A TOPSIS based Method for Gene Selection for Cancer Classification

I.M.Abd-El Fattah ,W.I.Khedr, K.M.Sallam,
[1] Department of Statistics, [3]Department of Decision support, [2]Department of information technology
[2,3] Faculty of Computers and Informatics, [1] Faculty of Commerce (Operation research)
Zagazig University

ABSTRACT
Cancer classification based on microarray gene expressions is an important problem. In this work a new gene selection technique is proposed. The technique combines TOPSIS (Techniques for Order Preference by Similarity to an Ideal Solution) and F-score method to select subset of relevant genes. The output of the combined gene selection technique is fed into four different classifiers resulting in four hybrid cancer classification systems. In the proposed technique some important genes were chosen from thousands of genes (most informative genes). After that, the microarray data sets were classified with a K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM) and Naive Bayes (NB).The goal of this proposed approach is to select most informative subset of features/genes that give better classification accuracy.

General Terms
Feature selection, classification, machine learning techniques, and data mining.

Keywords-component
TOPSIS, Gene Selection, Cancer classification, Neural Network, Decision Tree, Naive Bayes and K-Nearest Neighbour.

1. Introduction
The goal of microarray data classification is to build an efficient and effective model that can discriminate between normal and abnormal conditions, or classify tissue samples into different classes of diseases. However, gene selection process is a major problem because of the characteristics of microarray data; the huge number of genes compared to the small number of samples (high-dimensional data), irrelevant genes, and noisy data.

In other words, the problem of microarray data is the small amount of samples against the huge amount of features (genes in our case). To overcome this challenge, a gene selection method is used to select a subset of genes that increases the classifier's ability to classify samples accurately. The gene selection method has several advantages, such as improving classification accuracy, reducing the dimensionality of data, and removing irrelevant and noisy genes [1].

In order to improve the performance of classifier algorithms, it is preferred to reduce the dimensionality of the data by deleting redundant, irrelevant or noisy features [1]. In addition, selection of an optimal subset of features by exhaustive search is impractical and it consumes much time as the number of attributes increases, and a proper learning strategy must thus be devised. The relevance of good feature selection methods has been discussed in [2], but the recommendations in literature do not give evidence for a single best method for either feature selection or classification of microarray data [3].Many feature selection algorithms have been proposed in literature. All these methods search for optimal or near optimal subsets of features that optimize a given criterion. The rest of the paper is organized as follows. In Section 2, we provide background information on microarray data analysis and discuss some related works. Section 3 we discuss the feature selection types filters, wrappers and embedded models. Section 4 contains a detailed discussion of the proposed approach. Section 5 contains a detailed discussion as well a comparison between the proposed approach and several state-of-the-art methods. Finally, in Section 6 we conclude with some final remarks and suggest future research directions.

2. Microarray data
Microarray analyses were motivated by the search for useful patterns of tumour classification, disease state classification, discovery of new subtypes of disease or disease states, among others. Microarray data are generally having large number of features/genes in comparison to the number of samples or conditions (high dimensional data). There are many efficient methods for the analysis of microarray data such as clustering, classification and feature selection. Microarray technology measures the expression level of genes. That can be used in the diagnosis, through the classification of different types of cancerous genes leading to a cancer type [4]. A microarray data is commonly represented as an M × N matrix, where M is the number of samples and N is the number of genes in the experiment. Each cell in the matrix, as shown in Fig. 1, is the level of expression of a specific gene in a specific experiment. The last column to the right shows the label or class that each sample belongs to.

Microarray studies often involve the use of machine learning methods such as unsupervised and supervised learning. Unsupervised learning finds subgroups in the data based on a similarity measure between the expression profiles. Unsupervised learning studies how systems can learn to represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns[5]. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which is called a classifier [6].

	Gene1	Gene2	...	Gene N	Class
Sample1	F11	F12	...	F1N	C1
Sample2	F21	F22	...	F2N	C2
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
SampleM	FM1	FM2	...	FMN	CM

Figure 1: Matrix representation of a microarray dataset.

The expression profile of a sample i can be represented as a row vector

$$G_i = [\, f_{i1}, f_{i2}, \dots, f_{in}]$$

The expression profile of a gene j can be represented as a row vector

$$G_j = \begin{bmatrix} f_{1j} \\ f_{2j} \\ f_{mn} \end{bmatrix}$$

3. Feature Selection Methods

Feature selection techniques study how to identify and select most informative features for building classification models which can interpret data better. Feature selection reduces dimensionality of data, so the cost of computation is reduced. It also improves the classification accuracy and the comprehensibility of the models by eliminating redundant and noisy features. Feature selection and feature extraction are different. The former selects the most informative genes and maintains the original meanings of the selected features, which is desirable in some domains, while the latter creates new features by combining the original features.

The feature selection algorithms that have been proposed in literature search for optimal or near optimal subsets of features that optimize a given criterion.

Feature selection methods divided into three classes: filters, wrappers, and embedded methods based on whether the selection method depends on the classifier or not [2].

3.1 Filter Methods

Filter methods select subsets of features as a pre-processing step, independently of the classifier [7]. Features are rated by their discriminative powers with regard to targeted classes.

Feature ranking approaches are used rank features by certain ranking criterion. This ranking is used as the base of selection mechanism. In the proposed feature selection methods, feature ranking approach is attractive because it is simple, scalable, and shows good empirical success. Computationally, feature ranking is efficient since it requires only the computation of n scores and sorting these scores.

As is shown in Fig. 2 the filters have three main phases: feature subset generation, measurement, and testing. In feature subset generation stage, a feature subset is generated. Next the measurement step is performed, which measures the information of the current feature subset. These two steps will be performed iteratively until the measurement matches the stop criterion.

The final feature subset would contain the most informative features. Finally, the testing step is done by a classifier algorithm, like Decision tree (DT) [8], Naive Bayes (NB) Classifiers [8], K-Nearest neighbour (KNN) [9] or Support vector machine (SVM) [10].

3.2 Wrapper Methods

Wrapper methods use the classifier algorithm of interest to evaluate subsets of features according to their classification accuracy [7] .Wrappers can often find a small subset of feature with high accuracy because the features match well with the learning algorithms. Even though, wrappers typically require more computation effort, it is argued that filters have better generalization properties since it is independent of any specific classifier.

Fig. 3 represents the procedure of wrappers. It is the same as that of filters except that the measurement stage is replaced by a classifier algorithm. The wrapper model uses the predictive accuracy of a predetermined mining algorithm to give the quality of a selected feature subset, generally producing features better suited to the classification task at hand. However, it is time consuming for high dimensional data [2].

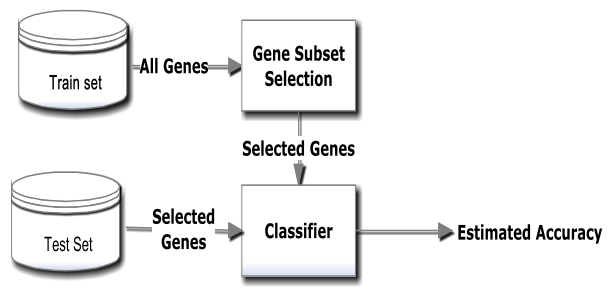

Fig.2. the Filter Method

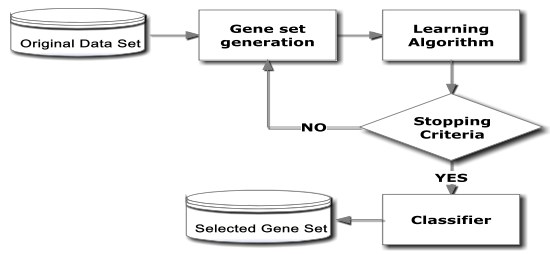

Fig3.The wrapper Method

3.3 Embedded Methods

Embedded methods use feature subset search and evaluation in the process of building a classifier model as shown in Fig. 4. The search is guided by the learning algorithm [2]. Embedded methods are computationally less intensive than wrappers and they account for feature dependencies.

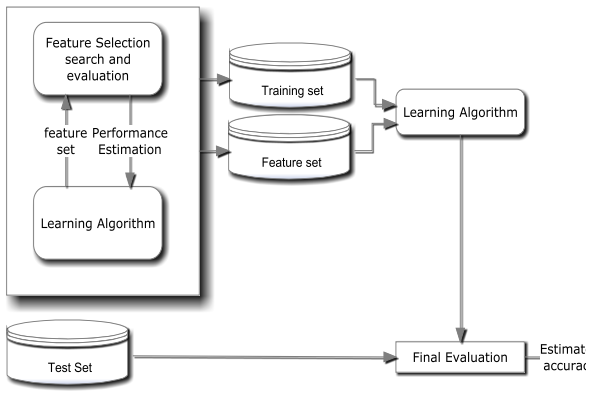

Fig4. The Embedded Method

4. Technique for Order Performance by Similarity to Ideal Solution TOPSIS

According to this technique, the best alternative would be the one nearest to the positive ideal solution and farthest from the negative ideal solution [11] . The positive ideal solution is a solution that maximizes the benefit criteria and minimizes the cost criteria, whereas the negative ideal solution is a solution that maximizes the cost criteria and minimizes the benefit criteria. In brief, the positive ideal solution is composed of all best values attainable from the criteria, while the negative ideal solution consists of all worst values attainable from the criteria [12].

The TOPSIS method consists of the following steps:

Step 1: Establish a decision matrix for the ranking. The structure of the matrix can be expressed as follows:

$$x = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{pmatrix}$$

Fig.5. the Matrix representation

where rows denotes the genes i, i = 1,2, . . ., m; column represents j^{th} sample, j = 1, 2,. . . , n; and x_{ij} is a gene expression level of gene i and a sample j.

Step 2: Calculate the normalized decision matrix R= $[r_{ij}]$. The normalized value r_{ij} is calculated as:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum x_{ij}^2}} \qquad i = 1,2,\dots,n\,;\, j = 1,2,\dots,m$$

Step 3: Calculate the weighted normalized decision matrix by multiplying the normalized decision matrix by its associated weights.

The weighted normalized value v_{ij} is calculated as:

$$v_{ij} = w_j \times r_{ij} \quad i = 1,2,\dots,n\,;\, j = 1,2,\dots,m$$

Where wj represents the weight of the jth sample. Here I calculate w_j using geometric mean by using the following formula:

$$w_j = \sqrt[m]{\prod_{i=1}^{n} x_{ij}}$$ Where j=1, 2... m.

Step 4: Determine the positive-ideal and negative-ideal solutions.

$$A^* = \{v_1^*, v_2^*, \dots, v_n^*\} = \{\left(\max_i v_{ij} \,|i \in I'\right), \left(\min_i v_{ij} \,|i \in I''\right)\}$$

$$A^- = \{v_1^-, v_2^-, \dots, v_n^-\} = \{\left(\min_i v_{ij} \,|i \in I'\right), \left(\max_i v_{ij} \,|i \in I''\right)\}$$

Step 5: Calculate the separation measures, using the n-dimensional Euclidean distance. The separation of each gene from the positive-ideal solution (D_i^*) is given as:

$$D_i^* = \sqrt{\sum_{j=1}^{n} (v_{ij} - A_j^*)^2} \quad i = 1,2,\dots,m$$

-Similarly, the separation of each gene from the negative-ideal solution (D_i^-) is as follows:

$$D_i^- = \sqrt{\sum_{j=1}^{n} (v_{ij} - A_j^-)^2} \quad i = 1,2,\dots,m$$

Step 6: Calculate the relative closeness to the ideal solution and rank the performance order. The relative closeness of the gene Aj can be expressed as:

$$CC_i^* = \frac{D_i^-}{D_i^* + D_i^-} \quad i = 1,2,\dots,m$$

-Where the CCi* index value lies between 0 and 1. The larger the index value means the better the performance of the gene.

5. Method

5.1 Colon cancer dataset

The Colon data set contains 62 samples collected from cancer patients. Among them 40 tumours and 22 normal are from healthy parts of the colons of the same patients as shown in table 1. Two thousands out of around 6500 genes were selected based on the confidence in the measured expression levels.

Table 1.Colon data set

Dataset	#of Genes	# of samples	# of classes	# of positive samples	#of negative samples
colon	2000	62	2	22	40

5.2 Gene selection methods

Not all genes in microarray data needed for classification. Microarray data consist of large number of genes in small samples. So some genes were needed to be chosen (most informative genes) [13]. This process is referred to as gene selection (feature selection) in machine learning.

5.2.1 T-test statistics

The t-test [14],[15] is a statistical hypothesis test used to determine the significance of the difference between the means of two independent samples. It assumes normally distributed populations. For unequal variance and unequal or equal sample sizes the t statistic is calculated as follows:

$$t = \frac{x_1^- - x_2^-}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where x_1^- and x_2^- are the means of the two samples; s_1^2 and s_2^2 are the variance estimates of the two samples; n1 and n2 are the sample sizes of the two samples.

The degrees of freedom df can be calculated as

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

The t distribution can be used with t statistic and df degrees of freedom as parameters to calculate the corresponding p-value. Features are then ranked according to p-values. The smaller the p-value the more the relevant the feature.

5.2.2 Signal-to-Noise (SNR) based Gene Selection Method

SNR is a calculated ranking number for each gene to define how well this gene is most informative [16]. For each gene, this work has normalized the gene expression data by subtracting the means of the two classes and then dividing by summation of the standard deviation of the two classes. Every sample is labelled as either a normal or a cancer sample. Signal-to-Noise Ratio (SNR) is computed as follows.

$$SNR = \frac{|\mu_i - \mu_j|}{\sigma_i + \sigma_j}$$

The SNR is the value of the ith gene; μ_i and μ_j denote the averaged expression intensities of the ith gene over the samples each belonging to normal case and tumour case respectively. Also, σ_i and σ_j denote the standard deviations of normal case and tumour case.

5.2.3 Fisher's ratio (FR)

Fisher-ratio [17] is a ratio between-class distances to with-in class distances, and it is defined as:

$$FR = \frac{(m_1 - m_2)^2}{(v_1 + v_2)}$$

Where m_1 and m_2 are the average of the expression level of a particular gene in normal class and tumor class; v_1 and v_2 are the corresponding variances. In the experiments, given a microarray dataset, we compute the FR value for each gene and rank them in descending order. Genes with higher ranking have more discriminative power for classifying samples into categories.

5.2.4 Information gain (IG)

Information Gain (IG) is a filter method used to pre-select genes and finally produces a subset of informative genes [18]. Information gain is given by

$$infoGain = H(Y) - H(Y|x)$$

Where X and Y are features, and

$$H(Y) = -\sum_{y \in Y} p(y) \log_2\big(p(y)\big).$$

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2(p(y|x)).$$

In our case, given a microarray dataset, we compute the information gain for each gene and rank them in descending order by their information gain value. The higher information value a gene has, the more effective the gene used to classify the training data.

5.3 Classifiers

5.3.1 Decision Trees (DT's)

Decision tree classifier is a method commonly used in machine learning. It aims to create a model that predicts the value of a target variable based on several input variables [6],[8],[19]. It is a tree where each non-terminal node represents a test or decision on the considered data item. Choice of a certain branch depends upon the outcome of the test. To classify a particular data item, we start at the root node and follow the assertions down until we reach a terminal node (or leaf node). A decision is made when a terminal node is approached [20].

5.3.2 Support Vector Machine (SVM)

Support vector machine (SVM) is a new generation learning system based on recent advances in statistical learning theory [10]. It is an algorithm that attempts to find a linear separator (hyper-plane) between the data points of two classes in multidimensional space. SVMs are well suited to dealing with interactions among features and redundant features.

5.3.3 Neural Networks (NN)

Neural networks (NN) [6] are those systems modelled based on the human brain working. As the human brain consists of millions of neurons that are interconnected by synapses, a neural network is a set of connected input/output units in which each connection has a weight associated with it.

The network learns in the learning phase by adjusting the weights so as to be able to predict the correct class label of the input [21].

5.3.4 K-Nearest Neighbor (KNN)

The purpose of the algorithm is to classify a new object based on attributes and training samples. The K-NN method consists of a supervised learning algorithm where the result of a new query instance is classified based on the majority of the K-nearest neighbour category. The K-NN method is easy to implement since only the parameter K (number of nearest neighbours) needs to be determined. The parameter K is an important factor affecting the performance of the classification process.

K-NN classifies a new sample based on the minimum distance from the test samples to the training samples. The Euclidean distance was used for calculations in this paper [13]. If an object is near to the K nearest neighbours, the object is classified into the K-object category.

5.3.5 Naive Bayes (NB)

The NB classifier is a probabilistic algorithm based on Bayes' rule and the simple assumption that the feature values are conditionally independent given the class [22]. Given a new sample observation, the classifier assigns it to the class with the maximum conditional probability estimate.

5.4 Proposed method

The four proposed classification modules (J48, SVM, NB and 3NN) receive pre-processed original high dimensionality microarray dataset as its input. The system first step is reducing the total number of genes in the input dataset to small subset using TOPSIS Method and F-score test technique as a combined gene selection technique. Then the selected data will be the data used by the chosen classifier to assign new samples into their correct classes instead of using the original full data set. At this point we can measure and record the test accuracy of the classifier. The workflow of the proposed approach is shown in Fig. 5.

In the proposed approach, first we rank gene according to TOPSIS Method and select 250 top ranked genes from this method and after that we use F distribution as follows:

- Calculating the mean of the expression values for each of the n genes (μn1 for the first class and μn2 for the second class).
- Calculating the standard deviation of the expression values for each of the n genes (σn1 for the first class and σn2 for the second class).
- Obtaining the absolute differences between the calculated means (| μn1 - μn2 |).
- Calculating the F value for each of the selected genes from TOPSIS METHOD by the following:
  (F = |μn1 - μn2 |/(σn1 + σn2)).
- Arrange the genes according to their F value and choose the most informative genes.

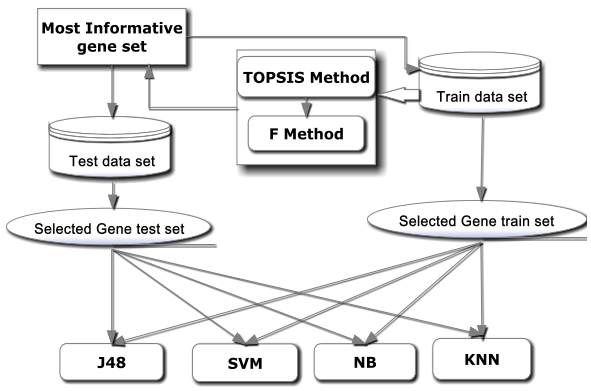

Fig5. Proposed Method

After finishing the gene selection step we use four different classifiers to test the accuracy of the proposed method. One of the proposed systems achieved the highest classification accuracies on the used dataset with a considerable number of genes. Although the others proposed system couldn't reach the same results, it contributes in validating the proposed gene selection technique.

6. Experimental Result

- We started experiment by evaluating performance accuracies of the classifiers, SVM, KNN, and decision tree (J48) on the dataset using 10-fold Cross Validation (CV) without using feature selection algorithm. The result for the four classifiers is shown in table 3.
- After feature selection, the selected feature subsets were evaluated using four common classification algorithms SVM, KNN, J48, and NB using 10-fold CV.
- We test the accuracy of the proposed approach using different number of genes as showed in table 3 and fig.6 using 10-fold cross validation. We see that when using data set which has 100 genes we see that KNN has the highest accuracy than other method. When using 50 and 20 gene data set we see that SVM has the highest accuracy than the other methods. When the number of reduced data set is 10 SVM and 3NN have higher accuracy than other classifiers.
- When comparing the proposed approach using 10 genes and using SVM or 3NN as classifiers with the method used in [23] ,[24] ,[25] we see that the proposed approach give higher classification accuracy than others as shown in table 4 and Fig.7.

Table 2 accuracy of the classifiers with the original Colon data set, and the reduced data set using 10 fold cross validation.

	Classifiers			
# of used genes	**SVM**	**3NN**	**J48**	**Naïve Bayes**
Original data set	85.4	79.1	82.3	60.2
100	74.2	85.5	80.6	70.9
50	87	85.4	77.4	72.5
20	87.1	85.5	82.3	79.1
10	88.7	88.7	82.3	85.4

Table 4. Comparison of the proposed method and other method used in [23], [24] and [25].

	Method				
	Proposed with SVM	**Proposed with 3NN**	**[23]**	**[25]**	**[24]**
Colon	88.7	88.7	85.4	80.08	83.05

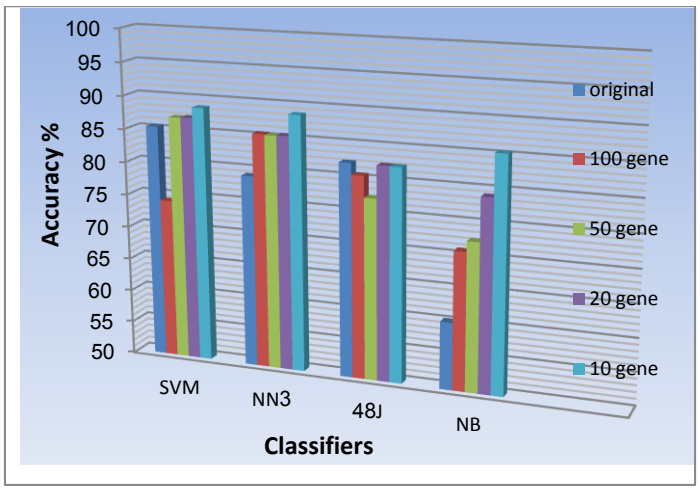

Fig.6 compared result of colon data set

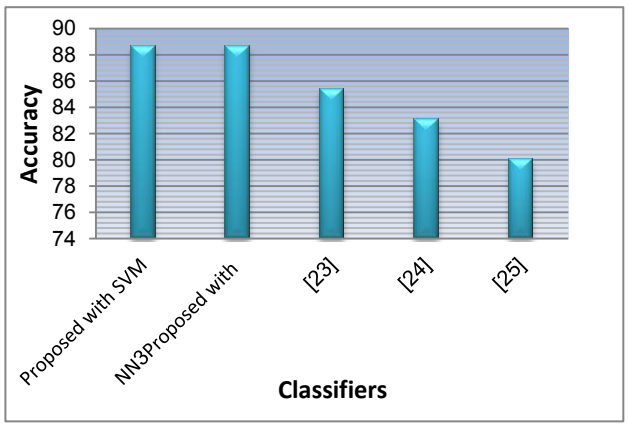

Fig.7 comparison of our method with [23, 24, 25]

7. Conclusion and future work

The process of classifying microarray data must enclose two main steps; implementing an effective gene selection technique and choosing a powerful classifier. These two steps form the

workflow of this paper trying to go through each step details and validating its outputs.

The goal of the proposed approach is to extract informative outcome from feature selection method to find a subset of informative features that have smaller size and better classification accuracy compare to individual method. In this paper, we introduced four new hybrid classification systems for classifying cancer samples using microarray gene expression datasets. The main target of the proposed systems is to get the highest accuracy when classifying the samples using a small subset of informative genes. A combination of two gene selection techniques is introduced to solve the problem of the microarray high dimensionality. This combined technique presents high performance as it reduces the number of genes. SVM, KNN, J48 and NB were chosen for classification as they are very efficient binary classification techniques and usually give good results by attenuating their variety of attributes. The four systems were a result of integrating the proposed gene selection technique once with SVM resulting in the First CS, another time with KNN resulting in the Second CS, another time with J48 resulting in third CS and another time with NB resulting in fourth CS. The four systems CS were applied public microarray datasets colon. When the number of reduced gene data set is 10 we see that the accuracy is higher than the other reduced data set number. Comparison with others verifies the efficiency of the proposed method.

8. References

[1] I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann, 2005.

[2] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," Bioinformatics, vol. 23, pp. 2507-2517, 2007.

[3] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," Bioinformatics, vol. 21, pp. 631-643, 2005.

[4] A. C. Lorena, I. G. Costa, and M. C. de Souto, "On the complexity of gene expression classification data sets," in Hybrid Intelligent Systems, 2008. HIS'08. Eighth International Conference on, 2008, pp. 825-830.

[5] M. A. Ranzato, F. J. Huang, Y.-L. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, 2007, pp. 1-8.

[6] S. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," Frontiers in Artificial Intelligence and Applications, vol. 160, p. 3, 2007.

[7] B. Krishnapuram, L. Carin, and A. Hartemink, "1 Gene expression analysis: Joint feature selection and classifier design," Kernel Methods in Computational Biology, pp. 299-317, 2004.

[8] Y. Lu and J. Han, "Cancer classification using gene expression data," Information Systems, vol. 28, pp. 243-268, 2003.

[9] Y. Song, J. Huang, D. Zhou, H. Zha, and C. Giles, "Iknn: Informative k-nearest neighbor pattern classification," Knowledge Discovery in Databases: PKDD 2007, pp. 248-264, 2007.

[10] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Machine learning, vol. 46, pp. 389-422, 2002.

[11] G.-W. Wei, "Extension of TOPSIS method for 2-tuple linguistic multiple attribute group decision making with incomplete weight information," Knowledge and information systems, vol. 25, pp. 623-634, 2010.

[12] T. Yang and C.-C. Hung, "Multiple-attribute decision making methods for plant layout design problem," Robotics and computer-integrated manufacturing, vol. 23, pp. 126-137, 2007.

[13] S.-B. Cho and H.-H. Won, "Machine learning in DNA microarray analysis for cancer classification," in Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics 2003-Volume 19, 2003, pp. 189-198.

[14] G. D. Ruxton, "The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test," Behavioral Ecology, vol. 17, pp. 688-690, 2006.

[15] D. C. Montgomery, G. C. Runger, and N. F. Hubele, Engineering statistics: Wiley, 2009.

[16] J.-H. Hong and S.-B. Cho, "Lymphoma cancer classification using genetic programming with SNR features," Genetic Programming, pp. 78-88, 2004.

[17] K. Mao, P. Zhao, and P.-H. Tan, "Supervised learning-based cell image segmentation for p53 immunohistochemistry," Biomedical Engineering, IEEE Transactions on, vol. 53, pp. 1153-1163, 2006.

[18] G. Forman, "An extensive empirical study of feature selection metrics for text classification," The Journal of Machine Learning Research, vol. 3, pp. 1289-1305, 2003.

[19] W. Du and Z. Zhan, "Building decision tree classifier on private data," 2002.

[20] G. Stiglic, S. Kocbek, I. Pernek, and P. Kokol, "Comprehensive Decision Tree Models in Bioinformatics," PloS one, vol. 7, p. e33812, 2012.

[21] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," Neural networks: the official journal of the International Neural Network Society, vol. 21, p. 427, 2008.

[22] K. Al-Aidaroos, A. Bakar, and Z. Othman, "Medical Data Classification with Naive Bayes Approach," 2012.

[23] B. Liu, Q. Cui, T. Jiang, and S. Ma, "A combinational feature selection and ensemble neural network method for classification of gene expression data," BMC bioinformatics, vol. 5, p. 136, 2004.

[24] T. Paul and H. Iba, "Identification of informative genes for molecular classification using probabilistic model building genetic algorithm," in Genetic and Evolutionary Computation–GECCO 2004, 2004, pp. 414-425.

[25] J. Liu and H. Iba, "Selecting informative genes using a multiobjective evolutionary algorithm," in Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on, 2002, pp. 297-302.