

Article

Efficient Machine Learning Models for the Uplift Behavior of Helical Anchors in Dense Sand for Wind Energy Harvesting

Le Wang ^{1,2}, Mengting Wu ³, Hongzhen Chen ¹ , Dongxue Hao ⁴, Yinghui Tian ⁵ and Chongchong Qi ^{3,*} 

¹ State Key Laboratory of Hydraulic Engineering Simulation and Safety School of Civil Engineering, Tianjin University, Tianjin 300350, China

² State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116081, China

³ School of Resources and Safety Engineering, Central South University, Changsha 410083, China

⁴ School of Civil Engineering and Architecture, Northeast Electric Power University, Jilin 132012, China

⁵ Department of Infrastructure Engineering, The University of Melbourne, Melbourne 3000, Australia

* Correspondence: chongchong.qi@csu.edu.cn

Abstract: Helical anchors are widely used in engineering to resist tension, especially during offshore wind energy harvesting, and their uplift behavior in sand is influenced by many factors. Experimental studies are often used to investigate these anchors; however, scale effects are inevitable in $1 \times g$ model tests, soil conditions for in situ tests are challenging to control, and centrifuge tests are expensive and rare. To make full use of the limited valid data and to gain more knowledge about the uplift behaviors of helical anchors in sand, a prediction model integrating gradient-boosting decision trees (GBDT) and particle swarm optimization (PSO) was proposed in this study. Data obtained from a series of centrifuge tests formed the dataset of the prediction model. The relative density of soil, embedment ratio, helix spacing ratio, and the number of helices were used as input parameters, while the anchor mobilization distance and the ultimate monotonic uplift resistance were set as output parameters. A GBDT algorithm was used to construct the model, and a PSO algorithm was used for hyperparameter tuning. The results show that the optimal GBDT model accurately predicted the anchor mobilization distance and the ultimate monotonic uplift resistance of helical anchors in dense fine silica sand. By analyzing the relative importance of influencing variables, the embedment ratio was found to be the most significant variable in the model, while the relative density of the fine silica sand soil, the helix spacing ratio, and the number of helices had relatively minor influence. In particular, the helix spacing ratio was found to have no influence on the capacity of adjacent helices when $S/D > 6$.

Keywords: helical anchor; sand; artificial intelligence techniques; gradient-boosting decision trees; particle swarm optimization



Citation: Wang, L.; Wu, M.; Chen, H.; Hao, D.; Tian, Y.; Qi, C. Efficient Machine Learning Models for the Uplift Behavior of Helical Anchors in Dense Sand for Wind Energy Harvesting. *Appl. Sci.* **2022**, *12*, 10397. <https://doi.org/10.3390/app122010397>

Academic Editors: Yosoon Choi and Jieun Baek

Received: 13 September 2022

Accepted: 8 October 2022

Published: 15 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Helical anchors have a century-old application history in engineering practice [1,2] and have the potential to be used to support offshore floating structures [3] due to their advantages of rapid installation and high uplift capacity. A typical helical anchor comprises a central steel shaft and several helical steel plates welded onto the shaft, as shown in Figure 1. In designing floating structures, one needs to accurately predict the monotonic tensile capacity of the helical anchor, which is affected by the relative density of the soil, the embedment ratio of the helices, the soil friction angle, the dilation angle, the number of helices, the installation effect, and many other factors; the monotonic tensile capacity of the helical anchor is difficult to predict.

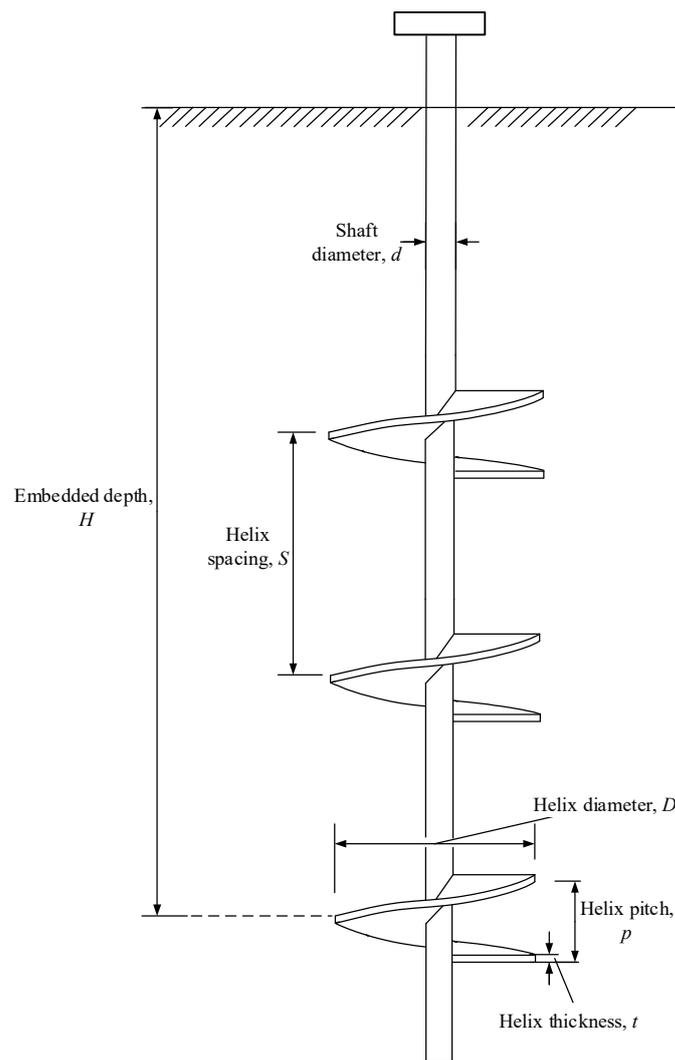


Figure 1. Helical anchor definitions.

The uplift capacity of a helical anchor in sand is generally predicted through experiments, including field tests [4,5], small-scale model tests [6–13], and centrifuge tests [10,14–18]. Experimental studies are relatively costly and time-consuming, especially field tests and centrifuge tests; additionally, it is challenging to obtain reliable data due to the complex operating environment. For small-scale studies, the experimental parameters are relatively more controllable; however, the size effect hinders the prediction accuracy. Thus, making full use of existing and reliable experimental results is of great significance for accurately predicting the uplift capacity of helical anchors in sand.

Given the advanced capabilities of AI technologies, they have been applied to a variety of related problems in civil engineering, such as predicting the bearing capacity of monopiles [19,20] and pile settlements [21], and determining the frictional resistance of driven piles [22]. Specifically, Alzabeebee et al. used multiobjective evolutionary polynomial regression to identify strong correlations between factors; their study highlights the influence of undrained cohesion and effective stress on adhesion and provides a guarantee for optimized bored pile designs under undrained conditions [23]. Furthermore, a novel model that predicts the friction capacity of driven piles was proposed based on a multiobjective genetic algorithm for evolutionary polynomial regression [24]. Goh et al. constructed a Bayesian neural network model to estimate the adhesion factor of undrained side resistance [25]. In addition, Zhang and Goh compared backpropagation neural networks and multiple adaptive regression splines to explore pile drivability [26]. Moayed

et al. established a variety of nonlinear models to estimate the ultimate bearing capacity of shallow footing under double-layer soil conditions [27]. Furthermore, Mosallanezhad and Moayedi [28] proposed an AI-technology-based model to predict the uplift capacity of helical piles in sand using 41 $1 \times g$ model tests. However, the above studies have at least three shortcomings: (1) the application of AI in geotechnical engineering problems is mostly based on artificial neural networks (ANN), but the “black box” nature of these models leads to poor interpretability [29]; (2) the feasibility of using other advanced models, such as tree-based ensemble algorithms, to predict the monotone tensile capacity of helical anchors in sand is still unknown; and (3) the scale effect, which significantly influences the strength of sand [30], was not considered in previous research [28].

To address these issues, in this study, the results of centrifugation tests were used to prepare the dataset, and a more robust AI technique, the gradient-boosting decision tree (GBDT) approach, was used to predict the monotonic tensile capacity of helical anchors in sand. GBDT is an ensemble learning technique that can improve the predictive accuracy by combining many decision tree models [31]. This approach is robust in non-relationship modeling, even when the dataset contains outliers, and the relative importance of the influencing variables can be investigated. Having been used for a wide range of applications on various datasets, the GBDT method has been shown to achieve relatively good performance compared with other AI techniques such as ANN [32,33]. Since the predictive performance of GBDT on a specific dataset is influenced by its hyperparameters, the particle swarm optimization (PSO) approach was employed for hyperparameter optimization in this work.

In this study, a novel method was proposed to predict the monotonic tensile capacity of helical anchors in sand. The GBDT method was used for nonlinear relationship modeling, while PSO was employed for GBDT hyperparameter tuning. The predictive performance was evaluated, and the relative importance of influencing variables was investigated. This study leads the way for the application of GBDT and PSO in capacity estimation of helical anchor, which will promote the utilization of the helical anchor in various geotechnical problems, such as in offshore wind power generation [34].

2. Study Background

2.1. Gradient-Boosting Decision Tree

The GBDT approach is composed of a classification and regression tree (CART) and a boosting algorithm. The details of these elements are briefly described below.

2.1.1. Boosting

Boosting is a serial-generated serialization algorithm in which each training procedure is the correction of the previous one. Boosting pays attention to the samples with errors in judgment during the previous training, and gives weight to the samples with errors [35]. There are two ways to update the weight of samples: one is to reweigh the samples, while the other is to resample the original samples according to their weights. The core concept of boosting is to build weak learners one by one and transform them into strong learners through multiple iteration accumulation [36]. In this paper, a gradient-boosting approach was used in which the negative gradient value of the current model on all samples was calculated through continuous iteration. This negative gradient value was then used as the residual approximation to construct a new weak evaluator for fitting [37]. The final output of the gradient-boosting algorithm was accumulated from multiple weak learners [38].

2.1.2. Classification and Regression Tree

CART is a basic classification and regression approach that consists of (decision) root nodes, leaf nodes, and directed edges, as shown in Figure 2 [39,40]. The root nodes denote features or attributes, while categories are shown by the leaf nodes [41]. In this paper, the CART for regression problems was used, whose core rationale is to identify the best partitioning point and confirm the output results of the leaf nodes. After the spatial region of the training set is effectively divided according to a particular method, the

regression tree is then constructed by determining the values of each subregion. The specific process is as follows:

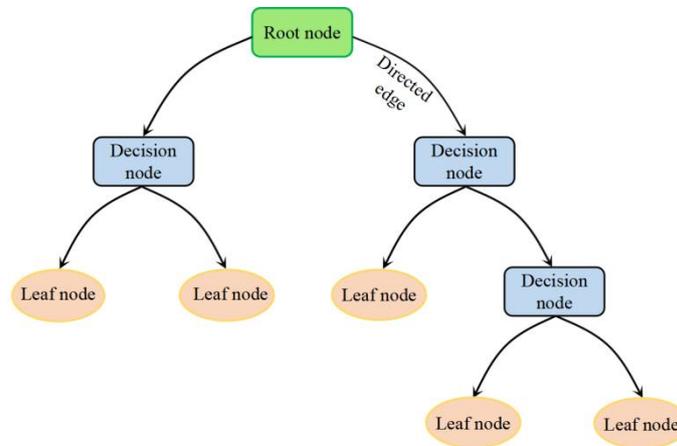


Figure 2. A typical CART architecture.

(1) Determine the optimal value of the variable, j , and the segmentation point, s , to minimize Equation (1).

$$\min_{j,s} [\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2] \tag{1}$$

where x_i and y_i represent the input and output variable, respectively; c_1 and c_2 are the mean values of the interval; R_1 and R_2 are some subregions of the input space.

(2) As shown in Equations (2) and (3), the selected pair (j, s) is used to divide the region and output the corresponding value:

$$R_1(j, s) = \{x | x^{(j)} \leq s\}, R_2(j, s) = \{x | x^{(j)} > s\} \tag{2}$$

$$\hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m(j,s)} y_i, x \in R_m, m = 1, 2 \tag{3}$$

where $x^{(j)}$ is the j -th variable, and s is the value it takes. The optimal value, \hat{c}_m , is the mean of the corresponding output, y_i , of all input instances, x_i , on region R_m .

Repeat steps (1) and (2) until the termination condition is met. In addition, if the input space is divided into M regions, such as $R_1, R_2 \dots R_m$, then a complete regression tree can be constructed according to Equation (4), as shown in Figure 2.

$$f(x) = \sum_{m=1}^M \hat{c}_m I(x \in R_m) \tag{4}$$

where I stands for the impurity of the decision tree.

2.2. Particle Swarm Optimization

PSO is a population-based swarm evolutionary computing technology used to find the optimal solution of multidimensional problems [42]. Based on adaptability to the environment, this algorithm can guide individuals in the group to optimal areas in the search space [43]. The PSO algorithm starts by initializing the particle group (including random position and velocity). The fitness value of each particle is then calculated according to the fitness function. Finally, the PSO algorithm compares the fitness of each particle with the corresponding value of the single historical optimal position (pbest) and the global optimal position (gbest) [44]. If the current particle has a higher fitness value, it will be

used to update and replace the previous position [44]. At the same time, the velocity and position of each particle will be updated according to Equations (5) and (6). The above steps are then repeated until either the algorithm reaches a maximum number of iterations or the increment of the optimal fitness value is below a given threshold.

$$v_{id}^k = wv_{id}^{k-1} + c_1r_1(pbest_{id} - x_{id}^{k-1}) + c_2r_2(pbest_{id} - x_{id}^{k-1}) \tag{5}$$

$$x_{id}^k = x_{id}^{k-1} + v_{id}^{k-1} \tag{6}$$

where v_{id}^k and v_{id}^{k-1} represent the flight velocity of the $k-1$ and k -generation particles, respectively; x_{id}^k and x_{id}^{k-1} denote the d -dimensional components of the position vectors of the $k-1$ and k -generation iteration particles, respectively [45]; c_1 and c_2 denote the acceleration constant, which is used to adjust the maximum stride length of learning; r_1 and r_2 are two random functions in the range of $[0, 1]$. w is the non-negative inertia weight for adjusting the search range of solution space. The overall flowchart of the PSO algorithm is shown in Figure 3.

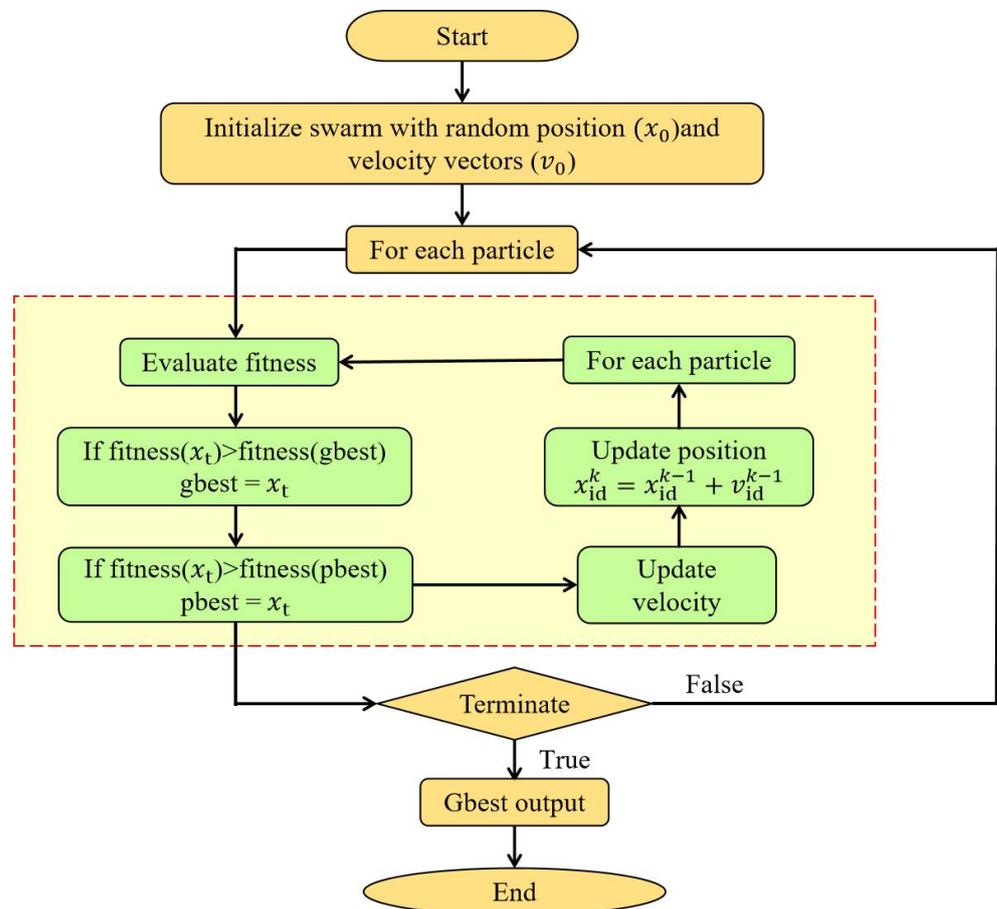


Figure 3. Overall flowchart of the PSO algorithm.

3. Methods

In this section, the PSO and GBDT algorithms are combined to learn from the datasets and make predictions. Figure 4 illustrates the GBDT-PSO methodology, which consists of three parts: the establishment of dataset, the tuning of hyperparameters, and the evaluation and interpretation of the model.

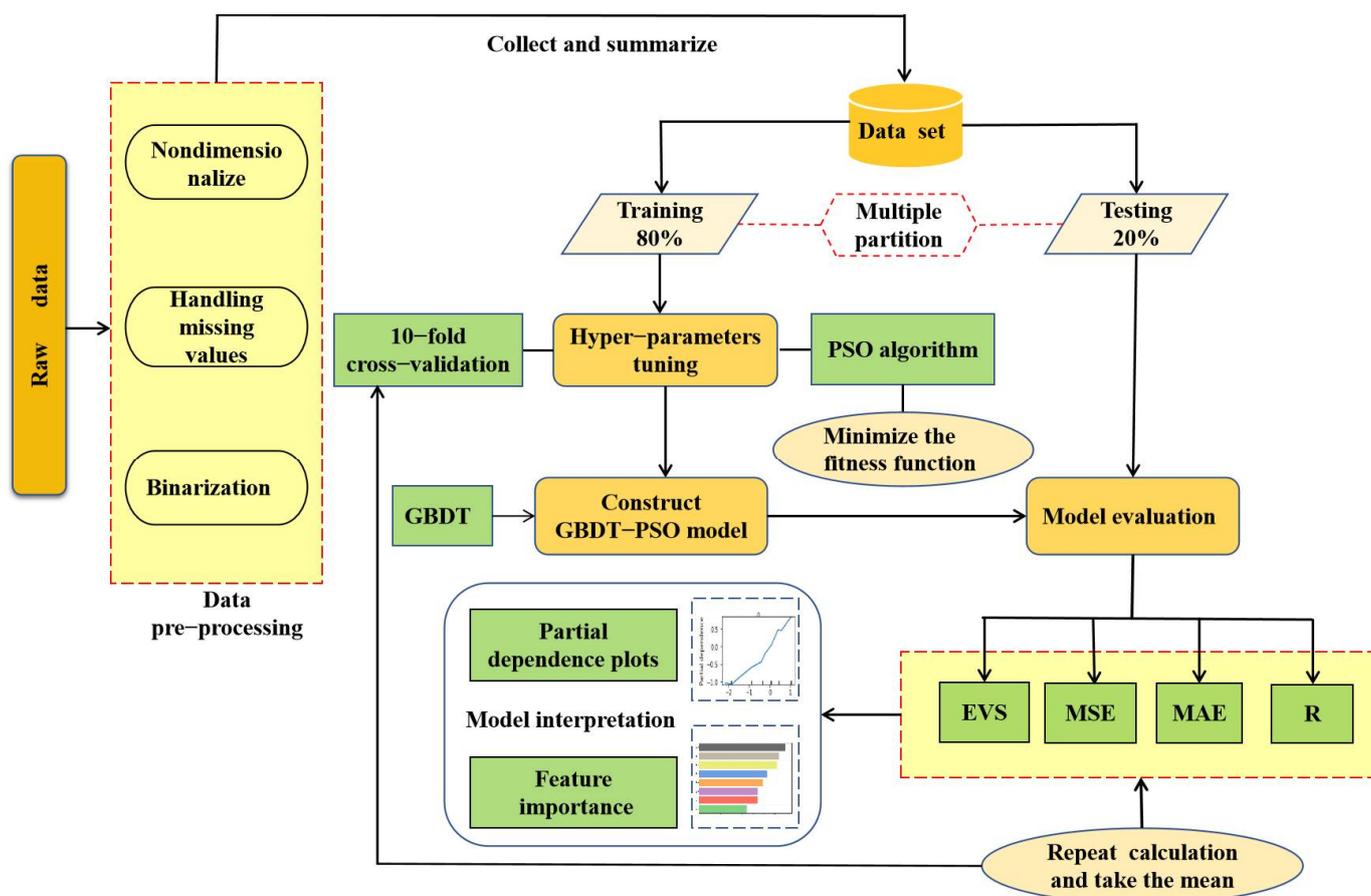


Figure 4. General process of GBDT-PSO model construction and evaluation.

3.1. Established Dataset

To eliminate the influence of the scale effect, the data used in this study were obtained from centrifuge experiments [23] conducted at the University of Western Australia (UWA) with an acceleration of $20 \times g$. In these centrifuge experiments, fine, dry silica sand [46,47] was used and prepared by air pluviation to control the relative density of soil in the range of $\sim 85\text{--}96\%$ at single gravity. The model anchors used in the centrifuge experiments, as shown in Figure 5a, consisted of a shaft, which can be divided into several sections with different lengths, plates, and anchor caps; all the anchors were manufactured from aluminum. Model anchors were placed with a “wished-in-place” approach at relevant embedment depths to avoid the installation effect. This was achieved by pausing pluviation when the soil reached the targeted height for the lowermost helix or plate. At this point, the plate or helix with the first shaft extension segment was located carefully on the sample surface, and pluviation commenced until the sample height reached the targeted location for the next helix. The information about the dimension of helical anchors and the UWA sand used in these centrifuge experiments are shown in Table 1. The details about the setup are shown in Figure 5b. During the experimental process, the model anchors were pulled by applying a force on the anchor cap through a hook; the force was then captured by a load cell attached to the hook. The relevant parameters and results of 33 experiments are listed in Table 2. In the predictive models, the relative density of soil, D_r , embedment ratio, H/D , helix spacing ratio, S/D , and the number of helices, n , were treated as input parameters; the anchor mobilization distance, u_p , and the ultimate monotonic uplift resistance, Q_u , were treated as output parameters. Table 3 contains a summary of the statistical description of the model’s inputs and outputs. The authors note that variations in D_r in the current study were relatively small as we are focusing primarily on dense sand. In addition, the difficulty

involved in data collection for loose and medium–dense sand makes the above selection reasonable.

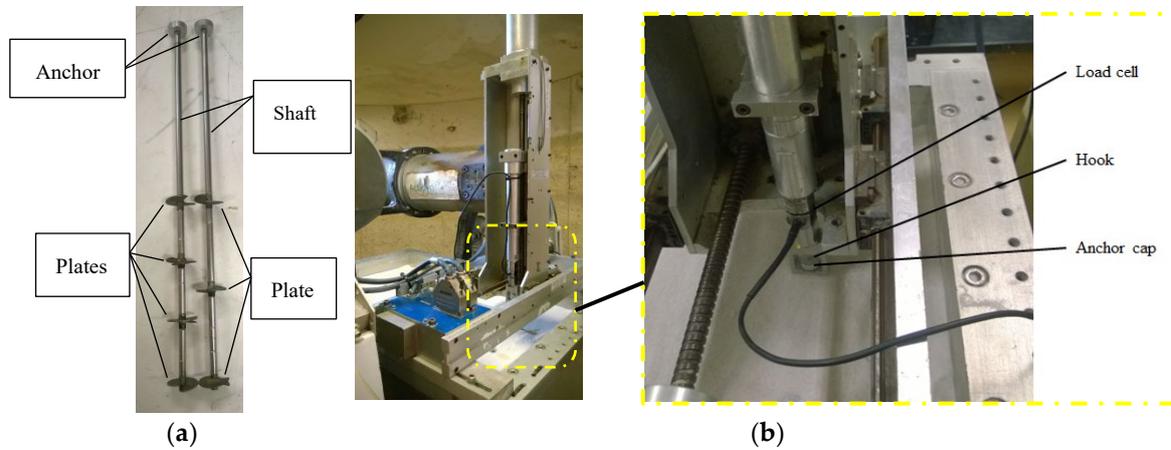


Figure 5. The test setup: (a) for the details about model anchors; (b) for the loading equipment.

Table 1. The information about the helical anchor and soil [18].

Helical Anchor		Soil	
Dimensions	Value	Properties	Value
Helix diameter, D (mm)	20	Specific gravity, G_s	2.65
Helix pitch, p (mm)	5	Median grain size, d_{50} (mm)	0.25
Helix thickness, t (mm)	2	Coefficient of uniformity, C_u	1.87
Shaft diameter, d (mm)	4.7	Coefficient of curvature, C_c	0.938
Number of helix, n	0, 1, 2, 3, 4	Maximum void ratio, e_{max}	0.703
Helix spacing, S	1.5, 2, 3, 4.5, 6 D	Minimum void ratio, e_{min}	0.516
		Critical state friction angle,	31

Table 2. Experimental program and key results used in this paper [18].

Experiment Number	N	S/D	H/D	D_r (%)	U_F / D	Q_u (kN)
T1	1	0	3	85.8	0.050	22.9
T2	1	0	6	85.8	0.140	108.7
T3	1	0	9	85.8	0.204	236.2
T4	1	0		85.8	0.242	357.5
T5	1	0	12	85.4	0.238	313.4
T6	1	0	2	86.7	0.032	9.9
T7	1	0		86.4	0.055	22.1
T8	1	0	3	96.2	0.067	22.9
T9	1	0	4	86.7	0.091	42.9
T10	1	0		86.4	0.128	108.1
T11	1	0	6	96.2	0.146	121.7
T12	1	0	7.5	90.0	0.180	161.6
T13	1	0		86.4	0.170	176.4
T14	1	0	8	96.4	0.188	217.6
T15	1	0		88.8	0.201	249.9
T16	1	0	9	96.1	0.192	270.3
T17	1	0		96.2	0.166	260.0
T18	1	0	10	96.4	0.190	309.6
T19	1	0	10.5	90.0	0.201	271.8
T20	1	0		85.4	0.227	322.1
T21	1	0	12	91.7	0.209	364.9
T22	2	1.5	7.5	88.7	0.153	158.8

Table 2. *Cont.*

Experiment Number	N	S/D	H/D	D_r (%)	u_p/D	Q_u (kN)
T23	2	1.5	12	86.6	0.224	383.1
T24	2	3	9	89.3	0.198	240.7
T25	2	3	12	86.7	0.201	412.0
T26	2	4.5	10.5	89.3	0.198	320.9
T27	2	4.5	12	86.6	0.220	370.7
T28	2	6	9	96.2	0.190	264.7
T29	2	6	12	86.7	0.214	384.2
T30	3	1.5	9	89.3	0.168	222.9
T31	3	3	12	88.8	0.211	459.5
T32	3	1.5	12	96.1	0.209	512.6
T33	4	2	12	90.0	0.254	489.4

Table 3. Statistical description of inputs and outputs.

Variables	Type	Standard Deviation	Maximum	Minimum	Mean	Kurtosis	Skewness
N	Input	0.783	4.000	1.000	1.515	1.844	1.464
S/D	Input	1.815	6.000	0.000	1.152	1.363	1.473
H/D	Input	3.107	12.000	2.000	8.788	−0.490	−0.705
D_r	Input	3.992	96.400	85.400	89.724	−0.959	0.737
u_p/D	Output	0.057	0.254	0.032	0.174	0.652	−1.129
Q_u	Output	137.747	512.600	9.900	248.182	−0.720	−0.098

After the normalization and binarization of the dataset, we divided the whole dataset into two parts: a training set comprising 80% of the data and a testing set comprising the remaining 20%. The data in the training set were used for training the GBDT–PSO model and hyperparameters tuning, while the model performance was evaluated using the validation set [48]. The authors note here that, although the dataset size seems relatively small, it is still by far the largest available dataset regarding helical anchors with multiple helices. Laboratory experiments involving helical anchors with multiple helices are cumbersome; thus, collecting validated data is extremely challenging. Considering the number of input features, which was four in the current study, the dataset size was feasible (as shown in the modelling results section).

3.2. Hyperparameters Tuning

Before constructing the GBDT–PSO model, it is necessary to first preadjust and determine several important hyperparameters, because the prediction performance of the model will vary greatly with different hyperparameters combinations. However, using traditional methods such as a learning curve and grid search to adjust these parameters individually is tedious and time-consuming. Accordingly, this paper uses the PSO algorithm to take the R value in the training set as the fitness function and maximize it during the evolution of PSO. Thus, the hyperparameters were tuned and the optimal parameters could be determined. Table 4 summarizes the selected hyperparameters to be tuned and their tuning ranges.

Table 4. Hyperparameters description and their tuning ranges.

Hyperparameters	Explanation	Type	Tuning Range
Max_depth	The maximum depth of the CART	Integer	3–15
Min_samples_split	The minimum number of samples required to split an internal node	Integer	2–15
Min_samples_leaf	The minimum number of samples at the leaf node	Integer	1–15
Max_RT	The maximum number of CART models in GBDT	Integer	50–2000
Learning rate	The learning rate shrinks the contribution of each CART model	Float	0.01–1
Max_features	The number of features to consider during tree splitting	Float	0.4–1

In the process of hyperparameter tuning, K -fold cross-validation was used as an indicator to evaluate model performance. The initial data were evenly divided into K folds, and each fold was successively used as the validation set while the rest of the $K-1$ folds were used for training to obtain K models [49]. Finally, the model performance was tested using the validation set. However, due to the randomness of the partitioning of the dataset, the model evaluation obtained by single K -fold cross-validation is not convincing. Therefore, the dataset was divided repeatedly 30 times in this study, and the mean value of the K -fold cross-validation was used as the final evaluation result.

3.3. Evaluation and Interpretation of the Model

In this paper we were selected the following elements to evaluate the model's performance: the explained variance score (EVS); the mean squared error (MSE); the mean absolute error (MAE); the correlation coefficient (R); the ratio of predictions with an error within the range of $\pm 20\%$ of the total predictions (a20-index); and the percentage of prediction within an error range of $\pm 30\%$ (P30) [49–53]. The EVS value range is $[0, 1]$; the closer this value is to 1, the more the independent variable can explain the observed variance change of the dependent variable. The MSE is used to calculate the mean value of the sum of error squares of sample points corresponding to the original data and the fitting data. The MAE is a measure of how close the predicted value is to the true value [54], and R is used to describe the degree of linear correlation between the predicted y value and the real y value. The variation ranges of a20-index and P30 are $[0, 1]$; the closer the values are to 1, the more predicted values fall within the acceptable error range, and the better the model performance. The calculation of the six evaluation indexes is shown in Equations (7)–(12):

$$\text{EVS} = 1 - \frac{\text{Var}\{y_i - \hat{y}_i\}}{\text{Var}\{y_i\}} \quad (7)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

$$\text{R} = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (10)$$

$$\text{a20-index} = \frac{\text{Per20}}{n} \quad (11)$$

$$\text{P30} = \frac{N30}{n} \quad (12)$$

where Var is the variance; \hat{y}_i is the predicted value; y_i is the desired target response; \bar{y} is the mean of the target response; $\bar{\hat{y}}$ is the mean of the target response; n is the number of samples; n is the number of test samples used for model evaluation; *per20* is the number of predicted values within $\pm 20\%$ error; and *N30* is sample points with $\pm 30\%$ deviation from the experimental value.

In addition, the importance score and partial dependence plots were used to further evaluate the correlation between outputs and inputs. The importance score is a single indicator that describes the importance degree of inputs to outputs while the partial dependence plots indicate the dependence, such as variation tendency, between the output and the variation of the inputs. A detailed explanation of the importance score and partial dependence plots can be found in our previous studies [52].

4. Results, Discussion, and Concluding Remarks

4.1. Results of the Hyperparameters Tuning

Taking R as the fitness function and using the data of the training set to tune the hyperparameters of the model, the tuning results are shown in Figure 6. As expected, the accuracy of the model was significantly improved. Specifically, for u_p , R increased from 0.75 to 0.82, while Q_u increased from 0.89 to 0.93. The determined optimization hyperparameters are summarized in Table 5.

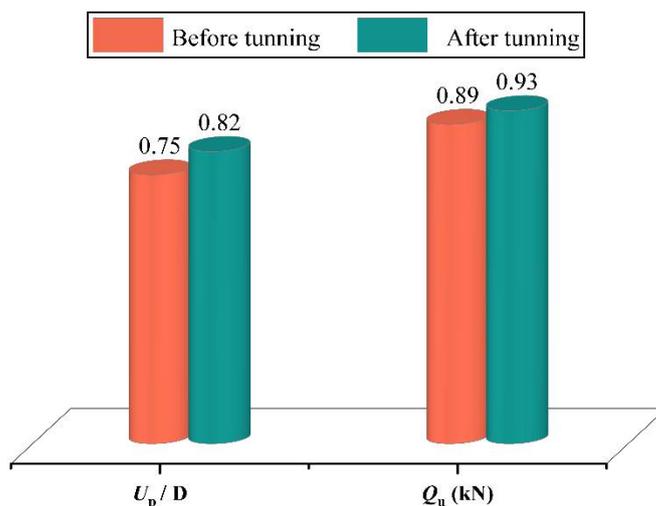


Figure 6. Comparison of model performance before and after hyperparameter tuning.

Table 5. Optimized hyperparameters for both datasets.

Hyperparameters	Q_u Dataset	u_p Dataset
Max_depth	4	15
Min_samples_split	7	7
Min_samples_leaf	1	1
Max_RT	1319	873
Learning rate	0.061	0.890
Max_features	1	1

4.2. Results of the Optimum GBDT Model

In the existing literature, machine learning predictions for Q_u and u_p have not been attempted. Thus, the prediction accuracy achieved in this study cannot be compared with previous results. However, GBDT modeling has been widely verified to display better performance than commonly used machine learning techniques in many problems [36,37,55–57]. Furthermore, the performance of GBDT was further increased using PSO in this study, improving the representativeness of the current work.

To verify the efficiency of the proposed optimization model, we applied it to the testing set for evaluation (Figure 7). For u_p , the R value achieved on the testing set was above 0.8, while the R on the testing set for Q_u is as high as 0.92. Furthermore, the MAE and MSE values of the two datasets were small, implying the high generalization ability of the model. However, there were clear differences in the MSE and MAE of Q_u on the training and testing sets, which will lead to weak predictive ability for Q_u . In addition, the EVS of u_p was between 0.6 and 0.7, and the predicted and sample values were more dispersed than those of Q_u (0.85).

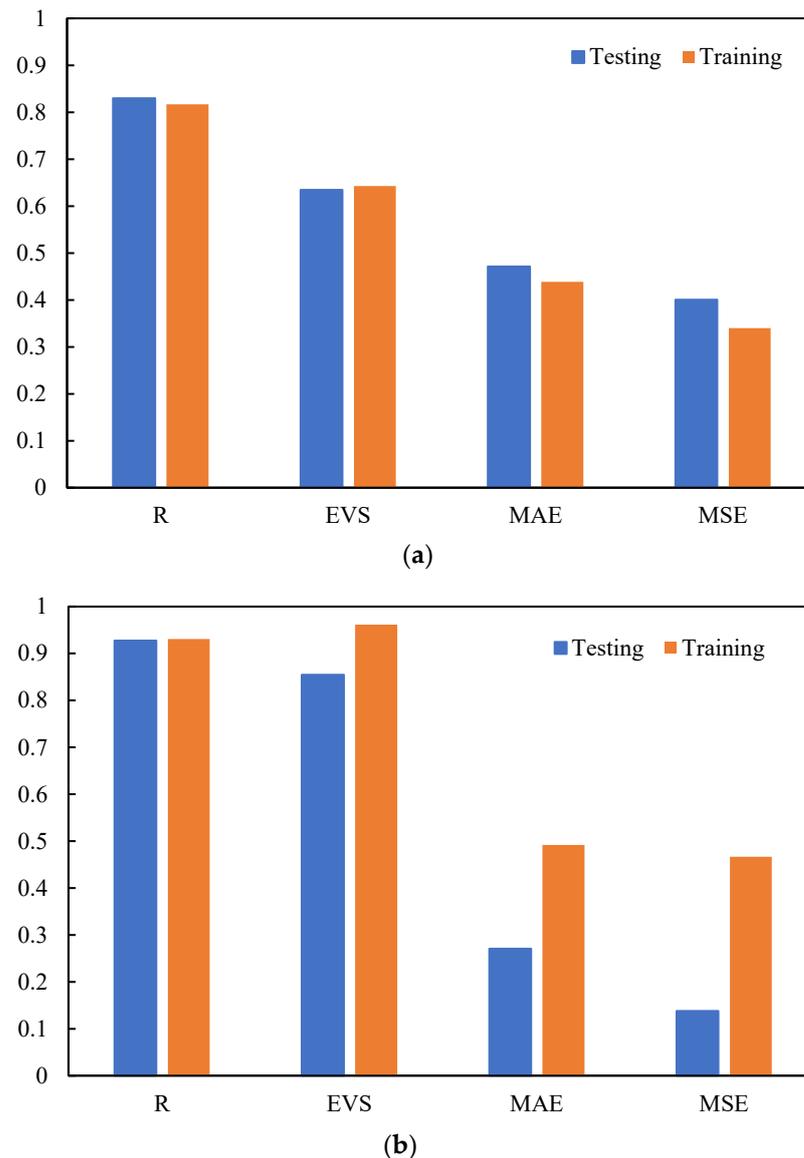
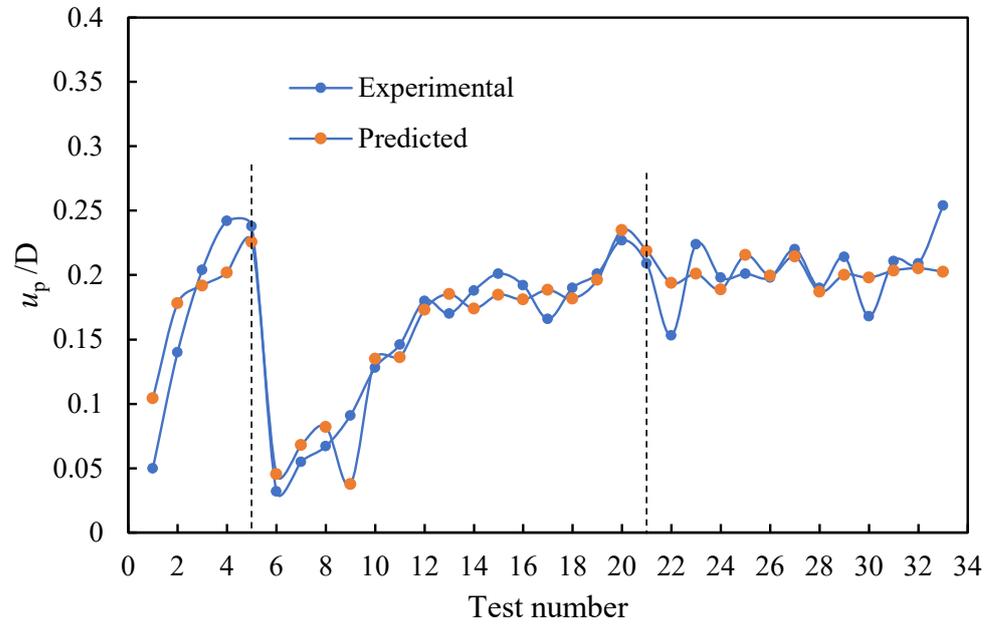


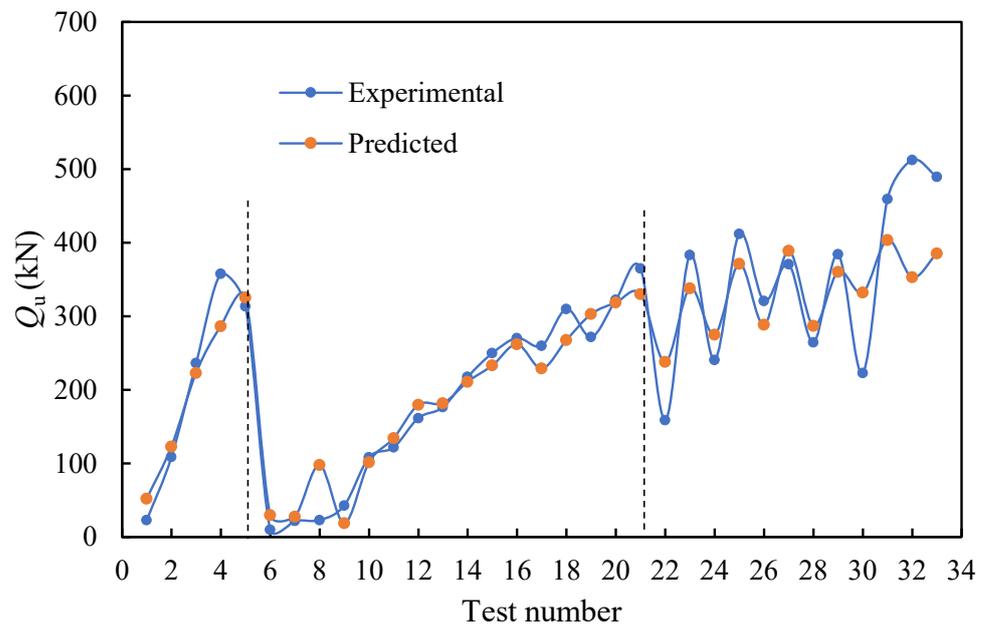
Figure 7. Evaluation of optimized GBDT models on testing and training sets: (a) for u_p , (b) for Q_u .

The experimental and predicted Q_u and u_p values were further compared, as shown in Figure 8. The variation trends and the numerical value of Q_u and u_p were predicted well except for some special cases, such as T33 for u_p and T32 for Q_u . The R value was 0.941 and 0.913 on the whole datasets of Q_u and u_p , respectively. Furthermore, Figure 8 shows that there are three different parts to the results: (1) T1–T5 consist of five tests concerning single-plate anchors embedded under different depths with similar D_r —in this part, u_p and Q_u increase with the embedment depth; (2) T6–T21 consist of ten different embedment depths of single-helix anchors in sand with similar D_r —in this part, u_p and Q_u also increase with the embedment depth and the relevant size is close to that for single-plate anchors in part (1); (3) T22–T33 comprise multiple-helix anchors with different H/D , S/D , and n , with double-helix anchors with different H/D and S/D to those analyzed in T22–T29. Adjacent helices were found to have no influence on each other for Q_u and u_p when $S/D > 6$ in dense fine silica sand; this outcome differs from the threshold value result of $S/D > 9$ in Hao et al. [23]. Combined with the finding by Ilamparuthi et al. [16] (i.e., the relevant influence of circular-plate anchors in loose sand is smaller than that in dense sand), the findings of this study have potentially important implications for optimization, Q_u , for multiple-helix anchors if they can be verified by other experimental studies in the future.

With the increasing number of helices, the differences between the predicted and actual values fluctuate, which may be caused by the relatively small available data volume for multiple-helix anchors.



(a)



(b)

Figure 8. Performance of the optimal GBDT model on the experimental and predicted value: (a) evaluation of experimental and predicted u_p ; (b) evaluation of experimental and predicted Q_u .

To further verify the model results, the cumulative frequency of the predicted values within a specific error range was repeatedly calculated (30 times) using the a20-index and P30, with the results shown in Figure 9. Specifically, for u_p , the mean values of the a20-index and P30 were 0.724 and 0.819, respectively; the equivalent values for Q_u were 0.70 and 0.824. All these results are close to 1, indicating good generalization performance of the model.

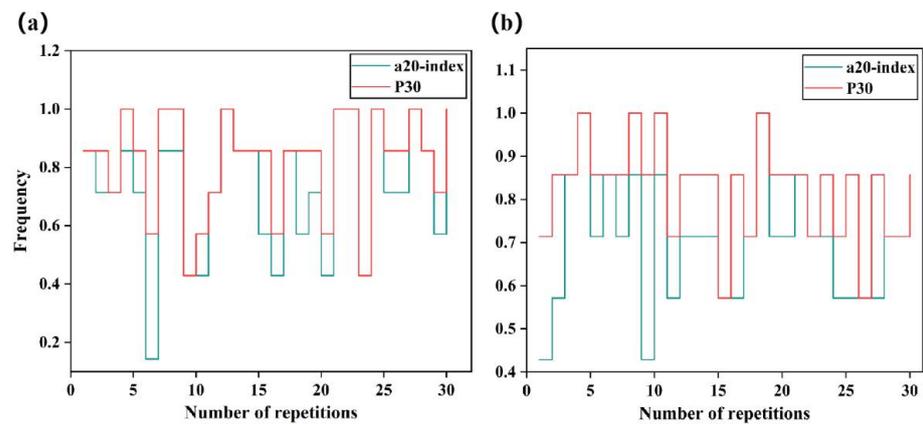


Figure 9. Repeated evaluation of model performance based on cumulative frequency of error: (a) for u_p , (b) for Q_u .

4.3. Relative Importance of Influencing Variables

The importance of influencing variables was analyzed for a deeper understanding of Q_u and u_p . As indicated before, partial dependence plots and the relative importance score were chosen as the methods for interpreting the importance of influencing variables [58,59].

In Figure 10, partial dependence plots of four influencing variables of Q_u and u_p are shown. Normally, the significance of influencing variable is reflected by the output response when an influencing variable is changed. As shown, there was an almost linear growth relation between Q_u and H/D . The finding is quite understandable as the self-weight stress of soil increases with the increase in H/D when D remains unchanged. An almost linear initial growth relation is also shown between u_p and H/D ; however, the growth rate gradually decreases with increasing H/D and tends toward a stable value when H/D reaches a certain size. The output response to variation in the inputs indicates that H/D was the most significant variable for both Q_u and u_p . The importance of the remaining three input variables was similar, although both Q_u and u_p exhibit slow growth increases with increasing n . The authors also note here the relatively small influence of Dr (%) in the current study was due to the limited change in Dr . If the data for loose and medium-dense sand were also included, their influence on Q_u and u_p would be significantly greater.

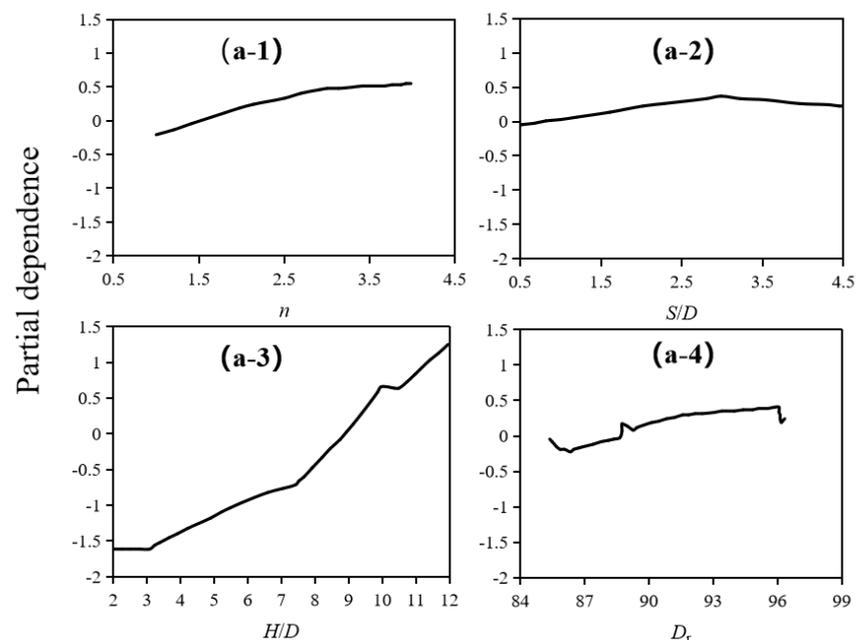


Figure 10. Cont.

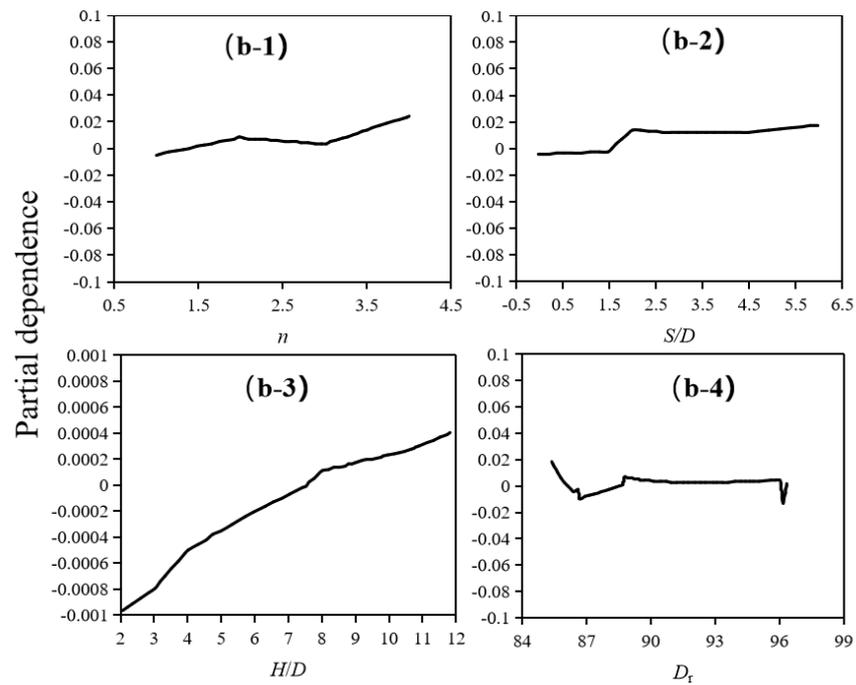
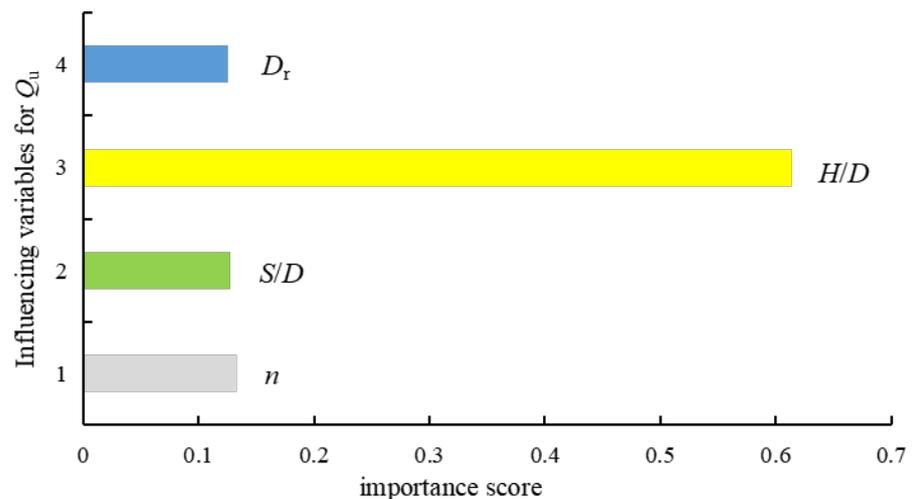


Figure 10. Partial dependence plots of the influencing variables in the optimal GBDT model for predicting: (a-1-a-4) Q_u , (b-1-b-4) u_p .

Figure 11 shows the relative importance score of four influencing variables with the summation of all importance scores being scaled to one; H/D was the most sensitive variable for both Q_u and u_p . The importance scores of H/D were 0.614 and 0.622 for Q_u and u_p , respectively. The influence of H/D on Q_u is consistent with results from previous studies [24,26,60] and the influence of H/D on u_p also agrees with the findings of previous research [26,61]. The other three variables show almost equal importance scores for Q_u . The above results agree well with the partial dependence plots. For u_p , the n parameter had an importance score of 0.058, indicating it is also a non-negligible influencing variable.



(a)

Figure 11. Cont.

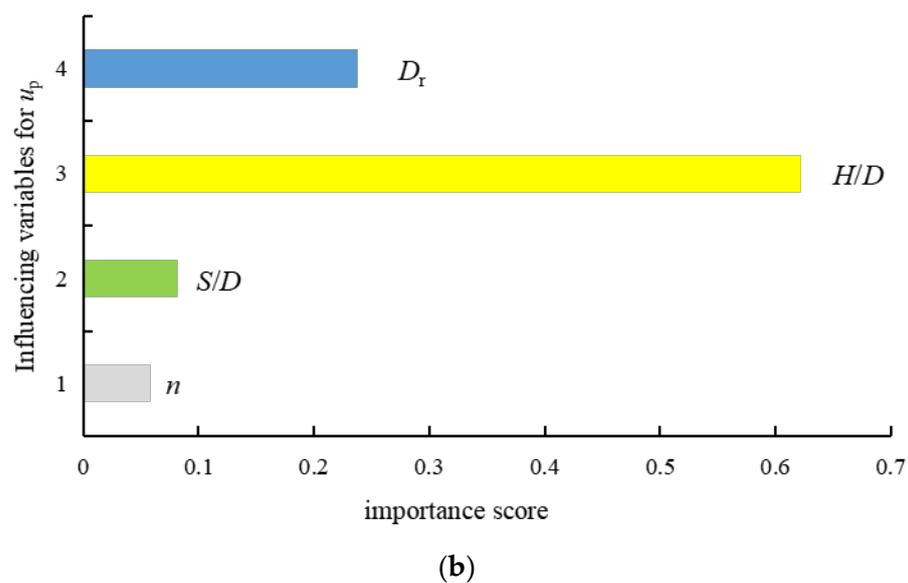


Figure 11. Importance score of influencing variables (1: n ; 2: S/D ; 3: H/D ; 4: D_r): (a) Q_u , (b) u_p .

In general, the partial dependence plots and relative importance scores of influencing variables in this study highlight some important outcomes and indicate future potential experimental studies concerning Q_u and u_p in dense sand. These findings have particular importance for optimizing Q_u and u_p and represent a useful tool for helical anchor design.

4.4. Superiority and Limitations

The primary strength of this study is the proposal of a novel strength prediction model based on GBDT and PSO for the Q_u and u_p prediction of helical anchors in dense fine silica sand. The proposed method was robust in predicting Q_u and u_p and the established model provides new insights into the importance of influencing variables. Broadly, the low-cost, less time-consuming, and non-destructive prediction of Q_u and u_p in this work will help to promote the utilization of helical anchors in engineering.

In terms of limitations, the analysis in this work was focused on dense fine silica sand; thus, the lack of results for Q_u and u_p in medium–dense sand and loose sand is a clear limitation. In addition, data regarding the influence of soil particle size were not included in the present study. Further experimental studies, which could provide more data, are thus required.

5. Conclusions

In this study, based on relevant centrifuge test data, a GBDT–PSO prediction model was constructed to explore the nonlinear relationship between four input variables (e.g., soil relative density) and the ultimate monotonic uplift resistance, Q_u , and the anchor mobilization distance, u_p . A 30-fold cross-validation approach was applied, and six external indicators (R, EVS, MSE, MAE, a20-index, and P30) were used to verify the performance of the optimal GBDT model. In addition, partial dependence plots and importance scores were selected to investigate the parameters' relative importance. The specific conclusions of the work are as follows:

- (1) PSO was efficient in the hyperparameter tuning of GBDT models with maximum R values of 0.987 on the Q_u dataset and 0.957 on the u_p dataset; these results were achieved in the first 10 PSO iterations.
- (2) The optimal GBDT–PSO model has a high generalization ability. For Q_u and u_p , the R values on the testing set were up to 0.93 and 0.82, respectively, and the a20-index values were 0.70 and 0.724, indicating that the model has predictive ability for the lifting behavior of spiral anchors in sand.

- (3) The embedment ratio, H/D , was found to be the most important variable; however, the helix spacing ratio, S/D , was found to have less influence on the capacity of adjacent helices when $S/D > 6$. The influence of other input variable is not obvious.

The prediction results for multiple-helix anchors have non-negligible differences with experimental results, which may be a result of insufficient data. With the filling of more centrifuge test results of in situ test results, the accuracy of prediction could be markedly improved, and the present model can be used for more robust and economic designs of helical anchors in sand.

Author Contributions: Conceptualization, L.W. and C.Q.; methodology, C.Q. and M.W.; validation, L.W., D.H. and C.Q.; formal analysis, M.W. and H.C. investigation, L.W. data curation, D.H. writing—original draft preparation, L.W. and C.Q. writing—review and editing, all authors; supervision, Y.T. funding acquisition, L.W., D.H. and C.Q.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University] grant number [HESS-2120], [State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology], grant number [LP2120], and [National Natural Science Foundation of China], grant number [52078108].

Institutional Review Board Statement: Not applicable, this research not involving humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The first three authors acknowledge the support by State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University (HESS-2120), State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology (LP2120) and National Natural Science Foundation of China (52078108).

Conflicts of Interest: The author declares no conflict of interest.

Nomenclature

D_r	relative density of soil
H	embedment depth of lowest helix
D	helix diameter
d	shaft diameter
S	helix spacing
n	number of helices
t	thickness of helix
p	the pitch of helix
u_p	the anchor mobilization distance
Q_u	the ultimate monotonic uplift resistance
N_γ	the anchor uplift capacity factor
γ	the unit weight of sand
A	the projected area of a single helix or plate
AI	artificial intelligence
ANN	artificial neural network
GBDT	gradient-boosting decision trees
PSO	particle swarm optimization
CART	classification and regression tree
EVS	explained variance score
MSE	mean squared error
MAE	mean absolute error
R	correlation coefficient

References

1. Lutenegeger, A. Behavior of multi-helix screw anchors in sand. In Proceedings of the 2011 Pan-Am CGS Geotechnical Conference, Toronto, ON, Canada, 1–10 October 2011.
2. Merifield, R.S. Ultimate Uplift Capacity of Multiplate Helical Type Anchors in Clay. *J. Geotech. Geoenviron. Eng.* **2011**, *137*, 704–716. [[CrossRef](#)]
3. Kwon, O.; Lee, J.; Kim, G.; Kim, I.; Lee, J. Investigation of pullout load capacity for helical anchors subjected to inclined loading conditions using coupled Eulerian-Lagrangian analyses. *Comput. Geotech.* **2019**, *111*, 66–75. [[CrossRef](#)]
4. Tucker, K. Uplift capacity of drilled shafts and driven piles in granular materials. In *Foundations for Transmission Line Towers*; Geotechnical Special Publication 8; ASCE: New York, NY, USA, 1987; pp. 142–159.
5. Sutherland, H. Uplift resistance of soils. *Geotechnique* **1988**, *38*, 493–516. [[CrossRef](#)]
6. Baker, W.H.; Kondner, R.L. Pullout Load Capacity of a Circular Earth Anchor Buried in Sand. *Highw. Res. Rec.* **1966**, *108*, 1–10.
7. Murray, E.; Geddes, J. Uplift of Anchor Plates in Sand. *J. Geotech. Eng.* **1987**, *113*, 202–215. [[CrossRef](#)]
8. Ghaly, A.; Hanna, A.; Hanna, M. Uplift Behavior of Screw Anchors in Sand. I: Dry Sand. *J. Geotech. Eng.* **1991**, *117*, 773–793. [[CrossRef](#)]
9. Ghaly, A.; Clemence, S. Pullout Performance of Inclined Helical Screw Anchors in Sand. *J. Geotech. Geoenviron. Eng.* **1998**, *124*, 617–627. [[CrossRef](#)]
10. Tagaya, K.; Scott, R.; Aboshi, H. Pullout Resistance of Buried Anchor in Sand. *Soils Found.* **1988**, *28*, 114–130. [[CrossRef](#)]
11. Ilamparuthi, K.; Dickin, E.A.; Muthukrisnaiah, K. Experimental investigation of the uplift behaviour of circular plate anchors embedded in sand. *Can. Geotech. J.* **2002**, *39*, 648–664. [[CrossRef](#)]
12. Liu, J.; Liu, M.; Zhu, Z. Sand Deformation around an Uplift Plate Anchor. *J. Geotech. Geoenviron. Eng.* **2012**, *138*, 728–737. [[CrossRef](#)]
13. Wang, L.; Zhang, P.; Ding, H.; Tian, Y.; Qi, X. The uplift capacity of single-plate helical pile in shallow dense sand including the influence of installation. *Mar. Struct.* **2020**, *71*, 102697. [[CrossRef](#)]
14. Ovesen, N.K. Centrifuge tests of the uplift capacity of anchors. In Proceedings of the 10th International Conference on Soil Mechanics and Foundation Engineering, Stockholm, Sweden, 15–19 June 1981; Volume 1, pp. 717–722.
15. Dickin, E.A. Uplift Behavior of Horizontal Anchor Plates in Sand. *J. Geotech. Eng.* **1988**, *114*, 1300–1317. [[CrossRef](#)]
16. Levesque, C.L. Centrifuge Modelling of Helical Anchors in Sand. Ph.D. Thesis, The University of New Brunswick, Saint John, NB, Canada, 2002.
17. Tsuha, C.H.C.; Aoki, N.; Rault, G.; Thorel, L.; Garnier, J. Evaluation of the efficiencies of helical anchor plates in sand by centrifuge model tests. *Can. Geotech. J.* **2012**, *49*, 1102–1114. [[CrossRef](#)]
18. Hao, D.; Wang, D.; O’Loughlin, C.D.; Gaudin, C. Tensile monotonic capacity of helical anchors in sand: Interaction between helices. *Can. Geotech. J.* **2019**, *56*, 1534–1543. [[CrossRef](#)]
19. Park, H.; Cho, C. Neural Network Model for Predicting the Resistance of Driven Piles. *Mar. Georesour. Geotechnol.* **2010**, *28*, 324–344. [[CrossRef](#)]
20. Momeni, E.; Nazir, R.; Armaghani, D.J.; Maizir, H. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. *Measurement* **2014**, *57*, 122–131. [[CrossRef](#)]
21. Baziar, M.H.; Aziakandi, A.S.; Kashkooli, A. Prediction of pile settlement based on cone penetration test results: An ANN approach. *KSCE J. Civ. Eng.* **2015**, *19*, 98–106. [[CrossRef](#)]
22. Suman, S.; Das, S.K.; Mohanty, R. Prediction of friction capacity of driven piles in clay using artificial intelligence techniques. *Int. J. Geotech. Eng.* **2016**, *10*, 469–475. [[CrossRef](#)]
23. Alzabeebee, S.; Chapman, D.N. Evolutionary computing to determine the skin friction capacity of piles embedded in clay and evaluation of the available analytical methods. *Transp. Geotech.* **2020**, *24*, 100372. [[CrossRef](#)]
24. Alzabeebee, S.; Zuhaira, A.A.; Al-Hamd, R.K.S. Development of an optimized model to compute the undrained shaft friction adhesion factor of bored piles. *Geomech. Eng.* **2022**, *28*, 397–404.
25. Goh, A.T.; Kulhawey, F.H.; Chua, C. Bayesian Neural Network Analysis of Undrained Side Resistance of Drilled Shafts. *J. Geotech. Geoenviron. Eng.* **2005**, *131*, 84–93. [[CrossRef](#)]
26. Zhang, W.; Goh, A.T. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci. Front.* **2016**, *7*, 45–52. [[CrossRef](#)]
27. Moayedi, H.; Moatamediyan, A.; Nguyen, H.; Bui, X.-N.; Bui, D.T.; Rashid, A.S.A. Prediction of ultimate bearing capacity through various novel evolutionary and neural network models. *Eng. Comput.* **2020**, *36*, 671–687. [[CrossRef](#)]
28. Mosallanezhad, M.; Moayedi, H. Developing hybrid artificial neural network model for predicting uplift resistance of screw piles. *Arab. J. Geosci.* **2017**, *10*, 479. [[CrossRef](#)]
29. Javadi, A.; Asr, A.A.; Johari, A.; Faramarzi, A.; Toll, D. Modelling stress–strain and volume change behaviour of unsaturated soils using an evolutionary based data mining technique, an incremental approach. *Eng. Appl. Artif. Intell.* **2012**, *25*, 926–933. [[CrossRef](#)]
30. Schiavon, J.; Tsuha, C.; Thorel, L. Scale effect in centrifuge tests of helical anchors in sand. *Int. J. Phys. Model. Geotech.* **2016**, *16*, 185–196. [[CrossRef](#)]
31. Elith, J.H.; Graham, C.P.; Anderson, R.; Dudík, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; et al. Novel methods improve prediction of species’ distributions from occurrence data. *Ecography* **2006**, *29*, 129–151. [[CrossRef](#)]

32. Olson, R.S.; Cava, W.L.; Mustahsan, Z.; Varik, A.; Moore, J.H. Data-driven advice for applying machine learning to bioinformatics problems. In *Biocomputing 2018*; World Scientific: Kohala Coast, HI, USA, 2018; pp. 192–203. [\[CrossRef\]](#)
33. Zhou, J.; Li, X.; Mitri, H.S. Comparative performance of six supervised learning methods for the development of models of hard rock pillar stability prediction. *Nat. Hazards* **2015**, *79*, 291–316. [\[CrossRef\]](#)
34. Byrne, B.W.; Houlsby, G.T. Helical piles: An innovative foundation design option for offshore wind turbines. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **2015**, *373*, 20140081. [\[CrossRef\]](#)
35. Ren, Q.; Ding, L.; Dai, X.; De Schutter, G. Prediction of Compressive Strength of Concrete with Manufactured Sand by Ensemble Classification and Regression Tree Method. *J. Mater. Civ. Eng.* **2021**, *33*, 04021135. [\[CrossRef\]](#)
36. Tyrallis, H.; Papacharalampous, G. Boosting algorithms in energy research: A systematic review. *Neural Comput. Appl.* **2021**, *33*, 14101–14117. [\[CrossRef\]](#)
37. Zou, Y.; Chen, Y.; Deng, H. Gradient Boosting Decision Tree for Lithology Identification with Well Logs: A Case Study of Zhaoxian Gold Deposit, Shandong Peninsula, China. *Nonrenew. Resour.* **2021**, *30*, 3197–3217. [\[CrossRef\]](#)
38. Chou, J.P.E.; Chiu, C.; Farfoura, M.; Al-Taharwa, I. Optimizing the Prediction Accuracy of Concrete Compressive Strength Based on a Comparison of Data-Mining Techniques. *J. Comput. Civ. Eng.* **2011**, *25*, 242–253. [\[CrossRef\]](#)
39. Du, X.; Xu, H.; Zhu, F. A data mining method for structure design with uncertainty in design variables. *Comput. Struct.* **2020**, *244*, 106457. [\[CrossRef\]](#)
40. Kiani, J.; Camp, C.; Pezeshk, S. On the application of machine learning techniques to derive seismic fragility curves. *Comput. Struct.* **2019**, *218*, 108–122. [\[CrossRef\]](#)
41. Carrizosa, E.; Molero-Río, C.; Morales, D.R. Mathematical optimization in classification and regression trees. *TOP* **2021**, *29*, 5–33. [\[CrossRef\]](#)
42. Dimou, C.; Koumousis, V. Reliability-Based Optimal Design of Truss Structures Using Particle Swarm Optimization. *J. Comput. Civ. Eng.* **2009**, *23*, 100–109. [\[CrossRef\]](#)
43. Bai, W.; Wang, Z.; Liu, H.; Yu, D.; Chen, C.; Zhu, M. Optimisation of the finite-difference scheme based on an improved PSO algorithm for elastic modelling. *Explor. Geophys.* **2020**, *52*, 419–430. [\[CrossRef\]](#)
44. Yan, J.; Gao, Y.; Yu, Y.; Xu, H.; Xu, Z. A Prediction Model Based on Deep Belief Network and Least Squares SVR Applied to Cross-Section Water Quality. *Water* **2020**, *12*, 1929. [\[CrossRef\]](#)
45. Jia, B.; Wu, J.; Du, J.; Ji, Y.; Zhu, L. A prediction model for the secure issuance scale of Chinese local government bonds. *Kybernetes* **2020**, *50*, 1125–1143. [\[CrossRef\]](#)
46. Chow, S.H.; O’Loughlin, C.D.; Corti, R.; Gaudin, C.; Diambra, A. Drained cyclic capacity of plate anchors in dense sand: Experimental and theoretical observations. *Géotech. Lett.* **2015**, *5*, 80–85. [\[CrossRef\]](#)
47. Zhu, F.Y.; Bienen, B.; O’Loughlin, C.; Cassidy, M.J.; Morgan, N. Suction caisson foundations for offshore wind energy: Cyclic response in sand and sand over clay. *Géotechnique* **2019**, *69*, 924–931. [\[CrossRef\]](#)
48. Qi, C.; Fourie, A.; Chen, Q.; Zhang, Q. A strength prediction model using artificial intelligence for recycling waste tailings as cemented paste backfill. *J. Clean. Prod.* **2018**, *183*, 566–578. [\[CrossRef\]](#)
49. Khan, M.A.; Zafar, A.; Farooq, F.; Javed, M.F.; Alyousef, R.; Alabduljabbar, H. Geopolymer Concrete Compressive Strength via Artificial Neural Network, Adaptive Neuro Fuzzy Interface System, and Gene Expression Programming With K-Fold Cross Validation. *Front. Mater.* **2021**, *8*, 621163. [\[CrossRef\]](#)
50. Alzabeebee, S.; Alshkane, Y.M.; Al-Taie, A.J.; Rashed, K.A. Soft computing of the recompression index of fine-grained soils. *Soft Comput.* **2021**, *25*, 15297–15312. [\[CrossRef\]](#)
51. Alzabeebee, S.; Alshkane, Y.M.; Rashed, K.A. Evolutionary computing of the compression index of fine-grained soils. *Arab. J. Geosci.* **2021**, *14*, 2040. [\[CrossRef\]](#)
52. Alzabeebee, S.; Mohammed, D.A.; Alshkane, Y.M. Experimental Study and Soft Computing Modeling of the Unconfined Compressive Strength of Limestone Rocks Considering Dry and Saturation Conditions. *Rock Mech. Rock Eng.* **2022**, *55*, 5535–5554. [\[CrossRef\]](#)
53. Alzabeebee, S. Explicit soft computing model to predict the undrained bearing capacity of footing resting on aggregate pier reinforced cohesive ground. *Innov. Infrastruct. Solut.* **2022**, *7*, 105. [\[CrossRef\]](#)
54. Zhang, L.; Wu, X.; Ji, W.; AbouRizk, S.M. Intelligent Approach to Estimation of Tunnel-Induced Ground Settlement Using Wavelet Packet and Support Vector Machines. *J. Comput. Civ. Eng.* **2017**, *31*, 04016053. [\[CrossRef\]](#)
55. Jin, X.; Zhu, X.; Li, S.; Wang, W.; Qi, H. Predicting soil available phosphorus by hyperspectral regression method based on gradient boosting decision tree. *Laser Optoelectron. Prog.* **2019**, *56*, 131102. [\[CrossRef\]](#)
56. Ye, Y.; Xiong, Y.; Zhou, Q.; Wu, J.; Li, X.; Xiao, X. Comparison of machine learning methods and conventional logistic regressions for predicting gestational diabetes using routine clinical data: A retrospective cohort study. *J. Diabetes Res.* **2020**, *2020*, 4168340. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Jun, M.-J. A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: The case of the Seoul metropolitan area. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 2149–2167. [\[CrossRef\]](#)
58. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C.J. *Classification and Regression Trees*, Wadsworth Statistics; Probability Series; Wadsworth: Belmont, CA, USA, 1984. [\[CrossRef\]](#)
59. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)

-
60. Giampa, J.; Bradshaw, A.; Schneider, J. Influence of Dilation Angle on Drained Shallow Circular Anchor Uplift Capacity. *Int. J. Géoméch.* **2017**, *17*, 04016056. [[CrossRef](#)]
 61. Wang, J.; Haigh, S.K.; Forrest, G.; Thusyanthan, N.I. Mobilization Distance for Upheaval Buckling of Shallowly Buried Pipelines. *J. Pipeline Syst. Eng. Pract.* **2012**, *3*, 106–114. [[CrossRef](#)]