Improving Amphetamine-Type Stimulants Drug Classification Using Binary Whale Optimization Algorithm as Relevant Descriptors Selection Technique

Norfadzlia Mohd Yusof[1], Azah Kamilah Muda[2(✉)], Satrya Fajri Pratama[2], and Ajith Abraham[3]

[1] Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
norfadzlia@utem.edu.my

[2] Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
{azah,satrya}@utem.edu.my

[3] Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, Auburn, WA, USA
ajith.abraham@ieee.org

Abstract. Swarm intelligence (SI) has become a popular choice to optimize the wrapper feature selection technique. It has attracted this research to employ a binary whale optimization algorithm (BWOA) to solve the molecular descriptors selection problem in ATS drugs classification. This effort is to enhance the learning and prediction ability of the classifier to generate good classification results. S-shaped transfer functions are adopted to generate BWOA, which are then consolidated in the wrapper feature selection with a k-Nearest Neighbor (k-NN) classifier. Our goal is to investigate the influence of different sigmoid transfer functions in BWOA on the selection of significant molecular descriptors and classification accuracy. Several metrics and Wilcoxon's rank-sum test are utilized for performance evaluation. Experimental results reveal that the BWOA-S5 offers performance advantages with the lowest fitness value, fast convergence, high classification accuracy and, small feature subset. Furthermore, the generalization of the optimal molecular descriptor subset is ratified by six different classifiers.

Keywords: Binary whale optimization algorithm · Transfer function · Descriptors selection · Drug classification

1 Introduction

The synthetic drug market has grown rapidly in recent years, with global seizures of Amphetamine-type Stimulants (ATS) increasing by more than fourfold from 60 tonnes in 2008 to 261 tonnes in 2017 [1]. As been reported in the "2019 World Drug Report" by the United Nations Office on Drugs and Crime (UNODC), there were about 271 million

The original version of this chapter was revised: The misspelt co-author name "Satrya Fajr Pratama" has been changed to "Satrya Fajri Pratama". The correction to this chapter is available at https://doi.org/10.1007/978-3-030-96302-6_65

A. Abraham et al. (Eds.): SoCPaR 2021, LNNS 417, pp. 424–432, 2022.
https://doi.org/10.1007/978-3-030-96302-6_39

illicit drug users in 2016 between the ages of 15 and 64 years [2]. ATS includes a range of synthetic psychostimulants, including methamphetamine, amphetamine, and ecstasy that are widely abused. They are easy to produce, cheap to buy, and hard to control.

UNODC has recommended methods for identifying and analyzing ATS drugs [3]. Each of the methods has its pros and cons that must be considered. Some of the downsides of existing analytical methods involve expensive facilities that entail well-trained skilled technicians, complex testing processes, lengthy running time, not updated analytical methods (libraries), and inconsistent results from different test kits [4–6]. Moreover, the presents of new ATS drugs on the illegal drug market provides a steep challenge for analytical toxicologists.

Feature selection is generally classified into filter and wrapper [7]. The filter approach is based on independent feature assessments, using essential data properties in identifying how significantly particular features are [8]. It is computationally efficient than the wrapper approach. In contrast, the wrapper approach is highly dependent on the performance of the machine learning classifier to find for and select the relevant features that results in computational overhead [9]. However, the wrapper approach produces more accurate results in terms of classification than the filter approach but suffers from being computational costly [9, 10]. This research considered the wrapper feature selection for the optimal selection of molecular descriptors in the ATS drug classification problem.

Feature selection is NP-hard computational problem. It able be a very complex and computationally intensive task, particularly with large datasets. [11, 12]. Swarm intelligence (SI) one of the branches of metaheuristic algorithms is established to be an efficient solution to resolve feature selection problems in several applications [13, 14]. One of the applications is as a molecular descriptors selection approach in the cheminformatics domain whereby these SI algorithms have been successfully employed: grasshopper optimization algorithm (GOA) [15], firefly algorithm (FA) [16], and salps swarm algorithm (SSA) [17].

The No Free Lunch theorem [18] states that none algorithm can solve all optimization problems, which drives our attempts to employ WOA [34] as wrapper feature selection in ATS drug classification. WOA is widely used in many fields because of its simple structure, less parameter for tuning, ease of implementation, and good performance [19, 20].

This work may become an alternative approach for resolving the problem in current ATS drug analysis and identification methods. Specific molecular descriptors called 3D Exact Legendre Moment Invariants (3D-ELMI) [21] that were formulated to represent the 3D molecular structure of ATS drugs are employed. It contains 1185 features to describe the drug instance. As we know, high-dimensional descriptors can degrade the ATS drug classification results. For that reason, BWOA algorithms are utilized in this study for the selection of important descriptors. In this research, five S-shaped transfer functions were used to transform native WOA to binary WOA.

The principal focus of this study is to examine the effect of various S-shaped transfer functions on BWOA that selects a less number of molecular descriptors while achieving the same or greater classification accuracy when employing all features. Next, the comparative analysis is conducted between BWOA and native WOA according to the selected performance metrics and statistical analysis.

Section 2 introduces the binary whale optimization algorithm (BWOA) concepts. Section 3 explains the 3D Exact Legendre Moment Invariants (3D-ELMI) molecular descriptors as the representation of ATS drugs molecular structure. Section 4 reveals the necessary material and methods used in the study. Section 5 discusses the empirical results for native WOA and five proposed BWOAs. Lastly, Sect. 6 concludes with some suggestions for future research.

2 Binary Whale Optimization Algorithm (BWOA)

The solutions (feature subsets) for feature selection problems are formed of binary which persuades the custom WOA to the binary version of WOA [14]. Similar to WOA, the position of search agents (whales) in the binary whale optimization algorithm (BWOA) is continually updated to any location in the search space by following the best search agent discovered so far. Then search agents' real position is converted to binary values. The conversion approach applied in this study is the transfer function. The probability definitions by this approach force search agents or whales to travel in a binary space by updating each position (element) in the search agent to 1 (relevant/selected feature) and 0 (irrelevant/not selected feature) [22] to generate the solution.

2.1 Transfer Functions

There are various families of transfer functions employed by researchers in the feature selection domain such as S-shaped [23], V-shaped [24], and quadratic[25] transfer functions. Among the SI algorithms that have successfully employed this approach is binary particle swarm optimization [22], binary grasshopper algorithm [26], binary grey wolf optimization algorithm [27], and binary harris hawk optimization algorithm [25]. This study employed five S-shaped transfer functions in [23, 28] to generate BWOA and the mathematical formulations are shown in Table 1. In sigmoid functions the position is updated based on Eq. 1 proposed by Kennedy and Eberhart in [29]:

$$x(t+1) = \begin{cases} 1, & \textit{if } rand < T(x(t+1)) \\ 0, & \textit{otherwise} \end{cases} \tag{1}$$

rand is a random number, $rand \in \{0, 1\}$ and $x(t+1)$ is the new whale's position.

2.2 Fitness Function

There are two objectives to achieve by feature selection which are maximizing the classification accuracy and minimizing the selected feature size [30]. The fitness function in Eq. 2 is used in the feature selection problem which combines the two objectives into a single objective problem. It is also designed to have a balance between the two objectives. The fitness value is determined from each search agent's solution over iterations. The optimal feature subset is the one with the least fitness value. Since the wrapper feature

Table 1. The utilized S-shaped transfer functions.

Identifier	Transfer function
S1	$T(x) = \frac{1}{1+e^{-x}}$
S2	$T(x) = \frac{1}{1+e^{-2x}}$
S3	$T(x) = \frac{1}{1+e^{-x/2}}$
S4	$T(x) = \frac{1}{1+e^{-x/3}}$
S5	$T(x) = \frac{1}{1+e^{-10(x-0.5)}}$

selection technique is used, a learning algorithm (classifier) is involved in evaluating the selected feature subset.

$$\downarrow Fitness = \alpha \cdot E + \beta \frac{|SF|}{|F|} \tag{2}$$

where α, and β are constant parameters which manage the trade-off between classification accuracy and subset length respectively. $\alpha \in (1, 0)$, and $\beta = (1 - \alpha)$. E indicates the classification error rate. $|SF|$ represents the selected feature size, and $|F|$ is the original feature size in the dataset. Since classification performance is our primary goal, α is assigned to 0.99 [31] to be the most important metric.

3 3D-Exact Legendre Moment Invariants (3D-ELMI) Molecular Descriptors Dataset

3D-ELMI molecular descriptors were introduced in 2017 by Pratama with the motivation of implementing an image processing technique, Moment Invariants (MI) [21]. The molecular descriptors were calculated on 7190 drug instances with an equal number of ATS drug compounds (pihkal.info database) and non-ATS drug compounds (ChemSpider database). The explanation regarding the preparation of the 3D molecular descriptors can be attained in Pratama et al. [32]. The 3D-ELMI molecular descriptors produce fixed attributes consisting of 1186 feature attributes and 1 class attribute with a binary label for each drug compound as outlined in Table 2.

4 Method

A binary whale-based wrapper feature selection algorithm is employed in this research. In this algorithm, the k-NN classifier with the Euclidean distance matric is utilized (where k = 5 [33]) as the feature evaluator in the wrapper method. The k-NN classifier was chosen because it has a high processing speed and can generally generate good results. BWOA algorithm works as a feature search and selector to optimize the wrapper method. For experiments, the hold-out validation strategy with a stratified random sampling is

Table 2. Attributes description.

Attribute	Attribute type	No. of attribute	Description
Molecule id	String	1	Molecule id refers to the reference id of the drug molecule in the original database
Feature*n*	Numeric (real numbers)	1185	Calculated 3D-ELMI value. *n* is feature number from 1 to 1185
Class	Nominal {0, 1}	1	0: non-ATS and 1: ATS

applied where 80% of the descriptors are for training and the residual 20% is for testing purposes. Somewhat, the Molecule id attribute is excluded during the feature selection process. All experimental works are repeated ten times (different random seeds) to obtain statistically meaningful results. The final results are viewed as the average of evaluation metrics [17, 27] that were recorded from the testing data in each run. All algorithms are implemented in Matlab R2019b that run on the Windows 10 platform in the Intel Core i7-6700 machine, 3.40 GHz CPU, and 16 GB of RAM.

5 Results and Discussion

Analysis of the experimentation results undertaken by us is presented in this section. Note that, the best results from each approach are emphasized in bolded text. Table 3 presents the worst fitness, best fitness, average fitness, standard deviation achieved, and average computation time over ten different runs of native WOA and five BWOA algorithms. As can be observed, BWOA-S5 scored the minimal worst, best, average fitness values among other algorithms. This exhibits that BWOA-S5 is most capable of selecting significant features. It is shown that all algorithms have achieved low standard deviation values that imply consistent fitness values were obtained by every algorithm in each run. BWOA-S4 is shown to be more stable and produce consistent results with the lowest standard deviation of 0.00589. The convergence curves of these six algorithms are presented in Fig. 1 to more intuitively reveal and compare the optimization precision and convergence rate of each algorithm. Based on the convergence curve and the average computational time in Table 3, we acknowledged that BWOA-S5 converged faster and deeper to seek out the optimal solution. This reveals that the transfer function affects the balance between exploitation and exploration in the native WOA algorithm and highlights that BWOA-S5 has improved the convergence speed of the native WOA.

Table 4 summarizes the average results of accuracy, size of selected descriptors, and classification time taken after the implementation of feature selection techniques. The obtained results specify that the classification accuracy has increased between 25.63% to 29.54% with WOA-based feature selection techniques. On the other hand, the number of selected descriptors were decreased to 27.75% (WOA), 52.62% (BWOA-S1), 56.66% (BWOA-S2), 53.39% (BWOA-S3), 48.13% (BWOA-S4), and 22.93% (BWOA-S5) from the original size. Additional advantages gained from the small descriptors are the reduction of more than 47% of classification time taken by the k-NN classifier to learn

and make a prediction. As a result, when evaluating the most informative descriptors subset, BWOA-S5 contributed to the optimal performance with the highest classification accuracy and the lowest number of selected descriptors which outperformed WOA, BWOA-S1, BWOA-S2, BWOA-S4, and BWOA-S5.

The statistical analysis results of the Wilcoxon signed-rank test for all possible pairs of the algorithms are presented in Table 5. The results describe whether the difference in classification accuracies between the two respective algorithms are statistically significant or not. If the computed *p*-value is less than 0.05, 1 is stated in the table, indicating there exists a significant difference in the two algorithms' performance. Otherwise, 0 is stated for not significant. It is clearly shown that the BWOA-S5 attained 1 for all *p*-values, indicating that its superiority is statistically significant compared to the other compared algorithms in the pair-wise Wilcoxon signed-ranks test.

Table 3. Results of best, worst, average, and standard deviation (Std) of fitness values and Average Computational Time (ACT) in seconds obtained by WOA and five proposed BWOA algorithms.

Algorithm	Worst	Best	Average	Std	ACT
WOA	0.22389	0.19585	0.20917	0.00783	493.94
BWOA-S1	0.22229	0.19998	0.21249	0.00609	1060.68
BWOA-S2	0.23009	0.20086	0.21344	0.00826	1211.66
BWOA-S3	0.22105	0.19340	0.21318	0.00804	983.52
BWOA-S4	0.21864	0.19733	0.21200	**0.00589**	943.36
BWOA-S5	**0.20403**	**0.17095**	**0.18570**	0.01170	**392.20**

Table 4. Averaged results of classification accuracy, descriptors size, and classification time in seconds for No Feature Selection (NFS), WOA, and BWOA algorithms.

Algorithm	Accuracy (%)	Descriptors size	Classification time
NFS	62.89	1185	3.23
WOA	79.15	328.90	0.85
BWOA-S1	79.07	623.60	1.58
BWOA-S2	79.01	671.40	1.60
BWOA-S3	79.01	632.70	1.71
BWOA-S4	79.11	614.70	1.65
BWOA-S5	**81.47**	**271.70**	**0.74**

Table 5. *P*-values of the Wilcoxon signed-rank test of classification accuracy results.

	WOA	BWOA-S1	BWOA-S2	BWOA-S3	BWOA-S4	BWOA-S5
WOA	N/A	0	**1**	0	0	**1**
BWOA-S1	0	N/A	0	0	0	**1**
BWOA-S2	**1**	0	N/A	0	0	**1**
BWOA-S3	0	0	0	N/A	0	**1**
BWOA-S4	0	0	0	0	N/A	**1**
BWOA-S5	**1**	**1**	**1**	**1**	**1**	N/A

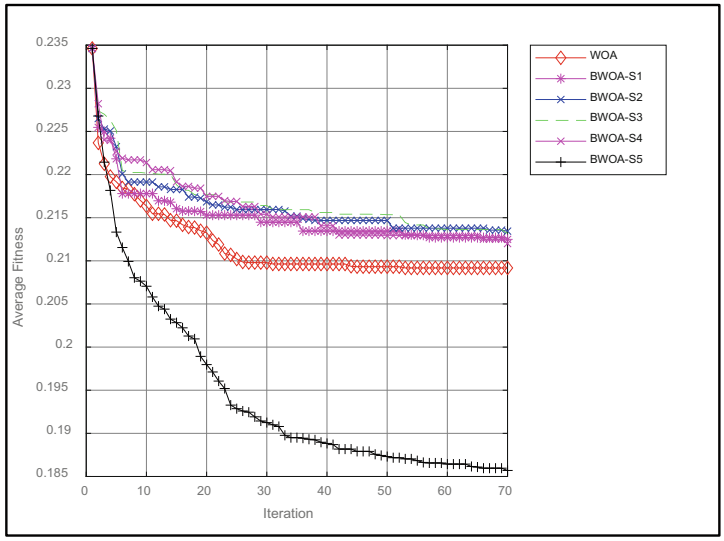

Fig. 1. Convergence curves of the average fitness of WOA and proposed five BWOA algorithms.

6 Conclusions and Future Works

This research has evidenced the benefit of transfer function implementation in BWOA to optimize the wrapper feature selection technique. The selection of transfer function is a crucial process to improve the native WOA. The performance measurement and statistical analysis signify that BWOA-S5 is highly proficient to fulfill the two feature selection objectives: attained the lowest and informative molecular descriptors and increase ATS drug classification accuracy. Further experimental investigations are needed to optimize dominating parameters in WOA: the vector a interval values, maximum iterations, and the number of search agents. Experimenting with other families of a transfer function is another subject to explore. Moreover, the application of other classifiers as a features evaluator in the wrapper feature selection technique will be considered as an extension for future work.

Acknowledgments. The authors would like to acknowledge 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. United Nations Office on Drugs and Crime: The ATS market - 10 years after the 2009 Plan of Action (2019)
2. United Nations Office on Drugs and Crime (UNODC): World Drug Report 2019. United Nations publication, Sales No. E.19.XI.8 (2019)
3. United Nations Office on Drugs and Crimes: Recommended Methods for the Identification and Analysis of Amphetamine, Methamphetamine and their ring-substituted analogues in seized materials (2006)
4. Liu, L., et al.: Newly emerging drugs of abuse and their detection methods: an ACLPS critical review. Am. J. Clin. Pathol. **149**, 105–116 (2018)
5. Chung, H., Choe, S.: Amphetamine-type stimulants in drug testing. Mass Spectrom. Lett. **10**, 1–10 (2019)
6. Alsenedi, K.A., Morrison, C.: Determination of amphetamine-type stimulants (ATSs) and synthetic cathinones in urine using solid phase micro-extraction fibre tips and gas chromatography-mass spectrometry (2018). https://doi.org/10.1039/c8ay00041g
7. Venkatesh, B., Anuradha, J.: A review of feature selection and its methods. Cybern. Inf. Technol. **19**, 3–26 (2019)
8. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. **40**, 16–28 (2014)
9. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artif. Intell. **97**, 273–324 (1997)
10. Liu, H., Zhao, Z.: Manipulating data and dimension reduction methods: feature selection. Comput. Complex Theory Tech. Appl. **9781461418**, 1790–1800 (2012)
11. Li, J., et al.: Feature selection: a data perspective. ACM Comput. Surv. (2017). https://doi.org/10.1145/3136625
12. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A.: Feature selection for high-dimensional data. Progr. Artif. Intell. **5**(2), 65–75 (2016). https://doi.org/10.1007/s13748-015-0080-y
13. Saw, T., Hnin, P.: Swarm intelligence based feature selection for high dimensional classification: a literature survey. Int. J. Comput. **33**, 69–83 (2019)
14. Brezočnik, L., Fister, I., Podgorelec, V.: Swarm intelligence algorithms for feature selection: a review. Appl Sci. (2018). https://doi.org/10.3390/app8091521
15. Algamal, Z.Y., Qasim, M.K., Lee, M.H., Ali, H.T.M.: QSAR model for predicting neuraminidase inhibitors of influenza A viruses (H1N1) based on adaptive grasshopper optimization algorithm. SAR QSAR Environ. Res. **31**, 803–814 (2020)
16. Fouad, M.A., Tolba, E.H., El-Shal, M.A., El Kerdawy, A.M.: QSRR modeling for the chromatographic retention behavior of some β-lactam antibiotics using forward and firefly variable selection algorithms coupled with multiple linear regression. J. Chromatogr. A **1549**, 51–62 (2018)
17. Hussien, A.G., Hassanien, A.E., Houssein, E.H.: Swarming behaviour of salps algorithm for predicting chemical compound activities. In: 2017 IEEE 8th International Conference on Intelligent Computing and Information Systems ICICIS 2017, vol 2018-Janua, pp. 315–320 (2017)
18. Wolpert, D.H., Macready, W.G.: No free lunch theorems. IEEE Trans. Evol. Comput. **1**, 67–82 (1997)

19. Mohammed, H.M., Umar, S.U., Rashid, T.A.: A systematic and meta-analysis survey of whale optimization algorithm. Comput. Intell. Neurosci. **2019**, 1–25 (2019)
20. Rana, N., Latiff, M.S.A., Abdulhamid, S.M., Chiroma, H.: Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments. Neural Comput. Appl. **32**(20), 16245–16277 (2020). https://doi.org/10.1007/s00521-020-04849-z
21. Pratama, S.F.: Three-dimensional exact legendre moment invariants for amphetamine-type stimulants molecular structure representation. Universiti Teknikal Malaysia Melaka (UTeM) (2017)
22. Mirjalili, S., Lewis, A.: S-shaped versus V-shaped transfer functions for binary particle swarm optimization. Swarm Evol. Comput. **9**, 1–14 (2013)
23. Hussien, A.G., Hassanien, A.E., Houssein, E.H., Bhattacharyya, S., Amin, M.: S-shaped binary whale optimization algorithm for feature selection. In: Bhattacharyya, S., Mukherjee, A., Bhaumik, H., Das, S., Yoshida, K. (eds.) Recent Trends in Signal and Image Processing. AISC, vol. 727, pp. 79–87. Springer, Singapore (2019). https://doi.org/10.1007/978-981-10-8863-6_9
24. Hussien, A.G., Houssein, E.H., Hassanien, A.E.: A binary whale optimization algorithm with hyperbolic tangent fitness function for feature selection. In: 2017 IEEE 8th International Conference on Intelligent Computing and Information Systems, ICICIS 2017, vol. 2018-Janua, pp. 166–172 (2017)
25. Too, J., Abdullah, A.R., Saad, N.M.: A new quadratic binary Harris hawk optimization for feature selection. Electron **8**, 1–27 (2019)
26. Mafarja, M., Aljarah, I., Faris, H., Hammouri, A.I., Al-Zoubi, A.M., Mirjalili, S.: Binary grasshopper optimisation algorithm approaches for feature selection problems. Expert Syst. Appl. **117**, 267–286 (2019)
27. Emary, E., Zawbaa, H.M., Hassanien, A.E.: Binary grey wolf optimization approaches for feature selection. Neurocomputing **172**, 371–381 (2016)
28. Eid, H.F.: Binary whale optimisation: an effective swarm algorithm for feature selection. Int. J. Metaheurist. **7**, 67 (2018)
29. Kennedy, J., Eberhart, R.C.: A discrete binary version of the particle swarm algorithm. In: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, vol. 5, pp. 4104–4108 (1997)
30. Mafarja, M., Mirjalili, S.: Whale optimization approaches for wrapper feature selection. Appl. Soft Comput. J. **62**, 441–453 (2018)
31. Mafarja, M., et al.: Binary dragonfly optimization for feature selection using time-varying transfer functions. Knowl.-Based Syst. **161**, 185–204 (2018)
32. Pratama, S.F., Muda, A.K., Choo, Y.H., Abraham, A.: Preparation of ATS drugs 3D molecular structure for 3D moment invariants-based molecular descriptors. In: Abraham, A., Muhuri, P., Muda, A., Gandhi, N. (eds.) HIS 2017. Advances in Intelligent Systems and Computing, vol. 734, pp. 252–261. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76351-4_26
33. Mafarja, M., Jaber, I., Ahmed, S., Thaher, T.: Whale optimisation algorithm for high-dimensional small-instance feature selection. Int J. Parallel Emergent Distrib. Syst. **36**, 80–96 (2019)
34. Mirjalili, S., Lewis, A.: The whale optimization algorithm. Adv. Eng. Softw. **95**, 51–67 (2016)