j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 4, No:2. 2021

A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem

I. Al-Jadir *[1], W. A. Mahmoud [2]

[1&2] *colleg of engenaring, Uruk univarsity,, Baghdad, Iraq*

Ibrahim.amer@uruk.edu.au

Abstract Optimization methods are considered as one of the highly developed areas in Artificial Intelligence (AI). The success of the Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) has encouraged researchers to develop other methods that can obtain better performance outcomes and to be more responding to the modern needs. The Grey Wolf Optimization (GWO), and the Krill Herd (KH) are some of those methods that showed a great success in different applications in the last few years. In this paper, we propose a comparative study of using different optimization methods including KH and GWO in order to solve the problem of document feature selection for the classification problem. These methods are used to model the feature selection problem as a typical optimization method. Due to the complexity and the non-linearity of this kind of problems, it becomes necessary to use some advanced techniques to make the judgement of which features subset that is optimal to enhance the performance of classification of text documents. The test results showed the superiority of GWO over the other counterparts using the specified evaluation measures.

Crossref 10.36371/port.2020.2.9

Keywords: *Optimizatio; Machine Learning; Swarm Intelligenc;, Grey Wolf Optimization.*

1. INTRODUCTION

Optimization is one of the best and most effective methods deployed to be used to solve real-world problems that are recently not limited to the academic and the scientific research only. For instance, such methods have been used for medical purposed, business, education, industry and others. The Evolutionary Algorithms such as the Genetic Algorithms (GAs), Genetic Programming (GP) and Evolutionary Strategies were among the first methods developed in this area. By time more and more methods have been invented in order to come with the changing needs in life and to solve more and more complicated problems. The Grey Wolf Optimization is one of those promising methods [1]. Recently, several effective feature selection techniques have been developed in the literature, and applied for English language text categorization classification.

Gray wolf optimization (GWO) is a one of the recently invented techniques in optimization, that suggests the gray wolf members have the capability to successfully reproduce 3 more members than hunting in pack while 2 gray wolf members (female and male) have an extra space and management authority of other members in the group [6]. Gray wolf's optimization method is one of the biologically inspired methods that mimics the hunt processes of a group of gray wolf in the forest [1][2].

The krill herd algorithm (KHA) is a new metaheuristic search algorithm based on simulating the herding behavior of krill individuals using a Lagrangian model. This algorithm was developed by Gandomi and Alavi (2012) and the preliminary studies illustrated its potential in solving numerous complex engineering optimization problems [2]. Text labeling and classification is one of the essential computational tasks in machine learning applications due to the increased amounts of large amount of text documents available in the digital forms. In this process, feature selection (FS) is challenging phase due to thousands of possible feature sets will be considered in text classification. Text feature selection is the process of performing dimensionality reduction and analyzing a large amount of natural language text in the data mining discipline. It aims to detect useful patterns and trends from the text. Many methodologies have been developed for performing this task such as dimensionality reduction that includes feature selection and feature extraction [3][4].

The main challenges of conducting text mining operations is the handling the increasing number of data and the hidden relationships between them. In data mining terms, pre-processing is the first step that should be considered before conducting any post-processing methods [5]. Due to the high dimensionality of data such as the text data, it become necessary to consider the dimensionality reduction step as

Al-Jadir, I. & A. Mahmoud, W., Ibraheem Al-Jadir. (2021). A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem. *Journal port Science Research, 4(2),* pp.125–131. https://doi.org/10.36371/port.2021.2.9

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 4, No:2. 2021

data mining pre-processing step in order to funnel down unwanted information. The existing methods used for classification still have many challenges; due to the huge increasing amount of data [6][7]. Thus, it is essential to continue to improve and enhance these approaches and techniques which are supposed to deal with high dimensional data such as textual data. In figure 1 we notice the feature selection process is conducted as an iterative process. Figure 2, represents the original and the reduced spaces of features.

In this paper, it is suggested that a selected number of optimization methods are used to enhance the classification systems via feature selection. As wrapper methods, these optimization methods can be used to iteratively select the optimal subset of features. Thus, these techniques can assist in improving the text mining post processes as they deal with cases where imprecise, and uncertain data representations are existing. The aim of this paper is to present a method intended to reduce the extra text that affects any data mining or machine learning process with the objectives listed below:

To develop a text feature selection method to reduce the original feature subsets into smaller feature space with an eliminated chance of falling in local optima.

The problem of missing class labels of text features is also handled in this research.

This paper is organized as follows, in section 2 an explanation of the algorithm is given while in section 3 we explain the proposed method and in sections 4 and 5 we explain the test results and the conclusions.

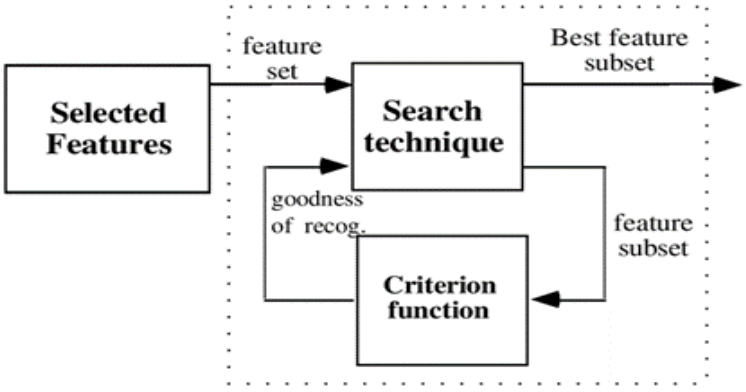

***Figure* 1**: the feature selection system [8]

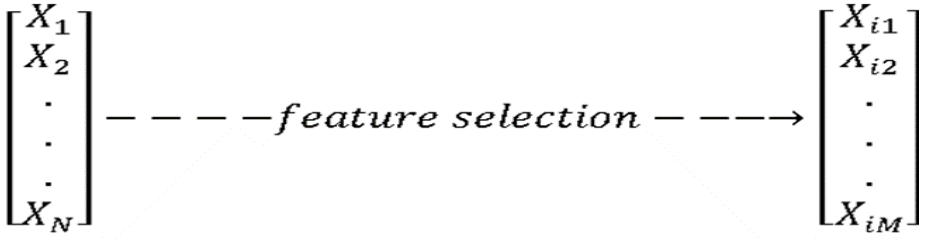

Figure 2: *Feature selection where N is the number of the original features and M is the number of reduced features i.e. m<n [9]*

2. THE PROPOSED METHODS

2.1 Krill Herd Optimization

The Krill Herd algorithm (KH) is based on the idea of krill individual's movements and foraging (Figure 3). The objective function of the KH is calculated by measuring the minimum distance of each krill from the food source, at the same time, the entire herd density is also taken into account. The krill position is determined mainly by three factors, first its distance from the food, the impact of movement generated by other krill individuals and the krill physical diffusion. These three factors can be mathematically represented as [10][11]:

The movement of the krill herd that induced by other individuals from the herd with the aim of keeping the swarm as dense as possible.

- Forage activities.
- Random diffusion.

Lagrange model can be used to generalize n-dimensional search space. The predation can remove krill from the herd and that reduces the density of the krill swarm. Also, that disturbs the way of the swarm to the food source. This is considered the initialization of the swarm. According to the

Al-Jadir, I. & A. Mahmoud, W., Ibraheem Al-Jadir. (2021). A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem. *Journal port Science Research,* 4*(2),* pp. 125–131. https://doi.org/10.36371/port.2021.2.9

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 4, No:2. 2021

three factors, a Lagrange model can be generated as is shown in equation (1).

$$\frac{dX_i}{dt} = N_i + F_i + D_i \quad (1)$$

Where: Ni the parameter that determines the motion induced by other krill individuals whereas Fi represents the foraging activity. Lastly, Di represents the diffusion of any particular krill individual.

Krill individuals strive to keep the swarm condense and move as one unit due to the impact of their mutual effects. The estimation of the swarm direction αi can be determined by measuring the local swarm density, a target swarm density, and a repulsive swarm density. Therefore, in equation (2) these three densities are used to calculate the motion induced by other krill individuals. For the first parameter Ni it can be updated by applying equation (2).

$$N_i^{new} = N^{\max} .\alpha_i + N_i^{old} .\omega_n \quad (2)$$

Where,

$$\alpha_i = \alpha_i^{local} + \alpha_i^{t\arg et}, \quad (3)$$

Nmax is the max speed induced. The ω, is the inertia weight in the range [0,1]. Niold is the final motion induced. And αilocal is the effect of the surrounding krill individuals on the motion of a particular krill while αitarget is the effect of the best krill that has the best fitness value.

The effect of the surrounding krills αilocal can be obtained from the following equations: 7

$$\alpha_i^{\text{local}} = \sum_{j=1}^{NN} \hat{K}_{ij} \hat{X}_{ij} \quad (4)$$

$$\hat{X}_{ij} = \frac{x_j - x_i}{\left\| x_j - x_i \right\| + \varepsilon} \quad (5)$$

$$\hat{K}_{ij} = \frac{K_i - K_j}{K^{worst} - K^{best}} \quad (6)$$

Where kworst, kbest are worst and best fitness scores in the swarn while ki, kj are the fitness scores of the ith and the jth krills. The Xi and Xj are the locations of ith and jth krills.

Lastly, ε is a small positive value added to denominator to prevent any singularities. The target effect that determines the effect of the best krill that has the best fitness value can be calculated as follows:

$$\alpha_i^{t\arg et} = C^{best} .\hat{K}_{i,best} .\hat{X}_{i,best} \quad (7)$$

Where;

Cbest is the effective coefficient of the krill that has the best fitness value to the ith krill. Cbest can be calculated as follows:

$$C^{best} = 2\left(rand + \frac{1}{I^{\max}} \right) \quad (8)$$

where rand is a random number between zero and one I is the iteration counter and Imax is the number of iterations.

kˆbest, is computed as same as kˆij, still the fitness value of jth krill individual kj, is substituted by the best fitness value.

Xˆi,best, is also computed as same as Xˆij, but Xj is substituted by the Xbest that represents the best fitness value.

2.2 Foraging Motion (Fi)

This factor can be computed in terms of food location and previous experience of where food was located. The foraging motion can be calculated as follows:

$$F_i = \beta_i^{food} + \beta_i^{best}, \quad (9)$$

where βfood is the food attraction that is used to attract krills to global optimum while βbest is the effect of the current best krill individual.

2.3 Physical Diffusion

The physical diffusion is an arbitrary process which is computed as a function of the maximum diffusion speed and a random directional vector as follows:

$$D_i = D^{\max} \times \delta \quad (10)$$

Where, Dmax is the maximum diffusion speed and δ is a random directional vector with the values ranging between [1,-1].

2.4 The Motion Process of KH

According to the Ni, Fi, and Di, the krill positions can be calculated during the time interval Δt as is shown in the following equation.

Al-Jadir, I. & A. Mahmoud, W., Ibraheem Al-Jadir. (2021). A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem. *Journal port Science Research*, 4*(2)*, pp.125–131. https://doi.org/10.36371/port.2021.2.9

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 4, No:2. 2021

$$x_i(t+\Delta t) = x_i(t) + \Delta t \frac{dX_i}{dt} \quad \text{.....(11)}$$

After the position updates of the krill individuals, the reproduction mechanisms are used, which are the crossover and the mutation. The krill distribution can be visualized as shown in Figure 4.

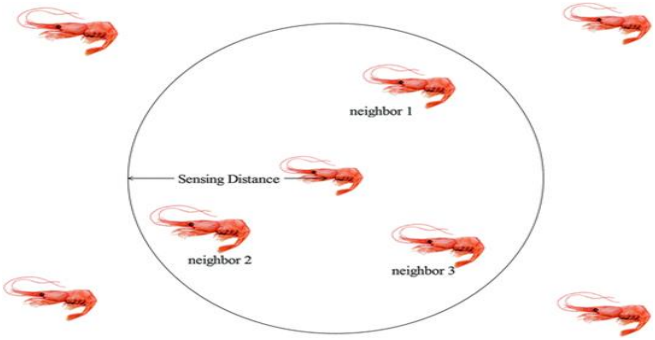

Figure 4: *krill herd distribution and distancing [11]*

2.2 Grey Wolf. Optimization

The hunting social team's attitude of the gray wolf member can be represented formally using the mathematical equations with the assistance of the optimal wolves for obtaining the optimal solution [14-17]. The w wolves are following the other prominent wolf members for the hunt process. The major steps used for hunting are listed below:

Approach the prey

Encircle and harass the prey until the stopping condition is achieved.

Attack the prey.

Prey Encirclement

To hunt, grey wolf is chasing and encircling the prey. In mathematics, it is designed as shown in Equations (1) and (2):

$$\vec{D} = \left|\vec{C}\vec{X_p}(t) - \vec{X}(t)\right| \qquad (12)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}.\vec{D} \qquad (13)$$

In this contact t is the designated present iteration and t + 1 is representing the upcoming iteration. $\vec{X}(t)$ is the feature vector of gray wolves and $\vec{X}_p(t)$ is the feature vector of pray. $\vec{A}$ and $\vec{C}$ are factors that they are represented as:

$$\vec{A} = 2.\vec{a}.\vec{r_1} - \vec{a} \qquad (14)$$

$$\vec{C} = 2.\vec{r_2} \qquad (15)$$

where r1 and r2 are the randomized vectors in the period of 0 and 1 of ~a are decreased in a linearly based manner from two to zero over the course of iteration.

The Hunt Process

Hunt can be represented in a mathematical basis as shown in Equation (7). Grey wolf encircling around for hunting when the locations of the pray is determined. Hunt mechanism is guided by α, β, δ grey wolf members. In the total searching space, there is no hint regarding the locations of the pray, merely a presumption is formed that α, β and δ wolf members have enough information relating to the pray locations. Thus, the best candidates gained are reserved and other candidates are eliminated, that means the w gray wolf members still in an update process of their locations on the basis of the optimal solutions.

$$\vec{X}(t+1) = (\vec{X}_1 + \vec{X}_2 + \vec{X}_3)/3 \qquad (16)$$

Eexploitation and Exploration are achieved during the search and attack processes to the prey. Random aspect in this process helps avoiding being stuck in the local minimum.

Dataset

Five datasets are utilized for evaluating the performance of the proposed krill herd based method. These datasets are listed below:

D1, and D2 Datasets areNews transcripts from the news broadcasted in the (http:// www. bernama. Com /bernama/v8/index.php) website.

D3, is Classic 3, a benchmark dataset containing 3 classes which are the MED, CISI and CRAN. Downloaded from: http://www.dataminingresearch.com/index.php/2010/09/classic3-classic4-datasets/.

D4 is a 20 News groups dataset, this dataset consisting of 20000 items and these items are gathered from 20 new channels. This dataset is available at the Machine Learning Repository1. A 2- class sub datasets is utilized in this research. This sub dataset is containing the talk Politics Mid-east and talk.

Reuters-21578, this dataset is available at Machine Learning Repository2. The edition used has only the label documents

Al-Jadir, I. & A. Mahmoud, W., Ibraheem Al-Jadir. (2021). A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem. *Journal port Science Research, 4(2),* pp. 125–131. https://doi.org/10.36371/port.2021.2.9

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 4, No:2. 2021

and single labelled documents. Moreover, the number of data chosen for each class is 200.

The details of the datasets are shown in Table (1).

Table 1 *datasets used in experiments*

D#	#Classes	Instances	Features
D1	six	Almost 200	Almost 4000
D2	ten	Almost 2000	Almost 15000
D3	ten	Almost 2000	Almost 13000
D4	two	Almost 1000	Almost 10000
D5	Fifty	Almost 1500	Almost 6000

Test Results

The experimental tests are made at first by utilizing the rendered attributes via the use of different methods. In order to conduct the comparison, the entire number of features are used under the name ALL, Feature Selection using GWO, KH, Harmony based optimization method (FS-HS-TC), the Feature-Selection-Genetic-Algorithm-Document-Clustering (FS-GA-TC) and feature selection using the modified Differential Evaluation method (DE) and differential evolution with simulated annealing (DE-SA) and PSO are utilized. The test results using the external evaluation measures showed that our method achieved the best results compared to the other state–of–the–art methods and outperformed other methods. This observation comes from using an intelligent method of classification with our proposed feature selection method. The GWO showed superiority over the other compared methods that used the traditional classification methods.

Table 2: *The comparison of using different methods*

		Min		Max		Mean	
	Method	F-Macro	F-Micro	F-Macro	F-Micro	F-Macro	F-Micro
Data1	GWO	0.95	0.97	0.98	0.98	0.92	0.95
	KHA	0.88	0.85	0.89	0.86	0.85	0.86
	PSO	0.84	0.84	0.87	0.88	0.83	0.85
	ALL	70.7	70.5	70.9	70.8	70.2	70.3
	DE	0.24	0.26	0.78	0.80	0.63	0.66
	FS-GA-TC	0.34	0.36	0.88	0.89	0.66	0.69
	FS-HS-TC	0.37	0.39	0.65	0.70	0.54	0.58
	FS-DE-SA	0.54	0.57	0.86	0.87	0.71	0.73
Data2	GWO	0.89	0.88	0.90	1.00	0.98	0.98
	KHA	0.46	0.51	0.69	0.73	0.55	0.59
	PSO	0.85	0.84	0.87	0.87	0.85	0.84
	ALL	0.98	0.96	0.99	0.99	0.89	0.99
	DE	0.53	0.52	0.65	0.62	0.64	0.62
	FS-GA-TC	0.40	0.48	0.60	0.68	0.47	0.53
	FS-HS-TC	0.49	0.53	0.72	0.76	0.60	0.64
	FS-DE-SA	0.42	0.48	0.76	0.79	0.64	0.69
Data3	GWO	0.91	0.97	0.92	0.98	0.92	0.98
	KHA	0.54	0.68	0.54	0.69	0.54	0.69
	PSO	0.78	0.76	0.78	0.78	0.75	0.74
	ALL	0.71	0.72	0.74	0.74	0.72	0.72
	DE	0.53	0.53	0.56	0.57	0.55	0.54
	FS-GA-TC	0.50	0.67	0.51	0.67	0.51	0.67
	FS-HS-TC	0.52	0.68	0.53	0.68	0.53	0.68
	FS-DE-SA	0.52	0.68	0.52	0.68	0.52	0.68
Data4	GWO	0.94	0.91	0.96	0.96	0.98	0.97
	KHA	0.92	0.92	0.94	0.94	0.95	0.95

Al-Jadir, I. & A. Mahmoud, W., Ibraheem Al-Jadir. (2021). A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem. *Journal port Science Research*, 4*(2)*, pp.125–131. https://doi.org/10.36371/port.2021.2.9

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 4, No:2. 2021

	PSO	0.95	0.94	0.96	0.96	0.94	0.94
	ALL	0.62	0.62	0.67	0.66	0.65	0.65
	DE	0.20	0.21	0.34	0.37	0.26	0.28
	FS-GA-TC	0.18	0.20	0.43	0.47	0.29	0.32
	FS-HS-TC	0.18	0.23	0.28	0.31	0.24	0.27
	FS-DE-SA	0.20	0.23	0.29	0.33	0.24	0.27
Data5	GWO	0.81	0.84	0.95	0.88	0.86	0.90
	KHA	0.80	0.88	0.93	0.76	0.78	0.90
	PSO	0.80	0.81	0.90	0.89	0.85	0.87
	ALL	46.4	46.6	55.5	53	0.62	0.64
	DE	0.10	0.14	0.31	0.33	0.20	0.23
	FS-GA-TC	0.10	0.14	0.38	0.41	0.22	0.24
	FS-HS-TC	0.14	0.16	0.33	0.35	0.21	0.24
	FS-DE-SA	0.12	0.14	0.42	0.45	0.21	0.26

This dataset has been chosen because they are a challenging dataset in terms of the variety of these subjects, consistent and the number of classes is varied, i.e., the number of classes is different from one dataset to another. On the other hand, these datasets are limited in their size. It is highly recommended to upgrade our proposed method dealing with big data—this number of datasets is used in our experiment to show how robust the system is working.

CONCLUSION

In this paper we introduced the use of a Grey Wolf Optimization (GWO) for the document classification using the selected subset of features generated by this method. The performance of the proposed GWO methods was compared with other optimization algorithms applied on the same documented classification for this purpose. The experiments were dedicated to test each one of those methods on the same data under the same conditions. Based on the test results given in Table 2 they reflect the superiority of the Grey Wolf Optimization over the other methods that use the external evaluation measures of classification.

REFRENCES

[1] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, *172*, 371–381. https://doi.org/10.1016/j.neucom.2015.06.083

[2] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021, July 1). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*. Association for Computing Machinery. https://doi.org/10.1145/3457607

[3] Gandomi, A. H., & Alavi, A. H. (2012). Krill herd: A new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, *17*(12), 4831–4845. https://doi.org/10.1016/j.cnsns.2012.05.010

[4] Wang, G. G., Guo, L., Gandomi, A. H., Hao, G. S., & Wang, H. (2014). Chaotic Krill Herd algorithm. *Information Sciences*, *274*, 17–34. https://doi.org/10.1016/j.ins.2014.02.123

[5] Wang, Z., Zheng, L., Wang, J., & Du, W. (2019). Research on Novel Bearing Fault Diagnosis Method Based on Improved Krill Herd Algorithm and Kernel Extreme Learning Machine. *Complexity*, *2019*. https://doi.org/10.1155/2019/4031795

[6] Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, *8*, 54776–54788. https://doi.org/10.1109/ACCESS.2020.2980942

Al-Jadir, I. & A. Mahmoud, W., Ibraheem Al-Jadir. (2021). A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem. *Journal port Science Research*, 4*(2)*, pp. 125–131. https://doi.org/10.36371/port.2021.2.9

[7] Mafarja, M., Qasem, A., Heidari, A. A., Aljarah, I., Faris, H., & Mirjalili, S. (2020). Efficient Hybrid Nature-Inspired Binary Optimizers for Feature Selection. *Cognitive Computation*, *12*(1), 150–175. https://doi.org/10.1007/s12559-019-09668-6

[8] Jensi, R., & Jiji, G. W. (2016). An improved krill herd algorithm with global exploration capability for solving numerical function optimization problems and its application to data clustering. *Applied Soft Computing Journal*, *46*, 230–245. https://doi.org/10.1016/j.asoc.2016.04.026

[9] Qin, Z. (2017). Random fuzzy mean-absolute deviation models for portfolio optimization problem with hybrid uncertainty. *Applied Soft Computing Journal*, *56*, 597–603. https://doi.org/10.1016/j.asoc.2016.06.017

[10] Hankins, G. D. V., & Miller, D. A. (2011, March). A review of the 2008 NICHD research planning workshop: Recommendations for fetal heart rate terminology and interpretation. *Clinical Obstetrics and Gynecology*. https://doi.org/10.1097/GRF.0b013e31820a015b

[11] Al-Tashi, Q., Abdul Kadir, S. J., Rais, H. M., Mirjalili, S., & Alhussian, H. (2019). Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection. *IEEE Access*, *7*, 39496–39508. https://doi.org/10.1109/ACCESS.2019.2906757

[12] Alelyani, S., Tang, J., & Liu, H. (2019). Feature Selection for Clustering: A Review. In *Data Clustering* (pp. 29–60). Chapman and Hall/CRC. https://doi.org/10.1201/9781315373515-2

[13] Bharti, K. K., & Singh, P. K. (2015). Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. *Expert Systems with Applications*, *42*(6), 3105–3114. https://doi.org/10.1016/j.eswa.2014.11.038

[14] Aljarah, I., Mafarja, M., Heidari, A. A., Faris, H., & Mirjalili, S. (2020). Clustering analysis using a novel locality-informed grey wolf-inspired clustering approach. *Knowledge and Information Systems*, *62*(2), 507–539. https://doi.org/10.1007/s10115-019-01358-x

[15] Aljarah, I., Mafarja, M., Heidari, A. A., Faris, H., Zhang, Y., & Mirjalili, S. (2018). Asynchronous accelerating multi-leader salp chains for feature selection. *Applied Soft Computing Journal*, *71*, 964–979. https://doi.org/10.1016/j.asoc.2018.07.040

[16] Chantar, H., Mafarja, M., Alsawalqah, H., Heidari, A. A., Aljarah, I., & Faris, H. (2020). Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification. *Neural Computing and Applications*, *32*(16), 12201–12220. https://doi.org/10.1007/s00521-019-04368-6

[17] Nayak, J., Chandrasekhar, G. T., Naik, B., Pelusi, D., & Abraham, A. (2020, June 1). Special issue on "Soft computing techniques: applications and challenges" neural computing and applications. *Neural Computing and Applications*. Springer. https://doi.org/10.1007/s00521-020-04902-x

[18] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, *172*, 371–381. https://doi.org/10.1016/j.neucom.2015.06.083

[19] Mahmoud, H. Y., Hasanien, H. M., Besheer, A. H., & Abdelaziz, A. Y. (2020). Hybrid cuckoo search algorithm and grey wolf optimiser-based optimal control strategy for performance enhancement of HVDC-based offshore wind farms. *IET Generation, Transmission and Distribution*, *14*(10), 1902–1911. https://doi.org/10.1049/iet-gtd.2019.0801

[20] Nikoobakht, A., Aghaei, J., & Mardaneh, M. (2018, September 30). Retraction: Optimal transmission switching in the stochastic linearised SCUC for uncertainty management of the wind power generation and equipment failures [Gener. Transm. Distrib., 12, 1, (2018) (3780-3792)] doi: 10.1049/iet-gtd.2017.0617. *IET Generation, Transmission and Distribution*. Institution of Engineering and Technology. https://doi.org/10.1049/iet-gtd.2018.0329

[21] El-Fergany, A. A., & Hasanien, H. M. (2015). Single and Multi-objective Optimal Power Flow Using Grey Wolf Optimizer and Differential Evolution Algorithms. *Electric Power Components and Systems*, *43*(13), 1548–1559. https://doi.org/10.1080/15325008.2015.1041625