RFE and mutual info based binary particle swarm optimization for gene selection and cancer classification.

Samkit Jain,Rashmi Maheshwari, Vinod kumar Jain,aadhar gupta and Kusum Kumari Bharti
PDPM Indian Institute of Information Technology, Design and Manufacturing
Jabalpur , MadhyaPradesh

(1911603,1921601,vkjain,1911001,kusum)@iiitdmj.ac.in

ABSTRACT

DNA Microarray technology can be used for simultaneous learning of the expression levels of thousands of genes. It finally produces microarray data that contain information of biological, diagnostic, and prognostic use for researchers. It is generally used for cancer classification problems. Identification of the informative and relevant genes has always been an important step in microarray data analysis. The cancer datasets contains more number of features and less number of samples which increases the complexity of dataset. The main problem addressed in the section is the selection of small subset of relevant from the thousands of genes which contribute for causal of disease. The process of selection becomes more difficult as samples are very less and genes are more which are irrelevant and noisy in nature. This selection process is difficult due to the availability of a small number of sample compared with the huge number of genes, many of which are irrelevant and noisy. Therefore, here we propose a two step method "RFE-INFO-BPSO" which is combination of wrapper methods. Firstly we apply Recursive Feature Elimination(RFE) and Mutual Information to reduce the dimensionality. Then apply binary particle swarm optimization to select a near-optimal(small) subset of informative genes and than dataset is reformed based on the BPSO gene subset. The above reformed dataset is used for the cancer classification using various classifiers like Support Vector Machine(SVM), Naive Bayes(NB) Classifier, Decision Tree (DT) and Neural Network(NN). Experimental results show that the performance of the proposed method is better than other classification methods in terms of accuracy and the number of selected genes.

Keywords: Cancer classification, Microarray data analysis, RFE, Mutual Information, BPSO, classification method

1. INTRODUCTION

In the recent years, machine learning has evolved to be an extremely powerful tool for data analysis in the field of computer science, biological data, automation of real-time applications etc. The various applications of machine learning in recent decade in biological field are- chemical formulations, drug discovery, tumor classification etc. Prediction models are very interesting applications of machine learning which are being used in different biological applications [1]. Statistical models are developed for identification and classification of cancer causing tissues and normal tissues. These models use gene expression profile which is one of challenging problems [2].

The branch of feature selection that selects subset of relevant and significant genes for cancer classification and prediction is called as Gene Selection process. The several issues related with microarray datasets as as follows [3]: (i) Selection of subset of informative genes from high-dimensional dataset is non-deterministic polynomial-time (NP)-Hard problem. Thus, evolutionary methods and bio-inspired algorithms are used widely. (ii) The second issue is curse of sparsity, in which number of samples are small and number of features (genes) are very large which makes data sparse. (iii) The third issue is the high complexity of gene expressions data which arises due to high correlation between genes and interactions among them. This makes process of selecting highly relevant genes a challenging task. The process of identifying infected genes will be easier if the number of informative genes will be lesser [4, 5].

Many gene selection methods have already been proposed by various authors to solve the challenges. These challenges can be roughly divided into three categories as filter model, wrapper model and hybrid model [6]. The filter model relies on the statistical characteristics of training data rather than an learning algorithms. These methods can be fast be has poor performance.

The wrapper methods use a well defined learning algorithm to find the optimized subset of genes/features. These models are generally bio-inspired or are evolutionary which depends upon its population. The population requires to be evaluated using established learner which can be improved in various iterations. The computational cost of this approach is high for high dimensional dataset [7, 8]. The performance of wrapper approaches are considered better than filter methods since there is interaction between solution and predictors. The main benefit of wrapper method is

it wraps up both search space and classification in the same method. The search method explores all the subset possible from given set of genes [9].

The SVM-RFE is one of the wrapper method used widely for gene selection in biomedical fields. The application of SVM is limited in biomedical data as it is not designed to evaluate prediction models and predictor variables [10]. The method is applicable for both binary classes and multi-class classification [11, 12]. The method is applied to reduce the dataset dimensionality and and apply as input to BPSO algorithm.

Over past few years, Particle Swarm Optimization (PSO) has been applied to many genes data for their prognostic detection [13, 14, 15]. It uses the intelligence of swarm and given fitness criteria to find the solution of an optimization problem by generating better subset of features. After a fixed number of iterations, it can quickly converge towards a global optimum. The group of data points are treated as swarm of particles, which are allowed to navigate in the search space with the controlled velocity. After each iteration, every particle updates its position and velocity to achieve global optimum solution. The Binary version of PSO (BPSO) for discrete problems was given by Kennedy and Eberhart [16] in which particles are represented having position digit 0 or digit 1. the sigmoid probability function is then used to assign new position and velocity to all the particles.

Many machine learning classifiers are available that produce good classification accuracy like SVM, Decision Tree, Naive bayes etc. Among all classification methods, apart from machine learning model deep learning models also show good performance and draw the more attention. These models intrinsically learn a high level representation of the data so they avoid the laborious work [17]. Also they have stronger expressive power than conventional shallow structure. Computer vision, speech recognition are the some field where Deep learning models achieved the promising performance. Deep learning models have also been widely used in the area of bioinformatics, including biomedical imaging [18]. However, the application of deep learning is rare in tumor classification because here we have very small number of sample to train the model.

1.1 Organization

In the following section, we discuss about the related work from the past. Section Proposed Methodology discusses the approach of finding optimal relevant genes and applying the classification models for predictions. The subsections included explains about the experimental setup and the dataset used. In the following section results of the experiments are discussed. The section Conclusions And Future Work concludes this work.

2. RELATED WORK

DNA microarray technology is a prospective and a clinical tool for diagnosis of cancer and its classification. A greater challenge to the microbiology is the measurement of thousands of genes for biological samples using microarray. To solve the issue of high dimensionality and get better results from experiments in past decades many statistical and machine learning methods have been applied to get most relevant genes for cancer classification. These models are divided into supervised and unsupervised learning models.

In unsupervised models a hierarchical clustering [19], two-way clustering [20] and self-organizing maps (SOMs)[21] algorithms are used for classification of different types of tumor.

In supervised classification, traditional statistical methods (discriminant analysis, Gaussian and logistic classifiers, etc.) [22, 23, 24], and various machine learning techniques like Support Vector Machines (SVMs) [25, 26, 27], Neural networks (ANNs) [28, 29], K-nearest neighbor (K-NN)[21, 23, 14] are widely used class prediction techniques in cancer classification.

The high dimension data having small number of samples and large number of features(genes) are microarrays [9]. To achieve the improvement in classification, analysis and cost efficient various gene selection methods have been proposed and defined. They help to select relevant genes, removing noisy and redundant genes.

For feature selection univariate filters and multivariate filters are used. Univariate filters uses the P-metric [21, 29] and t-score [30]. But they are less efficient than multivariate models. Examples of multivariate filters are Correlation-based Feature Selection (CFS) [4], Fast Correlation-Based Filter (FCBF) [34], Minimum Redundancy-Maximum Relevance (mRMR) [31], Uncorrelated Shrunken Centroid (USC)[32] algorithm, ReliefF [33], Weight Local Modularity [34] etc.

Wrapper approaches perform better than filter methods in terms of feature selection by evaluating the classification accuracy using some induction algorithm [35]. But these methods require greater computational cost and suffer risks of overfitting. Some authors have performed influential studies and implemented randomized and population based wrappers like Particle Swarm Optimization (PSO)[13, 36, 37, 15] and Genetic Algorithms (GA)[14, 38, 39] for gene selection purpose. Xi et al [40] used Binary Quantum-Behaved Particle Swarm Optimization for gene selection purpose. Huang et al. [41] for diagnosis of breast cancer proposed a new fruit fly optimization algorithm enhanced support vector machine.

Recently, hybrid framework has been used by the authors which combines the benefits of both filter and wrapper methods to reduce the computational cost and overfitting issue. A pre-treatment is provided with the filter algorithm for getting better results of cancer classification. Ruiz et al. [42] proposed a hybrid algorithm BIRS (Best Incremental Ranked Subset) which reduces the number of genes but with low classification accuracy. Zhu et al. [43] presented Markov Blanket-Embedded Genetic Algorithm (MBEGA) for gene selection. Shen et al.[13] developed tabu search and PSO using the hybrid framework on microarray data, but the results were not efficient. Li et al.[36] also proposed a new hybrid of PSO and Genetic Algorithms (GA) but the results were not satisfactory. They had low accuracy and high number of selected genes.

SVM-RFE is the method used by various authors for gene selection purpose. It is backward elimination algorithm. The wrapper methods produce better results than filter methods. Thus, the main purpose of RFE is to utilize non-linear kernels and produce good results for classification [10, 11]. The method produce the ranking such that the most relevant genes must not be eliminated and are specifically ranked subsets. This produce a highly optimized subset of genes for better performance.

Further Chuang et al. [44] applied gene selection and cancer classification using binary PSO (BPSO). BPSO is an improvement in the PSO algorithm done for discretiation[16]. The values for velocity and position are improved with iterations using knowledge from the swarm. For many datasets, they found lower classification error rates. Also the method is cumbersome taking more time to execute. The methods suffer from higher cost of computation cost and high risk of overfitting.

Indu Jain et.[9] al propose a hybrid model for gene selection and cancer class determination which attempts to address the drawbacks mentioned above.It integrates the benefits of fast and efficient dimensionality reduction of multi-variate filter (CFS) and Binary Particle Swarm Optimization approach.This hybrid model operates in two phases. In the first phase gene selection is performed to select a subset of genes by CFS, than gene optimization is done using improved BPSO-NB wrapper method. The improved BPSO (iBPSO) also eliminates the problem of local convergence of traditional BPSO. The main objective of the work was to achieve better classification accuracy using fewer number of highly predictive genes.

3. PROPOSED METHODOLOGY

The Figure 1 shows the overall structure and pseudo-code of the RFE-INFO-BPSO approach respectively. DNA microarray experiments have been proven to be a great boom in the field of microbiology. The experiments performed on them can count thousands of gene expressions related to tissue samples. These data is stored in the form of matrices of microarray by computer scientist to perform experiments and get results which can help for further exploration. Suppose there are N samples of biological tissues in the dataset containing DNA microarray. The row vector of the dataset can be represented as (SM_j, C_i) where $SM_j \epsilon EM$ which is set of combination of total M expression genes for N samples. $C_i \epsilon 1, 2, 3, ..., l$ denotes the class label associated with SM_j gene expression profile. For dataset M is very large for N samples (M>=N). The dataset is the composed from M-dimensional gene expression profile having N samples, having l target classes for the given samples. There exist association between genes expressions and the target classes to which is used for the experiments.

The dimensionality of the dataset is high thus, it also consist of irrelevant, redundant and noisy genes in the expressions. The higher dimension of dataset does not only makes the computation process complex but also degrades performance of computation of many learning algorithms. Hence, feature selection techniques are required to minimize irrelevant and redundant gene expression from the dataset and available data.

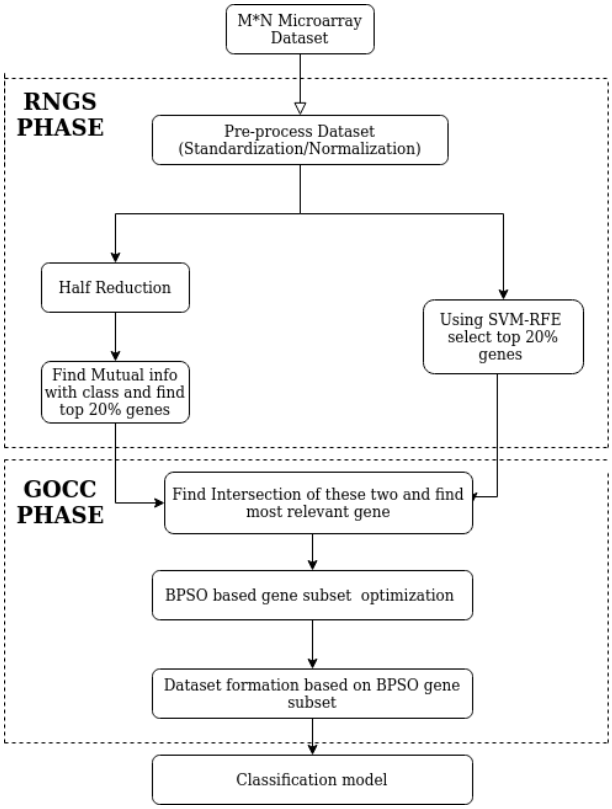

Figure 1: View of RFE-INFO-BPSO method

The proposed methodology works in two phases which are as follows and described in the subsections below -

- **Relevant and Non-redundant Gene Selection (RNGS) phase**- It consist of four steps as- data pre-processing, half reduction, mutual information with class and application of SVM-RFE. These steps are discussed in detail in the following subsection.
- **Gene Optimization and Cancer Classification (GOCC) phase**- In this phase, application of BPSO is done to get the optimal number of relevant genes.

3.1 RNGS phase

The microarray dataset has the property of having low samples and high dimensional features. The main task is to select low dimensional subset of biological gene which are highly relative to class for the prediction efficiently. The cancer classification is highly sensitive towards the classification process. It depends upon the fact that out of high dimensional genes subset most relevant subset must be identified for detection of disease and have high accuracy.

In the data pre-processing step, the missing values of genes are replaced with their mean values [9].

The RNGS phase consist of four steps defined as follows. In this phase, the first step is data pre-procesing which helps to deal with the missing values of the dataset. In the second step, we do the half reduction of the pre-processed received

from previous data. In the third step, the half reduced data of previous step is applied as input to find mutual information of genes with the target class. In the fourth step, method SVM-RFE is applied to produce the result of RNGS phase.

- **STEP 1- Data Pre-processing :**In this phase, we start with a data pre-processing step where in each dataset, we replaced the missing values of gene expressions with mean values. The standardization of whole dataset is done with respect to having a mean value equal to zero and standard deviation equal to one by using following equation [9]:

$$g^{new} = \frac{g-\mu}{\sigma}$$

Here g^{new} represent the transformed value for a gene expression g, μ is the mean value and σ is the standard deviation for the given vector.

- **STEP 2- Half Reduction :** After getting normalized pre-processed data, the correlation matrix of size M*M is generated, which gives the correlation of each gene with every other gene. Now we reduce M genes into $\frac{M}{2}$ genes, that is reduce their number to 50% (reduce to half). For performing this reduction, we find the second highest correlation value of genes with other genes. The reason for choosing second highest correlation value is the first highest correlation of gene will always exist with itself that will be 1. Suppose gene M_i has second highest correlation with gene M_j (correlation between M_i and M_i will be one), we have to remove one of the gene from these two because of their high correlation value. When the genes are highly correlated, they create redundancy in the information. This process is repeated until we get the half number of genes ($\frac{M}{2}$) from total genes (M).

- **STEP 3- Mutual Information With class :** After the reduction of genes to half, we find the mutual information of each gene with the class label. Mutual Information is calculated between the two variables. It measures the reduction in the uncertainty for a variable given a known value of other variable. In other words, it is the amount of information one can obtain from one random variable given the other one. The mutual information between two random variables X and Y can be formalised as follows:

$$I(X;Y) = H(X) - H(X|Y)$$

Where I(X ; Y) is the mutual information for X and Y, H(X) is the entropy for X and H(X | Y) is the conditional entropy for X given Y. Than we select top 20% gene which have highest mutual information with class and store it separately.

- **STEP 4- Apply SVM-RFE :** It is Support Vector Machine based algorithm. It produce the great result accuracy in case of multi-class classification problems. The computation ranking weights for all features is done and sorting of the features is done according to weight vectors as the classification basis [45]. The process performs the backward removal of features. The steps for feature set selection using SVM-RFE are as follows:

  (i) Use the dataset to train the classifier.
  (ii) The ranking weights for all features are computed.
  (iii) The features with smallest weight are deleted.

  The process is repeated and performed till we select the top 20% gene using SVM-RFE and store it separately.

The selected 20% genes from STEP 3 and STEP 4 are intersected to form reformed dataset D i.e. subset G1, (G1$\subseteq$ G) which is given as input to the next phase for optimal gene selection.

3.2 GOCC phase

In the above RNGS phase, we got a reduce dimensional gene subset G1($\subseteq$ G).The reformed dataset D contains N samples,each of dimensionality m(m â M). Now for the further optimization for the gene subset we are using the wrapper approch Binary Particle Swarm Optimization (BPSO). Particle Swarm Optimization (PSO) is a wrapper approach that is motivated by the simulation of social behavior of the organisms such as bird flocking and fish schooling. It was developed by Kennedy and Eberhart in (1995)[9]. BPSO is the discrete version of the original PSO.
In a PSO, a swarm of particles moves in a feature space to find an optimal solution for a given objective function with in a limited number of iterations. Each particle has two vector- (i) first position vector and (ii) second one is velocity vector for directing its movement in the search space. In each iteration, every particle updates its current position and velocity based on the best positions obtained by the particle itself or by the whole swarm denoted as p^{best} and g^{best} respectively.

Suppose, in a swarm there are W particles and vector X_i denotes a gene subset of n dimensions. The particles in the swarm is defined as a string of n binary values where a bit value 1 denotes the presence of that gene whereas a bit value 0 denotes absence of the gene in the subset. The position and velocity vector of the k-th particle at t-th iteration is represented as [9]:

$$X_k^t = (x_{k1}^t, x_{k2}^t, x_{k3}^t, \ldots, x_{km}^t).\ \text{k=1,2,...W; t=1,2,...T}$$
$$V_k^t = (v_{k1}^t, v_{k2}^t, v_{k3}^t, \ldots, v_{km}^t).\ \text{k=1,2,...W; t=1,2,...T}$$

Now, in each iteration each particle updates its position and velocity based on p^{best} or g^{best}according to the following equations [9] :

$$V_{kd}^{t+1} = V_{kd}^t * \text{w} + \text{c1} * \text{r1} * (p_{best,k}^t - p_{kd}^t) + \text{c2} * \text{r2} * (g_{best}^t - p_{kd}^t).$$

$$p_{kd}^{t+1} = \begin{cases} 1, \text{if } (Sig(V_{kd}^{t+1} > r3) \\ 0, \qquad \text{otherwise} \end{cases}$$

Finally a logistic Sigmoid function is used as follows [9]:

$$\text{Sig}(V_{kd}^{t+1})=\frac{1}{(1+exp^{-}v_{kd}^{t+1})}$$

$$Sig(V_{kd}^{t+1}) \in [0,1]$$

w^t inertia weight controls the exploration of the search space in the current iteration.Inertia weight component w^t is linearly decreasing from Wmax to Wmin in each iteration t,that is calculated as following equation:

$$w^t = \Big\{ Wmax - \frac{(Wmax-Wmin)*t}{T}.$$

where X is the Number of particles, T is the Maximum number of iterations,Vmax is the Maximum velocity ,w is the inertia weight where, Wmax denote maximum bound on inertia weight, Wmin minimum bound on inertia weight, c1 and c2 are the acceleration constants in the interval [0,2], r1, r2, and r3 are random values in the range [0,1].
Also the velocity(v_{kd}^t)has limited in the range(-vmax to vmax)to limit it from overflying.

$$p_{best,k}^t = (pbest_k^1, pbest_k^2, pbest_k^3, ..., pbest_k^m)$$
$$g_{best}^t = (gbest^1, gbest^2, ..., gbest^m)$$

represent the best previous position of the i-th particle and the global best position of the swarm (all particles), respectively.

Now,The performance of the gene subset is evaluated by the neural network classifier.Neural network contains the layers of interconnected nodes.In which we have visible and hidden layer input,out Each node is a perceptron and is similar to a multiple linear regression. The perceptron feeds the signal produced by a multiple linear regression into an activation function that may be nonlinear.

In a multi-layered perceptron (MLP), perceptrons are arranged in interconnected layers. The input layer collects input patterns. The output layer has classifications or output signals to which input patterns may map. For instance, the patterns may comprise a list of quantities for technical indicators about a security; potential outputs could be âbuy,â âholdâ or âsell.â

3.3 Experimental Setup

The algorithms are implemented in Python version 3.7 and Matlab(2019). We have used Google Colaboratory and Jupyter Notebook for our experiment. The configuration of machine is as follows: Processor- intel core i5 3.30GHz*4, RAM-8GB, OS- Ubuntu 18.04 LTS, Graphics- RadeOn, Disk- 1000 GB, OS Type- 64 bits.

3.4 About Dataset

We have taken five benchmark cancer datasets of microarray gene expression data, which are obtained from `http://csse.szu.edu.cn/staff/zhuzx/Datasets.html` [43]. These dataset are Leukemia, CNN(Central Nervous System), SRBCT (Small Round Blue Cell Tumor), Lung and Lymphoma Cancer dataset. The small summary of the dataset is given in Table 1.

The descriptive summary of each dataset, includes number of observed samples, number of genes per sample and number of classes. The datasets used in the experiments have large number of dimensions (thousands of genes) and consist of two and more than two classes, which are appropriate to show the effectiveness of our approach. Thus these datasets represent binary and multi-class cancer classification problems.

Descriptive summary of microarray datasets				
Datasets	**No. of total genes**	**No. of samples**	**No. of classes**	**Class name**
SRBCT	2308	83	4	$b'1'$, $b'2'$, b'3', b'4'
Leukemia	7130	72	3	b'B-cell', b'AML', b'T-cell'
hline CNS	7129	60	2	b'1', b'0'
Lymphoma	4026	62	3	b'DLBCL', b'FL', b'CLL'
Lung	12,601	203	2	b'1', b'2'

Table 1

4. RESULTS

In this section, we evaluate our method (RFE-INFO-BPSO) on five different microarray dataset namely - SRBTC, CNS, Lymphobia, Lung and Leukemia_3c on the basis of classification accuracy and number of gene selected. After reformed dataset we predict the classification accuracy using different classifier like SVM, DT, NB and NN for binary and multi-class classification.

we randomly partition the dataset into training and testing set as 80% and 20% respectively.

The parameters used in the BPSO method of GOCC phase are listed in Table 2. The parameters listed provides the optimal number of genes for the dataset.

In Table 3 represents the average number of genes selected for the five microarray dataset obtained by RFE-INFO-BPSO (↓ W) for hundred iteration of the algorithm. The table shows that the results obtained from the proposed method are optimal for some dataset like SRBTC, CNS, Lymphoma as compared to the other methods.

In Table 4, the classification accuracy of method using classifiers SVM, DT, NB and NN.

5. CONCLUSIONS AND FUTURE WORK

In this paper,a two phase hybrid model named RFE-BPSO-NN method is proposed and evaluated. Since feature selection is more believable than feature extraction, an excellent feature selection method SVM-RFE is used along with Mutual Information to reduce the dimensionality of gene expression data. It eliminates the irrelevant and redundant genes and reduce the dimensionality of the dataset. For gene optimization and cancer classification phase, BPSO(↓W) has

Parameters for BPSO algorithm.	
Parameters	**Values**
Total number of iterations (T)	100
Swarm size (W)	30
Vmax	6
Vmin	-6
Wmax	0.9
Wmin	0.1
c1	2
c2	2

Table 2

been used that gives a low dimensional gene subset which are most relevant for cancer classification. Classification is done with the help of Neural network, Support Vector Machine, Decision Tree and Naive Bayes classifiers and predict the classification accuracy for different datasets. The proposed method can serve as a good pre-processing tool to help optimize the feature selection process. The method provides better selected genes and classification accuracy for some datasets. At the same time it tries to keep computational resources needs to be minimum.

The drawback of proposed method can be high cost and more computational resources as the method is combination of two wrapper approaches. In future, it can be tried to be applied to problems in other areas of bioinformatics, other tissue related disease and more tumorous diseases. The future scope is more work can be done on reducing the computational cost and and running time of algorithm.

6. REFERENCES

[1] J. He, M.-a. Sun, Z. Wang, Q. Wang, Q. Li, H. Xie, Characterization and machine learning prediction of allele-specific dna methylation, Genomics 106 (6) (2015) 331–339.

[2] J. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, et al., Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nature medicine 7 (6) (2001) 673–679.

[3] Y. Ai-Jun, S. Xin-Yuan, Bayesian variable selection for disease classification using gene expression data, Bioinformatics 26 (2) (2010) 215–222.

[4] V. Piras, K. Selvarajoo, The reduction of gene expression variability from single cells to populations follows simple statistical laws, Genomics 105 (3) (2015) 137–144.

[5] Y. Qi, X. Yang, Interval-valued analysis for discriminative gene selection and tissue sample classification using microarray data, Genomics 101 (1) (2013) 38–48.

[6] V. Bolón-Canedo, N. Sánchez-Marono, A. Alonso-Betanzos, J. M. Benítez, F. Herrera, A review of microarray datasets and applied feature selection methods, Information Sciences 282 (2014) 111–135.

[7] H. M. Alshamlan, G. H. Badr, Y. A. Alohali, Genetic bee colony (gbc) algorithm: A new gene selection method for microarray cancer classification, Computational biology and chemistry 56 (2015) 49–60.

[8] A. M. Taha, A. Mustapha, S.-D. Chen, Naive bayes-guided bat algorithm for feature selection, The Scientific World Journal 2013 (2013).

[9] I. Jain, V. K. Jain, R. Jain, Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification, Applied Soft Computing 62 (2018) 203–215.

[10] H. Sanz, C. Valim, E. Vegas, J. M. Oller, F. Reverter, Svm-rfe: selection and visualization of the most relevant features through non-linear kernels, BMC bioinformatics 19 (1) (2018) 1–18.

[11] H. Chai, C. Domeniconi, An evaluation of gene selection methods for multi-class microarray data classification, in: Proceedings of the Second European Workshop on Data Mining and Text Mining in Bioinformatics, 2004, pp. 3–10.

[12] Y. Liang, F. Zhang, J. Wang, T. Joshi, Y. Wang, D. Xu, Prediction of drought-resistant genes in arabidopsis thaliana using svm-rfe, PLoS one 6 (7) (2011) e21750.

[13] Q. Shen, W.-M. Shi, W. Kong, Hybrid particle swarm optimization and tabu search approach for selecting genes for tumor classification using gene expression data, Computational Biology and Chemistry 32 (1) (2008) 53–60.

[14] L. Li, L. G. Pedersen, T. A. Darden, C. R. Weinberg, Computational analysis of leukemia microarray expression data using the ga/knn method, in: Methods of Microarray Data Analysis, Springer, 2002, pp. 81–95.

[15] R. Xu, G. C. Anagnostopoulos, D. C. Wunsch, Multiclass cancer classification using semisupervised ellipsoid artmap and particle swarm optimization with gene expression data, IEEE/ACM Transactions on Computational Biology and Bioinformatics 4 (1) (2007) 65–77.

[16] J. Kennedy, R. C. Eberhart, A discrete binary version of the particle swarm algorithm, in: 1997 IEEE International conference on systems, man, and cybernetics. Computational cybernetics and simulation, Vol. 5, IEEE, 1997, pp. 4104–4108.

[17] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE transactions on pattern analysis and machine intelligence 35 (8) (2013) 1798–1828.

[18] S. Min, B. Lee, S. Yoon, Deep learning in bioinformatics, Briefings in bioinformatics 18 (5) (2017) 851–869.

[19] C. M. Perou, S. S. Jeffrey, M. Van De Rijn, C. A. Rees, M. B. Eisen, D. T. Ross, A. Pergamenschikov, C. F. Williams, S. X. Zhu, J. C. Lee, et al., Distinctive gene expression patterns in human mammary epithelial cells and breast cancers, Proceedings of the National Academy of Sciences 96 (16) (1999) 9212–9217.

[20] G. Getz, E. Levine, E. Domany, Coupled two-way clustering analysis of gene microarray data, Proceedings of the National Academy of Sciences 97 (22) (2000) 12079–12084.

[21] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard,

Average number of genes selected for 5 microarray datasets using BPSO						
Datasets	**SVM**	**Random Forest**	**FCFS**	**PSO-DT**	**MBEGA**	**RFE-BPSO-NN**
SRBCT	2308	2308	82	874	60.7	26
Leukemia	7129	7129	51	1468	15.8	15
CNS	7129	7129	28	1486	20.5	13
Lymphoma	4026	4026	105	1346	34.3	23
Lung	12,533	12,533	119	1657	14.1	27

Table 3

Classification Accuracy using RFE-INFO-BPSO				
Datasets	**SVM**	**DT**	**NB**	**NN**
SRBCT	88.23	94.11	**100**	94.11
Leukemia-3c	80	66.66	93.33	93.33
CNS	50	66.66	75	**75**
Lymphoma	**100**	71	78.57	**100**
Lung	95.12	70.17	90.24	92.68

Table 4

M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, science 286 (5439) (1999) 531–537.

[22] Y.-J. Li, L. Zhang, M. C. Speer, E. R. Martin, Evaluation of current methods of testing differential gene expression and beyond, in: Methods of Microarray Data Analysis II, Springer, 2002, pp. 185–194.

[23] E. P. Xing, M. I. Jordan, R. M. Karp, et al., Feature selection for high-dimensional genomic microarray data, in: Icml, Vol. 1, Citeseer, 2001, pp. 601–608.

[24] M. Xiong, X. Fang, J. Zhao, Biomarker identification by feature wrappers, Genome Research 11 (11) (2001) 1878–1887.

[25] Y. Tang, Y.-Q. Zhang, Z. Huang, Development of two-stage svm-rfe gene selection strategy for microarray expression data analysis, IEEE/ACM Transactions on Computational Biology and Bioinformatics 4 (3) (2007) 365–381.

[26] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, D. Haussler, Support vector machine classification and validation of cancer tissue samples using microarray expression data, Bioinformatics 16 (10) (2000) 906–914.

[27] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, et al., Multiclass cancer diagnosis using tumor gene expression signatures, Proceedings of the National Academy of Sciences 98 (26) (2001) 15149–15154.

[28] F. Fernández-Navarro, C. Hervás-Martínez, R. Ruiz, J. C. Riquelme, Evolutionary generalized radial basis function neural networks for improving prediction accuracy in gene classification using feature selection, Applied Soft Computing 12 (6) (2012) 1787–1800.

[29] K.-B. Hwang, D.-Y. Cho, S.-W. Park, S.-D. Kim, B.-T. Zhang, Applying machine learning techniques to analysis of gene expression data: cancer diagnosis, in: Methods of microarray data analysis, Springer, 2002, pp. 167–182.

[30] W. Li, Y. Yang, How many genes are needed for a discriminant microarray data analysis, in: Methods of microarray data analysis, Springer, 2002, pp. 137–149.

[31] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, Journal of bioinformatics and computational biology 3 (02) (2005) 185–205.

[32] R. Tibshirani, T. Hastie, B. Narasimhan, G. Chu, Diagnosis of multiple cancer types by shrunken centroids of gene expression, Proceedings of the National Academy of Sciences 99 (10) (2002) 6567–6572.

[33] Y. Wang, F. Makedon, Application of relief-f feature filtering algorithm to selecting informative genes for cancer classification using microarray data, in: Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004., IEEE, 2004, pp. 497–498.

[34] G. Zhao, Y. Wu, Feature subset selection for cancer classification using weight local modularity, Scientific reports 6 (2016) 34759.

[35] N. Ibrahim, H. Hamid, S. Rahman, S. Fong, Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy, Pertanika Journal of Science and Technology 26 (2018) 329–340.

[36] S. Li, X. Wu, M. Tan, Gene selection using hybrid particle swarm optimization and genetic algorithm, Soft Computing 12 (11) (2008) 1039–1048.

[37] K.-H. Chen, K.-J. Wang, K.-M. Wang, M.-A. Angelia, Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data, Applied Soft Computing 24 (2014) 773–780.

[38] E. B. Huerta, B. Duval, J.-K. Hao, A hybrid ga/svm approach for gene selection and classification of microarray data, in: Workshops on Applications of

Evolutionary Computation, Springer, 2006, pp. 34–44.

[39] C.-P. Lee, Y. Leu, A novel hybrid feature selection method for microarray data analysis, Applied Soft Computing 11 (1) (2011) 208–213.

[40] M. Xi, J. Sun, L. Liu, F. Fan, X. Wu, Cancer feature selection and classification using a binary quantum-behaved particle swarm optimization and support vector machine, Computational and mathematical Methods in Medicine 2016 (2016).

[41] H. Huang, S. Zhou, J. Jiang, H. Chen, Y. Li, C. Li, et al., A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features, BMC bioinformatics 20 (8) (2019) 290.

[42] R. Ruiz, J. C. Riquelme, J. S. Aguilar-Ruiz, Incremental wrapper-based gene selection from microarray data for cancer classification, Pattern Recognition 39 (12) (2006) 2383–2392.

[43] Z. Zhu, Y.-S. Ong, M. Dash, Markov blanket-embedded genetic algorithm for gene selection, Pattern Recognition 40 (11) (2007) 3236–3248.

[44] L.-Y. Chuang, C.-S. Yang, K.-C. Wu, C.-H. Yang, Gene selection and classification using taguchi chaotic binary particle swarm optimization, Expert Systems with Applications 38 (10) (2011) 13367–13377.

[45] M.-L. Huang, Y.-H. Hung, W. Lee, R.-K. Li, B.-R. Jiang, Svm-rfe based feature selection and taguchi parameters optimization for multiclass svm classifier, The Scientific World Journal 2014 (2014).