Chemometrics and Intelligent Laboratory Systems 229 (2022) 104635

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Improving Amphetamine-type Stimulants drug classification using chaotic-based time-varying binary whale optimization algorithm

Norfadzlia Mohd Yusof [a], Azah Kamilah Muda [b,*], Satrya Fajri Pratama [b], Ramon Carbo-Dorca [c], Ajith Abraham [d]

[a] Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100, Durian Tunggal, Melaka, Malaysia
[b] Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100, Durian Tunggal, Melaka, Malaysia
[c] Institut de Química Computacional i Catàlisi, Universitat de Girona, 17071, Girona, Catalonia, Spain
[d] Machine Intelligence Research Labs (MIR Labs) Scientific Network for Innovation and Research Excellence, Auburn, WA, USA

ARTICLE INFO

Keywords:
Drug classification
Descriptors selection
Binary whale optimization algorithm
Time-varying transfer function
Logistic-tent chaotic map

ABSTRACT

A new chaotic time-varying binary whale optimization algorithm ($CBWOA_{TV}$) is introduced in this paper to optimize the feature selection process in Amphetamine-type Stimulants (ATS) and non-ATS drugs classification. Two enhancement methods were introduced in this study to provide a fit balance between exploration and exploitation in standard WOA. Firstly, a non-linear time-varying modified Sigmoid transfer function is used as the binarization method. Second, a hybrid Logistic-Tent chaotic map is employed to substitute the pseudorandom numbers of the probability operator in standard WOA. Specific high-dimensional molecular descriptors of ATS and non-ATS drugs were employed to evaluate the efficiency of the proposed algorithm. Experimental results and statistical analysis indicate that the proposed $CBWOA_{TV}$ algorithm can prevent the problem of stagnation and entrapment in local minima in WOA. As a result, optimal descriptors were selected contributing to enhanced classification performance.

1. Introduction

Descriptors selection has been popular research in cheminformatics since it is an essential pre-processing step in the computational chemistry model. The process is important due to the presence of new molecular descriptors that produce a large number of descriptors for each chemical compound. It may cause performance degradation in computation models, especially those that employed machine learning algorithms [1]. Descriptor selection is a well-known non-polynomial (NP) hard combinatorial optimization problem, especially when dealing with high-dimensional data [2–4]. This issue could well be addressed most effectively and efficiently through the development of computational intelligence technologies such as the swarm intelligence (SI) algorithm [5]. SI algorithms have been a popular choice recently in the cheminformatic domain for descriptors selection and some of the works are listed in Table 1.

In 2016, Mirjalili and Lewis presented the Whale Optimization Algorithm (WOA) algorithm in the SI domain. This algorithm is designed based on the hunting behavior of humpback whales which is known as bubble-net foraging [14]. The initial WOA algorithm is developed in a continuous version. The binary version of WOA is required to solve binary optimization problems [15–17].

The performance of swarm-based algorithms is highly dependent on the balance between exploitation and exploration. Premature convergence is caused by excessive exploitation and less exploration, while greater exploration and less exploitation may provoke difficulties to reach the optimal solution [18,19]. In the feature selection domain, the stagnation and premature convergence in WOA can lead to an inadequate selection of significant features and poor classification performance [20]. Various methods applied by researchers to improve the WOA convergence problem. The two methods that will be discussed and implemented in this study are the transfer function and chaos theory.

A transfer function is one of the binarization methods popularly used in the SI domain [21]. This method is straightforward and does not enhance the complexity of the SI algorithm. Besides that, the transfer function can improve the exploration and exploitation of the SI algorithm therefore the selection of an appropriate transfer function is important. Various transfer functions have been utilized and examined

* Corresponding author.
E-mail address: azah@utem.edu.my (A.K. Muda).

https://doi.org/10.1016/j.chemolab.2022.104635
Received 2 July 2022; Received in revised form 27 July 2022; Accepted 27 July 2022
Available online 10 August 2022
0169-7439/

Table 1
Some of the works of SI-based descriptors selection in the cheminformatics domain.

Ref	Application	Algorithm
[6]	Drug classification	Binary particle swarm optimization algorithm (BPSO), BWOA, and binary manta-ray optimization (BMRFO)
[7]	Drug classification	Chaotic dragonfly algorithm (CDA)
[8]	QSAR modeling	Hybrid Harris hawks optimization with cuckoo search and chaotic map (CHHO–CS)
[9]	QSAR modeling	Harris hawks optimization (HHO) algorithm
[10]	QSAR modeling	Salps algorithm
[11]	QSAR modeling	Seagull optimization algorithm (SOA)
[12]	QSAR modeling	Binary grasshopper optimization algorithm (BGOA)
[13]	QSAR/QSPR classification modeling	Binary pigeon optimization algorithm (BPO)

by several researchers in BWOA for feature selection problems in supervised data classification [22–27].

Another improvement method in the SI domain is the application of chaos theory. Chaos is a non-linear system with deterministic dynamic behavior [28–30]. Chaos has the property of randomness, ergodicity, and non-repetition, and is highly sensitive to its initial conditions and parameters [29]. These chaotic system characteristics can enrich the searchability of the SI algorithm. Each chaotic map has its unique formulation. Implementation of different chaotic maps within the SI algorithm will yield different results. Table 2 listed some of the works that hybrid chaotic maps with SI algorithms to solve various problems.

WOA algorithm has operators to balance between exploration and exploitation, unlike other SI algorithms such as particle swarm optimization (PSO) that apply one equation to update the agent's position, which can increase the chance trapping in local optima. Several operators in WOA are assigned with pseudorandom numbers, for example, the operator p where it is the probability to choose either the shrinking encircling or the spiral model mechanism during the optimization process. Therefore, Kaur and Arora developed a Chaotic WOA (CWOA) for global optimization in 2018 which adopted chaos theory to substitute the operator p with a set of chaotic numbers [34]. In the work, ten chaotic maps listed in Table 2 were adopted with the same initial value. Superior efficiency was displayed by CWOA with Tent chaotic map. CWOA was also implemented in the feature selection domain by Sayed et al. [35]. The authors employed chaos set for each pseudorandom-based operator in standard WOA. Like in former research, ten chaotic maps were examined within several CWOA variations using ten benchmark datasets from the University of California Irvine (UCI) Machine Learning repository. The chosen datasets contain a feature size between 10 and 82. Experimental results disclose the Tent map with modifications of exploration operators in CWOA-PI outperforms with the highest performance. Moreover, CWOA was customed to resolve engineering problems [39,40]. Table 3 lists several chaos implementation techniques in WOA from the works of literature.

Table 2
Previous works of the chaotic-based SI algorithms in various applications.

Ref.	Algorithms	Application	Initial point	Chaotic map
[7]	Chaotic dragonfly	Feature selection	0.7	Gauss
[31]	Chaotic crow search	Feature selection	0.7	Sine
[32]	Chaotic salps swarm	Feature selection	0.7	Tent
[33]	Chaotic binary particle swarm optimization	Feature selection	0.48	Tent
[34]	Chaotic whale	Global optimization	0.7	Tent
[35]	Chaotic whale	Feature selection	0.7	Tent
[29]	Chaotic salp swarm	Feature selection, Global optimization	0.7	Logistic
[36]	Chaotic grasshopper	Global optimization	Not stated	Circle
[37]	Chaotic antlion	Parameter optimization	Random	Logistic
[38]	Chaotic bat	Global optimization	Random	Sinusoidal

Our study made the following contributions: (i) $CBWOA_{TV}$ algorithm is an improved version of WOA with the integration of time-varying modified Sigmoid transfer function and Logistic-Tent chaotic map is proposed. (ii) $CBWOA_{TV}$ algorithm is incorporated with wrapper feature selection algorithm to optimize wrapper feature selection technique. (iii) we show that implementation of the $CBWOA_{TV}$ algorithm advantages from a fast convergence speed and a robust search capability. (iv) we demonstrate the importance of descriptors selection by the $CBWOA_{TV}$ algorithm in providing better drug classification performance in terms of accuracy and speed. (v) we show the proposed $CBWOA_{TV}$ algorithm leads to quality solutions and can outperform the existing BWOA and CWOA-PI algorithms, as well as other state-of-the-art SI algorithms.

Our study is limited to several mechanisms that are: (i) $CBWOA_{TV}$ wrapper feature selection algorithm is developed to solve the descriptors selection problem in the drug analysis domain. (ii) A high-dimensional chemical dataset is used for algorithms evaluation. (iii) This study does not apply any data preprocessing to the chemical dataset.

The remainder of this paper is structured as follows. Section 2 and Section 3 briefly describe the whale optimization algorithm (WOA) and chaotic maps. The detailed description of the proposed CBWOA algorithm is explained in Section 4. Section 5 describes the experimental setting. Section 6 reports and discusses the experimental results. Finally, Section 7 concludes the paper and suggests possible future works.

2. Whale optimization algorithm (WOA)

In the initial stage, the WOA algorithm will assume the target prey as the best search agent that is near to the optimum. Then, other whales (search agents) will update their positions based on the best search agent. WOA swarming behavior is simulated in mathematical formulations below:

$$D = |C \cdot Whale^{*}(t) - Whale(t)| \tag{1}$$

$$Whale(t+1) = Whale^{*}(t) - A \cdot D \tag{2}$$

where t is the iteration number. $Whale(t)$ denotes the candidate search agent at iteration number t and $Whale^{*}(t)$ indicate as the best search agent (prey) so far. A and C are coefficient numbers mathematically formulated by Equations (3) and (4) below. D indicates the distance vector between the whale (search agent) and the prey (best search agent). In each iteration $Whale^{*}(t)$ is updated when there is a better solution.

$$A = 2 \cdot a \cdot r_1 + a \tag{3}$$

$$C = 2 \cdot r_2 \tag{4}$$

where r_1 and r_2 represent random vectors in [0, 1].

Table 3
Different techniques of chaotic map implementation in WOA for various applications in the literature.

Description	Application
Population initialization	Global optimization [41]
Shrinking circle mechanism	Feature selection [35,42], engineering [39,40,43,44]
Spiral shaped mechanism	Feature selection [35], engineering [44]
Probability parameter, p	Global optimization [34], feature selection [35,42], engineering [44,45]
Choosing a random search agent	Feature selection [35]
Convergence factor, a	Engineering [44]

$$a = 2 - t\frac{2}{maxIter} \tag{5}$$

a is a convergence factor that gradually decreases from 2 to 0 over iterations. t indicates the iteration number and *maxIter* is the maximum number of iterations.

The humpback whale's bubble-net behavior is designed based on two mechanisms:

(1) Shrinking encircling of prey: The humpback move in a shrinking encircling along a spiral-shaped path towards the prey by decreasing the value a in Equation (3) and Equation (5). The fluctuation range of A also decreases as a decreases. A is a random value in the interval $(-a, a)$
(2) Spiral updating position: A logarithmic spiral function is used to imitate the helix-shaped movement of the humpback whales between the candidate whale (search agent) $Whale(t)$, and the prey (best search agent), $Whale^*(t)$ so far. This procedure is mathematically expressed in Equation (6) and Equation (7).

$$D^* = |Whale^*(t) - Whale(t)|, \tag{6}$$

$$Whale(t+1) = D^* \cdot e^{bl} \cdot \cos(2\pi l) + Whale^*(t), \tag{7}$$

where b is a constant defining the logarithmic spiral shape and l is a random number in [-1, 1].

The humpback whale swims around its prey in a narrow circle, while swimming along a spiral path. To simulate this simultaneous behavior, an assumption of 50% probability is used to choose between these two mechanisms to update the whales' position. The mathematical formulation to model this behavior is established as follows:

$$Whale(t+1) = \begin{cases} Whale^*(t) - A \cdot D, & if\ p < 0.5, \\ D^* \cdot e^{bl} \cdot \cos(2\pi l) + Whale^*(t), & if\ p \geq 0.5, \end{cases} \tag{8}$$

where p is a random number in. $[0,\ 1]$.

Contradicting the exploitation phase, in the exploration phase a search agent position is updated following a randomly chosen search agent instead of the best search agent found so far. A, in Equation (3) with the random values greater than 1 or less than -1. will urge the search agent to move far away from the reference whale. With this mechanism and $|A| > 1$, it emphasizes exploration and allows WOA to perform a global search to overcome the problem of the local optima. Equation (9) and Equation (10) describe the mathematical formulation:

$$D = |C \cdot Whale_{rand} - Whale| \tag{9}$$

$$Whale(t+1) = Whale_{rand} - A \cdot D \tag{10}$$

where $Whale_{rand}$ indicates a whale that is randomly chosen from the current population.

3. Chaos theory

Many SI algorithms contain randomness parameters. The randomness is drawn randomly from a uniform or Gaussian distribution. From the optimization perspective, a chaotic map can help escape from the local optimum and enhance the convergence rate to attain the global optimal solution. A chaotic map is regularly used to define optimization parameters as well as to generate the initial population. The initial value of the chaotic index is usually determined randomly or by a fixed number between 0 and 1. However, it is important to note that the initial value may have a significant effect on the fluctuation pattern of the chaotic mapping [20]. Table 4 listed the commonly used chaotic maps in the literature specifically in the metaheuristic algorithms [46–48].

Table 4
The chaotic maps and their definition.

No.	Name of map	Equation	Range
1	Chebyshev	$p_{q+1} = \cos(q\cos^{-1}(p_q))$	$(-1, 1)$
2	Circle	$p_{q+1} = mod\left(p_q + r - (\frac{1}{2\pi})\sin(2\pi p_q), 1\right), l = 0.5\ and\ r = 0.2$	(0, 1)
3	Gauss/mouse	$p_{q+1} = \begin{cases} 1, & p_q = 0 \\ \frac{1}{mod(p_q, 1)}, & otherwise \end{cases}$	(0, 1)
4	Iterative	$p_{q+1} = \sin(\frac{l\pi}{p_q}), l = 0.7$	$(-1, 1)$
5	Logistic	$p_{q+1} = lp_q(1 - p_q), l = 4$	(0, 1)
6	Piecewise	$p_{q+1} = \begin{cases} \frac{p_q}{l}, & 0 \leq p_q < l \\ \frac{p_q - l}{0.5 - l}, & l \leq p_q < 0.5 \\ \frac{1 - p_q}{l}, & 1 - l \leq p_q < 1 \end{cases}$	(0, 1)
7	Sine	$p_{q+1} = \frac{l}{4}\sin(\pi p_q), l = 4$	(0, 1)
8	Singer	$p_{q+1} = \mu(7.86p_q - 23.31p_q^2 + 28.75p_q^3 - 13.302875p_q^4), \mu = 1.07$	(0, 1)
9	Sinusoidal	$p_{q+1} = lp_q^2\sin(\pi p_q), l = 2.3$	(0, 1)
10	Tent	$p_{q+1} = \begin{cases} \frac{p_q}{0.7}, & p_q < 0.7 \\ \frac{10}{3}(1 - p_q), & p_q \geq 0.7 \end{cases}$	(0, 1)

4. Proposed method

The framework of the proposed technique is depicted in Fig. 1. The process begins by inputting the original Three Dimensional Exact Legendre Moment Invariants (3D ELMI) molecular descriptors [49] into the $CBWOA_{TV}$ algorithm. $CBWOA_{TV}$ algorithm is responsible for searching and selecting the relevant descriptors. Since it employed the wrapper method, the k-Nearest Neighbor (k-NN) classifier with the Euclidean distance metric and k is set to 5 as in Refs. [22,50–53] is employed to evaluate the selected descriptors by $CBWOA_{TV}$. K-NN generates the classification error rate and passes it back to the $CBWOA_{TV}$ to calculate the fitness of the selected descriptors using the fitness function in Equation (14). The smaller fitness rate indicates the relevant descriptors have been selected. The process of finding relevant descriptors is repeated until the specified maximum iteration is reached. This proposed method is expected to accelerate and ease the learning process of the classifier after the elimination of irrelevant descriptors from original molecular descriptors. Finally, the optimal selected descriptors subset is validated using the same classifier to generate the final classification results for the assessment process.

4.1. The binary whale optimization algorithm (BWOA)

The binary version of the WOA is mandatory to generate binary solutions {0, 1} for the feature selection problem [54]. Similar to WOA, in BWOA the search agents (solutions) update their positions continuously to any point in the search space based on the best search agent discovered so far. Then the real position of whales is converted to binary values using a transfer function. This technique forces whales to move in a binary space by probability definition which updates each element (feature) in the solution (features subset) to 0 (not selected feature) or 1 (selected feature) [55].

4.1.1. Non-linear time-varying Sigmoid transfer function

In this work, we adopted a recently proposed transfer function in Ref. [56] as the binarization method. The transfer function in Equation (11) is applied to convert the continuous value of WOA's position to a

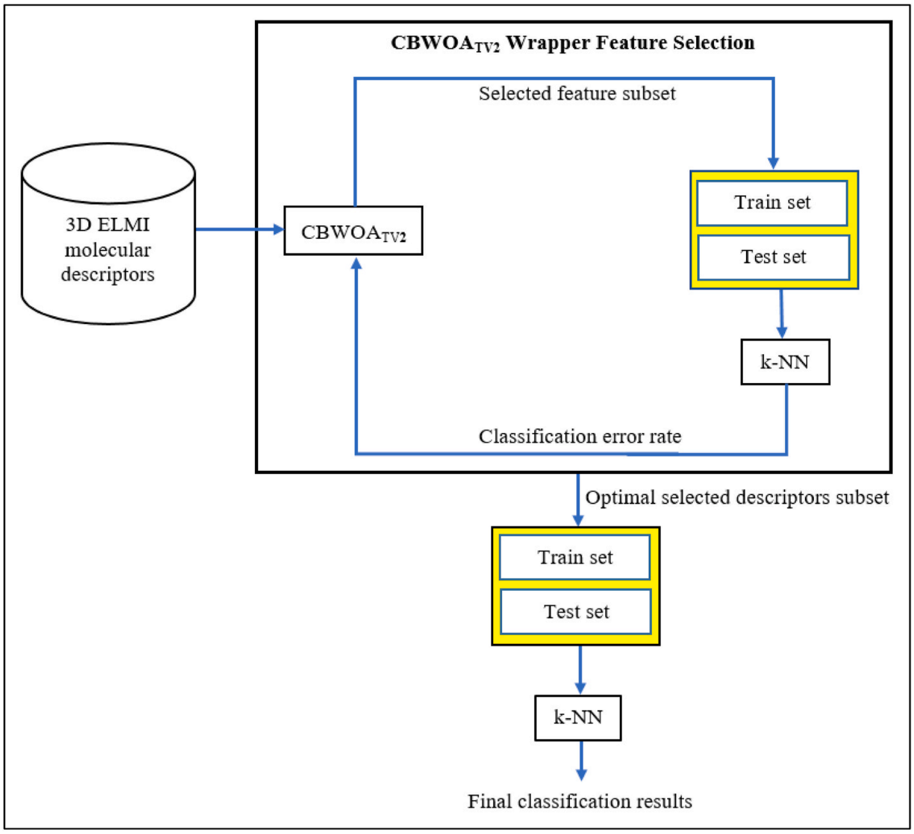

Fig. 1. Framework of the proposed CBWOA$_{TV}$ wrapper feature selection technique.

probability value:

$$Sigmoid\left(\overrightarrow{Whale}(t+1)\right) = \frac{1}{1+e^{-10\left(\overrightarrow{Whale}(t+1)/Tv-0.5\right)}} \tag{11}$$

where Tv is a time-varying control parameter that decreases over iterations.

The time-varying (Tv) is updated using a non-linear scheme as presented in Equation (12):

$$Tv(t) = Tv_{max} + (Tv_{min} - Tv_{max})\left(\frac{Itr_{t+1}}{Itr_{max}}\right)^{\alpha} \tag{12}$$

where Tv_{min}, Tv_{max} denote the minimum and maximum values of the control parameter, Itr_{t+1} is the current iteration while Itr_{max} signifies the maximum number of iterations. In this study, Tv_{max}, Tv_{min} and α is set to 4, 0.1, and 0.5 same as [56].

Lastly, the probability value in each element in the whale's position vector is converted to a binary value using Equation (13). This formulation was proposed by Kennedy and Eberhart in Ref. [42]:

$$\overrightarrow{Whale}(t+1) = \begin{cases} 1, & if\ rand < Sigmoid\left(\overrightarrow{Whale}(t+1)\right) \\ 0, & otherwise \end{cases} \tag{13}$$

4.1.2. Fitness function

The success of the feature selection algorithm is measured based on two objectives: increase classification accuracy and reduce features [57]. The fitness function used in the feature selection technique is designed to have a balance between these objectives. In a wrapper feature selection method, a classification algorithm is involved to evaluate the selected features. The classification error rate produced from the evaluation process is then used in the fitness function formulation in Equation (14). The fitness function is utilized by the optimization algorithm to evaluate the recommended feature subset. The best feature subset is the one with a small classification error rate and a less number of selected features. The small fitness value means the relevant feature subset has been selected.

$$\downarrow Fitness = \alpha \times CE + \beta \times \frac{|F_{select}|}{|F_{actual}|} \tag{14}$$

where CE represents the classification error rate calculated by the classifier. $|F_{select}|$ is the number of selected features, and $|F_{actual}|$ is the original feature size, α, and β are the two parameters corresponding to the importance of classification quality and subset length. $\alpha \in [1, 0]$ and $\beta = (1-\alpha)$ are adopted from Refs. [58,59]. For this work, the classification performance is considered to be the most important metric, thus α is seted to 0.99, thus β is 0.01 [58,60].

4.2. The chaotic time-varying binary whale optimization algorithm (CBWOA$_{TV}$)

In the original WOA, the pseudorandom numbers are assigned to parameter $p = rand()$ to help in controlling the exploration and exploitation. However, in this work, the operator p is taken from the sequence of Logistic-Tent chaotic vector, cp. As far as we know, no previous study has employed the Logistic-Tent in WOA [34]. The mathematical formulation of the Logistic-Tent map is presented in Equation (14).

$$p_{q+1} = \begin{cases} mod\left(rp_q\left(1-p_q\right)+\dfrac{(4-r)p_q}{2},1\right), & if\ p_q < 0.5 \\ mod\left(rp_q\left(1-p_q\right)+\dfrac{(4-r)\left(1-p_q\right)}{2},1\right), & if\ p_q \geq 0.5 \end{cases}, r \in [1,4] \tag{14}$$

```
Randomly generate an initial population of whales Whale_i (i = 1, 2, 3, N)
Initialize the generation counter t, max_iteration and WOA parameters
Generate Chaotic map cp with vector size of max_iteration
Evaluate fitness
Whale* = the best fit whale
while t < tMax do
for each solution in Whale population do
      p = cp(t)
       Update WOA parameters
       if p < 0.5 then
              if (|A| < 1) then
                     Update solution by Equation 2
                     Calculate the probability vector using Equation 11
                     Update the binary position using Equation 13
              else if (|A|≥ 1) then
                     Select a random solution, Whale_rand
                     Update solution by Equation 10
                     Calculate the probability vector using Equation 11
                     Update the binary position using Equation 13
              end if
       else p  ≥ 0.5
              Update solution by Equation 7
              Calculate the probability vector using Equation 11
              Update the binary position using Equation 13
       end if
end for
Determine the fitness value of each solution
Update Whale* if there is a better solution
t = t + 1
end while
return Whale*
```

Fig. 2. Pseudo-code of the $CBWOA_{TV}$ algorithm.

Table 5
Parameter settings.

Algorithm	Parameter	Value
All	Search agent (whale) size, N	15
	Iteration length, $tMax$	300
	No. of runs, R	10
	Problem dimension	1185
	Search domain	[0,1]
	α in the fitness function	0.99
	β in the fitness function	0.01
BWOA [6,22], BWOA-TV2 [26], CWOA-PI [35], $CBWOA_{TV}$	$\vec{a}$	2 to 0
MRFOv3 [61]	Somersault factor, S	2
BWOA-TV2 [26], $CBWOA_{TV}$	Tv_{max}	4
	Tv_{min}	0.1
TV_T-BPSO [62]	Tv_{max}	5
	Tv_{min}	1
$BHHO_{TV4}$ [63]	Tv_{max}	4
	Tv_{min}	0.01

In the equation, p represents the initial point, and q denotes the index of the chaotic sequence. In this study, we draw the initial value p_0 as in Refs. [20,34,35], which is 0.7. The pseudo-code of the $CBWOA_{TV}$ algorithm is presented in Fig. 2.

5. Experimental setting

5.1. Dataset

The 3D ELMI molecular descriptors dataset contains an equal sample size of 3595 ATS drug compounds and 3595 non-ATS drug compounds. Each drug is described by 1186 descriptors including molecule id and the class label (0 - non-ATS and 1 - ATS). However, the molecule id is excluded during experimentation.

In the experiments, the hold-out validation strategy with stratified data partitioning of 80% train set and 20% test set is applied to evaluate the proposed algorithm and comparative algorithms [51,57]. Stratified data partitioning is utilized to ensure the train set and test set are composed of a balanced number of ATS and non-ATS drugs sample. All algorithms are repeated ten times with different random seeds. Different random seeds are used to provide different data partitions in each different run. This is one way to inspect the robustness of the proposed algorithm. The experimental results are presented as the average metrics achieved from ten independent runs to obtain statistically valid results. All algorithms are developed using Matlab R2021a that runs on a PC with an Intel Core i7-6700 machine, 3.40 GHz CPU with Windows 10 operating system, and 16 GB of RAM. The parameter settings used in the experiments are shown in Table 5.

5.2. Performance measurement

Several performance metrics are used to validate the performance of the proposed $CBWOA_{TV}$ algorithm and its comparative algorithms which include the maximum fitness, minimum fitness, average fitness and its corresponding standard deviation, average size of selected descriptors, and average computational time (CT) in seconds. Then, the final classification performance is measured by the following metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

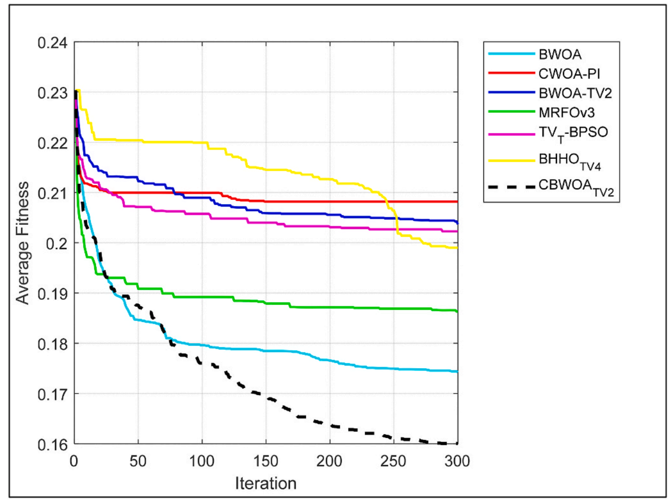

Fig. 3. Convergence curves of proposed $CBWOA_{TV}$ and other comparative SI algorithms.

$$Specificity = \frac{TN}{TN + FP}$$

$$G - mean = \sqrt{Sensitivity * Specificity}$$

$$AUC = \frac{Sensitivity + Specificity}{2}$$

The reliability of the proposed $CBWOA_{TV}$ algorithm is statistically analyzed by the non-parametric Wilcoxon signed-rank sum test [49]. A qualitative assessment is performed based on convergence curves depicted in Fig. 3.

6. Experimental results and discussion

Two different evaluations were reported in this section. The first evaluation aims to compare the $CBWOA_{TV}$ algorithm with other comparative SI algorithms which include BWOA [6,22], CWOA-PI [35], BWOA-TV2 [26], TV_T-BPSO [62], $BHHO_{TV4}$ [63], MRFOv3 [61]. The second evaluation is to compare the classification performance using the selected descriptors by the $CBWOA_{TV}$ algorithm with the classification performance of three different classifiers k-NN, support vector machine (SVM) with radial basis function (RBF) kernel, and random forest (RF) classifiers using the original descriptors.

6.1. Evaluation of CBWOATV and other SI algorithms

Initially, we have executed 11 variants of $CBWOA_{TV}$ implementing chaotic maps in Table 4 including the Logistic-Tent map. In this paper, only the best variant of $CBWOA_{TV}$ is reported and used for comparative analysis. From our experiments, $CBWOA_{TV}$ with a Logistic-Tent map is found to be superior to others.

Table 6
Comparison results based on fitness performance and average computational time (CT) in seconds.

Algorithm	Max	Min	Mean	Std	Avg. CT
BWOA	0.18182	0.16479	0.17422	0.00561	2278.5
CWOA-PI	0.22065	0.18954	0.20819	0.00981	3078.41
BWOA-TV2	0.21195	0.18715	0.20369	0.00706	11726.90
MRFOv3	0.19854	0.17982	0.18631	**0.00505**	677.6
TV_T-BPSO	0.21167	0.19210	0.20227	0.00557	7166.2
$BHHO_{TV4}$	0.21573	0.18333	0.19901	0.00930	**226.8**
$CBWOA_{TV}$	**0.17436**	**0.15141**	**0.16006**	0.00702	685.7

Table 7
Comparison results based on the average number of selected descriptors, average accuracy, average G-mean, and average AUC.

Algorithm	Avg. Nd	Avg. accuracy	Avg. G-mean	Avg. AUC
BWOA	184	82.6	82.8	82.8
CWOA-PI	204.10	79.14	79.40	79.46
BWOA-TV2	805	80.1	80.4	80.5
MRFOv3	**34**	81.2	81.3	81.4
TV_T-BPSO	588	80.1	80.4	80.5
$BHHO_{TV4}$	38	79.9	80.1	80.1
$CBWOA_{TV}$	98	**83.9**	**84.1**	**84.2**

Table 8
P-values of Wilcoxon signed-rank test based on mean fitness, average accuracy, average G-mean, and average AUC between $CBWOA_{TV}$ with other SI algorithms.

Algorithm	Mean fitness	Avg. accuracy	Avg. G-mean	Avg. AUC
BWOA	0.002	0.002	0.002	0.002
CWOA-PI	0.002	0.006	0.002	0.002
BWOA-TV2	0.002	0.002	0.002	0.002
MRFOv3	0.002	0.006	0.002	0.002
TV_T-BPSO	0.002	0.006	0.002	0.002
$BHHO_{TV4}$	0.002	0.002	0.002	0.002

The performance results in Table 6, Table 7, and the Wilcoxon signed-rank test statistical analysis in Table 8 show the superiority of the proposed $CBWOA_{TV}$ in convergence, robustness, stability, and obtaining relevant descriptors after attaining the lowest maximum, minimum, and mean fitness rate. All algorithms are seen to have a low standard deviation, but the MRFOv3 provides the lowest. $BHHO_{TV4}$ has the fastest convergence speed but failed to obtain the global optimal solution. In terms of the average size of selected descriptors, MRFOv3 exhibits the fewest descriptors but might neglect the significant one after obtaining the low average accuracy, G-mean, and AUC. CWOA-PI is the worst performer among all with a long computational time. It exhibits CWOA-PI not able to perform well with large feature dimensions.

Fig. 3 indicates that BWOA, CWOA-PI, BWOA-TV2, MRFOv3, TV_T-BPSO, and $BHHO_{TV4}$ were inferior to $CBWOA_{TV}$. $CBWOA_{TV}$ is seen to converge faster and deeper than other algorithms to find the optimal solution. This confirmed the two applied improvement methods influenced a satisfactory performance in the $CBWOA_{TV}$ algorithm.

Furthermore, Wilcoxon signed-rank tests based on the mean fitness, accuracy, G-mean, and AUC were carried out. The statistical analysis is

Table 9
Comparison results based on descriptors' length and average classification speed of the proposed technique with k-NN, SVM, and RF.

Algorithm	$CBWOA_{TV}$	Original - k-NN	Original - SVM	Original - RF
No. of descriptors	**98**	1185	1185	1185
Average CT (s)	**0.5**	3.2	21.6	111.6

Table 10
P-values of Wilcoxon signed-rank between the proposed technique with k-NN, SVM, and RF.

Algorithm	Original - k-NN	Original - SVM	Original - RF
Avg. accuracy	0.002	0.006	0.041
Avg. G-mean	0.002	0.002	0.037
Avg. AUC	0.002	0.002	0.037

to validate whether there are significant differences between $CBWOA_{TV}$ and comparative algorithms with a 0.05 significance level. The null hypothesis states that no significant difference between the two algorithms is accepted when the *p*-value is greater or equal to 0.05. Otherwise, the null hypothesis is rejected when the *p*-value is lower than 0.05. As seen in Table 8, all the comparative algorithms attained *p*-values less than 0.05 in all metrics when compared with $CBWOA_{TV}$, which established there are significant differences.

6.2. *Evaluation of $CBWOA_{TV}$ and other supervised classification algorithms*

This section presents the efficacy of $CBWOA_{TV}$ in reducing the number of descriptors and increasing the classification performance. The original dataset is fed to three different classifiers, k-NN, SVM (RBF), and RF. The classification performances of these three classifiers are compared with our proposed technique. Referring to the bar graph in Fig. 4, our proposed technique obtained the best scores for average accuracy, average G-mean, and average AUC. As displayed in Table 9, only 8% of the descriptor in the original dataset is found significant by $CBWOA_{TV}$. Fewer number descriptors reduced the time required for all classifiers to learn and make a prediction. The *p*-value of Wilcoxon signed-rank tests in Table 10 denote there are significant differences in classification performance after employment of $CBWOA_{TV}$ as a descriptors selector.

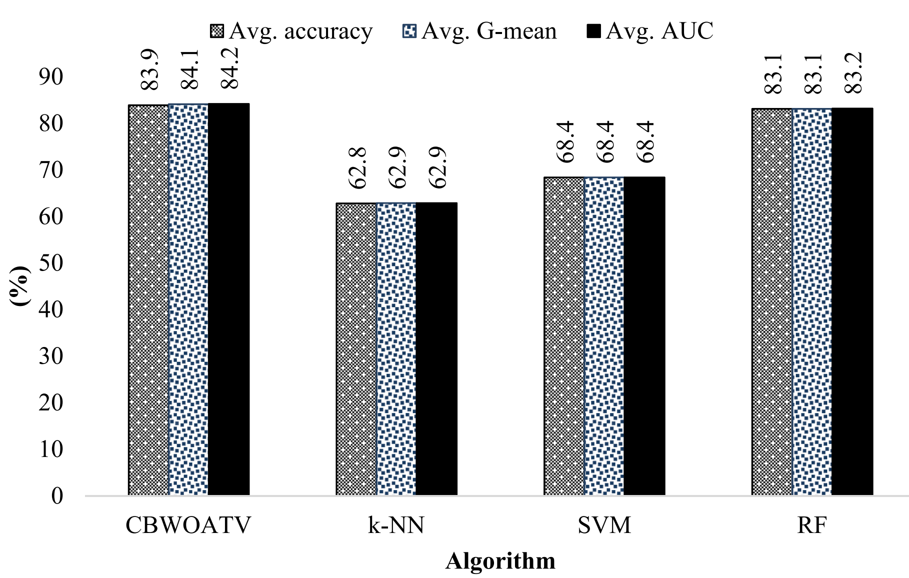

Fig. 4. Comparison of classification performance of our proposed technique and three classifiers.

7. Conclusions and future work

This work presents the CBWOA$_{TV}$ wrapper feature selection algorithm in finding relevant descriptors for the k-NN classifier to achieve better learning and obtain good prediction ability in classifying between ATS and non-ATS drugs. The implementation of a time-varying modified Sigmoid transfer function and Logistic-Tent map provide a fit balance between the exploration and exploitation stages of the WOA algorithm. The comparative analysis shows CBWOA$_{TV}$ outperforms BWOA, TV_T-BPSO, BHHO$_{TV4,}$ and MRFOv3 algorithms. The comparison results of CBWOA$_{TV}$ with k-NN, SVM(RBF), and RF classifiers have demonstrated the importance of descriptors reduction in improving the classification accuracy and accelerating the classification time. The application of different implementation strategies of the Logistic-Tent chaotic map is interesting to explore. Furthermore, the employment of the proposed CBWOA$_{TV}$ algorithm to other real-world problems [64,65] can be investigated and applied. For future improvement, we want to apply the 3D ELMI molecular descriptors to the recent metaheuristic algorithms such as Farmland Fertility [66], Artificial gorilla troops optimizer [67], and African vultures optimization algorithm [68].

Author statement

Norfadzlia Mohd Yusof: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – Original draft, Visualization. Azah Kamilah Muda: Conceptualization, Supervison, Project administration, Funding acquisition. Satrya Fajri Pratama: Conceptualization, Data Curation, Resource, Supervison. Ramon Carbo-Dorca: Wrting – Review & Editing. Ajith Abraham: Wrting – Review & Editing.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Universiti Teknikal Malaysia Melaka through the Fundamental Research Grant Scheme [FRGS/1/2020/FTMK-CACT/F00461] from the Ministry of Higher Education, Malaysia.

References

[1] M. Goodarzi, B. Dejaegher, Y Vander Heyden, Feature selection methods in QSAR studies, J. AOAC Int. 95 (2012) 636–651, https://doi.org/10.5740/jaoacint.SGE_Goodarzi.

[2] M.J. Goldanloo, F.S. Gharehchopogh, A Hybrid OBL-Based Firefly Algorithm with Symbiotic Organisms Search Algorithm for Solving Continuous Optimization Problems, Springer US, 2022.

[3] H. Mohammadzadeh, F.S. Gharehchopogh, A multi-agent system based for solving high-dimensional optimization problems: a case study on email spam detection, Int. J. Commun. Syst. 34 (2021) 1–48, https://doi.org/10.1002/dac.4670.

[4] H.R.R. Zaman, F.S. Gharehchopogh, An Improved Particle Swarm Optimization with Backtracking Search Optimization Algorithm for Solving Continuous Optimization Problems, Springer London, 2021.

[5] F.S. Gharehchopogh, An improved tunicate swarm algorithm with best-random mutation strategy for global optimization problems, J. Bionic. Eng. (2022), https://doi.org/10.1007/s42235-022-00185-1.

[6] N. Mohd Yusof, A. Kamilah Muda, S. Fajri Pratama, Swarm intelligence-based feature selection for amphetamine-type Stimulants (ATS) drug 3D molecular structure classification, Appl. Artif. Intell. 35 (2021) 914–932, https://doi.org/10.1080/08839514.2021.1966882.

[7] G.I. Sayed, A. Tharwat, A.E. Hassanien, Chaotic dragonfly algorithm: an improved metaheuristic algorithm for feature selection, Appl. Intell. 49 (2019) 188–205, https://doi.org/10.1007/s10489-018-1261-8.

[8] E.H. Houssein, M.E. Hosney, M. Elhoseny, et al., Hybrid Harris hawks optimization with cuckoo search for drug design and discovery in chemoinformatics, Sci. Rep. 10 (2020) 1–22, https://doi.org/10.1038/s41598-020-71502-z.

[9] E.H. Houssein, M.E. Hosney, D. Oliva, et al., A novel hybrid Harris hawks optimization and support vector machines for drug design and discovery, Comput. Chem. Eng. 133 (2020), 106656, https://doi.org/10.1016/j.compchemeng.2019.106656.

[10] A.G. Hussien, A.E. Hassanien, E.H. Houssein, Swarming behaviour of salps algorithm for predicting chemical compound activities, in: 2017 IEEE 8th International Conference on Intelligent Computing and Information Systems, ICICIS, 2017, pp. 315–320, 2017.

[11] E.H. Houssein, M.E. Hosney, D. Oliva, A hybrid seagull optimization algorithm for chemical descriptors classification, in: 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference, MIUCC, 2021, pp. 381–386, 2021.

[12] Z.Y. Algamal, M.K. Qasim, M.H. Lee, H.T.M. Ali, QSAR model for predicting neuraminidase inhibitors of influenza A viruses (H1N1) based on adaptive grasshopper optimization algorithm, SAR QSAR Environ. Res. 31 (2020) 803–814, https://doi.org/10.1080/1062936X.2020.1818616.

[13] Z.Y. Algamal, M.K. Qasim, M.H. Lee, H.T. Mohammad Ali, High-dimensional QSAR/QSPR classification modeling based on improving pigeon optimization algorithm, Chemometr. Intell. Lab. Syst. 206 (2020), 104170, https://doi.org/10.1016/j.chemolab.2020.104170.

[14] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Software 95 (2016) 51–67, https://doi.org/10.1016/j.advengsoft.2016.01.008.

[15] V. Kumar, D. Kumar, Binary whale optimization algorithm and its application to unit commitment problem, Neural Comput. Appl. 32 (2020) 2095–2123, https://doi.org/10.1007/s00521-018-3796-3.

[16] K.S. Reddy, L. Panwar, B.K. Panigrahi, R. Kumar, Binary whale optimization algorithm: a new metaheuristic approach for profit-based unit commitment problems in competitive electricity markets, Eng. Optim. 51 (2019) 369–389.

[17] A. Al-Ani, Ant colony optimization for feature subset selection, Int. J. Comput. Electr. Autom. Control. Inf. Eng. 1 (2007) 999–1002.

[18] F.S. Gharehchopogh, B. Farnad, A. Alizadeh, A modified farmland fertility algorithm for solving constrained engineering problems, Concurrency Comput. Pract. Ex. 33 (2021), https://doi.org/10.1002/cpe.6310.

[19] H. Mohmmadzadeh, F.S. Gharehchopogh, An efficient binary chaotic symbiotic organisms search algorithm approaches for feature selection problems, J. Supercomput. 77 (2021) 9102–9144, https://doi.org/10.1007/s11227-021-03626-6.

[20] W. Yang, K. Xia, S. Fan, et al., A multi-strategy whale optimization algorithm and its application, Eng. Appl. Artif. Intell. 108 (2022), https://doi.org/10.1016/j.engappai.2021.104558.

[21] S. Ghafori, F.S. Gharehchopogh, Advances in spotted hyena optimizer: a comprehensive survey, Arch. Comput. Methods Eng. (2021), https://doi.org/10.1007/s11831-021-09624-4.

[22] H.F. Eid, Binary whale optimisation: an effective swarm algorithm for feature selection, Int. J. Metaheuristics (IJMHeur) 7 (2018) 67, https://doi.org/10.1504/ijmheur.2018.10012912.

[23] A.G. Hussien, E.H. Houssein, A.E. Hassanien, A binary whale optimization algorithm with hyperbolic tangent fitness function for feature selection, 2017 IEEE 8th Int Conf Intell Comput Inf Syst ICICIS (2017) 166–172, https://doi.org/10.1109/INTELCIS.2017.8260031, 2017 2018-Janua.

[24] A.G. Hussien, A.E. Hassanien, E.H. Houssein, et al., S-shaped binary whale optimization algorithm for feature selection, in: Advances in Intelligent Systems and Computing, Springer Verlag, 2019, pp. 79–87.

[25] N. Mohd Yusof, A.K. Muda, S.F. Pratama, et al., Improved swarm intelligence algorithms with time-varying modified sigmoid transfer function for amphetamine-type Stimulants drug classification, Bioresour. Technol. Rep. (2019), 100310, https://doi.org/10.1016/j.chemolab.2022.104574.

[26] M.A. Kahya, S.A. Altamir, Z.Y. Algamal, Improving whale optimization algorithm for feature selection with a time-varying transfer function, Numer. Algebra Control Optim. 11 (2021) 87–98, https://doi.org/10.3934/naco.2020017.

[27] H. Zamani, M.-H. Nadimi-Shahraki, Feature selection based on whale optimization algorithm for diseases diagnosis, Int. J. Comput. Sci. Inf. Secur. 14 (2016) 1243–1247, https://doi.org/10.13140/RG.2.2.29065.88161.

[28] J. Too, A.R. Abdullah, Chaotic atom search optimization for feature selection, Arabian J. Sci. Eng. 45 (2020) 6063–6079, https://doi.org/10.1007/s13369-020-04486-7.

[29] G.I. Sayed, G. Khoriba, M.H. Haggag, A novel chaotic salp swarm algorithm for global optimization and feature selection, Appl. Intell. 48 (2018) 3462–3481, https://doi.org/10.1007/s10489-018-1158-6.

[30] Z.A. Dizaji, F.S. Gharehchopogh, A hybrid of ant colony optimization and chaos optimization algorithms approach for software cost estimation, Indian J. Sci. Technol. 8 (2015) 128–133, https://doi.org/10.17485/ijst/2015/v8i2/57776.

[31] G.I. Sayed, A.E. Hassanien, A.T. Azar, Feature selection via a novel chaotic crow search algorithm, Neural Comput. Appl. 31 (2019) 171–188, https://doi.org/10.1007/s00521-017-2988-6.

[32] A.E. Hegazy, M.A. Makhlouf, G.S. El-Tawel, Feature selection using chaotic salp swarm algorithm for data classification, Arabian J. Sci. Eng. 44 (2019) 3801–3816, https://doi.org/10.1007/s13369-018-3680-6.

[33] L.Y. Chuang, C.H. Yang, J.C. Li, Chaotic maps based on binary particle swarm optimization for feature selection, Appl. Soft Comput. J 11 (2011) 239–248, https://doi.org/10.1016/j.asoc.2009.11.014.

[34] G. Kaur, S. Arora, Chaotic whale optimization algorithm, J Comput Des Eng 5 (2018) 275–284, https://doi.org/10.1016/j.jcde.2017.12.006.
[35] G.I. Sayed, A. Darwish, A.E. Hassanien, A new chaotic whale optimization algorithm for features selection, J. Classif. 35 (2018) 300–344, https://doi.org/10.1007/s00357-018-9261-2.
[36] S. Arora, P. Anand, Chaotic grasshopper optimization algorithm for global optimization, Neural Comput. Appl. 31 (2019) 4385–4405, https://doi.org/10.1007/s00521-018-3343-2.
[37] A. Tharwat, A.E. Hassanien, Chaotic antlion algorithm for parameter optimization of support vector machine, Appl. Intell. 48 (2018) 670–686, https://doi.org/10.1007/s10489-017-0994-0.
[38] A.H. Gandomi, X.S. Yang, Chaotic bat algorithm, J. Comput. Sci. 5 (2014) 224–232, https://doi.org/10.1016/j.jocs.2013.10.002.
[39] C. Paul, P.K. Roy, V. Mukherjee, Chaotic whale optimization algorithm for optimal solution of combined heat and power economic dispatch problem incorporating wind, Renew. Energy Focus. 35 (2020) 56–71, https://doi.org/10.1016/j.ref.2020.06.008.
[40] F. Kavin, R. Senthilkumar, Chaotic whale optimized fractional order PID controller design for desalination process, Comput. Mater. Continua (CMC) 71 (2022) 2789–2806, https://doi.org/10.32604/cmc.2022.021577.
[41] S. Dong, C. Bu, Y. Wang, Improved whale optimization algorithm based on the tent chaotic mapping and nonlinear convergence factor, J. Phys. Conf. Ser. 1682 (2020), https://doi.org/10.1088/1742-6596/1682/1/012055.
[42] T. Thaher, A. Zaguia, S. Al Azwari, et al., An enhanced evolutionary student performance prediction model using whale optimization algorithm boosted with sine-cosine mechanism, Appl. Sci. 11 (2021), https://doi.org/10.3390/app112110237.
[43] D. Prasad, A. Mukherjee, V. Mukherjee, Temperature dependent optimal power flow using chaotic whale optimization algorithm, Expet Syst. 38 (2021) 1–14, https://doi.org/10.1111/exsy.12685.
[44] D. Yousri, D. Allam, M.B. Eteiba, Chaotic whale optimizer variants for parameters estimation of the chaotic behavior in Permanent Magnet Synchronous Motor, Appl. Soft Comput. J 74 (2019) 479–503, https://doi.org/10.1016/j.asoc.2018.10.032.
[45] D. Oliva, M. Abd El Aziz, A. Ella Hassanien, Parameter estimation of photovoltaic cells using an improved chaotic whale optimization algorithm, Appl. Energy 200 (2017) 141–154, https://doi.org/10.1016/j.apenergy.2017.05.029.
[46] F.S. Gharehchopogh, H. Gholizadeh, A comprehensive survey: whale Optimization Algorithm and its applications, Swarm Evol. Comput. 48 (2019) 1–24, https://doi.org/10.1016/j.swevo.2019.03.004.
[47] H.M. Mohammed, S.U. Umar, T.A. Rashid, A systematic and meta-analysis survey of whale optimization algorithm, Comput. Intell. Neurosci. (2019) 1–25, https://doi.org/10.1155/2019/8718571, 2019.
[48] F.S. Gharehchopogh, I. Maleki, Z.A. Dizaji, Chaotic Vortex Search Algorithm: Metaheuristic Algorithm for Feature Selection, Springer Berlin Heidelberg, 2021.
[49] S.F. Pratama, Three-Dimensional Exact Legendre Moment Invariants for Amphetamine-type Stimulants Molecular Structure Representation, Universiti Teknikal Malaysia Melaka (UTeM), 2017.
[50] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary grey wolf optimization approaches for feature selection, Neurocomputing 172 (2016) 371–381, https://doi.org/10.1016/j.neucom.2015.06.083.
[51] M. Mafarja, I. Aljarah, H. Faris, et al., Binary grasshopper optimisation algorithm approaches for feature selection problems, Expert Syst. Appl. 117 (2019) 267–286, https://doi.org/10.1016/j.eswa.2018.09.015.
[52] E. Emary, H.M. Zawbaa, K.K.A. Ghany, et al., Firefly optimization algorithm for feature selection, ACM Int Conf Proceeding (2015), https://doi.org/10.1145/2801081.2801091. Ser 02-04-Sept.
[53] H. Mohammadzadeh, F.S. Gharehchopogh, A novel hybrid whale optimization algorithm with flower pollination algorithm for feature selection: case study Email spam detection, Comput. Intell. (2020) 1–34, https://doi.org/10.1111/coin.12397.
[54] L. Brezočnik, I. Fister, V. Podgorelec, Swarm intelligence algorithms for feature selection: a review, Appl. Sci. 8 (2018), https://doi.org/10.3390/app8091521.
[55] S. Mirjalili, A. Lewis, S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization, Swarm Evol. Comput. 9 (2013) 1–14, https://doi.org/10.1016/j.swevo.2012.09.002.
[56] N. Mohd Yusof, A.K. Muda, S.F. Pratama, A. Abraham, A novel nonlinear time-varying sigmoid transfer function in binary whale optimization algorithm for descriptors selection in drug classification, Mol. Divers. (2022), https://doi.org/10.1007/s11030-022-10410-y.
[57] M. Mafarja, S. Mirjalili, Whale optimization approaches for wrapper feature selection, Appl. Soft Comput. 62 (2018) 441–453, https://doi.org/10.1016/j.asoc.2017.11.006.
[58] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary ant lion approaches for feature selection, Neurocomputing 213 (2016) 54–65, https://doi.org/10.1016/j.neucom.2016.03.101.
[59] M. Sharawi, M. Zawbaa H, E. Emary, H.M. Zawbaa, Feature selection approach based on whale optimization algorithm, in: 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI), Doha. Institute of Electrical and Electronics Engineers Inc., 2017, pp. 163–168.
[60] M. Mafarja, I. Aljarah, A.A. Heidari, et al., Binary dragonfly optimization for feature selection using time-varying transfer functions, Knowl. Base Syst. 161 (2018) 185–204, https://doi.org/10.1016/j.knosys.2018.08.003.
[61] K.K. Ghosh, R. Guha, S.K. Bera, et al., S-shaped versus V-shaped transfer functions for binary Manta ray foraging optimization in feature selection problem, Neural Comput. Appl. 33 (2021) 11027–11041, https://doi.org/10.1007/s00521-020-05560-9.
[62] M.J. Islam, X. Li, Y. Mei, A time-varying transfer function for balancing the exploration and exploitation ability of a binary PSO, Appl. Soft Comput. J 59 (2017) 182–196, https://doi.org/10.1016/j.asoc.2017.04.050.
[63] H. Chantar, T. Thaher, H. Turabieh, et al., BHHO-TVS: a binary harris hawks optimizer with time-varying scheme for solving data classification problems, Appl. Sci. 11 (2021), https://doi.org/10.3390/app11146516.
[64] A. Benyamin, S.G. Farhad, B. Saeid, Discrete farmland fertility optimization algorithm with metropolis acceptance criterion for traveling salesman problems, Int. J. Intell. Syst. 36 (2021) 1270–1303, https://doi.org/10.1002/int.22342.
[65] F.S. Gharehchopogh, B. Abdollahzadeh, An efficient harris hawk optimization algorithm for solving the travelling salesman problem, in: Cluster Comput 5, 2021, https://doi.org/10.1007/s10586-021-03304-5.
[66] T.S. Naseri, F.S. Gharehchopogh, A feature selection based on the farmland fertility algorithm for improved intrusion detection systems, J. Netw. Syst. Manag. 30 (2022), https://doi.org/10.1007/s10922-022-09653-9.
[67] B. Abdollahzadeh, F. Soleimanian Gharehchopogh, S. Mirjalili, Artificial gorilla troops optimizer: a new nature-inspired metaheuristic algorithm for global optimization problems, Int. J. Intell. Syst. 36 (2021) 5887–5958, https://doi.org/10.1002/int.22535.
[68] B. Abdollahzadeh, F.S. Gharehchopogh, S. Mirjalili, African vultures optimization algorithm: a new nature-inspired metaheuristic algorithm for global optimization problems, Comput. Ind. Eng. 158 (2021), 107408, https://doi.org/10.1016/j.cie.2021.107408.