ARTICLE IN PRESS

OPEN ACCESS

Journal of Multidisciplinary Applied Natural Science Vol. xx No. xx (xxxx) https://doi.org/10.47352/jmans.2774-3047.360

Check for updates

Research Article

Variable Selection in Kernel Ridge Regression based on Sparrow Search Algorithm with Application QSAR Modeling

Zainab Modhfer Ali Al-Shabaki and Zakariya Yahya Algamal*

Received : November 20, 2025 | Revised : January 3, 2026 | Accepted : January 8, 2026 | Online : March 1, 2026

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 quantitative structure-activity relationship (QSAR) modeling. This study investigates improved binary sparrow bird search algorithm (BSSA) 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 variable selection. Results demonstrate that all BSSA variants significantly outperform KRR in terms of mean squared error (MSE) and coefficient of determination. The quadratic-BSSA (Q-BSSA) variant consistently achieved the best predictive performance, reducing MSE by up to 30% and increasing the coefficient of determination to values above 0.95 on several datasets while selecting the fewest variables, reflecting effective and parsimonious variable selection. Furthermore, BSSA variants substantially decreased the computational time required for model training compared to KRR, with Q-BSSA exhibiting the lowest runtime across datasets. Statistical validation using the Wilcoxon signed-rank test confirmed that all BSSA 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, sparrow search algorithm, transfer function, variable selection

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 [1]-[3]. 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][5]. 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 [6]. 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 [7][8].

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 [9]-[11]. 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 [12]-[14].

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

Please cite this article as: Z. M. A. Al-Shabaki and Z. Y. Algamal. (2026). "Variable Selection in Kernel Ridge Regression based on Sparrow Search Algorithm with Application QSAR Modeling". *Journal of Multidisciplinary Applied Natural Science*. https://doi.org/10.47352/jmans.2774-3047.360

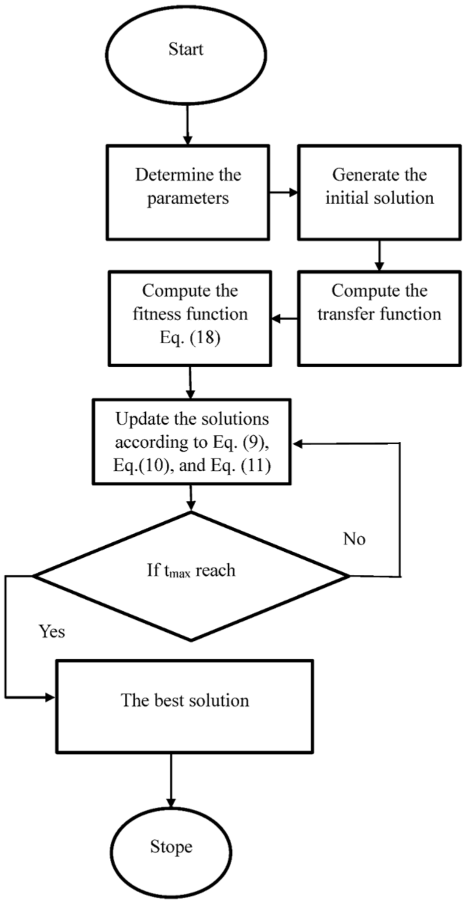

Figure 1. The flowchart of the proposed approach.

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

imbalance-bearing datasets [15]. Zhang and Suganthan [16] presented a co-trained KRR (co-KRR) to improve the generalization capabilities of the KRR model. Rakesh and Suganthan [17] Suggested a collection of KRR models for solving classification challenges. KRR is advantageous due to its cost-effectiveness and efficiency in knowledge-finding applications [18]. 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 [19].

Saunders et al. [20] introduced the concept of kernel ridge regression. We use KRR, 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 [21] [22].

Variable 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 [23]-[26]. 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 [27]-[30]. 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

Table 2. Prediction evaluation criteria for the training dataset.

Datasets	Methods	Selected variables	MSE_{train}	Q^2_{int}
CPD	KRR	-	3.208	0.864
	S-BSSA	25	2.671	0.881
	H-BSSA	22	2.508	0.889
	Q-BSSA	17	1.988	0.951
	Z-BSSA	21	2.369	0.907
COX2	KRR	-	4.315	0.878
	S-BSSA	31	3.778	0.895
	H-BSSA	28	3.615	0.903
	Q-BSSA	19	3.095	0.965
	Z-BSSA	26	3.476	0.921
FDA	KRR	-	3.666	0.885
	S-BSSA	27	3.129	0.902
	H-BSSA	24	2.966	0.911
	Q-BSSA	17	2.446	0.972
	Z-BSSA	20	2.827	0.928
Screen_U251	KRR	-	5.589	0.838
	S-BSSA	36	5.052	0.855
	H-BSSA	31	4.889	0.863
	Q-BSSA	23	4.369	0.925
	Z-BSSA	27	4.751	0.881
EPAFHM	KRR	-	6.034	0.825
	S-BSSA	44	5.497	0.842
	H-BSSA	41	5.334	0.851
	Q-BSSA	30	4.814	0.912
	Z-BSSA	39	5.195	0.868
DHFR	KRR	-	4.246	0.895
	S-BSSA	33	3.709	0.912
	H-BSSA	29	3.546	0.920
	Q-BSSA	20	3.026	0.982
	Z-BSSA	28	3.407	0.938
CAS_N6512	KRR	-	4.429	0.886
	S-BSSA	29	3.892	0.903
	H-BSSA	26	3.729	0.911
	Q-BSSA	18	3.209	0.973
	Z-BSSA	24	3.592	0.929

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.

In recent literature, there are several variable selections using meta-heuristic algorithms in QSAR modeling. Among them are: binary crow search algorithm [6][31], binary gravitational search algorithm [32], binary pigeon optimization algorithm [4], binary slime mould algorithm [33], and binary coyote optimization algorithm [2].

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 [34]. Its goal is to reveal a hidden linear pattern within the initial data [20][35]. Kernel ridge regression, an enhanced form of linear ridge regression that incorporates kernel tricks, can estimate nonlinear mappings [18][36]-[38].

The KRR employs the kernel technique in conjunction with ridge regressions [39][40]. 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 [41][42].

The dataset consists of M training data pairs, $(x_1,t_1),(x_2,t_2), \ldots, (x_M,t_M)$, where A represents 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 Equation (1) [43].

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

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

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

In Equation (2), λ represents a regularization parameter, which is usually determined using cross-validation method. The user must modify this parameter, which is a positive constant. It serves as a penalty for the squared error. The value of λ 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 Equation (3):

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

By differentiating G_{KRR} with respect to α and $(t_i - \alpha \bullet \phi(x_i))$, and equating the resulting β equations to zero, we obtain the output weight vector α as Equation (4).

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

Adding a positive I/λ value to the diagonal of a $\phi'\phi$ matrix results in a stable solution for the KRR's output weight vector α. We find that this stable solution has good generalization performance. In addition to ϕ, the matrix's rows include the mapped samples $\phi(x_i)$, where I is an identity matrix of dimension M x M. Furthermore Equation (5):

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

Consequently, we derive the explicit expression for the dual weight β solution in Equation (6).

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

We construct the kernel matrix K by obtaining its entries through a specific process in Equation (7).

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

In the end, the new test sample x_i yields the expected result in Equation (8).

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

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

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

Table 3. Prediction evaluation criteria for the testing dataset.

Datasets	**Methods**	MSE_{test}	Q^2_{ext}
CPD	KRR	3.886	0.850
	S-BSSA	3.349	0.867
	H-BSSA	3.186	0.875
	Q-BSSA	2.666	0.937
	Z-BSSA	3.047	0.893
COX2	KRR	4.993	0.864
	S-BSSA	4.456	0.881
	H-BSSA	4.293	0.889
	Q-BSSA	3.773	0.951
	Z-BSSA	4.154	0.907
FDA	KRR	4.344	0.871
	S-BSSA	3.807	0.888
	H-BSSA	3.644	0.897
	Q-BSSA	3.124	0.958
	Z-BSSA	3.505	0.914
Screen_U251	KRR	6.267	0.824
	S-BSSA	5.730	0.841
	H-BSSA	5.567	0.849
	Q-BSSA	5.047	0.911
	Z-BSSA	5.429	0.867
EPAFHM	KRR	6.712	0.811
	S-BSSA	6.175	0.828
	H-BSSA	6.012	0.837
	Q-BSSA	5.492	0.898
	Z-BSSA	5.873	0.854
DHFR	KRR	4.924	0.881
	S-BSSA	4.387	0.898
	H-BSSA	4.224	0.906
	Q-BSSA	3.704	0.968
	Z-BSSA	4.085	0.924
CAS_N6512	KRR	5.107	0.872
	S-BSSA	4.570	0.889
	H-BSSA	4.407	0.897
	Q-BSSA	3.887	0.959
	Z-BSSA	4.270	0.915

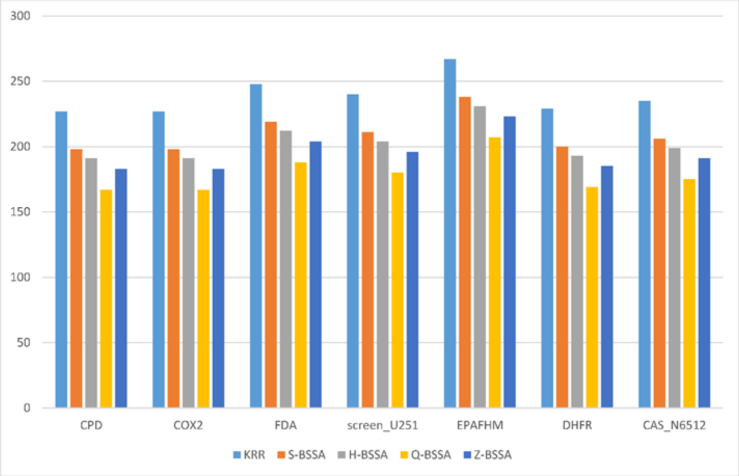

Figure 2. Computational time of the used methods across datasets.

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. SPARROW SEARCH ALGORITHM (SSA)

The SSA is a swarm intelligence optimization algorithm inspired by the foraging and anti-predation behavior of sparrows. It simulates the intelligent group behavior of sparrows in searching for food and avoiding predators [44]. SSA involves two primary types of sparrows: producers (leaders who search for food) and scroungers (followers who track producers to find food), with certain sparrows acting as scouts or explorers. The algorithm iteratively updates individuals' positions in the search space to find optimal solutions for complex optimization problems [45]. Mathematically, the SSA updates the positions of sparrows based on their roles and environmental conditions using several equations representing their search and avoidance behavior:

3.1. Position Update for Producers

$$y_{k.j}^{t+1}=\begin{cases} y_{k.j}^{t}\times\exp\left(\dfrac{-k}{\alpha\times iter}\right) & \mathrm{R}_2<ST \\ y_{k.j}^{t}+qL & \mathrm{R}_2\geq ST \end{cases} \tag{9}$$

Where $y^{t+1}{}_{k.j}$ indicates the individual position of the k^{th} sparrow on the j^{th} dimension in the $(t+1)^{th}$ generation; $y^{t}{}_{k.j}$ describes the position of the k^{th} sparrow on the j^{th} dimension in the t^{th} generation; α is randomly generated from [0,1]; iter is the total number of generations; R_2 is the alarm value that belongs to [0,1]; ST is the safety threshold generated from [0.5,1]; denotes a random digit following the standard distribution and L expresses the matrix of 1*dim , where dim indicates the dimension of the dataset [46].

3.2. Position Update for Scroungers

$$y_{k.j}^{t+1}=\begin{cases} q\times\exp\left(\dfrac{y_{worst}^{t}-x_{k.j}^{t}}{k^2}\right) & \mathrm{k}>\dfrac{n}{2} \\ y_{best}^{t+1}+\dfrac{1}{D}\sum\limits_{D=1}^{D}\left(rand\times\left|y_{k.j}^{t}-y_{best}^{t+1}\right|\right) & \mathrm{k}\leq\dfrac{n}{2} \end{cases} \tag{10}$$

Where $y^{t}{}_{worst}$ indicates the sparrow location having the worst fitness at the t^{th} generation and $y^{t}{}_{best}$ denotes the location of the producer with the best fitness in the $(t+1)^{th}$generation.

3.3. Alarm or Anti-predation Behavior Update

$$y_{k.j}^{t+1}=\begin{cases} y_{best}^{t}+\beta\left|y_{k.j}^{t}-y_{best}^{t}\right| & f_k>f_b \\ y_{k.j}^{t}+h\left(\dfrac{y_{k.j}^{t}-y_{worst}^{t}}{f_k-f_w+\varepsilon}\right) & f_k=f_b \end{cases} \tag{11}$$

Where $y^{t}{}_{best}$ represents the global optimal position in the t^{th} generation; β represents the step size control factor ,which follows the normal distribution ; h represents the move direction of sparrow individuals ,which is generated from[-1,1], and it is also a factor of the step size control; $\mathcal{E}$ is constant used to avoid division be zero errors; f_k indicates the fitness of the k^{th} sparrow; and the best and worst global are represented by f_b and f_w respectively.

The SSA algorithm iteratively updates sparrow positions by these rules, balancing exploration (searching new areas) and exploitation (refining solutions near the current best). The fitness of each individual is evaluated at each step, and individuals move towards better positions based on fitness values to converge to optimal or near-optimal solutions. SSA has demonstrated advantages such as simple structure, fewer parameters, quick convergence, and strong global search capabilities, making it suitable for continuous and discrete optimization problems, including variable selection, function optimization, and engineering.

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 [47]. Therefore, selection of variables is needed 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 variable selection paradigms into the kernelized paradigm [48]. 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 variable selection with the model training procedure, such as adding scarifying 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 [49][50]. 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 [51]. 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 variable selection. This renders them especially valuable to high-dimensional data

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

Pairwise comparison	ρ - Value
Q-BSSA vs KRR	0.0001
S-BSSA vs KRR	0.0013
H-BSSA vs KRR	0.0012
Z-BSSA vs KRR	0.0006

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

Pairwise comparison	ρ - Value
Q-BSSA vs KRR	0.0003
S-BSSA vs KRR	0.0017
H-BSSA vs KRR	0.0015
Z-BSSA vs KRR	0.0008

Table 6. The p-values for the Wilcoxon signed-rank test of the used methods result for Screen_U251 dataset.

Pairwise comparison	ρ - Value
Q-BSSA vs KRR	0.0008
S-BSSA vs KRR	0.0024
H-BSSA vs KRR	0.0019
Z-BSSA vs KRR	0.0011

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

Pairwise comparison	ρ - Value
Q-BSSA vs KRR	0.0008
S-BSSA vs KRR	0.0019
H-BSSA vs KRR	0.0017
Z-BSSA vs KRR	0.0011

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 SSA is put forward to resolve the perpetual optimization problems. Nevertheless, the optimization is not a continuous problem to conduct the variable selection. Binary SSA (BSSA) is modified to do selection of variables [52]. As opposed to SSA, the value of the position in BSSA is binary where value 1 implies the relevance of the variable and 0 the opposite. The dimension of individual members of SSA is the number of original variables in the model in variable selection.

To construct the binary version of the SSA. The S-shaped transfer function (sigmoidal function) (S-BSSA), the hyperbolic tan (V-shaped) function (H-BSSA), the quadratic transfer function (Q-BSSA), and the Z-shaped transfer function (Z-BSSA) were employed to design to construct the binary search space. These functions are defined as, respectively,

$$\text{S-BSSA}\left(y_{k.j}^{t+1}\right)=\frac{1}{1+e^{-y_{k.j}^{t+1}}} \tag{12}$$

$$\text{H-BSSA}\left(y_{k.j}^{t+1}\right)=\left|\tanh\left(y_{k.j}^{t+1}\right)\right| \tag{13}$$

$$\text{Q-BSSA}(y_{k.j}^{t+1})=\begin{cases}\left(\dfrac{y_{k.j}^{t+1}}{0.5\,y_{k.j,\max}^{t+1}}\right)^{1/2}, & if\ y_{k.j}^{t+1} < 0.5\,y_{k.j,\max}^{t+1}\\ 1, & \text{otherwise}\end{cases} \tag{14}$$

$$\text{Z-BSSA}(y_{k.j}^{t+1})=\sqrt{1-2^{y_{k.j}^{t+1}}} \tag{15}$$

The new position of sparrow of Eq. (12) and Eq. (13) are indicated as

$$\text{S-BSSA}(y_{k.j}^{t+1})=\begin{cases}0 & if\ \ r_1 < \text{S-BSSA}(y_{k.j}^{t+1})\\ 1 & if\ \ r_1 \geq \text{S-BSSA}(y_{k.j}^{t+1})\end{cases} \tag{16}$$

$$\text{H-BSSA}\left(y_{k.j}^{t+1}\right)=\begin{cases}\left(y_{k.j}^{t}\right)^{-1} & if\ \ r_2 < \text{H-BSSA}\left(y_{k.j}^{t+1}\right)\\ y_{k.j}^{t} & if\ \ r_2 \geq \text{H-BSSA}\left(y_{k.j}^{t+1}\right)\end{cases} \tag{17}$$

Where r_1 and r_2 are randomly generated from [0,1].

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

The population size $N_{sparrow}$ = 30 and the maximum number of iterations is $t_{\max}$ = 5000 . The value of λ ϵ {0.05.6} and the kernel parameter is selected using the cross-validation method.

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. The proposed approach flowchart is depicted in Figure 1.

5. RESULTS AND DISCUSSIONS

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 [3][4][53]. Table 1 listed the used data description.

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 Equations (19) and (20), respectively.

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

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

Furthermore, the test dataset was used to validate the model by computing the following criteria in Equations (21) and (22);

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

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

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% training and 30% testing. To achieve statistically significant values, this division is reiterated 25 times. All the resulted are obtained using R program. Therefore, the statistical indicators are gathered according to the general abilities and final outcomes .This part utilizes the radial basis function [43]. Tables 2 and 3 present the findings for each technique and dataset.

From Table 2, in terms of the number of selected variables, each BSSA 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-BSSA consistently selects the fewest variables across all datasets, ranging roughly from 17 variables to 30 variables, indicating a strong feature-selection capability. The S-BSSA, H-BSSA, and Z-BSSA 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 2 clearly seen that all BSSA variants improve the MSE_{train}, showing the benefit of selecting informative variables. The lowest MSE_{train} is consistently achieved by the Q-BSSA variant, which also selects the fewest variables. This suggests more efficient and effective variable selection leads to better generalization.

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

Pairwise comparison	ρ - Value
Q-BSSA vs KRR	0.0004
S-BSSA vs KRR	0.0014
H-BSSA vs KRR	0.0011
Z-BSSA vs KRR	0.0008

Table 9. The p-values for the Wilcoxon signed-rank test of the used methods result for CAS_N6512 dataset.

Pairwise comparison	ρ - Value
Q-BSSA vs KRR	0.0006
S-BSSA vs KRR	0.0017
H-BSSA vs KRR	0.0014
Z-BSSA vs KRR	0.0009

This is because the quadratic transfer functions map continuous positions to binary values while providing higher exploitation near boundaries, achieving probabilities closer to 1 or to 0 at extremes compared to S-BSSA, H-BSSA, and Z-BSSA. For example, in CPD dataset, KRR has a higher MSE_{train} of 3.208 and without variable selection. Q-BSSA reduces MSE_{train} to 1.988 with only 17 selected variables, indicating substantial performance gain through compact selected features. Comparing among BSSA variants, across all datasets, Q-BSSA yields the smallest MSE_{train}, indicating it selects an optimal subset of variables leading to the most accurate predictions. On the other hand, S-BSSA and H-BSSA show moderate MSE_{train} improvement. These variants select more variables than Q-BSSA but fewer than KRR. Their MSE_{train} tends to be higher than Q-BSSA but substantially lower than KRR without variable selection. This suggests a trade-off where S-BSSA and H-BSSA offer good but less aggressive variable reduction than Q-BSSA, leading to slightly larger errors. While Z-BSSA performs better than S-BSSA and H-BSSA but not as well as Q-BSSA. Z-BSSA sits in the middle in terms of MSE_{train}, usually selecting more variables than Q-BSSA but fewer than S-BSSA or H-BSSA. Its MSE_{train} values are close to but still above those of Q-BSSA, showing good but not optimal error minimization.

Among the BSSA 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-BSSA 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-BSSA excels at finding a compact yet highly predictive feature subset. Further, Z-BSSA also performs well but with slightly lower (Q^2_{int}) than Q-BSSA. Z-BSSA 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-BSSA and S-BSSA show progressively lower (Q^2_{int}) compared to Q-BSSA and Z-BSSA. These results indicate the important role of BSSA 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 BSSA 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-BSSA consistently achieves the best performance with the lowest MSE_{test} and highest (Q^2_{ext}) across datasets. This signifies that Q-BSSA'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-BSSA follows, showing competitive but slightly inferior performance compared to Q-BSSA. This suggests its search heuristics also achieve strong variable selection but possibly with less refinement or convergence efficiency. On the other hand, H-BSSA and S-BSSA represent intermediate and baseline BSSA variants, respectively, showing moderate gains over KRR but not matching the more advanced Q-BSSA and Z-BSSA variants in minimizing error or boosting explanatory power.

Figure 2 illustrates the computational time (in seconds) required by the KRR and its variants with different BSSA variable selection methods across seven datasets. KRR requires significantly more time than BSSA 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, BSSA variants consistently reduce computational time. S-BSSA, H-BSSA, Q-BSSA, Z-BSSA show markedly lower computational times compared to KRR, with Q-BSSA 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 BSSA variants, we conducted a pairwise comparison between the proposed method and each competing method to further validate the effectiveness of the BSSA 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 [54]-[56]. 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-BSSA, S-BSSA, H-BSSA, and Z-BSSA approach offers statistically significant performance improvements over KRR. These results rigorously demonstrate the value of BSSA variable selection in enhancing kernel-based regression modeling.

6. CONCLUSIONS

This study systematically evaluated the integration of BSSA variants with KRR for variable selection in high-dimensional QSAR modeling. The results across seven diverse biopharmaceutical datasets demonstrate that all BSSA variants, notably the Q-BSSA, substantially improve predictive performance, parsimony, and computational efficiency relative to standard KRR. Q-BSSA 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 17 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 BSSA 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 BSSA frameworks that incorporate adaptive learning strategies, such SSA with particle swarm optimization (PSO) or SSA-Genetic Algorithm, 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.

AUTHOR INFORMATION

Corresponding Author

Zakariya Yahya Algamal — Department of Statistics and Informatics, University of Mosul, Mosul-41002 (Iraq);
orcid.org/0000-0002-0229-7958
Email: zakariya.algamal@uomosul.edu.iq

Author

Zainab Modhfer Ali Al-Shabaki — Department of Statistics and Informatics, University of Mosul, Mosul-41002 (Iraq);
orcid.org/0009-0005-8353-0481

Author Contributions

All authors contributed equally.

Conflicts of Interest

The authors declare no conflict of interest.

ACKNOWLEDGEMENT

The authors would like to express their appreciation for all the support from the College of Computer Science and Mathematics, University of Mosul.

DECLARATION OF GENERATIVE AI

Not applicable.

REFERENCES

[1] H. Zhu, T. M. Martin, L. Ye, A. Sedykh, D. M. Young, and A. Tropsha. (2009). "Quantitative Structure-Activity Relationship

Modeling of Rat Acute Toxicity by Oral Exposure". *Chemical Research in Toxicology.* **22** (12): 1913-1921. 10.1021/tx900189p.

[2] A. Al-Fakih, M. Qasim, Z. Algamal, A. Alharthi, and M. Zainal-Abidin. (2023). "QSAR Classification Model for Diverse Series of Antifungal Agents Based on Binary Coyote Optimization Algorithm". *SAR and QSAR in Environmental Research.* **34** (4): 285-298. 10.1080/1062936X.2023.2208374.

[3] A. Al-Fakih, Z. Algamal, M. Lee, M. Aziz, and H. Ali. (2019). "QSAR Classification Model for Diverse Series of Antifungal Agents Based on Improved Binary Differential Search Algorithm". *SAR and QSAR in Environmental Research.* **30** (2): 131-143. 10.1080/1062936X.2019.1568298.

[4] Z. Y. Algamal, M. K. Qasim, M. H. Lee, and H. T. M. Ali. (2020). "High-Dimensional QSAR/QSPR Classification Modeling Based on Improving Pigeon Optimization Algorithm". *Chemometrics and Intelligent Laboratory Systems.* **206** : 104170. 10.1016/j.chemolab.2020.104170.

[5] Z. Y. Algamal and M. H. Lee. (2017). "A Novel Molecular Descriptor Selection Method in QSAR Classification Model Based on Weighted Penalized Logistic Regression". *Journal of Chemometrics.* **31** (10). 10.1002/cem.2915.

[6] A. Al-Fakih, Z. Algamal, and M. Qasim. (2022). "An Improved Opposition-Based Crow Search Algorithm for Biodegradable Material Classification". *SAR and QSAR in Environmental Research.* **33** (5): 403-415. 10.1080/1062936X.2022.2064546.

[7] B. Y. S. Li, L. F. Yeung, and K. T. Ko. (2015). "Indefinite Kernel Ridge Regression and Its Application on QSAR Modelling". *Neurocomputing.* **158** : 127-133. 10.1016/j.neucom.2015.01.060.

[8] J. Verma, V. M. Khedkar, and E. C. Coutinho. (2010). "3D-QSAR in Drug Design: A Review". *Current Topics in Medicinal Chemistry.* **10** (1): 95-115. 10.2174/156802610790232260.

[9] Z. Y. Algamal, M. H. Lee, A. M. Al-Fakih, and M. Aziz. (2017). "High-Dimensional QSAR Classification Model for Anti-Hepatitis C Virus Activity of Thiourea Derivatives Based on the Sparse Logistic Regression Model with a Bridge Penalty". *Journal of Chemometrics.* **31** (6). 10.1002/cem.2889.

[10] Z. Algamal, M. Qasim, and H. Ali. (2017). "A QSAR Classification Model for Neuraminidase Inhibitors of Influenza A Viruses (H1N1) Based on Weighted Penalized Support Vector Machine". *SAR and QSAR in Environmental Research.* **28** (5): 415-426. 10.1080/1062936X.2017.1326402.

[11] Z. Y. Algamal, M. H. Lee, and A. M. Al-Fakih. (2016). "High-Dimensional Quantitative Structure-Activity Relationship Modeling of Influenza Neuraminidase A/PR/8/34 (H1N1) Inhibitors Based on a Two-Stage Adaptive Penalized Rank Regression". *Journal of Chemometrics.* **30** (2): 50-57. 10.1002/cem.2766.

[12] Z. Y. Algamal, M. H. Lee, A. M. Al-Fakih, and M. Aziz. (2015). "High-Dimensional QSAR Prediction of Anticancer Potency of Imidazo[4,5-b]Pyridine Derivatives Using Adjusted Adaptive LASSO". *Journal of Chemometrics.* **29** (10): 547-556. 10.1002/cem.2741.

[13] M. Qasim, Z. Algamal, and H. M. Ali. (2018). "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.* **29** (7): 517-527. 10.1080/1062936X.2018.1491414.

[14] A. Tropsha. (2010). "Best Practices for QSAR Model Development, Validation, and Exploitation". *Molecular Informatics.* **29** (6-7): 476-488. 10.1002/minf.201000061.

[15] B. B. Hazarika and D. Gupta. (2023). "Affinity-Based Fuzzy Kernel Ridge Regression Classifier for Binary Class Imbalance Learning". *Engineering Applications of Artificial Intelligence.* **117** : 105544. 10.1016/j.engappai.2022.105544.

[16] L. Zhang and P. N. Suganthan. (2017).

"Benchmarking Ensemble Classifiers with Novel Co-Trained Kernel Ridge Regression and Random Vector Functional Link Ensembles". *IEEE Computational Intelligence Magazine.* **12** (4): 61-72. 10.1109/MCI.2017.2742867.

[17] K. Rakesh and P. N. Suganthan. (2017). "An Ensemble of Kernel Ridge Regression for Multi-Class Classification". *Procedia Computer Science.* **108** : 375-383. 10.1016/j.procs.2017.05.109.

[18] S. An, W. Liu, and S. Venkatesh. (2007). "Face Recognition Using Kernel Ridge Regression". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007).* 10.1109/CVPR.2007.383105

[19] C. Yan, Y. Li, W. Liu, M. Li, J. Chen, and L. Wang. (2020). "An Artificial Bee Colony-Based Kernel Ridge Regression for Automobile Insurance Fraud Identification". *Neurocomputing.* **393** : 115-125. 10.1016/j.neucom.2017.12.072.

[20] C. Saunders, A. Gammerman, and V. Vovk. (1998). "Ridge Regression Learning Algorithm in Dual Variables". *Proceedings of the 15th International Conference on Machine Learning (ICML 1998).*

[21] P. Exterkate, P. J. Groenen, C. Heij, and D. van Dijk. (2016). "Nonlinear Forecasting with Many Predictors Using Kernel Ridge Regression". *International Journal of Forecasting.* **32** (3): 736-753. 10.1016/j.ijforecast.2015.11.017.

[22] P. S. Kumar, S. Laha, and L. Kumaraswamidhas. (2023). "Assessment of Rolling Element Bearing Degradation Based on Dynamic Time Warping, Kernel Ridge Regression and Support Vector Regression". *Applied Acoustics.* **208** : 109389. 10.1016/j.apacoust.2023.109389.

[23] Y. Özüpak, F. Alpsalaz, and E. Aslan. (2025). "Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction". *Water, Air, & Soil Pollution.* **236** (7): 464. 10.1007/s11270-025-08122-8.

[24] E. Aslan. (2024). "Prediction and Comparative Analysis of Emissions from Gas Turbines Using Random Search Optimization and Different Machine Learning-Based Algorithms". *Bulletin of the Polish Academy of Sciences: Technical Sciences.* **72** (6): e151956. 10.24425/bpasts.2024.151956.

[25] M. A. Kahya, S. A. Altamir, and Z. Y. Algamal. (2020). "Improving Whale Optimization Algorithm for Feature Selection with a Time-Varying Transfer Function". *Numerical Algebra, Control and Optimization.* **11** (1): 87-98. 10.3934/naco.2020017.

[26] Z. Algamal, M. Lee, A. Al-Fakih, and M. Aziz. (2016). "High-Dimensional QSAR Modelling Using Penalized Linear Regression Model with L1/2-Norm". *SAR and QSAR in Environmental Research.* **27** (9): 703-719. 10.1080/1062936X.2016.1228696.

[27] E. Aslan and Y. Özüpak. (2024). "Detection of Road Extraction from Satellite Images with Deep Learning Method". *Cluster Computing.* **28** (1): 72. 10.1007/s10586-024-04880-y.

[28] Y. Andu, M. H. Lee, and Z. Y. Algamal. (2020). "Variable Selection of Yearly High-Dimension Stock Market Price Using Ordered Homogenous Pursuit LASSO". *AIP Conference Proceedings*. 10.1063/5.0019161

[29] N. A. Al-Thanoon, O. S. Qasim, and Z. Y. Algamal. (2025). "Improving the Binary Tree Growth Algorithm with Fuzzy Mutual Information for Feature Selection". 10.1063/5.0286531

[30] Z. A. Almishlih, O. S. Qasim, and Z. Y. Algamal. (2025). "Binary Arithmetic Optimization Algorithm Using a New Transfer Function for Fusion Modeling". *Fusion: Practice and Applications.* **18** (2): 157-168. 10.54216/FPA.180212.

[31] G. Y. I. Abdallh and Z. Y. Algamal. (2020). "A QSAR Classification Model of Skin Sensitization Potential Based on Improving Binary Crow Search Algorithm". *Electronic Journal of Applied Statistical Analysis.* **13** (1).

[32] A. Al-Fakih, Z. Algamal, M. Lee, M. Aziz, and H. Ali. (2019). "A QSAR Model for Predicting Antidiabetic Activity of

Dipeptidyl Peptidase-IV Inhibitors by Enhanced Binary Gravitational Search Algorithm". *SAR and QSAR in Environmental Research.* **30** (6): 403-416. 10.1080/1062936X.2019.1607899.

[33] A. A. Ewees, M. A. Al-Qaness, L. Abualigah, Z. Y. Algamal, D. Oliva, D. Yousri, and M. A. Elaziz. (2023). "Enhanced Feature Selection Technique Using Slime Mould Algorithm: A Case Study on Chemical Data". *Neural Computing and Applications.* **35** (4): 3307-3324. 10.1007/s00521-022-07852-8.

[34] S. Geman, E. Bienenstock, and R. Doursat. (1992). "Neural Networks and the Bias/Variance Dilemma". *Neural Computation.* **4** (1): 1-58. 10.1162/neco.1992.4.1.1.

[35] Z. Zhang, G. Dai, C. Xu, and M. I. Jordan. (2010). "Regularized Discriminant Analysis, Ridge Regression and Beyond". *Journal of Machine Learning Research.* **11** : 2199-2228.

[36] C. Orsenigo and C. Vercellis. (2012). "Kernel Ridge Regression for Out-of-Sample Mapping in Supervised Manifold Learning". *Expert Systems with Applications.* **39** (9): 7757-7762. 10.1016/j.eswa.2012.01.060.

[37] S. Zhang, Q. Hu, Z. Xie, and J. Mi. (2015). "Kernel Ridge Regression for General Noise Model with Its Application". *Neurocomputing.* **149** : 836-846. 10.1016/j.neucom.2014.07.051.

[38] S. Mahmood and Z. Algamal. (2025). "Kernel Ridge Regression Improving Based on Golden Eagle Optimization Algorithm for Multi-Class Classification". *Statistics, Optimization & Information Computing.* **15** (1): 354-371. 10.19139/soic-2310-5070-2569.

[39] Y. Zhang, J. Duchi, and M. Wainwright. (2013). "Divide and Conquer Kernel Ridge Regression". *Proceedings of the Conference on Learning Theory.*

[40] F. A. Y. Al-Taie, O. S. Qasim, and Z. Y. Algamal. (2024). "Improving Kernel Semi-Parametric Regression Model Based on a Bat Optimization Algorithm". 10.1063/5.0202591

[41] M. Ali, R. Prasad, Y. Xiang, and Z. M. Yaseen. (2020). "Complete Ensemble Empirical Mode Decomposition Hybridized with Random Forest and Kernel Ridge Regression Model for Monthly Rainfall Forecasts". *Journal of Hydrology.* **584** : 124647. 10.1016/j.jhydrol.2020.124647.

[42] Y. You, J. Demmel, C. J. Hsieh, and R. Vuduc. (2018). "Accurate, Fast and Scalable Kernel Ridge Regression on Parallel and Distributed Systems". *Proceedings of the 2018 International Conference on Supercomputing*. 10.1145/3205289.3205290

[43] J. Naik, R. Bisoi, and P. Dash. (2018). "Prediction Interval Forecasting of Wind Speed and Wind Power Using Modes Decomposition Based Low Rank Multi-Kernel Ridge Regression". *Renewable Energy.* **129** : 357-383. 10.1016/j.renene.2018.05.031.

[44] J. Xue and B. Shen. (2020). "A Novel Swarm Intelligence Optimization Approach: Sparrow Search Algorithm". *Systems Science & Control Engineering.* **8** (1): 22-34. 10.1080/21642583.2019.1708830.

[45] L. Sun, S. Si, W. Ding, J. Xu, and Y. Zhang. (2023). "BSSFS: Binary Sparrow Search Algorithm for Feature Selection". *International Journal of Machine Learning and Cybernetics.* **14** (8): 2633-2657. 10.1007/s13042-023-01788-8.

[46] L. Lian. (2024). "An Improved Sparrow Search Algorithm Using Chaotic Opposition-Based Learning and Hybrid Updating Rules". *Concurrency and Computation: Practice and Experience.* **36** (14). 10.1002/cpe.8101.

[47] S. W. Mahmood, G. T. Basheer, and Z. Y. Algamal. (2025). "Quantitative Structure-Activity Relationship Modeling Based on Improving Kernel Ridge Regression". *Journal of Chemometrics.* **39** (5). 10.1002/cem.70027.

[48] S. W. Mahmood, G. T. Basheer, and Z. Y. Algamal. (2025). "Improving Kernel Ridge Regression for Medical Data Classification Based on Meta-Heuristic Algorithms". *Kuwait Journal of Science.* **52** (3): 100408. 10.1016/j.kjs.2025.100408.

[49] Z. Y. Algamal. (2018). "A New Method for Choosing the Biasing Parameter in Ridge Estimator for Generalized Linear Model".

Chemometrics and Intelligent Laboratory Systems. **183** : 96-101. 10.1016/j.chemolab.2018.10.014.

[50] M. A. Kahya, S. A. Altamir, and Z. Y. Algamal. (2019). "Improving Firefly Algorithm-Based Logistic Regression for Feature Selection". *Journal of Interdisciplinary Mathematics.* **22** (8): 1577-1581. 10.1080/09720502.2019.1706861.

[51] Z. A. Almishlih, O. S. Qasim, and Z. Y. Algamal. (2025). "Design and Evaluation of a New Tent-Shaped Transfer Function Using the Polar Lights Optimizer Algorithm for Feature Selection". *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska.* **15** (2): 27-31. 10.35784/iapgos.6802.

[52] A. G. Gad, K. M. Sallam, R. K. Chakrabortty, M. J. Ryan, and A. A. Abohany. (2022). "An Improved Binary Sparrow Search Algorithm for Feature Selection in Data Classification". *Neural Computing and Applications.* **34** (18): 15705-15752. 10.1007/s00521-022-07203-7.

[53] M. Eklund, U. Norinder, S. Boyer, and L. Carlsson. (2012). "Benchmarking Variable Selection in QSAR". *Molecular Informatics.* **31** (2): 173-179. 10.1002/minf.201100142.

[54] E. Baş and G. Yildizdan. (2023). "Enhanced coati optimization algorithm for big data optimization problem". *Neural Processing Letters.* **55** (8): 10131-10199.

[55] J. Derrac, S. García, D. Molina, and F. Herrera. (2011). "A Practical Tutorial on the Use of Nonparametric Statistical Tests as a Methodology for Comparing Evolutionary and Swarm Intelligence Algorithms". *Swarm and Evolutionary Computation.* **1** (1): 3-18. 10.1016/j.swevo.2011.02.002.

[56] J. Ma, D. Xia, H. Guo, Y. Wang, X. Niu, Z. Liu, and S. Jiang. (2022). "Metaheuristic-based support vector regression for landslide displacement prediction: A comparative study". *Landslides.* **19** (10): 2489-2511.