

ABC-SVM: Artificial Bee Colony and SVM Method for Microarray Gene Selection and Multi Class Cancer Classification

Hala M. Alshamlan, Ghada H. Badr, and Yousef A. Alohal

Abstract—In this paper, we propose apply ABC algorithm in analyzing microarray dataset. In addition, we propose an innovative hybrid classification model, Support Vector Machine (SVM) with ABC algorithm, to measure the classification accuracy for selected genes. We evaluate the performance of the proposed ABC-SVM algorithm by conducting extensive experiments on six binary and multi-class microarrays dataset. Furthermore, we compare our proposed ABC-SVM algorithm with previously known techniques. The experimental results prove that ABC-SVM algorithm is promising approach for solving gene selection and cancer classification problems, and achieves the highest classification accuracy together with the lowest average of selected genes compared to previously suggested methods.

Index Terms—ABC, gene selection, microarray, and SVM.

I INTRODUCTION

Gene expression profiling or microarrays offer an efficient method of gathering data that can be used to determine the patterns of gene expression of many (if not all) genes in an organism in a single experiment. However, microarray dataset suffers from the curse of dimensionality, the small number of samples, and the level of irrelevant and noise genes, all of which makes the classification task for a given sample more challenging [1], [2]. Gene selection is the process of selecting the smallest subset of informative genes that are most predictive to its relative class using a classification model. This maximizes the classifier's ability to classify samples accurately. In this kind of study one aim is to identify the genes that contribute the most to cancer diagnosis, which would assist in drug discovery and early diagnosis. The optimal feature selection problem has been shown to be NP-hard [3]. Therefore, it is better to use heuristic approaches such as bio- inspired evolutionary algorithms in order to solve this problem.

Bio-inspired evolutionary algorithms such as GA, PSO, and ABC are more applicable and accurate than the wrapper gene selection method [4] because they are capable of searching for optimal or near-optimal solutions on complex and large spaces of possible solutions. Furthermore, they allow searching the solution space by considering multiple interacting attributes simultaneously, rather than by

considering one attribute at a time [4]. The artificial bee colony (ABC) algorithm that is introduced by Karaboga [5] is one bio-inspired evolutionary approach that has been used to find an optimal solution in numerical optimization problems. The algorithm is inspired by the behavior of honey bees when seeking a quality food source. The performance of ABC algorithms has been compared with other evolutionary methods such as genetic algorithm (GA), differential evolution (DE), evolution strategies (ES), particle swarm optimization, and particle swarm-inspired evolutionary algorithm (PS-EA) [6]-[8]. Numerical comparison results showed that the ABC algorithm is competitive. Due to its simplicity and ease of implementation, the ABC algorithm has captured much attention and has been applied to solve many practical optimization problems. Therefore, in this paper, we propose the application of the ABC algorithm to select the predictive and informative genes from microarray gene expression profile.

In this paper, we measure the efficiency of gene selection techniques using a support vector machine (SVM) as a classifier. An SVM displayed substantial benefits when compared to other classification approaches [9]. It is challenging to construct a linear classifier to separate the classes of data. An SVM addresses this problem by mapping the input space into a high-dimensional feature space; it then constructs a linear classification decision to classify the input data with a maximum margin hyperplane. An SVM has also been found to be more effective and faster than other machine learning methods, such as neural networks and k-nearest neighbour classifiers [10].

The proposed algorithm is tested using six binary and multi-class gene expression microarray datasets and is also compared with genetic algorithm when combined with SVM (GA-SVM), and particle swarm optimization when combined with SVM (PSO-SVM) algorithm. In addition, we compared it with other related algorithms that have been published recently. The experimental results show improvements in both the number of selected informative genes and cancer classification accuracy.

The rest of this paper is organized as follows: general background about Artificial Bee Colony (ABC) algorithm will be presented in Section II. The proposed Artificial bee colony and SVM algorithms (ABC-SVM) for gene selection and cancer classification using Microarray dataset explained in Section III. Then, Section IV outlines the experimental setup and provide results. Finally, Section V concludes our paper.

Manuscript received February 26, 2016; revised June 6, 2016.

The authors are with Computer Science Department, King Saud University, Saudi Arabia (e-mail: halshamlan@ksu.edu.sa, badrghada@hotmail.com, yousef@ksu.edu.sa).

II ARTIFICIAL BEE COLONY (ABC) ALGORITHM

The ABC algorithm is a new swarm intelligence algorithm that was invented in 2005 by Karaboga. It was inspired by the social life of bees and is used to solve the optimization problems [5]. The ABC algorithm is simple in concept, easy to implement, and has fewer control parameters; it has been widely used in many optimization applications such as protein tertiary structures (H.A. Bahamish, R. Abdullah, and R.A. Salam, 2009), digital IIR filters (N. Karaboga, 2009), artificial neural networks (D. Karaboga and B. Akay, 2005) and others. To the best of our knowledge, the ABC algorithm has not been applied before to select the predictive and informative genes from cancer microarray gene profiles. The ABC algorithm is a meta-heuristic evolutionary algorithm that simulates the search for food in a group of bees. The bees can be distributed at different distances (some of them quite far), in order to utilize the food resources [11]. The minimal model of foraging selection that leads to the emergence of collective intelligence of honey bee swarms consists of three essential components: food sources, employed foragers, and unemployed foragers. Furthermore, there are two basic behaviors: recruitment to a food source and abandonment of a food source [6]. All concepts are defined below along with the proposed variations when applied for gene selection in microarray data analysis.

- **Food Sources:** In order to select a food source, a forager bee evaluates several properties of the food source such as its closeness to the hive, richness of the energy, taste of its nectar, and the ease or difficulty of extracting this energy. For simplicity, the profitability of food source, expressed as the fitness of a food source, can be represented by only one quantity, although its fitness depends on many factors such as its proximity to the nest, its richness or concentration of its energy, and the ease of extracting this energy.
- **Employed foragers:** The honeybees who found the food source, also known as the employed bees, are equal to the number of food sources. An employed bee is employed at a specific food source that she is currently exploiting. She carries

information about this specific source and shares it with other bees waiting in the hive. The information includes the distance, the direction, and the profitability of the food source. The employed bee will become a scout when the food source has been exhausted.

- **Unemployed foragers:** A forager bee that looks for a food source to exploit is called an unemployed bees. There are two types of unemployed foragers: scouts, who conduct random searches of the environment, and onlookers who waiting in the nest and establish a food source using the information shared by the employed foragers.

The exchange of information among bees is the most important occurrence in the formation of collective knowledge. The most important part of the hive with respect to exchanging information is the dancing area. Communication among bees related to the quality of food sources takes place in the dancing area. This dance is called a Waggle dance.

III ABC-SVM ALGORITHM FOR MICROARRAY GENE SELECTION AND CANCER CLASSIFICATION

In this paper, we made some changes to the ABC algorithm representation in order to use it to solve the microarray gene selection problem. The representation of solution space (foods) for the ABC algorithm when applied on a microarray dataset illustrated in Fig. 1. The ABC algorithm first generates a random solution or initial the population of size SN, where SN denotes the size of population or the total number of food sources. When applying an ABC algorithm to gene selection for microarray data analysis, as illustrated in Fig. 1, each solution represents a group of numbers. These numbers are the indices of genes in the microarray gene expression profile (i.e., the position of food source). This is shown as x_{ij} , where i represents a particular solution ($i = 1, 2, \dots, SN$), and each solution is a D-dimensional vector ($j = 1, 2, 3, \dots, D$), where D represents the number of informative genes to be optimized in each solution. Each cell, x_{ij} , represents the corresponding gene index.

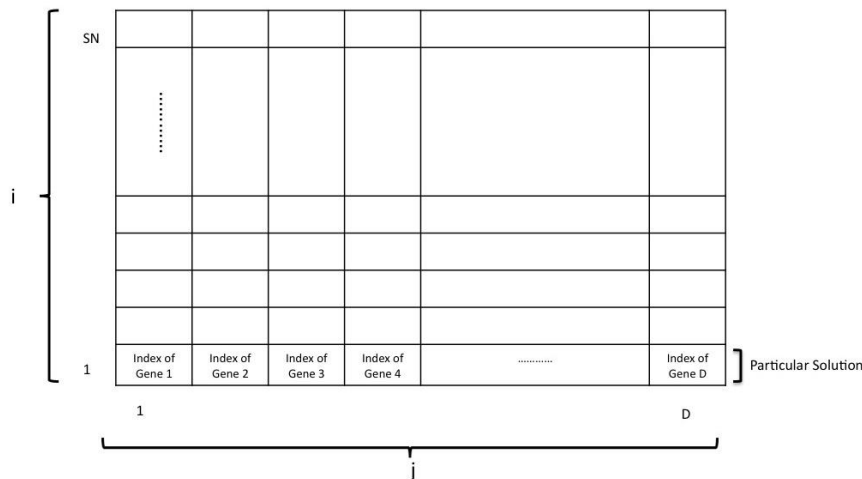


Fig. 1. The representation of solution space (Foods) for ABC algorithm when applied on Microarray Dataset). SN is the number of food sources, which is represent the solutions contain indices of genes in Microarray gene expression profile, and D represents the number of informative genes to be optimized for each solution. Each cell represent different genes index.

As show in Fig. 2, our proposed (ABC-SVM) algorithm consists of two main steps. The first one is gene selection step that is applied using Artificial Bee Colony (ABC) algorithm in order to predict the informative genes. The second step concerned in cancer classification from microarray dataset using the informative genes that is generated form first step. These steps illustrated in the following subsections:

A. Gene Selection Step (ABC Algorithm)

After initialization of random solutions (populations), the ABC algorithm starts searching for the optimal solution. In the ABC algorithm, each cycle of the search consists of three phases: (1) the employed bees phase: in which employed bees are sent to their food sources to evaluate the amount of nectar contained in each source; (2) the onlookers phase: in which, after receiving the nectar information for the food sources, the onlookers select the food source regions and evaluate the amount of nectar in the food sources; (3) the scouts bee phase: in which the scout bees are designated as such. The first half of the colony consists of the employed artificial bees and the second half includes the onlookers. For every food source, there is only one employed bee. In other words, the number of employed bees is equal to the number of food sources. The employed bee of an abandoned food source becomes a scout. The onlookers and employed bees carry out the exploitation process in the search space, while the scouts bees control the exploration process. In the next subsections, we will describe each of the phases and how can we apply them in gene selection for microarray data analysis.

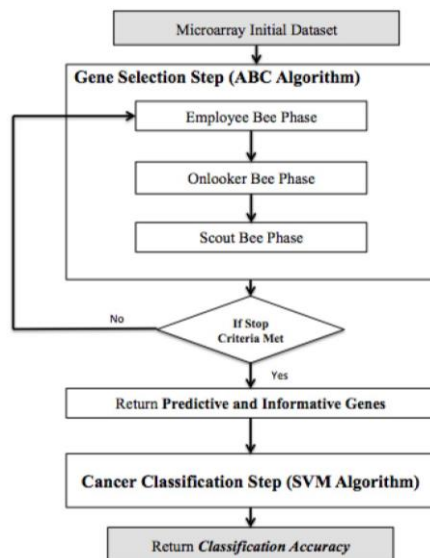


Fig. 2. The main steps and phases of proposed ABC-SVM algorithm for Microarray gene selection and cancer classification.

1) Employed Bee Phase: In this phase, the employee bees search around the solutions (food resources) at x_i and will search for the better genes index at the new location v_i . Identification of the new gene index takes place by the following equation [12]:

$$v_{ij} = x_{ij} + R_{ij}(x_{ij} - x_{kj}) \quad (1)$$

where $v_i = [v_{i1}, v_{i2}, \dots, v_{in}]$ is the new gene indices (location vector of the bees), $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]$ is

the current gene indices (location vector of the i_{th} bee), $k(k = j)$ is a correct random number in $[1, SN]$, and the SN is the number of the solutions (Artificial Bees). R_{ij} is a random number uniformly distributed in $[-1, 1]$. The random x_{ij} numbers selection from the microarray gene index is done by the following equation [12]:

$$x_{ij} = L_j + \text{rand}(0, 1) \times (U_j - L_j) \quad (2)$$

where, U_j and L_j are the top limit and the down limit of the x_i variable respectively, $U_j = (\text{Maximum_gene_index} - 1)$, and $L_j = 0$. While, $\text{rand}()$ is the random numbers function in $(0, 1)$. When the new index of the gene is identified, the optimization of it must be calculated based on the fitness function. In our problem, the fitness value fit_i determined according to the solution classification accuracy using an SVM classifier. If the new fitness value is better than the fitness value achieved thus far, then the bee moves to the new solution (food source) leaving the old one; otherwise it retains the old one.

2) Onlooker Bee Phase: After all employed bees complete this process, the information is shared with onlooker bees. An onlooker bee selects the genes depending on their winning probability value, which is similar to roulette wheel selection in genetic algorithm (GA).

3) Scout Bee Phase: Every bee (employee or onlooker) will search for better genes for a certain number of cycles (limit) and if the fitness value does not improve, then that particular bee becomes a scout bee. A solution which could not be improved through “limit” trials becomes a scout bee. A scout bee produces an index of genes randomly in the predefined search space. This will decrease the search efficacy because the best genes in the population often carry better information than others during the evolution process and the search space around the best genes could be the most promising region.

B. Cancer Classification Step (SVM Algorithm)

After the Artificial Bee Colony (ABC) algorithm finish all iteration, we use the informative and predictive genes that are generated from the ABC algorithm in the first step to train the SVM classifier. The SVM is applied again to classify the testing microarray dataset and return the classification accuracy.

The pseudo code for the proposed ABC-SVM algorithm is presented in Algorithm 1.

Algorithm 1 ABC-SVM algorithm

- 1: Set the parameter: Max Cycles, Colony Size and Limit.
- 2: Initialize the food sources.
- 3: Evaluate the food sources by calculate the fitness , which is the classification accuracy using SVM classifier.
- 4: Cycle $\leftarrow 1$
- 5: While Cycle $<$ MaxC ycles Do
- 6: Produce new solutions using employed bees.
- 7: Evaluate the new solutions by calculate the fitness (SVM classification accuracy).
- 8: Apply greedy selection process.
- 9: Calculate the probability values using fitness values.
- 10: Produce new solutions using onlooker bees based on the probability of food source.
- 11: Evaluate the new solutions by calculate the fitness (SVM classification accuracy).
- 12: Apply greedy selection process.

13: Determine abandoned solutions and generate new solutions randomly using scouts.
 14: Memorize the best solution found so far.
 15: Cycle \leftarrow Cycle + 1
 16: End While
 17: Return best solution (predictive and informative genes).
 18: Train the SVM classifier using selected genes.
 19: Classify Microarray dataset using SVM.
 20: Return the classification accuracy

IV EXPERIMENTAL SETUP AND RESULTS

A. Experiential Setup

A microarray dataset is commonly represented as an $N \times M$ matrix, where N is the number of the experimental samples and M is the number of genes involved in the experiments. Each cell in the matrix is the level of expression of a specific gene in a specific experiment. In this section, we evaluate the overall performance of gene selection methods using six popular binary and multi-class microarray cancer datasets, which were downloaded from <http://www.gems-system.org>. These datasets have been widely used to benchmark the performance of gene selection methods in bioinformatics field. The binary-class microarray datasets are: Colon [13], Leukemia [13], [14], and Lung [15]. Where the multi-class microarray datasets are: SRBCT [16], Lymphoma [17], and Leukemia [18]. In Table I, we present a detailed description of these six benchmark microarray gene expression datasets with respect to the number of classes, number of samples, and number of genes.

TABLE I: STATISTICS OF MICROARRAY CANCER DATASETS

Microarray datasets	Number of classes	Number of samples	Number of genes	Reference
Colon	2	62	2000	Alon <i>et al.</i> (1999) [13]
Leukemia1	2	72	7129	Golub <i>et al.</i> (1999)[14]
Lung	2	96	7129	Bear <i>et al.</i> (2002) [15]
SRBCT	4	83	2308	Khan <i>et al.</i> (2001) [16]
Lymphoma	3	62	4026	Aizadeh <i>et al.</i>
Leukemia2	3	72	7129	Scott <i>et al.</i> (2001) [18]

Binary class microarray datasets: The first binary class microarray dataset was obtained from cancerous and normal colon tissues. Among them, 40 samples are from tumors and 22 samples are from healthy parts of the colons of the same patients [13]. The second dataset was obtained from cancer patients with two different types of leukemia: acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). The complete dataset contains 25 AML and 47 ALL samples [14]. The last binary class dataset is the lung cancer microarray dataset [15], which includes 86 primary lung adenocarcinomas samples and 10 non-neoplastic lung samples. Each sample is described by 7,129 genes.

Multi-class microarray datasets: In our experiment, the small round blue cell tumors (SRBCTs), which contained 4 different childhood tumors were used, named so because of their similar appearance on routine histology,

which makes correct clinical diagnosis extremely challenging. However, accurate diagnosis is essential because the treatment options, responses to therapy, and prognoses vary widely depending on the diagnosis. The SRBCT datasets include 29 Ewing's sarcoma (EWS) samples, 18 neuroblastoma (NB) samples, 11 Burkitt's lymphoma (BL) samples, and 25 rhabdomyosarcoma (RMS) samples [16]. The second multi-class dataset was the lymphoma dataset, which contains the three most prevalent adult lymphoid malignancies. It contains 62 samples consisting of 4,026 genes spanning three classes, which include 42 Diffuse Large B-Cell Lymphoma (DLBCL) samples, 9 Follicular Lymphoma (FL) samples, and 11 B-cell Chronic Lymphocytic Leukemia (B-CLL) samples [17]. The last dataset was obtained from leukemia cancer patients with three different types of leukemia: acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), and mixed-lineage leukemia (MLL). The complete dataset contains 28 AML, 24 ALL, and 20 MLL samples [14].

Table II shows the control parameters for the ABC-SVM algorithm that was used in our experiments. The first control parameter is the bee colony size or population, with a value of 80. The second control parameter is the maximum cycle, which equal to the maximum number of generations. A value of 100 is used for this parameter. Another control parameter is the number of runs, which was used as stopping criteria and we used a value of 30 in our experiments, which has been shown to be acceptable. The last control parameter is the limit, which represents the maximum number of iteration allowed when the food source is not improved (exhausted). If the food source exceeds this limit, it will be selected by the scout bee. A value of 5 iterations is used for this parameter.

TABLE II: ABC-SVM CONTROL PARAMETERS

Parameter	Value
Colony Size	80
Max Cycle	100
Number of Run	30
Limit	5

In this study, we tested the performance of the proposed mRMR-ABC algorithm by comparing it with other standard bio-inspired algorithms, including ABC, GA, and PSO. We compared the performance of each gene selection approach based on two parameters: the classification accuracy and the number of predictive genes that have been used for cancer classification. We apply leave-one-out cross-validation (LOOCV) [19] in order to evaluate the performance of our proposed algorithm and the existing methods in the literature. LOOCV is very suitable to our problem, because it has the ability to prevent the "overfitting" problem [19]. It also provides an unbiased estimate of the generalization error for a stable classifiers such as the SVM classifier. In LOOCV, a single observation from the original sample is considered testing data, and the remaining observations are considered training data. This is repeated such that each observation in the sample is used once as the testing data. We implement a GA, PSO algorithm, and SVM using the Waikato Environment for Knowledge Analysis (WEKA version 3.6.10), an open source data mining tool [20]. Furthermore, in order to make

experiments more statistically valid, we conduct each experiment 30 times on each dataset. In addition, best, worst, and average results of the classification accuracies of the 30 independent runs are calculated in order to evaluate the performance of our proposed algorithm.

B. Experimental Results

TABLE III: THE PERFORMANCE OF THE PROPOSED ABC-SVM ALGORITHMS FOR COLON DATASETS

Number of genes	Classification Accuracy		
	Best	Mean	Worst
3	87.10%	85.91%	83.87%
4	87.10%	86.71%	85.48%
5	90.32%	87.98%	85.48%
6	90.32%	88.44%	85.48%
7	91.94%	90.20%	88.81%
8	91.94%	90.61%	88.81%
9	91.94%	90.95%	88.81%
10	93.55%	91.31%	88.81%
15	93.55%	91.38%	90.32%
20	95.61%	92.44%	90.32%

TABLE IV: THE PERFORMANCE OF THE PROPOSED ABC-SVM ALGORITHMS FOR LEUKEMIA 1 DATASETS

Number of genes	Classification Accuracy		
	Best	Mean	Worst
2	87.5%	86.45%	81.94%
3	88.88%	89.82%	83.33%
4	88.8%	91.15%	83.33%
5	91.99%	91.89%	87.5%
6	91.99%	92.04%	87.5%
7	93.05%	92.23%	87.5%
10	93.05%	92.38%	88.88%
13	93.05%	92.44%	88.88%
14	93.05%	92.51%	88.88%

TABLE V: THE PERFORMANCE OF THE PROPOSED ABC-SVM ALGORITHMS FOR LUNG DATASETS

Number of genes	Classification Accuracy		
	Best	Mean	Worst
2	88.54%	87.5%	84.37%
3	89.58%	88.54%	84.37%
4	91.66%	89.58%	87.5%
5	92.70%	90.03%	88.54%
6	94.79%	91.66%	88.54%
7	95.83%	92.18%	89.58%
8	97.91%	93.75%	91.66%

TABLE VI: THE PERFORMANCE OF THE PROPOSED ABC-SVM ALGORITHMS FOR SRBCT DATASETS

Number of genes	Classification Accuracy		
	Best	Mean	Worst
2	72.28%	69.87%	67.46%
3	73.34%	71.08%	68.67%
4	84.33%	81.92%	77.10%
5	87.95%	84.33%	77.10%
6	92.77%	87.99%	84.33%
8	93.97%	89.15%	84.33%
10	95.36%	91.56%	89.15%

In this section, we present and analyze the results that are obtained by our algorithm. The classification performance results for the binary-class microarray datasets: colon, leukemia1, and lung are shown in Table III, IV, and V, respectively. While, Table VI, VII, and VIII, respectively, present the comparison result for multi-class microarray datasets: SRBCT, lymphoma, and leukemia2.

TABLE VII: THE PERFORMANCE OF THE PROPOSED ABC-SVM ALGORITHMS FOR LYMPHOMA DATASETS

Number of genes	Classification Accuracy		
	Best	Mean	Worst
2	86.36%	86.36%	86.36%
3	89.39%	87.87%	86.36%
4	93.93%	89.39%	86.36%
5	96.96%	92.42%	90.90%

TABLE VIII: THE PERFORMANCE OF THE PROPOSED ABC-SVM ALGORITHMS FOR LEUKEMIA 2 DATASETS

Number of genes	Classification Accuracy		
	Best	Mean	Worst
2	84.72%	84.72%	84.72%
3	86.11%	85.23%	84.72%
4	87.5%	86.11%	84.72%
5	87.5%	86.45%	84.72%
6	90.27%	88.88%	86.11%
7	90.27%	89.22%	86.11%
8	91.66%	90.27%	88.88%
9	93.05%	91.46%	88.88%
10	93.05%	91.98%	88.88%
15	94.44%	92.78%	90.27%
18	95.83%	92.99%	90.27%
20	97.22%	93.15%	91.66%

In this research, we re-implement particle swarm optimization with SVM (PSO-SVM) and genetic algorithm with SVM (GA-SVM) in order to compare the performance of the mABC-SVM algorithm with the same parameters. We also compare it with published results for recent gene selection algorithms. Notably, all these algorithms have been combined with the SVM as a classification approach.

Table IX shows the experimental results of the ABC-SVM algorithm and other existing methods. To summarize, for all binary and multi-class microarray datasets, the existing methods tend to select few genes with high classification accuracy. In comparison, the ABC-SVM algorithm selects fewer genes than most of the existing methods with relatively high classification accuracy. Furthermore, for those methods selecting fewer genes than the ABC-SVM algorithm, their classification accuracy is less than that of the mRMR-ABC algorithm. Therefore, we can conclude that ABC-SVM is a promising approach for solving gene selection and cancer classification problems.

TABLE IX: THE CLASSIFICATION ACCURACY OF THE EXISTING GENE SELECTION ALGORITHMS UNDER COMPARISON WHEN COMBINED WITH THE SVM AS A CLASSIFIER FOR SIX MICROARRAY DATASETS. NUMBERS IN PARENTHESES DENOTE THE NUMBERS OF SELECTED GENES.

Algorithms	Colon	Leukemia1	Lung	SRBCT	Lymphoma	Leukemia2
ABC-SVM	95.61(20)	93.05(14)	97.91(8)	95.36(10)	96.96(5)	97.22(20)

GA-SVM	93.55(83)	91.99(51)	95.83(62)	92.77(74)	93.93(43)	94.44(57)
PSO-SVM	93.55(78)	95.83(53)	94.79(65)	93.97(68)	96.96(82)	95.83(61)
PSO-SVM [21]	85.48(20)	94.44(23)				
PSO-SVM [22]	87.01(2000)	93.06 (7129)				
GA-SVM [23]	93.55(12)					
mRMR-GA [24]				95(5)		
ESVM [25]			95.75(7)			

V CONCLUSION

In this research paper, we proposed applying ABC algorithm for microarray gene expression profile. In addition, we proposed to be combined with SVM as a classifier. It can be used to solve classification problems that deal with high dimensional datasets, especially microarray gene expression profile. Our proposed ABC-SVM algorithm consists of two main steps; gene selection step (ABC) and cancer classification step (SVM). The ABC algorithm is employed to select the predictive and informative genes from microarray dataset. Then, the SVM classifier was trained and tested using the selected genes and returned the classification accuracy. Extensive experiments were conducted using six binary and multi-class microarray datasets. The results showed that the proposed ABC-SVM algorithm outperforms the previously reported results. In the future, experimental results on more real and benchmark datasets to verify and extend this proposed algorithm. In addition, the mRMR-ABC algorithm can be considered as a general framework that can be used to solve various optimization problems.

REFERENCES

- [1] S. Ghorai, A. Mukherjee, S. Sengupta, and P. Dutta, "Multicategory cancer classification from gene expression data by multiclass nppc ensemble," in *Proc. International Conference on Systems in Medicine and Biology*, 2010, pp. 4–48.
- [2] L. Michael *et al.*, "Gene selection based on mutual information for the classification of multi-class cancer," in *Proc. the 2006 International Conference on Computational Intelligence and Bioinformatics*, Springer-Verlag, 2006, pp. 454–463.
- [3] N. Patrenahalli *et al.*, "A branch and bound algorithm for feature subset selection," *IEEE Transactions on Computers*, vol. 26, no. 9, pp. 917–922, 1977.
- [4] H. M. Alshamlan, G. H. Badr, and Y. A. Alohal, "The performance of bio-inspired evolutionary gene selection methods for cancer classification using microarray dataset," *International Journal of Bioscience, Biochemistry and Bioinformatics*, vol. 4, no. 3, pp. 166–170, 2014.
- [5] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Technical Erciyes University, Engineering Faculty, Computer Engineering Department, Tech. Rep., 2005.
- [6] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Applied Mathematics and Computation*, vol. 214, no. 1, pp. 108–132, 2009.
- [7] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm," *Journal of Global Optimization*, vol. 39, no. 3, pp. 459–471, 2007.
- [8] K. Dervis, and B. Basturk, "On the performance of artificial bee colony (abc) algorithm," *Appl. Soft Comput.*, vol. 8, no. 1, pp. 687–697, Jan. 2008.
- [9] H. Alshamlan, G. Badr, and Y. Alohal, "A comparative study of cancer classification methods using microarray gene expression profile," in *Proc. the First International Conference on Advanced Data and Information Engineering*, Springer Singapore, 2014, vol. 285, pp. 389–398.
- [10] X. Wang and O. Gotoh, "Microarray-based cancer prediction using soft computing approach," *Cancer Informatics*, vol. 7, pp. 123–139, 2009.
- [11] H. Li, K. Liu, and X. Li, "A comparative study of artificial bee colony, bees algorithms and differential evolution on numerical benchmark problems," *Computational Intelligence and Intelligent Systems*, vol. 107, pp. 198–207, 2010.
- [12] W.-L. Xiang and M.-Q. An, "An efficient and robust artificial bee colony algorithm for numerical optimization," *Computers & Operations Research*, vol. 40, no. 5, pp. 1256–1265, 2013.
- [13] U. Alon, N. Barkai, D. Notterman, K. Gish, S. Ybarra, D. Mack, and A. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences*, vol. 96, no. 12, pp. 6745–6750, 1999.
- [14] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, L. Coller, J. Downing, M. Caligiuri, C. Bloomfield, and E. Lander, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, 1999.
- [15] D. G. Beer, S. L. Kardia, C.-C. Huang *et al.*, "Gene-expression profiles predict survival of patients with lung adenocarcinoma," *Nature Medicine*, vol. 8, no. 8, pp. 816–824, 2002.
- [16] J. Khan, J. Wei, and M. Ringner, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, no. 6, pp. 673–679, 2001.
- [17] A. Alizadeh, M. Eisen, M. Davis *et al.*, "Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511, 2000.
- [18] S. A. Armstrong, J. E. Staunton, *et al.*, "Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia," *Nature Genetics*, vol. 30, no. 1, pp. 41–47, 2001.
- [19] A. Y. Ng, "'Preventing overfitting' of cross-validation data," *ICML*, vol. 97, 1997, pp. 245–253.
- [20] N. Z. University of Waikato (June 2014). Waikato environment for knowledge analysis. Available: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- [21] S. Qi, W.-M. Shi, K. Wei, and B.-X. Ye, "A combination of modified particle swarm optimization algorithm and support vector machine for gene selection and tumor classification," *Advances in Computer Science*, vol. 71, no. 4, pp. 157–162, 2007.
- [22] A. M. Javad and D. Giveki, "Automatic detection of erythematosquamous diseases using pso-svm based on association rules," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 603–608, 2013.
- [23] S. Peng, Q. Xu, X. B. Ling, X. Peng, W. Du, and L. Chen, "Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines," *FEBS Letters*, vol. 555, no. 2, pp. 358–362, 2003.
- [24] A. Amine, A. El Akadi, A. El Ouardighi, and D. Aboutajdine, "A new gene selection approach based on minimum redundancy-maximum relevance (mrmr) and genetic algorithm (ga)," in *Proc. IEEE/ACS International Conference on Computer Systems and Applications*, 2009, pp. 69–75.
- [25] H.-L. Huang and F.-L. Chang, "Esvm: Evolutionary support vector machine for automatic feature selection and classification of microarray data," *Biosystems*, vol. 90, no. 2, pp. 516–528, 2007.

Hala M. Alshamlan is a lecturer in Information Technology Department, King Saud University Riyadh, Saudi Arabia. Currently, she is a Ph.D. candidate in Computer Science Department, King Saud University. Her research of Interest in bioinformatics, artificial intelligent, data mining, and machine learning. And, she has many publications in these areas.

Ghada H. Badr completed her Ph.D in 2006 in computer science at Carleton University, School of Computer Science, Ottawa, Canada. She was the winner of the University Senate Medal for outstanding research achievements. From 2006 to 2007, she worked as a research associate at the National Research Council (NRC) of Canada for the language technology group in Gatineau, Canada, in the field of Machine Translation. In 2007, she won the prestigious NSERC Postdoctoral Fellowship in Canada. Since 2007 till 2011, she worked as a postdoctoral fellow at the University of Ottawa, Ottawa, Canada, where her research focused in the field of Data mining in bioinformatics. She worked, and still working, on another research project in other bioinformatics aspects, where it involves

discovering and localizing interacting RNA secondary structures. At KSU, she was able to establish the Bioinformatics Research group (BioInG), where she is the coordinator for the group since Fall 2012. Through the group she was able to attract a lot of researchers from different departments and to develop a lot of activities and workshops.

Yousef A. Alohali is associated professor in Computer Science Department at King Saud University. He completed his Ph.D in computer science at Concordia University, Montreal, Canada. Her research of interested included: artificial intelligence: human computer interaction, data mining, office automation, intelligent solutions, and optimization techniques.