California Bearing Ratio Prediction Based On LSSVR And Multiple Optimizers

Chenyue Ma, Qiang Shao*, and Mengtian Yin

Hebei University of Water Resources and Electric Engineering foundation department, Cangzhou 061001, China

Corresponding author. E-mail: yinhemeilingge@163.com

Received: Feb. 04, 2024; Accepted: Jun. 03, 2024

The California bearing ratio (CBR) as a widespread index is employed in Soil mechanics and geotechnical structures such as bridge abutments, earth dams, highway embankments, and so on. Generally, this index can be determined through laboratory or field tests. However, CBR determining procedure is time and cost-consuming. Therefore, the present study represents different hybrid models established on Least Square Support Vector Regression (LSSVR) combined with three meta-heuristic algorithms (namely, Arithmetic optimization algorithm, Bald eagle search optimization, and Seagull Optimization Algorithm) to estimate CBR values in a way that is cheap and quick to perform, and more accurate in solving real-world problems. 70% of developed hybrid models were allocated to the train stage and remained 30% were defined as testing models. The impression of five inputs (lime sludge percentage, lime percentage, maximum dry density, curing period, and optimum moisture content) on the predicted CBR values were considered. Finally, to assess the accuracy of the two-category created models, a comparison study between predicted and observed results is made by using five statistical indexes. The observations have shown that LSBE (LSSVR combined with Bald Eagle search optimization) model has strong potential for predicting the CBR.

Keywords: California Bearing Ratio; Least Square Support Vector Regression; Arithmetic optimization algorithm; Bald eagle search optimization; Seagull Optimization Algorithm

http://dx.doi.org/10.6180/jase.202505_28(5).0015

1. Introduction

CBR introduced as the ratio that defines correlation of the resistance against the piercing of a piston in a compacted soil with a specific velocity (1.27 mm/min) and in rock with the same circumstances and the same penetration depth [1]. To express differently, CBR is described as the correlation between the applied stress, with fixed moisture range, and the speed regulation to run down a standard penetration piston to outreach the desired depth and the resistance of squashed rock for the same penetration depth value [1, 2]. CBR tests for both unsoaked and soaked conditions of soil can be run in the laboratory or field according to ASTM D 1883 - 99 (2003) and ASTM D 4429-93 (2003), respectively [1, 3, 4]. The determination of the *CBR* in both laboratory and field is time-consuming and costly or the disturbance of samples may reduce the accuracy of these measurements. as well as the incompetent technicians testing the soil samples, thus, the construction work progress is seriously affected [5–7]. Thus, prediction models are an economical and rapid solution to predict the *CBR*. Several previous soundings have made efforts to approximate the CBR based on soil parameters such as maximum dry density (MDD), soil grain size, Atterberg's limits, etc. For instance, to predict the CBR of compacted soil, a correlation was established between the CBR and liquid limit and plasticity index [8]. Others used MDD, optimum moisture content (OMC), and Plasticity Index (PI)

and established a linear model for establishing a correlation between CBR and those soil parameters of cohesive type [9]. Furthermore, a vast number of previous works utilized one-variable and multiple-variable regression approach intending to approximate the CBR considering aforementioned primary parameters [10, 11]. However, all proposed formulas neither have high accuracy in predicted results nor have any generalized solution [1, 6].

Support vector machine (SVM), minimizes an upper limit of the generalization error so it has a strong capability for prediction and classification [12, 13]. Support vector regression (SVR) as a nonlinear regression can relieve the local minima and over-fitting issues thus its solution is globally optimum and enough stable. Least squares support vector machine (LSSVM) was established by Suykens and Vandewalle [14] to reduce the computational complexity. least squares support vector regression (LSSVR) represents one category of LSSVM which by resolving a system of equations instead of quadratic programming improves the speed of calculations significantly so it can be employed as a prediction model in real-world problems [15]. For instance, Rad and Ayubirad [16] used the LSSVR model considering Coupled Simulated Annealing (CSA) to establish a nonlinear correlation between the compressive strength of ultra-high-performance concrete and eight input factors. The training and testing findings were assessed via the coefficient of determination (R^2) and Root Mean Squared Error (RMSE). They proved that $ANNs$ and CSA-LSSVR models are strongly capable of estimating the compressive strength of $UHPC$.

Likewise, several hybrid techniques can get around the drawbacks of linear models in estimating [17, 18]. In order to demonstrate that hybrid models outperform single models in terms of anticipated outcomes accuracy, Xue and Wei [19] evaluated the findings of Least square support vector machined (LSSVM) optimized using genetic algorithm, Adaptive Neuro-Fuzzy Inference System (ANFIS), Fuzzy Inference System (FIS), and Artificial Neural Network (ANN). In another study, Lei et al. [20] evaluated six swarm intelligence optimization techniques, including the seagull optimization algorithm (SOA), which has the capacity of two models without optimization in predicting uniaxial compressive strength, against several hybrid-developed models created by neural networks. Furthermore, other authors improved existing meta-heuristic algorithms such as studies which are done by Agushaka and Ezugwu [21] and Dhiman et al. [22] for arithmetic optimization and Seagull Optimization Algorithms, respectively.

The summary of the major contribution of this research is as follows:

- First, a representation of the dataset including 5 input variables and the target value of CBR is presented.
- Secondly, in four discrete subsections, a description of LSSVR as the main approach and three optimizers along with their formulation has been presented.
- Then, the results of developed hybrid models in two categories of training and testing part were evaluated through five statistical indexes and the results were examined according to the observed test results of CBR expressed as various diagrams.
- Finally, all the results concluded and the model with strong potential in predicting the CBR value is proposed to be used in practical applications.

2. Materials and methodology

Dataset Description

In the current investigation influence of five various input variables on the CBR value was taken into account. The variables are namely: lime sludge percentage (LS (%)), lime percentage (LI (%)), optimum moisture content (OMC (%)), curing period (CP (day)), and maximum dry density $(MDD(g/cc))$. The dataset contains 73 samples [23]. The statistical properties including maximum, minimum, average as well as standard deviation values related to dependent input variables and CBR as the target and independent variable are reported in Table 1. Fig. 1 shows the histogram of used variables.

2.1. Least Square Support Vector Regression (LS-SVR)

Support vector machine (SVM) by Vapnik [13] is a powerful learning approach employed for function approximation, nonlinear classification, and density estimation for nonlinear classification. SVM for regression (SVR) target is to find the linear regression function in higher dimensional space in such a way that the estimated function is as flat as possible, also it deviates the least from training data [16]. In this case, $\{x_i, y_i\}_{i=1,2,\ldots,l}$ are given training dataset, and the regression function is expressed as:

$$f(x) = Q^S x + d \quad \text{with} \quad Q, x \in R^n \& d \in R \tag{1}$$

In most cases, the SVR algorithm transforms the input dataset to a high-dimensional space because of the nonlinear correlation linking the input and output. Suppose that ψ is the mapping function that takes x of input space and maps it to feature space. Therefore, linear predictors in transformed feature space can be formulated as follows:

$$f(x) = Q^S \psi(x) + d \tag{2}$$

Table 1. The statistical measures of the data

Statistical Properties	Dataset Components					
	LI	*LS*	*CP*	*OMC*	*MDD*	CBR
Max.	20	15	1.39	32.65	45	156
Min.	0	0	1.14	13.7	4	2.2
Ave.	5.63	4.931	1.242	24.058	24.109	52.98
St. Dev.	4.029	5.803	0.058	3.88	15.872	34.387

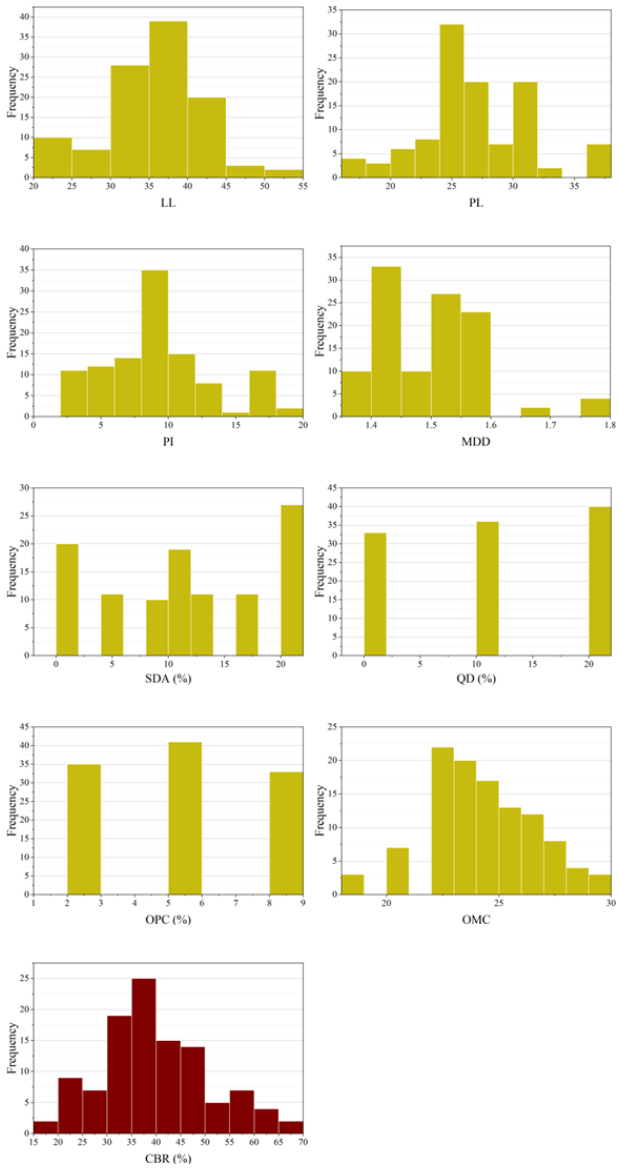

Fig. 1. Histogram of used variables

Here, Q and d show the weight vector of m-dimension and bias term, respectively. To find $f(x)$ in a way that the variance between the target and the function value, for all the training data, becomes less than ε, the ε-SVR has been used. In this case, the convex optimization problem with the following constraints represents the problem of finding the regression function:

$$\min \frac{1}{2}\|Q\|^2 \begin{cases} y_i - Q^S\psi(x_i) - d \le \varepsilon \\ Q^S\psi(x_i) + d - y_i \le \varepsilon \end{cases} \tag{3}$$

Regarding that, the existence of a f function with ε accuracy and the feasibility of convex optimization are not guaranteed. δ and δ_i^* as slack variables have been introduced. Finally, the convex optimization problem can be reformulated as:

$$\min \frac{1}{2}\|Q\|^2 + c\sum_{i=1}^{l}(\delta_i + \delta_i^*) \quad \begin{cases} y_i - Q^S\psi(x_i) - d \le \varepsilon + \delta_i \\ Q^S\psi(x_i) + d - y_i \le \varepsilon + \delta_i^* \\ \delta_i, \delta_i^* \ge 0 \end{cases} \tag{4}$$

Here $c \ge 0$, and it is the trade-off between tolerance limit for deviations greater than ε and the smoothness of the function f.

The $LS-SVM$ algorithm is an upgraded series of SVM, adapting quadratic optimization problems into a set of linear simple equations so it reduces the computational problems [24, 25]. A simplified formulation is achieved by solving the following problem using LSSVR:

$$\min \frac{1}{2}\|Q\|^2 + \frac{\Omega}{2}\sum_{i=1}^{l} e_i^2 \tag{5}$$

Such that $y_i = Q^S\psi(x_i) + d + e_i$,

Ω acts as a regulation parameter and e_i represents the trade-off between the evenness of the function and the minimization of training error. The obtained Lagrangian function is as follows:

$$LF = \frac{1}{2}\|Q\|^2 + \frac{\Omega}{2}\sum_{i=1}^{l} e_i^2 - \sum_{i=1}^{l}\alpha_i\left(y_i - Q^S\psi(x_i) - d - e_i\right) \tag{6}$$

Where, α_i shows the Lagrangian multiplier. Moreover, dual and primal variables must be set to zero for solving the optimization issue.

$$\frac{\partial LF}{\partial Q} = 0 \rightarrow Q = \sum_{i=1}^{l}\alpha_i\psi(x_i) \tag{7}$$

$$\frac{\partial LF}{\partial d} = 0 \rightarrow \sum_{i=1}^{l} \alpha_i = 0 \tag{8}$$

$$\frac{\partial LF}{\partial e_i} = 0 \rightarrow \alpha_i = \Omega e_i \tag{9}$$

$$\frac{\partial LF}{\partial \alpha_i} = 0 \rightarrow Q^S \psi(x_i) + d + e_i - y_i = 0 \tag{10}$$

The Lagrangian function is now modified using Q and e to create the following linear system:

$$\begin{bmatrix} 0 & 1^S \\ 1 & \lambda + \Omega^{-1} I \end{bmatrix} \begin{bmatrix} d \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{11}$$

Where:
$\alpha = [\alpha_1 + \alpha_2, \ldots, \alpha_L]^S$,
$e = [e_1 + e_2, \ldots, e_L]^S$,
$\lambda = VV^S$,
$V = [\psi(x_1), \psi(x_2), \ldots, \psi(x_L)]^S$,
$y = [y_1 + y_2, \ldots, y_L]^S$,
$1 = [1 + 1, \ldots, 1]^S$.

Currently, the matrix λ may be formed as Eq. (12) employing the kernel method, and the answer to Eq. (11) can then be recast as Eq. (13).

$$\lambda_{ij} = \psi(x_i)^S \psi\left(x_j\right)^S = K\left(x_i, x_j\right) \tag{12}$$

$$c = \frac{1^T \left(\lambda + \Omega^{-1} I\right)^{-1} y}{1^T \left(\lambda + \Omega^{-1} I\right)^{-1} I} \tag{13}$$

$$\alpha = \left(\lambda + \Omega^{-1} I\right)^{-1} (y - d1) \tag{14}$$

RBF (v^2) can be chosen as a kernel function corresponding to an infinite-dimensional function space. Moreover, employing achieved d and α, the optimal regression function can be expressed as:

$$f(x) = \sum_{i=1}^{l} \alpha_i K(x_i, x) + d \tag{15}$$

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2v^2}\right) \tag{16}$$

Where, v^2 is the kernel-squared bandwidth.

In the previous research, different optimum values have been introduced for Ω and v^2. For example, Hsu et al. [26] suggested 1 and $1/K$ (K is the number of input patterns) for Ω and v^2, respectively. Aiyer et al. [27] suggested trial and error procedure which can be very time-consuming in the case of Ω and ν^2 estimation while they can have any positive real values. Thus, optimization methods can be used to optimize these parameters for the LSSVR model.

2.2. Arithmetic optimization algorithm (AOA)

Arithmetic originates from number theory which is classified as one of the fundamental sections of modern mathematics. Multiplication, Subtraction, Division, and Addition operators as conventional calculation measures are used to survey the numbers [28]. These basic operators are utilized to present the greatest element subjected to particular measures from some group of solutions. The principal inspiration for the Arithmetic optimization algorithm emerges from utilizing AOs in solving Arithmetic problems. Generally, the optimization process in population-based algorithms including AOA consists of exploration and exploitation phases which refer to extensive coverage of defined search space and accuracy improvement of obtained solutions, respectively [29]. Three main phases of AOA are presented in the subsequent parts.

2.2.1. *Initialization phase*

The beginning stage in the optimization procedure of AOA is a random set of alternative solutions (X). In this case, the greatest alternative in each time iteration is supposed to be the optimum solution or in the range of its neighborhood.

$$X = \begin{bmatrix} x_{1,1} & \cdots & \cdots & x_{1,j} & x_{1,n-1} & x_{1,n} \\ x_{2,1} & \cdots & \cdots & x_{2,j} & \cdots & x_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N-1,1} & \cdots & \cdots & x_{N-1,j} & \cdots & x_{N-1,n} \\ x_{N,1} & \cdots & \cdots & x_{N,j} & x_{N,n-1} & x_{N,n} \end{bmatrix} \tag{17}$$

Before the AOA process starts, one of the exploration or exploitation phases should be selected. So, the Math Optimizer Accelerated (MOA) function which represents the function value at the ith iteration is computed by Eq. (18).

$$\mathrm{MOA}\left(B_{\mathrm{Iter}}\right) = \mathrm{Min} + B\_\mathrm{Iter} \times \left(\frac{\mathrm{Max} - \mathrm{Min}}{M\_\mathrm{Iter}}\right) \tag{18}$$

Where M_Iter is the maximum number of iterations and $1 < B_{\mathrm{I\,Iter}} <$ M_Iter. Max is the maximum amount of the accelerated function, and Min represents the minimum value.

2.2.2. *Exploration phase*

The exploration search mechanism includes mathematical computation via either Division (DO) or Multiplication (MO) operators which got highly distributed values or decisions. Thus, they cannot easily get nearer to the desired outcome and they may just find a solution that is very close to the best possible outcome after several iterations to support the exploitation. Eq. (19) models two main search strategies in the exploration phase and represents position

updating equations:

$$x_{i,j}\left(B\_\text{Iter}+1\right)= \begin{cases} \text{best}\left(x_j\right) \div (MOP+\varepsilon) \times \left(\left(UB_j-LB_j\right) \times \lambda + LB_j\right), r2<0.5 \\ \text{best}\left(x_j\right) \times MOP \times \left(\left(UB_j-LB_j\right) \times \lambda + LB_j\right), \text{otherwise} \end{cases} \tag{19}$$

Here, best $\left(x_j\right)$ is the j th position in the current best solution and $x_{i,j}\left(B_\text{Iter}\right)$is the j th position of the ith solution. ε is a small integer, LB_j and UB_j represent the upper and lower limit of j th location, respectively, λ is a control parameter, and MOP shows the mathematical optimizer probability. Function value of $MOP\left(B\_Iter\right)$ is represented as follows:

$$\text{MOP}\left(B_{\text{Iter}}\right)=1-\frac{B\_I\,\text{Iter}^{1/\beta}}{\text{Iter}^{1/\beta}} \tag{20}$$

Here, β denotes a sensitive parameter that defines the precision of exploitation during the iterations.

2.2.3. *Exploitation phase*

Subtraction (SO) or Addition (AO) operators in the mathematical calculations got highdense results so in the exploitation phase they can easily approach the target after several iterations. Eq. (21) models two main search strategies in the exploitation phase and represents position updating equations. Exploitation operators (SO and AO) evade getting trapped in the local investigation space, which facilitates the procedure for the optimal outcome.

$$x_{i,j}\left(B_{\text{Iter}}+1\right) = \begin{cases} \text{best}\left(x_j\right)-MOP \times \left(\left(UB_j-LB_j\right) \times \lambda + LB_j\right), r3<0.5 \\ \text{best}\left(x_j\right)+MOP \times \left(\left(UB_j-LB_j\right) \times \lambda + LB_j\right), \text{otherwise} \end{cases} \tag{21}$$

2.3. Bald eagle search optimization (BES)

In 2020, BES algorithm was developed by Alsattar et al. [30]. This optimization algorithm mainly inspires by the eagle's intelligent social behavior in hunting [31]. The bald eagle's hunting is splitted up into three phases including selecting space, exploring in space, and swooping. Through these three phases, the bald eagle chooses the space with the higher number of praise, then it starts searching for prey inside the selected space, and finally, it begins swinging from the supreme location [32].

2.3.1. *Selecting-space phase*

In this phase, bald eagles attempt to opt for the best area search space with a high amount of available food. The mathematical definition of this stage is presented as follows:

$$X_{\text{new},i}=X_{\text{best}}+\beta * r\left(X_{\text{mean}}-X_i\right) \tag{22}$$

Where β denotes the element controlling the location alters; r is the randomly chosen number in the range of $[0,1]$.$X_{\text{new},i}$, X_{best}, X_{mean}, and X_i are a new position, the bestselected position so far, the mean gap between all locations of the bald eagles in the population, and the present location of the eagle, respectively.

2.3.2. *Searching-in-space phase*

In this stage, the bald eagle searches in diverse directions within the selected space to find prey. During this stage, it determines the optimum location for hunting and swooping. The mathematical definition of this stage is presented as follows:

$$X_{\text{new},i}=X_i+f(i) * \left(X_i-X_{i+1}\right)+g(i) * \left(X_i-X_{\text{mean}}\right) \tag{23}$$

$$g(i)=\frac{gr(i)}{(\max|gr|)} \tag{24}$$

$$f(i)=\frac{fr(i)}{(\max|fr|)} \tag{25}$$

$$gr(i)=r(i)\cdot\sin(\varphi(i)) \tag{26}$$

$$fr(i)=r(i)\cdot\cos(\varphi(i)) \tag{27}$$

$$\varphi(i)=\beta.\pi.\,\text{rand} \tag{28}$$

$$r(i)=\varphi(i)+S.\,\text{rand} \tag{29}$$

Here, β defines the corner between point search and S defines the number of search attempts. rand is a value in the range of $[0,1]$.

2.3.3. *Swooping phase*

Finally, in this stage, all of the bald eagles begin swinging from the superior location to their predefined prey. The mathematical definition of its behavior in this phase is presented as follows:

$$\begin{aligned} X_{\text{new},i}=\ &\text{rand}.\,X_{\text{best}}+g1(i)\cdot\left(X_i-B1.X_{\text{mean}}\right) \\ &+f1(i)\cdot\left(X_i-B2\cdot X_{\text{best}}\right) \end{aligned} \tag{30}$$

$$g1(i)=\frac{gr(i)}{(\max|gr|)} \tag{31}$$

$$f1(i)=\frac{fr(i)}{(\max|fr|)} \tag{32}$$

$$gr(i)=r(i)\cdot\sin(\varphi(i)) \tag{33}$$

$$fr(i)=r(i)\cdot\cos(\varphi(i)) \tag{34}$$

$$\varphi(i)=\beta.\pi.\,\text{rand} \tag{35}$$

$$r(i)=\varphi(i) \tag{36}$$

Here, $B1, B2\epsilon[1,2]$.

The whole appearance of the BES algorithm is available in Alsattar et al. [30] and the subsequent pseudocode.

Set initial values randomly: X_i for n point;
Calculate the fitness values of the initial point: $F\left(X_i\right)$;
WHILE (the termination criteria have not been fulfilled)
Selecting space

For (each point i in the population)
$X_{new} = X_{best} + \beta * r\left(X_{mean} - X_i\right)$
If $F\left(X_{new}\right) < F\left(X_i\right)$
$X_i = X_{new}$
If $F\left(X_{new}\right) < F\left(X_{best}\right)$
$X_{best} = X_{new}$
End If
End If
End For
Searching in space
For (each point i in the population)
$X_{new} = X_i + f(i) * \left(X_i - X_{i+1}\right) + g(i) * \left(X_i - X_{mean}\right)$
If $F\left(X_{new}\right) < F\left(X_i\right)$
$X_i = X_{new}$
If $F\left(X_{new}\right) < F\left(X_{best}\right)$
$X_{best} = X_{new}$
End If
End If
End For
Swooping
For (each point i in the population)
$X_{new} = rand.\ X_{best} + g1(i) \cdot \left(X_i - B1 \cdot X_{mean}\right) + f1(i) \cdot \left(X_i - B2 \cdot X_{best}\right)$
If $F\left(X_{new}\right) < F\left(X_i\right)$
$X_i = X_{new}$
If $F\left(X_{new}\right) < F\left(X_{best}\right)$
$X_{best} = X_{new}$
End If
End If
End For
Set $k := k + 1$;
END WHILE

2.4. Seagull Optimization Algorithm (SOA)

Laridae, so-called Seagulls, are sea birds that live all over the world [22]. The seagull optimization algorithm (SOA) was developed by Dhiman and Kumar [33]. It is mainly triggered by the attacking and migration manner of seagulls [34].

During migration in the flock, seagulls follow special conditions [22, 34]:

- to prevent collisions between nearby search agents,
- to stay nearby to the neighborhoods with the optimal and richest sources of energy,
- and to stay in the vicinity of the fittest (lead) seagull.

mathematical model related to this behavior is applicable in finding a solution for realworld problems. For calculating the location of search agents, an additional variable is introduced as follows:

$$X_s = A * P_s^t \tag{37}$$

Here, X_s position vector of the seagull in a way that it does not collide with other agents in the exploration space. P_s^t is the current position of an individual, and A is defined by:

$$A = 2 - 2.\left(\frac{t}{t_{\max}}\right) \tag{38}$$

The following equation formulates the movement in a way that follows the first and second conditions above:

$$Y_s = S.\left(P_{\text{best}}^t - P_s^t\right) \tag{39}$$

Here, Y_s is updated position of the seagull in P_s^t which moves toward the best position $\left(P_{\text{best}}^t\right)$. S is calculated by:

$$S = 2.A^2 \cdot r \tag{40}$$

Where r is in the range of (0.1). To stay nearby the fittest (lead) seagull, the position can be updated as follows:

$$U_s = |X_s + Y_s| \tag{41}$$

In the attacking phase, seagulls exploit the information they collected in the previous phase. In this stage, it performs a spiral movement rendered through x, y, and z planes utilizing the subsequent equations:

$$X' = R \cdot \cos(m) \tag{42}$$

$$Y' = R \cdot \sin(m) \tag{43}$$

$$Z' = R.m \tag{44}$$

$$R = W \cdot d^{mv} \tag{45}$$

Here, R represents the width of each coil of the spiral. m is in the domain of $[0 \leq m \leq 2\pi]$. W and v are constant vales that determine the spiral shape.

Finally, the updated position of each seagull in the search space is determined by:

$$P_s^t = P_{\text{best}}^t + U_s \cdot X' \cdot Y' \cdot Z' \tag{46}$$

2.5. Performance evaluation metrics

The following indexes are utilized in this paper to assess the effectiveness of advanced models:

- coefficient of determination $\left(\mathrm{R}^2\right)$ is the degree of the linear relation among the projected and experiential definite values. It is introduced as given in Eq. (47).

$$R^2 = \left(\frac{\sum_{i=1}^{n}\left(t_i - \bar{w}\right)\left(v_i - \bar{v}\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(v_i - w\right)^2\right]\left[\sum_{i=1}^{n}\left(v_i - \bar{v}\right)^2\right]}}\right)^2 \tag{47}$$

- Root Mean Square Error (RMSE) is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_i - w_i)^2} \tag{48}$$

- Normalized Root Mean Square Error (NRMSE) is defined by:

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_i - w_i)^2}}{\frac{1}{n}\sum_{i=1}^{n}(w_i)} \tag{49}$$

- The mean of the absolute values of the error between the estimated and observed actual values is known as the mean absolute error (MAE), as shown in Eq. (50).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|v_i - w_i| \tag{50}$$

In all the five equations:
n : quantity of samples,
v_i : predicted value,
$\bar{v}$: average predict value,
w_i : empirically measured value
$\bar{w}$: average empirically measured value.
T-statistical test (T_{state}) is defined as Eq. (51):

$$T_{\text{state}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right.}} \tag{51}$$

In this formula, $\bar{x}_1$ and $\bar{x}_2$ are the average of the two categories being compared, s^2 is the pooled standard error of the two categories, and n_1 and n_2 are the number of data entries per group.

2.6. Training Process

Algorithms are structured to optimize model performance by iteratively adjusting hyperparameters. In this process, optimizers assign random positions to the cost function, which are then fed into the hyperparameters. The determined hyperparameters and the values are presented in Table 2. Through training iterations, the model trains itself with given hyperparameters (random positions) then generates predictions and assesses error parameters. The optimal value is determined by minimizing this error, and subsequently, it is returned to the optimizer for further refinement.

Table 2. The results were obtained from the hybridized models

	Hyperparameter	
	C	Gama
LSAO	8	0.1
LSBE	20	0.04
LSOA	9.355	0.145

3. Results and discussion

The findings of the developed models to estimate the CBR value are demonstrated as follows. Here, the results of assessing the experimentally measured records and those estimated by LSAO, LSBE, and LSSO models in the training (70% of models) and testing (30% of models) phases are presented in Table 3 and various diagrams.

In this section, Table 3 reports a thorough assessment of model productivity by means of five statistical evaluators R^2, $RMSE$, $NRMSE$, MAE, and T_{state} . In terms of R^2, it is evident that the best result is for LSBE in both the train and test part, in which R^2 is 0.9977 and 0.978 , respectively. All three models' test portion R^2 values are less than the training portion, indicating insufficient training. The greatest outcomes in the train and test phases are clearly associated with LSAO2 and LSAO1, respectively, when one compares the values of the Tstatistical test T_{state} , where a smaller SI indicates the highest model accuracy. Three types of errors exist: RMSE, MAE, and NRMSE. These show that LSBE with lower error values performs best in the training phase; in the testing phase, LSBE again has the lowest results for RMSE and MAE evaluators; however, LSAO1 has the lowest result for NRMSE, making LSBE the second-best model.

Table 3. The results were obtained from the hybridized models

Phase	Metric	Models		
		LSAO	LSBE	LSSO
Train	R^2	0.9924	0.9977	0.9843
	$RMSE$	3.203	2.133	4.926
	MAE	2.657	1.518	4.266
	$NRMSE$	0.062	0.041	0.096
	T_{state}	0.913	0.224	0.154
Test	R^2	0.9720	0.978	0.966
	$RMSE$	4.863	4.613	5.999
	MAE	3.788	3.403	4.137
	$NRMSE$	0.209	0.221	0.272
	T_{state}	0.120	0.4585	0.871

Fig. 2 illustrates Scattered representations of the correlation between the test results and estimated values of CBR. The figures provided correspond to their two evaluation sets, R^2 and RMSE. Since RMSE acts as a dispersion controller for the outcomes, a lower value of this evaluator corresponds to a larger density and a less distributed set of results. Furthermore, the learning and validation points are moved closer to the centerline by the R^2 metric. The centerline at coordinate $Y = X$ refers to a straight line that passes through the origin and has a slope of 1 . This line

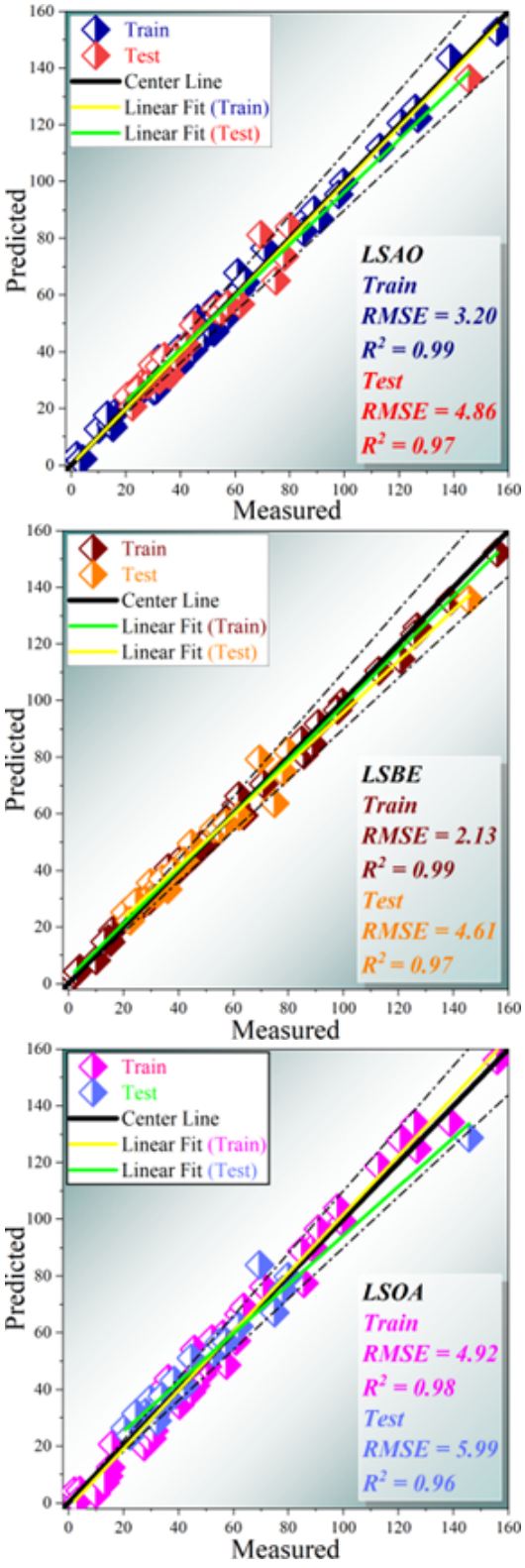

Fig. 2. Predicted and measured values provide the foundation of the scatter plot

represents the expected or ideal relationship between two variables. The points below and above those lines represent underestimation and overestimation. Fig. 2 shows three developed models in the train and test phases. Approximately the R^2 and RMSE of the training part in all three models, especially for LSBE, are in the convenient region. In the testing phase, results are not as proper as in the training phase. In this case, LSBE, LSAO, and LSSO demonstrate the best results, respectively.

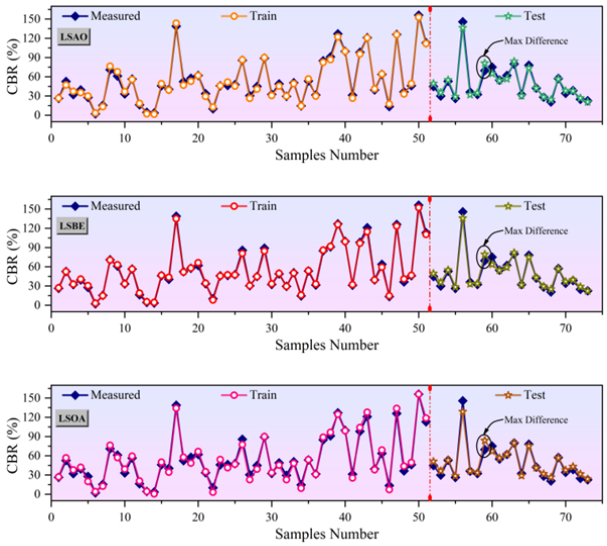

Fig. 3. A line-symbol plot comparing the observed and anticipated values

According to Fig. 3, there is a remarkable conformity among observed and estimated CBR values in all 3 models, especially for LSBE, which demonstrates the workability of the proposed developed models in estimating the *CBR* value. All in all, the evaluation of the proposed models suggests that the projected models have precise performance with a modest error in the CBR prediction procedure, providing that they are enough efficient to be used in practical applications.

The violin plot depicted in Fig. 4 illustrates the distribution of errors conforming to the normal distribution across discrete training and testing intervals for three distinct constructed models. Within this framework, it becomes apparent that the accuracy of outcomes diminishes as the propensity for error dispersion increases. During the training phase, the LSBE hybrid model demonstrated superior performance, outperforming both the LSAO and LSSO models, which attained second and third positions, respectively. However, a notable discrepancy emerges during the testing phase, wherein the LSSO model exhibits a wider spectrum of errors compared to LSBE and LSAO, albeit

LSBE and LSAO showcase relatively comparable outcomes. Across all three models, errors tend to converge around 0 percent, particularly evident within the LSAO and LSBE training segments. Nevertheless, an observation worth noting pertains to the LSBE model, wherein the training phase diagram displays a comparable spread of errors as observed in the testing diagram. This observation suggests that the training phase manifests a broader trend of error dispersion than what is encapsulated within the testing phase diagrams.

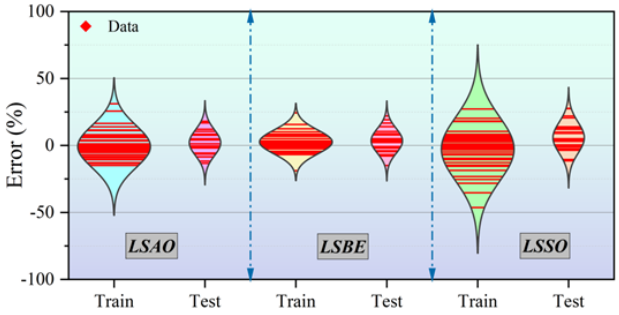

Fig. 4. The error percentage violin plot of the models that were provided

For each of the three constructed models, the testing phase's maximum error values are often less than the training phase's. When the prediction errors of the constructed models are compared in Fig. 5, they show that, for LSBE, errors range from less than 40% to less than 60%, which is considerably less than 60% and 90% for LSAO and LSSO, respectively. As a result, it can be said that in real-world situations, it performs best when predicting the CBR value.

4. Conclusion

In this study, the development and evaluation of three hybrid models involved fusing Least Square Support Vector Regression (LSSVR) with three distinct optimization algorithms: the Arithmetic Optimization Algorithm, the Bald Eagle Search Optimization, and the Seagull Optimization Algorithm. These hybrid models were meticulously crafted to address the time and cost constraints associated with traditional California bearing ratio (CBR) determination methods, aiming to offer a solution that is not only expedient but also accurate for real-world applications in soil mechanics and geotechnical engineering. To gauge the efficacy of these models, five crucial factors were considered, recognizing their significance in accurately predicting CBR values. Subsequently, the performance of each model was meticulously scrutinized across two critical phases: the training phase, where the models were honed and refined, and the testing phase, where their predictive capabilities were put

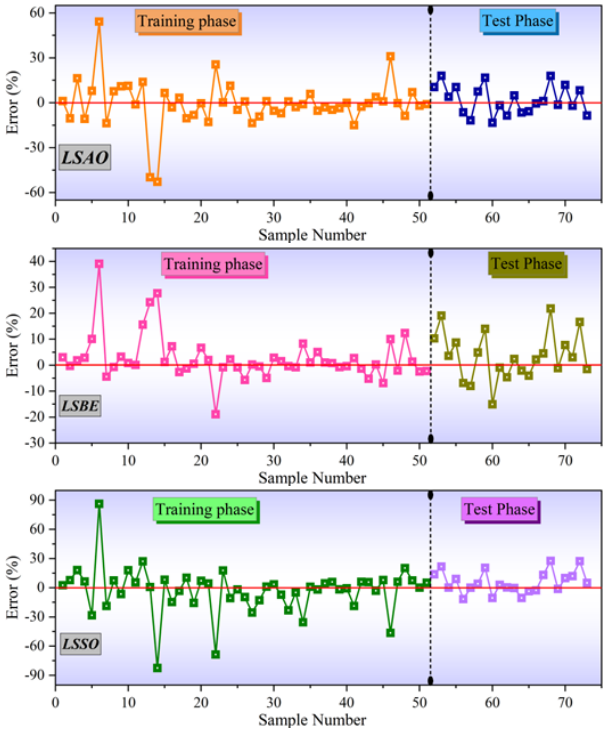

Fig. 5. The line-symbol plot for error percentage of developed hybrid models

to the ultimate test. The findings reveal a notable trend in the performance of the LSBE model, which emerged as the frontrunner during the training phase, showcasing superior predictive accuracy with significantly lower errors across various metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Normalized RMSE (NRMSE). Even during the testing phase, LSBE continued to demonstrate commendable performance, exhibiting the lowest RMSE and MAE values, although the NRMSE metric favored the LSAO model, indicating its slightly superior accuracy in this aspect. Moreover, the determination coefficient (R^2) values further underscored the robustness of the LSBE model, consistently yielding the highest values in both the training and testing phases. However, it is worth noting that while the models performed admirably during the training phase, there was a noticeable decrease in R^2 values during the testing phase, hinting at potential limitations in the training process that warrant further investigation. Despite these nuances, the hybrid models exhibited remarkable conformity in predicting CBR values, as evidenced by the strong agreement between the anticipated results and the experimentally observed values. Among them, the LSBE model stood out for its unparalleled accuracy, consistently delivering predictions with minimal errors, often hovering around the

0% mark. This impressive performance underscores the practical utility of the LSBE model and its suitability for deployment in real-world scenarios, where precision and efficiency are paramount.

nomenclature

Acronyms

T_stat	*T*-statistical test
AOA	Arithmetic Optimization Algorithm
BESO	Bald Eagle Search Optimization
CBR	California Bearing Ratio
CP	Curing Period
LI	Lime Percentage
LS	Lime Sludge Percentage
LSSVR	Least Square Support Vector Regression
MAE	Mean Absolute Error
MDD	Maximum Dry Density
ML	Machine Learning
NRMSE	Normalized Root Mean Square Error
OMC	Optimum Moisture Content
R^2	Coefficient of Determination
RMSE	Root Mean Square Error
SOA	Seagull Optimization Algorithm

References

[1] B. Yildirim and O. Gunaydin, (2011) *"Estimation of California bearing ratio by using soft computing systems"* **Expert Systems with Applications *38***: 6381–6391.

[2] M. Aytekin, (2000) *"Soil mechanics"* **Trabzon, Turkey: Academy Publishing house**:

[3] M. Bayat, M. Bayat, M. Kia, H. R. Ahmadi, and I. Pakar, (2018) *"Nonlinear frequency analysis of beams resting on elastic foundation using max-min approach"* **Geomechanics engineering *16***: 355–361.

[4] H. R. Ahmadi, A. M. Allahyari, and H. M. Allahyari, (2022) *"Identifying Damages in girders of Bridges Using Square Time-Frequency Distribution and Neural Network"* **Amirkabir Journal of Civil Engineering *54***: 3287–3312.

[5] L. S. Ho and V. Q. Tran, (2022) *"Machine learning approach for predicting and evaluating California bearing ratio of stabilized soil containing industrial waste"* **Journal of Cleaner Production *370***: 133587.

[6] T. Taskiran, (2010) *"Prediction of California bearing ratio (CBR) of fine grained soils by AI methods"* **Advances in Engineering Software *41***: 886–892.

[7] A. Ghorbani and H. Hasanzadehshooiili, (2018) *"Prediction of UCS and CBR of microsilica-lime stabilized sulfate silty sand using ANN and EPR models; application to the deep soil mixing"* **Soils and foundations *58***: 34–49.

[8] W. P. M. Black, (1962) *"A method of estimating the California bearing ratio of cohesive soils from plasticity data"* **Geotechnique *12***: 271–282.

[9] R. S. Patel and M. D. Desai. "CBR predicted by index properties for alluvial soils of South Gujarat". In: 2010, 79–82.

[10] M. Alawi and M. Rajab, (2013) *"Prediction of California bearing ratio of subbase layer using multiple linear regression models"* **Road Materials and Pavement Design *14***: 211–219.

[11] Y. Erzin and D. Turkoz, (2016) *"Use of neural networks for the prediction of the CBR value of some Aegean sands"* **Neural Computing and Applications *27***: 1415–1426.

[12] K.-j. Kim and H. Ahn, (2012) *"A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach"* **Computers Operations Research *39***: 1800–1811.

[13] V. Vapnik. *The nature of statistical learning theory*. Springer science business media, 2013.

[14] J. A. K. Suykens and J. Vandewalle, (1999) *"Least squares support vector machine classifiers"* **Neural processing letters *9***: 293–300.

[15] S. Wang, L. Yu, L. Tang, and S. Wang, (2011) *"A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China"* **Energy *36***: 6542–6554.

[16] M. A. Rad and M. S. Ayubirad, (2017) *"Comparison of artificial neural network and coupled simulated annealing based least square support vector regression models for prediction of compressive strength of high-performance concrete"* **Scientia Iranica *24***: 487–496.

[17] G. Xie, S. Wang, Y. Zhao, and K. K. Lai, (2013) *"Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study"* **Applied Soft Computing *13***: 2232–2241.

[18] G. G. Tejani, B. Sadaghat, and S. Kumar, (2023) *"Predict the maximum dry density of soil based on individual and hybrid methods of machine learning"* **Advances in engineering and intelligence systems *2***: 98–109.

[19] X. Xue and Y. Wei, (2020) *"A hybrid modelling approach for prediction of UCS of rock materials"* **Comptes Rendus. Mécanique *348***: 235–243.

[20] Y. Lei, S. Zhou, X. Luo, S. Niu, and N. Jiang, (2022) *"A comparative study of six hybrid prediction models for uniaxial compressive strength of rock based on swarm intelligence optimization algorithms"* **Frontiers in Earth Science *10***: 930130.

[21] J. O. Agushaka and A. E. Ezugwu, (2021) *"Advanced arithmetic optimization algorithm for solving mechanical engineering design problems"* **Plos one *16***: e0255703.

[22] G. Dhiman, K. K. Singh, M. Soni, A. Nagar, M. Dehghani, A. Slowik, A. Kaur, A. Sharma, E. H. Houssein, and K. Cengiz, (2021) *"MOSOA: A new multi-objective seagull optimization algorithm"* **Expert Systems with Applications *167***: 114150.

[23] M. Suthar and P. Aggarwal, (2018) *"Predicting CBR value of stabilized pond ash with lime and lime sludge using ANN and MR models"* **International Journal of Geosynthetics and Ground Engineering *4***: 1–7.

[24] U. Okkan and Z. A. Serbes, (2012) *"Rainfall–runoff modeling using least squares support vector machines"* **Environmetrics *23***: 549–564.

[25] Q. B. Pham, T.-C. Yang, C.-M. Kuo, H.-W. Tseng, and P.-S. Yu, (2019) *"Combing random forest and least square support vector regression for improving extreme rainfall downscaling"* **Water *11***: 451.

[26] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. *A practical guide to support vector classification*. 2003.

[27] B. G. Aiyer, D. Kim, N. Karingattikkal, P. Samui, and P. R. Rao, (2014) *"Prediction of compressive strength of self-compacting concrete using least square support vector machine and relevance vector machine"* **KSCE Journal of Civil Engineering *18***: 1753–1758.

[28] M. K. Habib and A. K. Cherri, (1998) *"Parallel quaternary signed-digit arithmetic operations: addition, subtraction, multiplication and division"* **Optics Laser Technology *30***: 515–525.

[29] L. Abualigah, A. Diabat, S. Mirjalili, M. A. Elaziz, and A. H. Gandomi, (2021) *"The arithmetic optimization algorithm"* **Computer methods in applied mechanics and engineering *376***: 113609.

[30] H. A. Alsattar, A. A. Zaidan, and B. B. Zaidan, (2020) *"Novel meta-heuristic bald eagle search optimisation algorithm"* **Artificial Intelligence Review *53***: 2237–2264.

[31] N. Tebbakh, D. Labed, and M. A. Labed, (2022) *"Optimal size and location of distributed generations in distribution networks using bald eagle search algorithm"* **Electrical Engineering Electromechanics**: 75–80.

[32] G. I. Sayed, M. M. Soliman, and A. E. Hassanien, (2021) *"A novel melanoma prediction model for imbalanced data using optimized SqueezeNet by bald eagle search optimization"* **Computers in biology and medicine *136***: 104712.

[33] G. Dhiman and V. Kumar, (2019) *"Seagull optimization algorithm: Theory and its applications for large-scale industrial engineering problems"* **Knowledge-based systems *165***: 169–196.

[34] N. Panagant, N. Pholdee, S. Bureerat, K. Kaen, A. R. Yıldız, and S. M. Sait, (2020) *"Seagull optimization algorithm for solving real-world design optimization problems"* **Materials Testing *62***: 640–644.