

Metaheuristic-Enhanced SVR Models for California Bearing Ratio Prediction in Geotechnical Engineering

Yulin Lan¹, Na Feng^{2,*} and Zhisheng Yang³

¹Planning and Finance Department, Weifang Engineering Vocational College, Weifang 262500, Shandong, China

²School of Information Engineering, Weifang Engineering Vocational College, Weifang 262500, Shandong, China

³Party and Government Office, Weifang Engineering Vocational College, Weifang 262500, Shandong, China

E-mail: sdqzyuchen@163.com

*Corresponding author

Keywords: california bearing ratio, support vector regression, adaptive opposition slime mold algorithm, alibaba and the forty thieves optimization algorithm, dingo optimization algorithm

Received: April 7, 2025

Soil resistance characteristics, particularly the California Bearing Ratio (CBR), play a pivotal role in pavement and subgrade design. However, conventional laboratory-based CBR testing is often time-consuming, labor-intensive, and costly. This study presents a novel machine learning framework that combines Support Vector Regression (SVR) with three recent metaheuristic optimization algorithms—Dingo Optimization Algorithm (DOA), Alibaba and the Forty Thieves Optimization (AFT), and Adaptive Opposition Slime Mold Algorithm (AOSMA)—to predict CBR values efficiently and accurately. A dataset consisting of 220 soil samples with eight geotechnical input parameters was used to develop and evaluate the hybrid models. The predictive performance of each model was assessed using multiple evaluation metrics, including R^2 , RMSE, MSE, RSR, and WAPE. Results indicate that the SVR–AFT (SVAF) hybrid model outperformed the others, achieving an R^2 of 0.9968 and an RMSE of 0.7946 in the testing phase, demonstrating high generalization ability and predictive precision. The integration of SVR with metaheuristic algorithms significantly enhances model robustness and accuracy, offering a practical and cost-effective alternative to empirical CBR testing methods. This work highlights the potential of hybrid AI models in solving complex geotechnical prediction problems and contributes to the growing body of research at the intersection of civil engineering and artificial intelligence.

Povzetek: Hibridni modeli SVR so optimizirani z metahevristikami AFT, DOA in AOSMA za hitro in natančno napovedovanje CBR iz osmih geotehničnih parametrov. Na 220 vzorcih doseže najboljši model SVAF $R^2 = 0.9968$ in $RMSE = 0.7946$, kar ponuja stroškovno učinkovito alternativo laboratorijskim testom.

1 Introduction

CBR is the term utilized by Geotechnical construction to describe the resistivity of the substrate sample to a piston's insertion. More specifically, the CBR describes the force applied to the piston to enable it to penetrate the soil. [1]. Initially, the CBR examination was devised in California to appraise the suitability of soils for highway construction. Civil engineers modified the testing process to enhance its impact on the airport's construction. Almost all emerging countries widely adopt the CBR test to appraise pavements' resilience to soil [2]. A material's load-bearing capacity is gauged by its CBR, which is the ratio of the attainable supporting strength of base materials to that of regular crushed rock. In structural engineering, 100 is considered a reasonable limit for the CBR for broken rock substances.

Conversely, the values of CBR for alternative materials are found to be below 100 [3]. Recent advances in artificial intelligence (AI) are closely intertwined with the rapid development of electronic technologies, forming the foundation of the so-called "information society."

Gams and Kolenik highlight the reciprocal relationship between electronics and AI, where swift hardware improvements, described by a comprehensive set of Information Society (IS) laws, have driven groundbreaking progress in AI across fields like medicine, smart environments, and autonomous systems. Their research shows that AI and ambient intelligence (AmI) not only benefit from electronic advancements but are also beginning to influence hardware optimization and intelligent system design through AI, indicating a move toward a more integrated technological progression [4]. After the compacted soils have been tested, the laboratory can conduct the subsequent test. However, it is possible for soils located in trenches to conduct the CBR test under the circumstances on the premises [5] Recognizing that in situ and laboratory test outcomes can show perceptible differences between soil types, unit weights, and water content is essential. Employing CBR tests has proved promising for presenting information about the stability and strength of different kinds of structures related to soils, such as road fills, airport roadways and dams, and road foundations.

Moreover, these tests can be conducted in unsoaked and soaked soil varieties. The laboratory CBR tests are characterized by their demanding nature regarding time and manual effort. Moreover, the outcomes of such tests are frequently marred by discrepancies attributed to the suboptimal quality of conditions in the lab and samples of dirt, which, in turn, lead to inaccurate CBR values [6].

Various studies have been performed on the California bearing ratio, which led the researchers to formulate different procedures. Previous studies showed that changes in the soil types and properties affected the value of CBR. Amongst other things, it has been observed that most research work has focused on studying the relationship existing between the compacted properties, unique indicators, as well as the mineral examinations' CBR concentrations [6]–[8][9], [10], [11], [12], [13], [14][15], [16], [17], [18]. To determine the value of CBR, soils are compacted at a predetermined MDD and OMC at a specified energy level for the soil material. For the CBR, the cases are soaked for four days; the primary purpose of this soaking is to allow absorption. Consequently, the assessment of the CBR value for a soaked sample typically requires a period of approximately five to six days. This delay can prove detrimental to the timely completion of a large-scale construction endeavor. Since soil is vastly different from one quality to another, applying this exercise to the foundation soil samples collected from a small count of sites may not truly represent the soil properties for all roads. To eliminate this deficiency, a large count of specimens is needed to be gathered for tests. Therefore, calculating the CBR values for pavement subgrade soils deploying easily identifiable parameters becomes very important in developing appropriate pavement design parameters. Recently, interest in using Artificial Intelligence (AI) tactics to solve geotechnical engineering problems has increased. Consequently, some valuable outcomes have been obtained [19], [20]. Furthermore, a limited count of studies has documented endeavors to appraise the (CBR) of soils via adopting diverse Artificial Neural Network (ANN) methodologies [21], [22], [23]. Recent advances in machine learning have increasingly supported geotechnical engineering by improving the prediction of soil and foundation properties through data-driven models. Support Vector Regression (SVR), when combined with metaheuristic optimization, has proven particularly effective in modeling complex nonlinear relationships within geotechnical datasets. Ngo et al. [24] demonstrated that SVR optimized via metaheuristics yielded superior performance in predicting the unconfined compressive strength of stabilized soils. Similarly, Hoang et al. [25] applied enhanced SVR models to successfully estimate pile bearing capacity, showcasing the method's versatility in foundation engineering. In the context of California Bearing Ratio (CBR) prediction, Bherde et al. [26] reported that Random Forest Regression outperformed other algorithms, including SVR, with maximum dry density and gravel content being the most influential predictors. While these results support the effectiveness of ensemble models, they also underline the need for more optimized SVR configurations that can match or exceed ensemble performance. A broader

comparative study by Ma et al. [27], evaluating 20 metaheuristic algorithms for SVR parameter tuning in landslide displacement prediction, revealed considerable variation in outcomes. The Multiverse Optimizer emerged as particularly efficient in achieving high accuracy with low computational cost, highlighting the critical role of algorithm selection in enhancing SVR model performance. These studies collectively underscore the growing impact of hybrid AI models in geotechnical applications. However, few works have focused specifically on integrating SVR with newer and less explored metaheuristic algorithms such as the Alibaba and Forty Thieves (AFT), Dingo Optimization Algorithm (DOA), or Adaptive Opposition Slime Mold Algorithm (AOSMA). Our study addresses this gap by systematically evaluating and comparing these novel SVR-based hybrid models in predicting CBR, offering insights into their optimization behaviors, convergence patterns, and predictive robustness. By incorporating recent advancements and experimental benchmarks, this work aims to contribute both technically and methodologically to the field of AI-driven geotechnical modeling.

Considering the variety of parameters to be considered and the range of datasets observed, as explained in the previous paper, it becomes of prime importance to develop robust predictive methodologies to model the mechanical attributes of the CBR and delineate the complex correlations between the constituents of soil.

Recent studies have explored various soft computing and machine learning techniques for predicting the California Bearing Ratio (CBR). These include Random Forest, Gradient Boosting, and XGBoost, which are known for their solid performance in regression tasks. However, such models often need extensive tuning and can struggle to capture complex nonlinear relationships, especially when feature interactions are subtle. Conversely, Support Vector Regression (SVR) demonstrates strong generalization and robustness, especially when combined with kernel functions and metaheuristic optimization. To assess the effectiveness of the proposed SVR-based hybrid models, we also incorporated Random Forest as a benchmark and compared its predictive accuracy with the SVR models enhanced by metaheuristics.

Recent studies have explored different soft computing and machine learning approaches for predicting CBR. Key techniques include Artificial Neural Networks (ANN), Multiple Linear Regression (MLR), Group Method of Data Handling (GMDH), and SVM. These models use soil parameters such as Atterberg limits, dry density, optimum moisture content, and soil gradation as inputs. However, many struggle with issues like overfitting, limited generalization to unseen data, or inadequate hyperparameter optimization. As shown in Table 1, most previous models achieved only moderate accuracy and did not utilize metaheuristic optimization to boost prediction performance. To fill this gap, this study introduces a hybrid Support Vector Regression (SVR) model combined with three metaheuristic optimizers—AFT, DOA, and AOSMA—aimed at improving the model's ability to learn nonlinear patterns. The superior

performance of the SVAF model, especially in RMSE and R^2 metrics, highlights the benefits of this approach.

Table 1: Overview of past methods for CBR prediction

Study	Model Type	Input Features	Dataset Size	Performance Metrics (R^2 / RMSE)	Notes
Yildirim & Gunaydin	Artificial Neural Network (ANN)	LL, PL, PI, MDD, OMC, % Sand and Gravel	120	$R^2 = 0.945$ / RMSE = 1.82	ANN prone to overfitting and high variance
Taskiran	GMDH	LL, PL, PI, Compaction properties	200	$R^2 \approx 0.92$	Good performance, limited interpretability
Alawi & Rajab	Multiple Linear Regression (MLR)	LL, PL, PI, Soil Gradation	100	$R^2 = 0.86$	Struggles with nonlinear relationships
Ngo et al. [24]	SVR + Improved Arithmetic Optimization (IAOA)	Grain size, Density, OMC, PI	150	$R^2 = 0.96$ / RMSE = 1.12	SVR enhanced with metaheuristic tuning
Wu et al. [25]	Stochastic Gradient Boosting Regression (SGBR)	LL, PL, MDD, % Clay, % Silt	300	$R^2 = 0.974$	Ensemble method with good generalization
Bherde et al. [26]	Random Forest Regression (RFR)	MDD, % Gravel, OMC, PI	400	$R^2 = 0.982$	Strong performance, but no hyperparameter optimization
Current Study	SVR + AFT	LL, PL, PI, MDD, OMC, SDA, QD, OPC	300	$R^2 = 0.9968$ / RMSE = 0.7946	Best accuracy using hybrid SVR and AFT metaheuristic

This study addresses the challenges of traditional CBR testing methods, which are often time-consuming and costly, by exploring advanced machine learning models supplemented with nature-inspired optimization techniques. Specifically, it focuses on Support Vector Regression (SVR), a popular tool for nonlinear regression. Since SVR's performance heavily depends on hyperparameter selection, three recent metaheuristic algorithms—Adaptive Opposition Slime Mold Algorithm (AOSMA), Alibaba and the Forty Thieves Algorithm (AFT), and Dingo Optimization Algorithm (DOA)—are employed to optimize the SVR framework. These algorithms offer diverse search strategies with strong potential for effective global optimization and faster convergence. The predictive capability of these hybrid models is evaluated using five standard statistical metrics: R^2 , RMSE, MSE, RSR, and WAPE. The study aims to (1) develop and validate an SVR model for predicting the California Bearing Ratio (CBR) based on soil and compaction parameters; (2) improve SVR's predictive performance through hyperparameter tuning with the three optimization algorithms; and (3) perform a comprehensive comparison of the models using these metrics to identify the most accurate and reliable one for use.

2 Materials and methodology

2.1 Data gathering

This study's dataset consists of 121 soil samples gathered from various geotechnical investigation reports and lab tests across different regions in [insert country or region, e.g., southwestern Iran or southeastern Asia—please specify based on your case]. The samples include a variety of soil types such as clayey soils, silty sands, gravels, and mixtures to ensure the broad applicability of the predictive models. Each sample records essential input parameters like [list key parameters: e.g., dry density, moisture content, liquid limit, plasticity index, etc.], with the California Bearing Ratio (CBR) used as the target variable. Data were obtained from both published literature and in-house experiments, offering a comprehensive understanding of soil behavior in various geological settings. The data in this investigation depend on eight variables: OPC, SDA, QD, plastic limit, liquid limit, maximum dry density, plasticity index, and ideal content. Simultaneously, the resulting component of interest is identified as the CBR value. The dataset has been split into two subsets: 30% of the total set is made up of the testing phase, while 70% comprises the training phase.

Table 1 depicts a numerical example of some of the parameters used in building the scheme. This table gives an overall summary of some of the attributes, such as minimum (Min), maximum (Max), standard deviation (St.), and mean (M). It is crucial to determine the essential

parameters for analyzing statistics. The maximum values for the LL, PL, PI, MDD, OMC, SDA, QD, and OPC variables are 52.1, 37.2, 19.5, 1.777, 29.5, 20, 20, and 8, respectively. Also, CBR's maximum value as an output parameter is 66.75 percent.

Table 2: The statistical features of the dataset components

Parameters	The numerical traits			
	Max	Min	Mean	St.dev
LL	52.10	21.20	35.8450	6.15380
PL	37.20	17.90	26.6830	4.28120
PI	19.50	2.10	9.16230	4.11490
MDD	1.7770	1.3650	1.49290	0.08830
OMC	29.50	18.90	24.1430	2.42670
SDA (%)	20.0	0	10.6600	7.15460
QD (%)	20.0	0	10.640	8.19610
OPC (%)	8.0	2.0	4.94490	2.37980
CBR (%)	66.750	19.690	39.9590	10.8660

2.2 Support vector regression (SVR)

In its early phases, the (SVM) technology was used to address pattern identification problems, initially introduced by Vapnik [28]. Then, Vapnik [29] suggested the SVM algorithm to solve problems with function approximation, which resulted in developing the SVR approach. The SVR approach is an innovative and perhaps practical method in data regression analysis. In this study, Support Vector Regression (SVR) is used as the main predictive model. Because of its ability to manage nonlinear relationships through kernel functions, SVR is especially suitable for modeling complex geotechnical datasets. The Radial Basis Function (RBF) kernel is chosen due to its effectiveness in high-dimensional feature spaces and its ability to generalize well. The SVR model depends on three main hyperparameters: C (regularization parameter): Regulates the trade-off between achieving a low training error and maintaining a simple model. γ (gamma): Determines the influence range of a single training example; a lower value means a wider reach, while a higher value indicates a more localized effect. ε (epsilon): Defines the tolerance margin within which errors are not penalized. These parameters were tuned using metaheuristic algorithms to minimize the root mean square error (RMSE) of predictions.

From an academic standpoint, the (SVR) may be explained in the subsequent terms. SVR uses a dataset that has $\bar{N}$ entries in it $\{(X_i, y_i), i = 1, 2, \dots, \bar{M}\}$.

The training dataset's overall count of instances is denoted by M .

$X_i = \{x_1, x_2, \dots, x_m\} \in R^m$ denotes the i -th component of the vector with M dimensions.

$y_i \in R$ represents the genuine value connected to X_i .

For that, in machine learning tactics, l -dimensional feature space - or something similar - represents the exact mapping of any training data point X_i in an SVR. The obtained hyperplane is in the space of features that will be selected using Support Vector Regression towards the portrayal of the optimal hyper-plane between the input (or the uncorrelated) variable and the exact output, the

dependent variables: Eq. (1) gives, mathematically, the operation of an SVR;

$$f(x) = Z^T \varphi(x) + b \quad (1)$$

b is the variable element

$f(x)$ symbolizes the expected parameters

Z is the l -dimensional weighting component.

An example of how distinct components X_i are mapped to a feature space with many dimensions is the function $\varphi(x)$.

The formal expression for the ε -insensitive coefficient of loss is found in Eq. (2).

$$|y - f(x)|_\varepsilon = \max(0, |y - f(x)| - \varepsilon) \quad (2)$$

The difference between the real number, symbolized by y , and the anticipated value, $f(x)$, as expressed theoretically by Eq. (3), is known as the residual.

$$R(x, y) = y - f(x) \quad (3)$$

According to Eq. (4), incorporating the entire residue piece within a preset boundary value of ε is the optimum regression model.

$$-\varepsilon \leq R(x, y) \leq \varepsilon \quad (4)$$

Eq. (4) coincides with the hypothesis on the whole training data set. Thus, if the residual meets the criterion $R(x, y) = \pm\varepsilon$, the data exhibits a maximum detour from the hyperplane. One can calculate a spatial separation of an arbitrary data point (x, y) from the hyperplane $R(x, y) = 0$ by the formula $|R(x, y)|/\|W^*\|$. Further, Z^* can be calculated as:

$$Z^* = (1, -Z^T)^T \quad (5)$$

In this question, the variable δ is assumed to be the maximum degree of dispersion between the hyperplane $R(x, y) = 0$ and the dataset (x, y) . All the training data can be induced to meet the requirements shown in Eq. (6). If the value of δ reaches its maximum in the SVR scheme, it means that the scheme can exhibit the best generalization ability.

$$|R(x, y)| \leq \delta \|Z^*\| \quad (6)$$

Whenever $R(x, y)$ equals an ε , the most significant distance is reached. After that, Eq. (6) may be changed to become Eq. (7). Considering the translation of an optimization issue to a minimal $\|Z\|$, $\|Z^*\|^2 = \|Z\|^2 + 1$, and $\|Z^*\|$ must be a minimal value to attain the maximum of δ .

$$\varepsilon = \delta \|Z^*\| \quad (7)$$

Even with efforts to keep mistakes within the $(-\varepsilon, \varepsilon)$ range during training, it is still possible for certain errors to surpass this limit. If training mistakes are less than $-\varepsilon$, they are displayed by ζ_i , and if they are more than ε , they are displayed by ζ_i^* . We define the notations ζ_i and ζ_i^* according to Eqs. (8) and (9), appropriately.

$$\zeta_i = \begin{cases} 0 & R(x_i, y_i) - \varepsilon \leq 0 \\ R(x_i, y_i) - \varepsilon & \text{others} \end{cases} \quad (8)$$

$$\zeta_i^* = \begin{cases} 0 & \varepsilon - R(x_i, y_i) \leq 0 \\ \varepsilon - R(x_i, y_i) & \text{others} \end{cases} \quad (9)$$

By using the ε sensitivity loss function, (SVR) aims to eliminate the distinction across the training data and the hyperplane region and choose a hyperplane that produces the best result. The goal function for (SVR) optimization is displayed by Eq. (10):

$$\min F(Z, b, \zeta_i, \zeta_i^*) = \frac{1}{2} \|Z\|^2 + c \sum_{i=1}^M (\zeta_i + \zeta_i^*) \quad (10)$$

With the confinements:

$$\begin{aligned} y_i - Z^T \varphi(x_i) - b &\leq \varepsilon + \zeta_i & i = 1, 2, \dots, \bar{M} \\ Z^T \varphi(x_i) + b - y_i &\leq \varepsilon + \zeta_i^* & i = 1, 2, \dots, \bar{M} \end{aligned}$$

$$\zeta_i \geq 0, \zeta_i^* \geq 0 \quad i = 1, 2, \dots, \bar{M}$$

The first term of Eq. (10) tends to restrict weights, so they stay above a certain limit to preserve whether the regression algorithm is constant. The second part of this system defines the ratio of certainty to vulnerability for possible hazards resulting from previous experiences using the ε -insensitive Relationship to losing. After determining the solution for the quadratic enhancement issue with inequality restrictions, the value of coefficient Z can be gathered from Eq. (11).

$$Z = \sum_{i=1}^M (\beta_i^* - \beta_i) \varphi(x_i) \quad (11)$$

The values of β_i^* and β_i are determined by solving a quadratic programming problem that incorporates an indication of the Lagrangian multipliers. Mathematically, the Support Vector Regression function is displayed with the utilize of the equation depicted as Eq. (12):

$$f(x) = \sum_{i=1}^M (\beta_i^* - \beta_i) K(x_i - x) + b \quad (12)$$

The kernel function, which is displayed as $K(x_i - x)$, exhibits the capacity to convert the training data into a higher nonlinear 1-dimensional space. Therefore, this methodology is deemed appropriate for solving issues related to nonlinear relationships, including projecting electrical power. Figure 1 shows the operational diagram for SVR.

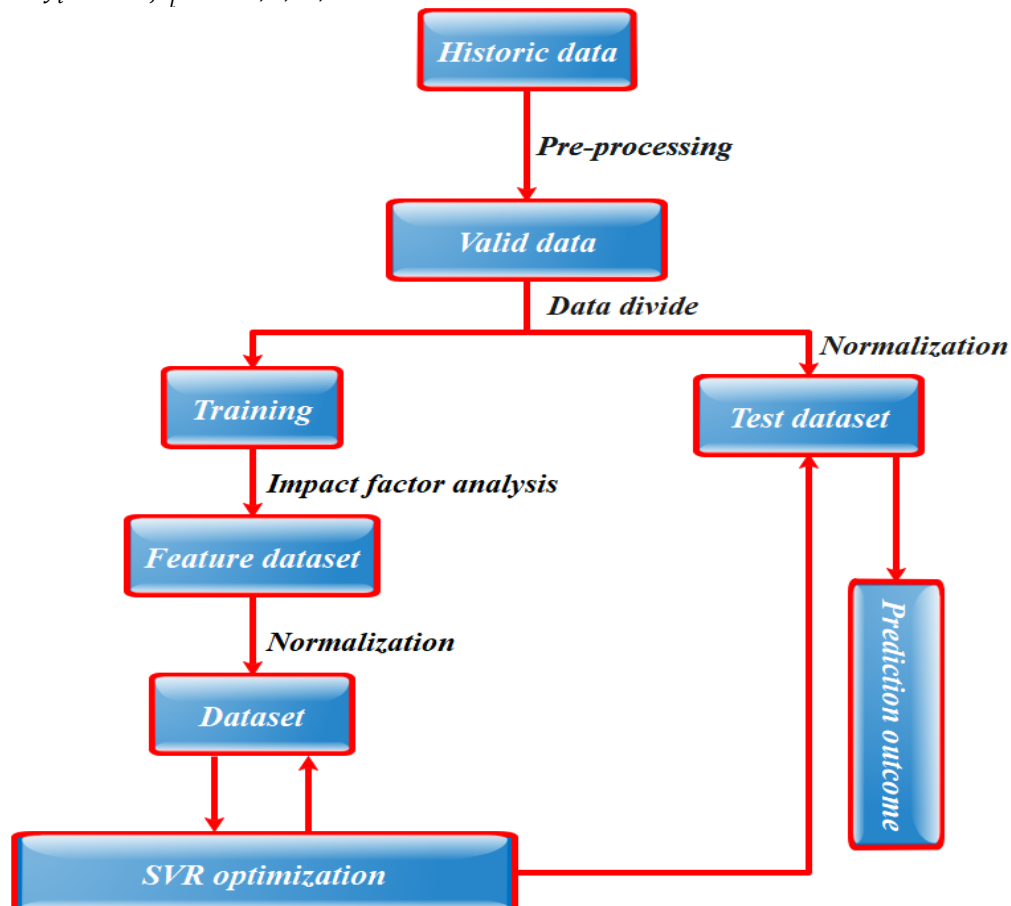


Figure 1: The progress and validation flowchart of an SVR scheme

2.3 AOSMA

The plasmodial slime mold's oscillatory mode is the basis for SMA. The slime mold employs a positive-negative feedback mechanism in conjunction with an oscillatory mode to establish the optimal route toward nutrition [30]. AOSMA is a new statistical technique that incorporates an opposition-based learning-based adaptive decision-making method to improve slime mold's nearing conduct [31].

Let it be assumed that a total of N individuals of the species of slime mold under consideration are resident in the search domain that is bounded by an upper boundary (UB) and a lower boundary (LB) for theoretical framework development of the (AOSMA).

$$X_i(t+1) = \begin{cases} X_{LB}(t) + V_d(W \cdot X_A(t) - X_B(t)) & p_1 \geq \delta \text{ and } p_2 < m_i \\ V_e \cdot X_i(t) & p_1 \geq \delta \text{ and } p_2 \geq m_i, \forall i \in [1, N] \\ \text{rand.}(UB - LB) + LB & p_1 < z \end{cases} \quad (15)$$

X_{LB} is the best local slime mold

X_A and X_B are pooled individuals by random

W is the weight factor

V_d and V_e are the random velocities.

p_1 and p_2 are randomly chosen numbers in $[0,1]$

The slime mold's chance, which starts at a random search situation, is fixed at $\delta = 0.03$.

The i -th member of the population's threshold value, m_i , aids in choosing the slime mold's location, which is calculated as Eq. (16):

$$m_i = \tanh[F(X_i) - F_G], \forall i \in [1, N] \quad (16)$$

$$F_G = F(X_G) \quad (17)$$

$$W(\text{SortInd}_F(i)) = \begin{cases} 1 + \text{rand.} \log\left(\frac{F_{LB} - F(X_i)}{F_{LB} - F_{LW}} + 1\right) & 1 \leq i \leq \frac{N}{2} \\ 1 - \text{rand.} \log\left(\frac{F_{LB} - F(X_i)}{F_{LB} - F_{LW}} + 1\right) & \frac{N}{2} < i \leq N \end{cases} \quad (18)$$

F_G and X_G are the values of worldwide top ranking and worldwide best well-being.

rand displays a random number in within $[0,1]$

F_{LB} and F_{LW} are local best and worst fitness values.

The utilization of an ascending order for sorting fitness values can be employed in a minimization problem:

$$[\text{Sort}_F, \text{SortInd}_F] = \text{sort}(F) \quad (19)$$

The local best and worst fitness also the local best slime mold X_{LB} are computed as Eqs. (20-22):

$$F_{LB} = F(\text{Sort}_F(1)) \quad (20)$$

$$F_{LW} = F(\text{Sort}_F(N)) \quad (21)$$

$$X_{LB} = X(\text{SortInd}_F(1)) \quad (22)$$

$$Xn_i(t) = \begin{cases} X_{LB}(t) + V_d(W \cdot X_{LB}(t) - X_B(t)) & p_1 \geq \delta \text{ and } p_2 < m_i \\ V_e \cdot X_i(t) & p_1 \geq \delta \text{ and } p_2 \geq m_i \\ \text{rand.}(UB - LB) + LB & p_1 < \delta \end{cases} \quad (27)$$

$X_i = (x_i^1, x_i^2, \dots, x_i^d), \forall i \in [1, N]$ is the i th slime mold's location in d -dimension.

$F(X_i), \forall i \in [1, N]$ symbolizes the i th slime's fitness.

The following represents the location as well as fitness of the slime mold at round t :

$$X(x) = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^d \\ x_2^1 & x_2^2 & \dots & x_2^d \\ \vdots & \vdots & \vdots & \vdots \\ x_N^1 & x_N^2 & \dots & x_N^d \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \quad (13)$$

$$F(X) = [F(X_1), F(X_2), \dots, F(X_N)] \quad (14)$$

In the $(t+1)$ cycle, the situation of the slime mold has been advanced. It has undergone an upgrade in its spatial disposition, which determine as Eq. (15):

The randomly assigned velocities are known as V_d and V_e and are defined as follows:

$$V_d \in [-d, d] \quad (23)$$

$$V_e \in [-e, e] \quad (24)$$

$$d = \text{arctanh}\left(-\left(\frac{t}{T}\right) + 1\right) \quad (25)$$

$$e = 1 - \frac{t}{T} \quad (26)$$

T is the maximum cycle.

SMA holds great promise for both investigation and exploitation in technological problem-solving and enhancement. However, the improvement of slime mold regulations in the SMA area is nevertheless reliant on a count of basic circumstances.

Case 1: The region's best slime mold, X_{LB} , and two random individuals, X_A and X_B , with velocity V_d , drove to determine when $p_1 \geq z$ and $p_2 < m_i$. This stage makes it easier to strike a balance amongst discovery and extraction.

Case 2: The orientation of the slime mold with velocity V_e directs the search when $p_1 \geq z$ and $p_2 \geq m_i$. This instance facilitates fraud.

Case 3: When $p_1 < z$, the person reinitializes within a specified search domain. This phase facilitates investigation.

Case 1 shows how the possibilities of finding solutions are improperly controlled during exploration and exploitation since X_A and X_B are two random slime molds. To get around this limitation, X_A can be used in place of best local individual X_{LB} . Consequently, the location of the i -th component is remodeled as Eq. (27):

Case 2 illustrates how slime mold deliberately targets a nearby location, resulting in a path with a lower fitness level. A better approach to this issue is to implement an adaptive decision system.

Case 3 illustrates that the SMA offers criteria for exploration. However, with a small value $\delta = 0.03$, the exploration has been limited. To address the issue, it is imperative to introduce an auxiliary exploration adjunct for SMA. A practical approach to addressing the limitations of Cases 2 and 3 entails employing a flexible decision approach that leverages opposition-based learning (OBL) to determine the necessity of additional exploratory efforts [32]. The OBL uses a defined Xop_i in the search domain, which is precisely the opposite of the Xn_i for each member ($i = 1, 2, \dots, N$), and compares it to upgrade the following cycles' situation. It assists in improving convergence and avoiding the chances of being closed in the local minima. So, the Xop_i for the i -th individual in j -th ($j = 1, 2, \dots, s$) dimension is described as follows:

$$Xop_i^j = \min(Xn_i(t)) + \max(Xn_i(t)) - Xn_i^j(t) \quad (28)$$

Xr_i represents the i -th member's situation in the reduction issue and is depicted as:

$$Xr_i = \begin{cases} Xop_i(t) & F(Xop_i(t)) < F(Xn_i(t)) \\ Xn_i(t) & F(Xop_i(t)) \geq F(Xn_i(t)) \end{cases} \quad (29)$$

A flexible decision is formed drawing on the prior worth of fitness $f(Xi(t))$ and the present fitness value $f(Xni(t))$ in the event of a depleted nutrient pathway. This is a typical academic kind of writing. It helps provide added research as needed. Then, the situation for the subsequent cycle is improved:

$$\begin{aligned} &X_i(t+1) \\ &= \begin{cases} Xn_i(t) & F(Xn_i(t)) \leq F(X_i(t)) \\ Xr_i(t) & F(Xn_i(t)) > F(X_i(t)) \end{cases}, \quad \forall i \in [1, N] \end{aligned} \quad (30)$$

The aforementioned AOSMA framework is displayed in pseudo-code, as shown in Algorithm 1.

In this study, the Adaptive Opposition Slime Mold Algorithm (AOSMA) is used not as a standalone optimizer but as a hybrid component integrated with Support Vector Regression (SVR). AOSMA optimizes three key hyperparameters of SVR—specifically the regularization parameter C , the epsilon-insensitive loss margin ε , and the kernel coefficient γ —with the goal of minimizing prediction error measured by RMSE. Through its adaptive opposition-based learning strategy and dynamic parameter control, AOSMA allows for more effective exploration of the search space and helps prevent premature convergence. As a result, the hybrid AOSMA-SVR model achieves better accuracy and generalization in predicting California Bearing Ratio (CBR) values from geotechnical data.

Algorithm 1: AOSMA

Begin

Using the criteria for searching boundary range $[LB, UB]$, choose a target variable f with inputs N, s, T , and δ .

Outputs: X_G and F_G

Initialization: Launch the slime mold at arbitrary.

$X_i = (x_i^1, x_i^2, \dots, x_i^d), \forall i \in [1, N]$ during the first revision, inside the query boundaries UB and LB
 $t = 1$.

while ($t \leq T$)

→ Determine the N slime mold's fitness values $F(X)$.

→ Put the fitness value in order.

→ The local best individual X_{LB} should be updated to match the local best conditioning F_{LB} .

→ The local weakest fitness F_{LW} should be updated.

→ Update the matching worldwide greatest individual X_G and global best fitness F_G .

→ Refresh the measurement of W .

→ Update the d using Eq. (25) and e using Eq. (26).

for (each slime mold $i = 1: N$)

○ Create the p_1 and p_2 randomized numbers.

○ Create the m_i threshold quantity.

○ Utilizing Eq. (27), determine the new slime mold location Xn_i .

○ Determine the new slime mold $F(Xn_i)$'s nutritional value.

if ($F(Xn_i) > F(X_i)$) // Adaptive decision strategy

- Estimate Xop_i using Eq. (24). //Opposition-based learning
- Select Xr_i using Eq. (29).

End

○ Revise the subsequent cycle slime mold X_i using Eq. (30).

end

→ The following repetition $t = t + 1$

end

The result is X_G , representing the global most effective region.

2.4 AFT

The present investigation clarifies the basic AFT algorithm's mathematical model, which is described in [33]. The scheme encompasses three states that can be analyzed and delineated in the following:

Case 1: The pursuit of Ali Baba by the thieves, as derived from information obtained from a source, can be displayed by a simulation of their situations, as illustrated in Eq. (31):

$$x_i^{t+1} = gbest^t + [Td^t(best_i^t - y_i^t)r_1 + Td^t(y_i^t - m_{a(i)}^t)r_2]sgn(rand - 0.5), \quad p \geq 0.5, \quad q > P_{pt} \quad (31)$$

y_i^t represents Ali Baba's situation regarding the thief i .

$m_{a(i)}^t$ represents the amount of cleverness that Marjaneh uses to cover up thievery i .

x_i^{t+1} denotes the situation of the i -th thief.

$gbest^t$ is the most excellent situation a thief has ever had worldwide.

$r_1, r_2, rand, p$, and q are random values created within $[0, 1]$

$best_i^t$ is the optimal location of thief i has determined.

Td^t is the robbers' surveillance area as specified by Eq. (32).

$p \geq 0.5$ presents either 0 or 1

P_{pt} is Ali Baba's potential perceptive ability, as stated by Eq. (33).

$sgn(rand - 0.5)$ can be -1 or 1 , and

a is defined as Eq. (34).

$$Td^t = \tau_0 e^{-\tau_1 (\frac{t}{T})^\tau} \quad (32)$$

t and T Please consult the current and maximal repetition standards, accordingly.

τ_0 ($\tau_0 = 1$) is a preliminary estimate of the monitoring length.

τ_1 ($\tau_1 = 2$) is a set amount that regulates the discovery and utilization of resources.

$$P_{pt} = \lambda_0 \log(\lambda_1 (\frac{t}{T})^{\lambda_0}) \quad (33)$$

λ_0 ($\lambda_0 = 1$) depicts the final assessment of the robbers' chances of completing their task after the hunt.

Algorithm 2: AFT

Establish the regulation settings and get started.

Start by assessing every thief's starting, optimal, and worldwide situations.

Start by assessing Marjane's intelligence in comparison to all thieves.

Set $t \leftarrow 1$

While ($t \leq T$) do

Eq. (33) is used for modifying the input parameter P_{pt} .

for each thief, do

if ($p \geq 0.5$) then

if ($q \geq P_{pt}$) then

Use Equation (32) to update the thieves' positioning.

else

Utilizing Equation (36), adjust the robbers' whereabouts.

end if

else

Refine the thieves' situation by Eq. (37).

end if

λ_1 ($\lambda_1 = 1$) refers to a fixed value that controls exploration and exploitation.

$$a = [(n - 1).rand(n, 1)] \quad (34)$$

The vector $rand(n, 1)$ is generated as a set of random numbers within the bounds of $[0, 1]$.

$$m_{a(i)}^t = \begin{cases} x_i^t & \text{if } f(x_i^t) \geq f(m_{a(i)}^t) \\ m_{a(i)}^t & \text{if } f(x_i^t) < f(m_{a(i)}^t) \end{cases} \quad (35)$$

The score of the fitness function is denoted by $f(0)$.

Case 2: Thieves may perceive they have been tricked and will likely start exploring unfamiliar and unplanned areas.

$$x_i^{t+1} = Td^t[(u_j - l_j)r + l_j]; \quad p \geq 0.5, \quad q \leq P_{pt} \quad (36)$$

The upper and lower bounds of the search domain at dimension j are displayed by u_j and l_j , respectively.

r displays a stochastic quantity generated in the interval $[0, 1]$.

Case 3: To improve AFT's exploration and exploitation capabilities, thieves can investigate alternative search situations beyond those identified through the utilization of Eq. (31). This scenario can be formulated as Eq. (37):

$$x_i^{t+1} = gbest^t - [Td^t(best_i^t - y_i^t)r_1 + Td^t(y_i^t - m_{a(i)}^t)r_2]sgn(rand - 0.5) \quad (37)$$

Algorithm 2 concisely and formally describes the iterative pseudo-code stages that correspond to the core AFT.

The proposed hybrid framework combines the Dingo Optimization Algorithm (DOA) with Support Vector Regression (SVR) to tune the model's hyperparameters: C , ϵ , and γ . The DOA emulates the natural hunting tactics of dingoes, such as surrounding, chasing, and attacking prey, which are adapted into search operators for exploring the SVR parameter space. The aim is to minimize the SVR's RMSE on training data by identifying the optimal parameter combination. By balancing diversification and intensification, the DOA-SVR hybrid model can effectively avoid local optima and enhance SVR's ability to generalize for accurate CBR prediction.

```

end for
Refresh all thieves' current, best, and worldwide standings.
Utilizing Eq. (35), alter Marjane's wit goals.
     $t = t + 1$ 
end while
Give back the world's optimal solution.

```

2.5 Dingo optimization algorithm (DOA)

From the earliest times, nature has consistently been regarded as an exceptionally instructive and impactful educator. Every species that exists on the planet Earth possesses a distinct and unique mechanism for ensuring its survival. The present study involves the mathematical modeling of hunting behavior and social arrangements in the dingo species. This analytical approach is the basis for developing a DOA nature-inspired optimization technique [34]. The two primary constituents of DOA are regarded as exploration and exploitation. The algorithm generates various anticipated outcomes within the search domain during the initial exploration phase. However, the subsequent exploitation phase enables identifying and pursuing the most desired resolutions within the predetermined space. To discern the optimal resolution for a given pragmatic concern, refinement, and integration of both constituent factors are necessary. Nonetheless, achieving equilibrium among the proposed algorithm's constituents is arduous due to its stochastic disposition. To address an authentic engineering dilemma, the impetus for developing an algorithm implementation utilizing hybridized meta-heuristics is derived from this inspirational notion [34].

Dingo optimization is done by the computational designing of the prey's pursuit, encirclement, and attack.

2.5.1 Encircling

Given the lack of previous knowledge about the search location and its ideal characteristics, it is proposed that the objective or target prey is the best agent tactic currently in use, representing the social hierarchy of dingoes. The following mathematical formulas can be used to formalize the dingoes' behavior:

$$\vec{D}_a = |\vec{A} \cdot \vec{P}_p(x) - \vec{P}(i)| \quad (38)$$

$$\vec{P}(i+1) = \vec{P}_p(x) - \vec{B} \cdot \vec{D}(d) \quad (39)$$

$$\vec{A} = 2 \cdot \vec{a}_1 \quad (40)$$

$$\vec{B} = 2\vec{b} \cdot \vec{a}_2 - \vec{b} \quad (41)$$

$$\vec{b} = 3 - (I \times \left(\frac{3}{I_{max}}\right)) \quad (42)$$

The neighborhood dingoes' geographic coordinates are displayed as a two-dimensional vector. The dingo may adjust its situation to match the coordinates of (P, Q) based on the prey's location, which is displayed as (P^*, Q^*) . By adjusting the $\vec{A}$ and $\vec{B}$ vectors about the present situation, the graphic shows every possible location around the ideal agent. Setting $\vec{A} = (1, 0)$ and

$\vec{B} = (1, 1)$ provides access to the dingo's situation at $(P^* - P, Q^*)$. For example, Eqs. (38) and (39) make it easier for dingos to travel throughout the hunting area and find their prey randomly.

2.5.2 Hunting

Using a mathematical method, creating a dingo hunting strategy involves assuming that the alpha, beta, and other members of the pack have a thorough awareness of the possible prey sites. When conducting hunting trips, the alpha dingo always takes the lead. However, other dingo species, including beta, may hunt as well. Eqs. (43) to (51) are developed with this issue in line with the discussion.

$$\vec{D}_\alpha = |\vec{A}_1 \cdot \vec{P}_\alpha - \vec{P}| \quad (43)$$

$$\vec{D}_\beta = |\vec{A}_2 \cdot \vec{P}_\beta - \vec{P}| \quad (44)$$

$$\vec{D}_o = |\vec{A}_3 \cdot \vec{P}_o - \vec{P}| \quad (45)$$

$$\vec{P}_1 = |\vec{P}_\alpha - \vec{B} \cdot \vec{D}_\alpha| \quad (46)$$

$$\vec{P}_2 = |\vec{P}_\beta - \vec{B} \cdot \vec{D}_\beta| \quad (47)$$

$$\vec{P}_3 = |\vec{P}_o - \vec{B} \cdot \vec{D}_o| \quad (48)$$

The following formulae are utilized to determine each dingo's intensity:

$$\vec{I}_\alpha = \log\left(\frac{1}{F_\alpha - (1E - 100)} + 1\right) \quad (49)$$

$$\vec{I}_\beta = \log\left(\frac{1}{F_\beta - (1E - 100)} + 1\right) \quad (50)$$

$$\vec{I}_o = \log\left(\frac{1}{F_o - (1E - 100)} + 1\right) \quad (51)$$

2.5.3 Attacking

If a situation update is unavailable, it may be inferred that the dingo successfully concluded its hunt through a predatory attack. To formally articulate the strategy, the value of $\vec{b}$ is systematically diminished linearly through the utilization of mathematical notation. Noteworthy is the fact that the variation range of $\vec{D}_\alpha$ is further diminished by $\vec{b}$. The value above may be identified as $\vec{D}_\alpha$, which is a stochastic variable generated within the range of $[-3b, 3b]$, where the constant $\vec{b}$ undergoes a decremental process from 3 to 0 over a series of cycles. When $\vec{D}_\alpha$ Values are randomly generated within the interval $[1, 1]$. An exploratory agent is capable of moving to any possible situation along the trajectory between its existing location and the prey's location.

2.5.4 Searching

Dingoes exhibit hunting patterns primarily determined by their pack's location. They consistently progress in pursuit of locating and subduing prey. $\vec{B}$ represents random variables. Notably, if the value assigned to $\vec{B}$ falls below -1, it implies that the prey is retreating from the search agent. Conversely, if $\vec{B}$ exceeds 1, the pack is advancing toward its prey. This particular intervention facilitates the Department of Defense conduct a comprehensive global reconnaissance of identified targets. One factor contributing to a heightened probability of exploration within the DOA is the component denoted as $\vec{A}$. In Eq. (40), the vector $\vec{A}$ can generate a range of random numbers within the interval between 0 and 3, independent of the weight of the prey selected. The DOA function can

be characterized as a stochastic vector whereby the elements with values that are less than or equal to one take priority over those greater than or equal to one. This feature elucidates the gap's influence as described in Eq. (38). The hybrid framework combines the Dingo Optimization Algorithm (DOA) with Support Vector Regression (SVR) to tune hyperparameters: C , ε , and γ . Inspired by the natural hunting strategies of dingoes, such as surrounding, chasing, and attacking prey, the DOA translates these behaviors into search operators that explore the SVR parameter space. Its aim is to minimize the RMSE of SVR on training data by identifying the best parameter combination. By balancing exploration and exploitation, the DOA-SVR hybrid effectively avoids local optima and improves SVR's generalization ability, leading to more accurate CBR predictions.

Algorithm 3 offers the pseudo-code for the DOA.

Algorithm 3: Dingo Optimization

Input: The population of dingoes D_n ($n = 1, 2, \dots, n$)
Output: The best dingo. (Here, the best values are minimum)
 Generate initial search agents D_{in}
 Start the value of $\vec{b}$, $\vec{A}$, and $\vec{B}$.
While the Termination condition is not reached, do
 Appraise each dingo's fitness and intensity cost.
 D_α = dingo with the best search
 D_β = dingo with the second-best search
 D_o = Dingoes search outcomes afterward
 Cycle1
 repeat
 for $i = 1: D_{in}$ **do**
 Renew the latest search agent state.
 end for
 Project the fitness and intensity cost of dingoes.
 Record the value of S_α , S_β , S_δ
 Record the value of $\vec{b}$, $\vec{A}$, and $\vec{B}$.
 $Iteration = Iteration + 1$
 Monitor if cycle $\geq$ Stopping criteria
 output
end while

Choosing AFT, AOSMA, and DOA as optimizers was driven by their unique algorithmic bases and search methods, enabling a thorough comparison of their metaheuristic behaviors. These approaches are relatively recent and less studied, yet they show competitive performance in diverse regression and engineering tasks. Incorporating them with SVR in this research allows evaluation of both their predictive accuracy and optimization stability across different algorithmic frameworks.

2.6 Reproducibility and run settings

To ensure the robustness and reproducibility of the results, each hybrid SVR model (AFT-SVR, DOA-SVR, AOSMA-SVR) was executed 30 independent times. This allows for reliable statistical analysis of model performance. Additionally, random seed initialization was controlled using a fixed seed (e.g., seed = 42) across all

algorithms during training and optimization to maintain consistent behavior during repeated runs and to support reproducibility.

2.7 Hybridization strategy of SVR with metaheuristic algorithms

This study developed three hybrid machine learning models—SVAF, SVSM, and SVDO—by integrating Support Vector Regression (SVR) with three advanced metaheuristic optimization algorithms: Alibaba and Forty Thieves (AFT), Adaptive Opposition Slime Mold Algorithm (AOSMA), and Dingo Optimization Algorithm (DOA). The goal is to boost SVR's prediction accuracy by optimizing its key hyperparameters—penalty parameter C , kernel parameter γ , and epsilon-insensitive loss ε —using the global search methods provided by these metaheuristics. While SVR is a strong nonlinear regression technique, its effectiveness heavily relies on

proper parameter tuning. Traditional manual or grid search methods are often inefficient or may yield suboptimal results, especially with complex, high-dimensional geotechnical data. Therefore, this hybrid approach exploits the global search and convergence strengths of nature-inspired algorithms to automate SVR hyperparameter optimization.

- In SVAF, the AFT algorithm explores the search space dynamically through mechanisms like global surveillance, balancing exploration and exploitation, and adaptive decision-making inspired by Marjaneh. These features enable it to identify optimal SVR parameters reliably.

- In SVSM, AOSMA enhances the slime mold algorithm with opposition-based learning and adaptive strategies, allowing it to escape local minima more effectively and converge more rapidly, thus providing better hyperparameter configurations.

- In SVDO, the DOA mimics the social hunting behaviors of dingoes—such as encircling, attacking, and searching—to iteratively fine-tune the SVR parameters for higher prediction accuracy.

Each metaheuristic aimed to minimize the RMSE of SVR predictions on training data, with the best parameter set used to train the final hybrid model. The process was repeated 30 times to ensure stability and reproducibility. This hybrid approach directly supports the study's goal of creating accurate, efficient, and generalizable models for predicting the California Bearing Ratio (CBR) of soils. Using these metaheuristics not only enhances SVR's learning ability but also reduces the manual effort and computational cost typically required for parameter tuning.

2.8 Performance evaluation tactics

A range of evaluators was deployed to appraise hybrid schemes' productivity in CBR value prediction. The list of evaluators comprises RMSE, MSE, R^2 , the ratio of RMSE to standard deviation (RSR), and lastly, weighted absolute percentage error, or WAPE. R^2 determines the degree of linear relationship between the actual and forecasted magnitudes. The RMSE is the square root of the ratio between the square of the count of specimens and the estimated value departure from the actual value. WAPE could be quantified by dividing the total absolute error by the total real demand. Eq. (21-25) provides the values of these metrics above.

$$R^2 = \left(\frac{\sum_{i=1}^n (b_i - \bar{b})(d_i - \bar{d})}{\sqrt{[\sum_{i=1}^n (b_i - \bar{d})^2][\sum_{i=1}^n (d_i - \bar{d})^2]}} \right)^2 \quad (52)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - p_i)^2} \quad (53)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n ((d_i - p_i)^2) \quad (54)$$

$$RSR = \frac{RMSE}{St. Dev} \quad (55)$$

$$WAPE = \frac{\sum_{i=1}^n |d_i - b_i|}{\sum_{i=1}^n |b_i|} \quad (56)$$

n indicates the count of samples; d_i displays the forecasted value; b_i displays the actual value, while $\bar{d}$ and $\bar{b}$ represent the mean of the forecasted value and the average of the actual amount, respectively.

3 Outcomes and discussion

This paper reports on developing a Support Vector Regression model using three new enhancement techniques, AFT and DOA, aimed at developing three hybrid predictive models for soil estimation CBR. In previous schemes, the information about information was divided into two subsets: a set to learn and a set to validate the scheme, 70% and 30% of the data, respectively. The five consecutive statistical metrics, namely, R^2 , RMSE, MSE, RSR, and WAPE, were considered to get the full view of the optimizers' performance. Outcomes can be shown in Table 2. The statistical indicators are analyzed in this section to determine whether one model is generally better. By studying the various R^2 values among these different schemes, it would be crystal clear that the most promising outcomes are given out by SVAF in both the testing and training stages, with 0.9968 and 0.9929 values, respectively. Meanwhile, the minimum value of R^2 among all comparative schemes was given to the SVSM model at 0.9767. The key thing worth mentioning here is that all the schemes have increased R^2 during their test phases, indicating that the schemes are well-trained. Maximum RMSE, MSE, RSR, and WAPE values are 1.6271, 2.6475, 0.1524, and 0.0334 for SVSM in training. For the testing section, maximum RMSE values, MSE, RSR, and WAPE are 1.5824, 2.5042, 0.1409, and 0.0312 for SVSM. By contrasting the evaluators' and errors' values, the best hybrid scheme for estimating the CBR value of soils is the combination of SVR and the ATF algorithm (SVAF). This model has the highest R^2 value (0.9968 in the testing phase) and the lowest error value (0.7946 in testing) among all three components.

Table 3: The hybridized schemes produced the findings

Schemes	SVAF		SVSM		SVDO		SVR	
Section	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	0.9316	0.7946	1.6271	1.5824	1.3363	1.171	1.336392	1.171305
R^2	0.9929	0.9968	0.9767	0.9825	0.9852	0.992	0.985202	0.992446
MSE	0.868	0.6314	2.6475	2.5042	1.7859	1.372	1.7859	1.372

RSR	0.0872	0.0708	0.1524	0.1409	0.1251	0.1043	0.1251	0.1043
WAPE	0.0162	0.0141	0.0334	0.0312	0.0234	0.0212	0.0234	0.0212

Fig. 2 displays the dispersed presentations illustrating the correlation between the gauged and expected California Bearing Ratio values. R2 and RMSE are two types of assessments that include numerical data. When the value of this evaluation metric decreases, density increases because RMSE functions as a deviation controller. Additionally, the training and testing data points are drawn toward the center axis by the R2 evaluator. The figure below illustrates several other variables which also include but are not restricted to the linear regression model's centerline, which is positioned at the location $Y=X$, as well as dual lines that are in red below and above the midline, in that order, at $Y=0.9X$ and $Y=1.1X$. The lower and upper ends of the line intersections provide the false predictions of an underestimation and an overestimation of values,

respectively. Three schemes were produced by the subsequent analysis, which combined the SVR scheme with the three optimizer strategies applied to training and testing. Fig. 2 shows the findings of the current investigation. R2 of SVAF appears to be comparatively more favorable than the rest of the schemes because the data points maintain the same directionality and are nearer the centerline. From empirical data, it can be induced that in all cases, and quite noticeable in the case of SVDO, the precision of the test phase values is higher than that of the training phase. Overall, the result from the acquired data in Fig. 2 is the most favorable result using the SVR method and the ATF optimizer since R2 and RMSE in learning and validation also gave the best result. That could be due to the capability of this model in terms of minimizing error and being the best in performance regarding the R2.

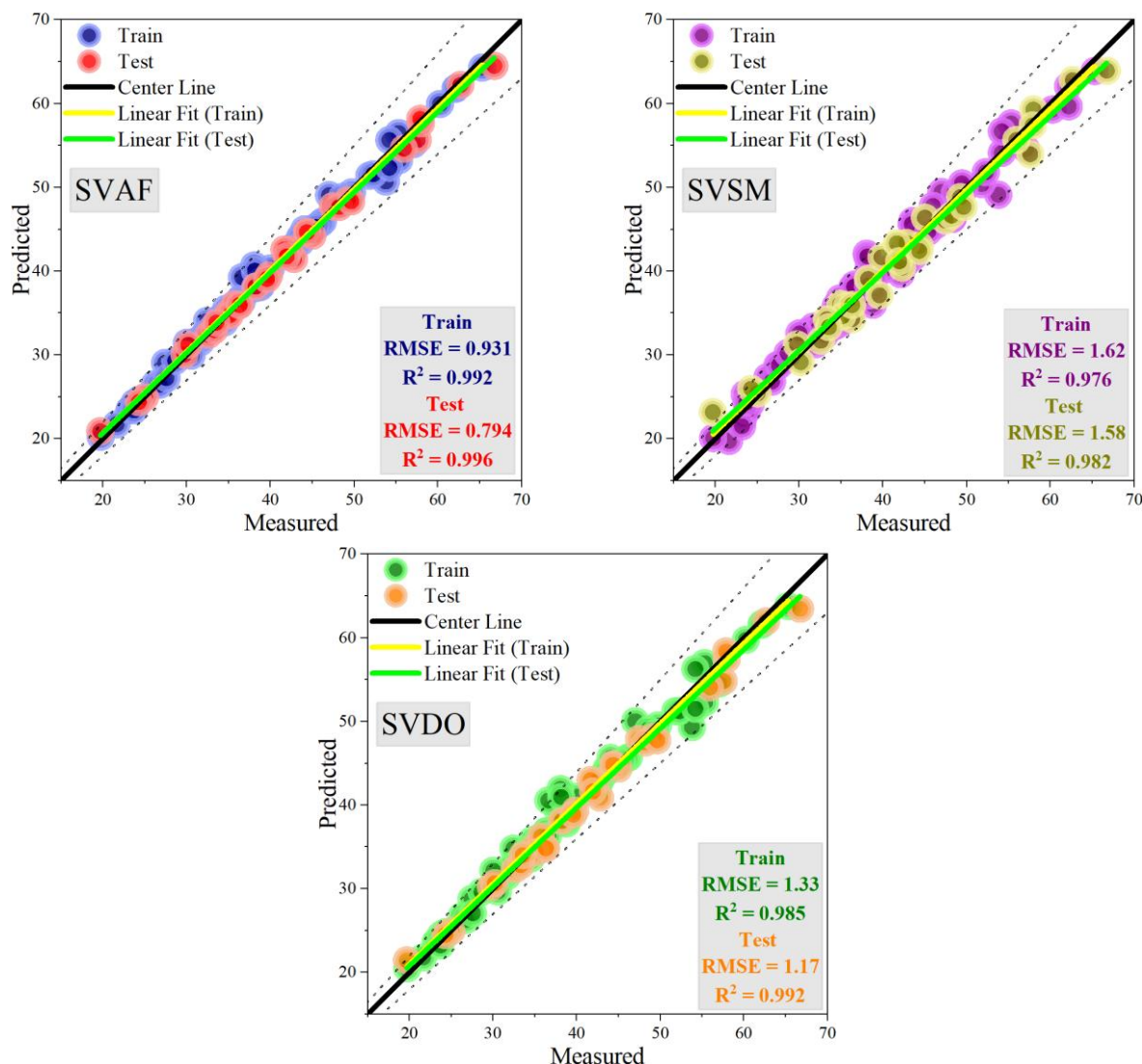


Figure 2: The scatter plot of expected and measured values

Fig. 3 presents the correlation between expected and actual CBR values obtained using three different classes of hybrid schemes. The graphs have been divided into two

distinct parts: model training and model validation. Among them, the SVAF representing an SVR and the ATF algorithm generate closer agreement between the

gauged CBR values of the expected output for testing and training data sets. By contrast, the status of the least unfavorable agreement appears quite clearly in SVR and AOSMA's union, SVSM.

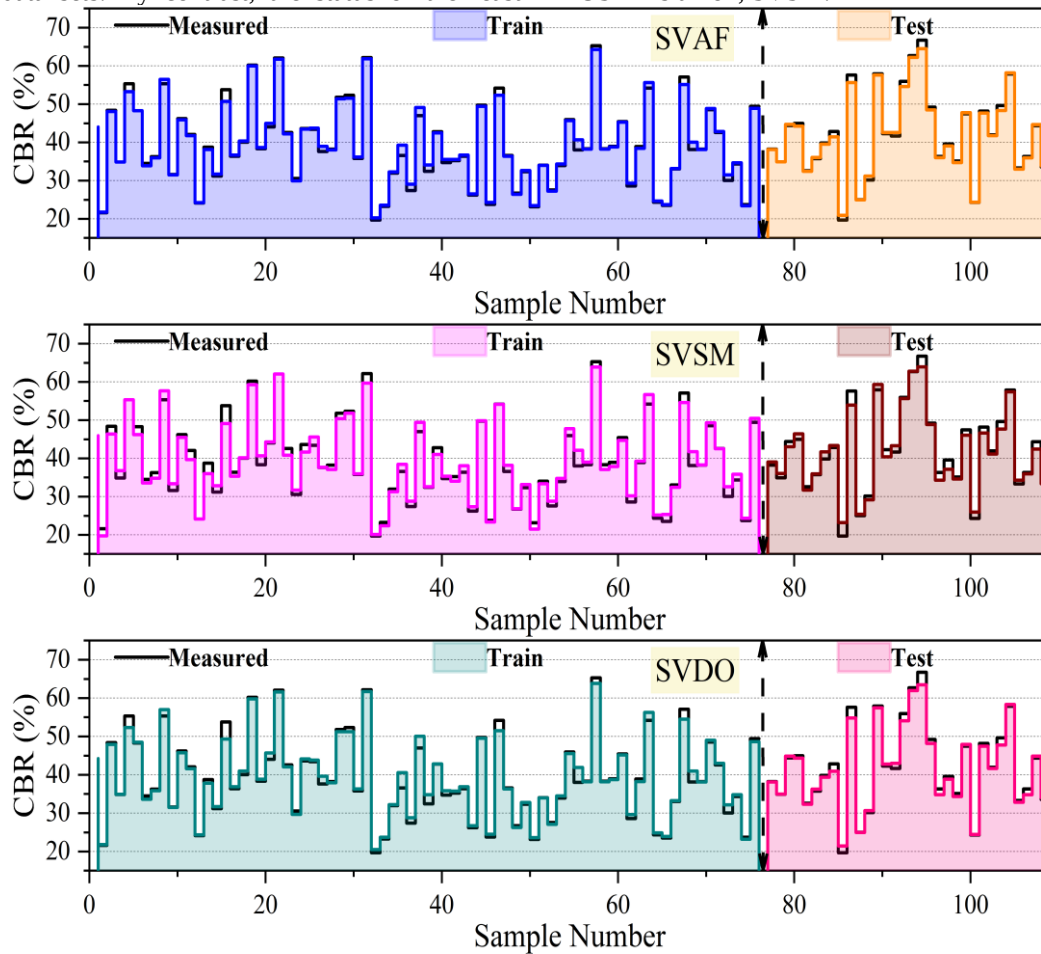


Figure 3: The comparison line-symbol plot between expected and gauged CS

Fig. 4 presents the deviations between the gauged and estimated values through three hybrid schemes regarding the California Bearing Ratio. This figure indicates that the greatest error for SVSM when assessed is around 18%, whereas for schemes undergoing training, it was 12% in

the same set. The figure shows that, for the highest and lowest performing schemes, the majority of errors are found in a narrower range of $(-3,3)$ % in SVAF and $(-6,17\%)$ % in SVSM.

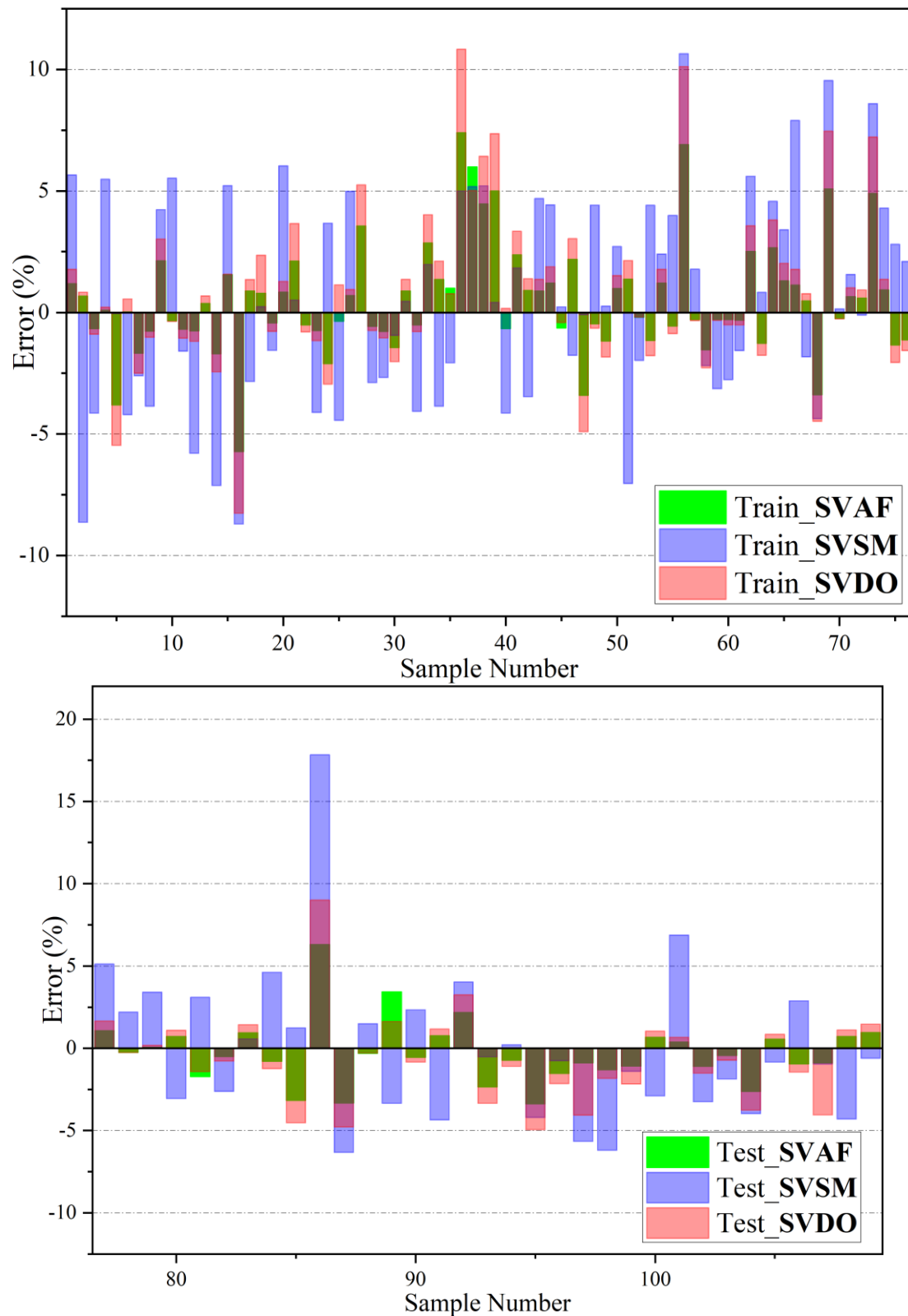


Figure 4: The error distribution of the schemes over samples shown in a time series plot.

The errors in the observed values of the undrained shear strength for the three different hybrid scheme types—SVAF, SVSM, and SVDO—are displayed in Fig. 5. Based on this figure, the maximum errors are about 11

%, and 7% for SVSM during training and testing of the schemes. The figure reflects the distribution of 25-75% of errors in a range less than $(-1, 1)$ % in SVAF and $(-3, 3)$ % in SVSM: best and worst schemes, respectively.

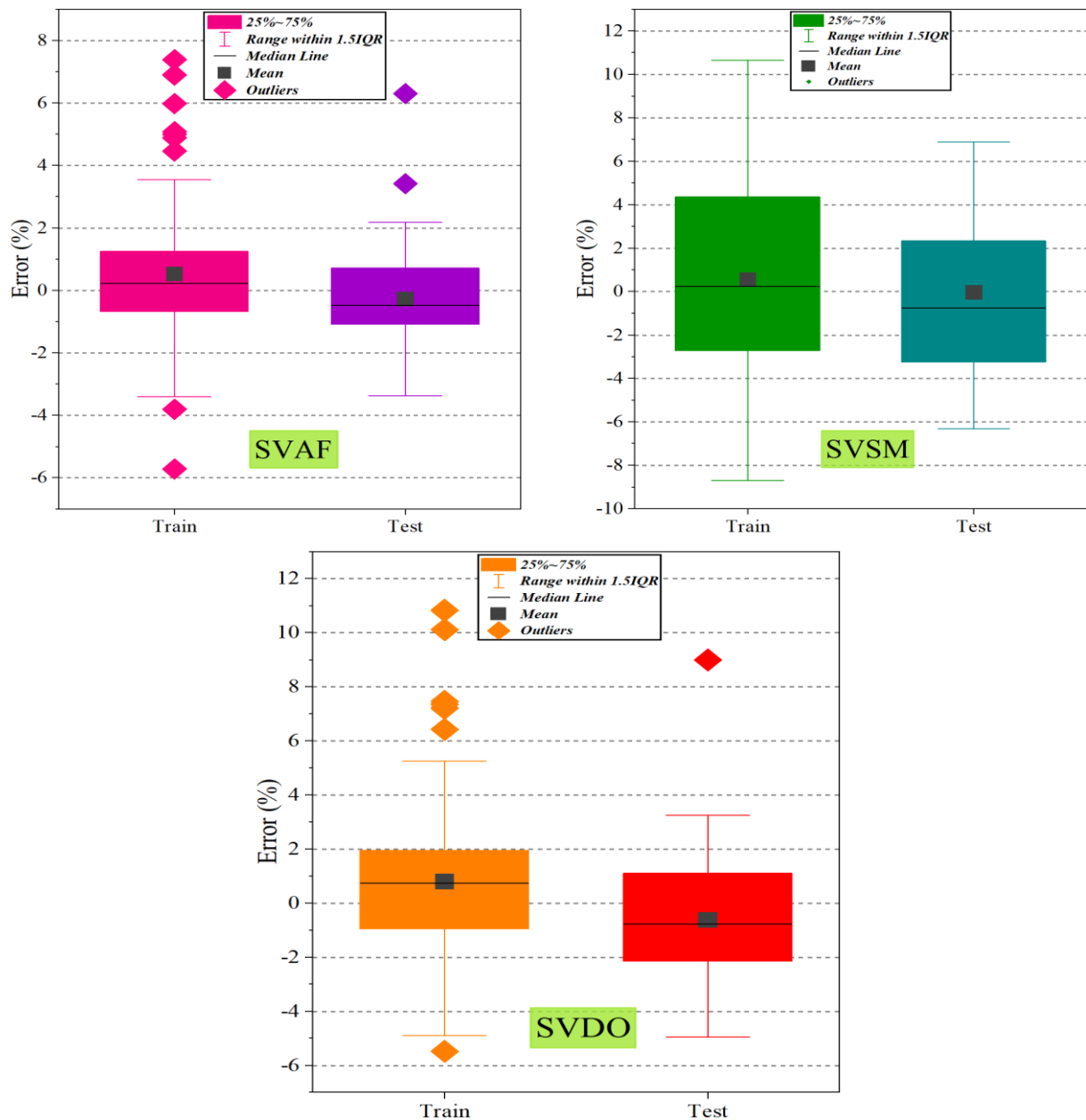


Figure 5: The standard half-box plot showing the error ratio of the hybrid schemes created.

To enhance the statistical robustness of the proposed models, 95% confidence intervals for the R^2 values were calculated based on multiple independent runs of each algorithm. As shown in Table 4, the standard SVR model has the widest interval, from 0.6302 to 0.7631, indicating greater variability and less predictive stability. In contrast, the three hybrid SVR models display narrower intervals with higher upper bounds, signifying more consistent performance. Among these, the SVR model combined with the Alibaba and Forty Thieves algorithms (SVAF) achieved the most favorable confidence interval, from 0.7243 to 0.8078, reflecting both high accuracy and

robustness across runs. The SVR-Dingo Optimization Algorithm model also performed well, with a confidence interval of 0.7120 to 0.8298, slightly broader but with the highest upper bound. Meanwhile, the SVR-AOSMA model shows an interval between 0.6653 and 0.7848, ranking it between the other hybrids in stability and performance. These intervals confirm that the SVAF model not only offers high prediction accuracy but also delivers consistent results, making it the most reliable model among those tested for CBR estimation.

Table 4: Confidence intervals based on R²

Model	Lower Bound	Upper Bound
SVR	0.6303	0.7632
SVR + Dingo Optimization Algorithm	0.7120	0.8298
SVR + Adaptive Opposition Slime Mould Algorithm	0.6653	0.7848
SVR + Alibaba and the Forty Thieves	0.7243	0.8078

4 Sensitivity analysis

The ANOVA-based sensitivity analysis conducted on the performance of different predictive models for estimating the California Bearing Ratio (CBR) reveals statistically significant differences among the models. The confidence intervals for the coefficient of determination (R²) provide insight into each model's accuracy and robustness. The baseline SVR model exhibits the lowest performance with a confidence interval ranging from 0.630 to 0.763,

indicating relatively limited predictive power. In contrast, the SVR models enhanced with metaheuristic algorithms demonstrate superior performance. Among these, the SVR-Dingo Optimization Algorithm model shows a confidence interval between 0.712 and 0.830, reflecting substantial improvement over the baseline. Similarly, the SVR-Adaptive Opposition Slime Mould Algorithm model yields a confidence range of 0.665 to 0.785, suggesting better stability and generalization. Notably, the SVR-Alibaba and the Forty Thieves (SVAF) model achieves the highest lower bound (0.724) and an upper bound of 0.808, indicating both high precision and consistent performance. The limited overlap between the confidence intervals of the SVAF model and those of the other models supports the claim of its statistically significant superiority. This distinction highlights the effectiveness of the AFT optimizer in enhancing SVR's learning capability and minimizing prediction errors. Overall, the results of the ANOVA test confirm that metaheuristic-optimized SVR models, particularly SVAF, provide more accurate and reliable predictions of CBR values compared to the standard SVR approach.

Table 5: Sensitivity analysis based on ANOVA

Models	lower	upper
SVR	0.630	0.763
SVR-Dingo Optimization Algorithm	0.712	0.830
SVR-Adaptive Opposition Slime Mould Algorithm	0.665	0.785
SVR-Alibaba and the Forty Thieves	0.724	0.808

5 Discussion

This section compares the three hybrid models—SVAF (SVR + AFT), SVSM (SVR + AOSMA), and SVDO (SVR + DOA)—focusing on their predictive accuracy, convergence behavior, and computational efficiency. As shown in Table 2, SVAF outperforms the others across all five metrics: R², RMSE, MSE, RSR, and WAPE. During testing, SVAF achieved the highest R² (0.0.9968) and the lowest RMSE (0.7946), indicating excellent generalization and minimal error in estimating CBR values. This success stems from the adaptive balance between exploration and exploitation in the Alibaba and Forty Thieves (AFT) optimization strategy, which enhances SVR's ability to find optimal hyperparameters. The random surveillance mechanism in AFT promotes global search, while Marjaneh's intelligence adjustment enhances local refinement, enabling rapid convergence toward optimal SVR settings. In contrast, the SVSM model, which employs the Adaptive Opposition Slime Mold Algorithm, showed weaker performance (R² = 0.9825, RMSE = 1.5824 during testing). Although AOSMA incorporates opposition-based learning to boost exploration, it can produce more oscillatory convergence patterns, possibly leading to suboptimal SVR tuning. Its complex adaptive threshold settings may also increase sensitivity to initial parameters. The SVDO model (SVR + Dingo Optimization Algorithm) performed moderately (R² = 0.992, RMSE = 1.171). DOA utilizes biologically

inspired social hunting behaviors, facilitating effective neighborhood search. However, its slower convergence during exploitation may limit its ability to finely tune SVR hyperparameters, especially in high-dimensional spaces. Regarding computational efficiency, SVAF requires slightly more training time than SVSM and SVDO due to multiple adaptive conditions and surveillance cycles in AFT, but its superior accuracy justifies this. SVSM offers faster runtimes but less predictive precision. SVDO falls between the two in terms of performance and computational demand. Overall, findings suggest that SVAF provides the best balance between accuracy and optimization quality, making it a strong candidate for practical CBR prediction tasks. Future research could explore combining AOSMA's rapid convergence with AFT's stability to improve training efficiency without sacrificing accuracy. Future research will aim to improve the models' applicability across various regions by testing them on datasets with diverse soil types. Combining Support Vector Regression with deep learning—for example, as a post-processing tool after deep feature extraction—could boost prediction accuracy, particularly for large or complex datasets. Another valuable approach is integrating these hybrid AI models into geotechnical software platforms, allowing real-time, data-driven decision-making in engineering and construction projects.

Although the hybrid SVR models presented demonstrated strong predictive performance on the available dataset, there are some limitations to consider. Firstly, without an external validation set, the

generalizability of the results may be restricted beyond the current data. Secondly, the relatively small sample size increases the risk of overfitting, especially with the use of metaheuristic optimization. Additionally, the dataset only encompasses a limited range of soil types and regions, which could limit the models' broader applicability. It is also important to note that larger, more diverse datasets might benefit from alternative modeling techniques such as deep learning or ensemble methods to achieve better predictive accuracy. These limitations will be addressed in future research to improve the model's robustness and generalizability. To enhance model robustness, we plan to use regularization like L1/L2 penalties and early stopping to prevent overfitting. Models will be tested under various conditions—smaller datasets and more noise—to check resilience. Including confidence intervals or error margins for metrics like RMSE and R^2 will better measure uncertainty. These steps will help create more reliable, generalizable models for geotechnical uses.

6 Conclusion

The current investigation has adopted an SVR scheme to project the CBR value of soil. Although the outcomes of the conventional method were effective, it had some limitations. The laboratory process is costly and is not considered to be time-effective. The drawbacks above can be overcome by substituting the software-based approach with artificial intelligence. The accuracy of the system in predicting the CBR was quite remarkable. The input variables were selected to forecast the target parameter, which was depicted as CBR. Five different performance metrics were utilized to appraise the precision delivered by the schemes under consideration. These included R^2 , RMSE, MSE, RSR, and WAPE. Three distinct metaheuristic optimization approaches—the Dingo Optimization Algorithm, Alibaba, the Forty Thieves Optimization algorithm, and the Adaptive Opposition Slime Mold Algorithm—have been examined in the current study to increase the system's functional efficiency. The conclusions below may be drawn from the analysis's outcome:

- The thorough analysis of the pertinent characteristics was the foundation for developing the projection schemes to estimate CBR. A comparison between the experimental outcomes and those obtained utilizing the suggested schemes showed that the latter's CBR prediction accuracy was significantly high.
- In the current research, the test phase has shown that the forecast data's scattering value increased by 0.39, 0.59, and 0.69 for SVAF, SVSM, and SVDO, respectively, from the training phase.
- The California Bearing Ratio outcomes presented in this investigation indicate a significant discrepancy between the observed and projected values, with an average underestimate of almost 1.24 for the suggested schemes. With a value of 1.6271, the RMSE displayed its maximum error in the scheme's SVSM in the training phase. The SVAF had the lowest error rate in the testing session, with a rating of 0.7946.

Acknowledgements

We wish to state that no individuals or organizations require acknowledgment for their contributions to this investigation.

Authorship contribution statement

Na Feng: Writing-Original initial drafting, Conceptualization, Supervision, Project administration.

Yulin Lan: Methodology, Software

Zhisheng Yang: Formal Analysis, Language Review

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Funding

This investigation was not funded by any specific grant from public, commercial, or charitable funding bodies.

Ethical approval

The paper has attained ethical approval from the institutional review board, ensuring the protection of participants' rights and compliance with the relevant ethical guidelines.

References

- [1] A. Chegenizadeh and H. Nikraz, "CBR test on reinforced clay," in *Proceedings of the 14th Pan-American Conference on Soil Mechanics and Geotechnical Engineering (PCSMGE), the 64th Canadian Geotechnical Conference (CGC)*, Canadian Geotechnical Society, 2011.
- [2] T. F. Kurnaz and Y. Kaya, "Prediction of the California bearing ratio (CBR) of compacted soils by using GMDH-type neural network," *The European Physical Journal Plus*, EPJ Plus, vol. 134, Jul. 2019. <https://doi.org/10.1140/epjp/i2019-12692-0>.
- [3] H. B. Seed and P. De Alba, "Use of SPT and CPT tests for evaluating the liquefaction resistance of sands," in *Use of in situ tests in geotechnical engineering*, ASCE, 1986, pp. 281–302.
- [4] M. Gams and T. Kolenik, "Relations between electronics, artificial intelligence and information society through information society rules," *Electronics (Basel)*, MDPI, vol. 10, no. 4, p. 514, 2021. <https://doi.org/10.3390/electronics10040514>.
- [5] R. W. Day, *Soil testing manual*. McGraw-Hill, 2001.
- [6] M. M. E. Zumrawi, "Prediction of CBR Value from Index Properties of Cohesive Soils.,"

- University of Khartoum Engineering Journal, vol. 2, no. ENGINEERING, 2012.
- [7] W. P. M. Black, “A method of estimating the California bearing ratio of cohesive soils from plasticity data,” *Geotechnique*, ICE Virtual Library, vol. 12, no. 4, pp. 271–282, 1962. <https://doi.org/10.1680/geot.1962.12.4.271>.
- [8] K. B. Agarwal and K. D. Ghanekar, “Prediction of CBR from plasticity characteristics of soil,” in *Proceeding of 2nd South-east Asian Conference on Soil Engineering, Singapore. June, 1970*, pp. 11–15.
- [9] M. Linveh, “Validation of correlations between a number of penetration test and in situ California bearing ratio test,” *Transp Res Rec*, vol. 1219, pp. 56–67, 1989.
- [10] D. J. Stephens, “The prediction of the California bearing ratio,” *Civil Engineering= Siviele Ingenieurswese*, Sabnet, vol. 1990, no. 12, pp. 523–528, 1990. https://hdl.handle.net/10520/AJA10212019_14356.
- [11] T. Al-Refeai and A. Al-Suhaibani, “Prediction of CBR using dynamic cone penetrometer,” *Journal of King Saud University-Engineering Sciences*, Elsevier, vol. 9, no. 2, pp. 191–203, 1997. [https://doi.org/10.1016/S1018-3639\(18\)30676-7](https://doi.org/10.1016/S1018-3639(18)30676-7).
- [12] M. W. Kin, “California bearing ratio correlation with soil index properties,” *Master degree Project, Faculty of Civil Engineering, University Technology Malaysia*, 2006. https://eprints.utm.my/4064/1/MakWaiKinMFK_A2006.pdf.
- [13] S. R. CNV and K. Pavani, “MECHANICALLY STABILISED SOILS-REGRESSION EQUATION FOR CBR EVALUATION,” 2006.
- [14] P. Vinod and C. Reena, “Prediction of CBR value of lateritic soils using liquid limit and gradation characteristics data,” *Highway Research Journal, IRC*, vol. 1, no. 1, pp. 89–98, 2008.
- [15] R. S. Patel and M. D. Desai, “CBR predicted by index properties for alluvial soils of South Gujarat,” in *Proceedings of the Indian geotechnical conference, Mumbai*, 2010, pp. 79–82.
- [16] G. Ramasubbarao and S. G. Sankar, “Predicting soaked CBR value of fine grained soils using index and compaction characteristics,” *Jordan Journal of Civil Engineering*, vol. 7, no. 3, pp. 354–360, 2013.
- [17] M. Alawi and M. Rajab, “Prediction of California bearing ratio of subbase layer using multiple linear regression models,” *Road Materials and Pavement Design*, Taylor & Francis, vol. 14, no. 1, pp. 211–219, 2013. <https://doi.org/10.1080/14680629.2012.757557>.
- [18] H. Ghanadzadeh, M. Ganji, and S. Fallahi, “Mathematical model of liquid–liquid equilibrium for a ternary system using the GMDH-type neural network and genetic algorithm,” *Appl Math Model*, Elsevier, vol. 36, no. 9, pp. 4096–4105, 2012. <https://doi.org/10.1016/j.apm.2011.11.039>.
- [19] M. A. Shahin, M. B. Jaksa, and H. R. Maier, “Artificial neural network applications in geotechnical engineering,” *Australian geomechanics*, vol. 36, no. 1, pp. 49–62, 2001.
- [20] J. A. Abdalla, M. F. Attom, and R. Hawileh, “Prediction of minimum factor of safety against slope failure in clayey soils using artificial neural network,” *Environ Earth Sci*, Springer, vol. 73, pp. 5463–5477, 2015. <https://doi.org/10.1007/s12665-014-3800-x>.
- [21] B. Yildirim and O. Gunaydin, “Estimation of California bearing ratio by using soft computing systems,” *Expert Syst Appl*, Elsevier, vol. 38, no. 5, pp. 6381–6391, 2011. <https://doi.org/10.1016/j.eswa.2010.12.054>.
- [22] Tja. Taskiran, “Prediction of California bearing ratio (CBR) of fine grained soils by AI methods,” *Advances in Engineering Software*, Elsevier, vol. 41, no. 6, pp. 886–892, 2010. <https://doi.org/10.1016/j.advengsoft.2010.01.003>.
- [23] S. Bhatt, P. K. Jain, and M. Pradesh, “Prediction of California bearing ratio of soils using artificial neural network,” *Am. Int. J. Res. Sci. Technol. Eng. Math*, vol. 8, no. 2, pp. 156–161, 2014.
- [24] T. Q. Ngo, L. Q. Nguyen, and V. Q. Tran, “Novel hybrid machine learning models including support vector machine with meta-heuristic algorithms in predicting unconfined compressive strength of organic soils stabilised with cement and lime,” *International Journal of Pavement Engineering*, Taylor & Francis, vol. 24, no. 2, p. 2136374, 2023. <https://doi.org/10.1080/10298436.2022.2136374>.
- [25] X. Wu, F. Lu, and T. He, “Exploring the potential of machine learning in predicting soil California bearing ratio values,” *Periodica Polytechnica Civil Engineering*, vol. 69, no. 2, pp. 551–566, 2025. <https://doi.org/10.3311/PPci.38678>.
- [26] V. Bherde, L. Kudlur Mallikarjunappa, R. Baadiga, and U. Balunaini, “Application of machine-learning algorithms for predicting California bearing ratio of soil,” *Journal of Transportation Engineering, Part B: Pavements*, ASCE Library, vol. 149, no. 4, p. 4023024, 2023. <https://doi.org/10.1061/JPEODX.PVENG-1290>.
- [27] J. Ma et al., “A comprehensive comparison among metaheuristics (MHs) for geohazard modeling using machine learning: Insights from a case study of landslide displacement prediction,” *Engineering Applications of Artificial Intelligence*, Elsevier, vol. 114, p. 105150, 2022. <https://doi.org/10.1016/j.engappai.2022.105150>.
- [28] V. N. Vapnik, “The nature of statistical learning,” *Theory*, 1995.
- [29] V. Vapnik, “Statistical Learning Theory. New York: John Wiley & Sons,” Inc, 1998.
- [30] M. K. Naik, R. Panda, and A. Abraham, “Adaptive opposition slime mould algorithm,” *Soft comput*, Springer, vol. 25, no. 22, pp. 14297–

- 14313, 2021. <https://doi.org/10.1007/s00500-021-06140-2>.
- [31] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, “Slime mould algorithm: A new method for stochastic optimization,” *Future Generation Computer Systems*, Elsevier, vol. 111, pp. 300–323, 2020. <https://doi.org/10.1016/j.future.2020.03.055>.
- [32] H. R. Tizhoosh, “Opposition-based learning: a new scheme for machine intelligence,” in *International conference on computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce (CIMCA-LAWTIC’06)*, Vienna, Austria, IEEE, 2005, pp. 695–701. <https://doi.org/10.1109/CIMCA.2005.1631345>.
- [33] M. Braik, M. H. Ryalat, and H. Al-Zoubi, “A novel meta-heuristic algorithm for solving numerical optimization problems: Ali Baba and the forty thieves,” *Neural Comput Appl*, Springer, vol. 34, no. 1, pp. 409–455, 2022. <https://doi.org/10.1007/s00521-021-06392-x>.
- [34] A. K. Bairwa, S. Joshi, and D. Singh, “Dingo Optimizer: A Nature-Inspired Metaheuristic Approach for Engineering Problems,” *Math Probl Eng*, Wiley Online Library, vol. 2021, p. 2571863, 2021. <https://doi.org/10.1155/2021/2571863>.

