronger is the gene, so the genes are sorted in decreasing order [30].

Each of the above explained criteria can be used to obtain the ranking of genes and from them the top ranked most informative genes are selected. The hierarchy of factors used for gene selection by Adaptive Analytic Hierarchy Process (A2HP) is shown in Fig. 2 as tree structure.

3.3. Adaptive Analytic Hierarchy Process (A2HP) gene selection

In Adaptive Analytic Hierarchy Process (A2HP) deals with high dimensional data, firstly some preprocessing is done on the gene expression dataset which includes *K nearest neighbor*

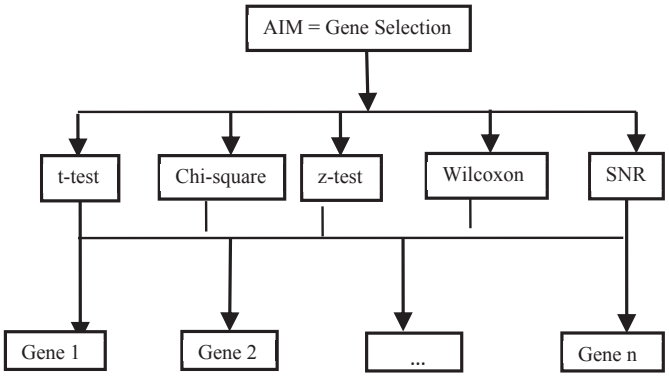

Fig. 2. Hierarchy structure.

(Knn) imputation method for imputing missing data and various filters to remove gene profiles with low absolute values, low entropy expression values and small profile variance as shown in Fig. 3. These genes have bad quality due to quantization errors and poor spot hybridization, so there are not of much interest and will lead to false results [14]. The reduced dataset obtained after applying these preprocessing steps is fed as input to the individual gene selection ranking methods.

The five criterion used here gives us the quantitative results. The information is synthesized to determine relative ranking of genes. At first the results are used to form the pairwise comparison matrix for the criteria. Earlier for solving the conventional problems Saaty rating scale [1,9] is used to make pairwise comparison matrix. It prioritizes the alternatives using eigen vectors but, it suffers from the problem of right and left Eigenvector inconsistency [31]. But here in A2HP

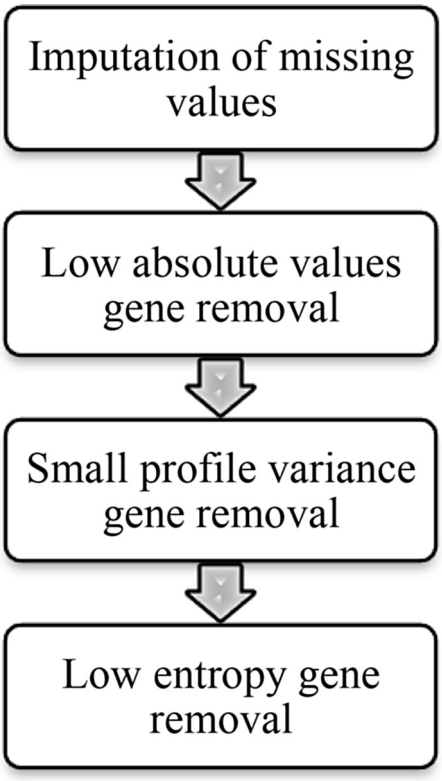

Fig. 3. The Block diagram of preprocessing steps for Proposed Method.

scale [1,5] is used for marking importance of one criteria over other. Suppose $Z = (z_{ij})$ is the m × m dimension pairwise matrix which shows the relative importance of one criterion over another, where m is the number of individual gene selection methods used for ranking of genes. Here each element z_{ij} represents the relative importance of criterion i over criterion j. The reciprocal characteristics represent the following:-

$$z_{ij} = \frac{1}{z_{ji}}, \forall i \neq j, \quad i,j \epsilon [1,m],$$

$$z_{ij} = 1, \quad \forall i \epsilon [1,m]$$

if criterion i is absolutely more important than criterion j, then take $z_{ij} = 5$. On the other hand j must be absolutely less important than criterion i and $z_{ji} = 1/5$. where $z_{ij} = 1$, this represents that the two criterion are equally important. The higher the value, the more important the criterion. Element z_{ij} that is greater than 1 is known as superior element while the element lesser than 1 is called inferior element [31].

3.3.1. Ranking procedure

For all pair of two genes i and j, prepare a pairwise comparison matrix $M = (m_{ij})$, where each element m_{ij} represents the relative importance of gene i over gene j with respect to specific criteria. After constructing comparison matrices, calculate eigen vectors, which gives the ranking of genes. The method to calculate eigen vectors from the comparison matrix is given in Table 1.

The eigen vector is used to get ranking of genes. Obtain each criterion information for each gene in a matrix. Suppose $Y = (\varepsilon_{ij})$ is a n × m performance matrix which gives the value (eigen vector) of each gene for a particular criterion, where n is the number of genes and m is the number of criterion used. Table 2 shows the n × m performance matrix.

Finally the ranking of genes is obtained by multiply this n × m performance matrix with m × 1 vector of criterion ranking. The top ranking genes are the informative genes obtained and these can be further used for various purposes like classification of disease, etc. The methodology proposed in this work is shown in Fig. 4.

4. Datasets and results

4.1. Experimental datasets

Two datasets of different organisms are taken for experiments which includes yeast and mouse. The datasets are obtained from NCBI Gene Expression Omnibus (GEO) repository [32]. The gene expression dataset of yeast contains

Table 1
Method to calculate eigen vector.

M	G_1	…	G_n	Sum of values	Eigen vector
G_1	m_{11}	…	m_{1n}	$S_1 = m_{11} + \ldots + m_{n1}$	$\varepsilon_1 = \left(\frac{m_{11}}{S_1} + .. + \frac{m_{1n}}{S_n}\right)/\mathrm{n}$
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…
G_n	m_{n1}	…	m_{nn}	$S_n = m_{1n} + \ldots + m_{nn}$	$\varepsilon_n = \left(\frac{m_{n1}}{S_1} + .. + \frac{m_{nn}}{S_n}\right)/\mathrm{n}$

Table 2
Performance matrix.

	t-test	Chi-square	z-test	Wilcoxon	SNR
Gene 1	ε_{T1}	ε_{C1}	ε_{Z1}	ε_{W1}	ε_{S1}
…	…	…	…	…	…
Gene n	ε_{Tn}	ε_{Cn}	ε_{Zn}	ε_{Wn}	ε_{Sn}

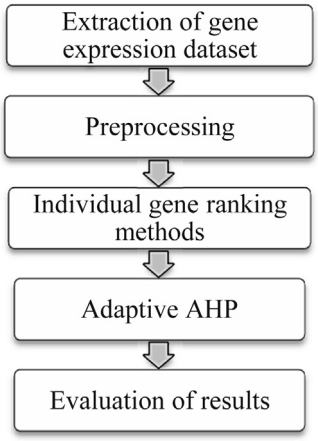

Fig. 4. The proposed informative gene selection methodology.

the expression levels of 6361 genes at 7 different time points. The gene expression dataset of mouse contains the expression levels of 500 genes and 26 samples.

4.2. Gene selection results

At first, various preprocessing steps are taken for the gene expression datasets as explained earlier, Knn is used to fill the missing values of genes and various filters are used to remove genes with low absolute values, low entropy and small profile variance. The number of genes obtained after applying preprocessing is detailed in Table 3 and graphically represented in Fig. 5a and b.

Then, this reduced dataset is fed to the five individual gene selection ranking methods. Each gene selection method results in different subsets of informative genes. These methods give ranking of genes. Tables 4 and 5 represent the overlap among the top genes selected by the six gene selection methods: t-test, Chi-square variance, z-test, Wilcoxon, SNR and A2HP.

4.3. Performance evaluation

The percentage of deduction in number of genes and time complexity are taken as the performance measure for the proposed methodology. Percentage of deduction is the measure to find the percentage of reduced number of genes after applying each individual gene selection method.

Table 3
Number of genes obtained after preprocessing.

Dataset	Total genes	Low absolute values	Small variance	Low entropy
Yeast Data (GDS37)	6361	6350	5715	4394
Mouse Data (GDS1406)	500	453	408	347

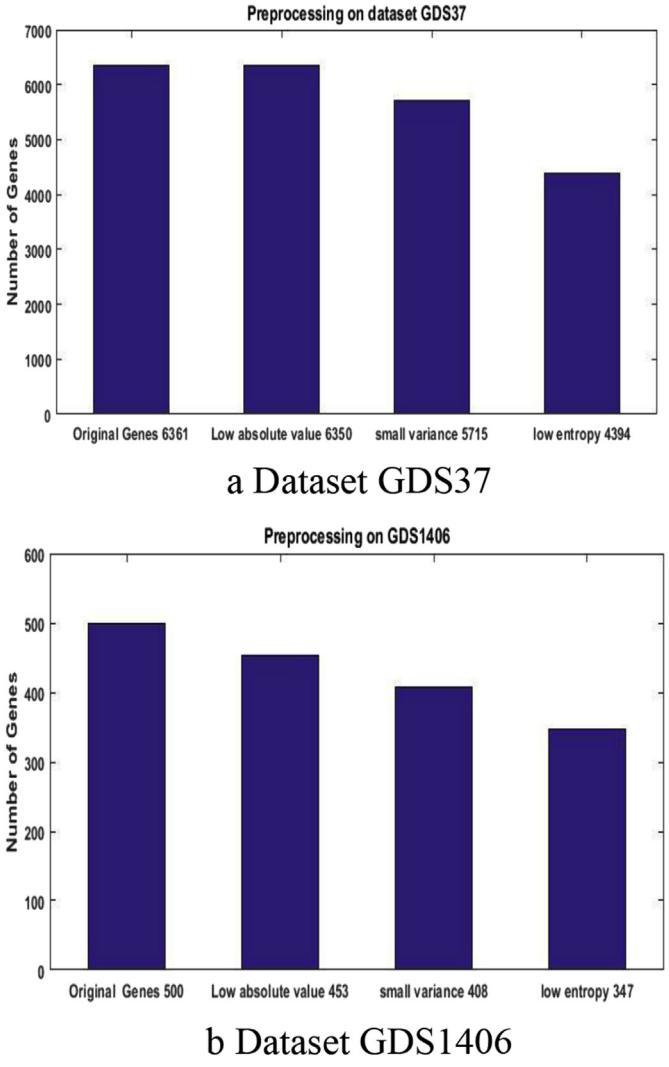

Fig. 5. Result of preprocessing filters on datasets.

Table 4
Overlap matrix for GDS37 for top 500 genes.

	t-test	Chi-square	z-test	Wilcoxon	SNR	A2HP
t-test	500	039	277	000	067	064
Chi-square	059	500	198	063	043	064
z-test	277	198	500	002	048	052
Wilcoxon	000	063	002	500	047	048
SNR	067	043	048	047	500	067
A2HP	064	064	052	048	067	500

Table 5
Overlap matrix for GDS1406 for top 50 genes.

	t-test	Chi-square	z-test	Wilcoxon	SNR	A2HP
t-test	50	04	10	07	03	11
Chi-square	04	50	41	06	06	40
z-test	10	41	50	07	06	47
Wilcoxon	07	06	07	50	13	08
SNR	03	06	06	13	50	06
A2HP	11	40	47	08	06	50

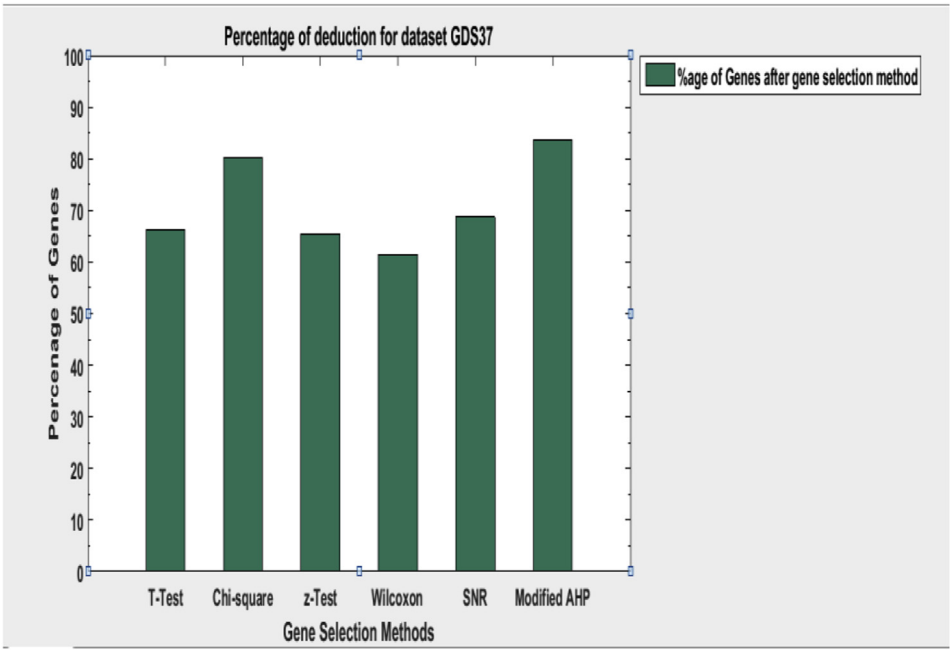

Fig. 6. Percentage of reduced number of genes for GDS37.

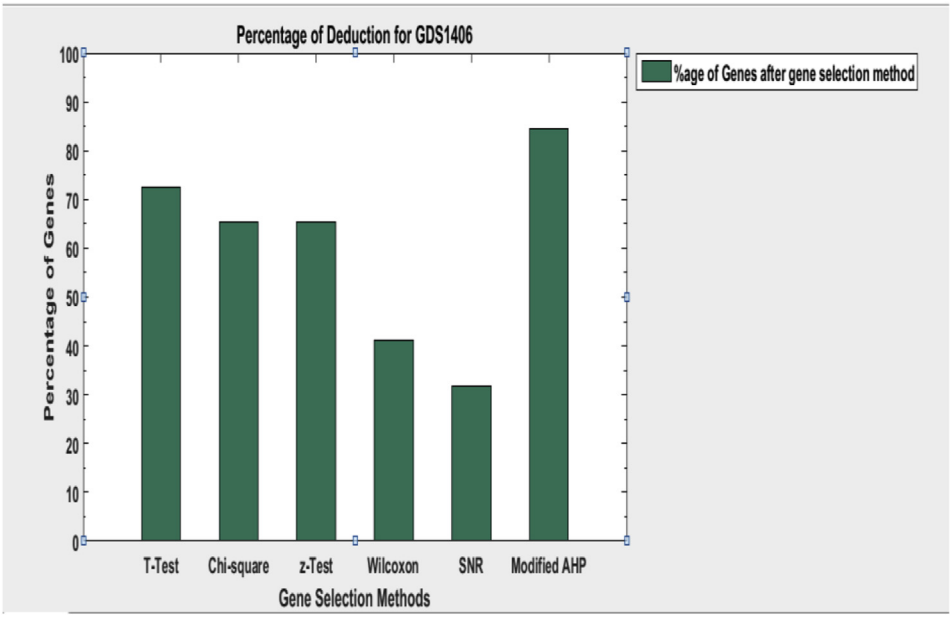

Fig. 7. Percentage of reduced number of genes for GDS1406.

Table 6
Percentage of reduced genes for both the datasets.

Gene selection methods	Reduction for GDS37 (in %)	Reduction for GDS1406 (in %)
t-test	66.20	72.6
Chi-square variance	80.30	65.4
z-test	65.47	65.4
Wilcoxon	61.43	41.2
SNR	68.71	31.8
A2HP	83.57	84.6

Table 7
Time complexity for both the datasets.

Gene selection methods	Time (in sec) for GDS37	Time (in sec) for GDS1406
t-test	16.36	1.33
Chi-square variance	14.70	1.95
z-test	08.66	1.61
Wilcoxon	13.30	1.80
SNR	53.65	5.75
A2HP	01.64	0.28

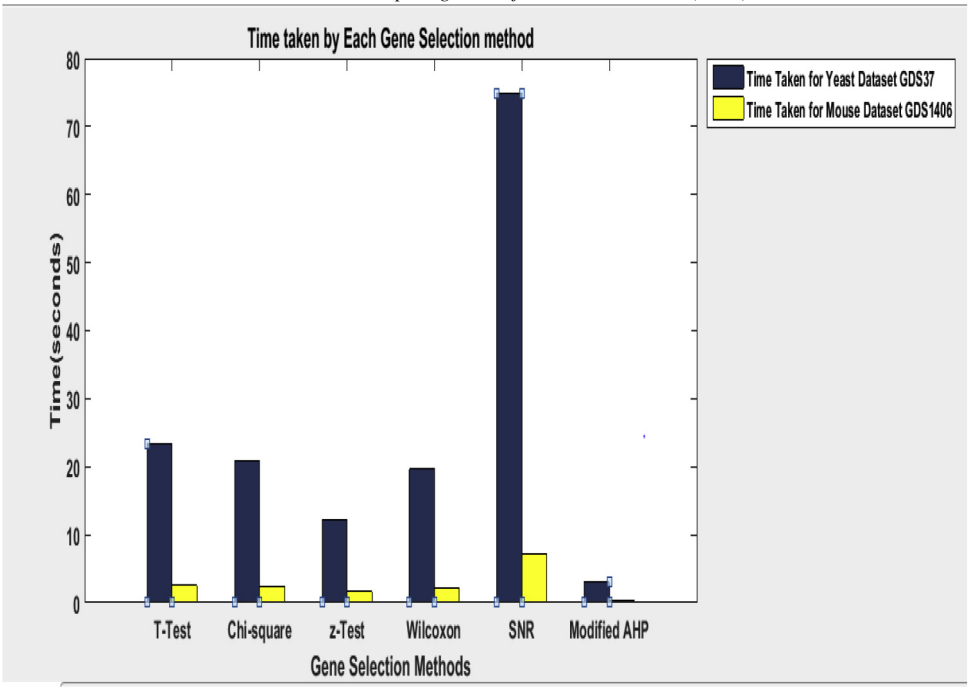

Fig. 8. Time complexity of each Gene Selection Method.

$$\%\ \text{of deduction} = \frac{\text{number of genes obtained after applying gene selection method}}{\text{total number of genes}} * 100$$

It is clearly shown that the proposed methodology yields better results in terms of number of genes selected for both the datasets. It is concluded that the proposed method gives small subset of informative genes when the high dimensional gene expression dataset is considered. Figs. 6 and 7 show the deduction in genes after applying gene selection methods.

Percentage of reduced genes for both the datasets is given in Table 6. Time complexity of different Gene Selection Methods is shown in Table 7. Time taken by each method is calculated and it is concluded that time taken by A2HP is minimum for both the datasets. While SNR method takes more time as compared to other methods for both the datasets. Fig. 8 gives the amount to time taken by each method and it is clear that the more the number of genes more the time required selecting informative genes.

Where in Table A1 depicts the names of top 30 genes selected by six gene selection methods t-test, chi-square variance test, z-test, wilcoxon, SNR and A2HP.

5. Discussion and conclusion

Already many methods have been proposed for dimensionality reduction and selection of informative genes from gene expression dataset but the composite methods which combine the outcomes of various individual gene selection methods provide better results.

In this work, Adaptive AHP is used which is a composite method and deals with both qualitative as well as quantitative criteria whereas conventional AHP deals only with qualitative factors and the results are derived from the experts, who have knowledge of their particular areas only. At first preprocessing of gene expression dataset is processed in order to make the dataset efficient and precise. After this five-individual gene selection ranking methods are applied on the gene expression dataset. Each method yields a different subset of reduced number of genes. All the resulted subsets obtained from these individual methods are then used by A2HP.

AHP is a multiple criteria decision making technique which has been used in many fields to solve decision problems. The problem in AHP is decomposed into a hierarchy of criteria and alternatives which are then represented in the form of a tree structure. A2HP utilizes the outcomes of individual five gene selection ranking methods to generate a stable and robust subset of genes. The experimental results show that the number of genes selected by using A2HP gives better time complexity and comparable results with the individual gene selection ranking methods.

There are number of directions to the future work. First, it can be enhanced by increasing the number of individual gene selection ranking methods. These methods can be enhanced to seven or more for more accurate results as the proposed method is a cumulative method which extracts the advantages of individual methods. Second, apart from these individual genes selection ranking methods, different methods can be used which may result in more informative genes selection. Third, focus can be laid on preprocessing steps to fill missing values of the genes, remove noise and those redundant genes which have different names but posses same functionality and also to deal with those genes that perform differently but have same names. There are multiple copies of genes with same name is present in the gene expression dataset which further need to be addressed.

Conflict of interest

The authors confirm that this article has no conflict of interest.

Appendix

Table A1
Top 30 genes selected in the Yeast Dataset (GDS37).

S. No.	T-test	Chi-square variance test	Z-test	Wilcoxon test	SNR	A2HP
1.	'PET9'	'GSY1'	'GAD1'	'TVP18'	'RSM25'	'SPO77'
2.	'MCP1'	'STR3'	'SPI1'	'DBP5'	'RPL5'	'TIR4'
3.	'DUG1'	'RTN2'	'HSP12'	'YPR1'	'YGR045C'	'ESBP6'
4.	'ADH5'	'OM45'	'MSC1'	'APC11'	'XBP1'	'NAM7'
5.	'ADH3'	'BNA2'	'PGM2'	'MKS1'	'STP4'	'CDC24'
6.	'COS8'	'SOL4'	'YGP1'	'MEC3'	'YOR387C'	'YIP3'
7.	'PDI1'	'HSP30'	'GSY2'	'YDL176W'	'ZIP1'	'TDA9'
8.	'RNR2'	'SFC1'	'HSP30'	'CIA1'	'ERP4'	'ATG5'
9.	'COS5'	'SPI11'	'HXT6'	'TCM62'	'GPN3'	'CWH41'
10.	'FRD1'	'NCE103'	'HSP26'	'DDC1'	'ROX1'	'LDB17'
11.	'COS7'	'GSY2'	'PIC2'	'BUB1'	'AIM20'	'GAL2'
12.	'PRM8'	'YMR206W'	'PRX1'	'RNA15'	'GAL83'	'SSM4'
13.	'TUF1'	'	'YNL200C'	'GUF1'	'BUD17'	'CCT2'
14.	'ATG9'	'YGL138C'	'	'DON1'	'SNU66'	'HIT1'
15.	'MC3'	'AGP21'	'GSY1'	'YJU2'	'PMT3'	'PXR1'
16.	'YGR026W'	'HSP12'	'HSP42'	'DAL7'	'VAC7'	'DOS2'
17.	'	'MSC1'	'SOL4'	'YDR360W'	'ZUO1'	'MAM33'
18.	'ALG6'	'HSP42'	'RGI1'	'	'YGL132W'	'YMC2'
19.	'MSN4'	'TMA10'	'YDC1'	'COX10'	'	'UTP15'
20.	'COS4'	'SPI1'	'HXT7'	'CLP1'	'SSN2'	'YMR087W'
21.	'YIR044C'	'YNL200C'	'MDH1'	'HMI1'	'CHS1'	'
	'TSA1'	'YLR312C'	'PRC1'	'YEL1'	'YME1'	'GIC1'
22.	'HAC1'	'AGP2'	'LSC2'	'ALP1'	'E. COLI	'YGR201C'
23.	'UFD2'	'HSP26'	'SDH2'	'YJL135W'	#40'	'YNL058C'
24.	'PTC3'	'MFB1'	'ALD4'	'RED1'	'TIF34'	'RNQ1'
25.	'YDR133C'	'PRX1'	'TMA10'	'YBR138C'	'JIP4'	'YHL018W'
26.	'SMP3'	'GAD1'	'PNC1'	'YOL024W'	'AXL1'	'CSM3'
27.	'RPD3'	'AIM17'	'COX5B'	'RMD6'	'YPT31'	'GIP3'
28.	'PST1'	'YPS6'	'QCR7'	'PHO23'	'TMN3'	'SCP160'
	'COS3'	'GAD1A'	'ACO1'	'AST2'	'CEM1'	'GEA1'
29.	'YLR241W'	'SPO74'	'PEP4'	'ZWF1'	'FLX1'	'SKI8'
30.				–	'PCP1'	

References

[1] Gill N, Singh S. Biological sequence matching using Boolean algebra vs. fuzzy Logic. IJCA 2011a;26:15–21.
[2] Gill N, Singh S. Multiple sequence alignment using Boolean algebra and fuzzy logic: a comparative study. Int J Comp Tech Appl 2011b;2:1145–52.
[3] Bansal P, Singh S, Bhola A. A review on Bayesian network techniques for inferring gene regulatory networks. IJARSE 2015;4:386–91.
[4] Gill N, Singh S, Aseri TC. Computational disease gene prioritization: an appraisal. J Comp Biol 2014;21:456–65.
[5] Grewal N, Singh S, Aseri TC. Effect of aggregation operators on network-based disease gene prioritization: a case study on blood disorders. IEEE/ACM Trans Comput Biol Bioinform 2016:1–9.
[6] Babu MM. Introduction to microarray data analysis. Comput Genomics Theory Appl 2004;17:225–49.
[7] Abhishek Singh S. A gene regulatory network prediction method using particle swarm optimization and genetic algorithm. IJCA 2013a;83:32–7.
[8] Abhishek Singh S. Approaches to gene regulatory network modeling. IJIEASR 2013b;2:35–8.
[9] Abhishek Singh S. Gene selection using high dimensional gene expression data: an appraisal. Curr Bioinform 2016;11:1–9.
[10] Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 1996;6:639–45.
[11] Singh RK, Sivabalakrishnan M. Feature selection of gene expression data for cancer classification: a review. Procedia Comput Sci 2015;50:52–7.
[12] Rathore S, Muhammad AI, Hussain M. A novel approach for automatic gene selection and classification of gene based colon cancer datasets. In: IEEE International Conference on Emerging Technologies (ICET); 2014. p. 42–7.
[13] Dheenathayalan K, Ramsingh J. Identifying significant genes from DNA microarray using genetic algorithm. In: IEEE International Conference on Intelligent Computing Applications (ICICA); 2014. p. 1–5.
[14] Nguyen T, Nahavandi S. Modified AHP for gene selection and cancer classification using type-2 fuzzy logic. IEEE Trans Fuzzy Syst 2016;24:273–87.
[15] Mahajan S, Bhola A, Singh S. Review on feature selection approaches using gene expression data. Imperial J Interdiscip Res 2016;2:356–64.
[16] Chandrashekar G, Ferat S. A survey on feature selection methods. Comput Electr Eng 2014;40:16–28.
[17] Lazar C, Taminau J, Meganck S, Steenhoff D, Coletta A, Molter C, et al. A survey on filter techniques for feature selection in gene expression microarray analysis. IEEE/ACM Trans Comput Biol Bioinform 2012;9:1106–19.
[18] Saeys Y, Iñaki I, Pedro L. A review of feature selection techniques in bioinformatics. Bioinformatics 2007;23:2507–17.
[19] Ang JC, Mirzal A, Haron H, Hamed HN. Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. IEEE/ACM Trans Comput Biol Bioinform 2016;13:971–89.

[20] Deepthi PS, Thampi SM. PSO based feature selection for clustering gene expression data. In: IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES); 2015. p. 1−5.
[21] Alshamlan HM, Ghada HB, Yousef AA. Genetic Bee Colony (GBC) algorithm: a new gene selection method for microarray cancer classification. Comp Biol Chem 2015;56:49−60.
[22] Banu PKN, Simon A. Informative gene selection − an evolutionary approach. In: IEEE International Conference on Current Trends in Information Technology (CTIT); 2013. p. 129−34.
[23] Nguyen T, Khosravi A, Creighton D, Nahavandi S. A novel aggregate gene selection method for microarray data classification. Pattern Recogn Lett 2015;60:16−23.
[24] Tung SL, Tang SL. A comparison of the Saaty's AHP and modified AHP for right and left eigenvector inconsistency. Eur J Oper Res 1998;106:123−8.
[25] Banuelas R, Antony J. Modified analytic hierarchy process to incorporate uncertainty and managerial aspects. Int J Prod Res 2004;42:3851−72.
[26] Jeanmougin M, De Reynies A, Marisa L, Paccard C, Nuel G, Guedj M. Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. PloS one 2010;5:e12336.
[27] Mantel N. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. JASA 1963;58:690−700.
[28] Sidak Z, Sen PK, Hajek J. Theory of rank tests. Academic Press; 1999.
[29] Oyeka IC, Ebuh GU. Modified Wilcoxon signed-rank test. Open J Stat 2012;2:172.
[30] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531−7.
[31] Lancucki A, Saha I, Lipinski P. A new evolutionary gene selection technique. In: IEEE Congress on Evolutionary Computation (CEC); 2015. p. 1612−9.
[32] Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, et al. NCBI GEO: mining millions of expression profiles—database and tools. Nucleic Acids Res 2005;33:D562−6.