Using ML Methods For Obtaining The Prediction Of Maximum Dry Density

Ying Liu and Liguo Teng*

Applied Technology College of Dalian Ocean University, Dalian 116300, China

Corresponding author. E-mail: tengliguo@126.com

Received: May 18, 2025; Accepted: Jul. 30, 2025

In geotechnical engineering, maximum dry density (MDD) is a fundamental concept, referring to the maximum mass per unit volume under specified conditions. MDD is pivotal for ensuring the stability of earthwork, including foundations and dams. It depends on factors like moisture, compaction force, grain size, and soil type. Accurate MDD prediction helps engineers ensure durability and reliability in civil engineering projects over time. This article, an original prediction method of MDD by means of the K-nearest neighbors (KNN) algorithm has been presented. Using the KNN method, accurate models can be developed that connect the MDD of treated soil to diverse natural soil attributes, including linear shrinkage, plasticity, particle size dispersion, and the type and content of stabilization additives. A large database comprising 187 samples representing various categories of soils from previously conducted stability tests has been utilized in this study to develop forecasting models and validate them. Besides, application of meta-heuristic strategies, specifically the Jellyfish Search Optimizer (JFO) and Black Widow Optimizing Algorithm (BWOA), further enhances the accuracy level of the KNN method in this work. Therefore, two hybrid models, KNJF and KNBW, are defined. Then, the data for the training, validation, and tests show excellent R2 values of 0.9895, 0.9681 and 0.9659, respectively, by the KNJF model. In addition, during the training phase KNJF has the best RMSE of 24.236. Everything taken into account, the KNJF model generalizes and predicts better than the models KNN and KNBW developed specifically for this investigation.

Keywords: K-nearest neighbor; Maximum Dry Density; Jellyfish Search Optimizer; Black Widow Optimization Algorithm(BWOA)

http://dx.doi.org/10.6180/jase.202605_29(5).0004

1. Introduction

Soil offers a diverse range of options in the countryside. However, not all soil types are suitable for every kind of soil construction. Each soil possesses its scientific method for dealing with force components, similar to other building materials. Soil is essential to the design of high-safety structures in civil engineering, particularly those that are in close contact with the ground, such as foundations, embankments, and soil-based buildings. These constructions need greater safety factors since soil and construction analysis are unpredictable [1–4]. Nevertheless, rock and soil are still necessary construction materials, whether left in their original state or improved with measures like consolidation, reinforcing, and compaction [5–7]. Mechanical soil compaction is a common technique used to enhance soil engineering attributes. The effectiveness of compaction is typically assessed by measuring the dry density about MDD [8]. Comprehending the features of soil compaction, particularly MDD, is crucial while working on construction projects such as dams, embankments for roads and trains, landfill liners, and retaining structure backfilling. MDD

offers information on soil quality following compaction but before treatment and acts as a performance indicator for stabilization [9–12]. To comprehend the complex relationships between soil characteristics and influencing factors and to do away with the need for comprehensive laboratory testing of MDD in each new building project, it is advisable to develop a framework for calculating MDD values. Such a model should take into account soil characteristics before stabilization, including texture, ductility, linear shrinkage, and the type and quantity of stabilizing additives [13, 14].

The laboratory test commonly employed to identify maximum unit weight on the dry side and optimum moisture content has also been known as the Proctor density test. The two commonly known tests are the standard Proctor (ASTM D698/AASHTO) and the justified Proctor (ASTM D1557/AASHTO T180). These tests, however, are time-consuming, expensive, and highly reliant on sample gathering personnel skill level and expertise, as well as result validation [15]. Machine learning (ML) and artificial intelligence (AI) tactics have gained importance in many fields, including geotechnical engineering. These techniques have displayed promise in identifying landslides, predicting floods, zoning groundwater potential, and predicting material properties [14]. AI techniques provide intelligent predictive schemes for geotechnical engineering issues, which necessitate developing precise models with extensive application areas [16–22].

Recently, the MDD of stabilized soils has also been forecast by several ML approaches. To project the MDD of stabilized soil, an artificial neural network (ANN) has been created, for example, Alavi et al. [13]. The multilayer perceptron (MLP) network architecture of the chosen models included layers containing neurons and connections. The performance analysis of the schemes suggested that the models provide satisfactory performance. Besides, Das et al. [23] has attempted applying an SVM (support vector machine) as well as three distinct ANN algorithms to forecast the MDD of soils stabilized by cement. Levenberg-Marquardt, Bayesian regularization, and differential evolution are models that have been applied. The writers had the opinion that the SVM outperforms the models based on ANN performance indicators. As research on projecting the MDD and UCS of stabilized soil, Saadat and Bayat [24] has implemented the application of nonlinear regression (NLR) and an adaptive neuro-fuzzy inference system (ANFIS). The writer had the opinion that an ANFIS UCS projection scheme outperforms nonlinear regression after it had conducted the two. The UCS and MDD of compacted soils have also been modeled by applying the ANN and ANFIS approaches [25]. Comparing the two models' effectiveness, the writers established that the ANFIS outperforms the ANN. In another study, Suman et al. [12] has implemented the application of several ML algorithms to project the MDD of soils stabilized by cement. The performance analysis of the choice of implemented ML algorithms, multivariate adaptive regression splines, and functional networks, against four models (Bayesian regularization, differential evolution, Levenberg-Marquardt, as presented by Das et al. [23]), SVM, and ANN orrespondingly, suggested that the implemented algorithms outshine models implemented by ANN, as well as those implemented by SVM.

Tejani et al. [26] introduced a novel tactic for predicting the MDD of soil stabilization blends using the Naive Bayes (NB) algorithm, incorporating soil features including particle size dispersion, plasticity, linear shrinkage, and stabilizer type and dosage. To improve accuracy, Artificial Rabbits Optimization (ARO) and Gradient-based Optimizer (GBO) were applied, producing three models: NB, NBAR, and NBGB. The NBAR model performed best with an R^2 of 0.9903 and RMSE of 34.563, showing the strength of mixid ML models in MDD prediction for geotechnical applications. Ali et al. [27] developed predictive models for Optimum Moisture Content (OMC) and MDD using six input variables: gravel (G), sand (S), fines (F), PL, LL, and plasticity index (PI). A dataset of 2,162 samples was used to evaluate four models: ANN, NLR, linear regression (LR), and multilinear regression (MLR). ANN achieved the highest R^2 for OMC (0.92), while MLR and LR had the lowest errors for MDD despite ANN yielding the highest R^2 (0.87). Sensitivity analysis showed PL was most influential for OMC and gravel content for MDD, affirming the practical value of ML in compaction prediction. Duc et al. [28] proposed ANN-MLP and RF models to estimate MDD and OMC, aiming to replace labor-intensive Proctor tests. Input features included silt, clay, LL, PL, PI, and specific gravity-variables strongly linked to compaction behavior. Evaluated using MAE, RMSE, and R^2, both models showed similar accuracy (R^2 of 0.829 for ANN-MLP and 0.827 for RF), demonstrating the reliability of ML in predicting compaction properties for embankment fill soils. Jia and Zheng [29] developed a theoretical MDD formula for over coarse-grained soils, incorporating particle density, original soil MDD, and giant particle mass fraction. Validation against Discrete Element Method (DEM) results—across seven mixed soil samples-revealed lower MDD values from DEM, especially as coarse particle proportion increased. To address this, a volume growth coefficient was introduced to capture pore volume changes due to particle embedding. The refined model aligned well with DEM outcomes, confirming its effectiveness for

Table 1. Quantitative features of MDD and inputs

Indexes	Variables						
	Intake						Goals
	Soil (%)	Cement (%)	Lime (%)	LL (%)	PL (%)	PI (%)	MDD (kg/m^3)
Max	100	30	30	102	58.24	70	2210
Min	70	0	0	18	12	0	1200
Avg	93.604	3.807	2.588	39.428	22.673	16.755	1780.61
St. Dev.	4.6366	4.316	4.086	16.763	9.412	12.694	227.508

coarse-grained soil applications.

This paper involves developing an ML model to forecast Maximum Dry Density (MDD) utilizing an experimental data source from reputed sources. Using a K-nearest neighbor (KNN) model, coupled with Jellyfish Search Optimizer (JFO) and BWOA, predicts the MDD value of the soil confidently. The KNN is utilized for prediction of MDD due to its ability to form accurate models by utilizing nearness relations. Using natural soil attributes, including particle size distribution, as well as stabilizing additives, KNN accurately extracts patterns, aiding accurate prediction of MDD in geotechnical engineering. The motivation for introducing the KNN model optimized by the JFO (KNJF) lies in its ability to enhance prediction accuracy by combining the simplicity and interpretability of KNN with the powerful global optimization capabilities of JFO. While KNN effectively captures local patterns based on proximity relationships within the data, its performance is highly dependent on the selection of optimal hyperparameters. By integrating the JFO meta-heuristic algorithm, the KNJF model dynamically optimizes these hyperparameters, enabling the model to adapt more effectively to the intricate, nonlinear links present in soil behavior data.

The next section will explain the data acquisition process, discussing the intricacies involved in utilizing the data. A detailed description of the basic concepts of the ML model, the architecture, as well as the building blocks, shall be included. In addition, the choice, application, and utilization of the optimization processes shall also be detailed, discussing approaches used to optimize performance. The performance evaluators, responsible for analyzing accuracy, as well as efficiency, shall also be covered. The findings shall be analyzed in Section 3, while an abridged summary, as well as conclusion, shall be included in the last section.

2. Materials and methodology

2.1. Data gathering

The proportions of lime and cement, PI, Liquid Limit (LL), and Plastic Limit (PL), are among the six elements in the sophisticated technique of calculating the MDD of chemically stabilized soil.

- The proportions of lime and cement represent the amounts of chemical stabilizers added to the soil to boost its engineering attributes.
- The LL is the water content at which soil alters from a plastic to a liquid state, indicating soil consistency.
- The PL is the water content at which soil begins to exhibit plastic behavior.
- The PI is computed as the disparity between LL and PL (PI = LL - PL) and measures the plasticity and moisture retention capacity of the soil.

These features collectively reflect both the chemical composition and geotechnical attributes of the stabilized soil. Together, they serve as key input variables in the sophisticated method used to project the MDD of chemically stabilized soil.

As seen in Table 1, the data acquisition process entails collecting 187 samples from various places and evaluating them in the lab. Preprocessing steps were carried out to ensure the quality and consistency of the input data before training the models. Specifically, the dataset was randomly permuted using the randperm operation, which helps eliminate any inherent ordering bias and ensures a fair distribution of data samples across training, validation, and testing subsets. Following randomization, feature normalization was applied to scale the input parameters to a uniform range. This step is pivotal in distance-based frameworks, like KNN, as it prevents attributes with larger numeric ranges from disproportionately influencing the model's predictions. In this study, the database was explicitly split into 70% training, 15% validation, and 15% testing subsets to facilitate scheme development, hyperparameter tuning, and performance evaluation, respectively.

Routine steps are conducted to compute the soil, lime, and cement percentages. In addition, values for LL, PL, and PI are found by applying the Atterberg Limits test. The

LL shows the water content when the chemically stabilized soil shifts from plastic to liquid, while the PL shows the transition of the soil from plastic to semi-solid. The PL vs. LL differentiation is used to find the PI [23, 30, 31]. These parameters play an essential role in understanding chemically stabilized soil behavior and form the foundation for estimating the MDD. Preprocessing processes, including random permutation, are executed to boost subsequent analysis accuracy. Results, maximum, minimum, standard deviation, and average values for samples are presented painstakingly in Table 1. Normalization, an essential preprocessing process in estimating MDD for chemically stabilized soil, boosts analysis accuracy by normalizing data values to an equal base. The adjustment keeps variable values with greater ranges from placing unequal weights on analysis. Through processes including the application of the 'randperm' function, data gets randomly permutated to provide equalization. Then, standard protocols measure Atterberg Limits values, as well as percentages of soil composition. Equitable representation by normalization ensures that every variable contributes an equal measure toward estimating MDD, thus aiding in chemically stabilized soil behavior understanding. By equalizing data ranges, normalization makes it possible to perform efficient correlation analysis across variables, describing influences toward MDD. Through this, engineers get empowered to optimize input values toward MDD's specified values.

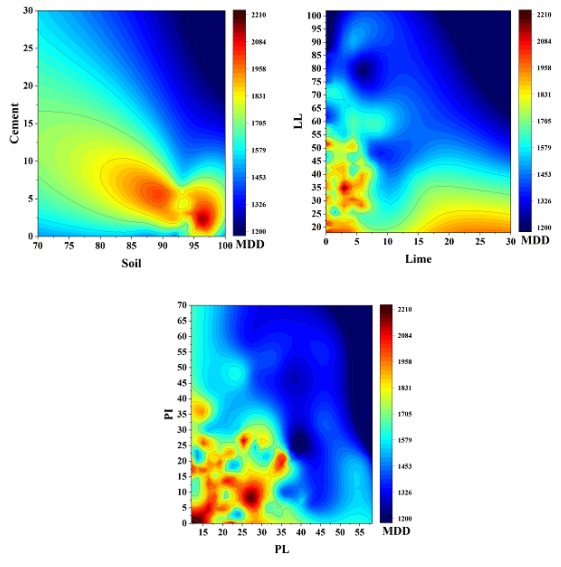

Fig. 1. The surface plot's representation of the link between input parameters and output

The MDD of chemically stabilized soil serves as the output variable, whereas Fig. 1 displays the complex correlation between the input variables, like the ratios of cement and lime, and the output variable. A close examination of the effect of variation in the values of the inputs on the MDD has been displayed by this graphical plot. The three-dimensional presentation by the surface plot shows the complexity of the correlation between different variables influencing the MDD. Every point on the surface has an attached MDD value together with an individual set of the inputs. The contours and gradients of the surface show areas where some inputs result in increased or decreased values of MDD, showing patterns and trends in data. The visualization can be used by scholars and engineers to get optimum values of inputs so that maximum MDD can be achieved or analyze how variation in values of inputs influences MDD under different scenarios. In addition, Fig. 2 demonstrates the correlation between inputs with each other and between input and output parameters. Drawing on this plot, the soil and cement have positive corelation with MDD, while the other parameters showed negative correlations.

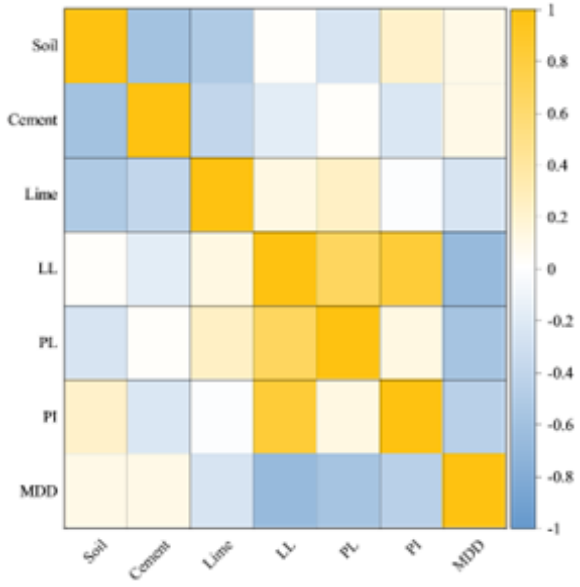

Fig. 2. The correlation between input parameters and output

2.2. ML Tactics

2.2.1. KNN-based

The KNN approach utilizes majority occurring feedback from the nearest K data points near the test point as prediction feedback for the model. Pre-normalization of these parameters must be addressed by applying Eq. (1) first.

$$x_{\text{normalization}} = \frac{x - \text{Min}}{\text{Max} - \text{Min}} \tag{1}$$

Next, Eq. (2) is used to compute the Euclidean distance between the data point as well as the test point.

$$H\left(x_i, x_j\right) = \left(\sum_{h=1}^{m} \left|x_i^{(h)} - x_j^{(h)}\right|^2\right)^{\frac{1}{2}} \quad (2)$$

Eq. (2) computes the distance H between the test point $\left(x_j\right)$ and the original data points (x_i) utilizing the Euclidean distance with m as the count of argument points. Because diverse parameters have distinct consequences on thermal comfort, even when the same value is modified, it is necessary to apply Eq. (3) to modify the Euclidean distance for all parameters to remove the disparate effects of indoor thermal comfort factors. For example, a 1 °C shift in air temperature has a greater impact than a 1% shift in air humidity.

$$H\left(x_i, x_j\right) = \left(\sum_{h=1}^{m} \left(w_h * \left|x_i^{(h)} - x_j^{(h)}\right|^2\right)\right)^{\frac{1}{2}} \quad (3)$$

The equation gives the weight (w_h) devoted to each interior temperature factor that influences thermal comfort.

The K nearest data points to a test point are found using distance calculations in KNN, and the test point's feedback is identified by which of them is provided the most frequently. Optimal K is chosen through cross-validation to balance sensitivity and accuracy. A small K makes the model overly sensitive to nearby points, introducing noise interference, while a large K may reduce accuracy.

2.2.2. JFO

The JFO, a recent swarm-oriented metaheuristic by Chou and Truong [32], emulates jellyfish foraging behavior in the ocean with three fundamental rules. The first involves jellyfish following currents in the water or moving in swarms, regulated by mechanisms for timing. The second rule has jellyfish seeking food, with areas rich in food attracting more jellyfish. According to the third rule, objective functions are assigned places and the amount of food discovered at each position determines the answer [33].

- Ocean current

In addition to detecting tiny planktonic creatures like fish eggs, larvae and phytoplankton, jellyfish can also sense ocean currents (G). Eq. (4) expresses this attribute:

$$G = X' - f_w u \quad (4)$$

G displays the ocean current direction, f_w denotes an attractiveness factor, and u indicates the central location of jellyfishes. The normal distribution of jellyfish indicates a higher probability of their presence within $\pm\beta z$ distance from the mean location. The length distribution coefficient G is denoted by $\beta(\beta > 0)$, and the standard deviation is determined by z. Assuming $[\mu \times \text{rand}\,\sigma(0,1)]$ displays standard deviation σ, the following can be derived using Eq. (5).

$$B(x) = \beta \times u \times c^x(0,1) \times c^z(0,1) \quad (5)$$

Eq. (5) is rewritten to simplify the computations as:

$$B(x) = \beta \times u \times c(0,1) \quad (6)$$

In addition, the new G is:

$$G = X' - \beta \times z \times c(0,1) \quad (7)$$

Each jellyfish's new location can be determined using Eq. (8):

$$X_i(h+1) = X_i(h) + c(0,1) \times G \quad (8)$$

The new and present positions of the $i_$th jellyfish are displayed above, respectively, by $X_i(h+1)$ and $X_i(h)$. After updating each jellyfish position by Eq. (7), a desirable position is assumed to be the currently located jellyfish. Furthermore, the timing method makes utilizes a timing function represented by Q(h) and a threshold constant. Similar to Eq. (9), the timing function is a random number varying from 0 to 1, showcasing a long-term, overall declining trend. The representation of the timing function is articulated in Eq. (9):

$$Q(h) = \left|\left(1 - \frac{h}{\text{MaxIter}}\right) \times (2 \times c(0,1) - 1\,|\right. \quad (9)$$

In this context, the time index is denoted by h, given by the repetition number, while MaxIter displays the peak tally of cycles [34].

2.2.3. BWOA

The BWOA is a meta-heuristic tactic inspired by the way black widow spiders reproduce, integrating evolutionary algorithms with distinct criteria [35]. It employs selection, reproduction, and mutation mechanisms, distinguishing itself through a unique balance between exploration and exploitation. Rooted in Darwinian principles, the algorithm mirrors the procreation behavior of Latrodectus mactans (black widow spiders) to generate novel solutions. The BWOA addresses intricate optimization problems by recognizing optimal solutions, generating progeny, and fostering exploration through mutation. This approach circumvents local optima and efficiently converges toward optimal solutions. Its ability to balance exploration and exploitation contributes to its remarkable performance in tackling complex optimization challenges, as validated by research [36, 37].

The following succinctly describes the main stages of BWOA:

- Initialization

In this stage, the population is the total count of widows of size M. Each widow is displayed by an array of $1 \times M_{\text{var}}$, which provides the problem's solution. This array is depicted as widow $= (x_1, x_2, \ldots, x_{M_{\text{var}}})$, where M_{var} is he optimization problem's dimension. Also, M_{var} is described as the number of threshold values needed to be attained by the scheme, while x_i displays the $i-th$ candidate resolution.

The fitness of a widow is attained by appraisal of the fitness function of f of each widow of the collection $(x_1, x_2, \ldots, x_{M_{\text{var}}})$. Then fitness $= f($ widow $)$, which can be displayed by: f itness $= (x_1, x_2, \ldots, x_{M_{\text{var}}})$. The next step in the reproductive process is the random selection of parent couples who participate in the mating phase, in which the female black widow either copulates or eats the male.

- Procreate

In procreation, an alpha β must be generated as a long widow array encompassing random numbers. Then offspring is generated by deploying α and Eq. (10) in which x_1 and x_2 signify parents, y_1 and y_2 signify offspring. The crossover outcome is processed and preserved.

$$\begin{aligned} y_1 &= \beta \times x_1 + (1-\beta) \times x_2 \text{ and} \\ y_2 &= \beta \times x_2 + (1-\beta) \times x_1 \end{aligned} \tag{10}$$

- Cannibalism

There are three different categories into which the phenomenon of cannibalism can be divided: filial, sibling, and sexual cannibalism, frequently noted when the offspring spiders consume their parent. Upon implementation of the cannibalism mechanism, the resulting population is subsequently assessed and assigned to a variable denoted as $p2$.

- Mutation

Mutation in the BWOA involves stochastically selecting individuals from the Mutep population for mutation. The mutated population, labeled p3, undergoes evaluation and is recorded within the novel population. The new population, derived from migrating p3 and p2, is sorted to identify the optimal widows based on the M_{var} dimension's threshold values.

2.3. Hybridization

This pioneering research utilizes a cutting-edge hybrid approach that combines the JFO and BWOA methodologies to boost the accuracy of predicting Moisture-Density Relationship (MDD). The KNN model's hybridization procedure smoothly integrates the special advantages of BWOA and JFO, leading to developing three distinctive mixed schemes: KNJF, KNBW, and KNN. The objective of this strategic integration is to leverage the synergies between BWOA as well as JFO, supporting the development and expansion of the KNN model's predictive powers in MDD estimation. The hybridization process unfolded through the subsequent steps:

- In the initial phase, a detailed integration of optimized parameters derived from BWOA and JFO into the KNN model was meticulously performed. This step, deemed pivotal in the hybridization process, encompassed the fusion of the distinct strengths inherent in each optimizer. The resulting hybrid models, KNJF and KNBW, were painstakingly created by balancing these optimized parameters, each of which embodies a distinct synergy obtained from the contributing algorithms.

- During the subsequent stage, an exhaustive database was meticulously assembled, serving as the foundational element for training three models. This database comprises a wide array of parameters associated with input variables and soil properties, skillfully chosen from a variety of literary sources to guarantee comprehensiveness. The inclusive selection is intended to take into account the complexities of the real world, strengthening the hybrid models' ability to project MDD in engineering applications.

2.4. Performance evaluation tactics

This article analyzes schemes deploying some metrics, like the correlation coefficient (R^2), root mean square error (RMSE), mean square error (MSE), median absolute percentage error (MAPE), and the ratio of RMSE (RSR), as previously mentioned. These metrics have formulae that can be found in the following equations. In the training, validation, and testing phases of the algorithm, a high R^2 value denotes exceptional performance. Conversely, as they indicate less scheme error, lower values of metrics like RMSE, RSR, MDAPE, and MSE are preferred. Eqs. (11) to (15) are utilized for computing these metrics.

$$R^2 = \left(\frac{\sum_{i=1}^{w} (a_i - \bar{a})(k_i - \bar{k})}{\sqrt{\left[\sum_{i=1}^{w} (a_i - a)^2\right]\left[\sum_{i=1}^{w} (k_i - \bar{k})^2\right]}} \right)^2 \tag{11}$$

$$RMSE = \sqrt{\frac{1}{w}\sum_{i=1}^{w} (k_i - a_i)^2} \tag{12}$$

$$MSE = \frac{1}{w}\sum_{i=1}^{w} k_i^2 \tag{13}$$

$$\text{MDAPE} = 100 \times \text{median}\left(\frac{|k_i - \bar{k}|}{|a_i - \bar{a}|}\right) \quad (14)$$

$$RSR = \frac{RMSE}{St.Dev.} \quad (15)$$

Here, a_i and k_i display the anticipated and empirical values, respectively. The mean values of the empirical samples and predicted are displayed by $\bar{a}$ and $\bar{k}$. Alternatively, w displays the count of samples being considered.

3. Outcomes and discussion

3.1. onvergence curve and Hyperparameter's results

Hyperparameters are those outer setups that span areas, including the strength of regularization and pace of learning, and play a crucial role in setting the schemes' behavior. Schemes' best performance revolves around the necessary hyperparameter tuning, an exercise that involves trial and error complemented by deploying enhancement techniques. Table 2 correctly records hyperparameter values for KNJF and KNBW schemes, reporting parameters including n neighbors, leaf size, and p. A detailed presentation plays an important role toward greater transparency and reproducibility by ML schemes, an important factor toward deeper examination, as well as exact replication, of schemes' setups.

Table 2. The outcomes of the KNN hyperparameters

Schemes	Hyperparameter		
	n_neighbors	leaf_size	p
KNJF	1	3	18
KNBW	0	94	7

3.2. Comparative results of predictive models

This section also compared the ability of various models to forecast MDD and presented the performance of their training, validation, and testing in Table 3. Among these models, the KNN model, frequently employed in lab modeling, proved efficient in forecasting MDD with satisfactory performance. In an attempt to optimize the KNN model's performance, algorithms included JFO and BWOA. The model was trained using 70% data, 15% validation data, and 15% test data to gauge the accuracy level of the scheme. In the training section, it seems from Table 3 that KNJF had the highest R^2 measure of 0.9895 and the lowest RMSE measure of 24.236. In comparison, the maximum RMSE by the KNN model during the period of testing was 61.067, whereas the minimum R^2 by the same model during the period of validation was 0.9447. $KNJF_{Train}$ had the best reported MSE result (587.39), while KNN_{Test} had the poorest result (3729.2). In terms of MDAPE, $KNJF_{Train}$ performed the best at 1.1314 , while KNN_{Test} performed the worst at 2.2823. Similarly, at 0.1029, the $KNJS_{Train}$ model performed best in RSR, whereas at 0.2472, the KNN_{Test} model performed worse. The optimal Testate value was derived from $KNJF_{Train}$, whereas the least favorable value was obtained from $KNN_{Test.}$ Overall, the results suggest that all models were well-trained, except for the KNN and KNBW models, whose parameters declined. Nevertheless, despite their initially elevated values, they remain dependable for making projections. Conversely, the KNJF model showed a little rise in R^2 along with drops in other metrics, indicating its improved prediction accuracy. Therefore, it can be concluded that the KNJF model is the most effective in predicting MDD, and its performance can be further improved by adjusting the scheme parameters.

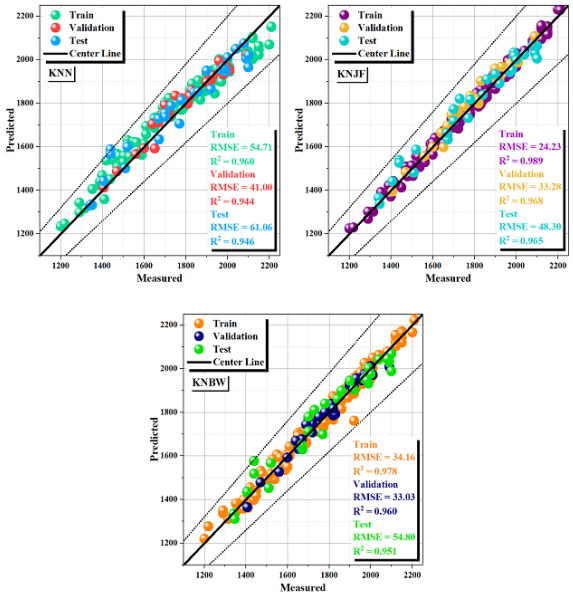

Fig. 3. The expected and measured values' scatter plot

A scatter map comparing the performance of mixed schemes according to two important parameters-RMSE and R^2-is displayed in Fig. 3. Whereas RMSE shows the degree of dispersion, R^2 displays the degree of agreement. For each of the three models, the outcomes demonstrate a substantial positive correlation between the expected and actual values, indicating a high degree of prediction accuracy. Specifically, the KNJF model, registering 0.9895 for R^2 and 24.236 for RMSE, demonstrated higher accuracy compared to the KNBW and KNN schemes. This is evident from the tight clustering of its data points around the center line, indicating a narrow range of dispersion. Con-

Table 3. Outcomes of schemes' assessment by evaluators

Schemes	Sections	Evaluators				
		R2	RMSE	MSE	MDAPE	RSR
KNN	Train	0.9600	54.718	2994.1	2.2346	0.2324
	Validation	0.9447	41.003	1681.3	1.3842	0.243
	Test	0.9461	61.067	3729.2	2.2823	0.2472
KNJF	Train	0.9895	24.236	587.39	1.1314	0.1029
	Validation	0.9681	33.288	1108.1	1.1416	0.1973
	Test	0.9659	48.307	2333.6	1.9175	0.1955
KNBW	Train	0.9789	34.167	1167.4	1.4665	0.1451
	Validation	0.9604	33.031	1091.1	1.3088	0.1957
	Test	0.9513	54.802	3003.3	2.1681	0.2218

versely, the KNBW and KNN schemes showed equivalent performance levels, with their data points scattered more broadly. Considering the correspondence of gauged and anticipated values of MDD values, concluded that this predictive scheme has the potential to be used for predicting MDD values of stabilized soils.

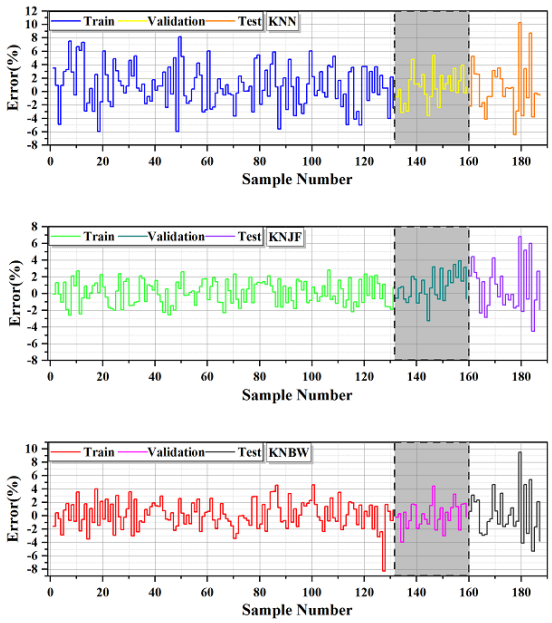

Fig. 4. The error percentage of created schemes

The error-frequency graph for the three different schemes, KNN, KNJF, and KNBW, is displayed in Fig. 4. With a maximum error rate of 7% throughout testing, the KNJF mixed scheme, a combination of KNN and JFO techniques stands out as the most promising. This result shows that the KNJF model has the ability to produce forecasts that are more accurate, which is a desired quality in any forecasting effort. As a result, the KNJF model demonstrates its ability to project MDD values with significant benefits, including time and cost reductions as compared to conventional techniques. This approach helps to boost productivity and efficiency in geotechnical engineering activities by simplifying the MDD prediction process. Its usefulness as a tool for maximizing resource allocation and accelerating crucial decisionmaking in soil stabilization projects is demonstrated by its practical advantages.

Fig. 5 shows the expected error rate percentages of MDD through a box normal plot. Schemes were checked across three phases, whereby some samples were kept reserved for validation, training, and testing. In KNJF, sample distribution was less scattered, whereby samples concentrated mostly between −3% and 3%. In KNBW, performance was poor, as displayed by increased sample dispersion. Besides, performance by KNN was higher than by the other two schemes across the three phases both in error value and dispersion.

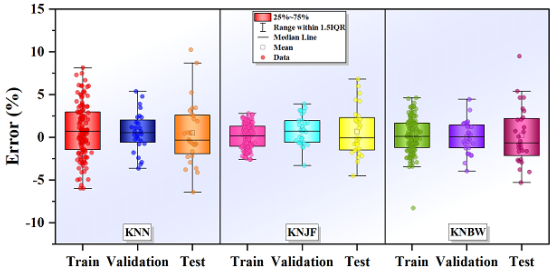

Fig. 5. The current schemes' error percentage

3.3. Sensitivity Analysis

Analyzing the frequency content of the scheme output is pivotal in understanding the complexities of sensitivity indices, including the First-Order Sensitivity Index (S1) and TotalOrder Sensitivity Index (ST). The work is

accomplished through the Fourier Amplitude Sensitivity Test (FAST), wherein sinusoidal functions are employed to study the sensitivity of changes in the inputs toward the output. These sensitivity indices find invaluable application in parameter importance analysis, model calibration, and uncertainty quantification, making them invaluable tools in model analysis and optimization.

In Fig. 6, a graphical presentation illustrates the effect of each input parameter on the anticipated values of MDD. This visualization helps reveal recognizable patterns in parameter importance, offering insights into the scheme's behavior. Among all parameters, the Plasticity Index (PI) exhibits the highest sensitivity, as indicated by its significantly greater ST and S1 values. From a geotechnical perspective, this is expected-PI directly reflects the clay content and plasticity characteristics of fine-grained soils, which strongly influence the dry density due to their effect on soil structure, water retention, and compaction behavior. Changes in PI thus have a pronounced impact on the predicted MDD. In contrast, other input parameters show minimal influence, playing a relatively minor role in affecting the scheme output. Understanding the comparative importance of the input variables through FAST analysis enables modelers and engineers to concentrate on the most influential factors, including PI, to enhance model calibration and improve predictive performance.

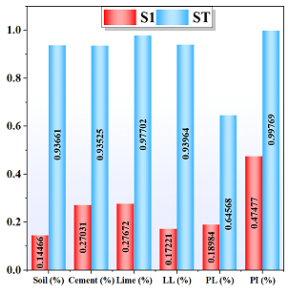

Fig. 6. Outcomes of FAST-oriented Sensitivity analysis for the input parameters' impact on MDD

4. Conclusion

One important soil property that has an effect on strength and stability includes maximum dry density (MDD), which has an effect due to various parameters, including moisture content, efforts of compaction, and type of soil. To guarantee the stability and longevity of buildings, precise measuring of MDD plays an inevitable role during construction and engineering. In this paper, an ML model applying the KNN method was presented to forecast the MDD's mechanical property. In this paper, three schemes arose due to the application by the study of the meta-heuristic algorithms Jellyfish Search Optimizer (JSO) and BWOA aimed at maximizing accuracy while reducing error. The performances of the schemes were gauged by various performance indicators, including R^2, RMSE, MSE, RSR, and MDAPE, in this study. It's noteworthy that KNJF models had the peak R^2 of 0.9895 during training, while the minimum R^2 of 0.9447 occurred during the validation period by the KNN model. The error indicators, RMSE, MSE, RSR, and MDAPE, proved KNJF schemes had fewer error indicators by containing values 24.236, 587.39, 1.1314, and 0.1029, representing RMSE, MSE, RSR, and MDAPE, respectively, showing greater performance by the schemes over schemes KNN, KNJF, and KNBW. In line with the findings of the study, the JFO optimizer partnered with the KNN in a way that achieved excellent prediction accuracy by MDD. Optimizing the KNN by applying the JFO optimizer gave an increase in prediction accuracy by an R^2 value by an increase of 2.5%. Despite these promising outcomes, some restrictions must be acknowledged. The database deployed may not fully capture the diversity of soil types and site-specific conditions encountered in practice, which may restrict the generalizability of the schemes. The sensitivity of the scheme performance to the choice of meta-heuristic optimizer introduces complexity that may affect interpretability. Furthermore, temporal variability, environmental factors, and long-term changes in soil behavior were not considered, which could impact realworld applicability. Additionally, the study did not employ extensive cross-validation or external validation across multiple databases, which may limit the robustness of the findings. Future exploration should focus on boosting scheme generalization by integrating larger and more diverse databases that reflect a broader range of geotechnical conditions. Incorporating feature selection techniques and explainable AI (XAI) methods could improve interpretability and transparency. Longitudinal data collection and site-specific calibration should also be explored to evaluate temporal consistency and improve field-level reliability. Finally, comparative studies involving other advanced ML algorithms and hybrid optimization frameworks can help refine the approach and expand its practical utility in geotechnical engineering.

5. Acknowledgment

This work is funded by the 2022 project of the 14th Five Year Plan for Education Science in Liaoning Province (No. JG22EB042).

References

[1] Z. S. Janjua and J. Chand, (2016) *"Correlation of CBR with index properties of soil"* **International Journal of Civil Engineering and Technology** *7*: 57–62.

[2] J. Duque, W. Fuentes, S. Rey, and E. Molina, (2020) *"Effect of grain size distribution on california bearing ratio (CBR) and modified proctor parameters for granular materials"* **Arabian Journal for Science and Engineering** *45*: 8231–8239. DOI: https://doi.org/10.1007/s13369-020-04673-6.

[3] M. Y. Khiabani, B. Sedaghat, P. Ghorbanzadeh, N. Porroustami, S. M. H. Shahdany, and Y. Hassani, (2023) *"Application of a hybrid hydro-economic model to allocate water over the micro-and macro-scale region for enhancing socioeconomic criteria under the water shortage period"* **Water Econ Policy**:

[4] S. Preethi, R. B. Tangadagi, M. Manjunatha, and A. Bharath, (2020) *"Sustainable effect of chemically treated aggregates on bond strength of bitumen"* **J. Green Eng** *10*: 5076–5089.

[5] E. G. Akpokodje, (1985) *"The stabilization of some arid zone soils with cement and lime"* **Quarterly journal of engineering geology** *18*: 173–180. DOI: https://doi.org/10.1144/GSL.QJEG.1985.018.02.06.

[6] F. Bell, (1996) *"Lime stabilization of clay minerals and soils"* **Engineering geology** *42*: 223–237. DOI: https://doi.org/10.1016/0013-7952(96)00028-2.

[7] A. H. Alavi, A. H. Gandomi, M. Gandomi, and S. S. S. Hosseini, (2009) *"Prediction of maximum dry density and optimum moisture content of stabilised soil using RBF neural networks"* **The IES Journal Part A: Civil Structural Engineering** *2*: 98–106. DOI: https://doi.org/10.1080/19373260802659226.

[8] A. Worku and D. Shiferaw, (2004) *"Prediction of maximum dry density of local granular fills"* **Zede Journal** *21*: 59–70.

[9] A. B. Ngowi, (1997) *"Improving the traditional earth construction: a case study of Botswana"* **Construction and Building Materials** *11*: 1–7. DOI: https://doi.org/10.1016/S0950-0618(97)00006-8.

[10] C. KS, Y. M. Chew, M. H. Osman, and M. G. SK. "Estimating maximum dry density and optimum moisture content of compacted soils". In: *international conference on advances in civil and environmental engineering*. *1*. 2015, 8.

[11] A. Bharath, M. Manjunatha, T. V. Reshma, and S. Preethi, (2021) *"Influence and correlation of maximum dry density on soaked unsoaked CBR of soil"* **Materials Today: Proceedings** *47*: 3998–4002. DOI: https://doi.org/10.1016/j.matpr.2021.04.232.

[12] S. Suman, M. Mahamaya, and S. K. Das, (2016) *"Prediction of maximum dry density and unconfined compressive strength of cement stabilised soil using artificial intelligence techniques"* **International Journal of Geosynthetics and Ground Engineering** *2*: 11. DOI: https://doi.org/10.1007/s40891-016-0051-9.

[13] A. H. Alavi, A. H. Gandomi, A. Mollahassani, A. A. Heshmati, and A. Rashed, (2010) *"Modeling of maximum dry density and optimum moisture content of stabilized soil using artificial neural networks"* **Journal of Plant Nutrition and Soil Science** *173*: 368–379. DOI: https://doi.org/10.1002/jpln.200800233.

[14] F. Masoumi, S. Najjar-Ghabel, A. Safarzadeh, and B. Sadaghat, (2020) *"Automatic calibration of the groundwater simulation model with high parameter dimensionality using sequential uncertainty fitting approach"* **Water Supply** *20*: 3487–3501. DOI: https://doi.org/10.2166/ws.2020.241.

[15] H. F. H. Ali, (2023) *"Utilizing multivariable mathematical models to predict maximum dry density and optimum moisture content from physical soil properties"* **Multiscale and Multidisciplinary Modeling, Experiments and Design** *6*: 603–627. DOI: https://doi.org/10.1007/s41939-023-00165-w.

[16] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, (2016) *"Machine learning basics"* **Deep learning**: 98–164.

[17] B. Mahesh, (2020) *"Machine Learning Algorithms—A Review"* **International Journal of Science and Research** *9*: 381–386. DOI: http://dx.doi.org/10.21275/ART20203995.

[18] Z.-H. Zhou. *Machine learning*. Springer nature, 2021. DOI: https://doi.org/10.1007/978-981-15-1967-3.

[19] G. Biau, (2012) *"Analysis of a random forests model"* **The Journal of Machine Learning Research** *13*: 1063–1095.

[20] K.-j. Kim, (2003) *"Financial time series forecasting using support vector machines"* **Neurocomputing** *55*: 307–319. DOI: https://doi.org/10.1016/S0925-2312(03)00372-2.

[21] M. I. Jordan and T. M. Mitchell, (2015) *"Machine learning: Trends, perspectives, and prospects"* **Science *349***: 255–260. DOI: https://doi.org/10.1126/science.aaa8415.

[22] B. Naeim, A. J. Khiavi, P. Dolatimehr, and B. Sadaghat, (2024) *"Novel optimized support vector regression networks for estimating fresh and hardened characteristics of SCC"* **Advances in Engineering and Intelligence Systems *3***: 110–123. DOI: https://doi.org/10.22034/aeis.2024.483317.1239.

[23] S. K. Das, P. Samui, and A. K. Sabat, (2011) *"Application of artificial intelligence to maximum dry density and unconfined compressive strength of cement stabilized soil"* **Geotechnical and Geological Engineering *29***: 329–342. DOI: https://doi.org/10.1007/s10706-010-9379-4.

[24] M. Saadat and M. Bayat, (2022) *"Prediction of the unconfined compressive strength of stabilised soil by Adaptive Neuro Fuzzy Inference System (ANFIS) and Non-Linear Regression (NLR)"* **Geomechanics and Geoengineering *17***: 80–91. DOI: https://doi.org/10.1080/17486025.2019.1699668.

[25] E. Kalkan, S. Akbulut, A. Tortum, and S. Celik, (2009) *"Prediction of the unconfined compressive strength of compacted granular soils by using inference systems"* **Environmental geology *58***: 1429–1440. DOI: https://doi.org/10.1007/s00254-008-1645-x.

[26] 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***: 95–106. DOI: https://doi.org/10.22034/aeis.2023.414188.1129.

[27] H. F. H. Ali, B. Omer, A. S. Mohammed, and R. H. Faraj, (2024) *"Predicting the maximum dry density and optimum moisture content from soil index properties using efficient soft computing techniques"* **Neural Computing and Applications *36***: 11339–11369. DOI: https://doi.org/10.1007/s00521-024-09734-7.

[28] M. N. Duc, A. H. Sy, T. N. Ngoc, and T. L. H. Thi. "An artificial intelligence approach based on multi-layer perceptron neural network and random forest for predicting maximum dry density and optimum moisture content of soil material in quang ninh province, vietnam". In: *CIGOS 2021, Emerging technologies and applications for green infrastructure: proceedings of the 6th international conference on geotechnics, civil engineering and structures*. Springer, 2021, 1745–1754. DOI: https://doi.org/10.1007/978-981-16-7160-9_176.

[29] M. Jia and Y. Zheng. "A prediction model of maximum dry density of over coarse-grained soil". In: *IOP Conference Series: Earth and Environmental Science*. ***1330***. IOP Publishing, 2024, 012050. DOI: https://doi.org/10.1088/1755-1315/1330/1/012050.

[30] A. H. Alavi, A. H. Gandomi, and A. Mollahasani. "A genetic programming-based approach for the performance characteristics assessment of stabilized soil". In: Springer, 2012, 343–376. DOI: https://doi.org/10.1007/978-3-642-23424-8_11.

[31] W. Z. Taffese and K. A. Abegaz, (2022) *"Prediction of compaction and strength properties of amended soil using machine learning"* **Buildings *12***: 613. DOI: https://doi.org/10.3390/buildings12050613.

[32] J.-S. Chou and D.-N. Truong, (2021) *"A novel metaheuristic optimizer inspired by behavior of jellyfish in ocean"* **Applied Mathematics and Computation *389***: 125535. DOI: https://doi.org/10.1016/j.amc.2020.125535.

[33] A. M. Shaheen, R. A. El-Sehiemy, M. M. Alharthi, S. S. M. Ghoneim, and A. R. Ginidi, (2021) *"Multi-objective jellyfish search optimizer for efficient power system operation based on multi-dimensional OPF framework"* **Energy *237***: 121478. DOI: https://doi.org/10.1016/j.energy.2021.121478.

[34] M. Farhat, S. Kamel, A. M. Atallah, and B. Khan, (2021) *"Optimal power flow solution based on jellyfish search optimization considering uncertainty of renewable energy sources"* **IEEE Access *9***: 100911–100933. DOI: https://doi.org/10.1109/ACCESS.2021.3097006.

[35] V. Hayyolalam and A. A. P. Kazem, (2020) *"Black widow optimization algorithm: a novel meta-heuristic approach for solving engineering optimization problems"* **Engineering Applications of Artificial Intelligence *87***: 103249. DOI: https://doi.org/10.1016/j.engappai.2019.103249.

[36] E. H. Houssein, B. E.-d. Helmy, D. Oliva, A. A. Elngar, and H. Shaban, (2021) *"A novel black widow optimization algorithm for multilevel thresholding image segmentation"* **Expert Systems with Applications *167***: 114159. DOI: https://doi.org/10.1016/j.eswa.2020.114159.

[37] S. Memar, A. Mahdavi-Meymand, and W. Sulisz, (2021) *"Prediction of seasonal maximum wave height for unevenly spaced time series by Black Widow Optimization algorithm"* **Marine Structures *78***: 103005. DOI: https://doi.org/10.1016/j.marstruc.2021.103005.