

## Research Article

# Improving cancer diseases classification using a hybrid filter and wrapper feature subset selection

Noor Muhammed Noori\* and Omar Saber Qasim

Department of Mathematics, University of Mosul, Mosul, Iraq

## More Information

\*Address for Correspondence: Noor Muhammed Noori, Department of Mathematics, University of Mosul, Mosul, Iraq, Tel: +964 7715381243; Email: NoorMuhammedNoori@uomosul.edu.iq

Submitted: 19 December 2019

Approved: 10 February 2020

Published: 11 February 2020

How to cite this article: Noori NM, Qasim OS. Improving cancer diseases classification using a hybrid filter and wrapper feature subset selection. *Ann Proteom Bioinform.* 2020; 4: 006-0011.

DOI: dx.doi.org/10.29328/journal.apb.1001010

Copyright: © 2020 Noori NM, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: Feature selection; Gray wolf optimization; Classification; Mutual information



## Abstract

In the classification of cancer data sets, we note that they contain a number of additional features that influence the classification accuracy. There are many evolutionary algorithms that are used to define the feature and reduce dimensional patterns such as the gray wolf algorithm (GWO) after converting it from a continuous space to a discrete space. In this paper, a method of feature selection was proposed through two consecutive stages in the first stage, the fuzzy mutual information (FMI) technique is used to determine the most important feature selection of diseases dataset through a fuzzy model that was built based on the data size. In the second stage, the binary gray wolf optimization (BGWO) algorithm is used to determine a specific number of features affecting the process of classification, which came from the first stage. The proposed algorithm, FMI\_BGWO, describes efficiency and effectiveness by obtaining a higher classification accuracy and a small number of selected genes compared to other competitor algorithms.

## Introduction

Mutual information (MI) is one of the filtering methods, which is calculated between two random variables using entropy. Entropy measures an average random variable the amount of information required to describe the random variable [1].

Gray wolf optimization (GWO) is a swarm intelligent technique developed by Mirjalili, et al. [2]. The hunting techniques and the social hierarchy of wolves are mathematically modeled in order to develop GWO and perform optimization. The GWO algorithm is tested with the standard test functions that indicate that it has superior exploration and exploitation characteristics than other swarm intelligence techniques [3]. Scientists have developed the basic algorithm GWO by converting the search algorithm from a continuous search space to a discrete search space. This modified algorithm is called a binary gray wolf optimization (BGWO) algorithm, which works in binary search areas and uses binary values equal to 0 or 1 [3,4].

Feature subset selection method is a procedure to reduce the number of unnecessary feature from the original feature set [5,6]. This method is used when we have a large number of features in a dataset, where the method of determining the feature to take a number of features necessary. The feature

selection can be applied in several areas and is intended to improve the accuracy of the classification. Two basic methods are used in selecting the feature: filtering approaches, wrapper approaches.

In this study, a new FMI\_BGWO algorithm is suggested to determine the best-feature selection. Our proposed algorithm can efficiently achievement the powerful points of both FMI and BGWO algorithms in finding the most important feature. The experimental results show the excellent performance of the proposed algorithm when the number of feature is high, and the sample size is low.

The remains of this paper are organized as follows. The GWO is presented in Section 2. In Sections 3 and 4, a brief description of feature selection and MI respectively. In Section 5, the proposed algorithm is explained. Section 6 covers the obtained results and their discussion. Finally, the most important general conclusions are mentioned, in section 7.

## Grey Wolf Optimization (GWO)

The gray wolf (*Canis lupus*) is part of the Canidae family. Gray wolves are types with a very rigorous social dominant hierarchy of leadership. Males and females in the pack are leaders, and they are called alpha [2]. Alpha wolf is also known as the dominant wolf because all its orders must be followed

by all wolves in the pack. Only alpha wolves are allowed to mate in the pack. This means that the order of the pack and its regularity is more important than its strength. Beta is the second level in the hierarchy of gray wolves.

Beta are wolves that help alpha wolves make decisions or other things about the pack. The wolf beta is the best candidate to be alpha in the case if one of the wolves of alpha becomes too big or in the case of the death of someone in command.

Omega is less gray wolves. The omega wolf always offers all the other dominant wolves on it. Omega acts as a scapegoat. The last wolves allowed to eat are omega. We may apparently notice that omega is not an important member of the group, but in fact, if omega is lost, the entire pack will face problems and internal fighting.

Delta is the fourth class in the pack. Delta wolves control omega, but they must undergo alpha and beta orders. Delta shall be responsible for guards, scouts, hunters, elders, and patients and each has its own specific responsibility [7].

### Mathematical modelling

The first best solution in GWO is alpha ( $\alpha$ ), the second-best solution is beta ( $\beta$ ), and the third-best solution is Delta ( $\delta$ ). While the rest of the solutions are Omega ( $\omega$ ), which follows the rest of the first three solutions [8].

**Encircling prey:** Gray wolves surrounded prey while hunting. The following equations will represent a mathematical model about the surrounding behavior:

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (1)$$

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \quad (2)$$

where  $t$  is the iteration,  $\vec{X}_p$  is the prey position,  $\vec{X}$  is the gray wolf position,

the  $\cdot$  operator indicates vector entry wise multiplication.

Where  $\vec{A}$  and  $\vec{C}$  are coefficient vectors calculated as follows:

$$\vec{A} = 2a \cdot r_1 - a \quad (3)$$

$$\vec{C} = 2r_2 \quad (4)$$

where components of  $\vec{a}$  are linearly decreased from 2 to 0 over the course of iterations and  $r_1, r_2, a$  are random vectors in  $[0,1]$ .

**Hunting:** Gray wolves have the ability to encircle prey and locate them. Hunting is performed by a complete pack based on, Information from the alpha, beta and delta wolves, so the updating for the wolves positions is as in the following equations:

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (5)$$

$$\vec{X}_1 = \left| \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha \right|, \vec{X}_2 = \left| \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta \right|, \vec{X}_3 = \left| \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \right| \quad (6)$$

$$\vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right|, \vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right|, \vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \quad (7)$$

The updating of the parameter  $\vec{a}$  through the following equation:

$$\vec{a} = 2 - t \cdot \frac{2}{Max_{iter}} \quad (8)$$

Where  $t$  is the iteration number and  $Max_{iter}$  is the total number of iteration allowed for the optimization [9]. The pseudo code of the GWO algorithm is displayed in figure 1.

```

Start
Generate an initial population of the gray wolf  $X_i$  ( $i=1, \dots, n$ )
Generate an initial value  $A, a$ , and  $c$ 
Find the fitness function of each search agent
 $x_\alpha$  = the first best search agent
 $x_\beta$  = the second-best search agent
 $x_\delta$  = the third best search agent
while ( $t < Max_{iter}$ . of iterations).
  for each search agent
    Update the position of the current search agent by equation (5)
  end for
  Update  $A, a$ , and  $c$ 
  Find the fitness function of each search agent
  Update  $x_\alpha, x_\beta$ , and  $x_\delta$ 
  set  $t=t+1$ 
end while
return  $x_\alpha$ 
    
```

Figure 1: The pseudo code of the GWO algorithm.

### Binary Gray Wolf Optimization (BGWO)

The positions of gray wolves constantly change in space to any point. Solutions are limited to binary values such as 0 or 1 in some special problems such as feature selection. In our research, we suggested feature selection through a binary GWO algorithm. Updating the equations of wolves are a function of the positions of three vectors is,  $X_\alpha, X_\beta$  and  $X_\delta$  which represents the best three solutions and which attracts each wolf towards it. At any given time, all solutions are at a corner of a hypercube and the solutions are grouped in binary form. Based on the basic GWO algorithm, the given wolf positions are updated, a binary restriction must be maintained according to equation (9).

The main update equation in the GWO algorithm is written as [3]:

$$X_i^{t+1} = Crossover(X_1, X_2, X_3) \quad (9)$$

Where  $(X_1, X_2, X_3)$  they are binary vectors representing wolves in bGWO and  $crossover(a, b, c)$  are an appropriate intersection between solutions  $(a, b, c)$  and  $(X_1, X_2, X_3)$ , and the wolves calculate alpha, beta and delta in order using equations (10), (13) and (16).



$$X_1^d = \begin{cases} 1 & \text{if } (X_\alpha^d + bstep_\alpha^d) \geq 1 \\ 0 & \text{other wise} \end{cases} \quad (10)$$

Where  $bstep_\alpha^d$  is a binary step in dimension  $d$ ,  $X_\alpha^d$  is a vector representing the position of the alpha wolf in dimension  $d$ .  $bstep_\alpha^d$  is calculated by the following equation:

$$bstep_\alpha^d = \begin{cases} 1 & \text{if } cstep_\alpha^d \geq rand \\ 0 & \text{other wise} \end{cases} \quad (11)$$

where  $cstep_\alpha^d$  is the continuous-valued step size for dimension  $d$ ,  $rand$  is a random number in the closed period  $[0,1]$  derived from the uniform distribution.  $cstep_\alpha^d$  can be calculated using the sigmoidal function by the following equation:

$$cstep_\alpha^d = \frac{1}{1 + e^{-10(A_\alpha^d D_\alpha^d - 0.5)}} \quad (12)$$

Where  $A_\alpha^d$  and  $D_\alpha^d$  are calculated using equations (3), and (5) in the dimension  $d$ .

$$X_2^d = \begin{cases} 1 & \text{if } (X_\beta^d + bstep_\beta^d) \geq 1 \\ 0 & \text{other wise} \end{cases} \quad (13)$$

Where  $bstep_\beta^d$  is a binary step in dimension  $d$ ,  $X_\beta^d$  represents the wolf position beta vector in dimension  $d$ .  $bstep_\beta^d$  is calculated by the following equation:

$$bstep_\beta^d = \begin{cases} 1 & \text{if } cstep_\beta^d \geq rand \\ 0 & \text{other wise} \end{cases} \quad (14)$$

Where  $cstep_\beta^d$  is the step size with a continuous value of dimension  $d$ , and ( $rand$ ) is a random number derived from the uniform distribution in the closed period  $[0,1]$ .

$cstep_\beta^d$  is calculated using the sigmoidal function by equation:

$$cstep_\beta^d = \frac{1}{1 + e^{-10(A_\beta^d D_\beta^d - 0.5)}} \quad (15)$$

Where  $A_\beta^d$  and  $D_\beta^d$  are calculated using equations (3), and (5) in the dimension  $d$ .

$$X_3^d = \begin{cases} 1 & \text{if } (X_\delta^d + bstep_\delta^d) \geq 1 \\ 0 & \text{other wise} \end{cases} \quad (16)$$

Where ( $X_\delta^d$ ) represents the location of the wolf delta ( $\delta$ ) in dimension  $d$ , and  $bstep_\delta^d$  is a binary step in dimension  $d$  calculated by:

$$bstep_\delta^d = \begin{cases} 1 & \text{if } cstep_\delta^d \geq rand \\ 0 & \text{other wise} \end{cases} \quad (17)$$

Where  $rand$  is a random number derived from the uniform distribution in the closed period  $[0,1]$ , and  $cstep_\delta^d$  is the step size with a continuous value of dimension  $d$  and can be calculated using the sigmoidal function by Eq. (18).

$$cstep_\delta^d = \frac{1}{1 + e^{-10(A_\delta^d D_\delta^d - 0.5)}} \quad (18)$$

Where  $A_i^d$  and  $D_\beta^d$  are calculated using equations (3), and (5) in the dimension  $d$ .

In the following equations, we will apply the intersection to each of the solutions  $a, b, c$ :

$$X_d = \begin{cases} a_d & \text{if } rand < 1/3 \\ b_d & \text{if } 1/3 \leq rand < 2/3 \\ c_d & \text{other wise} \end{cases} \quad (19)$$

where  $rand$  is a random number derived from the uniform distribution in the closed period  $[0,1]$ , and  $a_d, b_d$  and  $c_d$  represents the binary values of the first, second and third parameters of the dimension  $d$ ,  $X_d$  is the result of the exchange in dimension  $d$ .

BGWO algorithm is displayed in figure 2 [10]:

```

Start
Generate an initial population of the grey wolf  $X_i$  ( $i=1 \dots n$ )
Generate an initial value  $A, a$ , and  $c$ 
Find the fitness function of each search agent
 $x_\alpha$  = the first best search agent
 $x_\beta$  = the second-best search agent
 $x_\delta$  = the third best search agent
while ( $t < \text{Max}_{iter.}$  of iterations).
  for each search agent
    Calculate  $x_1; x_2; x_3$  using equations (10), (13), and (16)
    Apply the crossover method among  $x_1; x_2; x_3$  using the equation (9)
  end for
  Update  $A, a$ , and  $c$ 
  Find the fitness function of each search agent
  Update  $x_\alpha, x_\beta$ , and  $x_\delta$ 
  set  $t=t+1$ 
end while
return  $x_\alpha$ 
End
    
```

Figure 2: The pseudo code of the BGWO algorithm.

### Naïve bayes classifier

Naïve Bayes is one of the great models used in the classifications, which calculates the probability that a particular feature belongs to a particular category it is assumed that the constituent features of the search space are restricted and dependent on certain conditions [10]. Naïve Bayes performs mostly well in terms of simplicity of construction and ease of implementation. From the example that describes the feature vector which is  $(x_1, x_2, \dots, x_n)$ , we will look for the class  $G$  which increases the probability of:

$$P\left(\frac{X}{G}\right) = P(x_1, x_2, \dots, x_n | g) \quad (20)$$

The Naïve Bayes method will allow conditional independence between the features in the data by expressing this probability  $P\left(\frac{X}{G}\right)$  as a result of those probabilities [11,12]:

$$P\left(\frac{X}{G}\right) = \pi_{i=1}^n P\left(\frac{X}{G}\right) \quad (21)$$

### Feature selection

The feature selection method improves the performance of the algorithm by reducing the number of attributes used to describe a data set [6,13]. The purpose of features selecting in an algorithm is to reduce the number of genes used to improve classification and increase classification accuracy [14]. The feature identification algorithms consist of three parts:

1. **Search algorithm:** A subset of properties (features), which are part of the original features.
2. **Fitness function:** these input and digital assessment modes. The goal of the search algorithm is to draw attention to this function.
3. **Classifier:** It represents the required algorithm that uses the latest subset of genes (i.e., an algorithm that selects the most important features required) [15,16] (Figure 3).

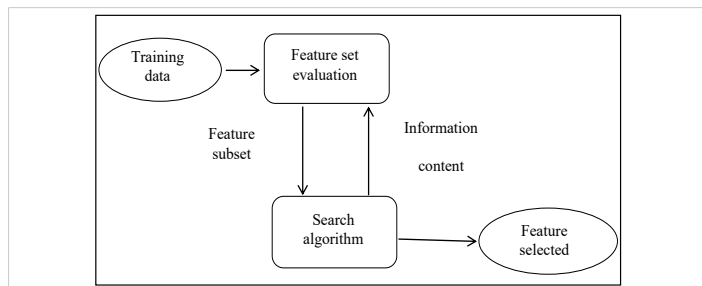


Figure 3: Filter control strategy.

One of the ways to select a feature on the predictive performance of a predefined algorithm is the wrapper method, which evaluates the quality and efficiency of certain features [17]. This is based on two-steps, based on a specific algorithm: (1) Searching for a subset of properties and features (2) Evaluate these specific features and attributes. In order to obtain learning performance or to reach some stop criteria we repeat (1) and (2) [ 12,18].

### Mutual Information (MI)

Mutual information technology is widely used in feature selection and also in Sequence alignments. Mutual information technology arranges features according to their importance to data [12]. Mutual information technology is used to find out the relationship between some features and category classifications through the following relationship:

$$MutualInfo(X_i, C) = \sum_{c \in C} \sum_{x_i \in X_i} P(X_i, y) \log \frac{P(X_i, y)}{P(X_i)P(y)} \quad (22)$$

Where  $C$  is the set of categories, and  $X_i$  is the set of features in position  $i$ .

In mutual information technology, if there is a close correlation between the category and the feature, then the

values of the mutual information will be large, but in the absence of a correlation between the feature and the category, the mutual information = 0. The mutual information can also be described as the degree of uncertainty in a feature that has been removed by knowing its category:

$$MutualInfo(X_i, C) = H(X_i) - H(X_i | C) \quad (23)$$

Where  $H(X_i)$  is an entropy in  $X_i$  and  $H(X_i | C)$  is an  $X_i$  entropy after knowing  $C$  so if knowledge of  $C$  provides a lot of information then  $H(X_i | C)$  will be low and therefore  $MutualInfo$  will be high. One disadvantage of the information exchanged favors rare features if the conditional probabilities are equal to  $P(X_{i,y})$  and this may not always be the desired result [19].

### The proposed algorithm

The proposed method for FMI\_BGWO is two basic stages. In the first stage, the fuzzy mutual information (FMI) method is used to determine the most important feature based on a fuzzy Mam dani model created by relying on two input variables that represent the size of the sample data and the number of genes per sample. Three linguistic variables (high, mean and low) were selected for both input and output. The output range was also fixed between (10-50) to select the number of genes arranged by the MI method and which will be submitted to the second phase. In the second stage, the BGWO algorithm is used to reduce and define a certain number of features, which came from the first stage. In the BGWO method, the fitness function (FF) is used to evaluate wolves' positions as follows:

$$FF = 0.1 * C + 0.9 * \left(\frac{q}{p - q}\right) \quad (24)$$

Where ( $q$ ) is the length of the selected gene subset, ( $p$ ) is the total number of features, ( $c$ ) is the accuracy of the classification model. The vector in figure 2 consists of a series of binary values 0 and 1 that represents a subset of features, in this vector, all features can be selected.

The second stage of the proposed algorithm FMI\_BGWO concentrates on the BGWO, specifically the wrapper in feature selection (Figure 4).

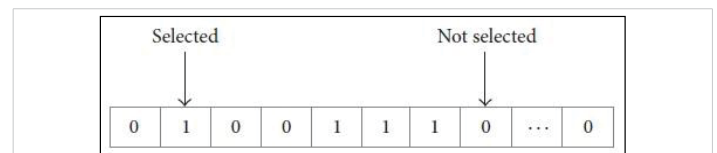


Figure 4: A sample of the gene subset solution.

Through the method (FMI) obtained the most important features, these features will be entered in the algorithm (BGWO) to reduce them as a second stage. Using an algorithm (BGWO), a certain number of features are arranged and selected in the final classification process using the Naive Bayes classifier. Figure 5 presents an elaborate flowchart of the proposed FMI\_BGWO structure.

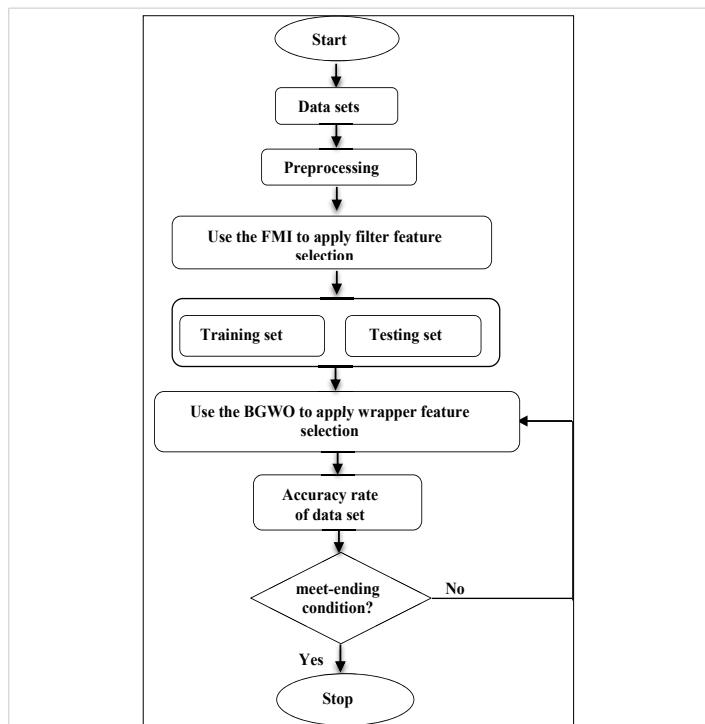


Figure 5: The architecture of the proposed FMI\_BGWO algorithm.

**Algorithm 1:** The proposed, FMI\_BGWO, algorithm.

- 1: Set the initial parameters:  $X_r, A, a, C, Max\_iteration$ .
- 2: Apply the FMI filter method to determine a subset of features by importance.
- 3: Set the initial positions Eq. (2)
- 4: Evaluate dataset by fitness function and set  $\vec{x}(r+1)$ .
- 5: Set iteration  $i$  from 1 to  $Max\_iteration$ .
- 6: Update gray wolf position and the parameter  $\vec{a}$  Eq. (5) and Eq. (8).
- 7: When  $i \leq Max\_iteration$  stop satisfied and return get the best global solution.
- 8: Choosing a subset of the features identified by the BGWO algorithm.
- 9: Return the good of features (selected features).

### Experimental results

The proposed algorithm FMI\_BGWO is evaluated, and its interest is compared with the other competitor algorithms.

#### Datasets

We have selected (3) different classification problems from the literature to verify the effectiveness of the proposed algorithm for classification problems. From a repository UCI, a data set was obtained [20].

The target variable is a binary variable that includes a set of data where the binary variable represents the condition of the sick person who has good = 1 and bad = 0.

The following table shows the overall description of the data set (Table 1).

Table 1: Description of the used datasets.

Dataset	# Samples	# Features	Target class
Leukemia	67	7130	2
Breast	168	2906	2

### Discussion and analysis

To correctly evaluate the proposed algorithm FMI-BGWO, the results were compared with the BGWO algorithm for feature selection.

The training and testing dataset for the proposed algorithm, FMI-BGWO, achieved the best results for the classification. For instance, in the breast dataset, the CA of the testing dataset is 82% by the FMI-BGWO which is higher than 100% by BGWO.

When comparing the BGWO algorithm with the proposed algorithm FMI-BGWO we note that the proposed algorithm FMI-BGWO has a clear advantage in terms of accuracy and classification capacity and BGWO worse than FMI-BGWO through the two datasets (Table 2).

Table 2: Classification performance on average of the algorithms used over 20 partitions (where the number in parentheses is the standard error).

Datasets	Methods	Number of all features	Training dataset				Testing dataset	
			CA	SE	SP	MCC	CA	
Leukemia	FMI-BGWO	21 (2.5386)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	
	BGWO	3550.9 (1.1511)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	
Breast	FMI-BGWO	17.5 (1.9003)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	
	BGWO	1447.5 (0.0412)	0.8286 (0.0366)	0.7646 (0.0486)	0.8563 (0.0352)	0.6072 (0.0822)	0.8839 (0.0478)	

### Conclusion

In this paper, to improve the performance of the classification of the big dataset, the FMI\_BGWO method has been proposed. After the feature selection process, datasets were sent to Naïve Bayes. The results of the FMI\_BGWO method were compared with the results of BGWO in table 2. Experimental results with a data set indicate that the proposed algorithm FMI\_BGWO has a better classification accuracy and a number of features less than BGWO.

### References

1. Estévez PA, Tesmer M, Perez CA, Zurada JM. Normalized mutual information feature selection. *IEEE Trans.* 2009; 20: 189–201.
2. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Adv Eng Softw.* 2014; 69: 46–61.
3. Emary E, Zawbaa HM, Hassanien AE. Binary grey wolf optimization approaches for feature selection. *Neurocomputing.* 2016; 172: 371–381.
4. Al-thanoon NA, Saber O, Yahya Z. Tuning parameter estimation in SCAD-support vector machine using firefly algorithm with application in gene selection and cancer classification. *Comput Biol Med.* 2018; 103: 262–268.  
**PubMed:** <https://www.ncbi.nlm.nih.gov/pubmed/30399534>
5. Zaffar M, Iskander S, Hashmani MA. A study of feature selection algorithms for predicting students academic performance. *Int J Adv Comput Sci Appl.* 2018; 9: 541–549.
6. Al-thanoon NA, Qasim OS, Algarnal ZY. Selection of Tuning Parameter



- in L1-Support Vector Machine via. Particle Swarm Optimization Method. 2020; 15: 310–318.
7. Singh N, Hachimi H. A new hybrid whale optimizer algorithm with mean strategy of grey wolf optimizer for global optimization. *Math Comput Appl.* 2018; 23: 14.
  8. Madadi A, Motlagh MM. Optimal control of DC motor using grey wolf optimizer algorithm. *Tech J Eng Appl Sci.* 2014; 4: 373–379.
  9. Emary E, Zawbaa HM, Grosan C. Experienced gray wolf optimization through reinforcement learning and neural networks. *IEEE Trans Neural Netw Learn Syst.* 2017; 29: 681–694.  
**PubMed:** <https://www.ncbi.nlm.nih.gov/pubmed/28092578>
  10. Manikandan SP, Manimegalai R, Hariharan M. Gene Selection from microarray data using binary grey wolf algorithm for classifying acute leukemia. *Curr Signal Transduct Ther.* 2016; 11: 76–83.
  11. Liu B, Blasch E, Chen Y, Shen D, Chen G. Scalable sentiment classification for big data analysis using naive bayes classifier. *IEEE International Conference.* 2013; 99–104.
  12. Chen J, Huang H, Tian S, Qu Y. Feature selection for text classification with Naïve Bayes. *Expert Syst Appl.* 2009; 36: 5432–5435.
  13. Ghamisi P, Benediktsson JA. Feature selection based on hybridization of genetic algorithm and particle swarm optimization. *IEEE Geosci Remote Sens Lett.* 2014; 12: 309–313.
  14. Alhafedh MAA, Qasim OS. Two-Stage Gene Selection in Microarray Dataset Using Fuzzy Mutual Information and Binary Particle Swarm Optimization. *Indian J Forensic Med Toxicol.* 2019; 13: 1162–1171.
  15. Karegowda AG, Jayaram MA, Manjunath AS. Feature subset selection problem using wrapper approach in supervised learning. *Int J Comput Appl.* 2010; 1: 13–17.
  16. Chuang LY, Chang HW, Tu CJ, Yang CH. Improved binary PSO for feature selection using gene expression data. *Comput Biol Chem.* 2008; 32: 29–38.
  17. Al-thanoon NA, Saber O, Yahya Z. Chemometrics and Intelligent Laboratory Systems A new hybrid fire fly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics. *Chemom Intell Lab Syst.* 2019; 184: 142–152.
  18. Zhao M, Fu C, Ji L, Tang K, Zhou M. Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Syst Appl.* 2011; 38: 5197–5204.
  19. Zhang JR, Zhang J, Lok TM, Lyu MR. A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. *Appl Math Comput.* 2007; 185: 1026–1037.
  20. Blake CL, Merz CJ. UCI repository of machine learning databases. University of California, Department of Information and Computer Science, Irvine, CA. 1998.