Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

High-dimensional QSAR classification modeling based on improving black hole algorithm

Hiba Sulaiman Dawood
Department of Statistics and Informatics, University of Mosul, Mosul, Iraq
E-mail: heba.sulaiman82@uomosul.edu.iq
Ala'a A. Hammodat
Department of Statistics and Informatics, University of Mosul, Mosul, Iraq
E-mail: allahamoodat@uomosul.edu.iq
Zakariya Yahya Algamal*
Department of Statistics and Informatics, University of Mosul, Mosul, Iraq
E-mail: zakariya.algamal@uomosul.edu.iq
ORCID: 0000-0002-0229-7958
*Corresponding Author
Zakariya Yahya Algamal
E-mail: zakariya.algamal@uomosul.edu.iq
Telephone number: +964 7701640834

Article Info
Page Number: 279-290
Publication Issue:
Vol 71 No. 3s2 (2022)

Article History
Article Received: 28 April 2022
Revised: 15 May 2022
Accepted: 20 June 2022
Publication: 21 July 2022

Abstract
High-dimensionality is one of the major problems which affect the quality of the quantitative structure-activity (property) relationship (QSAR/ QSPR) classification methods in chemometrics. Applying variable selection is essential to improve the performance of the classification task. Variable selection is well-known as an NP-hard optimization problem. Various evolutionary algorithms are dedicated to solving this problem in the literature. Recently, a black hole algorithm was proposed, which has been successfully applied to solve various continuous optimization problems. In this paper, a new time-varying transfer function is proposed to improve the exploration and exploitation capability of the binary black hole algorithm in selecting the most relevant descriptors (variables) in QSAR/ QSPR classification models with high classification accuracy and short computing time. Based on seven benchmark biopharmaceutical datasets, the experimental results reveal the capability of the proposed time-varying transfer function to achieve high classification accuracy with minimizing the number of selected descriptors and reducing the computational time.
Keywords: QSAR, black hole algorithm, evolutionary algorithm, transfer function, descriptors selection.

1. Introduction

In chemometrics, the quantitative structure-activity (property) relationship (QSAR/ QSPR) is a powerful and a promising model used to better understand the structural relationship between the chemical activity (property) and the chemical compounds by explicitly considering the mathematical, statistical, and informatical methods [1]. A common task in these models is the

selection of relevant descriptors (variables), where researchers try to determine the smallest possible set of descriptors that can still achieve good predictive performance [2-15]. A typical data in QSAR/ QSPR modeling consist of a small sample size of compounds (molecules) and a very large number of descriptors. Consequently, QSAR/ QSPR modeling is challenged by the high dimensionality of the descriptors.

In chemometrics, today, it is easily come out with thousands of molecular descriptors, such as Dragon 7, which is a commercial software. It can calculate 5270 molecular descriptors [16, 17]. In high dimensional QSAR/ QSPR modeling, where the number of descriptors, p, exceeds the number of compounds, n, the traditional statistical classification methods are not feasible [5, 18]. In addition, the large number of descriptors can degrade the generalizable performance of the used classifier or the prediction performance. Therefore, selecting descriptors that truly affect the biological activity is an attractive way in QSAR/ QSPR modeling [19].

Variable (Descriptor) selections can be reported as a non-polynomial (NP) hard problem. The objective of variable selection is to provide faster and more effective models, and also to avoid overfitting and the curse of dimensionality. Variable selection is a typical combinatorial optimization problem. A considerable effort has been devoted to developing variable selection procedures. With the development of computational intelligence, evolutionary algorithms, such as particle swarm optimization (PSO) [20], bat algorithm (BA) [21], and grey wolf optimization (GWO) [22], are the most effective and core technology to address high-dimensional data.

The black hole algorithm (BHA), which was proposed by Hatamlou [23], has certain outstanding merits, such as a simple computational process, simple implementation, and easy understanding with only a few parameters for tuning. Due to its good properties, BHA has become a useful tool for many real-world problems [23-31]. The BHA is inspired by the black hole phenomenon in the space. In the case of variable selection, the search space is modeled as an n-dimensional Boolean lattice, in which the selected variable is coded as 1 and the not selected variable is coded as 0. Therefore, a binary version of the BHA was proposed. The efficiency of the binary black hole algorithm (BBHA) is depending on the transfer function which is responsible to map a continuous search space to a discrete search space.

In this study, a new time-varying transfer function is proposed to improve the exploration and exploitation capability of the BBHA in selecting the most relevant descriptors in QSAR/ QSPR classification models with high classification accuracy and short computing time.

The rest of the paper is organized as follows: The explanation of the black hole algorithm and the proposed time-varying transfer function are given in Section 2. In section 3, the experimental setting is covered. Section 4, the results are summarized with their discussion. Finally, Section 5 contains a conclusion of this work.

2. Methodology

2.1 Variable selection

The variable selection method is a procedure that reduces or minimizes the number of variables and selects some subsets of original variables. Selecting the most relevant variables is one of the challenging tasks for large dataset. The variable selection has been proven to effectively remove irrelevant and redundant variables. In addition, it can improve the performance of classifiers and

reduce the computational time. Variable selection using evolutionary algorithm has been applied successfully in QSAR/QSPR classification [14, 15, 20, 32-39].

Let S be a dataset of n observations with D variables and A be the set of all the D variables. Variable selection is to choose d variables from D, where $d \le D$, so that the objective function, $f(\cdot)$, is maximized or minimized subject to the problem. In this way, variable selection becomes a combinatorial (discrete) optimization problem where the objective is to find the best variable subset $|B| = d$. By maximize the classification accuracy, the optimization problem is defined as

$$\begin{aligned} &\max\ f(X) \\ &\text{s.t. } X = (x_1, x_2, ..., x_d),\ x_j \in \{0,1\},\ j = 1,2,...,d \end{aligned} \tag{1}$$

2.2 Binary black hole algorithm

The black hole algorithm is one of the most recent evolutionary algorithms inspired from the black hole phenomenon in the space. This algorithm was introduced in 2013 by Hatamlou [23]. A black hole in space is what forms when a star of massive size collapses. The gravitational power of the black hole is too high that even the light cannot escape from it [24, 25, 27].

In BHA, the process begins by the initialization of the stars, which are acting as the population, in the search space. The initial population is randomly generated. The best objective function of an individual star is selected as the black hole and it starts absorbing populations (stars) surround it. Then, all the stars move towards the black hole. This movement can be formulated as follows

$$x_i^{t+1} = x_i^t + \theta \times (x_{BH} - x_i^t), \quad i = 1,2,...n_s, \tag{2}$$

where x_i^t and x_i^{t+1} are the locations of the ith star at iterations t and $t+1$, respectively. x_{BH} is the location of the black hole in the search space, θ is a random number in the interval $[0,1]$, and n_s is the total number of stars (candidate solutions in the search space). It is important to mention that, the black hole does not move, because it has the best objective value and then attracts all other stars [23, 26, 30, 31].

After determining the movement of the stars using Eq. (2), if the objective function value of a star is better than the value of the black hole, the star is then selected as the black hole. During moving stars towards the black hole, there is the possibility of crossing the event horizon (border of the black hole).

The radius of the event horizon (Schwarzschild radius) in the black hole algorithm is calculated by

$$R = \frac{f_{BH}}{\sum_{i=1}^{n_s} f_i}, \tag{3}$$

where f_{BH} and f_i is the objective function value of the black hole and the ith star, respectively. When the distance between a candidate solution and the black hole is less than Eq. (3), that candidate is collapsed and a new candidate is created and distributed randomly in the search space.

To perform the variable selection, a binary black hole algorithm was proposed [28-30]. Unlike the standard BHA, in which the solutions are updated in the search space towards continuous-valued

positions, in the BBHA, the search space is modeled as an n-dimensional Boolean lattice and the solutions are updated across the corners of a hypercube. In addition, as the problem is to select or not a given variable, a solution binary vector is employed, where 1 corresponds whether a variable will be selected to compose the new dataset, and 0 otherwise.
In any binary algorithm, where one uses the step vector to calculate the probability of changing positions, the transfer functions significantly impact the balance between exploration and exploitation [40, 41].

2.3 The proposed time-varying transfer function

In BBHA, the transfer function is used to map a continuous search space to a binary one, and the updating process is designed to switch positions of stars between 0 and 1 in binary search spaces. In order to build this binary vector, a transfer function in Eq. (4) can be used, in which the new solution is constrained to only binary values

$$x_i^t = \begin{cases} 1 & \text{if} \quad T\left(x\right) > r \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $r \in \left[0,1\right]$ is a random number, $T\left(x\right)$ is the transfer function. Two familiar transfer functions are usually used, namely the sigmoid transfer function (SIG), which belongs to the S-shaped family [42], and the inverse tangent hyperbolic transfer function (TH), which is belonging to the V-shaped family [42]. These two transfer functions are, respectively, defined as:

$$T_{\text{SIG}}\left(x_i^t\right) = \frac{1}{1+e^{-x_i^t}}, \tag{5}$$

and

$$T_{\text{TH}}\left(x_i^t\right) = \left|\frac{2}{\pi}\arctan\left(\frac{\pi}{2}x_i^t\right)\right|. \tag{6}$$

In optimization algorithm, it is expected that the focus of the early stages of the implementation the algorithm will be on exploration to avoid falling into the local point, but in later stages of implementation, the algorithm focuses more on exploitation to improve the quality of the solution [40, 41].

As in Mafarja, Aljarah, Heidari, Faris, Fournier-Viger, Li and Mirjalili [41] and Islam, Li and Mei [40], in this paper, a dynamic transfer function is proposed to improve the BBHA. In our proposed time-varying transfer function, (TV), a new control parameter λ is added in the original transfer function. This λ is a time-varying variable which starts with a large value and gradually decreases over time. The proposed λ is defined as

$$\lambda = \lambda_{\min} + (\lambda_{\max} - \lambda_{\min})\,\mathrm{e}^{-t}, \tag{7}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are, respectively, the minimum and maximum values of the control parameter λ. Accordingly, the two proposed transfer functions are defined as, respectively,

$$T_{\text{TV-SIG}}\left(x_i^t\right)=\frac{1}{1+e^{-\frac{x_i^t}{\lambda}}}, \tag{8}$$

and

$$T_{\text{TV-TH}}\left(x_i^t\right)=\left|\frac{2}{\pi}\arctan\left(\frac{\pi}{2}\frac{x_i^t}{\lambda}\right)\right|. \tag{9}$$

Figure 1 explains the behavior of the proposed time-varying transfer function for both SIG and TV, respectively. It is obvious that these proposed functions coverage to be a vertical line when iteration increasing.

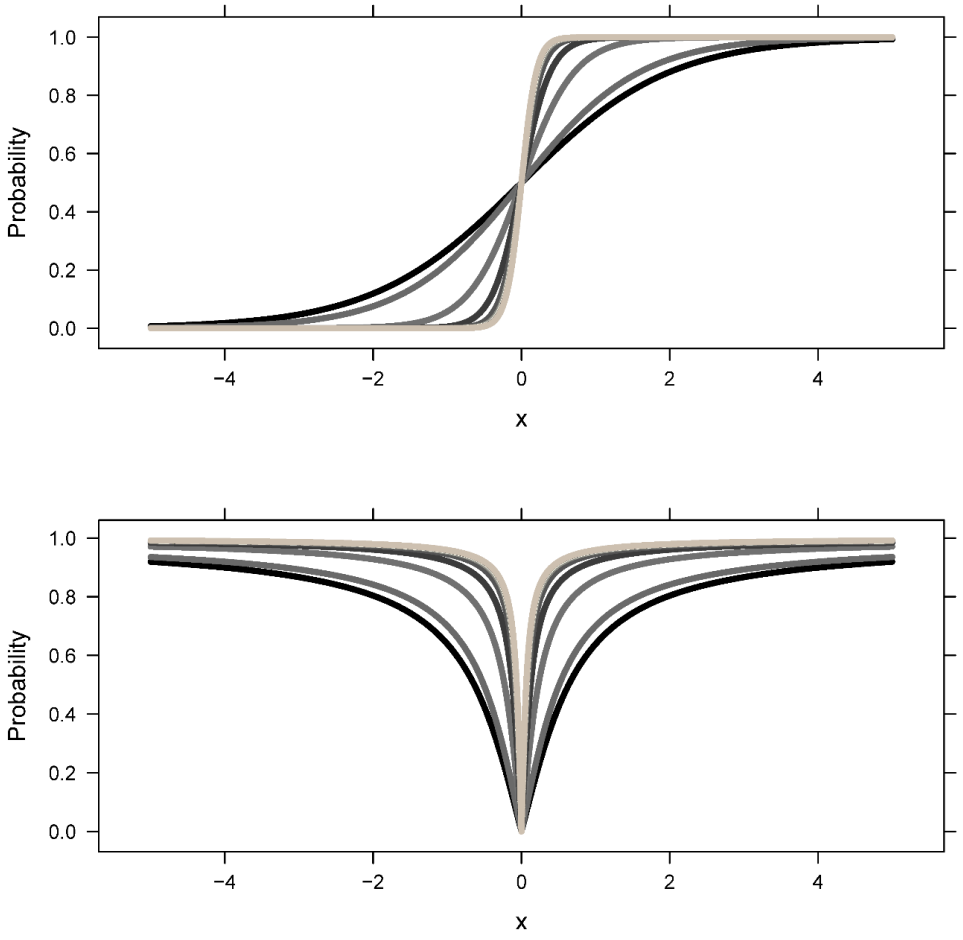

Figure 1: Explanation of the time-varying transfer function when $\lambda_{\max}=2$ and $\lambda_{\min}=0.1$ during 10 iteration. The top panel is the sigmoid transfer function and the bottom panel is the inverse tangent hyperbolic transfer function.

3. Experimental setting

3.1. Datasets

A set of seven benchmark biopharmaceutical datasets was used in our paper. These datasets include a binary class (active/inactive compound) with thousands of descriptors. These datasets have been used by Eklund, Norinder, Boyer and Carlsson [2] and Eklund, Norinder, Boyer and Carlsson [14]. Dragon software (version 6.0) was used to generate the molecular descriptors. To include consistent and useful descriptors, preprocessing steps were carried out as follows: First, those that had zero values for all molecules were discarded. Second, those that had a constant value for all molecules were excluded from the study. Then, descriptors in which 95% of their values were zeros were removed. And, finally, descriptors with a relative standard deviation of less than 0.001 were removed. The basic information about the datasets is described in Table 1.

Table 1: Datasets description

Dataset	End point	# compounds	# descriptors
COX2	cyclooxygenase-2	322 (154/168)	3449
CPD	Carcinogenic categorical activity in rats	1198 (581/617)	4073
DHFR	Dihydrofolate reductase	397 (201/196)	4411
EPAFHM	Fathead minnow acute toxicity	577 (297/280)	3682
FDA	FDA maximum recommended daily dose	1216 (581/635)	3957
cas_N6512	Ames mutagenicity	6512 (3171/3341)	4266
screen_U251	DTP human tumor U251 cell line screen	3743 (1922/1821)	3884

3.2. BBHA parameters initialization

There are four control parameters in our proposed time-varying transfer function: The number of stars (n_s) (Population size), the maximum number of iterations (t_{max}), and the minimum and maximum values of the control parameter λ of Eq. (7). The specific parameter values are outlined in Table 2. The position for each star is a vector of 0 and 1 values with size equals the number of the descriptors. Initially, the positions were randomly generated from a uniform distribution between 0 and 1. Further, the best fitness function that can combine the maximum classification performance and the minimum number of selected descriptors is preferable. The fitness function used in BBHA to evaluate each star position is defined as

$$\text{fitness} = 0.8 \times \text{CA} + 0.2 \times \left(\frac{d - \tilde{d}}{d} \right), \qquad (10)$$

where CA is the classification accuracy obtained from the training dataset, d represents the number of descriptors in the dataset, and $\tilde{d}$ represents the number of selected descriptors.

Table 2: Parameter setting for BBHA

Parameter	Value
n_s	30
t_{max}	500

λ_{max} in Eq. (7)	2
λ_{min} in Eq. (7)	0.1

4. Results and discussion

With the aim of correctly assessing the performance of our proposed time-varying transfer functions, $T_{\text{TV-SIG}}$ and $T_{\text{TV-TH}}$, comparative experiments with the original T_{SIG} and T_{TH} were carried out. In this study, the experiments were carried out using a support vector machine (SVM) with a linear kernel function. To obtain a reliable classification performance, for each dataset, 70% of samples is used as a training dataset and the remaining 30% of the samples are used as a testing dataset randomly. This partition repeated 50 times independently and the average classification accuracy

$(\text{CA} = (\text{TP}+\text{TN})/(\text{TP}+\text{FP}+\text{FN}+\text{TN}) \times 100\%)$ and the average number of selected descriptors are reported in Table 3.

It can be seen in Table 3 that, from among the four transfer functions, the $T_{\text{TV-TH}}$ function performs the best with average results of, overall the seven datasets, 96.07% and 94.27%, in terms of classification accuracy, for both training and testing datasets, respectively. Further, as it can be observed from Table 3, $T_{\text{TV-TH}}$ and $T_{\text{TV-SIG}}$ overtake the standard T_{TH} and T_{SIG}. Beside the high classification performance, the robustness is an important factor in evaluating a classifier. The standard deviation of all criteria for $T_{\text{TV-TH}}$ in all datasets is small. This shows that $T_{\text{TV-TH}}$ is a robust transfer function.

In terms of CA criterion, for both the training and testing datasets, the proposed $T_{\text{TV-TH}}$ outperformed T_{SIG}, T_{TH}, and $T_{\text{TV-SIG}}$ in all datasets. The most remarkable result for $T_{\text{TV-TH}}$ concerns the screen_U251 dataset. We obtain 98.56% and 96.77% accuracy with average 11 descriptors for training and testing datasets, respectively.

Moreover, Table 3 demonstrates that the $T_{\text{TV-TH}}$ and the $T_{\text{TV-SIG}}$ are significantly better than T_{SIG} and T_{TH} transfer functions in terms of the number of selected descriptors. $T_{\text{TV-TH}}$ selects descriptors approximately 2 times fewer than $T_{\text{TV-SIG}}$ and 4 times fewer than T_{TH}.

Comparatively speaking, in all of the datasets, T_{SIG} obtains worse classification accuracy compared with other transfer functions. In addition, $T_{\text{TV-TH}}$ selects the minimum number of the descriptors while keeping better classification performance.

Table 3: Average performance of the proposed time-varying transfer functions. The number in parentheses is the corresponding standard deviation.

Dataset	Method	Training dataset		Testing dataset
		# selected descriptors	CA	CA
COX2	T_{SIG}	25 ± 0.014	94.32 ± 0.013	92.84 ± 0.015
	T_{TH}	21 ± 0.011	95.14 ± 0.012	93.27 ± 0.014
	$T_{\text{TV-SIG}}$	21 ± 0.012	95.61 ± 0.012	94.07 ± 0.013

	$T_{\text{TV-TH}}$	**18 ± 0.011**	**96.53 ± 0.012**	**94.72 ± 0.013**
CPD	T_{SIG}	27 ± 0.021	94.96 ± 0.023	93.48 ± 0.024
	T_{TH}	24 ± 0.022	95.68 ± 0.022	93.91 ± 0.023
	$T_{\text{TV-SIG}}$	23 ± 0.019	96.25 ± 0.022	94.71 ± 0.022
	$T_{\text{TV-TH}}$	**21 ± 0.019**	**97.27 ± 0.021**	**95.38 ± 0.022**
DHFR	T_{SIG}	23 ± 0.033	93.68 ± 0.031	92.22 ± 0.032
	T_{TH}	19 ± 0.029	94.51 ± 0.030	92.63 ± 0.031
	$T_{\text{TV-SIG}}$	19 ± 0.030	94.97 ± 0.029	93.43 ± 0.031
	$T_{\text{TV-TH}}$	**16 ± 0.031**	**95.89 ± 0.029**	**94.08 ± 0.030**
EPAFHM	T_{SIG}	21 ± 0.018	95.15 ± 0.017	92.21 ± 0.018
	T_{TH}	17 ± 0.017	95.97 ± 0.017	93.63 ± 0.017
	$T_{\text{TV-SIG}}$	17 ± 0.017	96.44 ± 0.018	94.43 ± 0.018
	$T_{\text{TV-TH}}$	**14 ± 0.016**	**97.36 ± 0.016**	**95.78 ± 0.015**
FDA	T_{SIG}	32 ± 0.036	91.49 ± 0.034	90.01 ± 0.032
	T_{TH}	28 ± 0.034	92.31 ± 0.033	90.44 ± 0.031
	$T_{\text{TV-SIG}}$	28 ± 0.034	92.78 ± 0.031	91.24 ± 0.031
	$T_{\text{TV-TH}}$	**25 ± 0.033**	**93.72 ± 0.031**	**91.89 ± 0.030**
cas_N6512	T_{SIG}	34 ± 0.025	90.89 ± 0.026	89.43 ± 0.026
	T_{TH}	30 ± 0.023	91.73 ± 0.024	89.84 ± 0.025
	$T_{\text{TV-SIG}}$	29 ± 0.025	92.11 ± 0.025	90.64 ± 0.025
	$T_{\text{TV-TH}}$	**28 ± 0.021**	**93.17 ± 0.023**	**91.29 ± 0.022**
screen_U251	T_{SIG}	18 ± 0.016	96.35 ± 0.015	94.87 ± 0.016
	T_{TH}	14 ± 0.016	97.17 ± 0.014	95.33 ± 0.015
	$T_{\text{TV-SIG}}$	14 ± 0.015	97.64 ± 0.013	96.12 ± 0.013
	$T_{\text{TV-TH}}$	**11 ± 0.014**	**98.56 ± 0.011**	**96.77 ± 0.013**

To further highlight the efficiency of the proposed time-varying transfer functions, Table 4 displays the execution time in seconds, on average, for all considered transfer functions. The computational efficiency of $T_{\text{TV-TH}}$ and $T_{\text{TV-SIG}}$ are comparable to T_{TH} and T_{SIG}. It is noteworthy that $T_{\text{TV-TH}}$ has the fastest convergence speed beating the other three used transfer functions, where it requires the least amount of time to complete the optimized target.

Table 4: Average CPU time, in seconds, of the proposed time-varying transfer functions.

Dataset	T_{SIG}	T_{TH}	$T_{\text{TV-SIG}}$	$T_{\text{TV-TH}}$	SVM
COX2	237.21	208.59	118.57	**101.36**	371.58
CPD	361.21	332.59	242.57	**226.33**	406.33

DHFR	418.95	389.33	299.31	**282.11**	532.84
EPAFHM	259.78	231.16	141.14	**125.97**	310.79
FDA	314.19	284.59	194.57	**178.34**	428.09
cas_N6512	486.86	458.24	368.22	**351.01**	537.91
screen_U251	386.78	358.16	267.14	**249.93**	428.39

To further verify the effectiveness of the proposed time-varying transfer functions, the statistical paired t-test is performed to verify whether there is a significant difference between $T_{\text{TV-TH}}$ and $T_{\text{TV-SIG}}$, and the other two transfer functions in terms of the area under the curve (AUC). Table 5 displays the difference values between the $T_{\text{TV-TH}}$ and the other three functions and the p-values (in the parentheses). The bold values indicate a statistical significant taking significance level of $\alpha = 0.05$. It can be seen that there is a statistical difference between $T_{\text{TV-TH}}$ and each of T_{TH} and T_{SIG} for the seven dataset. On the other hand, there is a statistical difference between the $T_{\text{TV-TH}}$ and the $T_{\text{TV-SIG}}$ for CPD, FDA, cas_N6512, and screen_U251 datasets.

Table 5: Paired t-test results of $T_{\text{TV-TH}}$ and the other three functions in terms of AUC.

Dataset	T_{SIG}	T_{TH}	$T_{\text{TV-SIG}}$
COX2	**2.21 (0.000)**	**1.39 (0.004)**	0.92 (0.065)
CPD	**2.37 (0.000)**	**1.65 (0.007)**	**1.08 (0.041)**
DHFR	**2.28 (0.000)**	**1.45 (0.004)**	0.99 (0.071)
EPAFHM	**2.25 (0.000)**	**1.43 (0.001)**	0.96 (0.064)
FDA	**2.31 (0.000)**	**1.49 (0.002)**	**1.02 (0.037)**
cas_N6512	**2.38 (0.000)**	**1.54 (0.004)**	**1.16 (0.000)**
screen_U251	**2.32 (0.000)**	**1.48 (0.003)**	**1.01 (0.033)**

In summary, from the experimental results, it has been approved that $T_{\text{TV-TH}}$ and the $T_{\text{TV-SIG}}$ outperform T_{TH} and T_{SIG} with superiority of $T_{\text{TV-TH}}$ over $T_{\text{TV-SIG}}$ and achieved better results in the classification accuracy and the number of the selected descriptors. The reason behind outperforming the proposed time-varying transfer function over the T_{TH} and T_{SIG} is that $T_{\text{TV-TH}}$ and $T_{\text{TV-SIG}}$ have a good exploitation and exploration property by adding the time-varying parameter. Additionally, the proposed $T_{\text{TV-TH}}$ and $T_{\text{TV-SIG}}$, can achieve high classification performance with least number of descriptors in a short time.

5. Conclusion

In this paper, a new time-varying transfer function was proposed to improve the exploration and exploitation capability of the BBHA. It was evaluated in the context of variable selection in QSAR/ QSPR classification over several datasets and transfer functions. Depending on seven benchmark biopharmaceutical datasets, the experimental results show that the proposed time-varying transfer function, $T_{\text{TV-TH}}$ and $T_{\text{TV-SIG}}$, are capable of reducing the number of descriptors while maintaining

the high accuracy of classification with short computing time. The efficiency of $T_{\text{TV-TH}}$ in terms of AUC was also evaluated by utilizing paired t-test. The results show that the $T_{\text{TV-TH}}$ has significantly gained better classification performance compared to other used methods.

REFERENCES

[1] (!!! INVALID CITATION !!! (Abdo Mohammed Al-Fakih et al., 2016; Gupta et al., 2015; Nantasenamat et al., 2014; Pourbasheer et al., 2014)), pp.

[2] M. Eklund, U. Norinder, S. Boyer, and L. Carlsson, Benchmarking Variable Selection in QSAR, Mol Inform, 31 (2012), pp. 173-179.

[3] A.M. Al-Fakih, Z.Y. Algamal, M.H. Lee, H.H. Abdallah, H. Maarof, and M. Aziz, Quantitative structure-activity relationship model for prediction study of corrosion inhibition efficiency using two-stage sparse multiple linear regression, Journal of Chemometrics, 30 (2016), pp. 361-368.

[4] A.M. Al-Fakih, Z.Y. Algamal, M.H. Lee, M. Aziz, and H.T.M. Ali, QSAR classification model for diverse series of antifungal agents based on improved binary differential search algorithm, SAR QSAR Environ Res, 30 (2019), pp. 131-143.

[5] A.M. Al-Fakih, M. Aziz, H.H. Abdallah, Z.Y. Algamal, M.H. Lee, and H. Maarof, High dimensional QSAR study of mild steel corrosion inhibition in acidic medium by furan derivatives, International Journal of Electrochemical Science, 10 (2015), pp. 3568 - 3583.

[6] Z. Algamal, An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression, Electronic Journal of Applied Statistical Analysis, 10 (2017), pp. 242-256.

[7] Z.Y. Algamal, and M.H. Lee, Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification, Expert Systems with Applications, 42 (2015), pp. 9326–9332.

[8] Z.Y. Algamal, and M.H. Lee, Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification, Comput Biol Med, 67 (2015), pp. 136-145.

[9] Z.Y. Algamal, and M.H. Lee, A new adaptive L1-norm for optimal descriptor selection of high-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives, SAR QSAR Environ Res, 28 (2017), pp. 75-90.

[10] Z.Y. Algamal, and M.H. Lee, A two-stage sparse logistic regression for optimal gene selection in high-dimensional microarray data classification, Advances in Data Analysis and Classification, (2018), pp.

[11] Z.Y. Algamal, M.K. Qasim, and H.T.M. Ali, A QSAR classification model for neuraminidase inhibitors of influenza A viruses (H1N1) based on weighted penalized support vector machine, SAR QSAR Environ Res, 28 (2017), pp. 415-426.

[12] M.K. Qasim, Z.Y. Algamal, and H.T.M. Ali, A binary QSAR model for classifying neuraminidase inhibitors of influenza A viruses (H1N1) using the combined minimum redundancy maximum relevancy criterion with the sparse support vector machine, SAR QSAR Environ Res, 29 (2018), pp. 517-527.

[13] O.S. Qasim, and Z.Y. Algamal, Feature selection using particle swarm optimization-based logistic regression model, Chemometrics and Intelligent Laboratory Systems, 182 (2018), pp. 41-46.
[14] M. Eklund, U. Norinder, S. Boyer, and L. Carlsson, Choosing feature selection and learning algorithms in QSAR, J Chem Inf Model, 54 (2014), pp. 837-843.
[15] M. Goodarzi, B. Dejaegher, and Y.V. Heyden, Feature selection methods in QSAR studies, Journal of AOAC International, 95 (2012), pp. 636-651.
[16] R. Todeschini, V. Consonni, A. Mauri, and M. Pavan, Dragon 6, 2012.
[17] Dragon 7, www.kode-solutions.net, 2016.
[18] P. Filzmoser, M. Gschwandtner, and V. Todorov, Review of sparse methods in regression and classification with application to chemometrics, J. Chemom., 26 (2012), pp. 42-51.
[19] J. Garcia, P.R. Duchowicz, M.F. Rozas, J.A. Caram, M.V. Mirifico, F.M. Fernandez, and E.A. Castro, A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases, J. Mol. Graph. Model., 31 (2011), pp. 10-19.
[20] A. Khajeh, H. Modarress, and H. Zeinoddini-Meymand, Modified particle swarm optimization method for variable selection in QSAR/QSPR studies, Structural Chemistry, 24 (2013), pp. 1401-1409.
[21] S. Mirjalili, S.M. Mirjalili, and X.-S. Yang, Binary bat algorithm, Neural Computing & Application, 25 (2014), pp. 663–681.
[22] E. Emary, H.M. Zawbaa, and A.E. Hassanien, Binary grey wolf optimization approaches for feature selection, Neurocomputing, 172 (2016), pp. 371-381.
[23] A. Hatamlou, Black hole: A new heuristic optimization approach for data clustering, Information sciences, 222 (2013), pp. 175-184.
[24] R. Azizipanah-Abarghooee, T. Niknam, F. Bavafa, and M. Zare, Short-term scheduling of thermal power systems using hybrid gradient based modified teaching–learning optimizer with black hole algorithm, Electric Power Systems Research, 108 (2014), pp. 16-34.
[25] H. Bouchekara, Optimal power flow using black-hole-based optimization approach, Applied Soft Computing, 24 (2014), pp. 879-888.
[26] W. Gao, X. Wang, S. Dai, and D. Chen, Study on stability of high embankment slope based on black hole algorithm, Environmental Earth Sciences, 75 (2016), pp. 1381.
[27] A. Hatamlou, Solving travelling salesman problem using black hole algorithm, Soft Computing, 22 (2018), pp. 8167-8175.
[28] E. Pashaei, and N. Aydin, Binary black hole algorithm for feature selection and classification on biological data, Applied Soft Computing, 56 (2017), pp. 94-106.
[29] E. Pashaei, E. Pashaei, and N. Aydin, Gene selection using hybrid binary black hole algorithm and modified binary particle swarm optimization, Genomics, (2018), pp.
[30] C.C. Ramos, D. Rodrigues, A.N. de Souza, and J.P. Papa, On the study of commercial losses in Brazil: a binary black hole algorithm for theft characterization, IEEE Transactions on Smart Grid, 9 (2018), pp. 676-683.
[31] D.D. Warnana, Black hole algorithm for determining model parameter in self-potential data, Journal of Applied Geophysics, 148 (2018), pp. 189-200.

[32] T. Baghgoli, M. Mousavi, and B.M. Bababdani, Descriptor selection evaluation of binary gravitational search algorithm in quantitative structure-activity relationship studies of benzyl phenyl ether diamidine's antiprotozoal activity and Chalcone's anticancer potency, Chemometrics and Intelligent Laboratory Systems, 182 (2018), pp. 31-40.
[33] D.-S. Cao, S. Liu, L. Fan, and Y.-Z. Liang, QSAR analysis of the effects of OATP1B1 transporter by structurally diverse natural products using a particle swarm optimization-combined multiple linear regression approach, Chemometrics and Intelligent Laboratory Systems, 130 (2014), pp. 84-90.
[34] Q. Cao, and K. Leung, Prediction of chemical biodegradability using support vector classifier optimized with differential evolution, Journal of chemical information and modeling, 54 (2014), pp. 2515-2523.
[35] A. Khajeh, H. Modarress, and H. Zeinoddini-Meymand, Application of modified particle swarm optimization as an efficient variable selection strategy in QSAR/QSPR studies, Journal of chemometrics, 26 (2012), pp. 598-603.
[36] F. Liu, and Z. Zhou, A new data classification method based on chaotic particle swarm optimization and least square-support vector machine, Chemometrics and intelligent laboratory systems, 147 (2015), pp. 147-156.
[37] J. Shi, X. Hu, X. Zou, J. Zhao, W. Zhang, X. Huang, Y. Zhu, Z. Li, and Y. Xu, A heuristic and parallel simulated annealing algorithm for variable selection in near-infrared spectroscopy analysis, Journal of Chemometrics, 30 (2016), pp. 442-450.
[38] Y. Wang, J.J. Huang, N. Zhou, D.S. Cao, J. Dong, and H.X. Li, Incorporating PLS model information into particle swarm optimization for descriptor selection in QSAR/QSPR, Journal of Chemometrics, 29 (2015), pp. 627-636.
[39] Y.-P. Zhou, L.-J. Tang, J. Jiao, D.-D. Song, J.-H. Jiang, and R.-Q. Yu, Modified particle swarm optimization algorithm for adaptively configuring globally optimal classification and regression trees, Journal of chemical information and modeling, 49 (2009), pp. 1144-1153.
[40] M.J. Islam, X. Li, and Y. Mei, A time-varying transfer function for balancing the exploration and exploitation ability of a binary PSO, Applied Soft Computing, 59 (2017), pp. 182-196.
[41] M. Mafarja, I. Aljarah, A.A. Heidari, H. Faris, P. Fournier-Viger, X. Li, and S. Mirjalili, Binary dragonfly optimization for feature selection using time-varying transfer functions, Knowledge-Based Systems, 161 (2018), pp. 185-204.
[42] S. Mirjalili, and A. Lewis, S-shaped versus V-shaped transfer functions for binary particle swarm optimization, Swarm and Evolutionary Computation, 9 (2013), pp. 1-14.