IMPROVING WHALE OPTIMIZATION ALGORITHM FOR FEATURE SELECTION WITH A TIME-VARYING TRANSFER FUNCTION

MOHAMMED ABDULRAZAQ KAHYA*
and
SUHAIB ABDULJABBAR ALTAMIR
Department of Computer sciences
Education College for Pure Sciences
University of Mosul, Mosul, Iraq

ZAKARIYA YAHYA ALGAMAL
Department of Statistics and Informatics
College of Computer sciences and Mathematics
University of Mosul, Mosul, Iraq

(Communicated by Gerhard-Wilhelm Weber)

Abstract. Feature selection is a valuable tool in supervised machine learning research fields, such as pattern recognition or classification problems. Feature selection used to eliminate irrelevant and noise features that adversely affect results. Swarm algorithms are usually used in feature selection problem; these algorithms need transfer functions that change search space from continuous to the discrete. However, transfer functions are the backbone of all binary swarm algorithms. Transfer functions in the current formula cannot provide binary swarm algorithms with a fit balance between exploration and exploitation stages. In this work, a feature selection approach based on the binary whale optimization algorithm with different kinds of updating techniques for the time-varying transfer functions is proposed. To evaluate the performance of the proposed method, three of each chemical and biological binary datasets are used. The results proved that BWOA-TV2 has consistency in feature selection and it gives rise to the high accuracy of the classification with more congruent in the convergence. It worth mentioning that the proposed method is proved advance in performance over competitor optimization algorithms, such as particle swarm optimization (PSO) and firefly optimization (FO) that commonly used in this field.

1. Introduction. Classification of tumor cells, diseases or other patterns is becoming very important nowadays. As an essential step for a robust classification, features selection used as a relevant process to the subject under the study, and therefore feature selection has become a critical problem [2]. Features, genes or variables selection approaches based on artificial intelligence techniques have been widely used these days to supervised machine learning in order to select informative genes, related features, or instrumental variables that provide the robust accuracy
for the classifier. The main objective of these techniques is to enhance the classification via select the most informative features, to eliminate irrelevant features [23], to reduce over-fitting problem [6], and to avoid the famous issue “curse of dimensionality” [43]; all of these reasons helped the researchers to interpret the huge data. Practically, meta-heuristic algorithms were used effectively to solve different types of optimization problems [33]–[3]. In general, meta-heuristic algorithms can be divided into four main types: evolution-based, physics-based, swarm-based and human-based approaches [31]. Just recently, many researchers are interested in selecting features based on swarm algorithms for several reasons: ease of application, fast execution, and high accuracy. Chuang, Li-Yeh and et al [8] presented a feature selection method based on binary particle swarm optimization (BPSO) algorithm. The researchers achieved high classification rates with few numbers of features in 9 of the 11 for used data. Emary, E. and et al [9] suggested a binary firefly optimization (BFO) algorithm. The proposed method was used in features selection and proves advance over BPSO in various valuation indicators [9]. Mirjalili and Lewis [31] proposed The Whale Optimization Algorithm (WOA), WOA is a modern meta-heuristic, which is inspired by the behavior of humpback whales. The WOA can be observed as a newly succeeded approach that can outperform other well-regarded approaches. The WOA has been utilized to various real-world problems such as economic dispatch [1], numerical optimization [46], problems of truss and frame structures [18], prediction problems in Bioinformatics [5]. Mafarja, Majdi and et al [25] presented a binary WOA (BWOA) to test the effect of transfer functions on features selection. Researchers have successfully identified optimal transfer functions to select related features that give more accuracy. Moreover, the time-varying mechanisms with swarm algorithms have been proposed to improve algorithms performance as Table 1. The time-varying mechanisms basically have proposed to provide a perfect balance between the exploration and exploitation stages in swarm intelligence algorithms [26]–[27]. In fact, the ratio of exploration and exploitation is a challenge for researchers to avoid local optima problem. On the other hand, the BWOA with the current version cannot provide this balancing. Thus, time-varying transfer functions (TVTFs) with various updating techniques have been proposed for this purpose.

In this paper, the BWOA supported with the K-Nearest Neighborhood (KNN) classifier is proposed to evaluate the effects of updating techniques for the TVTFs on the efficiency of BWOA; via classification accuracy, the number of selected features, and the fitness function value. The proposed approach was benchmarked using several types of chemical and biological datasets. The rest of the paper is organized as follows: section 2 a brief overview of feature selection and WOA. The detailed description of the BWOA with TVTF is explained in section 3. The experimental results and discussion are listed in section 4. While the conclusions are presented in section 5.

2. Overview.

2.1. Feature Selection. One of the most important machine learning tools is the classification which considered as the main tool of supervised machine learning. Classification with high accuracy is an important task for researchers. Feature selection is one of the effective ways to increase classification accuracy. Feature selection is the process of selecting features that are relevant to the response variable of the subject of research under study. In general, feature selection techniques are
Table 1. Some swarm intelligence algorithms with various time-varying mechanisms.

| Algorithm used                  | Problem                      | Reference |
|---------------------------------|------------------------------|-----------|
| Binary dragonfly optimization   | Feature selection            | [26]      |
| BPSO                            | 0-1 knapsack                 | [13]      |
| PSO                             | Numerical optimization       | [44]      |
| BPSO                            | Multidimensional knapsack    | [7]       |
| BPSO                            | Feature selection            | [27]      |

divided into three kinds: filter, wrapper and embedded methods [20], [11]. Recently, intelligent techniques have widely used in features selection; the reason is due to the existence of many simple and easy of intelligent techniques that give affecting solutions [33], [38]–[4].

2.2. Objective Function of Features Selection. Let $D$ be a dataset of $N$ observations with $D$ features. The feature selection process is to select $d$ features from $D$ features, where $d \leq D$. The objective function for our model is to find an optimal features subset that gives maximum classification accuracy with a few the number of selected features. So, the objective function $f(x_1, x_2, ..., x_d), x_j = \{0, 1\}$ is maximized subject to the classification problem. Hence, the features selection process becomes a discrete optimization problem where the objective is to find the optimal features subset.

2.3. Whale Optimization Algorithm. In 2016, Mirjalili et al. proposed the whale optimization algorithm (WOA) which inspired from the humpback whales behavior in hunting method [31]. The hunting method is done through creating distinctive bubbles along a circle, loop or spiral (exploitation phase). This unique behavior of whales is called bubble-net feeding method. This strategy gives whales the possibility of detection of the prey location and encircles them (substitute for the best solution obtained until now). This strategy can be represented mathematically as presented in Equations (1) and (2):

$$\vec{D} = |\vec{C}.\vec{X}^*(t) - \vec{X}(t)|, \quad (1)$$

$$\vec{X}(t + 1) = \vec{X}^*(t) - \vec{A}.\vec{D}, \quad (2)$$

where $\vec{X}^*(t), \vec{X}(t)$ represent the best position and position vector solution, respectively and Equation (1) represents the distance between them. The symbol $t$ points to the iteration while $A$ and $C$ are coefficient vectors, they can be calculated according to the Equations (3) and (4):

$$\vec{A} = 2.\vec{a}.\vec{r} - \vec{a}, \quad (3)$$

$$\vec{C} = 2.\vec{r}, \quad (4)$$

where the vector $r$ is a random vector between zero to one while the vector $a$ in the iterations process decreases linearly from two to zero as Equation (5):

$$a(t) = 2(1 - \frac{t}{T}), \quad (5)$$

where $t$ is the current iteration while $T$ is the total number of iterations. The WOA has two phases: Exploitation and exploration. Exploitation phase is done through shrinking encircling mechanism and spiral updating. Shrinking encircling mechanism is done according to Equation (4). Spiral updating position is like mimic
the helix-shaped movement that we use to calculate the distance between the whale and the prey as below:

$$\vec{X}(t+1) = \vec{D} e^{b l} \cos(2\pi l) + \vec{X}^*(t),$$

where $b$ is a constant (set $b = 1$) and $l$ is a random number limited between -1 to 1. The later mathematical model is a simulation of the whale behavior around the prey. This behavior is either the shrinking encircling mechanism or the spiral model. The mathematical model assumes a probability of 50% for this behavior:

$$\vec{X}(t+1) = \begin{cases} \vec{D} e^{b l} \cos(2\pi l) + \vec{X}^*(t) & \text{if } \text{Pro.} > 0.5 \\ \vec{X}^*(t) - \vec{A} \vec{D} & \text{if } \text{Pro.} < 0.5 \end{cases}$$

where Pro. is a random number in a uniform distribution. Finally, the stage of searching for prey (exploration phase). While the exploration phase simulation of the humpback whales behavior is updating the positions of the current whales according to the best location (solution) presently. A whale is selected randomly from the population to allow the global search. Mathematically, the Equations (8) and (9) interpret this process as below:

$$\vec{D} = |\vec{C} \cdot \vec{X}_{\text{rand}} - \vec{X}|,$$

$$\vec{X}(t+1) = \vec{X}_{\text{rand}} - \vec{A} \vec{D},$$

where $\vec{X}_{\text{rand}}$ is a random whale selected from the population and $\vec{A}$ is a random vector with values greater than one or less than minus one.

3. **Binary WOA with time-varying transfer functions.** One of the most important applications of the binary optimization problem is a feature selection problem [25], [24]–[29], where a search vector can be changed to 1’s and 0’s. Discrete optimization algorithms are depending on the transfer function (TF) to change search space from continuous to discrete [19]. Kennedy and Eberhart [19] proposed the first sigmoid TF that has S-shaped as below:

$$\text{Sigmoid}(v^k_i(t+1)) = \frac{1}{1 + e^{-v^k_i(t+1)}},$$

where $v^k_i(t)$ is the $i^{th}$ search step vector of the $k^{th}$ dimension in the $(t+1)^{th}$ iteration. Now every element of the search vector after transformation is changed according to the Equation (11):

$$X^k_i(t+1) = \begin{cases} 1 & \text{if } \text{rand} < \text{Sigmoid}(v^k_i(t+1)) \\ 0 & \text{otherwise} \end{cases}$$

In the binary optimization problems, TFs were used to control for exploration and exploitation [30]. Therefore, it is normal to choose inappropriate TFs that will have a negative effect on the performance of the algorithms. The sigmoid TF was widely used in binary optimization algorithms, although it has some limitations. For example, it cannot provide enough balance between the two basic elements in the swarm algorithms represented by exploration and exploitation stages. Actually, the ratio of exploration ought to be greater than the exploitation ratio especially with the initiation of optimization algorithm in processing. Indeed, controlling the proportion of exploration and exploitation is very important to avoid local optima problem. On the other hand, the sigmoid function with the current shape cannot provide this balancing. Thus, to make sure that the exploration in early stages and
the exploitation in later stages should update the shape of the sigmoid function over the process of the optimization [26], [13].

3.1. Time-varying transfer functions. TVTF is a new version of TF with time-varying proposed by Islam et al. [13] as below:

$$\text{Sigmoid} \left( X_k^i(t+1)/Tv \right) = \frac{1}{1 + e^{-X_k^i(t+1)/Tv}}, \quad (12)$$

where $Tv$ is a control parameter of the time-varying, it is decreasing during iterations. Now every element of the search vector after transformation be changed according to Equation (13):

$$X_k^i(t+1) = \begin{cases} 1 & \text{if } \text{rand} < \text{Sigmoid} \left( X_k^i(t+1)/Tv \right) \\ 0 & \text{otherwise} \end{cases}$$  \quad (13)

TVTF has been proposed for a good balance and useful control of the exploration and exploitation ratios. In this study, three kinds of time-varying ($Tv$) update techniques have been presented and tested to evaluate the effect of this updating on the performance of the algorithm as below:

3.1.1. Linear update technique of the time-varying. Islam et al. [13] proposed this technique to achieve an influential balance between exploration and exploitation rates by controlling the time-varying parameter as below:

$$Tv_1(t) = Tv_{\text{max}} + (Tv_{\text{min}} - Tv_{\text{max}}) \frac{Itr_{t+1}}{Itr_{\text{Max}}},$$  \quad (14)

where $Tv_{\text{min}},Tv_{\text{max}}$ are the minimum and maximum values of the control parameter, $Itr_{t+1}$ is the current iteration while $Itr_{\text{Max}}$ is the maximum number of iterations.

3.1.2. Non-linear update technique of the time-varying. This technique was inspired by the approach used to adjust the inertial weight of particle swarm optimization proposed by Yang et al. [44] as follows:

$$Tv_2(t) = Tv_{\text{max}} + (Tv_{\text{min}} - Tv_{\text{max}}) \left( \frac{Itr_{t+1}}{Itr_{\text{Max}}} \right)^{\alpha},$$  \quad (15)

where alpha was set to be 0.5.

3.1.3. The decreasing shape of the time-varying update technique. This technique was proposed by the authors; it doesn’t need to adjust the parameter values by minimum or maximum as shown in Equation (16):

$$Tv_3(t) = \left( \frac{1}{\pi Itr_{t+1}} \right)^{\frac{1}{2}}.$$  \quad (16)

Inspecting the TVTF (see Figure 1 where $Tv_{\text{min}} = 0.1, Tv_{\text{max}} = 1$ within 1000 iterations with step size 25), it can be clearly seen the effect of the time-varying on the sigmoid function shape during iterations. This change in shape is very important to control the two basic stages in the swarm algorithms. Applying the update techniques for time-varying over TFs is very important to avoid trapped in local optima. On the other hand, this approach is used to explore numerous regions in the search space hoping to discover the global optima.
3.2. Fitness Function. In recent studies of the features selection problems [27]–
[11], the fitness function is calculated based on the classification accuracy with the
number of features selected over the training dataset. The researchers’ suggestions
were taken into account in the proposed fitness function as shown in the Equation
(17):

\[ \text{fitness} = 0.9 \times CA + 0.1 \times \left( \frac{F_{\text{tot}} - F_{\text{sel}}}{F_{\text{tot}}} \right), \]  

where CA is classification accuracy calculated by KNN classifier (K=3) [40] – [41]
and \( F_{\text{tot}} \) is a total number of database features while \( F_{\text{sel}} \) is the number of selected
features.

4. Results and discussion. In this section, two types of binary chemical and bio-
logical datasets were used (three of each type) in order to get an accurate evaluation
of the proposed method. Three various evaluation indicators were used: classification
accuracy (CA), the number of selected features, and the fitness function value.
The data are anti-hepatitis C virus activity of thiourea derivatives [21]–[17], antimicro-
bial agents (pMIC) [42], neuraminidase inhibitors of influenza a virus (H1N1)
[22], prostate cancer [36], leukemia [10] and breast cancer [39] The specifics of the
used data were listed in Table 2.

The used data were arbitrary split into training and test datasets by 70%, 30%
of the samples respectively. To reduce the effects of partitioning and for a fair
evaluation of the proposed approach, results obtained were averaged for experiments
executed over 20 partitioned times (whales population size is 20 and the number of
iterations 100) on training data using Matlab 2017b on a personal computer with Intel Corei7 processor, 2.5 GHz CPU and 16 GB of RAM.

The experimental results of the BWOA with $T_v^1$, $T_v^2$, $T_v^3$ and Sigmoid respectively listed in Table 3 and Table 4. The results include classification accuracy for both of the training and testing data that referred to above with a number of selected features.

Depending on the results in Table 3, the reader can observe the efficiency of using different update techniques for time-varying transfer functions on the algorithm performance in maximizing the accuracy of the classification and minimizing the dimensionality of the features. Specifically, the average of the classification accuracy and the number of selected features of BWOA-TV2 on all the experimental datasets is the best among other approaches, followed by BWOA-TV1. For H1N1 dataset, BWOA-TV2 and BWOA-TV1 achieved 98.34% and 98.25%, respectively, which are more accurate than BWOA-TV3 and BWOA-Sigmoid. Further, BWOA-TV2 and BOA-TV1 achieve perfect results in a reduction of the features number that are significantly different from BWOA-TV3 and BWOA-Sigmoid. The number of features selected by BWOA-TV2 is the smallest among the other approaches, which indicates that BWOA-TV2 is a beneficial to identify discriminatory and relevant features that give perfect classification accuracy. For instance, in anti-hepatitis C virus dataset, BWOA-TV2 selected 8.77 features on average compared to 10.87, 13.33 and 14.66 features out of 2559 for BWOA-TV1, BWOA-TV3, and BWOA-Sigmoid, respectively.

Concerning the testing data, Table 4 shows that the highest accuracies of BWOA-TV2 are 94.87%, 90.95%, 97.26%, 94.85%, and 92.32% on each dataset, respectively. Moreover, we can find out that BWOA-TV2 overcomes BWOA-TV1, BWOA-TV3, and BWOA-Sigmoid in overall performance over most datasets used. that represents the convergence of the algorithm, the accuracy of the classification and the number of selected features.

The perfect performance of the algorithms in this field is decided via method convergence, classification accuracy and the number of selected features. Proceeding from this principle, Figure 2 illustrates the updating techniques effectiveness of the time-varying on the convergence of method on the six used datasets. From Figure 2, the reader can observe significant differences in BWOA convergence depending on the updating techniques, where BWOA-TV2 shows the best convergence rate in most used datasets. On the other hand, one can see the flexibility of BWOA-Sigmoid in the convergence but it lacks the accuracy that achieved via BWOA-TV2. It is certainly noted that all proposed updating techniques for time-varying over sigmoid transfer function improve the accuracy of BWOA in features selection.

For more investigation of the efficacy of the proposed method in features selection, a comparison was executed between the proposed method and other various approaches (BPSO, BFO). Two various evaluation indicators were used for comparison: CA and #features. The proposed method was tested on six binary datasets and achieved superiority over other algorithms as BPSO and BFO that commonly used in this field. The basic settings of the competitor optimization algorithms were listed in Table 5. The performance of the BWOA, BFO, and BPSO over the various datasets were summarized in Table 6 and Table 7. We can remark from the Table 6 and Table 7 that the results of BWOA-TV2 is much better than BPSO and BFO in both of the CA and #features indicators, which proves that the proposed method is the better and the most efficient to identify relevant features.
Table 2. High-dimensional binary datasets.

| Datasets                        | #samples | #features | Class (+/-) | Data type |
|---------------------------------|----------|-----------|-------------|-----------|
| anti-hepatitis C virus          | 121      | 2559      | (31/90)     | chemical  |
| antimicrobial agents            | 212      | 3657      | (108/104)   | chemical  |
| H1N1                            | 479      | 2322      | (266/213)   | chemical  |
| Leukemia                        | 72       | 7129      | (47/25)     | biological|
| Breast Cancer                   | 38       | 7129      | (18/20)     | biological|
| Prostate Cancer                 | 102      | 12600     | (52/50)     | biological|

Table 3. Comparison between the influence of updating techniques over the proposed method in terms of average CA with standard deviation and #features according to the training data.

| Datasets                        | Methods                  | Indicator | BWOA-TV1 | BWOA-TV2 | BWOA-TV3 | BWOA-Sigmoid |
|---------------------------------|--------------------------|-----------|----------|----------|----------|--------------|
| anti-hepatitis C virus          | CA                       | 96.01(0.763) | 96.11(0.639) | 94.03(1.282) | 93.92(1.065) |
|                                 | #features                | 10.87     | **8.77** | 13.33    | 14.66    |              |
| antimicrobial agents            | CA                       | 92.05(1.045) | 93.99(0.987) | 91.57(1.084) | 92.65(0.885) |
|                                 | #features                | 12.60     | **10.53** | 16.20    | 11.76    |              |
| H1N1                            | CA                       | 98.25(0.973) | 98.34(1.002) | 97.46(1.078) | 96.89(1.541) |
|                                 | #features                | 9.83      | **7.20**  | 11.45    | 14.25    |              |
| Leukemia                        | CA                       | 96.56(1.289) | 97.21(0.587) | 96.29(1.972) | 93.29(1.18)  |
|                                 | #genes                   | 10.56     | **9.23**  | 13.37    | 16.34    |              |
| Breast Cancer                   | CA                       | 93.21(1.032) | 94.84(1.021) | 92.81(2.01)  | 92.17(1.49)  |
|                                 | #genes                   | 17.92     | **17.03** | 21.49    | 22.31    |              |
| Prostate Cancer                 | CA                       | 98.22(0.581) | 97.82(0.721) | 96.21(1.143) | 96.03(0.927) |
|                                 | #genes                   | 9.29      | **10.03** | 10.21    | 10.33    |              |

Table 4. Comparison between the TVTFs kinds in terms of average CA with standard deviation according to the testing data.

| Datasets                        | Methods                  | Indicator | BWOA-TV1 | BWOA-TV2 | BWOA-TV3 | BWOA-Sigmoid |
|---------------------------------|--------------------------|-----------|----------|----------|----------|--------------|
| anti-hepatitis C virus          | CA                       | 94.56(0.897) | 94.87(0.654) | 91.75(1.914) | 91.49(0.514) |
|                                 | #features                | 10.56     | **9.23**  | 13.37    | 16.34    |              |
| antimicrobial agents            | CA                       | 90.32(0.986) | 90.95(1.290) | 89.12(1.590) | 89.85(1.824) |
|                                 | #features                | 17.92     | **17.03** | 21.49    | 22.31    |              |
| H1N1                            | CA                       | 96.07(0.904) | 97.26(1.561) | 94.87(1.721) | 94.34(2.005) |
|                                 | #genes                   | 9.29      | **10.03** | 10.21    | 10.33    |              |
| Leukemia                        | CA                       | 90.91(0.730) | 92.32(0.928) | 89.95(0.230) | 89.27(0.296) |
|                                 | #genes                   | 9.29      | **10.03** | 10.21    | 10.33    |              |
| Breast Cancer                   | CA                       | 96.91(0.476) | 96.39(0.713) | 94.31(0.931) | 94.09(1.048) |
|                                 | #genes                   | 9.29      | **10.03** | 10.21    | 10.33    |              |

Table 5. Basic settings of the BPSO and BFO optimizers.

| BPSO                  | BFO                  |
|-----------------------|----------------------|
| $w = 0.5$             | $\alpha = 0.2$       |
| $c_1 = 1$             | $\beta_0 = 2$        |
| $c_2 = 3$             | $\gamma = 1$         |

5. Conclusion. In this paper, three different kinds of updating techniques for the time-varying transfer function (over sigmoid function) with BWOA were investigated to show the efficiency and effectiveness of the updating techniques for transfer
The experimental differentiation strategy was executed on six available high-dimensional binary datasets in terms of classification accuracy and the number of selected features. Moreover, the classifier used to evaluate the results is KNN model with K=3. The results confirmed that BWOA-TV2 has the ability to reduce the number of selected features and it gives rise to the high accuracy of the classification with more favorable in the convergence. On the other hand, BWOA-TV2 had achieved advancement over traditional rival methods (BFO and BPSO) in features selection domain.

Acknowledgments. The authors are very grateful to the University of Mosul/Iraq for their provided facilities, which helped to improve the quality of this work.
Table 6. Comparison between the proposed method and rival methods in terms of average CA with standard deviation and #features according to the training data.

| Datasets           | Indicator | Methods          | BWOA-TV2       | BFO-Sigmoid | BPSO-Sigmoid |
|--------------------|-----------|------------------|----------------|-------------|--------------|
|                    |           |                  | CA             |             |              |
| anti-hepatitis C   |           |                  | 96.11(0.639)   | 92.51(1.431)| 91.04(1.289) |
| virus              | #features |                  | 8.77           | 17.72       | 21.33        |
| antimicrobial agents|          |                  | 93.99(0.987)   | 90.07(1.129)| 89.91(1.787) |
|                    | #features |                  | 10.53          | 20.29       | 22.31        |
| H1N1               |           |                  | 98.34(1.002)   | 93.98(1.236)| 92.71(1.763) |
|                    | #features |                  | 7.20           | 17.33       | 19.83        |
| Leukemia           |           |                  | 97.21(0.587)   | 93.92(1.201)| 93.51(1.02)  |
|                    | #genes    |                  | 9.23           | 17.41       | 17.60        |
| Breast Cancer      |           |                  | 94.84(1.021)   | 91.92(0.921)| 91.27(0.907) |
|                    | #genes    |                  | 17.03          | 23.39       | 23.91        |
| Prostate Cancer    |           |                  | 97.82(0.721)   | 97.28(0.932)| 97.65(0.829) |
|                    | #genes    |                  | 10.03          | 10.92       | 10.41        |

Table 7. Comparison between the proposed method and rival methods in terms of average CA with standard deviation according to the testing data.

| Datasets           | Methods          | BWOA-TV2       | BFO-Sigmoid | BPSO-Sigmoid |
|--------------------|------------------|----------------|-------------|--------------|
|                    |                  | CA             |             |              |
| anti-hepatitis C   |                  | 94.87(0.654)   | 90.43(1.752)| 89.31(2.801) |
| virus              |                  | 90.95(1.290)   | 88.62(2.006)| 87.59(2.582) |
| antimicrobial agents|                | 97.26(1.561)   | 92.28(1.320)| 89.89(1.920) |
| H1N1               |                  | 94.85(1.051)   | 91.53(1.838)| 90.97(1.308) |
| Leukemia           |                  | 92.32(0.928)   | 89.57(1.534)| 89.49(1.395) |
| Breast Cancer      |                  | 96.39(0.713)   | 94.89(0.872)| 95.06(0.396) |

REFERENCES

[1] P. Adarshvijayan, S. K. Nandakumar, Priyadarshini and K. R. Devabalaji, Economic dispatch problem using whale optimization algorithm, *International Journal of Pure and Applied Mathematics*, 117 (2017), 253–256.
[2] Z. Y. Algamal and M. H. Lee, Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification, *Expert Systems with Applications*, 42 (2015), 9326–9332.
[3] I. I. Ali, Optimal location of SSSC based on PSO to improve voltage profile and reduce Iraqi grid system losses, *Engineering and Technology Journal*, 35 (2017), 372–380.
[4] H. Banati and M. Bajaj, Fire fly based feature selection approach, *International Journal of Computer Science Issues (IJCSCI)*, 8 (2011), 473.
[5] R. Barham and I. Aljarah, Link prediction based on whale optimization algorithm, in 2017 *International Conference on New Trends in Computing Sciences (ICTCS)*, IEEE, (2017), 55–60.
[6] M. L. Bermingham et al., Application of high-dimensional feature selection: evaluation for genomic prediction in man, *Scientific reports*, 5 (2015), 10312.
[7] M. Chih, C. J. Lin, M. S. Chern and T. Y. Ou, Particle swarm optimization with time-varying acceleration coefficients for the multidimensional knapsack problem, *Applied Mathematical Modelling*, 38 (2014), 1338–1350.
[8] L. Y. Chuang, H. W. Chang, C. J. Tu and C. H. Yang, Improved binary PSO for feature selection using gene expression data, *Computational Biology and Chemistry*, **32** (2008), 29–38.

[9] E. Emany, H. M. Zawbaa, K. K. A. Ghany, A. E. Hassanien and B. Parv, Firefly optimization algorithm for feature selection, in *Proceedings of the 7th Balkan Conference on Informatics Conference*, ACM, (2015), 26.

[10] T. R. Golub et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, **286** (1999), 531–537.

[11] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, Gene selection for cancer classification using support vector machines, *Machine Learning*, **46** (2002), 389–422.

[12] P. Hart, The condensed nearest neighbor rule (Corresp.), *IEEE Transactions on Information Theory*, **14** (1968), 515–516.

[13] M. J. Islam, X. Li and Y. Mei, A time-varying transfer function for balancing the exploration and exploitation ability of a binary PSO, *Applied Soft Computing*, **59** (2017), 182–196.

[14] M. Jassim, Improved PSO algorithm to attack transposition cipher, in *Proceedings of the International Conference on Future Networks and Distributed Systems*, ACM, (2017), 144–149.

[15] I. J. Kang et al., Design and efficient synthesis of novel arylthiourea derivatives as potent hepatitis C virus inhibitors, *Bioorganic and Medicinal Chemistry Letters*, **19** (2009), 6063–6068.

[16] I. J. Kang et al., Design, synthesis, and anti-HCV activity of thiourea compounds, *Bioorganic And Medicinal Chemistry Letters*, **19** (2009), 1950–1955.

[17] I. J. Kang et al., Synthesis, activity, and pharmacokinetic properties of a series of conformationally - restricted thiourea analogs as novel hepatitis C virus inhibitors, *Bioorganic and Medicinal Chemistry*, **18** (2010), 6414–6421.

[18] A. Kaveh and M. I. Ghazaan, Enhanced whale optimization algorithm for sizing optimization of skeletal structures, *Mechanics Based Design of Structures and Mach.*, **45** (2017), 345–362.

[19] J. Kennedy and R. C. Eberhart, A discrete binary version of the particle swarm algorithm, in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, IEEE, (1997), 4104–4108.

[20] R. Kohavi and G. H. John, Wrappers for feature subset selection, *Artificial Intelligence*, **97** (1997), 273–324.

[21] N. Khatri, V. Lather and A. K. Madan, Diverse classification models for anti-hepatitis C virus activity of thiourea compounds, *Chemometrics and Intelligent Laboratory Systems*, **140** (2015), 13–21.

[22] Y. Li, Y. Kong, M. Zhang, A. Yan and Z. Liu, Using support vector machine (SVM) for classification of selectivity of H1N1 neuraminidase inhibitors, *Molecular informatics*, **35** (2016), 116–124.

[23] C. Liao, S. Li and Z. Luo, Gene selection using wilcoxon rank sum test and support vector machine for cancer classification, in *International Conference on Computational and Information Science (CIS 2006)*, Springer, (2006), 57-C66.

[24] M. Mafarja, D. Eleyan, S. Abdullah and S. Mirjalili, S-shaped vs. V-shaped transfer functions for ant lion optimization algorithm in feature selection problem, in *Proceedings of the International Conference on Future Networks and Distributed Systems*, ACM, (2017), 21.

[25] M. Mafarja, I. Jaber, S. Ahmed and T. Thaher, Whale optimisation algorithm for high-dimensional small-instance feature selection, *International Journal of Parallel, Emergent and Distributed Systems*, (2019), 1–17.

[26] M. Mafarja et al., Binary dragonfly optimization for feature selection using time-varying transfer functions, *Knowledge-Based Systems*, **161** (2018), 185–204.

[27] M. Mafarja, R. Jarrar, S. Ahmad and A. A. Abusnaina, Feature selection using binary particle swarm optimization with time varying inertia weight strategies, in *Proceedings of the 2nd International Conference on Future Networks and Distributed Systems*, ACM, (2018), 18.

[28] M. Mafarja and S. Mirjalili, Whale optimization approaches for wrapper feature selection, *Applied Soft Computing*, **62** (2018), 441–453.

[29] M. M. Mafarja and S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, *Neurocomputing*, **260** (2017), 302–312.

[30] S. Mirjalili and A. Lewis, S-shaped versus V-shaped transfer functions for binary particle swarm optimization, *Swarm and Evolutionary Computation*, **9** (2013), 1–14.

[31] S. Mirjalili and A. Lewis, The whale optimization algorithm, *Advances in Engineering Software*, **95** (2016), 51–67.
[32] R. Y. M. Nakamura, L. A. M. Pereira, K. A. Costa, D. Rodrigues, J. P. Papa and X. S. Yang, BBA: a binary bat algorithm for feature selection, in 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images, IEEE, (2012), 291–297.
[33] O. S. Qasim and Z. Y. Algamal, Feature selection using particle swarm optimization-based logistic regression model, Chemometrics and Intelligent Laboratory Sys., 182 (2018), 41–46.
[34] D. Rodrigues et al., A wrapper approach for feature selection based on bat algorithm and optimum-path forest, Expert Systems with Applications, 41 (2014), 2250–2258.
[35] D. Rodrigues et al., BCS: A binary cuckoo search algorithm for feature selection, in 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013), IEEE, (2013), 465–468.
[36] D. Singh et al., Gene expression correlates of clinical prostate cancer behavior, Cancer cell, 1 (2001), 203–209.
[37] P. A. Vikhar, Evolutionary algorithms: A critical review and its future prospects, in 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICCC), IEEE, (2016), 261–265.
[38] X. Wang, J. Yang, X. Teng, W. Xia and R. Jensen, Feature selection based on rough sets and particle swarm optimization, Pattern Recognition Letters, 28 (2007), 459–471.
[39] M. West et al., Predicting the clinical status of human breast cancer by using gene expression profiles, Proceedings of the National Academy of Sciences, 98 (2001), 11462–11467.
[40] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016.
[41] X. Wu et al., Top 10 algorithms in data mining, Knowledge and Information Systems, 14 (2008), 1–37.
[42] J. J. Xing, Y. F. Liu, Y. Q. Li, H. Gong and Y. P. Zhou, QSAR classification model for diverse series of antimicrobial agents using classification tree configured by modified particle swarm optimization, Chemometrics and Intelligent Laboratory Systems, 137 (2014), 82–90.
[43] B. Xue, M. Zhang, W. N. Browne and X. Yao, A survey on evolutionary computation approaches to feature selection, IEEE Trans. Evol. Comput., 20 (2016), 606–626.
[44] C. Yang, W. Gao, N. Liu and C. Song, Low-discrepancy sequence initialized particle swarm optimization algorithm with high-order nonlinear time-varying inertia weight, Applied Soft Computing, 29 (2015), 386–394.
[45] X. S. Yang, Firefly algorithm, Nature-Inspired Metaheuristic Algorithms, 20 (2008), 79–90.
[46] B. Zeng, L. Gao and X. Li, Whale swarm algorithm for function optimization, in International Conference on Intelligent Computing, Springer, (2017), 624–639.

Received July 2019; 1st revision September 2019; final revision October 2019.
E-mail address: mohammedkahya@uomosul.edu.iq
E-mail address: suhaib.altamir@uomosul.edu.iq
E-mail address: zakariya.algamal@uomosul.edu.iq