Hyper Parameters Selection by using Metaheuristic Algorithm for Improving Support Vector Regression

Salih Mooaed Al Bakal [1*],Saifuldeen Dheyauldeen Alrefaee [2] ,Zakariya Yahya Algamal [3]

1. University of Mosul, Iraq, salih.mooaed@uomosul.edu.iq
2. University of Mosul, Iraq, saifldeen.alrefaee@uomosul.edu.iq
3. University of Mosul, Iraq, zakariya.algamal@uomosul.edu.iq

Article Info

Page Number: 1407 - 1425

Publication Issue:

Vol 71 No. 3 (2022)

Abstract

Support vector regression (SVR) is one of the most accuracy machine learning techniques. However, the performance of this method depends on the selection of its hyperparameters. In most of the research, Grid search algorithm was used to select these hyperparameters and consider this algorithm among the best algorithms. In this paper, crow search algorithm is used to select the best combination of hyperparameters and then use them in the SVR method. Experimental results, obtained by running on three datasets, show that the crow search algorithm is performance of SVR was better when using the grid search algorithm in terms of prediction. In addition, it was demonstrated that the crow's search algorithm was better for its speed to obtain the best set of parameters. Alongside, By experimental results the crow search algorithm improving confirm the efficiency of the proposed algorithm in improving the prediction performance and computational time compared to other nature-inspired algorithms.

Article History

Article Received*: 12 January 2022*

Revised*: 25 February 2022*

Accepted*: 20 April 2022*

Publication*: 09 June 2022*

Keywords: Support vector machine, Support vector regression, Crow search Algorithm.

1. Introduction

The SVR algorithm was developed in the 1960s in Russia by [1]. This algorithm is based on the statistical learning theory, and over the past three decades this theory was developed by [1-3]. Support vector machines consist of two main classes: Support Vector Classification (SVC) and Support Vector Regression (SVR). Both types passed through several stages until they settled on the model they are in today.[2] presented the Maximal Margin Classifier. Then a kernelized version was introduced using the kernel trick by[4]. This was followed by the Soft Margin version presented by[5]. Finally, the final version of Kernelized Soft Margin was stabilized, which is the combination of the three previous stages.

SVM has been used in many areas.[6] proposed a new Procedure for selecting a descriptor for the QSAR classification model by adding a new weight within the L1 criterion. Experimental results from the classification of neuraminidase inhibitors for influenza A (H1N1) viruses clarified that the proposed Procedure in the QSAR classification model works effectively and competitively compared to other penalized methods. Also in classification, and in Quantitative structure–activity relationship (QSAR) classification modeling, [7] suggested a two-stage classification approach is proposed by merge the minimum redundancy maximum relevancy criterion with the sparse support vector machine, and they show that the proposed method is able to effectively outperform other sparse alternatives methods. Also explain [8] a new selection of descriptors that truly affect biological activity and a QSAR classification model estimation method are proposed by merge the sparse logistic regression model with a bridge penalty for classifying the anti-hepatitis C virus activity of thiourea derivatives. In cancer classification, [9] used a firefly algorithm, that is a metaheuristic continuous algorithm to determine the parameter in PSVM with SCAD penalty.

The version of the regression SVM introduced by [10] instead of classification is called Support Vector Regression (SVR). Like the SVM classification, this regression model includes the hyperparameter and the Kernel trick. Then the SVM regression method was developed by[11].

The SVR method has recently been widely used by many researchers. [12] presented a study on RFID technology for positioning in indoor and indoor environments, in which the SVR method was used to improve the positioning accuracy of the LANDMARC algorithm. Whereas, Gaussian-Kalaman filter was used for noise resolution, positioning was predicted with SVR, thus improving positioning accuracy compared to traditional RFID positioning methods. [13] also presented a study on wastage in electricity in terms of generation, transmission and distribution, and accordingly analyzed the load data in electricity and predicted prices for consumers. The study included two models, the first is to predict the load of electricity and the second is to forecast the price of electricity. He improved the SVR method for predicting electricity prices, and this improvement was done by using a web search algorithm to select the best combination of hyperparameters that he used in the SVR method. They explained that the improved SVR method performs better at predicting electricity prices. [14] presented a study on the evaluation of shallow lands slip, where the study was applied to an area in Turkey - Trabzon Province (northeastern Turkey) where the SVR method was used, and it was compared with the logistic regression model widely used in this field. They demonstrated that the SVR method performed better at identifying potential landslide zones in that region.[15] also presented a study on maximum power point tracking of Solar energy. They used SVR to improves the traditional perturb and observation (P&O) method. They demonstrated that the SVR method improved tracking accuracy while shortening the convergence time. They attributed this to the fast and accurate prediction obtained through the SVR algorithm.

The SVM regression method has proven efficient in many areas. However, its efficiency depends on the selection of its hyperparameters. Many researchers have tried to define a certain pattern in selecting these hyperparameters, but there has not yet been agreement on a unified method for determining these hyperparameters. Many researchers justify this because the performance of the SVR method depends on the data being studied in addition to other factors [16]. In a study on chemical sensor arrays presented by [17], it was clarified that the performance of SVR depends greatly on the selection of its hyperparameters, and he used the grid search algorithm to determine the best combination of the salient parameters. Some researchers also explained that the performance of the SVR method depends on two issues, the first is how to assign hyperparameters to a specific data set, and the second issue is to increase the prediction performance by identifying features from a given input space [18]. [19] also showed that the overall performance of Regression SVM was better using the Vapnik's ε -insensitive loss function compared to the regression using 'least-modulus' loss ($\varepsilon = 0$). Note that the SVR method was used with the parameters being selected directly from the training data. Also, [20] proposed the opposition-based learning Harris hawks optimization algorithm (HHOA-OBL) to optimize the hyperparameters of the v-SVR with embedding the feature selection at the same time Recently,[21] used the black hole algorithm to optimize the hyperparameters of SVR. . [22] proposed a new hybrid algorithm and proposed particle swarm optimization to determine the tuning factor in PSVM.

Through the foregoing, it is noticed that the SVR method is one of the good methods, but its performance depends on the selection of its hyperparameters, and it is also noticed that in most of the research and studies the grid search algorithm was used as a method to choose the best combination of the hyperparameters and used in the SVR method, and this algorithm was compared and considered from The best algorithms for determining the best combination of salient parameters.

Algorithms inspired by nature have been used in several studies and the efficiency of these algorithms has been proven with an optimal solution. These algorithms were used in selecting the best parameters. Where it was estimated several parameters of the non-linear Hirota-Satsuma coupled KdV system by using a hybrid between the Modified Adomian decomposition method and the Firefly Algorithm [23].

Also presented [24] a hybrid algorithm between the statistical dependence and the binary dragonfly algorithm is presented, whereby the feature selection method in discrete space is modeled as a binary-based optimization algorithm, guiding binary dragonfly algorithm and using the accuracy of the k-nearest neighbors classifier on the dataset to verify it in the chosen fitness function.

Also, to estimate the three-parameter gamma distribution, a particle swarm optimization (PSO) is proposed and then to estimate the reliability and hazard functions. Where concluded that the proposed estimation method is considerably consistent in estimation compared to the maximum likelihood estimation method, in terms of log likelihood and mean time to failure [25].

In addition,[26] used the Harris hawks optimization algorithm (HHOA) to optimize the hyperparameter of v-SVR, They show that the proposed algorithm gave better results than other methods, in terms of prediction, number of selected features, and runtime. [27] used grasshopper optimization algorithm o optimize the hyperparameters of the SVR.

In this paper, a crow search algorithm is used to find the best combination of hyperparameters. it has been proven that another algorithm has surpassed the grid search algorithm, which is the crow search algorithm, and this was done by applying to several types of data with high dimensions, and it was proved that this algorithm is more efficient than the grid search algorithm. The comparison was made by the standard square root of the mean squares of error, and the comparison was made by the time taken for each algorithm to select the best combination of the hyperparameters that were used in the SVR method.

The rest of the paper is organized as follows: Section 1 included an introduction to SVM, its classifications, and researchers who worked on the class SVR. The second section introduces some details of the SVR method and the effect of its hyperparameters on its performance. Section 3 illustrates the crow's search algorithm and the steps to accomplish it. Section 4 Explanation Section 5 outlines the applications and the findings. Section 6 clarifies the conclusions.

2. Support vector Regression

As illustrated in Section 1, the SVM method is divided into two parts: the first is the classification method, which is symbolized by SVC, and the second is the regression method, which is symbolized by SVR. The SVR method is a support vector regression method. Whereas, SVR is one of the machine learning techniques, as it reduces experimental risks and reduces model complexity at the same time by building a predictive linear model [17] .

The SVR approach is based on a loss function that ignores errors within the epsilon intensive range. As it relies on Kernel functions, it is considered a non-parametric method. SVR and SVC treat training data as follows: SVR relies only on a subset of training data because the cost function ignores any training data close (within a threshold) to predict the model, whereas SVC relies on only a subset of training data, because the cost function is not concerned with the training points that fall within the margin range[28].

The SVR approach can be described by the training data set which is:

$$T = \{(x,y_i), i = 1,\ldots,N\}$$

Where N is the size of the training data, and $x_i \in \mathbb{R}^d$represents the feature vector. Also, $y_i \in \mathbb{R}$represents the target value. From the function:

$$f(w,x) = w.\varphi(x) + b \qquad (1)$$

SVR constructs a linear model after mapping training data to a high-dimension feature space in order to predict y by the function $f(w,x)$. Where w is the vector of weights, and $\varphi$is a high dimensional feature space that mapped the input space vector x, and that is a bias [29]. Using the ϵ-insensitive loss function the error prediction is calculated, where the loss function is represented by the following formula[17]:

$$\mathbf{L}(y,f(w,x) = \begin{cases} 0 & if|y_i - \mathrm{f}(w,x)| \leq \varepsilon \\ |y_i - \mathrm{f}(w,x)| - \varepsilon & otherwise \end{cases} \Bigg\} \qquad (2)$$

The aim of SVR is to find a value of no more than ε for each of the points of $x$resulting from the deviation of the function y from y, and at the same time this function is as flat as possible. This is done by minimizing the $\|w\|$ in order to maximize the value of the margin, and this is what the soft margin method adopts. Thus the problem can be formulated as a convex optimization problem[17]:

$$Minimize : \frac{1}{2}\|\mathrm{w}\|^2 \tag{3}$$

$$Subject\ to: \begin{Bmatrix} y_i - \mathrm{f}(w,x) - b \leq \varepsilon \\ \mathrm{f}(w,x) + b - y_i \leq \varepsilon \end{Bmatrix}$$

The deflection of training samples outside the epsilon intensive range is measured by introducing non-negative slack variablesξ_i^*, ξ_i [29]. In other words, these slack variables were introduced in order to withstand some training that fall outside the ε-insensitive tube. That is, it is a term that shows the difference between the output y values and the estimated value. The optimization problem becomes as follows [17]:

$$Minimize: \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi,\xi_i^*) \tag{4}$$

$$Subject\ to: \begin{Bmatrix} y_i - \mathrm{f}(w,x) - b \leq \varepsilon + \xi_i^* \\ \mathrm{f}(w,x) + b - y_i \leq \varepsilon + \xi \\ \xi_i^*,\xi \geq 0 \end{Bmatrix}$$

The trade-off between experimental error and model complexity is through the parameter C, where C is the penalty parameter and it is a regularization parameter [29]. ε is the tolerance error, and ξ_i^*, ξ are positive slack variables that represent the lowest and highest excess deviation from the value ε [29]. After transform optimization problem is into a dual problem, and by using the Lagrange equation and the Karush–Kuhn–Tucker (KKT) conditions. Thus, the solution is in the following equation, called the so-called support vector expansion function [16]:

$$\mathrm{f}(x) = \sum_{i=1}^{k}(\alpha_i - \alpha_i^*)\,\mathrm{K}(x_i,x_j) + b \tag{5}$$

Where α_i , α_i^* are Lagrange multipliers. And $\mathrm{K}(x_i,x_j)$ that is the Kernel function. Through the Kernel function, the calculation of the optimization problem is always in the input space rather than the mapping space, that is, a mapping space with a high dimension can be used [17]. In other words, the flatness of the function depends on the number of support vectors and does not depend on the dimensions of the input space[16]. In the case of linear regression, it is:

$$\mathrm{K}(x_i,x_j) = (x_i,x_j) \tag{6}$$

In non-linear regression, the radial basis function (RBF) is used [18], which will be used in this paper:

$$\mathrm{K}(x_i,x_j) = \exp(-\gamma\|x_i - x_j\|^2) \tag{7}$$

The performance and accuracy of SVR depends on good preparation of its meta-parameters which are c , ε , as well as dependent on the parameters of the Kernel functions[17].

The parameter C defines the trade-off between model complexity (flatness) and the degree to which deviations greater than ε are tolerated, where the goal is to minimize the empirical risk only, without concern for the model's complexity in formulating the optimization when the C value is very large (infinity) [19]. The higher value of the C parameter allows more support vectors to be selected, which increases the complexity of the model [17].

The width of the epsilon intensive region is controlled by the parameter ε, since the number of support vectors used in the formation of the regression function depends on the value of ε. Where less support vectors are chosen when large value of parameter ε is chosen. Also, the function will be more flat, and the estimates will be more stable[19].

The parameter of the Kernel function shown in the equation (7) represented by σ, the effect of this parameter is strong on the model. The high value of this parameter leads to the model being restricted excessively. If the σ value is very small, then the effect of the support vectors will be a very effective effect, and the complexity of the model cannot be controlled even through the parameter C, and this leads to the overfitting [17].

Thus, it is noted that each hyperparameter has a different effect on the SVR method, and an appropriate method must be chosen in order to choose a good combination of these hyperparameter in order to obtain good performance for the SVR method.

3. Crow search algorithm

The crow search algorithm is a recent algorithm proposed by [30] Which follows the smart behavior of crows in terms of hiding food and stealing others. Recent studies have shown that crows are birds that have a high Intelligence, as they have large brains in relation to the size of their body, and they have passed a Mirror's test of self-identification, in addition to having the ability to use the necessary tools during her daily activity[31], as well It has the ability to hide food in a specific place and refer to it at a later time within a period of time, which may be a later Season[30]. The basis of the work of crow search algorithm is that crows fly in a swarm and within a specific environment, that these crows hide their food in excess of their need in specific places and keep these places in their memories in order to restore this food when needed, these crows follow each other in order to steal food.

Many researchers have explained the crow's search algorithms and their mechanism of work. In this section of this research, this algorithm will be clarified as follows:

Assuming that crows fly within one swarm, let N be crows, and in the search area of d dimensions. In the crow search algorithm, there are several parameters that must be defined as follows, fl It represents the length of the flight, Ap the possibility that a particular crow may be perceived as being followed, $m_t(i)$The memory in which crow i stores his food in the t-th iteration. $x_t(i,j)$ Represents the location of crow i in the tth iteration. Where $i = 1,...,N$, $j = 1,...,d$, $t = 1,...,t_{\max}$.

The crow search algorithm works by tracing a specific crow to another crow, which is randomly chosen to find the location of the hidden food for that crow. This process has two probabilities:

The first probability is that the crow followed does not know that there is another crow followed, and therefore the first crow will know the location of the second crow's food, which leads to the update of the location of the first crow in its memory since the first location before the update is imposed randomly, the second probability it is represented by the knowledge and awareness of the second crow that it is followed by the first crow, which leads to the fact that the second crow will mislead the first crow so that the location of the hidden food belonging to the second crow is not known, and these two possibilities can be expressed in the following mathematical formula[32]:

$$x^{i,t+1} = \begin{cases} x^{i,t} + r_j \times fl^{i,t} \times (m^{i,t} - x^{i,t}) & r_j \geq AP^{i,t} \\ a\ random\ position & otherwise \end{cases} \tag{8}$$

Whereas, r_j is a random number from the uniform distribution located between [0,1]. And that $x^{i,t}$is the current position of crow i before updating on iteration t, and that $x^{i,t+1}$is the new position of crow i after updating on iteration t + 1. On this basis, crow positions i can be arranged within the area of a search environment with a d dimensions at the iteration t as follows[33]

$$x^{i,t} = [x_1^{i,t}, x_2^{i,t}, \ldots, x_d^{i,t}] \tag{9}$$

Also, a vector for crow's memory i can be arranged, in which the best place to hide food is kept within the area of a search environment with d dimensions when it is iteration t as follows[31] :

$$m^{i,t} = [m_1^{i,t}, m_2^{i,t}, \ldots, m_d^{i,t}] \tag{10}$$

Several researchers described the crow search algorithm as sequential steps, which can be illustrated as follows[34]:

1- Randomly create site for swarm of crows, let N be in the search area.

2-Crows site evaluation.

3-nitialize the memory of each crow.

4-If the upper iteration limit is not reached, do

5-Let it be i = 1: N and for all crows

6-One of the crows is randomly chosen, and the crow is j

7-Generate $r \in [0,1]$

8-If it is $r \geq Ap$then

9-$x^{i,t+1} = x^{i,t} + r_j \times fl^{i,t} \times (m^{i,t} - x^{i,t}) r_j \geq AP^{i,t}$

10-Else

11-$x^{i,t+1} = a\ random\ position$

12-check how feasible value of $x$is

13-calculate the new position $x^{i,t+1}$of each crow

14-update the memory of each crow

15-end

4. Crow search algorithm in determining SVR hyperparameters

In this section, the CSA-SVR model will be described in order to determine the best hyperparameters that can be used in the SVR method. The mechanism proposed in this paper is the process of initializing parameters using the crow search algorithm, and then these parameters are used in the SVR method. As specific initial values are specified for the parameters of the crow's search algorithm within a specified range, where the flight length parameter is determined, the probability parameter of the crow's perception that it is being followed, the crow's memory parameter of where the food is hidden, and finally the location of the food, which is initially imposed randomly. And a range of values are specified for each of the penalty parameter C, the epsilon parameter ε and the parameter of the Kernel functionσ, where the Radial basis function will be used. Then these parameters are entered in a frequency within a predetermined maximum. The best combination of these hyperparameters is obtained for use to train SVR to obtain an accurate prediction.

This mechanism can be described in the following steps:

1- The data is initially divided into the training group and the test group.
2- Determination of values for the parameters of the crow's search algorithm: the parameter of the flight length, the parameter of the probability of the crow's perception that it is being followed, the parameter of the crow's memory of where the food is hidden, and the location of the hiding of the food, which is initially imposed randomly. The size of the swarm, N crows, which is the size of the population.
3- Determine the maximum iteration, tmax=100.
4- Start with the value of the penalty parameter C within a range of values, a maximum and a minimum, from distribution as uniform [0,1].
5- Starting with the value of the epsilon parameter ε within a range of values, a maximum and a minimum, from distribution as uniform [0,1]

6- Starting with the value of the Kernel function parameter σ within a range of values, maximum and minimum, from distribution as uniform [0,1]
7- Set the time taken to select the best combination of hyperparameters using the crow's search algorithm.
8- Start applying the crow search algorithm and according to the values of the specified parameters.
9- Go to step 9 if the maximum number of iterations is exceeded, otherwise return to step 4.
10- After determining the best combination of the hyperparameters from step 9, it is used to train the SVR.
11- The model obtained from step 10 will be used for prediction.

5. Data Sets

Three datasets from the UCI Repository were used, which are described in Table 1. Both grid search and crow search algorithms were applied with the same settings for the values of the hyperparameters, and after obtaining the best combination of hyperparameters from each algorithm, they were used in the SVR method for each algorithm with computing the square root of the mean squares error of the training and test data. The time taken was calculated for each algorithm. This process was performed 10 times, each time 100 iterations were used to test the proposed method.

In order to compare the two algorithms accurately, Table 2 shows the average of ten times that were performed for each group of data.

Table 1: Description of the datasets used.

Dataset	Dataset Name	#Samples	#Features
Dataset 1	Melting Dataset	60	635
Dataset 2	Flu Dataset	108	2436
Dataset 3	Ipc50 Dataset	65	2541

6. Results and discussion

In this section, a model was trained using the SVR method when entering a combination of Hyperparameters obtained using the crow search algorithm, and then this performance was compared with the performance of the same method, but using another combination of Hyperparameters obtained using the grid search algorithm. The SVR method and the two algorithms were applied to three sets of data, and the criterion square root mean squares error was used for the training data and the test data to compare the two algorithms. The time taken for each algorithm was also calculated for comparison between them. A range of values for each Hyperparameters was used and entered into each algorithm in order to test all possible combinations between these Hyperparameters, and then choose the best combination to apply to the SVR method.

The parameters of the crow's search algorithm were determined based on previous studies, where the[35] in its study provided parameter values through which the best results could be obtained. Where he determined in his study the effect of the maximum repetition on the optimal solution, as he explained that the optimal solution was obtained at the 250 iteration, while the high iterations the algorithm was consuming a long time to obtain the optimal solution. He also showed that the value of the parameter of the best flight length was at 2. The parameter of Awareness probability was the best value at 0.1, and the best value for the size of the swarm (population) was 50. Also [31] specify the parameters of the crow's search algorithm by using the value of 2 for the flight length parameter, 0.1 for the Awareness probability parameter, and 30 for the swarm size (population). Also,[36] specify the parameters of the crow's algorithm by using the value of 2 for the flight length parameter, 0.1 for the Awareness probability parameter, and 20 for the swarm size (population). Determine [37] the parameters of the crow's search algorithm by using the value 2 for the flight length parameter, 0.1 for the Awareness probability parameter, and 20 for the swarm size (population).

In this study, the same values used in previous studies were used, where we drew the RMSE values for each repeat and it was noticed that it stabilized at approximately 60 repetitions, so the maximum frequency was determined to be 100. A range of values (0.1-2) was also used for the flight length parameter. 0.1 for the Awareness probability parameter, and 50 for the (population) swarm size.

The range used for the Hyperparameters c was within the period (1-10) with an increment of 1, and the range for the parameter ε was within the period (0.1-1) with an increment of 0.1, while the parameter of the Kernel function was within the period (0.1-1) with an increment of 0.1, where the kernels radial basis function (RBF) was used. Three data sets (Melting Data, Flu Data and Ipc50 Data) are used and they are of different dimensions. Each set of data two algorithms, grid search and crow search, were applied, with the same range specified for the hyperparameters. After each algorithm was selected for a specific combination of these hyperparameters, there were applied to the SVR method.

Tables 2, it is noticed that the value of the RMSE Train criterion and the value of the RMSE Test criterion were less for the crow search algorithm compared with the grid search algorithm for the three data sets: Melting, Flu and IPC50.

Table 2: Ten times (on average) based on training and test datasets.

Datasets	Algorithm	RMSE Train	RMSE Test
Dataset 1	Grid	36.7320± 14.16	51.1393± 4.89
	Crow	30.7852± 19.35	50.7240± 4.86
Dataset 2	Grid	0.5530± 0.124	1.1798± 0.076
	Crow	0.4271± 0.15	1.1597± 0.069
Dataset 3	Grid	0.2837± 0.242	1.1387± 0.369
	Crow	0.2061± 0.0869	1.1250± 0.084

The length of time to find hyperparameters is plotted for each algorithm. Therefore, from Figure 1, the crow search algorithm was distinguished from the Grid search algorithm because the time it took to select the best combination of the hyperparameters was less than the grid search algorithm, and this can be seen in Figure 1.

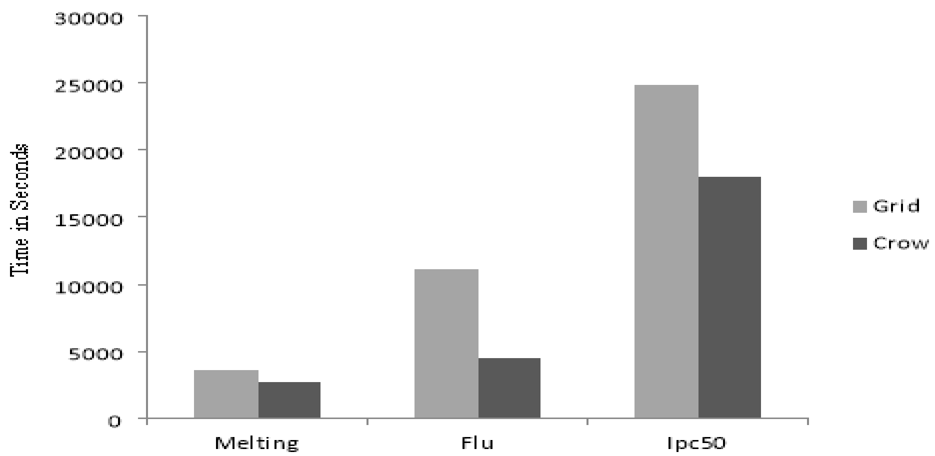

Figure 1: The time taken in seconds to select the best combination of hyperparameters

It can also be observed from Table 3 the average time spent for the two algorithms, as the crow search algorithm has outperformed the Grid search algorithm by having less time than the Grid search algorithm.

Table 3: Average time taken for the two algorithms, per minute.

Datasets	Algorithm	The average time in second
Dataset 1	Grid	3548.131
	Crow	2676.517
Dataset 2	Grid	11070.6
	Crow	4491.177
Dataset 3	Grid	24855.2
	Crow	17981.18

Also, a significant test of the time taken to find the best combination of the Hyperparameters of the two algorithms was performed. The test was significant and the result was shown in Table 4.

Table 4: Time t-test for two algorithms

Datasets	t-test	p-Value
Dataset 1	11.026	0.000
Dataset 2	10.34	0.000
Dataset 3	6.493	0.000

7. Conclusions

In this paper a proposal is made to use the crow search algorithm to select the best combination of hyperparameters and then use it in the SVR method instead of using the grid search algorithm. Comparison of the two algorithms was done using the RMSE Train and RMSE Test criteria. The results showed that the crow search algorithm had a lower value for the RMSE Train and RMSE Test criteria compared to the grid search algorithm. The process was performed 10 times and the crow search algorithm had lower values for the RMSE Train and RMSE Test criteria in a part of these times, while the grid search algorithm had lower values for the RMSE Train and RMSE Test criteria in another part of these times. Therefore, in order to definitively determine the best algorithm, the average values obtained from these ten times and for the three sets of data were found. It is through these values, the crow search algorithm outperformed the another nature-inspired algorithms by having the lowest value of the RMSE Train and RMSE Test criteria.

Moreover, the crow search algorithm was distinguished from the grid search algorithm in that the time it spent selecting the best combination of hyperparameters was less than the grid search algorithm. In addition, the time significance test was positive, meaning that there was a significant difference between the algorithms with respect to time.

Thus, it can be said that the crow search algorithm has outperformed the grid search algorithm in terms of choosing the best combination of hyperparameters that can be used in the SVR method; it also takes less time compared to nature-inspired algorithms. It is clear that the proposed algorithm is a promising algorithm for improving SVR performance due to its ability to reach the solution faster and with the best solution. This algorithm can be applied to other real applications.

Acknowledgments

This work was supported by the Mosul University, College of Computer Sciences and Mathematics, Republic of Iraq.

References

1. Vapnik, V., *Pattern recognition using generalized portrait method.* Automation and remote control, 1963. **24**: p. 774-780.
2. Vapnik, V. and A. Chervonenkis, *Theory of Pattern Recognition* 1974, Nauka, Moscow (German Translation: W. Wapnik & A. Tscherwonenkis, Theorie ….
3. Vapnik, V., *Estimation of dependences based on empirical data*. 2006: Springer Science & Business Media.
4. Boser, B.E., I.M. Guyon, and V.N. Vapnik. *A training algorithm for optimal margin classifiers*. in *Proceedings of the fifth annual workshop on Computational learning theory*. 1992.
5. Cortes-Vapnik, *Support-Vector Networks.* Machine Learning,, 1995. **20,**: p. 273-297.
6. Algamal, Z., M. Qasim, and H. Ali, *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, 2017. **28**(5): p. 415-426.
7. Qasim, M., 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, 2018. **29**(7): p. 517-527.
8. Algamal, Z.Y., et al., *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, 2017. **31**(6): p. e2889.
9. Al-Thanoon, N.A., O.S. Qasim, and Z.Y. Algamal, *Tuning parameter estimation in SCAD-support vector machine using firefly algorithm with application in gene selection and cancer classification.* Computers in biology and medicine, 2018. **103**: p. 262-268.
10. Vapnik, V., *The Nature of Statistical Learning Theory.* Springer, New York., 1995.
11. Vapnik, V., S.E. Golowich, and A.J. Smola. *Support vector method for function approximation, regression estimation and signal processing*. in *Advances in neural information processing systems*. 1997.

12. Xu, H., et al., *An RFID Indoor Positioning Algorithm Based on Support Vector Regression.* Sensors (Basel), 2018. **18**(5).

13. Zahid, M., et al., *Electricity Price and Load Forecasting using Enhanced Convolutional Neural Network and Enhanced Support Vector Regression in Smart Grids.* Electronics, 2019. **8**(2): p. 122.

14. Kavzoglu, T., E.K. Sahin, and I. Colkesen, *Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression.* Landslides, 2013. **11**(3): p. 425-439.

15. Tan, B., et al., *Improved perturb and observation method based on support vector regression.* Energies, 2019. **12**(6): p. 1151.

16. Li, S., H. Fang, and X. Liu, *Parameter optimization of support vector regression based on sine cosine algorithm.* Expert systems with Applications, 2018. **91**: p. 63-77.

17. Laref, R., et al., *On the optimization of the support vector machine regression hyperparameters setting for gas sensors array applications.* Chemometrics and Intelligent Laboratory Systems, 2019. **184**: p. 22-27.

18. Xu, S., et al., *An improved variable selection method for support vector regression in NIR spectral modeling.* Journal of Process Control, 2018. **67**: p. 83-93.

19. Cherkassky, V. and Y. Ma, *Practical selection of SVM parameters and noise estimation for SVM regression.* Neural networks, 2004. **17**(1): p. 113-126.

20. Ismael, O.M., O.S. Qasim, and Z.Y. Algamal, *Improving Harris hawks optimization algorithm for hyperparameters estimation and feature selection in v-support vector regression based on opposition-based learning.* Journal of Chemometrics, 2020. **34**(11): p. e3311.

21. Alrefaee, S.D., S.M. Al Bakal, and Z.Y. Algamal, *Hyperparameters optimization of support vector regression using black hole algorithm.* International Journal of Nonlinear Analysis and Applications, 2022. **13**(1): p. 3441-3450.

22. Al-Thanoon, N.A., O.S. Qasim, and Z.Y. Algamal, *A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics.* Chemometrics and Intelligent Laboratory Systems, 2019. **184**: p. 142-152.

23. Qasim, O.S., K.A. Abed, and A.F. Qasim, *Optimal Parameters for Nonlinear Hirota-Satsuma Coupled KdV System by Using Hybrid Firefly Algorithm with Modified Adomian Decomposition.* Journal of Mathematical and Fundamental Sciences, 2020. **52**(3): p. 339-352.
24. Qasim, O.S., M.S. Mahmoud, and F.M. Hasan, *Hybrid Binary Dragonfly Optimization Algorithm with Statistical Dependence for Feature Selection.* International Journal of Mathematical, Engineering and Management Sciences, 2020. **5**(6): p. 1420-1428.
25. Mahmood, S.W. and Z.Y. Algamal, *Reliability Estimation of Three Parameters Gamma Distribution via Particle Swarm Optimization.* Thailand Statistician, 2021. **19**(2): p. 308-316.
26. Ismael, O.M., O.S. Qasim, and Z.Y. Algamal. *A new adaptive algorithm for v-support vector regression with feature selection using Harris hawks optimization algorithm*. in *Journal of Physics: Conference Series*. 2021. IOP Publishing.
27. Algamal, Z.Y., et al., *Improving grasshopper optimization algorithm for hyperparameters estimation and feature selection in support vector regression.* Chemometrics and Intelligent Laboratory Systems, 2021. **208**: p. 104196.
28. Debasish Basak, S.P.a.D.C.P., *Support Vector Regression.* Neural Information Processing, 2007. **11**(10).
29. Nait Amar, M. and N. Zeraibi, *Application of hybrid support vector regression artificial bee colony for prediction of MMP in CO2-EOR process.* Petroleum, 2018.
30. Askarzadeh, A., *A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm.* Computers & structures, 2016. **169**: p. 1-12.
31. De Souza, R.C.T., et al. *A V-shaped binary crow search algorithm for feature selection*. in *2018 IEEE congress on evolutionary computation (CEC)*. 2018. IEEE.
32. Allaoui, M., B. Ahiod, and M. El Yafrani, *A hybrid crow search algorithm for solving the DNA fragment assembly problem.* Expert Systems with Applications, 2018. **102**: p. 44-56.
33. Liu, D., et al., *ELM evaluation model of regional groundwater quality based on the crow search algorithm.* Ecological Indicators, 2017. **81**: p. 302-314.
34. Gupta, D.S., Shirsh. Khanna, Ashish. Ella Hassanien, Aboul. de Albuquerque, Victor Hugo C., *Improved diagnosis of Parkinson's disease using optimized crow search algorithm.* Computers & Electrical Engineering, 2018. **68**: p. 412-424.

35. Abdelaziz, A.Y. and A. Fathy, *A novel approach based on crow search algorithm for optimal selection of conductor size in radial distribution networks.* Engineering Science and Technology, an International Journal, 2017. **20**(2): p. 391-402.

36. Rizk-Allah, R.M., A.E. Hassanien, and S. Bhattacharyya, *Chaotic crow search algorithm for fractional optimization problems.* Applied Soft Computing, 2018. **71**: p. 1161-1175.

37. Diego Oliva a , c., ∗, Salvador Hinojosa b , c , Erik Cuevas c , Gonzalo Pajares b , Omar Avalos c , d , Jorge Gálvez c , d, *Cross entropy based thresholding for magnetic resonance brain images using Crow Search Algorithm.* 2019.