I J C T A, 9(25), 2016, pp. 249-267 ISSN: 0974-5572

Multi-objective Biogeography-based Krill Herd (BBKH) Based Optimization for Feature Selection Using Gene Expression Data

S. Venkata Krishna Kumar* and R. Nedunchezhian**

ABSTRACT

Microarray data have a significant role to play in developing effective cancer diagnoses and classification System using Gene Expression Data. Nonetheless, micro array expression data generally have huge redundancy and are noisy, and it is only a subset of theirs which presents unique profiles for various classes of samples. Therefore, selection of genes that are hugely discriminative from the gene expression data has gone to provoke remarkable interest in the bioinformatics field. In this paper, a Krill herd (KH) is basically a new search heuristic technique. For the purpose of its performance improvement, a biogeography-based krill herd (BBKH) algorithm has been proposed for resolving complicated optimization jobs. The improvement is involved with the introduction of a novel krill migration (KM) operator having the krill updation to manage the optimization issues with more efficiency. The KM operator focuses on the usage and allows the krill to cluster around the solutions that are the best at the final run stage of the search on gene relevant to the classification. In the algorithm proposed, initially, the Infomax ICA algorithm is introduced for selecting 60 of the best gene expression data. Next, in order to make BBKH based optimization to suit the discrete issue is proposed on the basis of a binary migration model in addition to a binary mutation model using Krill herd (KH) algorithm. Then, multi-objective biogeography-based krill herd (BBKH) based optimization is called MOBBKH. Finally, the MOBBKH technique is employed for the selection of gene, and then the support vector machine is employed to be the classifier along with the leave-one-out cross-validation technique (LOOCV).To exhibit the effectiveness and resourcefulness obtained from the algorithm, this algorithm proposed is tested over ten of the gene expression dataset standards. Results obtained from experiments illustrate that the new technique has nearly good or at the least comparable performance with the earlier existing particle swarm optimization (PSO) algorithm along with the support vector machine (SVM) while taking the quality observed from the obtained solutions into consideration.

Index Terms: Gene expression data, gene selection, hybrid approach, multi-objective binary biogeography based optimization.

1. INTRODUCTION

The advancements seen recently in the technology of microarrays lets the scientists to be able to do the measurement of the expression levels of genes in the number of thousands at the same time in biological organisms and hence have rendered it feasible to have databases containing cancerous tissues to be created [1]. At last, it generates the gene expression data containing helpful information with regard to genomic, diagnostic, and prognostic to choose the informative genes which act as contributors to a cancerous state. But, the process of gene selection is a huge challenge due to the following features of the gene expression data: the great numbers of genes in comparison with the minute count of samples (high-dimensional data), genes that

* Research Scholar and Associate Professor, Department of Computer Science, PSG College of Arts and Science, Coimbatore, Tamil Nadu, India, *Email: svkkumars@gmail.com*

** Director-Research and Professor Department of Computer Science and Engineering, Kalaingar Karunanidhi Institute of Technology, Kannampalayam (post), Coimbatore.

are irrelevant, and data that are noisy [2]. In order to get over this problem, a gene selection technique is employed for selecting a subset containing the informative genes which again helps in the maximization of the classifier's capability of classifying samples with more accuracy. In the domain of computational intelligence, gene selection is referred to as feature selection. The gene selection contains multiple benefits.

1) It can be useful in maintaining or improving the classification accuracy. 2) It can be helpful in the dimensionality reduction of the data. 3) It can limit the time involved in computation. 4) It can eliminate the irrelevant as well as noisy genes.

In the gene classification context, feature selection is chiefly divided into: wrappers and filters [3]. Wrappers exploit the learning algorithm for estimating the quality or the genes suitability for the modeling issue. Few researchers have stated the conclusion that if the aim of the model is the minimization of the classifier error rate, and the cost of measurement for all the features stands equal, then the predictive accuracy of the classifier becomes the most significant fact. The second category, filters model, conducts the selection of the feature subset in addition to the classification in two different phases, and makes use of a measure which is easier and can be rapid for computing.

As mentioned in [4]–[5], the vital goal of the procedure of feature selection is the reduction of the gene numbers that are utilized in classification while maintaining acceptable classification accuracy. It means, the objectives of the feature selection are concerned with both the maximizing the classification performance and minimizing the number of features selected. These two goals are most of the times contradictory and hence the decision that is optimal requires to be done with a compromise made between them. In this regard, it would be very much better to consider the problem of feature selection to be a multi-objective problem instead of a single-objective problem. Nonetheless, many of the feature selections based on evolutionary algorithms that are available in the literature are in turn based on single objective that are often minimally general and are computation wise more expensive when compared to the approaches that are multi-objective. Hence, the evolutionary algorithm is desired to be used for developing a multi objective approach for the simultaneous maximization of the classification rate and minimization of the number of features that are selected.

In the recent times, several multi-objective based techniques, for instance NSGA-II [6] and SPEA [7], have been introduced for feature selections. In [8], a hybrid multi-objective optimization technique has been introduced to find a small set having the non-redundant disease related genes by making use of the multi-objective particle swarm optimization's algorithm. In [9], Spolaôr et al. uses a multi-objective genetic algorithm through the combination of various filter approaches criteria for the feature selection. This method depends on the general characteristics pertaining to the data for feature correlation, and carries out the search for the features subset. In [10], a multi-objective technique in hybrid GA and support vector machine classifier (GASVM) is proposed for the selection of gene along with the gene expression data classification. In [11], a technique for feature selection depending on the archived multi-objective simulated annealing (AMOSA) is designed for the prediction of the miRNA promoters making use of an SVM along with RBF kernel. In [12], the multi-objective genetic algorithm is deployed in the form of the feature selection strategy for the recognition of handwriting. A neural network is deployed for permitting the usage of a representational database for evaluating the fitness. In [13], a multi-objective evolutionary algorithm is proposed to solve the problem of gene selection in the sense of gene subset size minimization and performance maximization. In [14], the optimization of split modified radius-margin model selection criteria is then utilized over benchmark issues. In [15], the use of a multi-objective genetic algorithm is introduced for the gene selection of Microarray datasets. The classification task is accomplished by support vector machines. In [16], a technique for the features selection exploits a multi-objective genetic algorithm in which the number of features is then optimized to boost the classification accuracy. In [17], multi-objective genetic programming is applied to the well-known edge detection problem in image processing and detailed comparisons are

made with the canny edge detector. In [18], the application of multi-objective optimization is used in the areas of bioinformatics along with computational biology. This method is used to explain the reasons behind the use of multi-objective optimization in each application area and also to point the way to potential future uses of the technique. However, this field of study is still in its early days, a large number of future researches are needed for developing a multi-objective algorithm to be used for the selection of features.

2. RELATED WORKS

Wang, et al., [19] showed that a DNA micro array may be used to simultaneously identify the expressions belonging to several genes. Micro-array data habitually a surround a petite number of samples, it contains an innumerable number of gene expression levels in the form of a feature. It is a great challenge to select the relevant genes that are involved in several kinds of cancer .For getting the mining information regarding the genes from one of the micro-array data of cancer and for using in the reduction of dimensionality, the algorithm like feature selection algorithms has been assessed in a systematic manner. Selecting the suitable genes from the micro array data may be done making use of Wrappers, Filters and CFS (correlation-based feature selector) and then the machine learning algorithms like decision trees, naïve Bayes and support vector machines. The data set utilized in this work was from the microarray data of acute leukemia and lymphoma. The performance of classification obtained from this experiment indicates that rather a good accuracy can be got on micro-array data set of acute leukemia and diffuse large Bcell lymphoma rather than the results that are published. Also there is a huge probability of having the selection of suitable genes with more confidence by means of the usage of various combinations of classification and feature selection methodologies. The results from experiments got from this research work demonstrate that the gene selection carried out by filters, CFS, and wrappers, showed a similar kind of performance over the data set analysed. For the rapid analysis of the data, the filters and CFS are advised. However, for selecting a very less count of genes validation results, the wrapper approaches can be utilized.

Chu & Wang [20] expressed that the Micro array gene expression data usually has a great deal of dimensions. The classifier utilized here is a support vector machine (SVM for the cancer classification to be used in addition to the microarray gene expression data. The genes selection has been finished by using four efficient feature dimensionality reduction techniques, for example, principal components analysis (PCA), classseparability measure, Fisher ratio, and T-test. The data set that finds application here is the SRBCT, lymphoma data set and leukemia data set of the openly existing micro array gene expression data set. In order to perform the multi-group classification, a scheme of voting is applied by means of k (k “ 1) binary SVMs. The results indicated that the T-test’s genetic selection in comparison to the rest of the three approaches. In all of the three data sets, the SVMs came out with very good accuracies when used with very less number of genes when compared with the other earlier published techniques.

Huilin Xiong & Xue-Wen Chen [21] states that the new approach referred to as the kernel function, enhances the classifier performance in the case of genetic data. The utility of a kernel approach was investigated where it is dependent upon the optimization of a kernel model that is data-dependent. The K-nearest-neighbor (KNN) and support vector machine (SVM) can be employed in the role of a classifier for the purpose of performance assessment. The data set utilized here are namely, ALL-AML Leukemia Data, Breast-ER, BreastLN,, Colon Tumor Data, Lung Cancer Data and Prostate Cancer from micro array data. A kernel optimization technique was found out to be helpful in classifying the gene expression data. Then the performance is assessed during the application of the kernel that is optimized in the classification of the gene expression data. While being compared with KNN, SVM in the role as “oksvm”, along with optimized kernel yields better accuracy.

Shen & Tan [22] demonstrated the penalized logistic regression for assisting in the cancer classification. The penalized logistic regression in union with the two-dimension reduction techniques so that the accuracy

of classification and the speed of computation found improvement. Support vector machines and the least squares regression are selected for comparison. The technique known as the Recursive feature elimination (RFE) has been employed for the gene selection which is iteration-based that attempts to choose a subset of genes having the most relevancy with the cancer types.

In [23] is discussed about the Support vector machine (SVM) that has attained the popularity of becoming a tool for the tasks of machine learning that involves classification, regression or novelty detection. SVM is capable of calculating the maximum margin (hyper-plane that separates) between the data that has and that does not have the interested outcome in case they can be linearly separated. In order to enhance the performance with regard to the generalization of the SVM classifier, optimization strategy is brought into use. As per the authors, Optimization indicates selecting an element that is best from some kind of set containing alternatives that are available. Particle swarm optimization (PSO) is actually a population based random optimization method in which the promising solutions, known as particles, follow the current optimum particles and they fly through the problem space. Principal Component Analysis (PCA) is used for the reduction of the features seen of breast cancer, lung cancer and heart disease data sets and also an empirically made comparison done for the kernel selection employing PSO for SVM is helpful for achieving better performance. This paper is concentrated on SVM trained making use of linear, polynomial and radial basis function (RBF) kernels and then PSO is applied to every kernel for each of the data set in order to obtain a better accuracy.

3. PROPOSED METHODOLOGY

The overall proposed framework is shown in the Figure 1. In this work firstly, the Infomax ICA is employed for choosing the 60 of the top gene expression data. Next, in order to make the Krill Herd Algorithm With Migration Operator In Biogeography-Based Optimization (BBKH) to suit the discrete problem, as called BBKH, is introduced on the basis of a migration model. Finally, The MOBBKH technique is utilized for the selection of gene, and afterwards, support vector machine is utilized to act in the form of a classifier along with the leave-one-out cross-validation technique (LOOCV).

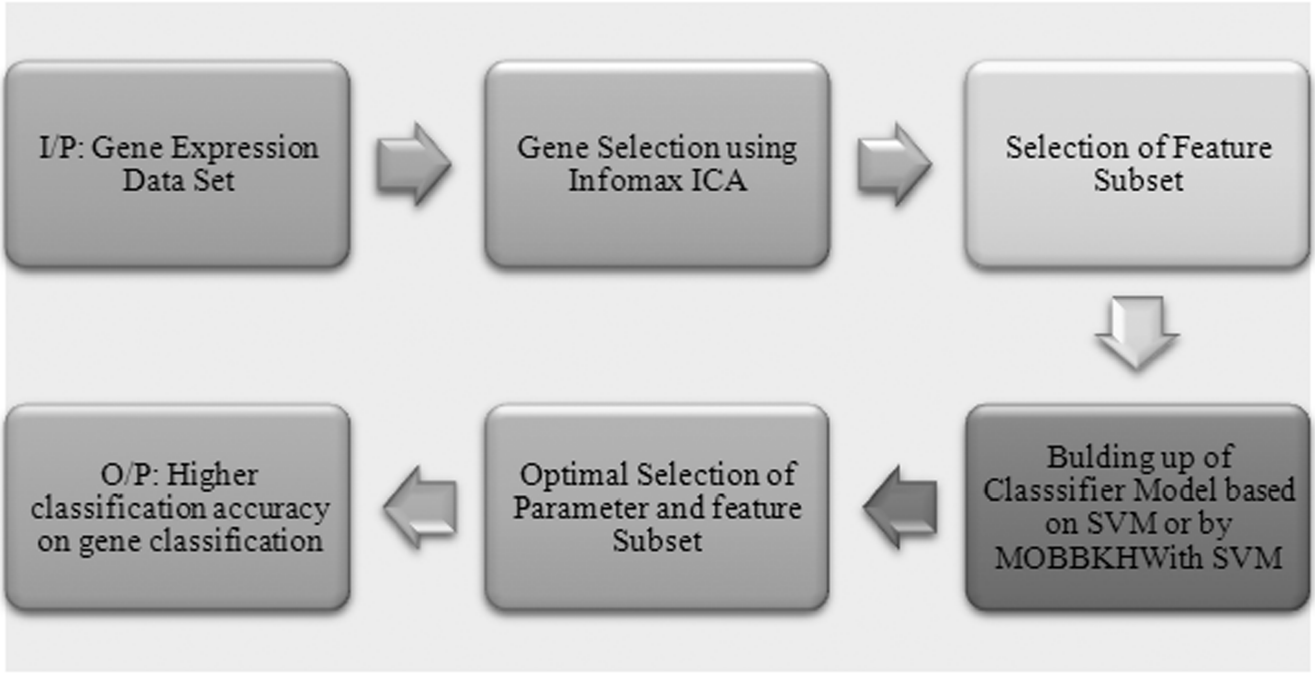

Figure 1: Overall Proposed Methodology

3.1. Infomax ICA for Preprocessing and Gene Selection

Recently, many successful ICA based applications of microarray data analysis methods are reported to extract gene expression modes [24]. In this section, the details regarding the ICA model for gene expression data are discussed. An improved version of ICA called Infomax ICA approach. In this work, Infomax ICA is used to select the genes and subset of tumor genes. An Infomax ICA method uses a mutual information gain to select the genes.

3.1.1. Infomax ICA

Mutual information is definite for a pair consisting of two random variables X and Y is defined as below:

$$I(X;\ Y) = H(X) - H(X|Y) \tag{1}$$

Where $H(X|Y)$ the conditional entropy of X taken Y on certain is values y and is the conditional entropy of X. Conditional entropy of x is defined as below:

$$H(X|Y) = H(X,\ Y) - H(Y) \tag{2}$$

where $H(X,\ Y)$ refers to the joint entropy of X and Y and $H(Y)$ indicates the entropy of Y. Mathematically, the entropy for a given random variable x and y is defined as:

$$H\left(x\right) = -\sum_x P\left(x\right)\log P\left(x\right) \tag{3}$$

$$H\left(y\right) = -\sum_y P\left(y\right)\log P\left(y\right) \tag{4}$$

$$H\left(X,Y\right) = -\sum_{x,y} P\left(x,y\right)\log P\left(x,y\right) \tag{5}$$

Where P(x) denotes the probability of X in the state x i,e gene. Entropy is used to measures the value of uncertainty. The lower value of information gives detail result of system. Consequently, going back, $I(X;\ Y)$ mutual information can be used to show that decrease of uncertainty about variable X following the Y's observation. By representation of infomax that tries to reduce mutual information and correspondingly searches for the components which are maximally independent. After the upadataion of $I(X;\ Y)$ result is proposed the following algorithm for computing the unmixing matrix W (referred to as InfoMax):

1. Initialize $W(0)$ (e.g. random)
2. $W\left(t+1\right) = W\left(t\right) + \eta\left(t\right) + \eta\left(t\right)\left(I - f\left(Y\right)Y^T\right)W\left(t\right)$
3. If not converged, return to step 2.

where t stands for an approximation step given, $\eta(t)$ is defined as the step size for the unmixing matrix updates, which is usually an exponential function or a constant value. $f(Y)$ a nonlinear function normally selection based on the type of gaussian distribution (either super or sub), I the identity matrix of dimensions $m \times m$ and T the transpose operator. In super-Gaussian distributions, $f(Y)$ is normally denoted as:

$$f(Y) = \text{Tanh}\ (Y) \tag{6}$$

and for the case of sub-Gaussian distributions, $f(Y)$ is denoted as:

$$f(Y) = Y - \text{Tanh}\ (Y) \tag{7}$$

The package InfoMax.nb is a realization of this algorithm.

3.1.2. Gene Selection using Infomax ICA

Gene expression dataset can be indicated by a $p \times n$ matrix x ($p >> n$), in which element x_{ij} refers to the expression level of the i^{th} gene in the j^{th} as say $\left(1 \le i \le p, 1 \le j \le n\right)$. Two dimensional vectors are used to

represent the gene samples in the dataset there are n dimensional vector and p dimensional vector. The *n* dimensional vector r_i, i.e., the i^{th} row of *X*, is the expression profile corresponding to the i^{th} gene. On the other side, the *p*-dimensional vector c_j, i.e., the j^{th} column of *X*, refers to the snapshot of the j^{th} as say (cell sample). Consider that the data samples have been preprocessed with p gene samples with n samples and normalized, i.e., each cell sample has a zero mean and unit standard deviation.

In the infomax ICA model for the gene expression data in ICA model is defined as Y = WX denotes the gene expression profiles by a new group of basis vectors. This result comes from following assumptions. Initially the gene expression profiles are evaluated by a combination containing hidden variables that are named as "expression gene of modes "(eigengenes). Following, the genes' responses for those variables can be estimated by linear functions. Gene expression profile q of Y matrix is defined by all of the eignegens in the rows of X and also by its linearly independent impacts on the expression profile q^{th} row of W eigengenes.

Finding the set of good basis profiles to characterize gene expression data thus the subsets of genes having relevance to cell samples classification can be chosen. Also the chief objective of this research work is to choose the genes subset which could have relevance for cell clustering. The selection of gene is obtained by Infomax ICA. The new gene selection process is in accordance with a ranking of the p genes. This ranking procedure is explained as below.

1. Initialize random variables of genes which is denoted as W
2. $W(t+1) = W(t) + \eta(t) + \eta(t)\left(I - f(Y)Y^T\right)W(t)$
3. If not converged, return to step 2 and choose another genes for the further procedure

By this way, the genes and the subset of the gene are selected for further processing.

3.2. Krill Herd Algorithm With Migration Operator In Biogeography-Based Optimization (BBKH) for optimizing the objective function

Krill herd (KH) is a new search heuristic technique. In order to have its performance improved, a biogeography- based krill herd (BBKH) algorithm is proposed for resolving sophisticated optimization tasks. The improvement deals with the introduction of a novel krill migration (KM) operator with the krill updating for managing the problems of optimization with more efficiency. The KM operator enforces the usage and allows the krill to be clustered all around the top solutions during the search's future run phase. The impacts of these improvements are tested by different well-defined standard functions. On the basis of the experimental results, this new BBKH approach outperforms the fundamental KH and the other optimization algorithms.

3.2.1. KH algorithm

By idealizing by the swarm behaviour of krill, KH [25] is a meta-heuristic optimization approach for solving optimization problems. In KH, the position is mainly affected by three actions [25]:

Movement affected by other krill; ii. Foraging action; iii. Physical diffusion

The Lagrangian model [25] is used in KH within predefined search space as Eq. (8).

$$\frac{dX_i}{dt} = N_i + F_i + D_i \tag{8}$$

where N_i refers to the motion generated by other krill individuals; F_i refers to the foraging motion, and D_i refers to the random diffusion corresponding to the ith krill individual.

In the first one, its direction, , is determined by the following divisions: target effect, local effect, along with an effect that is repulsive. In short, its definition is as below:

$$N_i^{new} = N^{\max}\alpha_i + \omega_n N_i^{old} \tag{9}$$

and N^{max}, ω_n and N_i^{old} refer to the maximum speed, the inertia weight, the last motion, correspondingly.

The second one can then be computed approximately by the two components: the location of food and its earlier experience. For the ith krill, it can be idealized below:

$$F_i = V_f \beta_i + \omega_f F_i^{old} \tag{10}$$

Where $\beta_i = \beta_i^{food} + \beta$ and $\mathbf{V}_f$ refers to the foraging speed, ω_f stands for the inertia weight, F_i^{old} indicates the last one.

The third part is essentially a random process. It is computed based on the maximum diffusion speed along with a random directional vector. Its expression is below:

$$D_i = D^{\max}\delta \tag{11}$$

Where D^{max} refers to the maximum diffusion speed, and d refers to the random directional vector and in addition, its arrays are actually random numbers. Herein, the position in KH from t to $t + \Delta t$ is expressed as follows:

$$X_i\left(t+\Delta t\right) = X_i\left(t\right) + \Delta t \frac{dX_i}{dt} \tag{12}$$

More detailed information about the KH method can be referred as in [25].

3.2.2. BBKH

As the search used in KH is mainly based on random walks, it cannot converge to the satisfactory function value all the time. To tackle with this drawback, Gandomi and Alavi [26] have done the addition of genetic reproduction strategies to the algorithm. It was shown that these mechanisms notably improve the performance of basic KH [26]. In addition, KH performs well on unimodal functions and several multimodal functions. But, at times KH's performance over complicated multimodal functions is quite a disappointment. In general, KH can implement exploration quickly and locate the promising region. While, KH has a poor exploitation ability. Therefore, another optimization technique which can carry on exploitation well is required for searching locally to force the method converge the best solution. Herein, in order to enhance the exploitation ability of KH, an improved habitat migration operator originally used in BBO performing local search, called krill migration (KM) operator, is then combined with KH in order to create an efficient biogeography-based krill herd (BBKH) approach. In BBKH, the KM operator is utilized for regulating the new solution that is produced by KH for each krill; while randomness is used in KH. This improved local search technique can increase the population diversity for preventing the early convergence and then make the krill search to be a small promising region with caution during the future stage of the method. The step of the KM operator used in BBKH is described in Algorithm 1. In BBKH, to begin with, due to its fast convergence, KH is applied to make the krill cluster to a limited area. After that, the KM operator with good exploitation is used to search locally to choose better krill. In essence, the KH in BBKH does the centralization on the exploration at an early stage of the optimization; while later, the KM operator enforces the usage and renders most krill to be clustered around the top solution at the later stage of the optimization. In this way, BBKH can fully explore the space with KH and exploit the useful information by KM operator. Hence, this technique can combat with the poor exploitation of the KH.

Algorithm 1 Krill migration (KM) operator

Begin
Select krill i (its position Xi) with probability based on ki
if rand (0, 1)< ki then
for j = 1 to d (all elements) do
Select Xj with probability based on lj
if rand (0, 1) < lj then
Randomly select an element r from Xj
Replace a random element in Xi with r
end if
end for j
end if
End.

Additionally, as in other algorithms, some type of elitism is added with the aim of keeping the optimal krill in the population at all times. It can forbid the optimal krill from being ruined by three motions and the KM operator. We must point out that, in BBKH, an elitism mechanism is employed for saving the characteristic of the krill to be with optimal fitness, and this way even in case KH and/or the KM operator corrupts its earlier good krill, we have saved it and can easily return it to its former condition.

Through the combination of KM operator and the focused elitism into the KH algorithm, the BBKH method has been developed as Algorithm 2. ki and li are functions of NP.

Algorithm 2 Biogeography-based KH method

Step 1: Initialization. Define the generation counter t = 1; initialize the population P of NP krill randomly; set Vf, Dmax, Nmax, Smax and pmod.
Step 2: Fitness evaluation. Evaluate each krill.
Step 3: While t < MaxGeneration do
Sort the krill from best to worst.
Store the best krill.
for i = 1 : NP (all krill) do
Conduct the three motion computation.
Update the krill position by Eq. (6).
Fine-tune Xi + 1 by using KM operator in Algorithm 1.
Evaluate each krill by Xi + 1.
end for i
Replace the worst krill with the best krill.
Sort the krill and find the current best.
t = t + 1;
Step 4: end while
Step 5: Output the best solutions.
End.

The final result of the KH optimization is decided as the best individual of the final iteration.

3.3. Krill Herd Algorithm with Migration Operator in Biogeography-Based Optimization (BBKH) and Support Vector Machines

While the SVM is brought into use, two issues require solution: the means to select the SVM's optimal input feature subset, and the desirable kernel parameters. These two issues are significant since the subset of features impacts the parameters C and r. On the basis of the assessment of the multi-objective Krill Herd Algorithm with Migration Operator in biogeography based optimization, BBKH algorithm contains the power to have both the optimal feature subset and the SVM parameters generated simultaneously. The research goal is the simultaneous optimization of the parameter and the feature subset, with no reduction in the rate of testing accuracy of the SVM.

In this work, develop a BBKH optimization along with SVM, referred to as BBKH SVM, for the determination of parameter and feature selection employing gene expression data. In spite of the feature selection parameters, two kernel function parameters, which are designated C and r, are necessary. The position of every individual is denoted by a binary (0/1) string excepting for the two dimensions of C and r. In the initial first phase, the BBKH algorithm generates a binary encoded individual in which each bit denotes a gene. In case a bit is 1, then it represents that this gene is maintained in the subset; otherwise if a bit is 0, it indicates an unselected feature. Hence, the length of the individual is equal to 2+D in the first microarray dataset. Afterwards, the fitness of every individual is evaluated through the accuracy of leave-one-out cross-validation technique (LOOCV). The leave-one-out cross-validation technique can be explained as below: if there are n data that has to be grouped, the data are then separated into one testing sample in addition to n-1 training samples. Every individual will then be chosen in the form of a testing sample in turns. The rest of the other n-1 individuals acts to be the training data set for the purpose of determining the model's prediction parameter. The solution representation can be described in Table 1.

Table 1
The solution representation

1	2	3	4	…….	D+2
P_c	P_r	F_1	F_2	……	F_D

As it can be observed in Table 1, P_c denotes the value of the C and P_r denotes the value of r. Both these two dimensions are integer coding schemes. With the aim of updating them in an effective manner, a new mutation method is proposed. Mutation vectors are then generated based on KH algorithm based on each of the population member or target vector in the current population. The mutation strategy can be explained as below:

$$H_i(SIV) = H_j(SIV) + rand * H_j(SIV) - H_i(SIV) \tag{13}$$

The range of searching of the parameter C of SVM is $[2^{-5}, 2^{15}]$, and the search ranging of parameter r of SVM is $[2^{-15}, 2^5]$. In the next part, we will use an example to demonstrate the process of BBKH SVM.

Given a data set contains seven records A1, A2, A3, A4, A5, A6, and A7, each of them has four features. Each record will be used as testing data in turn, and the other six records will be used as the training data. To make things clear, let A7 be the testing data, and A1–A6 data sets refer to the training data, as it can be observed in Fig. 1. Let the ith individual be randomly generated as 0,1,0,1,5,2 as can be seen in Figure. 2.As the value of the four features is 0, 1, 0, 1, the second and fourth rows are selected. Then the SVM uses the parameter $C = 2^5$ and $r = 2^{-2}$, yielding the higher classification accuracy rate across different datasets. The fitness value of the ith individual can be calculated as follows:

$$f_1 = SVM\_accuracy \tag{14}$$

$$f_2 = \left(\frac{D - R}{D} \right) \tag{15}$$

$$f = [f_1, f_2] \tag{16}$$

where *SVM_accuracy* is the SVM classification accuracy, D is the total number of the genes, and R is the number of selected genes.

In the next part, hybrid multi-objective binary biogeography based optimization and the system of feature selection and parameter optimization based on support vector machine will be introduced.

The process of the algorithm is shown in Figure. 3.

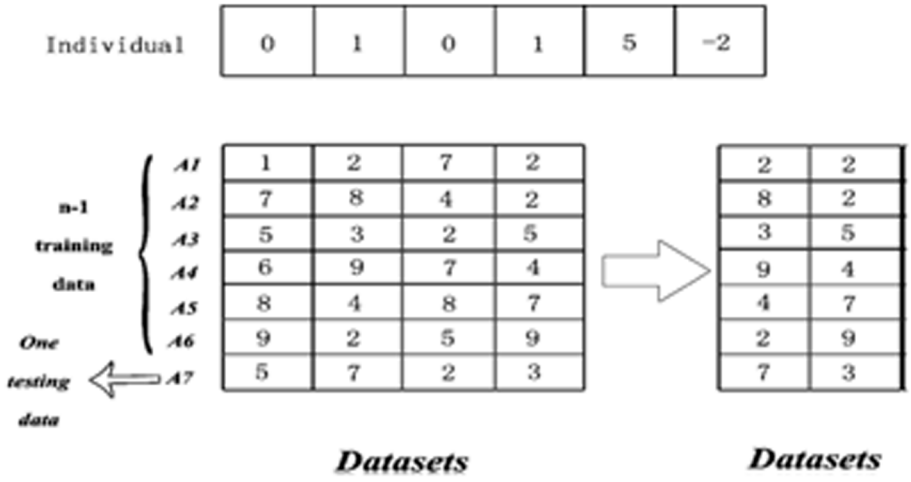

Figure 2: BBKH for the feature selection and parameter estimation of SVM

Step 1: The gene expression data are preprocessed by the Infomax ICA method. The 60 top genes with the highest scores are selected as the crude gene subset. The respective subsets in the testing portions are also chosen simultaneously.

Step 2: Conversion of genotype to phenotype. This step will do the conversion of the parameter C, r and feature subset from its genotype to a phenotype.

Step 3: the two objectives functions, and in (9) are calculated using Multi-Objective BBKH optimization.

Step 4: Optimization based on Multi-objective binary biogeography: In this step here, the algorithm seeks for solutions that are better with the aid of binary migration model and binary mutation model.

Step 5: Checking the termination criterion. The final feature subsets are selected, and then output the feature subset and the parameters C and r.

3.4. Computational Complexity of the MOBBKH SVM

Here we analyze the time complexity of the MOBBKH SVM model. Considering an iteration of the BBKH, a call will be invoked to the SVM. Tsand et al (2005) [27] shows that the data subroutines of standard SVM are $O(T^3)$, where T is the size of the dataset. Moreover, for each iteration, the non-dominated sorting is $O(M(2N)^2)$ and the assignment of crowding-distance is $O(M(2N)\log(2N))$, where N refers to the population

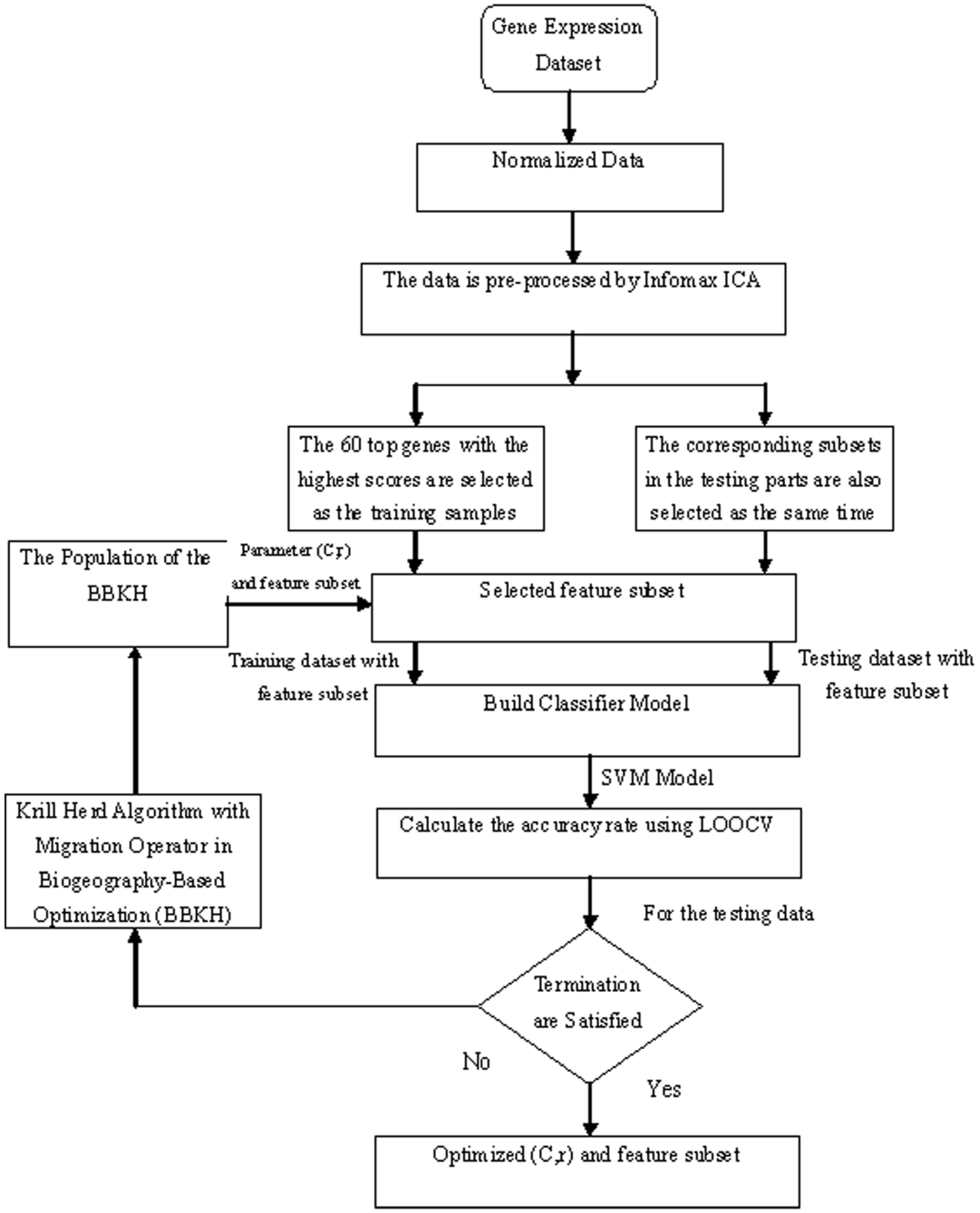

Figure 3: System architectures of the proposed BBKH-based feature selection and parameter optimization for the support vector machine

size and M indicates the number of objectives. Hence the complexity of the iteration, on an overall is $O(4\,MN^2\,T^3)$. Supposing that the total number of iterations is I, the time complexity of the algorithm is then $O(4\,IMN^2\,T^3)$. Specifically, there are two objectives in this problem, so the time complexity of the whole algorithm is $O(4 \times 2I\,N^2T^3)$, i.e., $O(IN^2T^3)$.

4. RESULT AND DISCUSSION

Selection of the relevant genes for fulfilling the classification of gene expression is a popular issue in bioinformatics. Here, in this work, ten of the gene expression data, inclusive of tumor samples, brain

Table 2
Format of Gene Expression Classification Data

Dataset Number	*Dataset Name*	*Number of Samples*	*Number of Genes*	*Number of Classes*
1	11_Tumors	60	9	5726
2	9_Tumors	174	11	12533
3	Brain_Tumors 1	90	5	5920
4	Brain_Tumors 2	50	4	10367
5	Leukemia 1	72	3	5327
6	Leukemia 2	72	3	11225
7	Lung_Cancer	203	5	12600
8	SRBCT	83	4	2308
9	Prostate_Tumor	102	2	10509
10	DLBCL	77	2	5469

tumor, leukemia, lung cancer, and prostate tumor samples, are employed in the tests and also their respective characteristics are provided in Table 2 that can then be downloaded from http://www.gems-sytem.org.The data sets are separated into one testing sample in addition to n-1 training samples. It is to be noted that every individual will then be chosen to be a testing sample in turns. The rest of the n-1 individuals will act as the data set for training.

For our algorithm, the MOBBKH SVM is written in Matlab-7.9 language development environment by means of the extension of the libsvm that is actually designed by Chang and Li [28]. All the experiments are conducted on a Pentium 3.0 GHz Processor along with memory of 1.0 GB. In this paper, we shall compare this algorithm with many benchmarked single objective algorithms: SVM Grid search, IBPSO, HPSOTS, PSO/GA[29], and a multi-objective algorithm NSGA-II [6].

The parameters of the relevant algorithms are set as follows: for the MOBBKH SVM, the population size is 50, the number of generations is 100, the habitat modification probability is 1, and mutation rate is 0.5. For the IBPSO [4], the number of particles comes to 100, the number of iterations is 100, and the values of and are both 2. For the IBPSO [28], the particle is 100, the number of iterations is 300, and the values of and are 2. For the hybrid PSO/GA [5], there involve two processes. For each process, the count of particles is 50. So the total count of particles comes to 100. The number of iterations is 10, the values of and is 2, the crossing rate is 0.985, and the mutation rate is 0.05. For the HPSOTS [30], the population size is 50, the size of tabu list is set as 5, and the number of iterations is 50. For the NSGA-II, the population size is 50, the number of generations is 100, the crossover rate is 0.7 and the mutation rate is 0.5. To make the experiments more accurate, each algorithm will be run ten times for each gene data. After that, an average result of the ten independent runs is received and they are compared.

As stated in [29], the objectives of feature selection should be both the maximization of the classification performance and the minimization of the number of features that are to be selected. That's why regard the problem of feature selection to be a multi-objective problem instead of a single-objective problem. Specifically, two objectives are considered, i.e., the accuracy of classification and the number of genes selected. Dissimilar to single objective algorithms that generate a single solution, our algorithm produces a set of solutions, referred to as a Pareto front, where no solution is dominated strictly by another solution for all of the objectives. Selecting one solution over others is quite a challenging task because the Pareto set can contain an unmanageable number of solutions. In this paper, a solution having the highest accuracy of classification is selected in a Pareto front. There are mainly two reasons for us to make such choice. At first, many of the single objective algorithms only take the classification accuracy into consideration. For the sake of simplicity with regard to the comparison done with these algorithms, priority has to be given to the

highest classification accuracy. Secondly, as it is mentioned in[8], classification accuracy is usually more important in many fields rather than the number of the genes selected, though the latter is also an important measure.

Tables 3 and IV list the experiment results of MOBBKH SVM on ten gene datasets. In these tables, #acc denotes the testing accuracy, and #selected gene denotes the number of genes selected for these gene expression data. From the tables, we can see that the results of the proposed algorithm are consistent on all datasets. For the Leukemia1, Leukemia2, SRBCT, and DLBCL datasets, MOBBBO, the proposed MOBBKH can attain 100% LOOCV accuracy having less than 10 selected genes. For the Brain_Tumor2 dataset, MOBBBO can obtain the 100% LOOCV accuracy for nine times. For the other datasets of Lung Cancer, Prostate_Tumor, and Brain_Tumor1, the BBKF algorithm can also provide more than 96% classification accuracies except for the 11_Tumors dataset (92.414%) and 9_Tumors dataset (80.5%). From the point of

Table 3
Experimental Results for each Run using MOBBKH based on LOOCV for 11_tumors, 9_tumors, brain_tumor1, Brain_tumor2, and Leukemia1

Run	*11_Tumors*		*9_Tumors*		*Brain_Tumors 1*		*Brain_Tumors 2*		*Leukemia 1*	
	Accuracy	*Selected Genes*	*Accuracy*	*Selected Genes*	*Accuracy*	*Selected Genes*	*Accuracy*	*Selected Genes*	*Accuracy*	*Selected Genes*
1	93.21	26	81.96	21	95.88	12	100	11	100	6
2	94.10	26	80.23	20	95.88	15	100	10	100	7
3	91.86	33	78.54	33	96.85	12	100	8	100	5
4	92.12	21	76.68	33	98.12	11	100	10	100	5
5	91.02	24	82.10	20	96.72	15	100	12	100	5
6	93.87	25	82	26	97.95	18	100	12	100	8
7	92.15	26	82	25	96.84	12	100	9	100	8
8	93.95	24	80.41	21	96.84	8	100	12	100	8
9	92.21	22	80	20	96.84	12	100	9	100	7
10	92.86	24	81.74	22	96.98	17	100	9	100	6
Average	92.414	25.1	80.5	24.1	96.667	13.2	99.8	10.2	100	6.5
S.D.	±1.00	±3.24	±1.76	±5.13	±1.00	±3.01	±0.63	±1.47	±0	±1.26

Table 4
Experimental Results for each Run using MOBBKH on Leukemia2, Lung_cancer, SRBCT, Prostate_tumor, and DLBCL

Run	*Leukemia 2*		*Lung_Caner*		*SRBCT*		*Prostate_Tumor*		*DLBCL*	
	Accuracy	*Selected Genes*	*Accuracy*	*Selected Genes*	*Accuracy*	*Selected Genes*	*Accuracy*	*Selected Genes*	*Accuracy*	*Selected Genes*
1	100	4	98.82	14	100	7	97.52	6	100	6
2	100	6	98.32	15	100	7	99.25	9	100	7
3	100	4	98.33	10	100	7	97.35	8	100	7
4	100	4	99.52	18	100	7	98.50	26	100	5
5	100	6	99.72	20	100	9	100	15	100	6
6	100	4	98.25	13	100	4	98.25	10	100	5
7	100	6	98.25	16	100	5	98.25	11	100	5
8	100	3	99.44	13	100	5	99.50	12	100	5
9	100	5	98.32	24	100	7	97.65	10	100	5
10	100	6	98.75	19	100	6	100	12	100	6
Average	100	4.8	98.47	16.2	100	6.4	98.3	11.9	100	5.7
S.D.	±0	±1.13	±0.54	±4.10	±0	±1.42	±1.13	±5.52	±0	±0.82

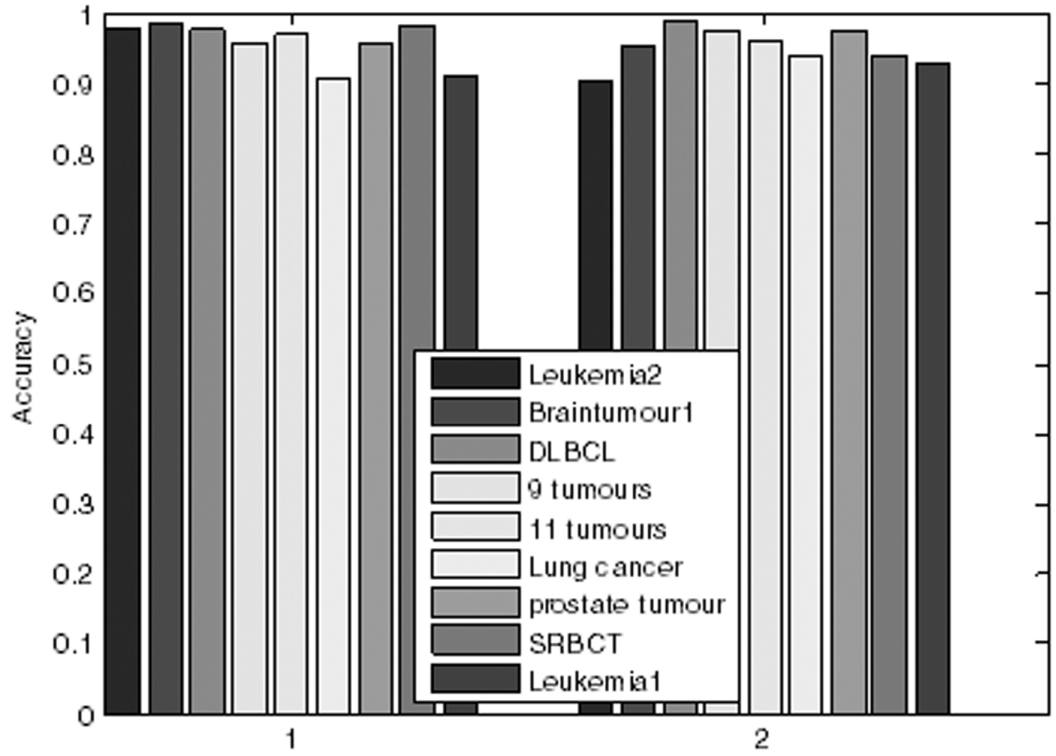

Figure 4: The accuracy obtained from MOBBBO in each of the independent runs

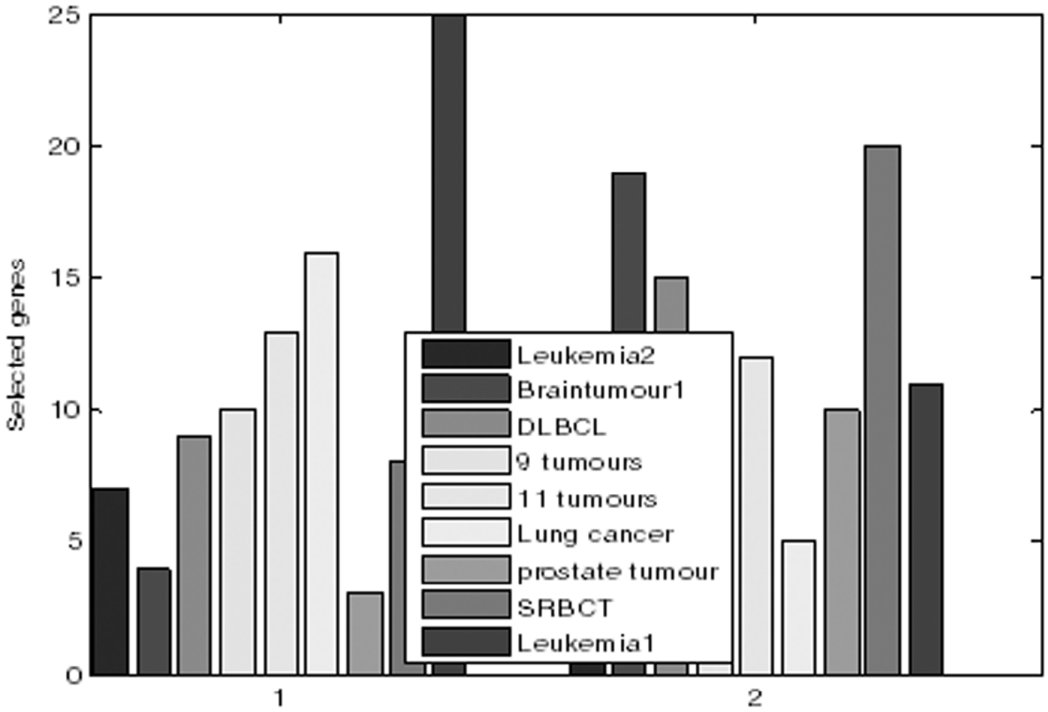

Figure 5: The number of selected genes obtained from MOBBKH in every independent run

perspective of the average accuracy in every independent run, the LOOCV accuracy and also the count of genes selected which are generated by proposed MOBBKH are illustrated in Figures 4 and 5.

The characteristics of the genes, selected genes, and percentage of selected genes are tabulated in Table 5. The average percentage of the genes selected is 0.0017. For the Leukemia1, Leukemia2, SRBCT, and DLBCL dataset, our algorithm can obtain 100% LOOCV accuracy though the percentage of the genes

selected is reduce to 0.0012, 0.0004, 0.0028, and 0.0010. So we can conclude that not all features are necessary for achieving the better classification accuracy. Figure. 5 shows the percentage of the selected genes. As it can be observed in this figure, we can find the Leukemia2 generates the smallest values, while the 11_Tumors generates the largest values.

With the purpose of showing the efficiency of every portion, two experiments are then conducted. The first experiment is with regard to showing the efficiency of the Fisher-Markov selector, and then the second experiment is aimed at demonstrating the effectiveness of MOBBKH. For the first experiment, we compare

Table 5
The genes, genes selected, and the percentage of gene selected

Dataset Name	*Total Number f Genes*	*Number of Selected Genes*	*Percentage of genes selected*
11_Tumors	5726	25.1	0.0044
9_Tumors	12533	24.1	0.0019
Brain_Tumors 1	5920	13.2	0.0022
Brain_Tumors 2	10367	10.2	0.0010
Leukemia 1	5327	6.5	0.0012
Leukemia 2	11225	4.8	0.0004
Lung_Cancer	12600	16.2	0.0013
SRBCT	2308	6.4	0.0028
Prostate_Tumor	10509	11.9	0.0011
DLBCL	5469	5.7	0.0010

Table 6
Comparison

Dataset Name	Evaluation	MOBBBO	Proposed MOBBKH with Infomax ICA
11_Tumors	Accuracy	92.414	95.1
	Genes	25	25
9_Tumors	Accuracy	80.5	81
	Genes	24.1	24.1
Brain_Tumors 1	Accuracy	96.667	97
	Genes	13.2	13.2
Brain_Tumors 2	Accuracy	99.8	99.9
	Genes	10.2	10.2
Leukemia 1	Accuracy	100	100
	Genes	6.5	6.5
Leukemia 2	Accuracy	100	100
	Genes	4.8	4.8
Lung_Cancer	Accuracy	98.473	98.5
	Genes	16.2	16.2
SRBCT	Accuracy	100	100
	Genes	6.4	6.4
Prostate_Tumor	Accuracy	98.33	98.5
	Genes	11.9	11.9
DLBCL	Accuracy	100	100
	Genes	5.7	5.7

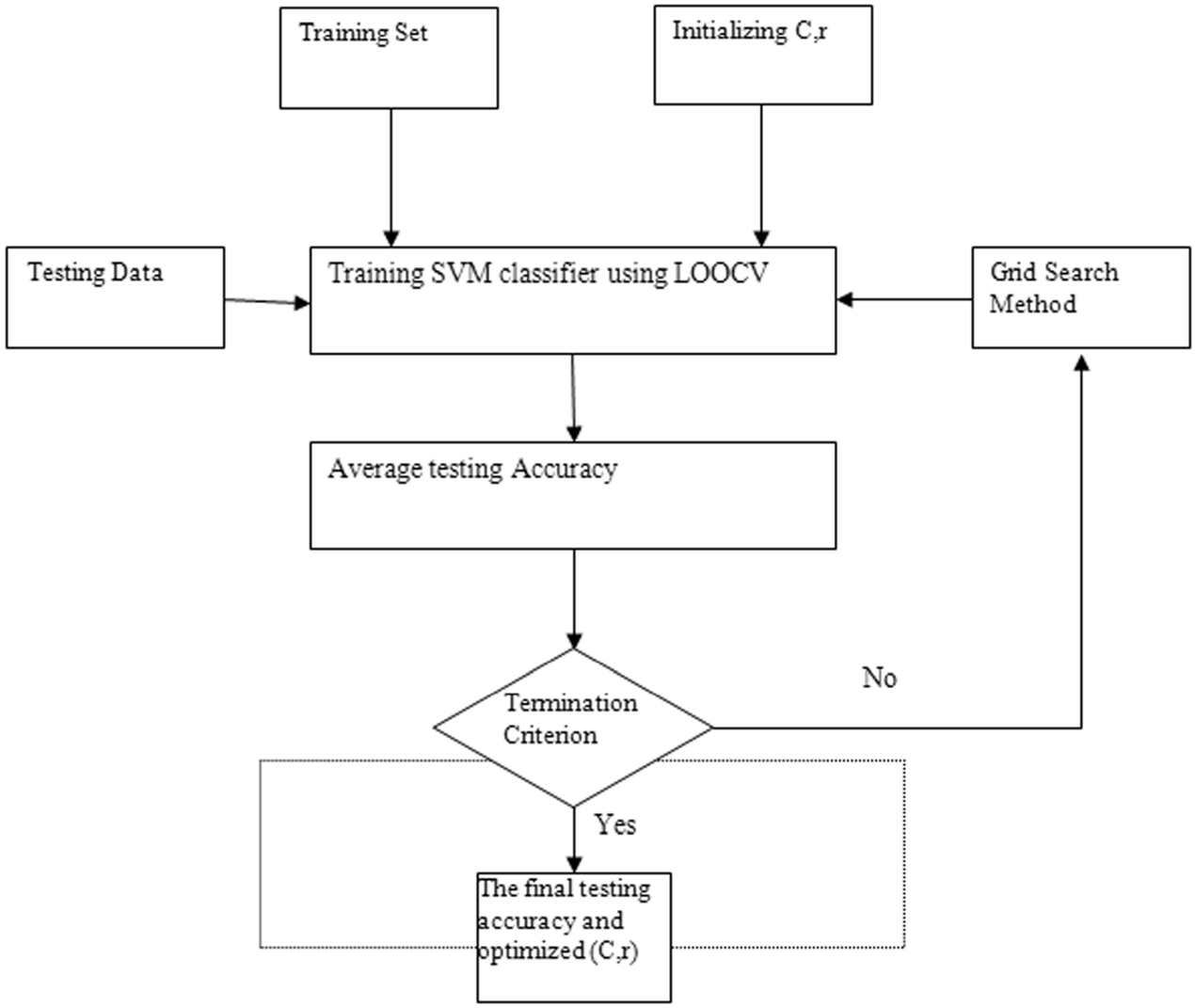

Figure 6: The process of grid algorithm combined SVM algorithm

the algorithm MOBBKH SVM with Fisher-Markov selector. The results of the accuracy and the number of genes selected are listed in Table 6.

From the table we can see for Leukemia1, Leukemia2, SRBCT, and DLBCL, both MOBBBO SVM and MOBBBO SVM without the Fisher-Markov selector can get 100% classification accuracy, while MOBBBO SVM can provide lower gene count. For 9_Tumors, MOBBBO SVM, and MOBBBO SVM without the Fisher-Markov selector can get 80.5% of classification accuracy, and MOBBBO SVM can be also used to yield a lower number of genes. For Brain_Tumors1, Brain_Tumors2, Lung_cancer, and Prostate_Tumor datasets, MOBBBO SVM can not only provide better classification accuracy, but also lower gene numbers. Only for the 11_Tumors, the MOBBBO without the Fisher-Markov selector can render the better classification accuracy that also shows that the Fisher-Markov selector is not desirable for all kinds of conditions. For the second experiment, we compare our approach with grid search SVM. Figure. 6 shows the process of grid search SVM.

In Table 7, the better results between two algorithms are highlighted using boldface. It is easy to see that both the accuracy of classification and the number of genes selected by MOBBBO SVM are superior to grid search SVM. This also proves the effectiveness of MOBBBO.

To further verify the efficiency of the proposed technique, we first compare our work with some single objective algorithms. Note that most of these single objective algorithms only care about the accuracy of

Table 7
Compared with MOBBKH SVM with grid search SVM

Dataset Name	*Evaluation*	*MOBBBO*	*Grid Search SVM*	*Proposed MOBBKH with Infomax ICA*
11_Tumors	Accuracy	92.414	89.08	95.1
	Genes	25	12533	25
9_Tumors	Accuracy	80.5	51.67	81
	Genes	24.1	5726	24.1
Brain_Tumors 1	Accuracy	96.667	90	97
	Genes	13.2	5920	13.2
Brain_Tumors 2	Accuracy	99.8	90	99.9
	Genes	10.2	10367	10.2
Leukemia 1	Accuracy	100	97.22	100
	Genes	6.5	5327	6.5
Leukemia 2	Accuracy	100	94.44	100
	Genes	4.8	11225	4.8
Lung_Cancer	Accuracy	98.473	95.07	98.5
	Genes	16.2	12600	16.2
SRBCT	Accuracy	100	98.80	100
	Genes	6.4	2308	6.4
Prostate_Tumor	Accuracy	98.33	93.14	98.5
	Genes	11.9	10509	11.9
DLBCL	Accuracy	100	96.10	100
	Genes	5.7	5469	5.7

classification. So we mainly compare the classification accuracy of these algorithms. But we also list the number of genes selected as a reference. As known generally, one among the essential criteria in the assessment of any new algorithm is the quality of the algorithm and its capability to generate similar outcomes when executed several time. Hence, the mean of the objective values and the count of genes of ten times are listed in Tables VIII and IX. As can be seen in Tables VIII and IX, this algorithm can get greater averages of the LOOCV classification accuracies on all of the datasets when compared with the other PSO algorithms [6]–[8] and the other SVM based algorithms [35], [36] except 11_Tumors. For the 11_Tumors dataset, IBPSO [8] can obtain better solutions ratherthan the other algorithms. For the Leukemia1 data, the algorithm Proposed MOBBKH, MOBBBO, IBPSO [8] and IBPSO [6] can all obtained 100% accuracy rate, while the IBPSO can obtain less number of genes compared with MOBBBO and Proposed MOBBKH. According to Tables VIII–IX, we can draw the result that even if only the classification accuracy is considered, MOBBBO algorithm again stands tall to be a high utility tool for the gene selection.

Based on the above analysis, the experimental results can be used to demonstrate the flexibility and reliability of the MOBBBO proposed for the purpose of feature selection. This algorithm can provide positive results when applied to the gene expression data with the limited number of features and samples. The reason may be that the multi-objective binary biogeography based optimization tends to be able to share their features with low *HSI* solutions, which can take up a lot of new features from high *HSI* solutions. In MOBBBO, a habitat is a vector which follows binary migration and binary mutation step to the optimal solution. The new candidate habitat is then generated from all of the solutions in the population by making use of the models of binary migration and binary mutation. Adhering to these rules, the MOBBBO algorithm at last generates a better subset for the purpose of gene classification.

Table 8
Compared with MOBBBO SVM with other earlier PSO

Dataset Name	*Evaluation*	*Proposed MOBBKH*	*MOBBBO*	*IBPSO [8]*	*IBPSO [6]*
11_Tumors	Accuracy	95	92.414	95.06	93.10
	Genes	26	25	240.9	2948
9_Tumors	Accuracy	81.23	80.5	75.50	78.33
	Genes	25	24.1	240.6	1280
Brain_Tumors 1	Accuracy	97.12	96.667	92.56	94.44
	Genes	14	13.2	11.2	754
Brain_Tumors 2	Accuracy	99.92	99.8	92.00	94.00
	Genes	11	10.2	9.10	1197
Leukemia 1	Accuracy	100	100	100	100
	Genes	7	6.5	3.5	1034
Leukemia 2	Accuracy	100	100	100	100
	Genes	5.2	4.8	6.7	1292
Lung_Cancer	Accuracy	99.1	98.473	95.86	96.55
	Genes	17	16.2	14.90	1897
SRBCT	Accuracy	100	100	100	100
	Genes	7.5	6.4	17.50	431
Prostate_Tumor	Accuracy	99	98.33	97.94	92.16
	Genes	13	11.9	13.60	1294
DLBCL	Accuracy	100	100	100	100
	Genes	6	5.7	6	1042

5. CONCLUSION

In this research work, a hybrid multi-objective binary biogeography based optimization along with support vector machine is introduced for selection of genes on the ten of the gene expression datasets. The results received from the experiments exhibit that the algorithm is capable to make the feature selection simplified by getting a less number of features which are needed in an effective manner and obtains a classification accuracy that is higher when compared with the other earlier available techniques. The algorithm proposed can get the accuracy to be highest in nine out of the ten problems of microarray dataset as the multi-objective approach in it can yield a unique solution in Pareto optimal set. In addition, the results also indicate that exists several irrelevant genes present in the gene expression data and few of them also do not have relevancy for a given type of cancer. For futuristic work, the algorithm proposed can find its application in some of the problems present in other fields.

REFERENCES

[1] S. Bandyopadhyay, R. Mitra, and U.Maulik, "Development of the human cancer microRNA network," BMC Silence. Vol. 1.No. 6, 2010.

[2] M. S.Mohamad, S.Omatu, M.Yoshioka, and S. Deris, "A three-stage method to select informative genes for cancer classification," Int. J. Innovative Comput. Inf. Control. Vol.6, No.1. pp. 117-126, 2010.

[3] M. S. Mohamad, S. Omatu, M. Yosioka, and S. Deris, "A cyclic hybrid method to select a smaller subset of informative genes for cancer classification," Int. J. Innov. Comput., Inf. Control. Vol. 5. No. 8. pp.2189–2202, 2009.

[4] L.Y. Chuang, H. W. Chang, C. J. Tu, and C. H.Yang, "Improved binary PSO for feature selection using gene expression data," Comput. Bio. Chem. Vol. 32. pp. 29–38, 2008.

[5] S. Li, X. Wu, and M. Tan, "Gene selection using hybrid particle swarm optimization and genetic algorithm," Soft Comput. Vol. 12. pp. 1039–1048, 2008.

[6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Trans. Evol. Comput. Vol. 6. No. 2. pp. 182–97, 2002.

[7] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms:A comparative case study and the strength Pareto approach," IEEE Trans. Evol. Comput. Vol. 3. No. 4. pp. 257–271, 1999.

[8] A. Mukhopadhyay and M. Mandal, "A hybrid multiobjective particle swarm optimization approach for non-redundant gene marker selection.' in Proc. 7th Int. Conf. Bio-Inspired Comput.: Theories Appl. 2013, Springer, India, pp. 205–216, 2013.

[9] N. Spolaôr, A. C. Lorena, and H.D. Lee, Multi-Objective Genetic Algorithm Evaluation in Feature Selection[C]// Evolutionary Multi-Criterion Optimization. Berlin/Heidelberg, Germany: Springer. pp.462–476, 2011.

[10] M. S. Mohamad et al., "A multi-objective strategy in genetic algorithms for gene selection of gene expression data," Artif. Life Robot. Vol. 13. No. 2. pp. 410–413, 2009.

[11] M. Bhattacharyya et al., "MicroRNA transcription start site prediction with multi-objective feature selection," Stat. Appl. Genet. Mol. Biol. Vol. 11. No. 1. pp. 1–25, 2012.

[12] L. S. Oliveira et al., "Feature selection using multi-objective genetic algorithms for handwritten digit recognition.' in IEEE Proc. 16th Int. Conf. Pattern Recog 2002, Vol. 1. pp. 568–571, 2002.

[13] J. Liu and H. Iba, "Selecting informative genes using a multiobjective evolutionary algorithm," in Proc. 2002 IEEE Congr. Evol. Comput. 2002. Vol. 1. pp. 297–302.

[14] T. Suttorp and C. Igel, Multi-objective optimization of support vector machines. inMulti-objectiveMachine Learning. Berlin/Heidelberg, Germany: Springerpp. pp.199–220, 2006.

[15] J. García-Nieto et al., "Sensitivity and specificity based multiobjective approach for feature selection:Application to cancer diagnosis," Inf. Process. Letter. Vol. 109. No. 16. pp. 887–896, 2009.

[16] U. Tekguc, H. Soyel, and H. Demirel, "Feature selection for person-independent 3D facial expression recognition using NSGA-II.' in Proc. 24th IEEE Int. Symp.Comput. Inf. Sci.2009 , pp. 35–38, 2009.

[17] Y. Zhang and P. I. Rockett, Feature extraction using multi-objective genetic programming. in Multi-Objective Machine Learning. Berlin/Heidelberg, Germany: Springer. pp. 75–99, 2006.

[18] J. Handl, D. B. Kell, and J. Knowles, "Multiobjective optimization in bioinformatics and computational biology," IEEE/ ACM Trans. Comput. Biol. Bioinformatics (TCBB), vol. 4, no. 2, pp. 279–292, 2007.

[19] Y.Wang, I. -V.Tetko, M. -A.Hall, E.Frank, A.Facius, K. -F.Mayer, and H. -W.Mewes, "Gene Selection From Microarray Data For Cancer Classification —A Machine Learning Approach," Comput Biol Chem. Vol.29 .No.1.pp.37-46, 2005.

[20] F. Chu And L. Wang, "Applications Of Support Vector Machines To Cancer Classification With Microarray Data," International Journal Of Neural Systems. Vol. 15.No. 6. pp.475–484, 2005.

[21] H.Xiong and X-W. Chen, "Optimized Kernel Machines For Cancer Classification Using Gene Expression Data," Proceedings Of The 2005 IEEE Symposium On Computational Intelligence In Bioinformatics And Computational Biology. pp.1-7, 2005.

[22] L. Shen and E.C. Tan, "Dimension ReductionBased Penalized Logistic Regression For Cancer Classification Using Microarray Data," IEEE/ACM Trans. Computational Biology And Bioinformatics. Vol. 2.No. 2. pp. 166-175, 2005.

[23] Das, S. Rekha, P.K. Panigrahi, K. Das, and D. Mishra. "Improving RBF kernel function of support vector machine using particle swarm optimization," International Journal.2012.

[24] G.Hori, M.Inoue, S.Nishimura, and H. Nakahara, "Blind gene classification based on ICA of microarray data.' International Conference on Independent Component Analysis and Signal Separation.ICA 2001, pp. 332–336, 2001.

[25] G.G.Wang, A.H.Gandomi, X.S.Yang, and A.H. Alavi, "A new hybrid method based on krill herd and cuckoo search for global optimization tasks," Int. J. Bio-Inspir. Comput. 2012.

[26] Available online: http://www.inderscience.com/info/ingeneral/forthcoming.php?jcode=ijbic (October 2015).

[27] A. H. Gandomi, and A. H. Alavi, "Krillherd: a new bio-inspired optimization algorithm," Communications in Nonlinear Science and Numerical Simulation. Vol.17. No.12.pp.4831–4845, 2012.

[28] C. C. Chang and C. J. Lin, "LIBSVM—Alibraryfor support vector machines" [Online]. Available: http:// www.csie.ntu.edu.tw/~cjlin/libsvm/

[29] I. W. Tsang, J. T. Kwok, and P. M. Cheung, "Core vector machines: Fast SVM training on very large data sets," J. Mach. Learn. Res. pp. 363–392, 2005.

[30] M. S. Mohamad, S. Omatu, S. Deris, and M. Yoshioka, "A modified binary particle swarm optimization for selecting the small subset of informative genes from gene expression data," IEEE Trans. Inf. Technol. Biomed. Vol. 15.No. 6. pp. 813–822, 2011.

[31] Q. Shen, W. M. Shi, and W. Kong, "Hybrid particle swarm optimization and Tabu search approach for selecting genes for tumor classification using gene expression data," Comput. Biol. Chem. Vol. 32. pp. 53–62, 2008.