A Novel Hybrid Grey Wolf Optimization Algorithm Using Two-Phase Crossover Approach for Feature Selection and Classification

Mukesh Nimbiwal, Jyoti Vashishtha

University of Science and Technology,
Guru Jambheshwar,
India

{mukeshnimbiwal, jyoti.vst}@gmail.com

Abstract. Data mining process can be hampered by high dimensional large datasets, so feature selection become a mandatory task in prior for dimensionality reduction of datasets. Main motive of feature selection process is to choose most informative features and use them to maximize the classification accuracy. This work introduces a novel two phase crossover operator with grey wolf algorithm to solve the problem of feature selection. Two phase crossover improves the exploitation part. First phase crossover is used for feature selection and second phase used for adding some more important information and improve the classification accuracy. The KNN classifier improved the classification accuracy which is most famous classifier based on wrapper method. Ten-fold crossover validation is used to defeat the over-fitting problem which is always a milestone in the way of accuracy. Experiments are applied using various datasets and results prove that proposed algorithm outperform and provide better results.

Keywords. ALO (Ant Lion algorithm), BGOA (binary grasshopper approach), FS (feature selection), GWO (Grey Wolf Optimization), KNN (K-Nearest neighbor), PSO (Particle Swarm Optimization), TCGWO (Two-Phase Crossover Grey Wolf Optimization), WOA (Whale Optimization algorithm).

1 Introduction

SWARM optimization techniques have become the first choice of researchers for classification and feature selection. Nowadays, data mining has become a key field for research. The main motive of data mining process is to obtain the knowledge and use it for any type of conclusion and result. In the field of data mining, computational cost is always affected by the multi-dimensions of the data. Each data set contains many sets of samples. These samples provide information about a specific case which is known as features. Along with multi-dimension, one major limitation is irrelevant and redundant features. Large number of features has become limitation for traditional machine learning methods. So, the main purpose is dimension reduction and redundancy elimination has been accomplished by using feature selection methods. The main tenacity of FS is to obtain best subset of features which preserves high classification accuracy to represent original data-set features.

Various feature selection methods have been used for classification purpose. Three major classes of these methods are filter, wrapper and embedded. Filter method evaluate selected features based on characteristics whereas wrapper method provide a classification method which assess the selected subset of features. Filter approach is better than wrapper approach when we want a result in less time. On the other hand, when we talk about accuracy, wrapper approach becomes the best choice of researchers while it is a time-consuming approach. The main objective of feature selection are reducing the noisy features which helps to maximize the classification accuracy.

Many problems may occur in the way of finding best subset of features. Random search, depth search, breadth search or hybrid approaches are some searching methods for determining relevant

feature subset selection. Comprehensive search is not applicable for huge datasets because it is not possible to select best subset of features in 2d solution where d is the size of features. In the field of optimization, metaheuristic algorithm plays the main role to determine good solutions in adequate time. These algorithms have reduced computation cost by taking place of exhaustive search which was time consuming also.

On the other side, most of metaheuristic approaches suffer from lack of diversity, local optimum and imbalanced between explorative or exploitative capabilities of algorithm. So, in the field of research, these problems can be a milestone in the way of finding best solutions. These problems motivate us to find a method to overcome these feature selection problems. Grey Wolf optimization has a lot of possibilities to overcome these problems [1]. Two phase crossover operator is effectively works to advance the exploitation of algorithm. Main objective of the work are:

– A novel version of Grey Wolf Optimization (GWO) is adapted.
– Influence of two different transformation functions is evaluated on various datasets.
– Two phase crossover operator is implemented by using small probabilities.
– TCGWO is compared with other traditional metaheuristic.

This work is organized as follows. Section 2 explores literature of feature selection field. Section 3 explores the standard GWO algorithm and its working. Section 4 provide proposed algorithm named as TCGWO (Two-Phase Crossover Grey Wolf Optimization). Section 5 introduces the conclusion of work and future suggestions in this field.

2 Literature Work

Nature inspired algorithm provides a huge contribution for solving feature selection and classification problems. Various reviews exist in literature [2] produced a whale optimization algorithm which explored wrapper method to find the optimal subset.

Sigmoid function introduced by [3] for feature selection problems. Exploitation capabilities are enhanced by introducing simulated annealing. This hybrid approach enhanced the best solution after each iteration. In 2018 [4] crossover and mutation operator are used in whale optimization algorithm. The problem of local optima stagnation and slow convergence is tackled by introducing chaotic search with whale optimization algorithm by [5].

In 2019, [6] a BGOA (binary grasshopper approach) is proposed which is used two mechanism for improvisation. First mechanism employs V-shaped function for transformation function and second mechanism uses mutation operator to improve the classification accuracy.

ALO (Ant Lion algorithm) is improved by [7] applying two binary variants as S-shaped and V-shaped function. A novel binary butterfly algorithm is proposed by [8] which is based on transformation function. Quality of the solution is improved by applying simulated annealing in firefly approach in [9].

Three different variations are applied on Grey Wolf optimization to increase the performance by[10]. In [11], Grey Wolves are divided into two parts named as dominant and omega. Different approaches are applied on each part. Moreover, various classifiers i.e. decision tree, KNN, and Neural Network are used with GWO to improve the classification accuracy for Alzheimer detection [12]. Various discussed metaheuristic approaches are already existed which are proposed for feature selection [13, 14, 15, 16, 17]. Different algorithm are improved the task of feature by hybridizing various swarm based optimization algorithm in literature which have analyzed to develop this method [18-22].

3 GWO (Grey Wolf Optimizer)

Grey Wolves are wonderful hunters which have high level intelligence in catching their prey. They hunt in organized manner. This new metaheuristic has been proposed by [1] in last few years back which mimic the behavior of Grey Wolves in searching, encircling and hunting their prey. There exist four groups in their community which are divided according to their capabilities.

These four groups are named as Alpha, Beta, Delta and Omega. These are solutions of GWO where alpha, beta and delta are first, second, and

third best solution respectively. Group omega is weakest solution of GWO.

Many problems may occur in the way of finding best subset of features. Random search, depth search, breadth search or hybrid approaches are some searching methods for determining relevant features subset. Comprehensive search is inappropriate for large datasets because it is not possible to select best subset of features in 2^d solution where d is the size of features. In the field of optimization, algorithm plays the main role to obtain good solutions in adequate time.

They have reduced computation cost by taking place of exhaustive search which was time consuming also. On other side, most of s approach suffers from lack of diversity, local optimum and imbalanced between explorative or exploitative capabilities of algorithm. So, in the field of research, these problems can be a milestone in the way of finding best solutions. Problems faced by motivate us to find a method to overcome these feature selection problems. Grey Wolf optimization has a lot of possibilities to overcome these problems. Two phase crossover operator is effectively works to enhance the exploitation phase of algorithm.

Encircling step of GWO is:

$$D=|\vec{C}.\vec{X}_p-\vec{X}(t)|, \qquad (1)$$

$$\vec{X}(t+1)=|\vec{X}_p(t)-\vec{A}.\vec{D}|, \qquad (2)$$

where $\vec{X}_p$ and $\vec{X}$ are the positions of prey and grey wolves respectively. (t) is the number of iterations. $\vec{A}$ and $\vec{C}$ are coefficient vectors where:

$$\vec{A}=|2\vec{a}.\overrightarrow{rand1}-\vec{a}|, \qquad (3)$$

$$\vec{C}=2.\overrightarrow{rand2}, \qquad (4)$$

where $\overrightarrow{rand1}$ and $\overrightarrow{rand2}$ are random vector between 0 and 1. $\vec{a}$ is linearly decreasing number from 2 to 0 at each iteration:

$$a = 2 - t \times (2/i), \qquad (5)$$

where i determine maximum number of iteration. Position of wolves [X, Y] changes with respect to prey position $[X^*, Y^*]$.

Updating formula for α, β and δ are as:

$$\vec{X}(t+1) = (\vec{X}_1 + \vec{X}_2 + \vec{X}_3)/3, \qquad (6)$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1.\vec{D}_\alpha , \vec{D}_\alpha = |\vec{C}_1.\vec{X}_\alpha - \vec{X}|, \qquad (7)$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2.\vec{D}_\beta , \vec{D}_\beta = |\vec{C}_2.\vec{X}_\beta - \vec{X}|, \qquad (8)$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3.\vec{D}_\delta , \vec{D}_\delta = |\vec{C}_3.\vec{X}_\delta - \vec{X}|, \qquad (9)$$

Algorithm steps are as follows:

1. Population initialization with n grey wolves is X_i where i= 1, ..., n.
2. Parameters initialization as a, A, and C.
3. Evaluate fitness value for every grey wolf.
4. Define $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$ as three best solutions.
5. Determine t=0
6. While (t< iterations)
7.   For each grey wolf
8.   Update the position of current grey wolf using (3.6)
9.   End for
10.   Update a, A and C
11.   Evaluate fitness values of every grey wolf
12.   Update best values as $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$
13.   t = t + 1
14. End while
15. Return the $\vec{X}_\alpha$ as best value.

4 Proposed Algorithm

4.2 Initialization

Initialization phase will generate population of n number of wolves randomly. These are search agents where every agent could be a possible solution. Each agent has a dimension which is called as d. d is the amount of features in the given dataset. In almost all the classification problems, one common task is feature selection.

Feature selection is a process to select a few features which helps in maximizing the classification accuracy. So firstly, we specify some features and leave some feature.

We configure these with 0's and 1's.

4.2 Evaluation

In multi objective environment, problem must achieve a best solution. So feature selection is also a multi objective problem which can be accomplished by two major tasks:

Table 1. Data-sets description

Sr. No.	Datasets names	Number of classes	Number of features	Number of Samples
1	Breast cancer Coimbra	2	9	116
2	Breast cancer Tissue	6	9	106
3	Climate	2	20	540
4	German	2	24	1000
5	Vehicle	4	18	846
6	WineEW	3	13	178
7	Zoo	7	16	101
8	Lung Cancer	2	21	226
9	HeartEW	5	13	270
10	Parkinson	2	22	195

Table 2. Comparison of GWO-S with two phase crossover TCGWO

Datasets	GWO-S Algorithm		TCGWO1		TCGWO2		Full	
Datasets	Number of selected features	Classification accuracy	Number of selected features	Classification accuracy	Number of selected features	Classification accuracy	Number of selected features	Classification accuracy
Breast cancer Coimbra	4	0.7354	4	0.7354	4	0.7354	9	0.3715
Breast cancer Tissue	4	0.3296	4	0.3296	4	0.3300	9	0.1995
Climate	8	0.9201	7	0.9201	6	0.9312	20	0.8868
German	19	0.7402	15	0.7550	15	0.7550	24	0.6858
Vehicle	13	0.7250	9	0.7314	10	0.7340	18	0.6570
WineEW	11	0.9342	7	0.9420	7	0.9420	13	0.6642
Zoo	10	0.9595	9	0.9595	8	0.9595	16	0.9400
Lung Cancer	11	0.8763	7	0.8796	6	0.8810	21	0.8312
HeartEW	8	0.8123	6	0.8306	6	0.8402	13	0.6620
Parkinson	10	0.8576	8	0.8601	6	0.8605	22	0.7420

– Select minimum number of features.

– Maximize classification accuracy.

In this way, first milestone is to evaluate the fitness function which is:

$$\text{Fitness} = \alpha\,\gamma_r\,(D) + \beta\,(|s|)/(|d|), \tag{10}$$

where $\gamma_r\,(D)$ indicates classification error rate and it is examine by using KNN classifier [12].

|s| and |d| are cardinalities of selected feature subset's length and of all features of each sample, where α and β are weight parameters. Also α ϵ [0,1] and β = 1–α.

The prominent impact and weight is assigned to classification accuracy. But by considering the evaluation function as classification we are neglecting the possibilities of reduced features. KNN classifier is very famous classifier because of its simple implementation facilities. It has become a common practice that data set is divided into two parts as training and testing data. It is essential for testing data to prescribe its K-nearest neighbor from training data. For this purpose the Euclidean distance formula is:

$$Euc_D = \sum_{h-1}^{d}(trf_h - tsf_h), \tag{11}$$

where trf_h and tsf_h are just single feature. h is a variable. d indicates the number of features in every sample. Over- fitting problem may occur when we try to train the classifier. To overcome this problem, we use cross validation. We applied k-fold cross validation by assign the value of k as 10. Classifier is trained on k-1 partitions.

These partitions will be combined together. After this process classifier will be applied on test data to anticipate the class label for every partition. Incorrect predictions will be calculated on percentage base which is called classification percentage error rate.

4.3 Transformation Function

Grey wolf optimization approach generates agent's position as continuous values. So, because of binary nature, it is not possible to apply GWO. So this transformation function is responsible to change these continuous values into binary values. Sigmoid function and tanh functions are used to convert these continuous values in binary values. Formula to obtain this function is:

$$X_{s_i} = \frac{1}{1+e^{-x_i}} \quad , \quad X_b = \begin{cases} 0 \; if \; rand < X_{s_i} \\ 1 \; if \; rand \geq X_{s_i} \end{cases} \quad (12)$$

$$X_{v_i} = |tanh(X) , \quad X_b = \begin{cases} 0 \; if \; rand < X_{v_i} \\ 1 \; if \; rand \geq X_{v_i} \end{cases} \quad (13)$$

In this formula X_{s_i} and X_{v_i} are continuous values and i=1, …, d. X_{binary} Value can be 0 to 1. Rand is a random number which can be compare with X_{s_i} and X_{v_i}.

4.4 Two Phase Crossover

Next phase of proposed algorithm is exploitation and to accomplish this task a two phase crossover operator is used. The first phase will diminish the selected features by preserving classification accuracy. Second phase will increase classification accuracy because it will add some more important features. Proposed algorithm can be explained by following steps:

1. Initialize a population of n grey wolves X_i where i= 1, . . . , n
2. Each grey wolf is a vector and it has d dimensions that are equal to the size of the problem.
3. Initialize parameters of GWO as a, A and C
4. Fitness calculation for each grey wolf
5. Assign the best three grey wolves as $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$ respectively.
6. Define t=0
7. Repeat step 8 to 14 till t complete all the iterations
8. Update each grey wolf position by using equation 6.
9. Update a, A and C
10. Change these continuous value using equations 12 and 13
11. Fitness evaluation for all grey wolf again
12. Update $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$ and increase value of t by 1
13. Apply two phase crossover operator on $\vec{X}_\alpha$
14. Update the fitness of $\vec{X}_\alpha$
15. Return the best value as $\vec{X}_\alpha$

Step 13 has the following steps:

First Phase Crossover

1. Take $\vec{X}_\alpha$ as input for first crossover with $\vec{X}_\beta$
2. Define vector one-position to store the location of selected features in $\vec{X}_\alpha$.
3. Determine $X_{crossover1} = \vec{X}_\alpha$.
4. For I = 1 to length of one-position
5.   Generate a random number r
6.   If (r <M_p)
7. $X_{crossover1}$[one-position[i]] = 0 while keeping other features
8.   Fitness_crossover = the fitness of $X_{crossover1}$
9.   If (fitness_crossover < fitness)
10.   Fitness = fitness_crossover
11. $\vec{X}_\alpha$ = $X_{crossover1}$
12.   Endif
13.   Endif
14. Endfor

Second Phase Crossover

1. Determine vector zero-position to store the location of unselected features in $\vec{X}_\alpha$
2. Define $X_{crossover2}$
3. For j = 1 to length of zero-position
4.   Generate random number r
5.   If (r <M_p)
6. $X_{crossover2}$[zero_position[j]] = 1 while keeping the other features
7.   Fitnessf _crossover = the fitness of $X_{crossover2}$
8.   If (fitness_crossover < fitness)
9.   Fitness = fitness_crossover
10. $\vec{X}_\alpha$ = $X_{crossover2}$
11.   Endif
12.   Endif
13. Endfor
14. Output: the improved value of $\vec{X}_\alpha$

Table 3. Comparison of TCGWO with various metaheuristic based on classification accuracy

Datasets	PSO	bWOA	FA	bGWOA	Full	TCGWO
Breast cancer Coimbra	0.7354	0.7354	0.7354	0.7354	0.3715	0.7354
Breast cancer Tissue	0.3200	0.3200	0.3300	0.3300	0.1995	0.3300
Climate	0.9202	0.9220	0.9270	0.9239	0.8868	0.9312
German	0.7410	0.7395	0.7430	0.7440	0.6858	0.7550
Vehicle	0.7256	0.7301	0.7261	0.7290	0.6570	0.7340
WineEW	0.9310	0.9320	0.9420	0.9350	0.6642	0.9420
Zoo	0.9595	0.9595	0.9595	0.9595	0.9400	0.9595
Lung Cancer	0.8761	0.8764	0.8764	0.8864	0.8312	0.8810
HeartEW	0.8215	0.8160	0.8402	0.8402	0.6620	0.8402
Parkinson	0.8605	0.8590	0.8605	0.8570	0.7420	0.8605

Table 4. Comparison of average fitness values

Datasets	PSO	bWOA	FA	bGWOA	Full	TCGWO
Breast cancer Coimbra	0.2802	0.2740	0.2820	0.2740	0.3220	0.2670
Breast cancer Tissue	0.7041	0.7012	0.6980	0.6890	0.8040	0.6821
Climate	0.0890	0.0865	0.0838	0.0890	0.1210	0.0806
German	0.2750	0.2723	0.2704	0.2720	0.3230	0.2602
Vehicle	0.2870	0.2843	0.2873	0.2874	0.3512	0.2750
WineEW	0.0799	0.0810	0.0752	0.0790	0.0849	0.0642
Zoo	0.0528	0.0484	0.0528	0.0460	0.0860	0.0492
Lung Cancer	0.1380	0.1379	0.1362	0.1372	0.1775	0.1182
HeartEW	0.2175	0.2056	0.2058	0.2061	0.3438	0.1770
Parkinson	0.1662	0.1565	0.1672	0.1651	0.5630	0.1406

5 Experimental Study

We have applied the proposed approach on Windows 10 64-bit operating system; intel i3 CPU; 6 GB RAM.

5.1 Used Datasets

To check the strength of TCGWO algorithm, we have used 10 datasets which are taken from UCI repository. Table 1 describe the data sets which determine attributes names number of classes, number of feature and number of samples.

5.2 Parameters Setting

Performance of TCGWO is compared with various well known metaheuristic algorithms named as:

– Binary whale optimization algorithm (bWOA) [23],
– Particle swarm optimization algorithm (PSO) [24],

– Binary grey wolf optimization algorithm (bGWOA) [25],

– Flower algorithm (FA) [26].

Number of iterations is set to 30 for all experiments and also 10-fold cross validation strategy is applied. Most famous wrapper approach of feature selection is KNN classifier because of its supervised learning capabilities.

Training and testing data is divided in 9:1, where 9 is for training and 1 is used for testing data. Training data is used to train the KNN classifier.

5.3 The Effect of The Two-Phase Crossover Operator on the Proposed Algorithm

Various research work in this field already concluded that the performance of GWO-S id more effective than other techniques so we have compared the GWO-S with first and second phase crossover. In this experiment, it is clarified that the proposed algorithm provides better results than GWO-S.

Table 2 describes this comparison in very efficient way. Bold parts are the capacity of the proposed algorithm which declare that novel approach outperform the other algorithms in feature selection and classification accuracy.

5.4 Comparison of the Proposed Algorithm and other Metaheuristic Algorithms

In next experiment, we provide comparison results of TCGWO with various most famous swarm optimization approaches (FA, PSO, bWOA, bGWOA) in the form of classification accuracy. Selection. Bold values in table 3 determines that these results are provide equal or better results in comparison to the previously proposed algorithms. Fitness value is another very important parameter of the comparison. Table 4 presents the fitness values results. Algorithm having lower fitness value declared the best performance of the same. Results shows that TCGWO algorithm has lower fitness values with respect to other algorithms.

If average fitness value is calculated than it also declared the better performance of proposed algorithm. Most of the results justify that proposed algorithm perform in better way in the field of classification accuracy and feature. Robustness

and quality of novel TCGWO algorithm is improved by using sigmoid function and also two phase crossover operator is used to improve the exploitation phase of TCGWO.

KNN classifier increases the quality of solution because of its best wrapper method and effective learning capabilities. These statistical analysis show the superiority of the TCGWO in term of feature selection, fitness values and classification accuracy.

6 Conclusion

This work introduces a new approach to handle the feature selection issues with grey wolf algorithm. It is improved by using crossover operator in two phase manner. Results declare that crossover operator improves the results of classification accuracy. These two phase follow the small probabilities to improve the results. KNN classifier plays great effort to train the model. Over fitting problem is solved by using 10-fold cross validation.

TCGWO provide better results in comparison of FA, WA, MVO and PSO algorithms.

References

1. **Mirjalili, S., Mirjalili, S.M., Lewis, A. (2014).** Grey wolf optimizer. Advances in Engineering Software, Vol. 69, pp. 46–61. DOI: 10.1016/j.advengsoft.2013.12.007.

2. **Tawhid, M.A., Dsouza, K.B. (2018).** Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems. Applied Computing and Informatics. Vol. 16, No. 1, pp. 117–136. DOI: 10.1016/j.aci.2018.04.001.

3. **Eid, H.F. (2018).** Binary whale optimization: An effective swarm algorithm for feature selection. International Journal of Metaheuristics, Vol. 7, No. 1, pp. 67–79. DOI: 10.1504/IJMHEUR. 2018.091880.

4. **Zheng, Y., Li, Y., Wang, G., Chen, Y., Xu, Q., Fan, J., Cui, X. (2018).** A novel hybrid algorithm for feature selection based on whale optimization algorithm. IEEE Access, Vol. 7, pp. 14908–14923. DOI: 10.1109/ACCESS. 2018.2879848.

5. **Sayed, G.I., Darwish, A., Hassanien, A.E. (2018).** A new chaotic whale optimization algorithm for features selection. Journal of Classification, Vol. 35, No. 2, pp. 300–344. DOI: 10.1007/s00357-018-9261-2.

6. **Mafarja, M., Aljarah, I., Faris, H., Hammouri, A.I., Ala'M, A.Z., Mirjalili, S. (2019).** Binary grasshopper optimization algorithm approaches for feature selection problems. Expert Systems with Applications, Vol. 117, pp. 267–286. DOI: 10.1016/j.eswa.2018.09. 015.

7. **Mafarja, M., Eleyan, D., Abdullah, S., Mirjalili, S. (2017).** S-shaped vs. V-shaped transfer functions for ant lion optimization algorithm in feature selection problem. Proceedings of the International Conference on Future Networks and Distributed Systems, 21, pp. 1–7. DOI: 10.1145/3102304.3102325.

8. **Arora, S., Anand, P. (2019).** Binary butterfly optimization approaches for feature selection. Expert Systems with Applications, Vol. 116, pp. 147–160. DOI: 10.1016/j.eswa.2018.08. 051.

9. **Zhang, L., Mistry, K., Lim, C.P., Neoh, S.C. (2018).** Feature selection using firefly optimization for classification and regression models. Decision Support Systems, Vol. 106, pp. 64–85. DOI: 10.1016/j.dss.2017.12.001.

10. **Tu, Q., Chen, X., Liu, X. (2019).** Hierarchy strengthened grey wolf optimizer for numerical optimization and feature selection. IEEE Access, Vol. 7, pp. 78012–78028. DOI: 10.1109/ACCESS.2019.2921793.

11. **Wang, H., Jing, X., Niu, B. (2017).** A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data. Knowledge-Based Systems, Vol. 126, pp. 8–19. DOI: 10.1016/j.knosys.2017. 04.004.

12. **Shankar, K., Lakshmanaprabu, S.K., Khanna, A., Tanwar, S., Rodrigues, J.J., Roy, N.R. (2019).** Alzheimer detection using group grey wolf optimization based features with convolutional classifier. Computers and Electrical Engineering, Vol. 77, pp. 230–243. DOI: 10.1016/j.compeleceng.2019.06.001.

13. **Hafez, A.I., Zawbaa, H.M., Emary, E., Hassanien, A.E. (2016).** Sine cosine optimization algorithm for feature selection. IEEE. International Symposium on Innovations in Intelligent Systems and Applications INISTA. pp. 1–5. DOI: 10.1109/INISTA.2016. 7571853.

14. **Jain, K., Purohit, A. (2017).** Feature selection using modified particle swarm optimization. International Journal of Computer Applications, Vol. 161, No. 7, pp. 8–12.

15. **Li, T., Dong, H., Sun, J. (2019).** Binary differential evolution based on individual entropy for feature subset optimization. IEEE Access, Vol. 7, pp. 24109–24121. DOI: 10.1109/ACCESS.2019.2900078.

16. **Wang, H., Jing, X., Niu, B. (2017).** A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data. Knowledge-Based Systems, Vol. 126, pp. 8–19. DOI: 10.1016/j.knosys.2017. 04.004.

17. **Kashef, S., Nezamabadi-Pour, H. (2015).** An advanced ACO algorithm for feature subset selection. Neurocomputing, Vol. 147, pp. 271–279. DOI: 10.1016/j.neucom.2014.06.067.

18. **Cho, J.H., Lee, D.J., Park, J.I., Chun, M.G. (2013).** Hybrid feature selection using genetic algorithm and information theory. International Journal of Fuzzy Logic and Intelligent Systems, Vol. 13, No. 1, pp. 73–82. DOI: 10.5391/ijfis.2013.13.1.73.

19. **Ghamisi, P., Benediktsson, J.A. (2015).** Feature selection based on hybridization of genetic algorithm and particle swarm optimization. IEEE Geoscience and Remote Sensing Letters, Vol. 12, No. 2, pp. 309–313. DOI: 10.1109/LGRS.2014.2337320.

20. **Hafez, A.I., Hassanien, A.E., Zawbaa, H.M., Emary, E. (2015).** Hybrid monkey algorithm with krill herd algorithm optimization for feature selection. IEEE International Computer Engineering Conference (ICENCO), pp. 273–277. DOI: 10.1109/ICENCO.2015.7416361.

21. **Tawhid, M.A., Dsouza, K.B. (2018).** Hybrid binary dragonfly enhanced particle swarm optimization algorithm for solving feature

selection problems. Mathematical Foundations of Computing, Vol. 1, No. 2, pp. 181–200.

22. **Tawhid, M.A., Dsouza, K.B. (2018).** Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems. Applied Computing and Informatics, Vol. 16, No. 1, pp. 117–136. DOI: 10.1016/j.aci.2018.04.001.

23. **Hussien, A.G., Hassanien, A.E., Houssein, E. H., Bhattacharyya, S., Amin, M. (2019).** S-shaped binary whale optimization algorithm for fea- ture selection. Recent Trends in Signal and Image Processing Singapore, Springer Vol. 727, pp. 79–87. DOI: 10.1007/978-981-10-8863-6_9.

24. **Emary, E., Zawbaa, H.M., Hassanien, A.E. (2016).** Binary grey wolf optimization approaches for feature selection. Neurocomputing, Vol. 172, pp. 371–381. DOI: 10.1016/j.neucom.2015.06.083.

25. **Kennedy, J., Eberhart, R.C. (1997).** A discrete binary version of the particle swarm algorithm. Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation, IEEE International Conference on: Vol. 5 pp. 4104–4108. DOI: 10.1109/ICSMC. 1997.637339.

26. **Emary, E., Zawbaa, H.M., Hassanien, A.E. (2016).** Binary grey wolf optimization approaches for feature selection. Neurocomputing, Vol. 172, pp. 371–381. DOI: 10.1016/j.neucom.2015.06.083.

27. **Yang, X.S. (2012).** Flower pollination algorithm for global optimization. International Conference on Unconventional Computing and Natural Computation, Springer, Vol. 7445. pp. 240–249. DOI: 10.1007/978-3-642-32894-7_27.