Improved Multi-Verse Optimizer Feature Selection Technique with Application to Phishing, Spam, and Denial of Service Attacks

Malek Alzaqebah[1], Sana Jawarneh[2], Rami Mustafa A Mohammad[3], Mutasem K. Alsmadi[4] and Ibrahim ALmarashdeh[5]

[1] Department of Mathematics, College of Science, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, Saudi Arabia.
[1]Basic and Applied Scientific Research Center, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, Saudi Arabia.
[2]Computer Science Department, Community College Dammam, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, Saudi Arabia.
[3]Computer Information Systems Department, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, Saudi Arabia. rmmohammad@iau.edu.sa
[4, 5]Department of MIS, College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, Saudi Arabia.

Abstract: Intelligent classification systems proved their merits in different fields including cybersecurity. However, most cybercrime issues are characterized of being dynamic and not static classification problems where the set of discriminative features keep changing with time. This indeed requires revising the cybercrime classification system and pick a group of features that preserve or enhance its performance. Not only this but also the system compactness is regarded as an important factor to judge on the capability of any classification system where cybercrime classification systems are not an exception. The current research proposes an improved feature selection algorithm that is inspired from the well-known multi-verse optimizer (MVO) algorithm. Such an algorithm is then applied to 3 different cybercrime classification problems namely phishing websites, spam, and denial of service attacks. MVO is a population-based approach which stimulates a well-known theory in physics namely multi-verse theory. MVO uses the black and white holes principles for exploration, and wormholes principle for exploitation. A roulette selection schema is used for scientifically modeling the principles of white hole and black hole in exploration phase, which bias to the good solutions, in this case the solutions will be moved toward the best solution and probably to lose the diversity, other solutions may contain important information but didn't get chance to be improved. Thus, this research will improve the exploration of the MVO by introducing the adaptive neighborhood search operations in updating the MVO solutions. The classification phase has been done using a classifier to evaluate the results and to validate the selected features. Empirical outcomes confirmed that the improved MVO (IMVO) algorithm is capable to enhance the search capability of MVO, and outperform other algorithm involved in comparison.

Keywords: Feature Selection, Classification, Multi-Verse Optimizer (MVO) Algorithm, phishing, Spam, and Denial of Service Attacks.

1. Introduction

Machine Learning based approaches have increasing growth recently in many research areas [1-3], the dataset's features are fed into a machine learning to generate predictive model to predict upcoming behaviors. Feature selection is a fundamental preprocessing phase before building intelligent classifiers, it reduces the number of features in objective to eliminate the features that misled the classification process which affects the performance of the classifier, hence selecting the most revealing features with a minimum features number simultaneously with highest possible accuracy are the objectives of feature selection. Several feature selection methods are existing in literature [4], one of these methods is wrapper methods. wrapper methods iteratively pick subset of features to come up with the subset which produces the highest prediction accuracy [5]. Optimization algorithms were used to pick a subset of features in objective to maximizes the classification accuracy. Some wrapper methods that based on nature-inspired algorithms were proposed recently [6]. Nature-inspired algorithms simulate the process observed in the natural phenomena. optimizing the solution of a problem by recruiting a collection of agents, simulating natural phenomena. Nature-inspired approaches were extensively used for optimization problems [7] including ant colonies, bird flocks, and fish schools etc. Several researchers normally resolve various optimization problems by making use of one of these techniques including Genetic Algorithms [8], Differential Evolution [9, 10], Particle Swarm Optimization [8, 11], Firefly Algorithms [12], Cuckoo Search [13], Multi-Objective Optimization [14], and others [15, 16]. The multi-verse optimizer (MVO) is a new developmental technique influenced by the notions of the white/black holes which is a well-known multi-verse theory,. Ewees et al, [17] presents a unique chaotic MVO algorithm (CMVO) to prevent the drawbacks of the MVO, where the chaotic maps are employed to enhance the effectiveness of MVO. The CMVO is utilized to find a solution for the feature selection problem, where they apply on five benchmark datasets to estimate the accomplishment of CMVO algorithm. Faris et al, [18] propose MVO for choosing the optimal features and improving the parameters values. Where the MVO is used to control the key parameters of SVM and discover the best possible group of features for that [18]classification system. Hans and Kaur [19] present the Multi-Verse Optimization that mimics the concept of Multi-Verse in Physics and is similar to the collaboration between the several universes. They proposed the binary versions of MVO with two main objectives: firstly, to reduce irrelevant and unsuitable attributes from the dataset and after that to obtain better classification accuracies. The offered binary variants apply the idea of transformation functions for the mapping of a continuous variant of the MVO algorithm to its binary variations [19]. Aljarah et al, [20] discuss the theoretical basis, processes, and core strengths behind the Multi-Verse Optimization algorithm. Furthermore, a comprehensive literature review is carried out for discussing different versions of the MVO algorithm [20]. Machine learning (ML) approaches are able to predict the threads, security attacks automatically. In addition, ML showed to be very effective in several cybersecurity problems including Phishing website, spam emails and denial of service attacks (DoS) [21]. Spam emails can be defined as sending annoying, harmful, fraudulent, and misleading emails which are typically sent randomly from an individual or from a campaign that has no direct relationship with the recipient. The Denial of Service (DoS) attack overwhelms network system and cloud and online solutions by means of a distributed group of malicious computer systems to do damaging activities [22]. Such damaging activities are typically intended to harm the Information Technology (IT) infrastructure for an organization. On top of that, it can also be intended to harm the IT infrastructure of various governmental and public solutions that have been mainly designed to enhance the individual's way of life [21, 23]. Typically, DoS attempts might impact the availability and the accessibility of the targeted internet-based solution due to processing capability overload. On the other hand, Phishing attacks have received additional focus amongst other cybercrimes found in on the internet. The deceptive emails that request individuals to go to a fraudulent webpage which looks similar to the genuine webpage is considered the first step for beginning the phishing attempts [24]. This in fact a form of social engineering attempts where the users are targeted for obtaining their private information including usernames, passwords, and bank account credentials for committing further financial crimes. All of these attacks are regarded dynamic non-static classification problems due to the continuous change in the set of discriminative features that can be used for detecting such attacks [25-27]. Thus, this article proposes an advanced version MVO algorithm for feature selection to facilitate creating more robust classifiers.

2. Data Description

Three datasets were used in the experiments Phishing website dataset is obtained from the UCI [28]. Such dataset includes 4898 phishing instances and 6157 legitimate ones. The dataset includes 30 input features and one class variable. More information about this dataset can be found in [29]. The spam email dataset has also obtained from UCI repository and it comprise 57 features and one class variable. The dataset includes a total of 4601 instances where 2788 of them are legitimate emails and the remaining 1813 instances were spam emails. Yet, for the DoS attack, the well-known training dataset namely UNSW-NB15 is used in our experiments [21]. Such a dataset includes 29,175 instances in which 15,601 of them were normal traffic and the other 13,574 belong to DoS attack. This dataset includes 41 features. The table 1 depicts the description of these datasets [30].

Table 1: summary of the dataset's description.

Dataset Name	Instances	Attributes	Authors
Phishing Websites	11055	30	*
Spam	4601	57	*
KDD Denial of Service attack	29,175	41	*

3. The Proposed algorithm

The physical theory named multi-verse theory inspires researchers to develop the MVO algorithm for global Optimization [31]. multi-verse theory considered that more than one big bang is exist and each generate new universe. MVO is a population-based algorithm that uses the concepts of multi-verse theory's i.e., black hole, hole wormhole and white hole. Based on big bang theory, our universe generated by the massive explosion. The scientists believed that more than one big bang exists, and each generates a universe. In the multi-verse theory, the multiple universes interact with each other.

3.1. Original MVO

Multi-Verse Optimizer inspiration is based on three concepts of multi-verse, i.e., black hole and a wormhole, white hole. Scientists believed that the big bang is the white hole which is the main element of generation new universe [32-35]. Black holes have very high gravitational force behave, which works inversely to white wholes. While Wormholes links the parts of a universe. In the multi-verse theory, the wormhole works as a tunnel allowing objects to travel between the parts of the universe and between universes. The expansion of the universe is caused by the *inflation rate*, and the *Inflation speed* forms the suitability of life, as well as its forms the stars, black holes, white holes, wormholes, plants etc. In the multi-verse theory, that multiple universes interact and cooperate through the white holes, the black holes, and the wormholes toward reach the stability [36, 37] . The Interaction of the multiple universes through the white holes, the black holes, and the wormholes, is simulated to develop the MVO. MVO uses the black and white holes principles for exploration, and wormholes principle for exploitation [31]. In optimization problem each solution represents one universe, the fitness function value represents the inflation rate and the problem variables are the objects of the universe. Consequently, throughout the optimization process, the MVO follow following steps [31].

1. The high inflation rate has higher probability to have white whole, and lower probability to have black hole.
2. Universes that have high inflation rate are transferring objects over white holes.
3. Universes that have low inflation rate are getting objects over black holes.
4. The universes' objects are randomly moving to the best universe through wormholes.

Roulette wheel selection is performed in the original MVO algorithm to construct the mathematical model of the white holes, black holes and to modify the universes' objects. At each iteration of the MVO process, the universes(solutions) are sorted based on the associated inflation rates (fitness function value), and one is selected by the roulette wheel selection method to have a white hole. The following n×d matrix (*U*) represents the population with n universes (solutions), and d is the number of objects in the universes (problem variables).

$$U = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^d \end{bmatrix}$$

The values of x_i^j are assigned initially based on equation 1.

$$x_i^j = \begin{cases} x_k^j & r1 < Normalizedf\,(Ui) \\ x_i^j & r1 \geq Normalizedf\,(Ui) \end{cases} \quad (1)$$

where r1 is a number generated randomly in [0, 1], Normalizedf (Ui) is the fitness value (normalized inflation rate) of the universe i, and x_k^j denotes the parameter j of the universe k, which selected based on roulette wheel selection method. The pseudocode of this process is represented in Objects_Exchange phase figure 1 lines 14 to 19. The solutions in population U are updated based on equation 2 [31]:

$$x_i^j = \begin{cases} \begin{cases} x_j + TDR * r_4 & r_3 < 0.5 \\ x_j - TDR * r_4 & r_3 \geq 0.5 \end{cases} & r_3 < WEP \\ x_{RW}^j & r_2 \geq WEP \end{cases} \quad (2)$$

Where x_j is the element at position j of the best solution, three random numbers r2, r3 and r4 are generated between 0 and 1, x_{RW}^j indicates the element at position j of the selected solution by roulette wheel selection method. And the TDR and WEP are calculated based on the adaptive equations 3 and 4.

$$WEP = \min + l\left(\frac{\max - min}{MaxIter}\right) \quad (3)$$

$$\mathrm{TDR} = 1 - \frac{l^{1/p}}{\mathrm{MaxIter}^{1/p}} \quad (4)$$

Where l is the current iteration, and MaxIter is the maximum number of iterations, p is used to determine the depth of local search (exploitation); when p is high then more local search will be performed. The pseudocodes of this process represented in Update phase figure 1 lines 14 to 19. The pseudocode of MVO algorithm is represented in figure 1.

```
1  Initialization
2  PS: Population size,
3  Dim: problem dimensions
4  MaxIter=Maximum number of iterations.
5  U=generate a random population (Population size).
6  while iter <= MaxIte:
7       compute the fitness value for all solutions in the population.
8       compute Sorted_U =Sort U.
9       compute WEP based on equation # 3
10      compute TDR based on equation # 4
11      for solution s in U Do
12        BlackHoleIndex=i.
13        for each object indexed by j do
14            //Objects_Exchange phase
15            Initialize r1,r2,r3 and r4 by random numbers between 0 and 1.
16            if r1 < Normalizedf (Ui) then
17             WhiteHoleIndex = RouletteWheelSelection(-Ui);
18             U(BlackHoleIndex,j)= Sorted_U(WhiteHoleIndex,j);
19            end if
20           //Update_phase
21            if r2 < WEP then
22               r3 = rand([0, 1]);
23               r4 = rand([0, 1]);
24               if r3 < 0.5 then
25                   U(i,j)=Best universe(j) + TDR * r4;
26                else
27                    U(i,j)=Best_universe(j) - TDR * r4;
28               end if
29            end if
30        end for
31      end for
32 end while
```

Figure 1: pseudocode for the MVO [31].

3.2. MVO for Feature Selection

The solution in feature selection problem is represented by an array of (zeros and ones) of size n, where n is the number of all features in the dataset. the values of 0 mean the features at these locations are unselected features and the values of 1 mean the features at these locations are selected features. therefore, a version of the MVO algorithm must be adapted to be suited to the feature selection problem. The objectives of features selection might be treated as one of multi-objective optimization problems with that has 2 objectives; to minimize the amount of selected attributes and to maximizing the prediction accuracy. To deal with the multi-objective a fitness function is used with classifier k-nearest neighbors' algorithm [38] (KNN) to find the number of selected features and the error rate. Equation 5 shows the fitness function that applied in this paper [37].

$$Fitness = \alpha\gamma_r(D) + \beta\left|\frac{R}{N}\right|, \quad (5)$$

Where the error rate denoted as $\gamma_r(D)$, |R| represents size of the selected subset, and |N| denots the number of all features, additionally, α ∈ [0, 1] and β= (1 −α) represents the significance of error rate and the subset size [39]. Due to the nature of the solution in feature selection problem, in the Update_phase phase figure 1 have been modified to be suited to the feature selection problem. As it can be observed from the Update_phase figure 1, at each iteration all solutions in the population are updated with respect to the best solution to ensure the exploitation capability of the algorithm, but after a number of iterations the exploration of the search space will be lost and needs to be recovered, thus the lines at 24 and 28 are replaced by the pseudocode shown in figure 2. Two neighborhood operators (NBs) are proposed as follow:

NB1: Choose one random element from the current solution and replace it by a random binary value, which can be 0 or 1.

NB2: Choose one random element from the current solution and XOR the element with 1.

The function (NBS (rand) from figure 2) is used to select at random NB to be applied, to ensure the diversity of solutions. The value of TDR is decreased by increasing the number of iterations, so a random number (i.e.r4) is compared with TDR, which at the beginning of the iterations more probability is given to the exploration and by increasing the iterations number the probability will be increased.

Exploitation (local search)

```
Objects_Exchange phase
if (r4 > TDR)

   if r3 < 0.5 then
     U(i,j)=Best universe(j) + TDR * r4 ;
   else
     U(i,j)=Best_universe(j) - TDR * r4 ;

   end if
else
   U(i)=NBS(rand);
end if
```

Recover Exploration

Figure 2: Pseudocode of the updating procedure.

4. Experimental Results

4.1. Parameter settings

For the classifiers Parameters, the only classifier needs parameter is the KNN classifier, where the K is set to be five [39]. K_folds_cross_validation is used for generating the accuracy, where the K–1 employed for training and validation and the remaining for testing [40, 41]. The parameters were as follows: the maximum iterations= 100 and the number of solutions in the population is 10. The number of runs is 5 for each algorithm in this paper. We used an Intel Core i5 PC- 2.3 GHz CPU, 8 GB RAM to perform our experiments.

4.2. Results Comparisons and Discussions

The objectives in feature selection problem are to minimize the number of selected features and to maximizing the prediction accuracy. In addition, the objective function (fitness) is to be minimize (equation 5). The results for MVO and the improved version of MVO are listed in table 2.

Table 2: Results comparison between MVO and IMVO for all datasets.

Dataset		Phishing Website Dataset		Spam		DoS	
Method		MVO	IMVO	MVO	IMVO	MVO	IMVO
Accuracy	avg	92.10	**94.67**	96.85	**97.13**	99.42	**99.45**
	best	93.22	**94.84**	97.50	**98.15**	99.47	**99.59**
	std	0.01	**0.00**	0.005	0.007	0.001	0.001
selected features	avg	**17.40**	22.40	**28.60**	29.60	20.33	**20.00**
	best	**12**	20	**23**	**23**	**19**	**19**
	std	0.351	**0.182**	**0.328**	0.513	0.115	0.173
Objective Function	avg	0.075	**0.060**	0.03	**0.02**	0.0093	**0.0081**
	best	0.070	**0.057**	0.03	**0.01**	0.0085	**0.0076**
	std	0.003	0.002	0.004	0.002	0.0008	0.0005

Table 2 shows the comparisons of the best, average and standard deviation of classification accuracies, number of selected features and objective function performed by MVO and other approaches (PSO, MFO and WOA). As it can be seen from the table the Improved version of MVO results outperform the MVO in term of accuracy and objective function. As the IMVO is superior from the comparison in table 2, so IMVO is selected to be compared with other approaches using same datasets and same sittings, tables 3, 4 and 5. The algorithms used in the comparisons are selected because they perform well recently using different datasets; which are the particle swarm optimization (PSO) [8, 42], the Moth Optimization algorithm (MFO) [43, 44], and the whale optimization algorithm (WOA) [45, 46, 47, 48].

Table 3: Performance of IMVO PSO, MFO and WOA in Phishing Website Dataset.

Method		IMVO	PSO	MFO	WOA
Accuracy	avg	**94.67**	92.67	94.43	94.35
	best	**94.84**	93.85	94.66	94.57
	std	**0.00**	0.01	0.00	0.00
Selected Features	avg	**22.40**	22.80	24.20	25.20
	best	**20**	20	22	23
	std	**0.182**	0.295	0.164	0.164
Objective Function	avg	0.060	0.059	**0.058**	0.059
	best	**0.057**	0.058	**0.057**	0.058
	std	0.002	**0.001**	**0.001**	**0.001**

The results shown in table 3 confirm that the IMVO outperforms other algorithm in term of accuracy, the selected features and the best objective function for the phishing website dataset. In addition to demonstrate the results based on the box and whisker plot, figure 3 show that small difference between the lower and upper quartile, and also the minimum and maximum values, the median is represented by the horizontal line inside the box.

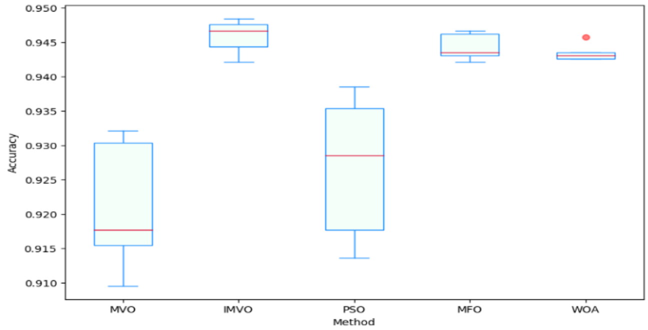

Figure 3: Box and Whisker plot for Phishing Website Dataset.

Table 4: Performance of MVO PSO, MFO and WOA on Spam Dataset.

Method		IMVO	PSO	MFO	WOA
Accuracy	avg	97.13	96.33	97.02	**97.20**
	best	**98.15**	96.74	97.61	97.39
	std	0.007	0.004	0.004	**0.002**
selected features	avg	29.60	39.20	35.20	**27.00**
	best	**23**	36	32	**23**
	std	0.541	0.216	**0.204**	0.255
Objective Function	avg	**0.02**	**0.02**	**0.02**	**0.02**
	best	**0.01**	0.02	0.02	0.02
	std	0.002	**0.001**	0.002	**0.001**

When the comparison between the algorithms has been done based on the Spam Dataset as in table 4, the IMVO get best results in terms of accuracy, selected feature, and objective function, comparing with the WOA which achieve good average results in term of accuracy and selected feature but in term of the objective function all the algorithm get same average results.

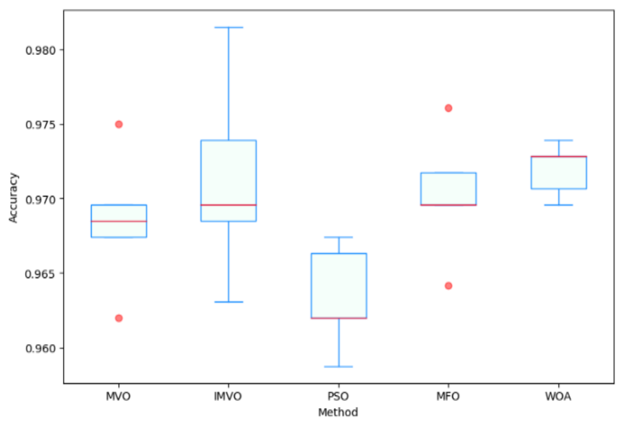

Figure 4: box and Whisker plot for Spam Dataset.

When the comparison done by the box and whisker plot in term of accuracy as figure 4, the IMVO show large difference between the maximum and minimum values, upper and lower quartile comparing with other algorithm, nonetheless it obtains the best value where it was 98.15.

Table 5: Performance of MVO PSO, MFO and WOA on DOC Dataset.

Method		IMVO	PSO	MFO	WOA
Accuracy	avg	**99.45**	99.37	**99.45**	99.31
	best	**99.59**	99.47	**99.59**	99.45
	std	0.001	0.001	0.002	0.001
selected features	avg	**20.00**	26.67	25.00	21.67
	best	**19**	26	24	**19**
	std	0.173	0.58	0.173	0.379
Objective Function	avg	**0.0081**	0.0084	0.0089	0.0083
	best	**0.0076**	0.0079	0.0087	0.0081
	std	0.0005	0.0005	0.0002	0.0003

Table 5 present the results based on the DOC dataset, where the IMVO achieve the best results and good average in the accuracy, selected features ad objective function terms. The MFO algorithm also get same results comparing with IMVO in term of accuracy.

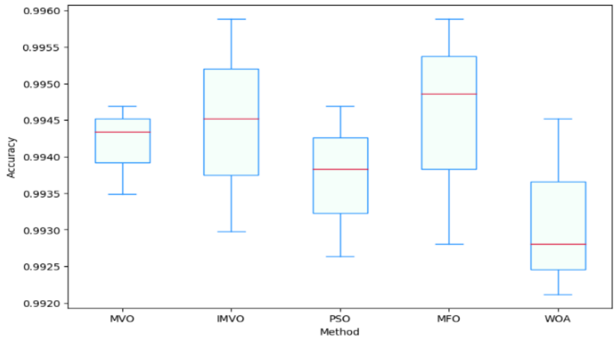

Figure 5: box and Whisker plot for DOC Dataset

Figure 5 illustrate the box and Whisker plot for DOC Dataset in term of the classified accuracy where the IMVO has medium difference between the maximum and minimum values, upper and lower quartile and the median in the middle of whisker box, but still it achieve the best maximum value.
To conclude the results obtained from all the presented algorithms in terms of classification accuracy, figures 6, 7 and 8 are plotted for the Phishing Website Dataset, Spam Dataset and DOC Dataset, respectively. It is obvious that the performance of the IMVO is comparable comparing with PSO, MVO, MFO and WOA, in terms of accuracy and object function, where Figure 6.A and Figure 6.B show the accuracy and the objective function of the algorithms and its clear the MFO in the second place followed by WOA and PSO in term of accuracy, respectively, and with respect to the object function, all algorithms almost same after the IMVO where it obtain the best result with large difference in spam dataset.

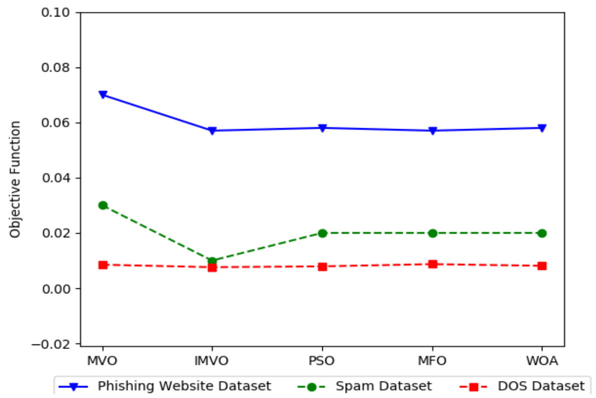

Figure 6.A: plotting the classification accuracy and the objective function for all Dataset.

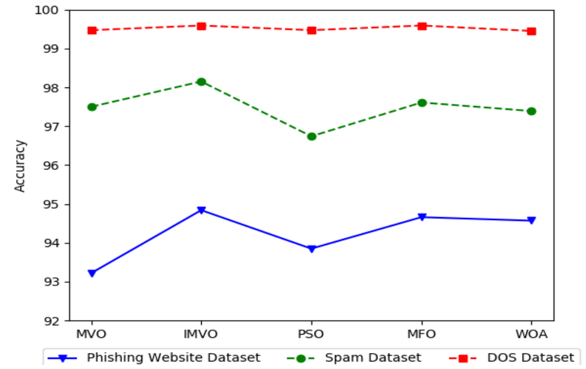

Figure 6.B: plotting the classification accuracy and the objective function for all Dataset.

5. Conclusion and Future Works

Classification algorithms are able to generate models from the data, which showed to be very valuable in many fields. Feature selection process aims to reduces the number of features in the data, to eliminate the features that affects the performance of the classification method. So, we proposed an improved version of multi-verse optimizer (MVO) algorithm for feature selection problem, in MVO all solutions in the population are updated with respect to the best solution to ensure the exploitation capability of the algorithm, but after a number of iterations the exploration needs to be maintain. In the IMVO one random neighborhood operator from the two proposed neighborhood operators is applied to maintain the diversity of solutions. More probability for the neighborhood operator is given to when the number of iterations increased. Experimental results show that IMVO is able to maintain the diversity of MVO, which leads to improve the search capability in MVO algorithm. Based on the results of IMVO it can be outperformed the MVO and other competitive algorithms. Appling the proposed algorithm on more complex datasets is the subject to our future work.

References

[1] Khanday A M U D, Rabani S T, Khan Q R, Rouf N and Din M M U. Machine learning based approaches for detecting COVID-19 using clinical text data. International Journal of Information Technology, 2020, 12(3): 731-739.

[2] Birkle P, Zouch M, Alzaqebah M and Alwohaibi M. Machine Learning-based Approach for Automated Identification of Produced Water Types from Conventional and Unconventional Reservoirs. In Petroleum Geostatistics 2019, pp. 1-5.

[3] Al-Batah M, Mrayyen S and Alzaqebah M. Arabic Sentiment Classification using MLP Network Hybrid with Naive Bayes Algorithm. Journal of Computer Science, 2018, 14(8): 1104-1114.

[4] Chandrashekar G and Sahin F. A survey on feature selection methods. Computers & Electrical Engineering, 2014, 40(1): 16-28.

[5] Kohavi R and John G H. Wrappers for feature subset selection. Artificial intelligence, 1997, 97(1-2): 273-324.

[6] Nguyen B H, Xue B and Zhang M. A survey on swarm intelligence approaches to feature selection in data mining. Swarm and Evolutionary Computation, 2020, 54: 100663.

[7] Fister Jr I, Yang X-S, Fister I, Brest J and Fister D. A brief review of nature-inspired algorithms for optimization. arXiv preprint arXiv:1307.4186, 2013.

[8] Shreem S S, Abdullah S, Nazri M Z A and Alzaqebah M. Hybridizing ReliefF, MRMR filters and GA wrapper approaches

for gene selection. J. Theor. Appl. Inf. Technol, 2012, 46(2): 1034-1039.

[9] Zhang Y, Gong D-w, Gao X-z, Tian T and Sun X-y. Binary differential evolution with self-learning for multi-objective feature selection. Information Sciences, 2020, 507: 67-85.

[10] Das S, Abraham A and Konar A. Automatic clustering using an improved differential evolution algorithm. IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 2007, 38(1): 218-237.

[11] Aljarah I and Ludwig S A. Mapreduce intrusion detection system based on a particle swarm optimization clustering algorithm. In 2013 IEEE congress on evolutionary computation, pp. 955-962.

[12] Marie-Sainte S L and Alalyani N. Firefly algorithm based feature selection for Arabic text classification. Journal of King Saud University-Computer and Information Sciences, 2020, 32(3): 320-328.

[13] Pandey A C, Rajpoot D S and Saraswat M. Feature selection method based on hybrid data transformation and binary binomial cuckoo search. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(2): 719-738.

[14] Karasu S, Altan A, Bekiros S and Ahmad W. A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. Energy, 2020, 212: 118750.

[15] Alzaqebah M, Jawarneh S, Alwohaibi M, Alsmadi M K, Almarashdeh I and Mohammad R M A. Hybrid Brain Storm Optimization algorithm and Late Acceptance Hill Climbing to solve the Flexible Job-Shop Scheduling Problem. Journal of King Saud University-Computer and Information Sciences, 2020.

[16] Alzaqebah M, Jawarneh S, Sarim H M and Abdullah S. Bees Algorithm for Vehicle Routing Problems with Time Windows. International Journal of Machine Learning and Computing, 2018, 8(3): 234-240.

[17] Ewees A A, Abd El Aziz M and Hassanien A E. Chaotic multi-verse optimizer-based feature selection. Neural Computing and Applications, 2019, 31(4): 991-1006.

[18] Faris H, Hassonah M A, Ala'M A-Z, Mirjalili S and Aljarah I. A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture. Neural Computing and Applications, 2018, 30(8): 2355-2369.

[19] Hans R and Kaur H. Binary Multi-Verse Optimization (BMVO) Approaches for Feature Selection. International Journal of Interactive Multimedia & Artificial Intelligence, 2020, 6(1).

[20] Aljarah I, Mafarja M, Heidari A A, Faris H and Mirjalili S. Multi-verse optimizer: theory, literature review, and application in data clustering. Nature-inspired optimizers, 2020: 123-141.

[21] Mohammad R M A, Alsmadi M K, Almarashdeh I and Alzaqebah M. An improved rule induction based denial of service attacks classification model. Computers & Security, 2020, 99: 102008.

[22] Bonguet A and Bellaiche M. A survey of denial-of-service and distributed denial of service attacks and defenses in cloud computing. Future Internet, 2017, 9(3): 43.

[23] Mohammad R M A and Abdulqader M M. Exploring Cyber Security Measures in Smart Cities. In 2020 21st International Arab Conference on Information Technology (ACIT), 28-30 Nov. 2020, pp. 1-7.

[24] Mohammad R M A. An Improved Multi-Class Classification Algorithm based on Association Classification Approach and its Application to Spam Emails. IAENG International Journal of Computer Science, 2020, 47(2).

[25] Mohammad R M A and Alqahtani M. A comparison of machine learning techniques for file system forensics analysis. Journal of Information Security and Applications, 2019, 46: 53-61.

[26] Mohammad R M A. A lifelong spam emails classification model. Applied Computing and Informatics, 2020.

[27] Mohammad R M A and Alsmadi M K. Intrusion detection using Highest Wins feature selection algorithm. Neural Computing and Applications, 2021.

[28] Mohammad R, Thabtah F A and McCluskey T. Phishing websites dataset. 2015.

[29] Mohammad R M, Thabtah F and McCluskey L. Predicting phishing websites based on self-structuring neural network. Neural Computing and Applications, 2014, 25(2): 443-458.

[30] Dua D and Graff C. UCI machine learning repository, 2017. URL http://archive. ics. uci. edu/ml, 2019, 7(1).

[31] Mirjalili S, Mirjalili S M and Hatamlou A. Multi-verse optimizer: a nature-inspired algorithm for global optimization. Neural Computing and Applications, 2016, 27(2): 495-513.

[32] Singh M P. A new approach for data clustering using Multi-verse optimizer algorithm, 2016.

[33] Abasi A K, Khader A T, Al-Betar M A, Naim S, Makhadmeh S N and Alyasseri Z A A. Link-based multi-verse optimizer for text documents clustering. Applied Soft Computing, 2020, 87: 106002.

[34] Abasi A K, Khader A T, Al-Betar M A, Naim S, Alyasseri Z A A and Makhadmeh S N. A novel hybrid multi-verse optimizer with K-means for text documents clustering. Neural Computing and Applications, 2020, 32: 17703-17729.

[35] Eardley D M. Death of white holes in the early universe. Physical Review Letters, 1974, 33(7): 442.

[36] Khoury J, Ovrut B A, Seiberg N, Steinhardt P J and Turok N. From big crunch to big bang. Physical Review D, 2002, 65(8): 086007.

[37] Barrow J D, Davies P C and Harper Jr C L. Science and ultimate reality: Quantum theory, cosmology, and complexity. 2004.

[38] Altman N S. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 1992, 46(3): 175-185.

[39] Emary E, Zawbaa H M and Hassanien A E. Binary ant lion approaches for feature selection. Neurocomputing, 2016, 213: 54-65.

[40] Rami Malkawi M A, Ali Al-Yousef and Abul-Huda B. The impact of the digital storytelling rubrics on the social media engagements. Int. J. Computer Applications in Technology, 2019, 59(3): 269-275.

[41] Mrayyen S, Al-Batah M S and Alzaqebah M. Investigation of Naive Bayes Combined with Multilayered Perceptron for Arabic Sentiment Analysis and Opinion Mining. International Journal of Mathematical Models And Methods in Applied Sciences, 2018, 12.

[42] Rostami M, Forouzandeh S, Berahmand K and Soltani M. Integration of multi-objective PSO based feature selection and node centrality for medical datasets. Genomics, 2020, 112(6): 4370-4384.

[43] Alzaqebah M, Alrefai N, Ahmed E A, Jawarneh S and Alsmadi M K. Neighborhood search methods with Moth Optimization algorithm as a wrapper method for feature selection problems. International Journal of Electrical and Computer Engineering, 2020, 10(4): 3672.

[44] Khurma R A, Aljarah I and Sharieh A. An Efficient Moth Flame Optimization Algorithm using Chaotic Maps for Feature Selection in the Medical Applications. In ICPRAM, pp. 175-182.

[45] Agrawal R, Kaur B and Sharma S. Quantum based whale optimization algorithm for wrapper feature selection. Applied Soft Computing, 2020, 89: 106092.

[46] Mafarja M M and Mirjalili S. Hybrid whale optimization algorithm with simulated annealing for feature selection. Neurocomputing, 2017, 260: 302-312.

[47] M. A. Khan, S. Khan, B. Shams, J. Lloret, distributed flood attack detection mechanism using ANN in wireless mesh networks, Security and Communication Networks, Vol. 9, No. 15, pp. 2715–2729, October 2016.

[48] N. Alrajeh, S. Khan, B. Shams, "Intrusion detection systems in wireless sensor networks: A Review", International Journal of Distributed Sensor Networks, pp. 1-7, Volume 2013.