FF-SVM: New FireFly-based Gene Selection Algorithm for Microarray Cancer Classification

1st Nada Almugren
Information Technology Department
King Saud Universty
Riyadh, Saudi Arabia
nada.almugren@hotmail.com

2nd Hala Alshamlan
Information Technology Department
King Saud Universty
Riyadh, Saudi Arabia
halshamlan@ksu.edu.sa

***Abstract*—Several bio-inspired evolutionary based feature selection algorithms for microarray data classification have been proposed in the literature and show a good performance. In this research we proposed a wrapper feature selection algorithm for classifying cancer microarray gene expression profile that uses FireFly algorithm along with SVM classifier named FF-SVM. Support vector machine SVM classifier with leave-one-out cross validation LOOCV are used to measure the classification accuracy for the selected gene subset. Five benchmark microarray datasets of binary and multi class are used to evaluate FF-SVM algorithm. To validate the result of the proposed algorithm we compare it with other related state-of-the-art algorithms. The experiment proves that the FF-SVM show high classification accuracy using small number of selected genes.**

***Index Terms*—Gene Selection, Cancer Classification, Microarray, Gene Expression Profile, FireFly, SVM.**

I. Introduction

DNA Microarray technology is a powerful tool that helps researchers monitor the gene expression level in an organism. Microarray data analysis is used to determine which genes are being differentially expressed. Differently expressed genes can be used in cancer diagnosing to differentiate between infected and uninfected tissues. Microarray data analysis provides valuable results which contribute towards solving gene expression profile problems. One the most important applications of Microarray data analysis is cancer classification. Classifying microarray data is challenging and considered as (NP)-Hard problem due to the high dimensionality found in a small sample size of gene expression data [1]. Also, gene expression data has a high complexity; genes are directly or indirectly correlated to each other [2].

Hence, most practical method to overcome these challenges is therefore a feature selection technique. The main idea behind the feature selection method is selecting the most informative and significant genes for the prediction (classification) problem. Several gene selection algorithms have been reported in the literature, fall in three categories filter, wrapper and hybrid. Many filter approach statistical algorithms have been used for dimension reduction to remove redundant and irrelevant genes without using any learning algorithms, e.g. Mutual Information [3], [4]. In addition to these filter approaches, several wrapper algorithms and machine learning algorithms have been applied [5], [6]. Wrapper methods have achieved better performance than filter methods [7]. The hybrid approach is also adopted in order to utilize the advantages of both the filter and wrapper approaches [6], [8]–[10].

978-1-7281-1462-0/19/$31.00 ©2019 IEEE

The aim of this study is to identify the most informative genes that contribute to cancer diagnosis. Therefore, we developed a new wrapper feature selection method for Microarray gene expression profiles to select the most informative genes that cause cancer. The proposed method consists of two phases: gene selections phase and classification phase. In the gene selection phase, the Firefly wrapper method was employed to find the optimal gene subset. In the classification phase, this optimal gene subset is tested based on a Support Vector Machine (SVM) classifier and the classification accuracy is obtained using leave-one-out cross validation (LOOCV). Five Microarray benchmark datasets of different cancer types was used to evaluate the proposed model. To validate the effectiveness of the proposed algorithm we compare it with other state-of-the-art algorithms. The experiment result shows the improvement in term of classification accuracy and the number of selected genes.

The rest of this paper is organized as follows: In Section 2, we briefly present background about FireFly algorithm and SVM classifier. This is followed by an explanation of our proposed *FF-SVM* algorithm in Section 3. Subsequently, Section 4 outlines the experimental setup and provides results. Finally, Section 5 draws conclusions of this paper.

II. Background

In this section, we briefly present a general background about FireFly algorithm (FFA) and Support Vector Machine (SVM) classification method.

A. FireFly Algorithm (FFA)

Firefly algorithm is bio-inspired global optimization method inspired by the flashing light patterns and behaviour of firefly

insects thus it simulates the attraction behaviour of fireflies [11]. Fireflies use their flashing pattern to attract other fireflies (from the opposite sex). Though, firefly algorithms development was based on three idealised rules: first, being the assumption that all fireflies are attracted to each other, regardless of their sex. Second, their attractiveness is based on their brightness ability, thus their attractiveness will decrease as the distance between them increases. As a result, it will always be the less bright fireflies that move towards the brighter ones. Third, a fireflys brightness is determined or affected by the objective functions form. Firefly algorithm is metaheuristic population based where each firefly represents a possible solution in the search space. This section presents the main behavior of artificial firefly algorithm. Two issues can be considered with the Firefly algorithm and they are; the variation of the brightness intensity, as well as how attractiveness is formulated. In the standard Firefly algorithm attractiveness is basically determined by brightness, which is associated with the objective function. Therefore, the brightness of firefly I at specific location x can be presented as $I(x) \propto f(x)$. Attractiveness β is also relative, in that it varies with the distance r_{ij} between firefly i and j. Thus, it differs from one firefly to the next, purely based on distance. Thus, the light intensity can be formulated in Equation 1.

$$I = I_0 e^{-\gamma r_{ij}^2} \tag{1}$$

Where the I_0 is the light intensity at the beginning. The Attractiveness β of the firefly is determined by the brightness. The attractiveness can be measured as shown in Equation 2 :

$$\beta = \beta_0 e^{-\gamma r_{ij}^2} \tag{2}$$

where β_0 is the attractiveness of the firefly at $r = 0$. γ is light absorption coefficient, which is fixed as 1.0 in FA. The distance between any two fireflies i and j at x_i and x_j is calculated in Equation 3 as Cartesian distance:

$$r_{ij} = ||x_i - x_j|| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{3}$$

The movement of the firefly i towards more attractive firefly j calculated using Equation 4 , as bellow:.

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}(x_j - x_i)^2 + \alpha(rand - 0.5) \tag{4}$$

In the Equation 4 the second term refer to the attraction, and the third term is the randomize term with α as randomization parameter, $\alpha \in [0, 1]$. β_0 is always set to 1 and rand is a random number between $[0, 1]$. In firefly the α is set to allow the variation in the solution. The value of γ characterizes the variation of the attractiveness, in many applications the value of γ varies from 0.01 to 100.

B. Support Vector Machine (SVM) Classifier

SVM is a supervised machine learning algorithm, which is typically used for classification purposes, based on statistical learning theory proposed by Vapnik [12]. The SVM has been used widely in many applications related to bioinformatic and show good performance. Moreover, SVM has been extensively used to classify microarray data and it improved the classification accuracy. Main advantage for using SVM to classify microarray data, it works well when classifying high dimensional data [13]. In addition, it works well when the number of features is greater than the number of samples. SVM is centred around searching for a hyperplane that optimally divides the tuples, from one class to another. The hyperplane is found using the support vector and the margin. The support vector is calculated from the vectors (data points) that define the hyperplane. The margin is the shortest distance between the hyperplane and the nearest point (on two sides). To make SVM classifier work for multi class problem, we have two approaches (One-against-One) and (One -against-Rest). We apply (One-against-One) approach because the performance is better than (One-against-Rest) approach. Also, because its training time is shorter.

In order to evaluate the performance of the SVM classifier we apply leave-one-out cross validation (LOOCV). LOOCV is a model evaluation method, it is equal to K-fold cross validation based on it logic where K equal to N, the number of sample in the dataset. LOOCV is work as follows, takes one sample as validation data (testing data) and the remaining as training. The process is repeated such that each sample in the dataset is used once as testing data. The aim of using LOOCV in our proposed algorithm is that it can prevent the problem of overfitting [14]. On the other hand, from the computational point of view LOOCV is considered as very expensive since the many time the training process is repeated.

III. Proposed Algorithm (FF-SVM)

In this section, the proposed FF-SVM algorithm will be described. The aim of FF-SVM algorithm is to find the most informative genes that maximize the SVM classifier performance. In order to deal with overfitting problem, we apply fireflies algorithm first to reduce the dimensionality of microarray. Then adopt SVM with small number of selected genes. The basic concept of the proposed method is to compare each firefly on the swarm to every other firefly and based on the brightness (fitness value) one best firefly will be chosen. The steps of FF-SVM algorithm can be described as follows:

Step 1: Swarm Initialization The firefly algorithm first initiates a population of n fireflies x_i ,$i = 1, 2, 3, .., n$ where n is the swarms size. The fireflies are positioned randomly in the search space. Every firefly x_i in the population represent a set of predefined number of features(genes) i.e. a possible

solution to gene selection problem.

Step 2: Calculate the Fitness Function After the initialization step, the light intensity of the initial swarm that associated with the fitness function $f(x_i)$. Figure 1 represent a sample of the initial population. Where the x_i represent a firefly i.e. one possible solution with its fitness $f(x_i)$. Each firefly contains D number of genes.

Step 3: Find the Best Firefly In this step the algorithm finds the best firefly that maximize the classification accuracy while keep the minimum number of selected genes. The idea behind this step is to compare each firefly in the population with every other firefly.

Given a fixed number of generation (iteration), the proposed algorithm starts the comparison. If the fitness of the firefly i is less than the fitness of the firefly j, then firefly i will move toward firefly j. The movement of the firefly is done using Equation 4. Because of the position of the firefly i is updated its fitness must also updated. Then the algorithm follows the same procedure for subsequent iteration.

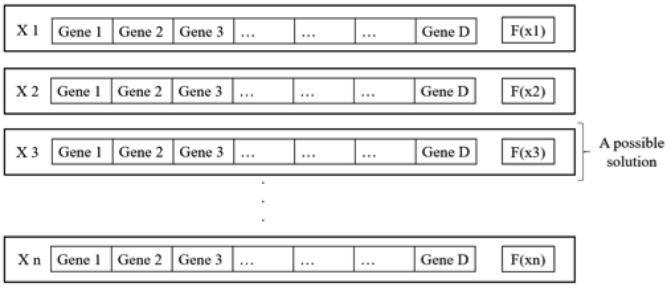

Fig. 1. The FireFly population sample. Where x_1 to x_n in the population represent a candidate solution that contain set of predefined number of genes D. The $f(x_1)$ to $f(x_n)$ is the fitness function for each firefly x_i, $i = 1, 2, ...n$.

Step 4: Ranking the Result Ranking the fireflies that resulting from the comparison step and return the best firefly. Figure 2 show the proposed algorithm flow chart.

The fitness function for the FF-SVM is to maximize classification performance i.e. classification accuracy while keeping minimum number of selected features. The classification accuracy is obtained using Leave One Out Cross Validation LOOCV. In Algorithm 1, we present pseudo code for FF-SVM algorithm.

IV. Experimental Result

In this section the datasets used to test the proposed algorithm as well as the experiment setup and experiment result are explained.

A. Microarray Dataset

In order to evaluate the proposed algorithm Five bunch mark microarray dataset have been used. The used datasets are of binary and multi class namely, SRBCT, Lung, Colon, Leukemia1

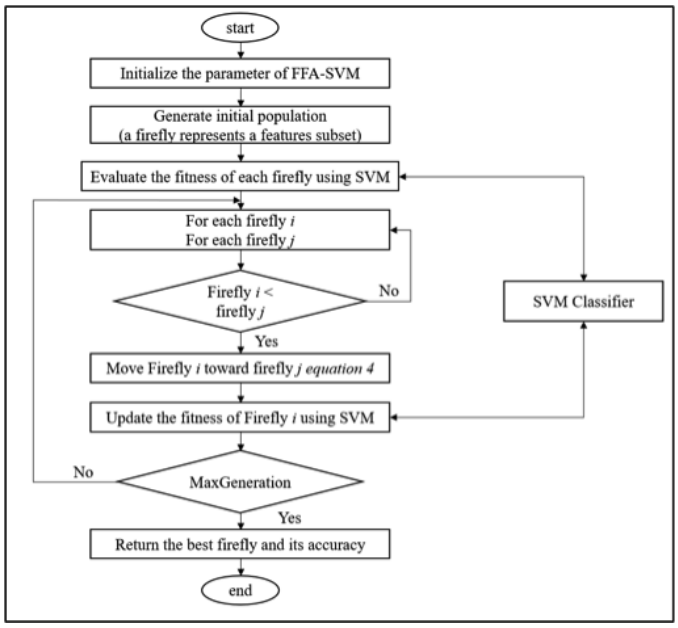

Fig. 2. Proposed (FF-SVM) algorithm flow chart.

Algorithm 1 FF-SVM Algorithm

Input:
Microarray dataset
Size of firefly i.e. number of features or genes, D.
Define light absorption coefficient, $\gamma \in [0.01, 100]$.
Define randomization parameter, $\alpha = 1$.
Define attractiveness parameter, $\beta_0 = 1$.
Set maximum number of iteration, MaxGeneration=25.
Output:
The best firefly and its fitness
Algorithm:
Generate initial population of n fireflies randomly $x_i, i = 1, 2, 3, .., n$;
Evaluate the fitness each firefly using objective function $f(x)$;
while $(t < MaxGeneration)$ **do**
  for $i = 1$ to n **do**
    for $j = 1$ to n **do**
      if $(f(x_i) < f(x_j))$ **then**
        Move firefly i towards j using
        Calculate the distance r
        Calculate the new position x_i of the firefly i
        Update the fitness of firefly i
      end if
      Evaluate new solutions and update light intensity
    end for
  end for
end while
Rank the fireflies and find best firefly

and Leukemia2. Table 1 present a detailed description of these datasets.

TABLE I
DESCRIPTION OF MICROARRAY DATASETS.

DATA SET	NO. CLASSES	NO. SAMPLE	NO. GENES	NO. SAMPLE IN EACH CLASS
LEUKEMIA2 [15]	3	72	7129	28 AML, 24 ALL, AND 20 MLL
SRBCT [16]	4	83	2308	29 EWS, 18 NB, 11 BL, AND 25 RMS
LUNG [17]	2	96	7129	86 CANCER AND 10 NORMAL
LEUKEMIA1 [2]	2	72	7129	25 AML AND 47 ALL
COLON [18]	2	62	2000	40 CANCER AND 22 NORMAL

B. Experement Setup

The proposed algorithm is written in Python development environment. All the experiments are performed in iMac desktop computer 4 GHs processor and 8 GB of RAM memory. In the experiment different parameter need to be set, the number of iteration, number of population and the number of run. The number of population set to be 70 fireflies i.e. 70 possible solution. One iteration gives possibilities to get the best firefly in one generation. Therefore, multiple iteration (generation) is give the possibilities to get the optimal feature set (firefly). Thus, the number of generation is set to 25. In order to make the result more accurate and statistically valid, the experiment is repeated 20 times for each dataset. The best, worst and average of the classification accuracy is calculated. The accuracy is obtained using SVM classifier with Leave One Out Cross Validation LOOCV.

C. Results and Analysis

In this section we present and analyze the result obtained from proposed algorithm. In addition, we will compare our result with other relevant algorithms

1) Experiment Result: The objective of feature selection is to maximize the classification performance while minimizing the number of selected features. For each dataset we implement the algorithm on different number of features. For example, we implement the FF-SVM in the Lung dataset using 2,5 and 10 number of genes (feature). The result of different experiments is presented in this section.

Table 2 show the best, worst and average classification accuracy of applying FF-SVM algorithm in Colon dataset. We can notice that the best average accuracy is obtained when the number of genes is 19. The other number of selected gene does not improve the classification accuracy.

TABLE II
COLON DATASET RESULT

NO. GENES	AVERAGE	BEST	WORST
10	92.9%	95.2%	90.3%
15	92.7%	95.3%	90.3%
17	92.6%	98.4%	90.3%
19	93.5%	98.4%	90.3%

Table 3 show the best, worst and average classification accuracy of applying FF-SVM algorithm in Leukemia1 dataset. The result show that in all targeted number of genes the best accuracy achieves 100%. The highest average accuracy is obtained when the number of selected genes is 11 with 99.5%.

TABLE III
LEUKEMIA1 DATASET RESULT

NO. GENES	AVERAGE	BEST	WORST
3	98.7%	100%	97.2%
5	97.8%	100%	94%
8	99.1%	100%	97.2%
10	98.9%	100%	97.1%
11	99.5%	100%	98.6%

Table 4 show the best, worst and average classification accuracy of applying FF-SVM algorithm in Lung dataset. The result shows that in all different number of selected genes the average accuracy is 100%.

TABLE IV
LUNG DATASET RESULT

NO. GENES	AVERAGE	BEST	WORST
2	100%	100%	100%
5	100%	100%	100%
10	100%	100%	100%

Table 5 show the best, worst and average classification accuracy of applying FF-SVM algorithm in SRBCT dataset. The highest accuracy obtained when the number of selected gene is 12 and 14. For the other number of selected genes the average accuracy is higher than 92.5%.

TABLE V
SRBCT DATASET RESULT

NO. GENES	AVERAGE	BEST	WORST
5	92.6%	95.2%	90.3%
7	94.3%	95.2%	92.3%
10	96.7%	98.8%	95.1%
12	97.5%	98.8%	96.4%
14	97.5%	98.8%	96.4%

Table 6 show the best, worst and average classification accuracy of applying FF-SVM algorithm in Leukemia2 dataset. the

highest accuracy obtained when the number of selected genes is 19. For the other numbers of selected gene, the accuracy is less than 91%.

TABLE VI
LEUKEMIA2 DATASET RESULT

NO. GENES	AVERAGE	BEST	WORST
5	81.33%	83.3%	79.2%
10	88.38%	97.2%	84.5%
13	90.55%	95.8%	87.5%
15	90.94%	94%	88.9%
19	92.58%	95.8%	90.3%

2) Comparison Result: To validate effectiveness of the FF-SVM algorithm we compare it with other wrapper-based and hybrid-based state-of-the-art algorithms. The algorithms used in the comparison employed evolutionary based wrapper algorithm for the feature selection. The comparison results are presented in Table 7 in term of classification accuracy and number of selected genes.

According to Table 7 the proposed method improved the classification accuracy and reduced the search spacy complexity. When comparing the result to wrapper approach the FF-SVM outperforms the other reported wrapper methods in four out of five datasets with small number of selected genes. However, the proposed method achieves more than 92.5% in all dataset and selects less than 22 genes. In leukemia2 dataset the PSO-SVM [5] gets 95.83% which is higher than the proposed algorithm but selects 61 genes which is relatively high number of selected genes compare to 19 gene in our proposed algorithm. For the Colon dataset the FF-SVM achieve the second-best performance. In the Leukemia1, lung and SRBCT datasets the FF-SVM obtained the highest accuracy with smallest number of selected genes.

However, when compared the result to hybrid approach the FF-SVM achieve better performance in Lung dataset. for the rest of the datasets the proposed algorithm obtained lower performance. Comparing to the other wrapper-based algorithm the FF-SVM achieve good performance in terms of the accuracy and the number of selected genes.

V. CONCLUSION

In this research we proposed FF-SVM feature selection algorithm for microarray gene expression profile. The proposed algorithm was combined with SVM classifier. Experiment result show that the proposed method achieved higher accuracy with smaller number of selected genes than other wrapper-based algorithm reported in the literature. On the other hand, when compare the algorithm to hybrid algorithms, the hybrid algorithms outperform the proposed algorithm in terms of the classification accuracy and the number of selected genes. Thus, as future work we will add filter phase to the proposed algorithm so that it will the microarray dataset will be filtered. Then, the filtered data will input to the FF-SVM algorithm in order to obtain higher performance.

REFERENCES

[1] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Transactions on computers*, no. 9, pp. 917–922, 1977.

[2] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, 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*, vol. 286, no. 5439, pp. 531–537, 1999.

[3] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowe, "A survey on filter techniques for feature selection in gene expression microarray analysis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 9, no. 4, pp. 1106–1119, 2012.

[4] R. Cai, Z. Hao, X. Yang, and W. Wen, "An efficient gene selection algorithm based on mutual information," *Neurocomputing*, vol. 72, no. 4-6, pp. 991–999, 2009.

[5] H. M. Alshamlan, G. H. Badr, and Y. A. Alohali, "Abc-svm: artificial bee colony and svm method for microarray gene selection and multi class cancer classification," *International Journal of Machine Learning and Computing*, vol. 6, no. 3, pp. 184–190, 2016.

[6] H. Lu, J. Chen, K. Yan, Q. Jin, Y. Xue, and Z. Gao, "A hybrid feature selection algorithm for gene expression data classification," *Neurocomputing*, vol. 256, pp. 56–62, 2017.

[7] H. Alshamlan, G. Badr, and Y. Alohali, "A comparative study of cancer classification methods using microarray gene expression profile," in *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*. Springer, 2014, pp. 389–398.

[8] F. V. Sharbaf, S. Mosafer, and M. H. Moattar, "A hybrid gene selection approach for microarray data classification using cellular learning automata and ant colony optimization," *Genomics*, vol. 107, no. 6, pp. 231–238, 2016.

[9] M. Dashtban, M. Balafar, and P. Suravajhala, "Gene selection for tumor classification using a novel bio-inspired multi-objective approach," *Genomics*, vol. 110, no. 1, pp. 10–17, 2018.

[10] R. Aziz, C. Verma, and N. Srivastava, "A novel approach for dimension reduction of microarray," *Computational biology and chemistry*, vol. 71, pp. 161–169, 2017.

[11] X.-S. Yang, "Firefly algorithms for multimodal optimization," in *International symposium on stochastic algorithms*. Springer, 2009, pp. 169–178.

[12] V. N. Vapnik, "Adaptive and learning systems for signal processing communications, and control," *Statistical learning theory*, 1998.

[13] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[14] A. Y. Ng *et al.*, "Preventing" overfitting" of cross-validation data," in *ICML*, vol. 97, 1997, pp. 245–253.

[15] S. A. Armstrong, J. E. Staunton, L. B. Silverman, R. Pieters, M. L. den Boer, M. D. Minden, S. E. Sallan, E. S. Lander, T. R. Golub, and S. J. Korsmeyer, "Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia," *Nature genetics*, vol. 30, no. 1, p. 41, 2001.

[16] 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*, vol. 7, no. 6, p. 673, 2001.

[17] D. G. Beer, S. L. Kardia, C.-C. Huang, T. J. Giordano, A. M. Levin, D. E. Misek, L. Lin, G. Chen, T. G. Gharib, D. G. Thomas *et al.*, "Gene-expression profiles predict survival of patients with lung adenocarcinoma," *Nature medicine*, vol. 8, no. 8, p. 816, 2002.

[18] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. 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.

[19] J. C. H. Hernandez, B. Duval, and J.-K. Hao, "A genetic embedded approach for gene selection and classification of microarray data," in *European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*. Springer, 2007, pp. 90–101.

TABLE VII

COMPARISON OF THE CLASSIFICATION ACCURACY OF OUR PROPOSED ALGORITHM (FF-SVM) WITH MORE RECENT STATE-OF-THE-ART GENE SELECTION ALGORITHMS WITH REGARD TO FIVE OF THE MICROARRAY DATASETS (THE NUMBER BETWEEN PARENTHESES REPRESENTS THE NUMBER OF SELECTED GENES).

		BINARY CLASS DATASE			MULTI CLASS DATASE	
	GENE SELECTION ALGORITHMS	COLON	LUNG	LEUKEMIA1	SRBCT	LEUKEMIA2
WRAPPER APPROCHE	FF-SVM	92.7(22)	100(2)	99.5(11)	97.5(14)	92.6(19)
	PSO-SVM [5]	93.55(78)	94.79(65)	95.83(53)	93.97(68)	95.83(61)
	GA-SVM [5]	93.55(83)	95.83(62)	91.99(51)	92.77(74)	94.44(57)
	ABC-SVM [5]	92.44(20)	93.7(8)	92.5 (14)	91.5(10)	93.1(20)
	ACO-SVM [19]	91.5(8)	-	-	-	-
	GA-SVM [19]	84.6(8)	-	91.5(5)	-	-
	MOBBA-LS [9]	-	-	-	85(6)	-
	HS-GA [20]	95.9(20)	-	97.5(20)	-	-
	BPSO-CGA [21]	99.964(214)	-	-	-	100(196)
	HPSO-LS [22]	84.38 (60)	-	89.28(100)	-	-
HYPRID APPROCHE	IDGA [23]	-	-	100(15)	100(18)	-
	IG/SGA [24]	85.48 (60)	100(9)	97.06 (3)	-	-
	CLA-ACO [8]	-	-	95.95(3)	-	-
	RFR-BBHA-BAGGING [25]	91.93(3)	-	-	-	-
	ICA+ABC [10]	98.14(16)	-	98.68(12)	97.33(15)	-
	SU-HSA [26]	87.53(9)	-	100(26)	99.89(37)	100(24)
	MRMR-ABC [27]	96.77 (15)	100 (8)	100 (14)	100 (10)	100 (20)
	MIMAGA [6]	83.41(202)	-	-	88 .64 (207)	-

[20] S. A. A. Vijay and P. GaneshKumar, "Fuzzy expert system based on a novel hybrid stem cell (hsc) algorithm for classification of micro array data," *Journal of medical systems*, vol. 42, no. 4, p. 61, 2018.

[21] L.-Y. Chuang, C.-H. Yang, J.-C. Li, and C.-H. Yang, "A hybrid bpso-cga approach for gene selection and classification of microarray data," *Journal of Computational Biology*, vol. 19, no. 1, pp. 68–82, 2012.

[22] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Applied Soft Computing*, vol. 43, pp. 117–130, 2016.

[23] M. Dashtban and M. Balafar, "Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts," *Genomics*, vol. 109, no. 2, pp. 91–107, 2017.

[24] H. Salem, G. Attiya, and N. El-Fishawy, "Classification of human cancer diseases by gene expression profiles," *Applied Soft Computing*, vol. 50, pp. 124–134, 2017.

[25] E. Pashaei, M. Ozen, and N. Aydin, "Gene selection and classification approach for microarray data based on random forest ranking and bbha," in *Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference on*. IEEE, 2016, pp. 308–311.

[26] S. S. Shreem, S. Abdullah, and M. Z. A. Nazri, "Hybrid feature selection algorithm using symmetrical uncertainty and a harmony search algorithm," *International Journal of Systems Science*, vol. 47, no. 6, pp. 1312–1329, 2016.

[27] H. Alshamlan, G. Badr, and Y. Alohali, "mrmr-abc: a hybrid gene selection algorithm for cancer classification using microarray gene expression profiling," *BioMed research international*, vol. 2015, 2015.