ISSN (Online) : 0974-5645
ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(1), 27–34, January 2015* DOI: 10.17485/ijst/2015/v8i1/52310

Multi-Objective Firefly Algorithm for Multi-Class Gene Selection

G. V. Manoharan[1*] and R. Shanmugalakshmi[2]

[1]Faculty of Information and Communication Engineering, Anna University, Chennai, Tamil Nadu, India; manoharangv@gmail.com
[2]Department of CSE, Government College of Technology, Coimbatore, Tamil Nadu, India

Abstract

Multiclass cancer classification is an emerging technique which presents the possibility of cancer identification using microarray data. For selecting genes in the multiclass gene categorization filter methods are frequently used. But the filter method is not applicable for some of the multiclass microarray data sets because of the rigorous heterogeneity of biological tissues and samples. So, for selecting genes decay the multiclass ranking statistics into class explicit statistics and then Pareto-front analysis is used. Also, to identify the Pareto-optimal set the non-dominant sorting genetic algorithm is suggested. But the drawback is this method does not scale with high complexity. Because, where the number of elements which are represented to mutation is large there is an exponential raise in search space size. So, in this manuscript an innovative technique is introduced which is called Multiobjective Firefly Algorithm for Multiclass Gene Selection (MFGS). A firefly has a tendency to be fascinated towards other fireflies with superior flash intensity. The multiple objective firefly algorithms intend to optimize two or more conflicting characteristics represented by fitness functions. In the multiple objective firefly method, a set of Pareto-optimal solutions are created which concurrently optimize the contradictory necessities of the multiple fitness functions. In the proposed method, the genes are selected by optimizing the number of fireflies in the multiple class-specific statistics. An experimental result shows that when compared to the existing method, there is less complexity, high classification accuracy of the proposed MFGS method.

Keywords: Filter Methods, Gene Selection, Multiobjective Firefly Algorithm, Pareto-Front Analysis

1. Introduction

Gene expression data is an appearance level of gene when DNA encoded into protein. By using the microarray techniques, the appearance level of the gene is evaluated. In the microarray gene-expression data sets, the selection of pertinent genes has become an essential part in order to enhance the identification of types of samples and also provides motivating biological insights. The gene selection methods are categorized as a filter, wrapper and embedded methods[1-3]. The filter method is most frequently used gene ranking method. Because the filter methods are very simple, easy to utilize and less computational effectiveness. By taking the correlation of gene expressions the filtering methods rank the genes.

For multiclass gene selection, some of the filter criteria like F-score[4], Kruskal-Wallis (KW)-score[5], mutual information[6], and entropy[7] are suggested. A generalization of signal-to-noise ratio is suggested for multiclass gene selection by considering the gene dominant index and gene dormant index[8]. But some of the filter methods are not considered the significant distinctiveness of many multiclass microarray data sets. Because there is severe heterogeneity of biological tissues and samples[9].

Due to the deficiency of strong features or enough samples some of the classes may be more complex to differentiate when compared to other classes. This class-explicit distinctiveness of data leads to a lesser-known problem in gene selection, referred to as the siren-pitfall problem[10]. In the existing method, the F-score and the

*Author for correspondence

KW - score are exaggerated by siren pitfalls by utilizing simulated data sets. To conquer the trouble of the pitfalls in the genetic selection process, a multiobjective perspective and a Pareto-Front Analysis (PFA)-based approach is suggested[22]. In this manuscript, a Multiobjective Firefly Algorithm for Multiclass Gene Selection (MFAMGS) is introduced to concurrently optimize the contradictory necessities of the multiple fitness functions.

The rest of the section is organized as follows: Section II describes the related work. Section III presents the Pareto-front analysis method. Section IV describes the proposed algorithm. Section V provides the evaluation and comparison analysis. Section V concludes the paper.

2. Previous Research

In microarray analysis, the selection of discriminatory genes is vital to develop the accuracy and also to reduce the computational complexity. There are two basic techniques for selection of genes which appeared in the machine learning techniques. The techniques are filtered and wrapper methods: In most of the work related to the DNA research area, filter techniques are used which is based on gene ranking. In this technique, the ranking of genes is accomplished based on the score value and the genes with the highest scores are chosen. These methods contain-test, Relief-F, information gain, Kruskal–Wallis rank and so on. These methods are simple to develop and understand. But the drawback is according to the unique contribution the genes are selected. The mutual information between genes is not considered so the result is not satisfied[11-13].

Tibshirani et al.[14,15] suggested the nearest shrunken centroid method which is used to recognize discriminant genes for multiple cancer categories by a shrinkage factor. In some of the Variable selection methods searching the class-specific genes for a specific cancer type and the genes, including significant information about the subtypes of cancers are eradicated by them. The classification of the nearest centroid method[16] is suggested by Dabney which considers the out class explicit features selection so that the error rate is low. Qi Shen et al. suggested a centroid based scoring method[11] to decide which genes differ most considerably between the classes and to choose genes for every cancer type. By using this parameter, the condition for the choosing gene for a particular class is to recognize genes that have shorter distances from centroid and large distances from the other class centroid.

There are some methods suggested for gene selection for improving accuracy and reduce the computational complexity.

A signal-to-noise ratio is used as a condition is suggested by Golub et al.[17] for calculating the association between a gene and a tumor subcategory. A gene shaving method is recommended by Hastie et al. with the principal component analysis. Lee et al. recommended the instructive gene subset by utilizing the technique of Bayesian learning[19]. For recursive elimination of features, the support vector machine is recommended by Guyon et al[20]. There are two categories in these methods; one is individual gene ranking approaches and gene subset ranking approaches[21]. Jin Hyun Park et al. suggested a new decisive factor[22] for evaluating the significance of individual genes by utilizing the mean and standard deviation of the distances from every attribute to the class centroid to consider the problem of gene selection. This technique follows the two steps: one is ranking and selection of genes and another one is validation of genes. This technique is suitable for multiple classification troubles.

3. Pareto Optimal Front Analysis

Selection of relevant genes is crucial in the microarray data analysis. So, in this method the multiclass ranking statistics are decomposed into class specific statistics. After that the Pareto front analysis is used for the selection of genes. Because in the microarray analysis due to the deficiency in the samples, some classes may be more complicated to differentiate when compared to other classes[22]. On the other hand, the class-specific statistics of some classes may dominate the overall ranking of the aggregation process which leads to siren pitfall problem[10] in gene ranking and impact the performance of classification. The main cause of the siren-pitfall trouble is that genes representing effortlessly distinguishable classes are strappingly represented in the data set, whereas genes pertinent to complex to separate classes are faintly represented. In this work, the two statistical parameters are used which is called an F-score and KW-score. Also, to overcome the problem of pitfalls in selection of genes, focus the two statistical conditions and a Pareto-front analysis method is used.

F-score is based on the F-statistics and it is most extensively used statistical test. It is defined as the ratio of between the intraclass and the interclass distances of the gene expression values[4]. The F-score of particular gene k is computed as,

$$F_k = \frac{\sum_{c=1}^{C}\sum_{s=1}^{m}\varphi(y_s = c)\left(\hat{x}_{kc} - \hat{x}_k\right)^2}{\sum_{c=1}^{C}\sum_{s=1}^{m}\varphi(y_s = c)\left(x_{ks} - \hat{x}_{kc}\right)^2} \quad (1)$$

In this equation, $\hat{x}_k$ denotes the average expression level of gene k across the entire samples, $\hat{x}_{kc}$ denotes the average expression level of the gene k of the samples belonging to cth class. φ represents the pointer function which is equal to 1 of the argument is true and 0 otherwise. The class specific statistic is determined by,

$$F_{kc} = \frac{m_c\left(\hat{x}_{kc} - \hat{x}_k\right)^2}{\sum_{c=1}^{C}\sum_{s=1}^{m}\varphi(y_s = c)\left(x_{ks} - \hat{x}_{kc}\right)^2} \quad (2)$$

F_{kc} represents the class-specific statistic which is computed by the ratio between the discrepancy of gene expressions of gene k of the samples belonging to class c and the variance of expressions of gene in the entire samples. This denotes the ranking statistics of gene k which is belonging to dissimilar classes.

The KW-score is another statistical parameter which is based on a nonparametric KW statistic that utilizes the rankings of gene expressions instead of their values. It is defined as the square of the differences between within-class average ranks and overall mean of the ranks. The KW-score of gene k is evaluated by,

$$KW_k = \frac{12}{m(m+1)}\left(\sum_{c=1}^{c} m_c\left(\hat{D}_{kc} - \hat{D}_k\right)\right)^2 \quad (3)$$

Where $\hat{D}_{kc}$ represents the average value of the ranking of expression value of gene k in the samples belongs to the cth class and $\hat{D}_k$ denotes the overall average rank of the expression value of gene k and m_1 represents the number of samples belonging to class c. The class specific statistics of gene k belonging to class c is given by,

$$KW_{kc} \propto m_c(\hat{D}_{kc} - \hat{D}_k)^2 \quad (4)$$

The set of class-specific statistics are represented $\{A_{kc}\}_{k=1,c=1}^{n,C}$. A_{kc} denotes a one-against-all statistic of gene k in class c and represents how discriminative gene k of class c is qualified to its capability to distinguish the other classes. The aggregating statistics are represented by A_i is rank statistics attained by aggregating class-specific statistics. This provides the overall capability of separability of classes:

$$A_k = \sum_{c=1}^{C} A_{kc} \quad (5)$$

The two statistical parameters are computed. For accomplishing the better classification performance, a group of genes optimizing individual class-specific statistics are required. As a result, the selection of genes is accomplished by recognizing the group of class-specific statistics, which are non-dominant by the rest of the class specific statistics. So, the Pareto front analysis method is suggested which is utilized to recognize the optimal set of genes to increase the classification performance.

The Pareto-front analysis is used to divide the entire genes into dissimilar Pareto fronts by utilizing class-specific statistics. A Pareto front is defined as the set of genes which does not dominate one another. The main intent to determine the Pareto fronts is to easily recognize the nondominating class-specific ranking statistics so that the pertinent genes were recognized for enhancing the performance of the classification. By determining the Pareto-optimal fronts, nondominating class-specific ranking statistics and thereby relevant genes for classification were identified.

There are two conditions to be satisfied if gene k is said to dominate another gene k'. The class specific statistic value of gene k is no inferior than gene k'. At least one of the class-specific statistic values of gene k are superior than that of a gene.

By using the Pareto analysis method the non-dominant genes are identified. If none of the genes dominate other genes, a pair of genes is said to be reciprocally nondominating. So, the nondominating gene set is generated. In this set, genes are not dominated by any other gene. In a given group of class-specific statistics, the Pareto fronts of genes are decided by performing all pair wise assessments and establish the nondominated genes. By using the Pareto front analysis, the genes are separated into ordered sets of Pareto fronts of class specific statistics. In the Pareto front, the genes are nondominated by other genes, referred as Pareto-optimal set. In the framework of gene selection, the Pareto-optimal set of genes contains the most significant genes for the classification as those genes have at least one class-specific statistics better than that of genes in the other fronts.

The nondominant sorting genetic algorithm is used to acquire the optimal set of Pareto fronts. In every Pareto front the genes are ranked by using the crowding distance which enlarges the diversity of class specific statistics of genes. In this measure, a smaller the value of crowding distance means that the gene is crowded by many other genes having close class-specific statistics. By utilizing the

crowding distance of each and every gene, the genes are sorted and ranked in the Pareto fronts.

4. Multiobjective Firefly Algorithm for Multiclass Gene Selection (MFGS)

Multiobjective Firefly Algorithm for Multiclass Gene Selection (MFGS) is an innovative technique which is used to optimize two or more conflicting characteristics represented by fitness functions. In this method, the genes are selected by optimizing the number of fireflies in the multiple class-specific statistics. A firefly algorithm is a metaheuristic algorithm which tends to be attracted towards other fireflies with superior flash intensity. But many search and optimization problems usually include multiple objectives. In contrast to the single objective optimization problems, the multi-objective optimization method is able to optimize one or more conflicting characteristics which are illustrated by objective functions. In this proposed method, a group of Pareto-optimal solutions are generated which concurrently optimize the conflicting necessities of the multiple fitness functions.

The Pareto fronts are a group of genes which does not dominate one another. If there are M objective functions, a gene k is said to dominate another gene k' if both constraints are satisfied:

(1) The gene k is no worse than k' in the entire M objective functions.

$$k < k' \Leftrightarrow A_{kc} \geq A'_k C, \quad \forall c \tag{6}$$

In this constraint A_{kc} represents the class specific statistic of gene k in class c.

(2) The gene k is sternly superior than k' in at least of the M objective functions.

$$A_{kc} > A'_k c, \exists c$$

In this constraint $A'_k c$ represents the class specific statistic of gene k' in class c. C is a class label.

Algorithm: A novel algorithm for Multiobjective Firefly Algorithm for Multiclass Gene Selection (MFGS)

1. Define the objective functions $f_1(x), \ldots f_K(x)$ *where* $x = \left(x_1, \ldots x_d\right)^T$
2. Generate initial population of fireflies x_i *where* $i = 1, \ldots n$
3. While(t < Max Generation)
4. for(i =1: n (all n fireflies)
5. for(j =1: n (all n fireflies)
6. // Determination of optimal set of Pareto sets
7. Check the conditions for non-dominant gene selection
8. // Constraints for Non-dominant genes computation
9. If gene $k < k'$ represents the gene k is dominating gene k' then
10. // Class specific statistic value of gene k is no worse than that of gene k'
11. $k < k' \Leftrightarrow A_{kc} \geq A'_k c, \quad \forall c$ // A_{kc} = class specific statistic of gene k in class c, $A'_k c$ = class specific statistic of gene k' in class c
12. //At least one of the class-specific statistic value of gene k is better than that of gene k'
13. $A_{kc} > A'_k c, \exists c,$ // A_{kc} = class specific statistic of gene k in class c, $A'_k c$ = class specific statistic of gene k' in class c
14. If PF_i dominates PF_j
15. Move firefly i towards j using the equation the equation (7)
16. 16Create new ones if the moves do not satisfy all the conditions
17. End if
18. End if
19. If there is no non-dominated genes can be found
20. Random weights are generated $w_p = p, \ldots P$
21. Identify the best solution g_*^t among all fireflies using (8)
22. Random walk around using (9)
23. End if
24. Update and pass the non-dominated solutions to next iterations
25. Find the current best approximation to the Pareto front
26. Update $t \leftarrow t + 1$
27. End while
28. For the Pareto fronts of genes and C classes
29. Let $Q = |P^c|$
30. Let $d_q = 0$ *for all* $q = 1, 2, \ldots Q$
31. For c = 1 to C do
32. Sort statistics A_{kc} in P^c in worse order
33. //Determination of crowding distance
34. $d_q = d_q \dfrac{A_{(q+1)c} - A_{(q-1)c}}{A_{q(c)}^{\max} - A_{q(c)}^{\min}}$

35. End for
36. Return crowding distance
37. Sort and rank the genes
38. Obtain ranked Pareto fronts

The following algorithm shows the Multiobjective Firefly Algorithm for Multiclass Gene Selection (MFGS). In this algorithm, the process begins with a suitable definition of objective functions with the constraints. Firstly, initialize a population of n fireflies so that they should disseminate among the search space as evenly as possible. This even dissemination is accomplished by sampling techniques via uniform distributions. The fixed number of iterations is defined; the iterations initiate with the assessment of brightness of all the fireflies and compare each pair of fireflies.

For the Pareto front analysis, the constraints are validated. If PF_i dominates PF_j the firefly i move towards j. In the given two fireflies, x_i and x_j the movement of firefly i is concerned to another more brighter firefly j is identified by,

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2}\left(x_j^t - x_i^t\right) + \alpha_t \in_i^t \tag{7}$$

After that an arbitrary weight vector is created so that an integrated best solution g_*^t can be acquired. For the next iteration, the non-dominated solutions are passed. At the end of a fixed number of iterations, in general n non-dominated solution points can be acquired to approximate the true Pareto front.

In order to do random walks more proficiently, we can discover the current best g_*^t which reduces an integrated objective through the weighted sum value.

$$\phi(x) = \sum_{p=1}^{P} w_p f_p, \quad \sum_{k=1}^{K} w_p = 1$$

In this equation, $w_p = \frac{\gamma_p}{P}$ in which the γ_p represents the arbitrary numbers drawn from a uniform distributed [0, 1]. To guarantee that $\sum_p w_p = 1$ the rescaling process is executed after creating K evenly distributed numbers. It is worth pointing out that the weights w_p should be selected arbitrarily at every iteration, so that the non-dominated solution can sample diversely along the Pareto front. Suppose, if a firefly is not dominated by others in the sense of Pareto front, the firefly moves

$$x_i^{t+1} = g_*^t + \alpha_t \tag{9}$$

In this equation, g_*^t the best solution found so far for a specified set of arbitrary weights.

Additionally, the randomness can be minimized as the iterations proceed, and this can be accomplished in a similar manner as that for simulated annealing and other random reduction techniques.

$$\alpha_t = \alpha_0 0.9^t \tag{10}$$

In this equation α_0 represents the initial randomness factor.

5. Performance Evaluation

For the experimental results, the performance of the existing and the proposed system is compared. The performance of the proposed approaches was estimated on the three real datasets such as a National Cancer Institute (NCI) Ross, lung, and NCI Staunton gene expression datasets. These are extensively used benchmark data sets to compute a gene ranking method, which consists of changeable numbers of classes and genes. These three data sets contained a huge number of genes, many of which had constant gene expression levels. National Cancer Institute is a dataset of gene expression summary of 60 National Cancer Institute (NCI) cell lines. These 60 human tumor cell lines are obtained from patients with leukemia, melanoma, along with, lung, colon, central nervous system, ovarian, renal, breast and prostate cancers. The lung dataset consists of data on 40 lung cancer patients, which is used to compare the outcome of two chemotherapy behavior in extending survival time.

In the existing method, Multiobjective genetic Algorithm for Multiclass Gene Selection (MGGS) is used. In the proposed system, Multiobjective Firefly Algorithm for Multiclass Gene Selection (MFGS) is introduced which optimizes two or more conflicting characteristics represented by fitness functions. The performance metrics such as precision, Recall, accuracy is compared for existing and proposed system.

5.1 Precision

Precision value is evaluated according to the retrieval of information at true positive prediction, false positive.

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)}$$

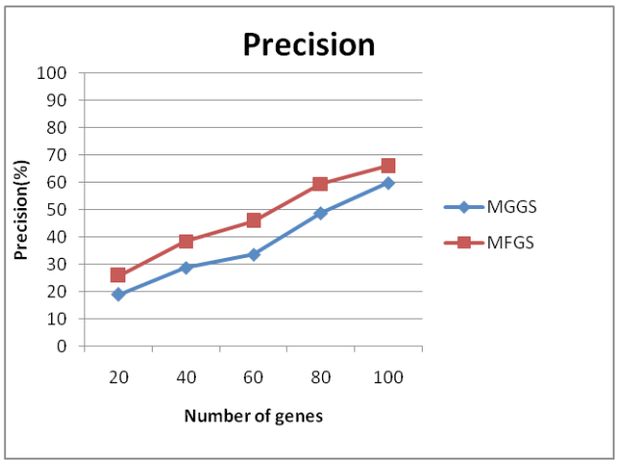

Figure 1. Precision

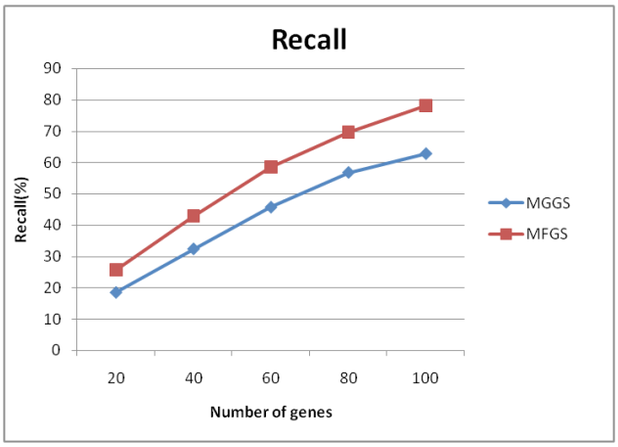

Figure 2. Recall

Table 1. Precision vs. Number of genes

SNO	Number of genes	MGGS	MFGS
1	20	18.9	25.7
2	40	28.7	38.4
3	60	33.6	45.8
4	80	48.7	59.4
5	100	59.7	65.9

The corresponding results of the MGGS and MFGS are evaluated for Precision. Figure 1 shows that when compared to MGGS the precision is improved in MFGS. In the existing research, for optimization the genetic algorithm is used. In the proposed research, the firefly algorithm is used to find the Pareto optimal set. Compared to the genetic algorithm, in the firefly algorithm, there is high precision value.

5.2 Recall

Recall value is evaluated according to the retrieval of information at true positive prediction, false negative.

$$Recall = \frac{True\ Positive}{(True\ positive + False\ negative)}$$

The corresponding results of the MGGS and MFGS are evaluated for Recall. Figure 2 shows that when compared to MGGS the recall is improved in MFGS. In the existing research, for optimization the genetic algorithm is used. In the proposed research, the firefly algorithm is used to find the Pareto optimal set. Compared to the genetic algorithm, in the firefly algorithm, there is high recall value.

Table 2. Recall vs. Number of datasets

SNO	Number of genes	MGGS	MFGS
1	20	18.6	25.7
2	40	32.5	42.9
3	60	45.8	58.5
4	80	56.9	69.7
5	100	62.9	78.2

5.3 Accuracy

Accuracy is evaluated as,

$$Accuracy = \frac{(True\ positive + True\ negative)}{(True\ Positive + False\ negative + @False\ positive + False\ negative)}$$

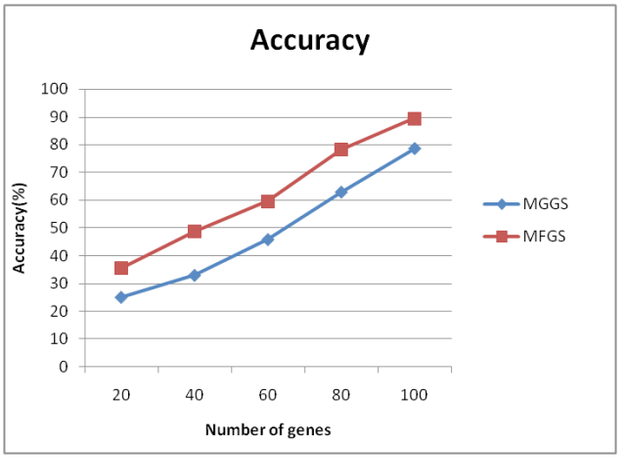

Figure 3. Accuracy

Table 3. Accuracy vs. Number of genes

SNO	Number of genes	MGGS	MFGS
1	10	25.2	35.6
2	20	32.9	48.9
3	30	45.8	59.7
4	40	62.9	78.3
5	50	78.6	89.5

The corresponding results of the MGGS and MFGS are evaluated for Accuracy. Figure 3 shows that when compared to MGGS the accuracy is improved in MFGS. In the existing research, for optimization the genetic algorithm is used. In the proposed research, the firefly algorithm is used to find the Pareto optimal set. Compared to the genetic algorithm, in the firefly algorithm, there is high accuracy.

6. Conclusion

In the microarray data analysis, multiclass cancer classification is a promising technique which provides the recognition of cancer. Selection of relevant genes is a challenging in the identification of cancer analysis. In this work, a novel technique called Multiobjective Firefly Algorithm for Multiclass Gene Selection (MFGS) is introduced which is used to optimize two or more conflicting characteristics represented by objective functions. In the proposed method a set of Pareto optimal set of genes is identified by using the Multiobjective Firefly Algorithm for Multiclass Gene Selection method. In the proposed method the selection of genes is chosen by optimizing the number of fireflies in the multiple class-specific statistics. When compared to the existing method there is less complexity and high classification accuracy of the proposed method. This proposed method is not suitable for multiple criteria. So, in future work, an efficient technique will propose to consider multiple criteria for enhancing the classification accuracy.

7. References

1. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. J Mach Learn Res. 2003; 3:1157–82.
2. Duan K-B, Rajapakse JC, Wang H, Azuaje F. Multiple SVM–RFE for Gene Selection in Cancer Classification with Expression Data. IEEE Trans. Nanobiosciences. 2005 Sep; 4(3):228–34.
3. Mundra PA, Rajapakse JC. SVM-RFE with MRMR Filter for Gene Selection. IEEE Trans Nanobiosciences. 2010 Mar; 9(1):31–7.
4. Dudoit S, Fridlyand J, Speed TP. Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data. J Am Statistical Assoc. 2002; 97(457):77–86.
5. Chen D, Liu Z, Ma X, Hua D. Selecting Genes by Test Statistics. J Biomed Biotechnol. 2005; 2:132–8.
6. Ding C, Peng H. Minimum Redundancy Feature Selection from Microarray Gene Expression Data. J Bioinformatics Comput Biol. 2005; 3:185–205.
7. Liu X, Krishnan A, Mondry A. An Entropy-Based Gene Selection Method for Cancer Classification Using Microarray Data. BMC Bioinformatics. 2005; 6:76.
8. Tsai Y-S, Lin C-T, Tseng G, Chung I-F, Pal N. Discovery of Dominant and Dormant Genes from Expression Data Using a Novel Generalization of SNR for Multi-Class Problems. BMC Bioinformatics. 2008; 9:425.
9. Clarke R et al. The Properties of High-Dimensional Data Spaces: Implications for Exploring Gene and Protein Expression Data. Nature Rev. Cancer. 2008; 8:37–49.
10. Forman G. A Pitfall and Solution in Multi-Class Feature Selection for Text Classification. Proc. 21st Int'l Conf. Machine Learning; 2004.
11. Shen Q, Shi W-M, Kong W. New Gene Selection Method for Multiclass Tumor Classification by Class Centroid. J Biomed Informat. 2009; 42(1):59–65.
12. Cho JH, Lee D, Park JH, Lee IB. New gene selection method for classification of cancer subtypes considering within-class variation. FEBS Lett 2003; 551:3–7.
13. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995; 270:467–70.
14. Tibshirani R, Hastie T, Narasimhan B, Chu G. Class prediction by nearest shrunken centroids, with applications to DNA microarrays. Stat Sci. 2003; 18:104–17.
15. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci USA. 2002; 99:6567–7572.
16. Dabney AR. Classification of microarrays to nearest centroids. Bioinformatics. 2005; 21:4148–54.
17. Golub TR., Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomeld CD, Lander ES. Science. 1999; 286:531–7.

18. Hastie T, Tibshirani R, Eisen MB, Alizadeh A. Levy R., Staudt L, Chan WC, Botstein D and Brown P. Genome Biol. 2001; 1:research0003.1–0003.21.
19. Guyon I, Weston J, Barnhill S, Vapnik V. Mach Learn. 2002; 46:389–422.
20. Lu Y, Han J. Inf Syst. 2003; 28:243–68.
21. Cho J-H, Lee D, Park JH, Lee I-B. New Gene Selection Method for Classification of Cancer Subtypes Considering within-Class Variation. FEBS Letters. 2003; 551:3–7.
22. Rajapakse JC, Mundra PA. Multiclass Gene Selection Using Pareto-Fronts. IEEE ACM Trans Comput Biol Bioinformatics. 2013; 10(1).