STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–11.
Published online in International Academic Press (www.IAPress.org)

Binary Spotted Hyena Optimization-Based Feature Selection for Kernel Ridge Regression in Quantitative Structure-Activity Relationship modeling

Zainab Modhfer Ali Al-Shabaki, Zakariya Yahya Algamal*

Abstract Variable selection plays a critical role in enhancing the predictive accuracy, interpretability, and computational efficiency of kernel ridge regression (KRR) models, especially when applied to high-dimensional datasets such as those used in QSAR modeling. This study investigates improved binary spotted hyena optimization algorithm (BSHO) variants incorporating different transfer functions for variable selection in KRR. The performance of these variants was extensively evaluated on seven benchmark biopharmaceutical datasets with thousands of molecular descriptors, comparing their prediction accuracy, variable subset compactness, and computational cost against baseline KRR without feature selection. Results demonstrate that all BSHO variants significantly outperform KRR in terms of mean squared error (MSE) and coefficient of determination. The Quadratic BSHO (Q-BSHO) variant consistently achieved the best predictive performance, reducing MSE by up to 30% and increasing coefficient of determination to values above 0.95 on several datasets while selecting the fewest variables, reflecting effective and parsimonious variable selection. Furthermore, BSHO variants substantially decreased computational time required for model training compared to KRR, with Q-BSHO exhibiting the lowest runtime across datasets. Statistical validation using the Wilcoxon signed-rank test confirmed that all BSHO variants provided statistically significant improvements over KRR. The findings highlight the efficacy of sophisticated binary metaheuristic algorithms for variable selection in kernel-based models, underscoring their potential in computational chemistry and related domains where high-dimensionality and nonlinear interactions complicate predictive modeling.

Keywords Kernel ridge regression, QSAR, variable selection, spotted hyena optimization algorithm, transfer function

DOI: 10.19139/soic-2310-5070-3338

1. Introduction

Quantitative Structure-Activity Relationship (QSAR) modeling is a computational and mathematical approach widely used in chemistry, biology, and drug design to predict the biological activity or properties of chemical compounds based on their molecular structures [33, 35, 36, 37]. The fundamental assumption of QSAR is that chemically similar molecules tend to exhibit similar biological effects, which allows the mathematical linking of chemical structures, expressed through molecular descriptors or physicochemical properties, to biological activities or other relevant properties [4]. QSAR models are typically constructed by first selecting a dataset of compounds with known biological activities and then extracting meaningful descriptors that quantitatively represent the molecular structure and physicochemical properties of these compounds [38, 39]. These descriptors can range from simple properties like molecular weight and lipophilicity to more complex topological, electronic, and geometric features. Using statistical or machine learning techniques, such as multiple linear regression, partial least squares, support vector machines, or artificial neural networks, the relationship between these descriptors (independent variables) and biological activity (dependent variable) is modeled, resulting in equations or algorithms that can predict the activity of untested compounds [16, 26].

*Correspondence to: Zakariya Yahya Algamal (Email: zakariya.algamal@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

QSAR plays a crucial role in drug discovery and chemical risk assessment by enabling virtual screening of large chemical libraries to identify promising candidates with desired biological effects and reduced toxicity, significantly reducing the time and cost of experimental testing. Its success depends heavily on the quality of the input data, the choice of descriptors, and robust model validation techniques, making it a powerful and essential tool in computational chemistry and cheminformatics [3, 22, 25].

One of the most popular kernels is Kernel Ridge Regression (KRR) that has similar generalization to that of Support Vector Machine (SVM) and training time and closed-form solutions. However, it becomes ineffective in the face of noise and imbalance-bearing datasets [13, 23, 29]presented a co-trained KRR (Co-KRR) to improve the generalization capabilities of the KRR model. [23] Suggested a collection of Kernel Ridge Regression (KRR) models for solving classification challenges.

KRR is advantageous due to its cost-effectiveness and efficiency in knowledge-finding applications [7]. Their second-to-none classification and regression performance attracted a lot of attention. Moreover, fraudulent claims must be detected before money is paid out and new fraud trends should be revealed as soon as possible. KRR provides an all-purpose multiclass categorization solution. Hence, we use knowledge representation and reasoning (KRR) methods to analyze detecting fraud activity in automotive insurance [27].

[24] introduced the concept of kernel ridge regression. We use Kernel Ridge Regression, as a regression approach for forecasting. The design specifically addresses scenarios where multiple predictors interact nonlinearly with the target variable. The KRR method can generate significantly more accurate predictions than classic linear and nonlinear strategies when working with various predictors based on main components [11, 15].

Feature selection (sometimes called variable selection) is an important stage of statistical modelling and machine learning, which is concerned with identifying and choosing the most significant variables (features or predictors) amongst a larger number of candidates to be used in the model. The main objective of variable selection is to maximize the performance of the model through minimization of overfitting, maximization of predictive ability and simplification of the model to make it easier to interpret and reduce computation cost. Issues of selection of variables have several significant objectives: they eliminate redundant variables or irrelevant variables that do not add value to the prediction capabilities of a model, they help to reduce the dimensions of the data, and they assist in the detection of the actual underlying relationships between the predictors and the response variable. It is particularly valued in situations where the data is high-dimensional where a large number of variables can be correlated or irrelevant.

2. Kernel Ridge Regression

Ridge regression is a linear regression approach that incorporates a sum-of-squares error (sse) function together with regularization, allowing for control over the bias-variance trade-off [12]. Its goal is to reveal a hidden linear pattern within the initial data [24, 32]. Kernel ridge regression, an enhanced form of linear ridge regression that incorporates kernel tricks, can estimate nonlinear mappings [7, 21, 30].

The kernel ridge regression employs the kernel technique in conjunction with ridge regressions [31, 34]. One major benefit of the KRR is that it uses a regularization and kernel technique to effectively capture the non-linear relationship, thereby addressing the problem of over-fitting in regression [6, 28].

The dataset consists of Mtraining data pairs,$(x_1, t_1), (x_2, t_2), \ldots, (x_M, t_M)$, where Arepresents the number of inputs. Additionally, there is a nonlinear mapping function $\phi(x_i)$ that converts the original input space into a higher-dimensional feature space. The linear regression model is denoted as (Naik, Bisoi, & Dash, 2018)

$$y_i = \alpha \bullet \phi(x_i), \quad i = 1, 2, \ldots, M \tag{1}$$

Here, $y_i$represents the output, and $\alpha$is the weight vector. KRR utilizes regularized least squares to minimize the objective function and generate the weight vectorα.

$$Minimize \quad G_{KRR} = \frac{1}{2}\|\alpha\|^2 + \frac{1}{2}\lambda \sum_{i=1}^{M} (t_i - \alpha \bullet \phi(x_i))^2 \tag{2}$$

In equation (2), λ represents a regularization parameter. The user must modify this parameter, which is a positive constant. It serves as a penalty for the squared error. The value of $\lambda$is selected to be equal to$\lambda = 2^c$, where $c$is a positive number. We prepare the network to forecast the time series data after training it and setting the output weights. Using Lagrange multipliers in Equation (2), we derive the following expression:

$$G_{KRR} = \frac{1}{2}\left\|\alpha\right\|^2 + \frac{1}{2}\lambda \sum_{i=1}^{M} \left(t_i - \alpha \bullet \phi(x_i)\right)^2 + \sum_{i=1}^{M} \beta_i \left(t_i - \alpha \bullet \phi(x_i) - \left(t_i - \alpha \bullet \phi(x_i)\right)\right) \tag{3}$$

By differentiating $G_{KRR}$with respect to $\alpha$and$(t_i - \alpha \bullet \phi(x_i))$, and equating the resulting βequations to zero, we obtain the output weight vector αas

$$\alpha = \left(\phi'\phi + \frac{I}{\lambda}\right)^{-1} \phi' T \tag{4}$$

Adding a positive $I/_{\lambda}$value to the diagonal of a $\phi'\phi$ matrix results in a stable solution for the KRR's output weight vectorα. We find that this stable solution has good generalization performance. In addition toϕ, the matrix's rows include the mapped samples$\phi(x_i)$, where $I$is an identity matrix of dimension$M \times M$. Furthermore

$$\alpha = \sum_{i=1}^{M} \beta_i \phi(x_i) = \phi' \beta \tag{5}$$

Consequently, we derive the explicit expression for the dual weight βsolution.

$$\beta = \left(\phi'\phi + \frac{I}{\lambda}\right)^{-1} T = \left(K + \frac{I}{\lambda}\right)^{-1} T \tag{6}$$

We construct the kernel matrix $K$by obtaining its entries through a specific process.

$$K(x_i, x_j) = \phi(x_i)\phi(x_j)' \tag{7}$$

In the end, the new test sample $x_i$yields the expected result.

$$h_i(x_i) = \sum_{i=1}^{M} \beta_i K(x_i, x_i) = K' \times (K + \frac{I}{\lambda})^{-1} T \tag{8}$$

The $K'$is defined as$\left[K(x_i, x_1), K(x_i, x_2), \ldots, K(xi, x_M)\right]'$.

Kernel functions that satisfy the Mercers' condition are widely recognized for their ability to enhance the computational capabilities of learning systems. They achieve this by mapping the data into a feature space with a large number of dimensions, which, in turn, makes the data linearly separable. This leads to improved accuracy, stability, and generalization for both regression and classification problems.

3. The Spotted Hyena Optimization (SHO)

The fact that the social bond that exists between spotted hyenas and the cooperative behavior that they exhibit hunting prey served as the inspiration for a recently developed innovative swarm-based algorithm that was given the name "Spotted hyena optimization (SHO) [13, 14]. This technique consists of three primary steps: encircling, hunting, and attacking prey. The mathematical modeling of the SHO algorithm is detailed in the following paragraphs.

Encircling prey:

Due to the unknown search space, the target prey is considered to be the current best candidate solution that is near the ideal solution. Once that intended prey has been identified, other hyenas will attempt to adjust their position toward the location of the prey [15, 16].

$$\vec{D}_h = \left| \vec{B} \cdot \vec{P}_p\left(x\right) - \vec{P}\left(x\right) \right| \tag{9}$$

$$\vec{P}\left(x+1\right) = \vec{P}_p\left(x\right) - \vec{E} \cdot \vec{D}_h \tag{10}$$

where, $\vec{D}_h$ is the distance between hyenas and their prey, x denotes the current iteration, $\vec{P}$ represents the position vector of the hyena, $\vec{P_p}$ denotes the position vector of the prey, and, $\vec{B}$ and $\vec{E}$ represent the coefficient vectors, which can be calculated as follows [17]:

$$\overrightarrow{B} = 2 \cdot \overrightarrow{rd}_1 \tag{11}$$

$$\overrightarrow{E} = 2\,\vec{h} \cdot \overrightarrow{rd}_2 - \vec{h} \tag{12}$$

$$\vec{h} = 5 - \left(I \times \frac{5}{Max_I} \right) \tag{13}$$

where, I denotes the iteration and Max_I is the maximum number of iterations.

In this scenario, the value of $\vec{h}$ is reduced linearly from 5 to 0 throughout the repetitions. The stability that exists between exploration and exploitation is maintained by its use. The random vector in the range [0,1] are denoted by $\overrightarrow{rd}_1$ and $\overrightarrow{rd}_2$. Adjustments are made to the values of $\overrightarrow{B}$ and $\overrightarrow{E}$ to allow spotted hyenas to go to other regions concerning their present location. Through the use of Equations (1) and (2), hyenas are able to update their location in a random manner around the prey [15].

Hunting prey

To simulate the hunting behavior of spotted hyenas, we use the assumption that the most effective search agent has information on the whereabouts of the prey. The rest of the search agents merge into a cluster of reliable acquaintance and adjust their location based on the most optimal search agent. The mathematical representation of the hunting mechanism can be described as follows [18]:

$$\vec{D}_h = \left| \vec{B} \cdot \vec{P}_h - \vec{P}_k \right| \tag{14}$$

$$\vec{C}_h = \vec{P}_k + \vec{P}_{k+1} + \cdots + \vec{P}_{k+N} \tag{15}$$

here, $\vec{P}_h$ denotes the location of the first observed spotted hyena, $\vec{P}$ shows the positions of the other spotted hyenas, whereas N represents the number of spotted hyenas, which is found using the following calculation [13]:

$$N = count_{nos} \left(\vec{P}_h, \vec{P}_{h+1}, \vec{P}_{h+2}, \ldots, \left(\vec{P}_h + \overrightarrow{M} \right) \right) \tag{16}$$

where $\overrightarrow{M}$ represents a randomly generated vector with a value of [0.5,1]. "nos" represents the number of solutions, while "$count$" refers to the total number of candidate solutions that, when combined with $\overrightarrow{M}$, closely resemble the best optimum solution in the search space. $\vec{C}_h$ represents a group or cluster of N optimal solution.

Attacking prey:

Mathematical modeling for assaulting the prey requires reducing the value of the vector $\vec{h}$. Over the course of repetitions, the value of the vector $\vec{h}$ may decline from 5 to 0, and this is achieved by decreasing the variance in the vector $\vec{E}$. When $|E| < 1$, the spotted hyena gang attacks their victim. The following is the mathematical expression for the assault on the prey [17]:

$$\vec{P}\left(x+1\right) = \frac{\vec{C}_h}{N} \tag{17}$$

where $\vec{P}\left(x+1\right)$ indicates the optimal solution, and additional spotted hyena placements are adjusted according to the best search agent's position (the hyenas who are closest to the prey). Typically, SHO algorithm aims to balance exploration and exploitation, which can lead to competitive performance. However, its computation time can vary based on parameter settings and problem complexity. It usually be slower compared to more established algorithm in certain scenarios, especially if it involves complex operations or a high number of iterations.

4. The suggested improvement

Although KRR is a flexible model, high dimensional data with a large number of features pose a challenge due to the fact that it is more expensive to compute, there is a risk of overfitting and the model is hard to interpret [19]. Therefore, selection of variables is needed in order to select and keep only the most informative variables that have significant contribution to the prediction of the outcome. Variable selection does not only enhance model accuracy, complexity reduction, interpretability and accelerated computation which is crucial especially in large scale applications.

The KRR methods of variable selection generally extend classical feature selection paradigms into the kernelized paradigm [19]. These are filter, wrapper and embedded. Filter methods are based on statistical indicators of feature relevance in isolation of the model, e.g. correlation with the response, or the dependency measures of a kernel. The KRR model is a black-box model in the wrapper method used as an evaluator to locate the best subsets of variables through heuristic or metaheuristic optimization algorithms. Embedded techniques combine feature selection with the model training procedure, such as sparsity-inducing penalties to the regularization scheme, but the natural result of the KRR is not a sparse solution.

The ability to search large and complicated spaces has made metaheuristic algorithms an important part of variable selection in statistical modeling and machine learning. Variable selection, which consists in finding the most informative set of variables in potentially high-dimensional data, is commonly posed as a combinatorial optimization problem [2, 14].

It is aimed at identifying a multiplicity of characteristics, which are the most effective according to model performance measures, including prediction accuracy, model interpretability, or model simplicity. The classical approaches to variable selection such as stepwise regression and filter-based approaches are challenged when dealing with nonlinear models or large data sets, which can easily be overtaken by local optima or involves exhaustive search, which is computationally infeasible.

The use of metaheuristic algorithms is an effective alternative due to its use of stochastic, population-based, and nature-inspired search methods with a balance between exploration and exploitation [40, 41, 42]. In variable selection, metaheuristics tests and optimizes candidate solution which can also be thought of as subsets of variables by comparing them based on fitness functions that quantify the quality of the subsets selected in a given prediction or statistical sense.

Perhaps the most important benefit of metaheuristics is that they can avoid local optima, which can be a problem in older methods of greedy or deterministic feature selection. This renders them especially valuable to high-dimensional data having nonlinear and intricate interactions amid characteristics. Flexibility in specifying custom objective functions is also offered by metaheuristics, which enables a variety of aspects of model quality (including accuracy, sparsity, and computational cost) to be considered. A widely applied metaheuristic algorithm for variable selection is SSA.

Originally SHO is put forward to resolve the perpetual optimization problems. Nevertheless, the optimization is not a continuous problem to conduct the variable selection. Binary SHO (BSHO) is modified to do selection of variables [5]. As opposed to SHO, the value of the position in BSHO is binary where value 1 implies the relevance of the variable and 0 the opposite. The dimension of individual members of SHO is the number of original variables in the model in variable selection.

To construct the binary version of the BSHO. The S-shaped transfer function (sigmoidal function) (S-BSHO), the hyperbolic tan (V-shaped) function (H-BSHO), the quadratic transfer function (Q-BSHO), and the Z-shaped transfer function (Z- BSHO) were employed to design to construct the binary search space.

The parameter configurations for our proposed improvement are presented as follows.

1. The population size $N_{\text{Spotted Hyena}} = 30$ and the maximum number of iterations is $\text{t}_{\max}= 5000$.
2. The total number of the variables in the model gives the number of positions of each sparrow. The positions are randomly assigned randomly with the uniform distribution using 0 and 1. The fitness function is defined as

$$f = \min\left[\frac{1}{n}\sum_{i=1}^{n}(z_{i,test} - \hat{z}_{i,test})^2\right], \tag{18}$$

where $z_{i,test}$ and $\hat{z}_{i,test}$ are, respectively, is the true value and the predicted value of the response variable in the testing dataset.

5. Results

This part uses seven standard biopharmaceutical datasets to show how well our suggested proposed improvement works compared to the KRR without variable selection.

5.1. *Description of the Data Set*

We have used a set of 7 benchmark biopharmaceutical datasets in our work. The sets are binary and the choices are an active or inert substance and they have thousands of descriptions. The datasets have been also used by Eklund, Norinder, Boyer, and Carlsson [1, 4, 10]. Table 1 listed the used data description.

Table 1. Description of the data

Data	**n**	**variables**
CPD	1198	4073
COX2	322	3449
FDA	1216	3957
screen_U251	3743	3884
EPAFHM	577	3682
DHFR	397	4411
CAS_N6512	6512	4266

5.2. *Evaluation criteria*

The prediction performance of the used methods was measured by the mean-squared error (MSE) and leave-one-out internal validation (${Q^2}_{int}$), which are defined by

$$MSE_{train} = \frac{\sum_{i=1}^{n_{train}} (z_{i,train} - \hat{z}_{i,train})^2}{n_{train}}, \tag{19}$$

and

$${Q^2}_{int} = 1 - \left[\frac{\sum_{i=1}^{n_{train}} (z_{i,train} - \hat{z}_{i,train})^2}{\sum_{i=1}^{n_{train}} (z_{i,train} - \bar{z})^2} \right], \tag{20}$$

respectively.

Furthermore, the test dataset was used to validate the model by computing the following criteria

$$MSE_{test} = \frac{\sum_{i=1}^{n_{test}} (z_{i,test} - \hat{z}_{i,test})^2}{n_{test}}, \tag{21}$$

and

$${Q^2}_{ext} = 1 - \left[\frac{\sum_{i=1}^{n_{test}} (z_{i,test} - \hat{z}_{i,test})^2}{\sum_{i=1}^{n_{test}} (z_{i,test} - \bar{z}_{train})^2} \right], \tag{22}$$

respectively, where ${Q^2}_{ext}$is the external validation, n_{train} and n_{test} represent the training and test sample sizes, the $z_{i,train}$, $z_{i,test}$, $\hat{z}_{i,train}$, and $\hat{z}_{i,test}$ stand for the experimental activity of the training dataset, test dataset, and their corresponding predicted values. While $\bar{z}$ and $\bar{z}_{train}$ represent the mean of all the activity values and the mean of the training activity values, respectively.

5.3. *Performance comparisons*

In order to validate the optimality of the results and maintain the abilities of algorithms, we resort to hold-out strategy in which all the datasets are randomly divided into 70 percent training and 30 percent testing. To achieve statistically significant values, this division is reiterated 25 times. Therefore, the statistical indicators are gathered according to the general abilities and final outcomes. This part utilizes the radial basis function (Naik et al., 2018). Tables 2 and 3 present the findings for each technique and dataset.

From Table 2, in terms of the number of selected variables, each BSHO variant selects a subset of variables out of the total available to build a more parsimonious model. This number varies by dataset and algorithm variant. The Q-BSHO consistently selects the fewest variables across all datasets, ranging roughly from 17 variables to 30 variables, indicating a strong feature-selection capability. The S-BSHO, H-BSHO, and Z-BSHO variants select more variables, usually between 20 and 44 variables, adjusting the balance between sparsity and model fit. KRR, by itself, uses all available variables in the dataset as predictors. It cannot reduce dimensionality or identify a subset of important features. This means KRR models are often more complex and may include noisy or irrelevant variables, which can worsen prediction accuracy and interpretability.

In terms of MSE_{train}, Table 1, it clearly seen that all BSHO variants improve the MSE_{train}, showing the benefit of selecting informative variables. The lowest MSE_{train} is consistently achieved by the Q-BSHO variant, which also selects the fewest variables. This suggests more efficient and effective variable selection leads to better generalization. For example, in CPD dataset, KRR has a higher MSE_{train} of 3.72 and without variable selection. Q-BSHO reduces MSE_{train} to 3.02 with only 19 selected variables, indicating substantial performance gain through compact selected features. Comparing among BSHO variants, across all datasets, Q-BSHO yields the smallest MSE_{train}, indicating it selects an optimal subset of variables leading to the most accurate predictions. On the other hand, S-BSHO and H-BSHO show moderate MSE_{train} improvement. These variants select more variables than Q-BSHO but fewer than KRR. Their MSE_{train} tends to be higher than Q-BSHO but substantially lower than KRR without variable selection. This suggests a trade-off where S-BSHO and H-BSHO offer good but less aggressive variable reduction than Q-BSHO, leading to slightly larger errors. While, Z-BSHO performs better than S-BSHO and H-BSHO but not as well as Q-BSHO. Z-BSHO sits in the middle in terms of MSE_{train}, usually selecting more variables than Q-BSHO but fewer than S-BSHO or H-BSHO. Its MSE_{train} values are close to but still above those of Q-BSHO, showing good but not optimal error minimization.

Among the BSHO variants, the differences in $Q^2{}_{int}$ values reflect their varying effectiveness in selecting the most informative variables to maximize the explained variance in the predicted outcomes. Q-BSHO usually attains the highest $Q^2{}_{int}$ (close to or above 0.95 for most datasets), indicating that the subset of variables it selects explains the largest proportion of variance in the response. This suggests Q-BSHO excels at finding a compact yet highly predictive feature subset. Further, Z-BSHO also performs well but with slightly lower $Q^2{}_{int}$ than Q-BSHO. Z-BSHO typically ranks second in $Q^2{}_{int}$ values, still substantially better than using all variables in KRR. These variant balances variable selection and fit, offering a good trade-off between sparsity and prediction. On the other hand, H-BSHO and S-BSHO show progressively lower $Q^2{}_{int}$ compared to Q-BSHO and Z-BSHO. These results indicate the important role of BSHO in variable selection, where precise selection can substantially improve model explanatory power compared to standard KRR without variable selection.

Based on the test dataset, Table 3, across all datasets, all BSHO variants outperform the baseline KRR in both MSE_{test} reduction and $Q^2{}_{ext}$ improvement. This trend reflects the critical advantage of integrating binary metaheuristic optimization for variable selection that identifies relevant variables and removes noise or irrelevant variables, leading to more accurate and generalizable predictive models. Q-BSHO consistently achieves the best performance with the lowest MSE_{test} and highest $Q^2{}_{ext}$ across datasets. This signifies that Q-BSHO's enhance global search capabilities, effectively balancing exploration and exploitation to find near-optimal subsets of variables that maximize explained variance and minimize prediction error. In addition, Z-BSHO follows, showing competitive but slightly inferior performance compared to Q-BSHO. This suggests its search heuristics also achieve strong variable selection but possibly with less refinement or convergence efficiency. On the other hand, H-BSHO and S-BSHO represent intermediate and baseline BSHO variants, respectively, showing moderate gains over KRR but not matching the more advanced Q-BSHO and Z-BSHO variants in minimizing error or boosting explanatory power.

Table 2. Prediction evaluation criteria for the training dataset

Datasets	Methods	# selected variables	MSE_{train}	$Q^{2}{}_{int}$
CPD	KRR	-	3.72	0.823
	S-BSHO	27	3.183	0.84
	H-BSHO	24	3.02	0.848
	Q-BSHO	19	2.5	0.91
	Z-BSHO	23	2.881	0.866
COX2	KRR	-	4.827	0.837
	S-BSHO	33	4.29	0.854
	H-BSHO	30	4.127	0.862
	Q-BSHO	21	3.607	0.924
	Z-BSHO	28	3.988	0.88
FDA	KRR	-	4.178	0.844
	S-BSHO	29	3.641	0.861
	H-BSHO	26	3.478	0.87
	Q-BSHO	19	2.958	0.931
	Z-BSHO	22	3.339	0.887
Screen_U251	KRR	-	6.101	0.797
	S-BSHO	38	5.564	0.814
	H-BSHO	33	5.401	0.822
	Q-BSHO	25	4.881	0.884
	Z-BSHO	29	5.263	0.84
EPAFHM	KRR	-	6.546	0.784
	S-BSHO	46	6.009	0.801
	H-BSHO	43	5.846	0.81
	Q-BSHO	32	5.326	0.871
	Z-BSHO	41	5.707	0.827
DHFR	KRR	-	4.758	0.854
	S-BSHO	35	4.221	0.871
	H-BSHO	31	4.058	0.879
	Q-BSHO	22	3.538	0.941
	Z-BSHO	30	3.919	0.897
CAS_N6512	KRR	-	4.941	0.845
	S-BSHO	31	4.404	0.862
	H-BSHO	28	4.241	0.87
	Q-BSHO	20	3.721	0.932
	Z-BSHO	26	4.104	0.888

Figure 1 illustrates the computational time (in seconds) required by the KRR and its variants with different BSHO variable selection methods across seven datasets. As shown in the Figure 1, KRR requires significantly more time than BSHO methods on every dataset. This is expected since KRR runs on the full set of variables without any preliminary selection, increasing dimensionality and computational burden during model fitting. Conversely, BSHO variants consistently reduce computational time. S-BSHO, H-BSHO, Q-BSHO, Z-BSHO show markedly lower computational times compared to KRR, with Q-BSHO achieving the lowest times overall across datasets. The reduction in computational time is attributable to the metaheuristic variable selection stage, which effectively removes irrelevant or redundant variables, shrinking the input space for KRR and thus lowering computation expenses.

To further highlight the performance of the BSHO variants, we conducted a pairwise comparison between the proposed method and each competing method to further validate the effectiveness of the BSHO variants in selecting the most relevant variables with strong prediction performance. This comparison utilized by Friedman test and the Wilcoxon signed-rank test.

The Wilcoxon signed-rank test was used to determine if two samples represent distinct populations. This nonparametric test's methods are similar to the paired t-test, and it is a pairwise test that aims to identify significant differences in the behaviours of the two methods [8, 9, 17].

Tables (4 – 9) present the outcomes of the Wilcoxon signed-rank test with a significance level of $\alpha = 0.05$. According to the results presented in these tables, all computed p-values were well below the conventional 0.05 threshold, confirming that each Q-BSHO, S-BSHO, H-BSHO, and Z-BSHO approach offers statistically significant

Table 3. Prediction evaluation criteria for the testing dataset

Datasets	Methods	MSE_{test}	$Q^2{}_{ext}$
CPD	KRR	4.398	0.809
	S-BSHO	3.861	0.826
	H-BSHO	3.698	0.834
	Q-BSHO	3.178	0.896
	Z-BSHO	3.559	0.852
COX2	KRR	5.505	0.823
	S-BSHO	4.968	0.84
	H-BSHO	4.805	0.848
	Q-BSHO	4.285	0.91
	Z-BSHO	4.666	0.866
FDA	KRR	4.856	0.83
	S-BSHO	4.319	0.847
	H-BSHO	4.156	0.856
	Q-BSHO	3.636	0.917
	Z-BSHO	4.017	0.873
Screen_U251	KRR	6.779	0.783
	S-BSHO	6.242	0.8
	H-BSHO	6.079	0.808
	Q-BSHO	5.559	0.87
	Z-BSHO	5.941	0.826
EPAFHM	KRR	7.224	0.77
	S-BSHO	6.687	0.787
	H-BSHO	6.524	0.796
	Q-BSHO	6.004	0.857
	Z-BSHO	6.385	0.813
DHFR	KRR	5.436	0.84
	S-BSHO	4.899	0.857
	H-BSHO	4.736	0.865
	Q-BSHO	4.216	0.927
	Z-BSHO	4.597	0.883
CAS_N6512	KRR	5.619	0.831
	S-BSHO	5.082	0.848
	H-BSHO	4.919	0.856
	Q-BSHO	4.399	0.918
	Z-BSHO	4.782	0.874

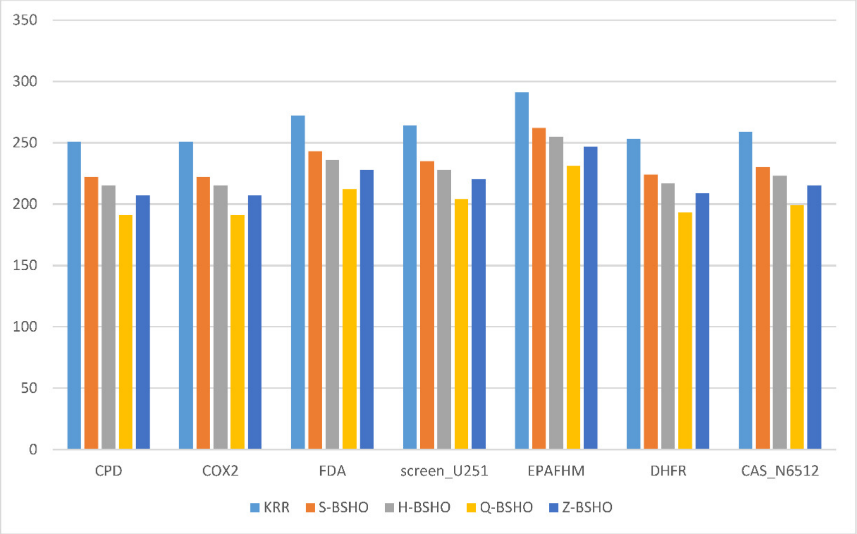

Figure 1. Computational time of the used methods across datasets.

performance improvements over KRR. These results rigorously demonstrate the value of BSHO variable selection in enhancing kernel-based regression modeling.

Table 4. p-values for the Wilcoxon signed-rank test of the used methods result for CPD dataset

Pairwise comparison	ρ-Value
Q-BSHO vs KRR	0.0002
S-BSHO vs KRR	0.0015
H-BSHO vs KRR	0.0017
Z-BSHO vs KRR	0.0005

Table 5. p-values for the Wilcoxon signed-rank test of the used methods result for FDA dataset

Pairwise comparison	ρ-Value
Q-BSHO vs KRR	0.0002
S-BSHO vs KRR	0.0021
H-BSHO vs KRR	0.0018
Z-BSHO vs KRR	0.0006

Table 6. p-values for the Wilcoxon signed-rank test of the used methods result for Screen$_U 251 dataset$

Pairwise comparison	ρ-Value
Q-BSHO vs KRR	0.0005
S-BSHO vs KRR	0.0021
H-BSHO vs KRR	0.0017
Z-BSHO vs KRR	0.0009

Table 7. p-values for the Wilcoxon signed-rank test of the used methods result for EPAFHM dataset

Pairwise comparison	ρ-Value
Q-BSHO vs KRR	0.0005
S-BSHO vs KRR	0.0014
H-BSHO vs KRR	0.0013
Z-BSHO vs KRR	0.0010

Table 8. p-values for the Wilcoxon signed-rank test of the used methods result for DHFR dataset

Pairwise comparison	ρ-Value
Q-BSHO vs KRR	0.0002
S-BSHO vs KRR	0.0018
H-BSHO vs KRR	0.0018
Z-BSHO vs KRR	0.0007

Table 9. p-values for the Wilcoxon signed-rank test of the used methods result for CAS$_N 6512 dataset$

Pairwise comparison	ρ-Value
Q-BSHO vs KRR	0.0004
S-BSHO vs KRR	0.0012
H-BSHO vs KRR	0.0013
Z-BSHO vs KRR	0.0007

6. Conclusion

This study systematically evaluated the integration of BSHO variants with KRR for variable selection in high-dimensional QSAR modeling. The results across seven diverse biopharmaceutical datasets demonstrate that all

BSHO variants, notably the Q-BSHO, substantially improve predictive performance, parsimony, and computational efficiency relative to standard KRR. Q-BSHO consistently achieved the lowest mean MSE_{train}, with reductions up to 30%, and the highest $Q^2{}_{int}$, while also reducing the number of selected variables to as few as 19 in some datasets and minimizing computational time. Wilcoxon signed-rank tests confirmed that these performance gains are statistically significant, establishing the advantage of metaheuristic-driven variable selection in kernel methods. Despite these advancements, several limitations remain. Like other swarm intelligence optimization algorithms, the BSHO approach can exhibit uneven initial population distribution, premature convergence, and a tendency to fall into local optima, particularly in highly complex or high-dimensional settings. Future work should focus on developing hybrid BSHO frameworks that incorporate adaptive learning strategies, opposition-based or chaos-enhanced mechanisms, and dynamic population diversity preservation to mitigate stagnation and further improve optimization robustness. Comparative studies with other state-of-the-art metaheuristic optimizers and extensions to multi-objective variable selection or deep kernel architectures are promising directions for expanding the practical impact of this approach.

REFERENCES

1. A. Al-Fakih, Z. Algamal, M. Lee, M. Aziz, and H. Ali, *QSAR classification model for diverse series of antifungal agents based on improved binary differential search algorithm*, SAR and QSAR in Environmental Research, vol. 30, no. 2, pp. 131–143, 2019.
2. Z. Y. Algamal, *A new method for choosing the biasing parameter in ridge estimator for generalized linear model*, Chemometrics and Intelligent Laboratory Systems, vol. 183, pp. 96–101, 2018.
3. Z. Y. Algamal, M. H. Lee, A. M. Al-Fakih, and M. Aziz, *High-dimensional QSAR prediction of anticancer potency of imidazo[4,5-b]pyridine derivatives using adjusted adaptive LASSO*, Journal of Chemometrics, vol. 29, no. 10, pp. 547–556, 2015.
4. Z. Y. Algamal, M. K. Qasim, M. H. Lee, and H. T. M. Ali, *High-dimensional QSAR/QSPR classification modeling based on improving pigeon optimization algorithm*, Chemometrics and Intelligent Laboratory Systems, vol. 206, p. 104170, 2020.
5. A. M. Alharthi, N. A. Al-Thanoon, A. M. Al-Fakih, and Z. Y. Algamal, *QSAR modelling of enzyme inhibition toxicity of ionic liquid based on chaotic spotted hyena optimization algorithm*, SAR QSAR Environ Res, vol. 35, no. 9, pp. 757–770, 2024.
6. M. Ali, R. Prasad, Y. Xiang, and Z. M. Yaseen, *Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts*, Journal of Hydrology, vol. 584, p. 124647, 2020.
7. S. An, W. Liu, and S. Venkatesh, *Face recognition using kernel ridge regression*, in 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007.
8. E. Baş and G. Yildizdan, *Enhanced coati optimization algorithm for big data optimization problem*, Neural Processing Letters, vol. 55, no. 8, pp. 10131–10199, 2023.
9. J. Derrac, S. García, D. Molina, and F. Herrera, *A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms*, Swarm and Evolutionary Computation, vol. 1, no. 1, pp. 3–18, 2011.
10. M. Eklund, U. Norinder, S. Boyer, and L. Carlsson, *Benchmarking variable selection in QSAR*, Molecular Informatics, vol. 31, no. 2, pp. 173–179, 2012.
11. P. Exterkate, P. J. Groenen, C. Heij, and D. van Dijk, *Nonlinear forecasting with many predictors using kernel ridge regression*, International Journal of Forecasting, vol. 32, no. 3, pp. 736–753, 2016.
12. S. Geman, E. Bienenstock, and R. Doursat, *Neural networks and the bias/variance dilemma*, Neural Computation, vol. 4, no. 1, pp. 1–58, 1992.
13. B. B. Hazarika and D. Gupta, *Affinity based fuzzy kernel ridge regression classifier for binary class imbalance learning*, Engineering Applications of Artificial Intelligence, vol. 117, p. 105544, 2023.
14. M. A. Kahya, S. A. Altamir, and Z. Y. Algamal, *Improving firefly algorithm-based logistic regression for feature selection*, Journal of Interdisciplinary Mathematics, vol. 22, no. 8, pp. 1577–1581, 2019.
15. P. S. Kumar, S. Laha, and L. Kumaraswamidhas, *Assessment of rolling element bearing degradation based on Dynamic Time Warping, kernel ridge regression and support vector regression*, Applied Acoustics, vol. 208, p. 109389, 2023.
16. B. Y. S. Li, L. F. Yeung, and K. T. Ko, *Indefinite kernel ridge regression and its application on QSAR modelling*, Neurocomputing, vol. 158, pp. 127–133, 2015.
17. J. Ma, D. Xia, H. Guo, Y. Wang, X. Niu, Z. Liu, and S. Jiang, *Metaheuristic-based support vector regression for landslide displacement prediction: A comparative study*, Landslides, vol. 19, no. 10, pp. 2489–2511, 2022.
18. S. W. Mahmood, G. T. Basheer, and Z. Y. Algamal, *Improving kernel ridge regression for medical data classification based on meta-heuristic algorithms*, Kuwait Journal of Science, vol. 52, no. 3, 2025.
19. S. W. Mahmood, G. T. Basheer, and Z. Y. Algamal, *Quantitative Structure–Activity Relationship Modeling Based on Improving Kernel Ridge Regression*, Journal of Chemometrics, vol. 39, no. 5, 2025.
20. J. Naik, R. Bisoi, and P. Dash, *Prediction interval forecasting of wind speed and wind power using modes decomposition based low rank multi-kernel ridge regression*, Renewable Energy, vol. 129, pp. 357–383, 2018.
21. C. Orsenigo and C. Vercellis, *Kernel ridge regression for out-of-sample mapping in supervised manifold learning*, Expert Systems with Applications, vol. 39, no. 9, pp. 7757–7762, 2012.
22. M. Qasim, Z. Algamal, and H. 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 and QSAR in Environmental Research, vol. 29, no. 7, pp. 517–527, 2018.

23. K. Rakesh and P. N. Suganthan, *An ensemble of kernel ridge regression for multi-class classification*, Procedia Computer Science, vol. 108, pp. 375–383, 2017.
24. C. Saunders, A. Gammerman, and V. Vovk, *Ridge regression learning algorithm in dual variables*, in Proceedings of the 15th International Conference on Machine Learning, 1998.
25. A. Tropsha, *Best practices for QSAR model development, validation, and exploitation*, Molecular Informatics, vol. 29, no. 6-7, pp. 476–488, 2010.
26. J. Verma, V. M. Khedkar, and E. C. Coutinho, *3D-QSAR in drug design-a review*, Current Topics in Medicinal Chemistry, vol. 10, no. 1, pp. 95–115, 2010.
27. C. Yan, Y. Li, W. Liu, M. Li, J. Chen, and L. Wang, *An artificial bee colony-based kernel ridge regression for automobile insurance fraud identification*, Neurocomputing, vol. 393, pp. 115–125, 2020.
28. Y. You, J. Demmel, C.-J. Hsieh, and R. Vuduc, *Accurate, fast and scalable kernel ridge regression on parallel and distributed systems*, in Proceedings of the 2018 International Conference on Supercomputing, 2018.
29. L. Zhang and P. N. Suganthan, *Benchmarking ensemble classifiers with novel co-trained kernel ridge regression and random vector functional link ensembles [Research Frontier]*, IEEE Computational Intelligence Magazine, vol. 12, no. 4, pp. 61–72, 2017.
30. S. Zhang, Q. Hu, Z. Xie, and J. Mi, *Kernel ridge regression for general noise model with its application*, Neurocomputing, vol. 149, pp. 836–846, 2015.
31. Y. Zhang, J. Duchi, and M. Wainwright, *Divide and conquer kernel ridge regression*, in Conference on Learning Theory, 2013.
32. Z. Zhang, G. Dai, C. Xu, and M. I. Jordan, *Regularized discriminant analysis, ridge regression and beyond*, The Journal of Machine Learning Research, vol. 11, pp. 2199–2228, 2010.
33. H. Zhu, T. M. Martin, L. Ye, A. Sedykh, D. M. Young, and A. Tropsha, *Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure*, Chemical Research in Toxicology, vol. 22, no. 12, pp. 1913–1921, 2009.
34. Mahmood, S., & Algamal, Z. Y. *Kernel ridge regression improving based on golden eagle optimization algorithm for multi-class classification* Statistics, Optimization Information Computing, vol.15, no.1, p. 354–371, 2026.
35. Salih, A. M., Algamal, Z., & Khaleel, M. A. *A new ridge-type estimator for the gamma regression model* , Iraqi Journal for Computer Science and Mathematics, vol . 5 , no .1 , p. 85-98.
36. Algamal, Z., & Ali, H. M. *An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression* , Electronic Journal of Applied Statistical Analysis, vol . 10 , no . 1 , 242-256 , 2017.
37. Qasim, M. K., Algamal, Z. Y., Ali, H. M. *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 and QSAR in Environmental Research, vol. 29, no.(7), p.517-527, 2018
38. Algamal, Z. Y., Lee, M. H. *Applying penalized binary logistic regression with correlation based elastic net for variables selection*, Journal of Modern Applied Statistical Methods, vol. 14, no.(1), p.15, 2015
39. Al-Taweel, Y., Algamal, Z. *Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model*, Periodicals of Engineering and Natural Sciences, vol. 8, no.(1), p.248-255, 2020
40. Al-Fakih, A. M., Algamal, Z. Y., Qasim, M. K. *An improved opposition-based crow search algorithm for biodegradable material classification* SAR and QSAR in Environmental Research, vol.33, no.5, p. 403-415, 2022.
41. Naziyah, A. A., & Algamal, Z. Y. *Jackknifed Liu-type estimator in Poisson regression model* Journal of the Iranian Statistical Society, vol.19, no.1, p. 21-37, 2020.
42. Al-Fakih, A. M., Qasim, M. K., Algamal, Z. Y., Alharthi, A. M., & Zainal-Abidin, M. H. *QSAR classification model for diverse series of antifungal agents based on binary coyote optimization algorithm* Statistics, SAR and QSAR in Environmental Research, vol.34, no.4, p. 285-298, 2023.