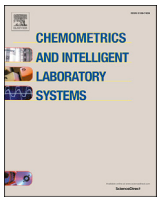
Chemometrics and Intelligent Laboratory Systems 226 (2022) 104574

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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



# Improved swarm intelligence algorithms with time-varying modified Sigmoid transfer function for Amphetamine-type stimulants drug classification

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:*
Descriptors selection
Time-varying transfer function
Binary optimization algorithm
Drug classification

ABSTRACT

Swarm-intelligence (SI) algorithms have received great attention in addressing various binary optimization problems such as feature selection. In this article, a new time-varying modified Sigmoid transfer function with two time-varying updating schemes is proposed as the binarization method for particle swarm optimization (PSO), grey wolf optimization algorithm (GWO), whale optimization algorithm (WOA), harris hawk optimization (HHO), and manta-ray foraging optimization (MRFO). The new binary algorithms, BPSO, BGWOA, BWOA, BHHO, and BMRFO algorithms are utilized for solving the descriptors selection problem in supervised Amphetamine-type Stimulants (ATS) drug classification task. The goal of this study is to improve the speed of convergence and classification accuracy. To evaluate the performance of the proposed methods, experiments were carried out on a specific chemical dataset containing molecular descriptors of ATS and non-ATS drugs. The results obtained showed that the proposed methods' performances on the chemical dataset are promising in near to optimal convergence, fast computation, increased classification accuracy, and enormous reduction in descriptor size.

## 1. Introduction

Amphetamine-type stimulant (ATS) drugs are one of the popular synthetic drugs of abuse. These substances were originally developed for pharmacological research, but, the underground chemist keep modifying the chemical structure of these compounds to evade legal regulation. The novelty of these substances makes them undetectable by traditional drug testing methods [1].

Nowadays, there are different devices and test kits available in the market for use in ATS drug testing [2–4]. However, several drawbacks were discovered within these methods such as long preparation and execution time, costly apparatus, and equipment, complex testing process, requiring well-trained technicians, unreliable and inconsistent outputs from different test kits, and outdated analytical methods. At present, computational methods have been shown as the promising techniques in the cheminformatics field such as in drug design and discovery [5], drug-non drug classification [6], molecular similarity analysis [7], toxicity prediction [8,9], and quantitative structure-activity/property relationships (QSAR/QSPR) analysis [10]. Computational methods also offer much cheaper and faster procedures. Molecular descriptors are an important component that has provided support for many modern computational models. Molecular descriptors are numerical indexes encoded from molecular structure representation of different dimensionalities (0D, 1D, 2D, 3D, or 4D). The higher the dimensionality the more information about the molecular features is stored in the descriptors.

Due to the rapid increment of chemical data, machine learning has become a promising tool to process big data at high volume, veracity, and velocity and with enormous flexibility [11]. In a previous study, Pratama et al. proposed an approach for ATS drug identification by employing a newly developed 3D image pre-processing technique called 3D Exact Legendre Moment Invariants (3D ELMI) as a feature extractor algorithm and several classification algorithms [12,13]. 3D ELMI is a molecular descriptor algorithm that is used to calculate descriptors of 7190 drug

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

https://doi.org/10.1016/j.chemolab.2022.104574
Received 23 March 2022; Received in revised form 13 April 2022; Accepted 26 April 2022
Available online 6 May 2022

**Table 1**
Recent works on SI-based descriptors selection in the cheminformatics domain.

| Ref | Swarm Intelligence Algorithms | Application |
|---|---|---|
| [5] | Harris hawks optimization (HHO) algorithm | QSAR modeling |
| [16] | Binary grasshopper optimization algorithm (BGOA) | QSAR modeling |
| [29] | Binary particle swarm optimization algorithm (BPSO) and genetic algorithm (GA) | QSAR modeling |
| [30] | Binary particle swarm optimization algorithm (BPSO), binary whale optimization algorithm (BWOA), and binary manta-ray optimization (BMRFO) | Drug classification |
| [31] | Chaotic dragonfly algorithm (CDA) | Drug classification |
| [32] | Binary pigeon optimization algorithm (BPO) | QSAR/QSPR classification modeling |
| [33] | Hybrid Harris hawks optimization with cuckoo search and chaotic map (CHHO–CS) | QSAR modeling |
| [34] | Seagull optimization algorithm (SOA) | QSAR modeling |

compounds (equal size of ATS and non-ATS drugs). It generates 1185 descriptors for each drug compound and is then used as input to the classification algorithms to perform the identification task. The experimental results found that the random forest (RF) classifier is superior in achieving the highest classification accuracy. According to Ref. [14], molecular classification involved three steps: feature extraction, feature selection, and classification. This study aims to improve the ATS drug classification performance by carrying out the feature selection as a data-preprocessing step before executing the classification task.

The existence of new molecular descriptors that generate high-dimensional descriptors has made the feature or descriptor selection step required in computational modeling. Descriptor selection is a well-known non-polynomial (NP) hard combinatorial search problem as the number of possible feature subsets grows exponentially with the increase of dimensionality. The traditional feature selection techniques are inefficient to handle medium or large descriptors or datasets. Therefore, the SI algorithm is one of the core technology to address this issue. SI algorithm is classified as a population-based optimization algorithm having several advantages [15]:

- Ease of implementation.
- Fewer operators compared to evolutionary approaches.
- Fewer parameters to tune.
- Retain information about the search space throughout the iteration.
- Regularly use memory to save the best solution obtained so far.

Some of the successful works that implemented SI algorithms in the descriptors selection problem are outlined in Table 1.

Descriptor selection has become popular research in the cheminformatics domain where researchers attempt to identify the lowest feasible number of descriptors that can provide good predictive performance [16]. The research area of incorporating and implementing SI algorithms to the feature selection problem is still active to date. According to the No Free Lunch (NFL) theorem, there is no universal algorithm that applies to all optimization problems [17]. Therefore, there is always an opportunity to come up with new metaheuristic-based feature selection algorithms to enhance the process of solving feature selection problems. Numerous new approaches are originated in the literature [18–22]. Motivated by this, this research proposed ten new SI-based feature selection algorithms to obtain the significance and discriminative 3D ELMI descriptors to enhance the classification performance.

The initial version of the SI algorithm produces a continuous solution and is only applicable to solve continuous optimization problems. Therefore, the binary version of the SI algorithm is mandatory to generate a binary solution for addressing binary optimization problems. Examples of binary optimization problems are feature selection [23], and traveling salesman problem (TSP) [24]. The common practice is to use the transfer function as a conversion method [25–27]. The implementation of a transfer function is straightforward and does not increase the complexity of the original algorithm. In addition, the utilization of a suitable transfer function will provide a good balance between the exploration and exploitation phases in the SI algorithm resulting in a better convergence and good classification accuracy. Several popular transfer functions used in the literature are listed in Table 2.

**Table 2**
Standard and commonly used transfer functions in the literature.

| (a) S-shape transfer functions | | (b) V-shape transfer functions | |
|---|---|---|---|
| Name | Formulation | Name | Formulation |
| **S1** | $S(x) = \frac{1}{1+e^{-x}}$ | **V1** | $V(x) = \lvert\tanh(x)\rvert$ |
| **S2** | $S(x) = \frac{1}{1+e^{-2x}}$ | **V2** | $V(x) = \left\lvert \mathrm{erf}\left(\frac{\sqrt{\pi}}{2}x\right)\right\rvert$ |
| **S3** | $S(x) = \frac{1}{1+e^{-x/2}}$ | **V3** | $V(x) = \left\lvert \frac{s}{\sqrt{1+x^2}}\right\rvert$ |
| **S4** | $S(x) = \frac{1}{1+e^{-x/3}}$ | **V3** | $V(x) = \left\lvert \frac{2}{\pi} arctan\left(\frac{\pi}{2}x\right)\right\rvert$ |
| **S5** | $S(x) = \frac{1}{1+e^{-10(x-0.5)}}$ | | |

In the present research, we introduced a time-varying modified Sigmoid transfer function with a linear time-varying updating strategy. We also adopted the transfer function that we have proposed in Ref. [28]. To evaluate the efficiency of the particular transfer functions, we integrate them into five continuous SI algorithms: particle swarm optimization algorithm (PSO), whale optimization algorithm (WOA), grey wolf optimization algorithm (GWO), harris hawk optimization algorithm (HHO), and manta-ray foraging optimization algorithm (MRFO). The characteristics of these SI algorithms are summarized in Table 3 while Table 4 and 5 present the transfer functions used to produce their binary versions.

In a few words, this paper introduces a new approach for tackling descriptors selection problem based on the PSO, BWO, GWO, HHO, and MRFO algorithms, and its main contributions can be summarized as follows:

1. A novel time-varying modified Sigmoid transfer function with a linear (TV1) time-varying updating scheme is introduced as a binarization technique for the metaheuristic algorithm.
2. Ten new binary variants of PSO, BWOA, GWO, HHO, and MRFO algorithms are developed by employing TV1 and our recently proposed transfer function (TV2) in Ref. [28]: $BPSO_{TV1}$, $BPSO_{TV2}$, $BWOA_{TV1}$, $BWOA_{TV2}$, $BGWO_{TV1}$, $BGWO_{TV2}$, $BHHO_{TV1}$, $BHHO_{TV2}$, $BMRFO_{TV1}$, $BMRFO_{TV2}$.
3. These proposed SI algorithms are adapted as a feature search for wrapper feature selection in a supervised binary classification task that differentiates ATS and non-ATS drugs.
4. The final results were assessed based on different performance metrics, including the average fitness, average classification accuracy, average fitness, average number of selected features, as well as the respective standard deviation values.
5. The significance of the proposed algorithms was validated against competitive algorithms using a Wilcoxon's rank-sum non-parametric statistical test at a significance level of $\alpha = 0.05$.

There are several limitations of this study which include:

1. The transfer functions are validated on five SI-based optimization algorithms only.
2. The algorithms are used to solve the descriptors selection problem in the drug analysis domain.
3. Only one chemical dataset is used for algorithms evaluation.
4. This study does not apply any data preprocessing to the dataset.

The remainder of this paper is structured as follows. Section 2 explains the concepts of the proposed new transfer functions and their

**Table 3**
Nature-inspired metaheuristic algorithm within SI category [35] is employed in the present study.

| Algorithm | Acronym | Subcategory | Type | Social behavior method | Year | Ref. |
|---|---|---|---|---|---|---|
| Particle swarm optimization | PSO | Movement | Flying | It is based on a group of birds randomly searching for food in the area. | 1995 | [36] |
| Grey wolf optimization | GWO | Terrestrial | Foraging | It mimics the leadership hierarchy and hunting mechanism of grey wolves in nature. | 2014 | [15] |
| Whale optimization algorithm | WOA | Aquatic | Foraging | It mimics the bubble-net feeding in the foraging behavior of the humpback whales:<br>1) encircling prey<br>2) bubble-net attacking method<br>3) search for prey | 2016 | [37] |
| Harris hawk optimization | HHO | Flying | Foraging | It is inspired by the cooperative behaviors and chasing style of Harris hawks in nature called surprise pounce. | 2019 | [38] |
| Manta ray foraging optimization | MRFO | Aquatic | Foraging | It mimics three unique foraging strategies of manta rays:<br>1 chain foraging,<br>2 cyclone foraging,<br>3 somersault foraging | 2020 | [39] |

**Table 4**
Existing S-shaped and Time-varying S-shaped transfer functions used within SI algorithm in feature selection applications.

| Algorithms | Ref. | Standard | | | | | Time-varying | | | | | Remarks |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | S1 | S2 | S3 | S4 | S5 | S1 | S2 | S3 | S4 | S5 | |
| BWOA | [40] | ✓ | ✓ | ✓ | ✓ | | | | | | | – |
| | [41] | | | | | ✓ | | | | | | – |
| | [42] | | | | | | ✓ | | | | | Linear, non-linear, and shape decreasing time-varying update schemes. |
| BHHO | [43] | ✓ | | | | | | | | | | – |
| BGWO | [44] | | | | | ✓ | | | | | | – |
| BPSO | [32] | ✓ | | | | | ✓ | | | | | – |
| | [46] | | | | | ✓ | | | | | | – |
| BMRFO | [46] | ✓ | | | | | | | | | | – |
| | [47] | ✓ | ✓ | ✓ | ✓ | | ✓ | | | | | – |
| **Proposed BPSO, BGWO, BWOA, BHHO and BMRFO** | | | | | | | | | | | ✓ | Linear and non-linear time-varying update schemes. |

**Table 5**
Existing V-shaped and Time-varying V-shaped transfer functions used within SI algorithms in feature selection applications.

| Algorithms | Ref. | Standard | | | | Time-varying | | | | Remarks |
|---|---|---|---|---|---|---|---|---|---|---|
| | | V1 | V2 | V3 | V4 | V1 | V2 | V3 | V4 | – |
| BWOA | [48] | ✓ | ✓ | ✓ | ✓ | | | | | – |
| BHHO | [49] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | – |
| BPSO | [32] | | | | ✓ | | | | ✓ | – |
| BMRFO | [47] | ✓ | ✓ | ✓ | ✓ | | | | | – |
| BHHO | [49] | ✓ | ✓ | ✓ | ✓ | | | | | – |

implementation in PSO, GWO, WOA, HHO, and MRFO algorithms. Section 3 describes the necessary material and methods used in the experiments. Results and discussions are presented in Section 4. Finally, Section 5 concludes and highlights some future works.

## 2. The binary version of PSO, GWO, WOA, HHO, and MRFO

[23,24,50]. In BPSO, BWOA, BGWOA, BHHO, and BMRFO, 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 search agents is converted to binary values using the proposed time-varying modified Sigmoid transfer function. This technique forces search agents to move in a binary space by probability definition which updates each element (feature) in the solution (features subset) to 1 (selected feature) or 0 (not selected feature) [51].

### 2.1. Time-varying modified Sigmoid transfer function

In this work, BGWO, BWOA, BHHO, and BMRFO employed a time-varying modified Sigmoid transfer function proposed by our previous research [28] which can convert all real values of position into probability values [0, 1]. The transfer function formulation is shown in Equation (11):

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

where *Agent, t, i, d* represent the search agent's position, current iteration number, order of the agent in the population, and search space dimension. *Tv* denotes as a control parameter of the time-varying that decreases during iterations. After converting the position to probability values the position vector can be updated with the probability of its position by Equation (13) which was proposed by Kennedy and Eberhart in Ref. [52]:

$$Agent_i^d(t+1) = \begin{cases} 1, \ if \ rand \ < Sigmoid\left(Agent_i^d(t+1)\right) \\ 0, \ otherwise \end{cases} \tag{12}$$

*rand* is a random number drawn from the uniform distribution [0, 1].

For BPSO, the transfer function is applied to the velocity, *v*, to convert to a probability value:

$$Sigmoid\left(v_i^d(t+1)\right) = \frac{1}{1+e^{-10\left(v_i^d(t+1)/Tv-0.5\right)}} \tag{13}$$

Finally, the new position is updated using Equation (14).

**Table 6**
Summary of 3D ELMI molecular descriptors dataset.

| Dataset | No. of attribute | No of. class attribute | No. of instance |
|---|---|---|---|
| 3D ELMI | 1186 | 1 | 7190 |

$$Agent_i^d(t+1) = \begin{cases} 1, & if\ rand\ < Sigmoid\left(v_i^d(t+1)\right) \\ 0, & otherwise \end{cases} \tag{14}$$

In this work, two time-varying updating schemes are utilized as follows:

The linear time-varying updating strategy, TV1 is presented in Equation (15) was employed in several works in the literature [32,42,49,53]:

$$\tau = \tau_{\max} + (\tau_{\min} - \tau_{\max})\left(\frac{Itr_{t+1}}{Itr_{\max}}\right) \tag{15a}$$

The non-linear time-varying updating strategy, TV2 is presented in Equation (16) and was employed in Refs. [28,42]:

$$\tau = \tau_{\max} + (\tau_{\min} - \tau_{\max})\left(\frac{Itr_{t+1}}{Itr_{\max}}\right)^{\alpha} \tag{16}$$

where $\tau_{min}$, $\tau_{max}$ are the minimum and maximum values of the control parameter, $Itr_{t+1}$ is the current iteration while $Itr_{max}$ is the maximum number of iterations. In this study, $\tau_{min}$, $\tau_{max}$ and $\alpha$ are set to 4 [49], 0.1, and 0.5 [42].

The update strategy for time-varying over TF is very critical to prevent entrapment in local optima. This strategy, on the other hand, is used to search for the global optima by exploring multiple regions in the search space.

### 2.2. Fitness function

The two aims of feature selection are to obtain a lower number of selected features and achieve higher classification accuracy [54]. Since a wrapper-based feature selection technique is used, a learning algorithm (classifier) is involved in the evaluation process. The best feature subset is the one with less number of selected features and a small classification error rate. The fitness function used in the feature selection technique is designed to have a balance between the two criteria. Recommended solutions are evaluated using the fitness function in Equation (15):

$$\downarrow Fitness = \alpha \times CE + \beta \times \frac{|F_{select}|}{|F_{actual}|} \tag{15b}$$

where $CE$ represents the classification error rate calculated by the k-Nearest Neighbour (k-NN) classifier [55] with the Euclidean distance metric and k is set to 5 [41,44,56,57]. $|F_{select}|$ is the length of the selected feature subset, and $|F_{actual}|$ are the original feature size in the dataset, $\alpha$, and $\beta$ 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]. As the classification performance is considered the most important metric in this work, thus $\alpha$ is seted to 0.99, thus $\beta$ is 0.01 adopted from literature [45,58,60,61].

## 3. Method and material

### 3.1. Dataset

In this study, a special chemical dataset was adopted in the experiments. The dataset contains 1187 descriptors to represent the 3D molecular structure of 3595 ATS dan 3595 non-ATS drugs compounds. The descriptors were generated by using a novel 3D Exact Legendre Moment Invariant molecular descriptors algorithm that was developed by Pratama in Refs. [12,13]. The descriptors comprise molecule id, 1185 moment invariants values, and a binary class label, 0 (non-ATS drug) and 1 (ATS drug). During experimentation, the molecule id descriptor is excluded. The summary of the dataset is presented in Table 6.

### 3.2. Methodology

This study presents an approach for improving ATS and non-ATS drug classification by proposing 10 novel binary SI algorithms to optimize the wrapper feature selection algorithms in selecting significance descriptors to enhance the classification performance. As illustrated in Fig. 1, the process begins with inputting the whole descriptors from a certain data partitioned into the proposed algorithms to perform the feature selection task. The descriptors searching and selection processes are executed iteratively and terminate when it reaches the maximum number of iteration. Since the wrapper method is utilized, the k-NN classifier is adopted to evaluate the selected descriptors subset. k-NN is a simple, easy to implement, and fast classification algorithm [56,62]. The classification error rate obtained by the k-NN classifier is transmitted to the proposed algorithms to obtain the fitness rate of the selected descriptors using the fitness function formulation in Equation (16). The small fitness rate indicates that significant features have been selected. Finally, the selected optimal descriptors are evaluated using the same classifier to obtain the final classification accuracy. The obtained results are then used for validation and comparison with other comparative algorithms.

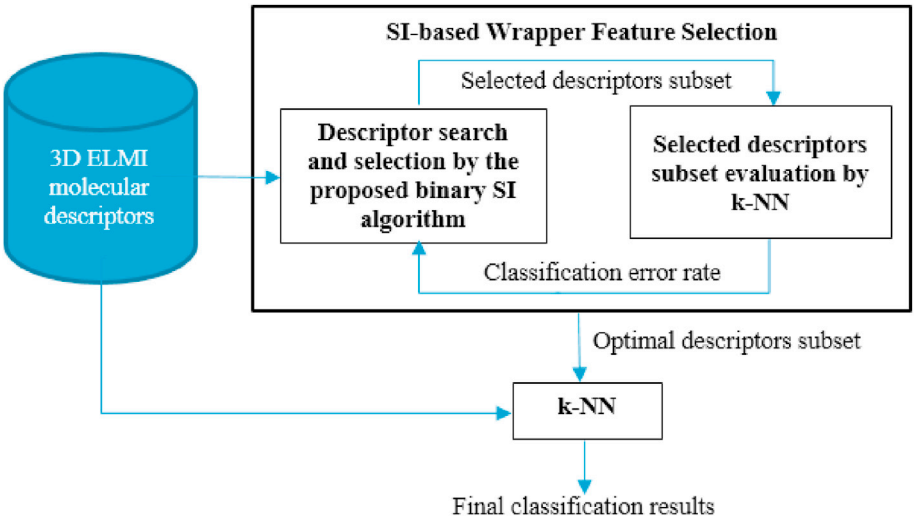



**Fig. 1.** Framework of the proposed binary SI-based wrapper descriptors selection technique.

**Table 7**
Parameter settings.

| Algorithm | Parameter | Value | Ref. |
|---|---|---|---|
| All | Search agent size, $N$ | 15 | |
| | Iteration length, $tMax$ | 150 | |
| | No. of runs, $R$ | 10 | |
| | Problem dimension, $D$ | 1185 | |
| | Search domain | [0,1] | |
| | $\alpha$ in the fitness function | 0.99 | [45,58,60, 61] |
| | $\beta$ in the fitness function | 0.01 | |
| BWOA. BWOA-TV2, BWOA $_{TV1}$, BWOA $_{TV2}$ | $\vec{a}$ | 2 to 0 | [46] |
| BPSO, BPSO $_{TV1}$, BPSO $_{TV2}$ | Inertia weight, $w$ | 0.9 to 04 | [46] |
| | Cognitive factor, $c_1$ | 2 | |
| | Social factor, $c_2$ | 2 | |
| | Maximum velocity, $v_{max}$ | 6 | |
| | Minimum velocity, $v_{min}$ | −6 | |
| bGWO2, BGWO $_{TV1}$, BGWO $_{TV2}$ | – | – | [44] |
| BHHO, BHHO$_{TV4}$, BHHO $_{TV1}$, BHHO $_{TV2}$ | Energy, $E$ | 2 to 0 | [49] |
| | $\beta$ in the levy flight pattern | 1.5 | |
| BHHO$_{TV4}$ | $\tau_{max}$ | 4 | [49] |
| | $\tau_{min}$ | 0.01 | |
| BMRFO, BMRFO $_{TV1}$, BMRFO $_{TV2}$ | Somersault factor, $S$ | 2 | [46] |

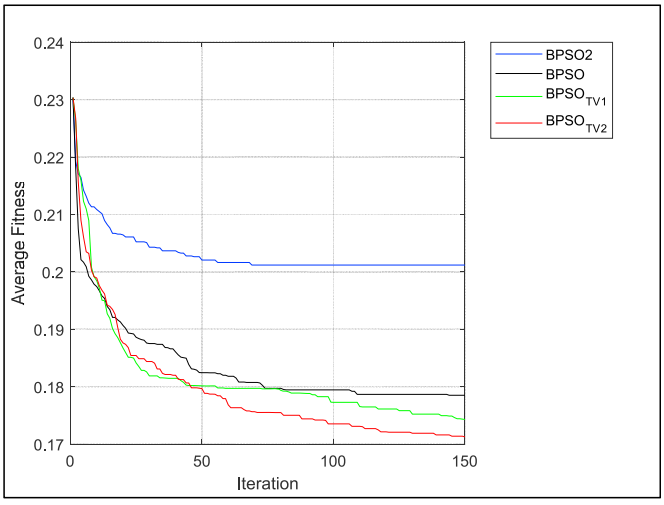



**Fig. 2.** Convergence curves of proposed BPSO$_{TV1}$ and BPSO$_{TV2}$ algorithms with other binary variants of PSO algorithm.

### 3.3. Parameter settings

In the experimental works, all algorithms are executed ten times with different random seeds. The usage of random seed is for reproducibility where all algorithms will start with the same random number and data partitions. For experimentation, the stratified hold-out validation strategy is implemented with data partitioned to 80% training set and 20% testing set. This partitioning was used in various previous works in the literature [54,56]. The training set is used to train the classifier during the optimization process while the testing set is used to evaluate the selected features. The experimental results are displayed 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. Table 7 outlines the parameter settings of the algorithms.

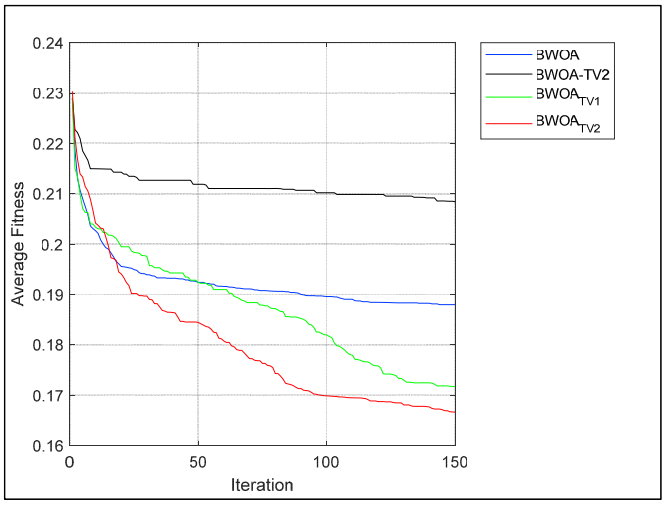



**Fig. 3.** Convergence curves of proposed BWOA$_{TV1}$ and BWOA$_{TV2}$ algorithms with other binary variants of WOA algorithm.

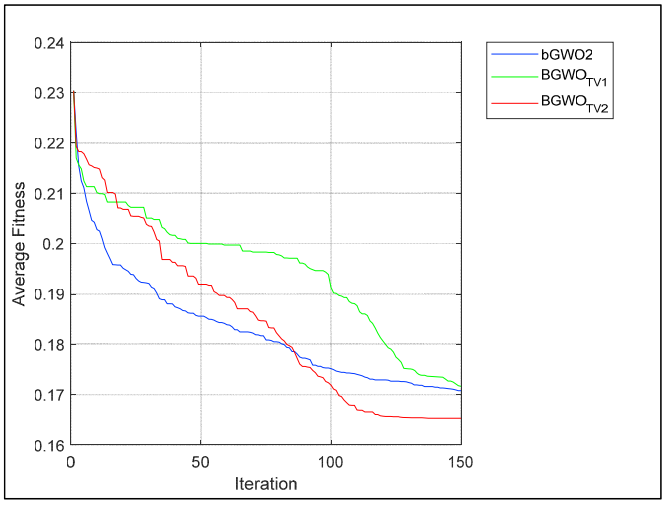



**Fig. 4.** Convergence curves of proposed BGWO$_{TV1}$ and BGWO$_{TV2}$ algorithms with other binary variants of GWO algorithm.

### 3.4. Performance measurement

Various metrics are applied to compare the proposed algorithms that including the average fitness, the standard deviation of fitness, the average accuracy, the standard deviation of accuracy, and the average length of the selected feature subset. The same classifier is utilized to perform the final classification using the optimal descriptors subset selected by all optimization algorithms. Furthermore, this study conducts a non-parametric statistical test, the Wilcoxon signed-rank test [63] with a significant level of 0.05 [64] on the proposed algorithm, and the comparative algorithms to verify whether there are significant differences in their performance. Also, a qualitative assessment is performed by analyzing the convergence behavior of the proposed algorithms with its comparative algorithms based on convergence curves visualized in Figs. 2, 3, 4, 5 and 6.

## 4. Experimental results and discussion

Table 8 present the detailed experimental results. The average fitness value achieved by the proposed algorithms is much lower when

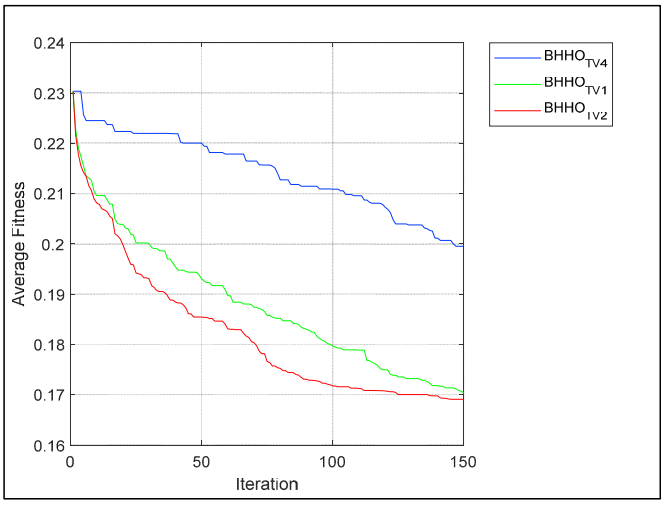



**Fig. 5.** Convergence curves of proposed BHHO$_{TV1}$ and BHHO$_{TV2}$ algorithms with other binary variants of HHO algorithm.

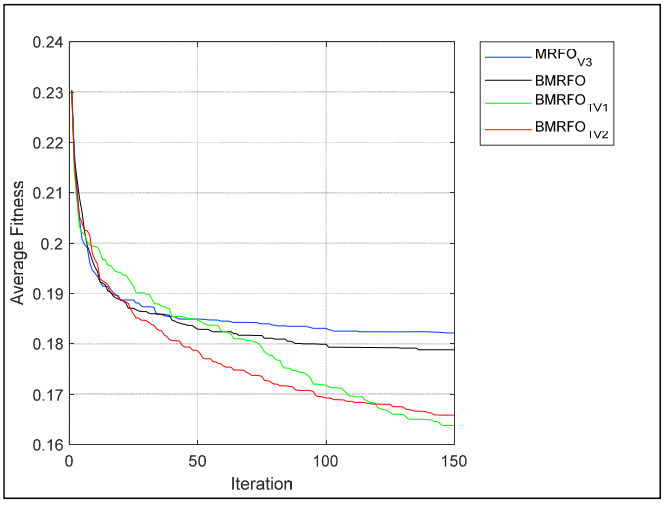



**Fig. 6.** Convergence curves of proposed BMRFO$_{TV1}$ and BMRFO$_{TV2}$ algorithms with other binary variants of MRFO algorithm.

compared with their existing binary version except for BGWO $_{TV1}$. It is confirmed by the convergence curves illustrated in Figs. 2–6. This disclosed the potential of TV2 in improving the convergence of BPSO, BWOA, BGWO, and BHHO. Meanwhile, TV1 is capable to help the BMRFO algorithm to escape from local optima more efficiently. BMRFO$_{TV1}$ is seen to provide the lowest fitness among all algorithms followed by BGWO$_{TV2}$ and BWOA$_{TV2}$.

By observing the combo bar charts in Fig. 7 most of the proposed algorithms have selected fewer descriptors than their existing binary versions compared. BHHO$_{TV4}$ is seen to select the smallest descriptors however does not able to obtain good classification accuracy. The proposed algorithms are shown to be outstanding in reaching a near-optimal solution with a low fitness rate, yielding good classification accuracy with the small size of selected descriptors except for BGWO $_{TV1}$ which achieved lower accuracy than bGWO.

Fig. 8 portrays another remarkable advantage of the proposed approaches in accelerating convergence speed when only consumes a short computational time than the comparative algorithms. This reveals the potential of the proposed techniques in evading stagnation and getting trapped in local optima.

**Table 8**
Performance results of ten proposed algorithms their existing binary variants.

| Algorithms | Avg. Fitness | Avg. Features | Avg. Accuracy | Avg. CT |
|---|---|---|---|---|
| BPSO $_{TV1}$ | 0.17428 (0.00633) | 44 | 82.43 (0.63902) | 222.68 |
| BPSO $_{TV2}$ | 0.17124 (0.00560) | 60 | 82.75 (0.56325) | 320.23 |
| BPSO2 [51,52] | 0.20120 (0.00626) | 583 | 80.17 (0.63161) | 3645.66 |
| BPSO [30] | 0.17851 (0.00876) | 81 | 82.04 (0.87965) | 460.12 |
| BWOA$_{TV1}$ | 0.17170 (0.00985) | 74 | 82.72 (1.00258) | 214.51 |
| BWOA $_{TV2}$ | 0.16664 (0.01164) | 102 | 83.25 (1.18008) | 354.26 |
| BWOA [30,41] | 0.18798 (0.00924) | 436 | 81.38 (0.88842) | 2641.15 |
| BWOA-TV2 [42] | 0.20841 (0.00511) | 720 | 79.56 (0.60099) | 4776.69 |
| BGWO $_{TV1}$ | 0.17164 (0.00804) | 107 | 82.75 (0.83448) | 194.58 |
| BGWO $_{TV2}$ | 0.16532 (0.00792) | 125 | 83.41 (0.81349) | 338.77 |
| bGWO2 [44] | 0.17078 (0.00749) | 250 | 82.96 (0.74250) | 1697.96 |
| BHHO $_{TV1}$ | 0.17036 (0.00730) | 69 | 82.85 (0.73871) | 373.06 |
| BHHO $_{TV2}$ | 0.16912 (0.00564) | 78 | 82.98 (0.57517) | 529.60 |
| BHHO$_{TV4}$ [49] | 0.19952 (0.00774) | 34 | 79.87 (0.78627) | 163.56 |
| BMRFO $_{TV1}$ | 0.16374 (0.00858) | 93 | 83.54 (0.87802) | 830.85 |
| BMRFO $_{TV2}$ | 0.16584 (0.00720) | 97 | 83.33 (0.75863) | 1093.78 |
| MRFOv3 [47] | 0.17883 (0.00599) | 94 | 82.02 (0.59433) | 2050.04 |
| BMRFO [30] | 0.18218 (0.00937) | 149 | 81.72 (0.92299) | 3117.68 |

Fig. 9 visualized and compared the performance between TV1 and TV2 more precisely in terms of accuracy, size of selected descriptors, and computational time. BMRFO achieved better classification accuracy with TV1 while the rest with TV2. TV1 is seen to select fewer descriptors compared to TV2. TV1 also provides faster convergence in all algorithms.

Table 9 outlines the average classification accuracies achieved by three classifiers: k-NN (k = 5) with Euclidian distance, support vector machine (SVM) with Radial Basis Function (RBF), and random forest (RF) with the original dataset. The results indicate that RF is the best classifier after achieving 83.11% classification accuracy. The drawback of the RF classifier is it required a long computation time compared to other classifiers. Overall, BWOA$_{TV2}$, BGWO$_{TV2}$, BMRFO$_{TV1}$, and BMRFO$_{TV2}$ have shown great performance in determining relevant descriptors that successfully increased the classification accuracy and beat the best classification result obtained by the RF classifier. The best performer algorithm is bestowed by BMRFO$_{TV1}$ after achieving 83.54% average classification accuracy with 93 average selected descriptors. It is seen to have improved the classification accuracy of k-NN, SVM(RBF), and RF by 32.84%, 22.12%, and 0.52% by using only 7.85% of the original descriptors. Moreover, the reduced descriptors enhanced the classifier learning and prediction time that only requires 0.32 s.

A Wilcoxon signed-rank test based on the average classification accuracy is conducted to validate whether there is a significant difference between proposed binary algorithms with existing ones. If the *p*-value is greater or equal to 0.05, then the null hypothesis that there is no significant difference is accepted. If the *p*-value is lower than 0.05, then the null hypothesis is rejected. Take note that a *p*-value that is equal to or larger than 0.05 is underlined. By observing Table 10, BPSO$_{TV2}$ shows a

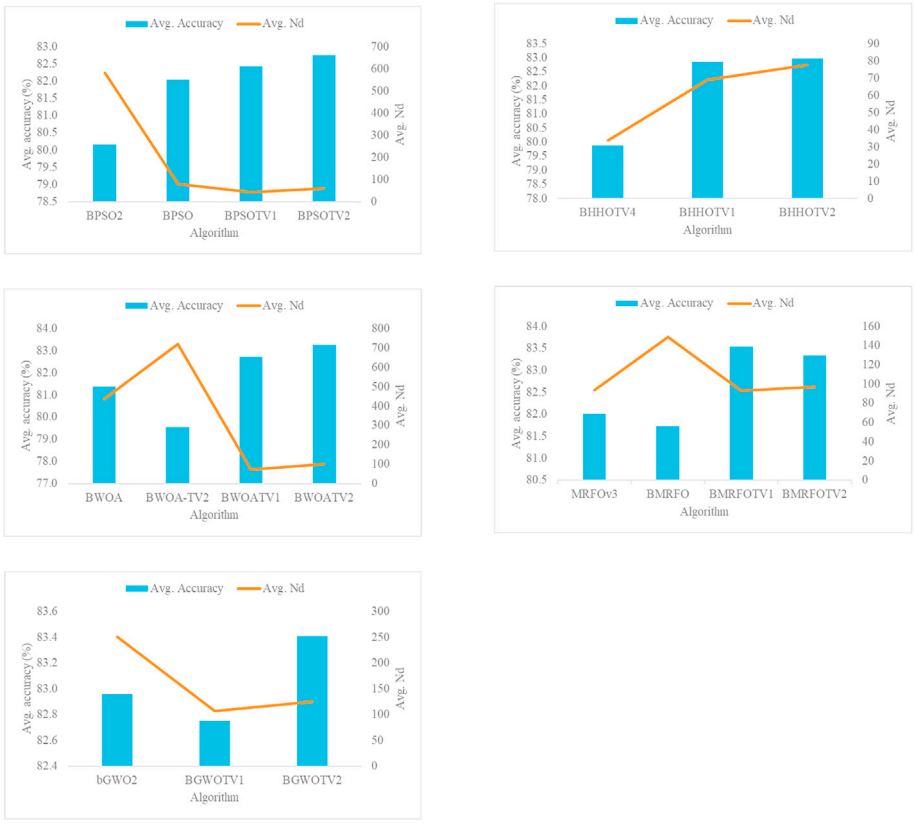



**Fig. 7.** Graphs of the average accuracy versus the average number of selected descriptors.


**Fig. 8.** Comparison results based on average computational time by proposed and comparative algorithms.

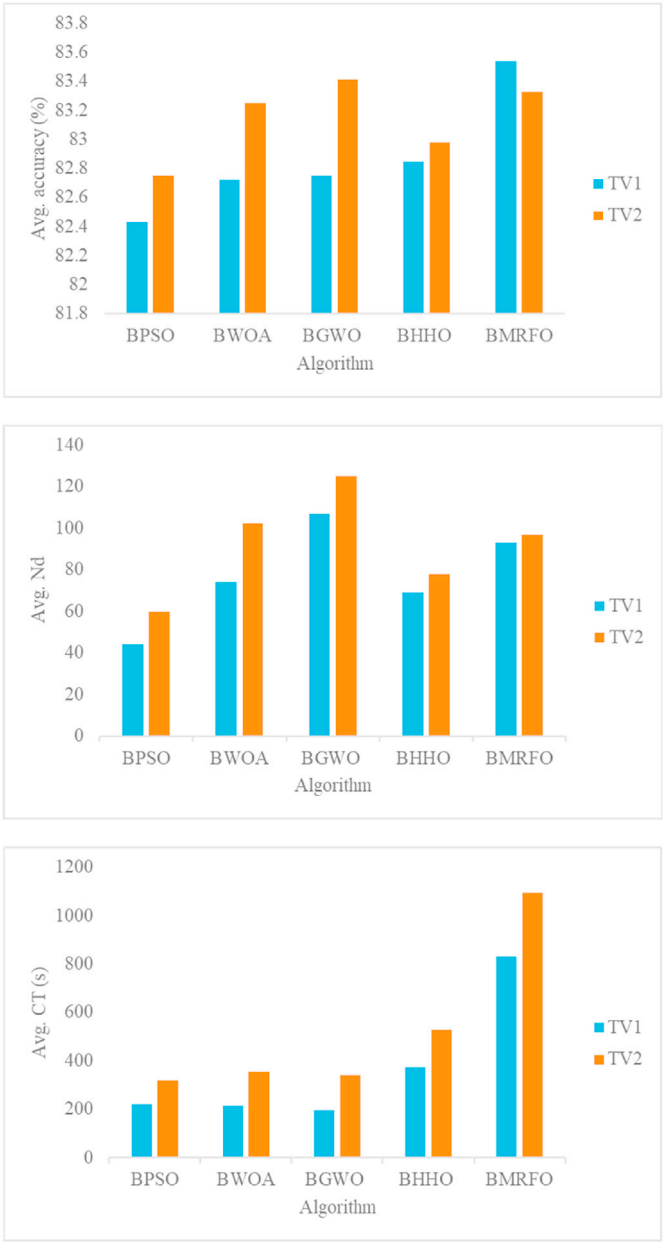



**Fig. 9.** Comparison results of average accuracy and average number of selected features between proposed time-varying binary SI algorithms with different update strategies.

**Table 9**
Classification accuracies before and after implementation of $BMRFO_{TV1}$ feature selection algorithm.

| Dataset (No. of features) | Original (1185 descriptors) | | $BMRFO_{TV1}$ (93 descriptors) | |
|---|---|---|---|---|
| Classifier | Avg. accuracy (%) | Avg. classification time (s) | Avg. accuracy (%) | Avg. classification time (s) |
| k-NN | 62.89 | 3.23 | **83.54** | **0.32** |
| SVM(RBF) | 68.41 | 21.62 | | |
| RF | 83.11 | 111.57 | | |

significant difference between BPSO2 and BPSO, while $BPSO_{TV1}$ shows a significant difference between BPSO2 only with p-values less than 0.05. Conversely, there is no significant difference between $BPSO_{TV1}$ and $BPSO_{TV1}$ with a *p*-value equal to 0.05. Similar results are displayed in

**Table 10**
Wilcoxon signed-rank test result for binary variants of PSO.

| Algorithms | BPSO2 | BPSO | $BPSO_{TV1}$ | $BPSO_{TV2}$ |
|---|---|---|---|---|
| BPSO2 | – | 0.002 | 0.002 | 0.002 |
| BPSO | 0.002 | – | 0.444 | 0.014 |
| $BPSO_{TV1}$ | 0.002 | 0.444 | – | 0.050 |
| $BPSO_{TV2}$ | 0.002 | 0.014 | 0.050 | – |

**Table 11**
Wilcoxon signed-rank test result for binary variants of WOA.

| Algorithms | BWOA | BWOA-TV2 | $BWOA_{TV1}$ | $BWOA_{TV2}$ |
|---|---|---|---|---|
| BWOA | – | 0.006 | 0.032 | 0.006 |
| BWOA-TV2 | 0.006 | – | 0.006 | 0.002 |
| $BWOA_{TV1}$ | 0.032 | 0.006 | – | 0.084 |
| $BWOA_{TV2}$ | 0.006 | 0.002 | 0.084 | – |

**Table 12**
Wilcoxon signed-rank test result for binary variants of GWO.

| Algorithms | bGWO2 | $BGWOA_{TV1}$ | $BGWOA_{TV2}$ |
|---|---|---|---|
| bGWO2 | – | 0.492 | 0.322 |
| $BGWO_{TV1}$ | 0.492 | – | 0.009 |
| $BGWO_{TV2}$ | 0.322 | 0.009 | – |

**Table 13**
Wilcoxon signed-rank test result for binary variants of HHO.

| Algorithms | $BHHO_{TV4}$ | $BHHO_{TV1}$ | $BHHO_{TV2}$ |
|---|---|---|---|
| $BHHO_{TV4}$ | – | 0.006 | 0.006 |
| $BHHO_{TV1}$ | 0.006 | – | 0.386 |
| $BHHO_{TV2}$ | 0.006 | 0.386 | – |

**Table 14**
Wilcoxon signed-rank test result for binary variants of MRFO.

| Algorithms | MRFOv3 | BMRFO | $BMRFO_{TV1}$ | $BMRFO_{TV12}$ |
|---|---|---|---|---|
| MRFOv3 | – | 0.492 | 0.002 | 0.002 |
| BMRFO | 0.002 | – | 0.006 | 0.002 |
| $BMRFO_{TV1}$ | 0.002 | 0.006 | – | 0.386 |
| $BMRFO_{TV2}$ | 0.002 | 0.002 | 0.386 | – |

**Table 15**
P-values of Wilcoxon signed-rank test based on average classification accuracy between $BMRFO_{TV1}$ with other SI algorithms.

| Algorithms | *p*-value |
|---|---|
| bGWO2 | 0.083 |
| $BGWO_{TV1}$ | 0.006 |
| $BGWO_{TV2}$ | 0.444 |
| BPSO2 | 0.002 |
| BPSO | 0.002 |
| $BPSO_{TV1}$ | 0.006 |
| $BPSO_{TV2}$ | 0.004 |
| $BHHO_{TV4}$ | 0.006 |
| $BHHO_{TV1}$ | 0.009 |
| $BHHO_{TV2}$ | 0.126 |
| BWOA | 0.002 |
| BWOA-TV2 | 0.002 |
| $BWOA_{TV1}$ | 0.004 |
| $BWOA_{TV2}$ | 0.286 |
| MRFOv3 | 0.002 |
| BMRFO | 0.002 |
| $BMRFO_{TV2}$ | 0.386 |

Table 11–14 which indicate there are significant differences in the average accuracies between proposed binary algorithms with existing binary algorithms. Table 15 outlines the *p*-values between the best performer algorithm, BMRFO $_{TV1}$ with the other seventeen comparative algorithms in this paper. The results denote that there are significant differences and improvements between BMRFO $_{TV1}$ and twelve algorithms except for bGWO2, $BGWO_{TV2}$, $BHHO_{TV2}$, $BWOA_{TV2}$, $BMRFO_{TV2}$.

## 5. Conclusions and future work

In this paper, the time-varying modified Sigmoid transfer function with two time-varying update strategies are introduced as binarization methods for PSO, GWO, WOA, HHO, and MRFO algorithms. These proposed binary SI algorithms are used in the descriptors selection problem to improve ATS drug classification accuracy. Among all the best proposed binary algorithms, $BMRFO_{TV1}$ was seen to be superior and significant in selecting relevant and discriminative descriptors and obtained high classification accuracy. This study opens research opportunities for other researchers to employ the transfer functions in other existing or new metaheuristic algorithms to solve various binary optimization problems. Extensive performance validation can be conducted on the proposed binary SI algorithms by testing them with benchmarked datasets from the UCI machine learning repository. Some of the improvement that can be considered in the proposed binary algorithms is to employ stronger classifiers as feature evaluator and to hybrid with chaos theory to enhance the ATS drug classification performance.

## 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.

## 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] K. Tamama, Synthetic drugs of abuse, Adv. Clin. Chem. 103 (2021) 191–214, https://doi.org/10.1016/bs.acc.2020.10.001.
[2] L. Harper, J. Powell, E.M. Pijl, An overview of forensic drug testing methods and their suitability for harm reduction point-of-care services, Harm Reduct. J. 14 (2017), https://doi.org/10.1186/s12954-017-0179-5.
[3] E. Lendoiro, C. Jiménez-Morigosa, A. Cruz, et al., An LC-MS/MS methodological approach to the analysis of hair for amphetamine-type-stimulant (ATS) drugs, including selected synthetic cathinones and piperazines, Drug Test. Anal. 9 (2017) 96–105, https://doi.org/10.1002/dta.1948.
[4] H. Chung, S. Choe, Amphetamine-type stimulants in drug testing, Mass Spectrom Lett 10 (2019) 1–10, https://doi.org/10.5478/MSL.2019.10.1.1.
[5] 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.
[6] S. Korkmaz, G. Zararsiz, D. Goksuluk, Drug/nondrug classification using Support Vector Machines with various feature selection strategies, Comput. Methods Progr. Biomed. 117 (2014) 51–60, https://doi.org/10.1016/j.cmpb.2014.08.009.
[7] M.D. Krasowski, S. Ekins, Using cheminformatics to predict cross reactivity of "designer drugs" to their currently available immunoassays, J. Cheminf. 6 (2014) 1–13, https://doi.org/10.1186/1758-2946-6-22.
[8] A. Karim, A. Mishra, M.A.H. Newton, A. Sattar, Efficient toxicity prediction via simple features using shallow neural networks and decision trees, ACS Omega 4 (2019) 1874–1888, https://doi.org/10.1021/acsomega.8b03173.
[9] G. Idakwo, J. Luttrell, M. Chen, et al., A review on machine learning methods for in silico toxicity prediction, J. Environ. Sci. Health Part C Environ. Carcinog. Ecotoxicol. Rev. 36 (2018) 169–191, https://doi.org/10.1080/10590501.2018.1537118.
[10] Y. Wang, J.J. Huang, N. Zhou, et al., Incorporating PLS model information into particle swarm optimization for descriptor selection in QSAR/QSPR, J. Chemom. 29 (2015) 627–636, https://doi.org/10.1002/cem.2746.
[11] Y.-C. Lo, S.E. Rensi, W. Torng, R.B. Altman, Machine learning in chemoinformatics and drug discovery, Drug Discov. Today 23 (2018) 1538–1546, https://doi.org/10.1016/j.drudis.2018.05.010.
[12] S.F. Pratama, Three-Dimensional Exact Legendre Moment Invariants for Amphetamine-type Stimulants Molecular Structure Representation, Universiti Teknikal Malaysia Melaka (UTeM), 2017.
[13] S.F. Pratama, A.K. Muda, Y.H. Choo, A. Abraham, Preparation of ATS drugs 3D molecular structure for 3D moment invariants-based molecular descriptors, in: Advances in Intelligent Systems and Computing, Springer International Publishing, 2018, pp. 252–261.
[14] A. Elsawy, M.M. Selim, M. Sobhy, A hybridised feature selection approach in molecular classification using CSO and GA, Int. J. Comput. Appl. Technol. 59 (2019) 165–174, https://doi.org/10.1504/IJCAT.2019.098034.
[15] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Software 69 (2014) 46–61, https://doi.org/10.1016/j.advengsoft.2013.12.007.
[16] 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.
[17] D.H. Wolpert, W.G. Macready, No free lunch theorems, IEEE Trans. Evol. Comput. 1 (1997) 67–82, https://doi.org/10.1007/978-3-662-62007-6_12.
[18] F.S. Gharehchopogh, I. Maleki, Z.A. Dizaji, Chaotic Vortex Search Algorithm: Metaheuristic Algorithm for Feature Selection, Springer Berlin Heidelberg, 2021.
[19] 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.
[20] A. Hosseinalipour, R. Ghanbarzadeh, A novel approach for spam detection using horse herd optimization algorithm, Neural Comput. Appl. (2022), 0123456789, https://doi.org/10.1007/s00521-022-07148-x.
[21] 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.
[22] N. Nayar, S. Ahuja, S. Jain, Swarm intelligence for feature selection: a review of literature and reflection on future challenges, in: Advances in Data and Information Sciences, Springer, 2019.
[23] M.H. Nadimi-Shahraki, M. Banaie-dezfouli, H. Zamani, S. Taghian, B-MFO : a binary moth-flame optimization for feature selection from medical datasets, Computers 10 (2021) 1–18, https://doi.org/10.3390/computers10110136.
[24] F.S. Gharehchopogh, B. Abdollahzadeh, An Efficient Harris Hawk Optimization Algorithm for Solving the Travelling Salesman Problem. Cluster Comput 5, 2021, https://doi.org/10.1007/s10586-021-03304-5.
[25] M. Mafarja, I. Jaber, S. Ahmed, T. Thaher, Whale Optimisation Algorithm for high-dimensional small-instance feature selection, Int. J. Parallel, Emergent Distributed Syst. 36 (2019) 80–96, https://doi.org/10.1080/17445760.2019.1617866.
[26] S. Taghian, M.H. Nadimi-Shahraki, H. Zamani, Comparative analysis of transfer function-based binary metaheuristic algorithms for feature selection, in: 2018 Int Conf Artif Intell Data Process IDAP vol. 2018, 2019, https://doi.org/10.1109/IDAP.2018.8620828.
[27] S.-S. Guo, J.-S. Wang, M.-W. Guo, Z-shaped transfer functions for binary particle swarm optimization algorithm, Comput. Intell. Neurosci. (2020), https://doi.org/10.1155/2020/6502807, 2020.
[28] 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.
[29] M. Goodarzi, W. Saeys, O. Deeb, et al., Particle swarm optimization and genetic algorithm as feature selection techniques for the QSAR modeling of imidazo[1,5-a]pyrido[3,2-e]pyrazines, inhibitors of phosphodiesterase 10A, Chem. Biol. Drug Des. 82 (2013) 685–696, https://doi.org/10.1111/cbdd.12196.
[30] N. Mohd Yusof, A.K. Muda, S.F. 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.
[31] 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.
[32] 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.
[33] 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) 1–23, https://doi.org/10.1038/s41598-020-71502-z.

[34] 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 vol. 2021, 2021, pp. 381–386.
[35] D. Molina, J. Poyatos, J. Del Ser, et al., Comprehensive taxonomies of nature- and bio-inspired optimization: inspiration versus algorithmic behavior, critical analysis and recommendations, Cognit Comput 12 (2020) 897–939, https://doi.org/10.1007/s12559-020-09730-8.
[36] R. Eberhart, J.K. Sixth, A new optimizer using particle swarm theory, Proc IEEE Symp Micro Mach Hum Sci Nagoys, Japan (1997) 39–43.
[37] 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.
[38] A.A. Heidari, S. Mirjalili, H. Faris, et al., Harris hawks optimization: algorithm and applications, Future Generat. Comput. Syst. 97 (2019) 849–872, https://doi.org/10.1016/j.future.2019.02.028.
[39] W. Zhao, Z. Zhang, L. Wang, Manta ray foraging optimization: an effective bio-inspired optimizer for engineering applications, Eng. Appl. Artif. Intell. 87 (2020), 103300, https://doi.org/10.1016/j.engappai.2019.103300.
[40] 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.
[41] 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.
[42] 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.
[43] T. Thaher, A.A. Heidari, M. Mafarja, et al., Binary harris hawks optimizer for high-dimensional, Low Sample Size Feature Selection (2020) 251–272, https://doi.org/10.1007/978-981-32-9990-0_12.
[44] 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.
[45] 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.
[46] N.M. Yusof, A.K. Muda, S.F. 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.
[47] 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.
[48] A.G. Hussien, D. Oliva, E.H. Houssein, et al., Binary whale optimization algorithm for dimensionality reduction, Mathematics 8 (2020) 1821, https://doi.org/10.3390/math8101821.
[49] 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.
[50] L. Brezočnik, I. Fister, V. Podgorelec, Swarm intelligence algorithms for feature selection: a review, Appl. Sci. 8 (2018), https://doi.org/10.3390/app8091521.
[51] 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.
[52] J. Kennedy, R.C. Eberhart, A Discrete binary version of the particle swarm algorithm, Proc. IEEE Int. Conf. Syst. Man Cybern. 5 (1997) 4104–4108, https://doi.org/10.1109/icsmc.1997.637339.
[53] Z. Beheshti, A time-varying mirrored S-shaped transfer function for binary particle swarm optimization, Inf. Sci. 512 (2020) 1503–1542, https://doi.org/10.1016/j.ins.2019.10.029.
[54] 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.
[55] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Statistician 46 (1992) 175–185, https://doi.org/10.1080/00031305.1992.10475879.
[56] 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.
[57] 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.
[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] B. Turkoglu, S.A. Uymaz, E. Kaya, Binary artificial algae algorithm for feature selection, Appl. Soft Comput. 120 (2022), 108630, https://doi.org/10.1016/j.asoc.2022.108630.
[61] A.A. Abd El-Mageed, A.G. Gad, K.M. Sallam, et al., Improved binary adaptive wind driven optimization algorithm-based dimensionality reduction for supervised classification, Comput. Ind. Eng. 167 (2022), 107904, https://doi.org/10.1016/j.cie.2021.107904.
[62] J. Too, A.R. Abdullah, N.M. Saad, A new quadratic binary harris hawk optimization for feature selection, Electron 8 (2019) 1–27, https://doi.org/10.3390/electronics8101130.
[63] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bull. 1 (1945) 80–83, https://doi.org/10.2307/3001968.
[64] B. Abdollahzadeh, F.S. Gharehchopogh, A multi-objective optimization algorithm for feature selection problems, Eng. Comput. (2021), https://doi.org/10.1007/s00366-021-01369-9.