



OPEN Prediction of the displacements of the pile tops and ground surface around piles based on machine learning algorithms

Penglin Li¹, Shaolong Guo²✉, Manman Liang² & Qun Lu²

The soil squeezing effect of pile groups may cause displacements and deformation at the pile tops and ground surface around piles. In severe cases, it can cause problems such as broken piles, cracking of adjacent buildings or cracking of pipes. Artificial intelligence provides a new way to predict horizontal displacements of the pile tops and ground surface around piles caused by soil squeezing effect. The adaptive boosting (AdaBoost) algorithm was applied to the back propagation (BP) neural network model to form the Adaboost-BP model, which improved the learning ability of the BP neural network. For small sample datasets, the prediction accuracy of AdaBoost-BP model, Random Forest (RF) model and Deep Neural Networks (DNN) model is higher than that of BP model. For large sample datasets, the prediction accuracy of various models has improved, but the BP model is lower than that of other models. Analysis shows that the horizontal distance and angle between the center of the bearing platform and the center of the pile tops (or ground surface monitoring points) are the two most important influencing factors. The resting time is also an important influencing factor. Moisture content, relative density, and internal friction angle have a more significant influence on the horizontal displacements of the pile tops and ground surface around piles than other soil property indexes. Quantile regression analysis shows that the horizontal displacements is negatively correlated with the horizontal distance, and positively correlated with the rest time and moisture content. The prediction accuracy of machine learning algorithms (such as DNN) is higher than that of the cylindrical hole expansion method.

Keywords Squeezing effects, Jacked piles, Field measurement, AdaBoost-BP Algorithms, DNN Algorithms

The squeezing effect¹ caused by the driving process of jacked piles may cause vertical and horizontal displacements of the soil in the adjacent area, leading to the floating and displacements of the piles. In severe cases, it can cause problems such as broken piles, cracking of adjacent buildings, road heave, pipeline deformation, etc. Excessive piles position deviation will lead to uneven load distribution and even structural instability. Therefore, predicting and controlling the squeezing effect of jacked piles is the key to ensure the quality of the project. It has important application value to carry out the prediction of the displacements of the pile tops and ground surface around piles.

In terms of theoretical analysis, classical theoretical analysis methods include cylindrical hole expansion method² and strain path method³. The cylindrical expansion method regards the process of piles driving as the expansion of a cylindrical hole in an elastic-plastic infinite medium, and the material follows the Tresca or Mohr Coulomb yield criterion. The theory holds that the resistance of piles driving is related to the deformation modulus and strength of soil, and it can reflect the nonlinear characteristics of soil and simulate the working state of actual piles. Chen et al.⁴ proposed a novel graph-based analysis method for analyzing the response of expanded cylindrical cavities in modified Cam clay under non-drainage conditions. Gao et al.⁵ uses the curve equation of quartic polynomial to simulate the boundary of pile hole. Based on the assumption that single-pile penetration can be simulated through a series of spherical cavity expansions, Li⁶ provided an analytical solution for cavity expansions near the slope. The solution provides a simplified and realistic theoretical method to predict the soil behaviors around the spherical cavity near the sloping ground.

¹China Harbor Engineering Co., LTD., Beijing 100027, China. ²Tianjin Chengjian University, Tianjin 300384, China. ✉email: dream2020@tcu.edu.cn

The theory of cylindrical hole expansion⁷ is the study of cylindrical hole expansion under the action of uniform internal pressure p . The schematic diagram of the cylindrical hole expansion method is shown in Fig. 1. When the internal pressure p increases, the cylindrical region around the cylindrical hole will enter the plastic state from the elastic state, and the plastic region will expand with the increase of the internal pressure p . The maximum radius of the plastic zone is R_p , and the corresponding limit expansion pressure is P_u . The soil outside the radius R_p still maintains the elastic equilibrium state. In order to compare with the calculation results of machine learning algorithms, the cylindrical hole expansion method was simultaneously used to calculate the displacements of the soil around the pile.

where R_u is the initial radius of cylindrical hole. R_p is the maximum radius of plastic zone. P_u is the limit expansion pressure value. r is the calculate the distance between point and piles center. r_0 is the initial radius of the pile. a is the cylindrical hole radius during expansion. u_p is the radial displacement of the boundary of the influence zone. σ_θ is the tangential stress of soil. σ_r is the radial stress of soil.

The displacements in the elastic region are:

$$u_r = \frac{(1 + \mu)c_0 \cos \varphi_0}{E} \left(\frac{R_p}{r}\right)^2 r \quad (1)$$

where μ is the Poisson's ratio. c_0 is the cohesion of the soil. φ_0 is the internal friction angle of the soil.

It should be noted that the elastic modulus E in the above calculation formulas is different from the compression modulus E_s of soil. E and E_s can be converted according to the following relationship:

$$E = E_s \left(1 - \frac{2\mu^2}{1 - \mu}\right) \quad (2)$$

Lu et al.⁸ used finite element method to simulate the complete process of continuous driving of a single pile. Luo et al.⁹ studied the influence law of shielding effect on soil displacements of jacked piles, and found that shielding effect had significant influence on soil displacements in front direction and back direction. Shao et al.¹⁰ used the finite element method to simulate the displacements of soft clay and the underlying gravel layer with the increase of soil depth and radial distance during pile driving. It is found that the lateral soil displacement is obvious in the area 1.0 m away from the prestressed high-strength concrete (PHC) pile axis, and decreases with the increase of the radial distance during pile driving. When the radial distance is above 4.0 m, the lateral displacement can be ignored.

Zhou et al.¹¹ monitored the driving process of three static press piles in saturated clay, and analyzed the variation laws of lateral displacement of soil around the piles, vertical uplift value of the ground, and pore water pressure with the depth of pile driving and distance from the pile center. Zhang et al.¹² found that the soil deformation caused by pile driving first increases and then decreases in depth, and the soil deformation decreases exponentially in the horizontal direction. Under the influence of compression, the width of the shear strain zone does not change with the increase of driving. Yuan et al.^{13,14} proposed a method for visualizing the soil displacement field around laterally loaded piles using transparent soil technology. The influence of passive piles on the three-dimensional ground deformation around laterally loaded piles and laterally loaded piles was studied through a series of model experiments.

The theoretical analysis method simplifies the piles group into a single pile, which is very different from the actual situation. The field test method needs more material resources, and the test cost is high. The conventional model test method cannot guarantee the similarity ratio, and the centrifuge test cost is very high. In the numerical simulation of soil squeezing effect of pile groups, the grid will have large deformation, which will cause convergence difficulties, and the error between the simulation results and the measured results is often large.

Considering that there are many factors affecting the soil squeezing effect of piles group, the mechanism is complex and has strong nonlinearity, which is difficult to be expressed directly by explicit function. Although empirical formulas can be used to establish the mathematical expressions between various influencing factors and soil squeezing effect indexes (such as excess pore water pressure, the displacements of the pile tops and ground surface around piles), the prediction accuracy and universality of these empirical formulas are often not ideal. The rapid development of artificial intelligence (AI) technology provides new tools for the progress

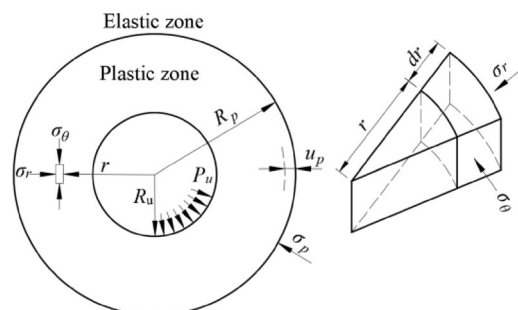


Fig. 1. The schematic diagram of the cylindrical hole expansion method.

of many industries. Machine learning is an important part of AI. In recent years, the application of machine learning in all walks of life has developed rapidly. Machine learning is an appropriate and effective method to solve engineering problems.

Many scholars^{15–19} have introduced algorithms with excellent nonlinear mapping ability such as back propagation (BP), adaptive boosting (Adaboost), deep neural networks (DNN), random forest (RF), extreme gradient boosting (XGBoost) and support vector machine (SVM) into pile foundation engineering, and established many prediction models of pile bearing capacity considering different influencing factors (such as the geometric parameters of the piles foundation, soil physical and mechanical parameters, standard driving test (SPT) value, resting time, etc.). Kordjazi et al.²⁰ established a SVM prediction model of pile bearing capacity based on 108 sample data sets including geometric parameters of pile foundation, pile load test and cone penetration test (CPT) test data. Shahin et al.²¹ established a recurrent neural networks (RNN) bearing capacity-settlement prediction model of pile foundation based on field load test and CPT test data. Moayed, et al.²² established a prediction model of the load-settlement relationship curve of pile foundation based on the in-situ CPT data set by feedforward neural network (FFNN) and focused time delay neural network (FTDNN). Tan et al.^{23,24} proposed an innovative hybrid machine learning model specifically for predicting the load–displacement characteristics of bored in-situ piles. This model establishes a complex relationship between key design parameters (diameter, length, SPT index and effective overlay pressure) and the load–displacement response of piles. Tram et al.^{25,26} addresses a robust predictive model for the axial load-bearing behavior of pre-bored grouted planted nodular (PGPN) piles. This model adopts a new hybrid method for predicting pile head settlement and has been applied in pile foundation engineering in Vietnam. Yuan et al.²⁷ examined the effects of coral sand particle size and rigid pile embedment depth on pile-soil interaction. The horizontal strain distribution in coral sand around piles under lateral load was disclosed.

At present, there are many research achievements on the prediction of pile foundation bearing capacity and pile foundation settlement based on machine learning algorithms. The influence laws of various influencing factors on the bearing capacity of pile foundations and the settlement of pile foundations were analyzed. However, there are few research results on machine learning prediction of the soil squeezing effect (pile foundation displacements, soil displacement around the pile) caused by the driving of piles group, and the analysis of the action laws of various factors affecting the soil squeezing effect of piles group is not thorough. Based on the machine learning prediction for the bearing capacity and settlement of the reference pile foundation, the relationship between the horizontal displacements of the pile tops and ground surface around piles and various influencing factors was established using the machine learning algorithm in this paper. In addition, the patterns of the effects of various influencing factors have been analyzed.

BP neural network is a multilayer feedforward neural network trained according to the error back propagation algorithm (Tiwari et al.²⁸), which is one of the most widely used neural network models. But BP neural network has slow convergence speed, low learning efficiency and easy to converge to a local minimum (Guo et al.²⁹). In addition, common machine learning algorithms also include DNN, RF, XGBoost, SVM, et al. (Lin et al.³⁰; Huang et al.³¹). As a representative of the Boosting series of algorithms, the Adaboost algorithm can improve the prediction accuracy of the model by gradually enhancing the model's performance. AdaBoost can effectively enhance the prediction accuracy of BP neural networks. Adaboost-BP has disadvantages such as easy overfitting and slow training speed. AdaBoost-BP has been applied in the prediction of foundation and base settlement and is expected to be extended to other engineering scenarios^{32,33}.

DNN can automatically learn more advanced and essential feature representations than shallow networks (such as BP networks), making it more suitable for handling complex tasks. DNN requires a large amount of data and computing power, is complex to train, and is easily affected by hyperparameters. RF effectively reduces the variance of a single decision tree through random sampling and random feature selection, and has a natural ability to resist overfitting. However, the RF algorithm often performs worse than the Boosting algorithm in dealing with complex problems. As an outstanding representative of the Boosting series of algorithms, XGBoost effectively reduces bias and variance and improves prediction accuracy through techniques such as gradient boosting, regularization, and weighted quantiles. The XGBoost algorithm has complex parameter tuning and is prone to overfitting on small data. AdaBoost-BP, RF and XGBoost are all integrated algorithms. Although DNN is not an integrated algorithm, its performance can be improved through the idea of integrated learning.

Each algorithm has its own advantages and disadvantages. In this study, multiple algorithms (AdaBoost-BP, DNN, RF and XGBoost) were used for displacements prediction, and the prediction results of different algorithms were compared. In addition, in order to compare with the calculation results of machine learning algorithms, the cylindrical hole expansion method was used to calculate the soil displacements around the pile.

The main influencing factors of the displacements of the pile tops and ground surface around piles

The influencing factors of soil squeezing effect of jacked piles mainly include soil properties, pile driving sequence, pile spacing, time effect, etc. (Sagaseta et al.⁶; Lu et al.⁸; Zhou et al.¹¹). Different soil properties have different responses to soil squeezing effect. For example, unsaturated soil and sand will be compacted under the squeezing effect (void ratio decreases), while saturated cohesive soil will cause lateral displacement and vertical uplift due to excess pore water pressure. The influence of pile driving sequence on soil squeezing effect is also obvious⁸. For example, the soil squeezing effect of pile driving from the four sides to the middle is usually more serious than that of pile driving from the middle to the four sides. In addition, the size of the pile (pile diameter, pile length), pile spacing and soil plug effect on pile driving will affect the soil squeezing effect of jacked piles⁸. The soil property index of cohesive soil includes physical property index, plasticity index and liquid index. The three most important indexes of soil property, namely relative density, moisture content, and density, can be directly measured in the laboratory (Guo et al.²⁹). Compression modulus, cohesion force and internal friction

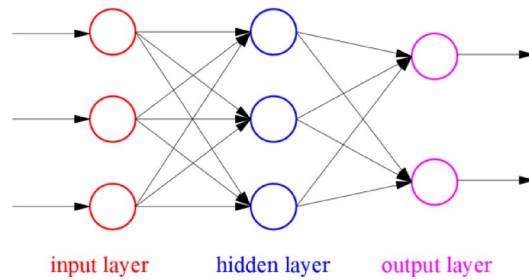


Fig. 2. Topological structure of BP neural network.

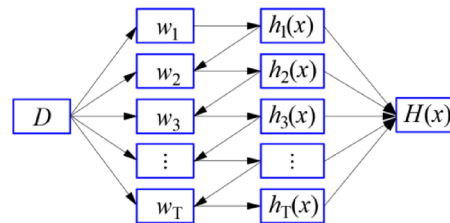


Fig. 3. Topological structure of the AdaBoost algorithm.

angle are also important factors affecting the displacement of the pile tops and ground surface around piles. In addition, factors such as pile diameter, pile length, bending stiffness of the pile body, number of piles, pile spacing, and time effect (the rest age after pile driving) also have a significant impact on the displacement of the pile tops and ground surface around piles.

In summary, there are 15 main factors affecting the displacements of pile tops of jacked piles group considered in this paper, which are respectively moisture content, natural density, relative density, compression modulus, cohesion, internal friction angle, pile diameter, pile length, bending stiffness of pile body, number of piles in each row in X direction, number of piles in each row in Y direction, pile spacing, resting time, the distance and orientation between the center of the pile tops and the center of the bearing platform. There are 15 main factors affecting the displacements of the ground surface around piles, which are moisture content, natural weight, relative density, compression modulus, cohesion, internal friction Angle, pile diameter, pile length, bending stiffness of pile body, number of piles in each row in X direction, number of piles in each row in Y direction, pile spacing, resting time, the distance and orientation between the monitoring points and the center of the bearing platform. Considering that the soil layer within the engineering site is multi-layered soil, referring to the method proposed by Liu et al.⁷, the average moisture content, average natural weight, average relative density, average compression modulus, average cohesion and average internal friction angle of the soil within the pile length range were obtained through the weighted average of soil layer thickness.

Introduction to machine learning algorithms

BP neural network

The classic BP neural network is composed of three layers: input layer, hidden layer and output layer (Guo et al.²⁹). The topology structure of the three-layer BP network is shown in Fig. 2.

The algorithm expression of BP neural network can be found in relevant references (Wang.³⁴).

The number of hidden layer units is largely dependent on experience (Tiwari et al.²⁸). The selection range was determined by using the method of Deng et al.³⁵.

$$\sum_{i=0}^n C_{n_1}^i > k \quad (3)$$

$$n_1 = \sqrt{n + m} + a \quad (4)$$

$$n_1 \geq \log_2 n \quad (5)$$

The parameters in Eqs. (3), (4) and (5) are shown in reference (Deng et al.³⁵).

AdaBoost algorithm

Adaboost combines multiple weak classifiers into a strong classifier by iteratively adjusting sample weights and weak classifier weights (Murmu et al.³⁶). The topology structure of AdaBoost algorithm is shown in Fig. 3.

The expression of the AdaBoost algorithm can be found in reference Guo et al.²⁹.

AdaBoost-BP algorithm

For each iteration (i.e., each weak predictor), a weak predictor is first trained using the current sample weights, and the prediction error of this predictor is calculated. If the prediction error of the samples in a certain round of iteration exceeds the set threshold, the sample weights will be updated based on the current weak predictor's weight. The specific approach is to adjust the model by increasing the weight of the incorrect samples and reducing the weight of the correct samples. In the next round of training, more attention will be paid to the samples with higher errors, and the weight of each sample will be added to the total error rate. That is, the sum of the prediction errors. Calculate the weight of the current weak predictor based on the sum of the prediction errors.

The calculation process is as follows:

Step 1: Randomly select m groups of training data in the sample space and initialize the distribution weights $D_t(i)$ of the data (Guo et al.²⁹):

$$D_t(i) = 1/m \quad (6)$$

Step 2: Determine network parameters (such as the number of nodes in the input layer, hidden layer, and output layer).

Step 3: Train to obtain the t -th BP weak predictor, denoted as $h_t(x)$. The cumulative error exceeds the weight $D_t(i)$ of the corresponding term δ_t to obtain the calculation error ε_t :

$$\varepsilon_t = \sum_{i=1}^m D_t(i) \quad (7)$$

Step 4: Calculate the weight α_t for $h_t(x)$ based on the error ε_t calculated in Step 3:

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad (8)$$

Step 5: Adjust the weight of training data.

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t}, & \varepsilon_t < \delta_t \\ e^{\alpha_t}, & \varepsilon_t \geq \delta_t \end{cases} \quad (9)$$

where:

$$Z_t = \sum_{i=1}^m D_t(i) \times \begin{cases} e^{-\alpha_t}, & \varepsilon_t < \delta_t \\ e^{\alpha_t}, & \varepsilon_t \geq \delta_t \end{cases} \quad (10)$$

Step 6: After training T rounds, obtain T weak predictors.

Step 7: Output strong predictor:

$$H(x) = \sum_{i=1}^T \frac{\alpha_t}{\sum_{t=1}^T \alpha_t} h_t(x) \quad (11)$$

Introduction to other machine learning methods

Besides AdaBoost algorithm, DNN, XGBoost, Bagging, RF and other algorithms are often used for prediction (Lin et al.³⁰). Currently, the DNN used in the analysis are mainly FFNN. The depth of the network refers to the number of hidden layers. Different from the traditional shallow neural network, DNN can extract features from low to high, learn features between data at a deeper level, extract the features of each layer, and establish a mapping relationship from the bottom signal to the top signal. DNN has a deep nonlinear structure that approximates arbitrary complex functions, which is an important feature of traditional shallow neural network, and has stronger ability to deal with complex, uncertain and fuzzy data. DNN can express larger and more complex functions. The topology structure of a DNN with three hidden layers is shown in Fig. 4. In the DNN structure, the layers are connected to each other.

Xgboost is an enhanced tree model (Guo et al.²⁹), which integrates many decision trees to form a stronger learner. RF is an ensemble learning method that belongs to a type of supervised learning algorithm (Chen et al.³⁷), and it is a classifier or regressor composed of multiple decision trees. In this paper, BP, AdaBoost-BP, DNN, Xgboost and RF algorithms are used to predict the squeezing effect.

Case analysis

Data sources

The length of the Metro Line 1 project in Bogota, the capital of Colombia, is about 23.9 km. The whole line of the project is a viaduct. The general feature of the landform of the project is that the groundwater level is shallow, the thickness of silt and peat soil is large, and there are often fine sand and clay intercalations. Most pile foundations of the project are PHC pipe piles. The project has a total of 6232 PHC pipe piles, and each static

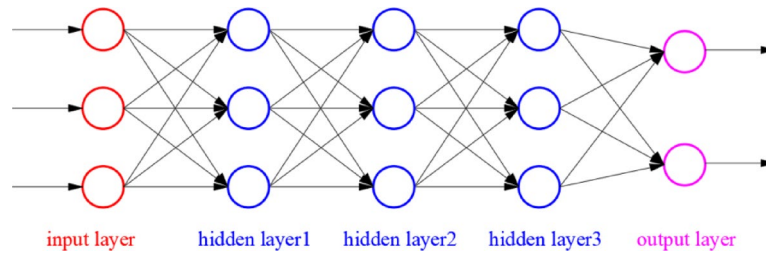


Fig. 4. Topology structure of DNN with three hidden layers.

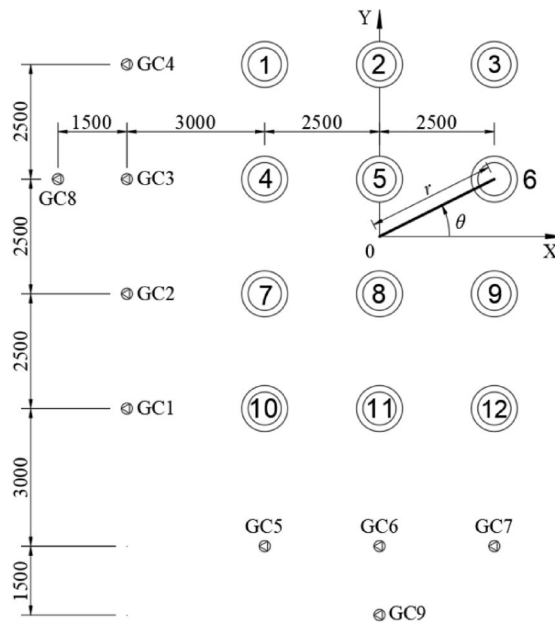


Fig. 5. Pile position layout diagram of 12 pile groups.

pile driver works an average of 3 piles per day. The model of PHC pipe pile is PHC-1000-140, with a diameter of 1000 mm and a wall thickness of 140 mm. The length of the piles is 15 to 48 m, and the concrete strength grade of the pile body is C60. The number of piles under the bearing platform is mainly 12, 16 and 20, and the pile spacing is 2.5 m. There are many different types of pipelines around the pile group (gas, water supply and drainage, cables, communications, etc.), and the materials include concrete, cast iron, PVC, ceramics, etc.). The distance between most pipelines and adjacent pile foundations is 0.4 m ~ 5 m. The buried depth of the pipeline is between 1-5 m. The geological survey data of the project is complete. The construction company monitored the horizontal displacements of the pile tops and ground surface around piles before and after the construction of PHC pipe piles, and obtained a large number of measured data.

The input of the prediction model of displacements of the pile tops is as follows: Feature1 is moisture content (%). Feature2 is natural weight (kN/m^3). Feature3 is the relative density (%). Feature4 is the compression modulus (MPa). Feature5 is the cohesion force (kPa). Feature6 is the internal friction angle ($^\circ$). Feature7 is the resting time (day). Feature8 is the horizontal distance r (m) between the center of the bearing platform and the center of the pile (Fig. 5); Feature9 represents the angle θ between the line connecting the center of the bearing platform and the center of the pile and the positive direction of the X-axis ($^\circ$, counterclockwise is positive, Fig. 5)). Feature10 is the pile diameter (mm); Feature11 is the pile length (m); Feature12 is the pile bending stiffness EI ($\text{N}\cdot\text{mm}^2$); Feature13 is the number of piles per row in the X direction; Feature14 is the number of piles per row in the Y direction; Feature15 is pile spacing (m). The output value is the horizontal displacements (mm) of the pile tops. Part of the sample data in this paper are shown in Table 1. The values of Feature10 ~ Feature15 are unchanged (1000, 30, $1.29\text{E}21$, 3,4,2.5 respectively). Through trial calculation, it is found that removing the variables (Feature10 ~ Feature15) will improve the calculation speed of the model and have no significant impact on the prediction accuracy of the model. So these variables were removed in the modeling of this case. The pile position layout of the 12 pile groups is shown in Fig. 5. The pile driving sequence is $5 > 8 > 2 > 11 > 4 > 6 > 7 > 9 > 10 > 1 > 3 > 12$.

Similarly, the input of the prediction model of displacements of the ground surface around piles is as follows: Feature1 is moisture content (%). Feature2 is natural weight (kN/m^3). Feature3 is the relative density (%). Feature4 is the compression modulus (MPa). Feature5 is the cohesion force (kPa). Feature6 is the internal friction angle

Sample no	Feature1 (%)	Feature2 (kg/m ³)	Feature3 (%)	Feature4 (MPa)	Feature5 (kPa)	Feature6 (°)	Feature7 (d)	Feature8 (m)	Feature9 (°)	Output value
1	68.2	15.6	2.62	13.2	20.2	23.6	5	4.5	123.7	63
2	68.2	15.6	2.62	13.2	20.2	23.6	8	3.8	90.0	65
3	68.2	15.6	2.62	13.2	20.2	23.6	7	4.5	56.3	31
4	68.2	15.6	2.62	13.2	20.2	23.6	7	2.8	153.4	73
5	68.2	15.6	2.62	13.2	20.2	23.6	8	1.3	90.0	12
6	68.2	15.6	2.62	13.2	20.2	23.6	7	2.8	26.6	34
7	68.2	15.6	2.62	13.2	20.2	23.6	6	2.8	206.6	122
8	68.2	15.6	2.62	13.2	20.2	23.6	8	1.3	270.0	59
9	68.2	15.6	2.62	13.2	20.2	23.6	6	2.8	333.4	102
10	68.2	15.6	2.62	13.2	20.2	23.6	6	4.5	236.3	73
11	68.2	15.6	2.62	13.2	20.2	23.6	7	3.8	270.0	161
12	68.2	15.6	2.62	13.2	20.2	23.6	5	4.5	303.7	44

Table 1. Partial samples of horizontal displacements at pile tops.

Sample no	Feature1 (%)	Feature2 (kg/m ³)	Feature3	Feature4 (MPa)	Feature5 (kPa)	Feature6 (°)	Feature7 (d)	Feature8 (m)	Feature9 (°)	Output value
1	86.9	15.0	2.6	21.3	15.3	25.5	3	7.2	249.7	82
2	86.9	15.0	2.6	21.3	15.3	25.5	3	6.8	270.0	69
3	86.9	15.0	2.6	21.3	15.3	25.5	3	7.2	290.3	57
4	86.9	15.0	2.6	21.3	15.3	25.5	3	7.2	110.3	41
5	86.9	15.0	2.6	21.3	15.3	25.5	3	6.8	90.0	30
6	86.9	15.0	2.6	21.3	15.3	25.5	3	7.2	69.7	23
7	86.9	15.0	2.6	21.3	15.3	25.5	3	6.7	325.7	44
8	86.9	15.0	2.6	21.3	15.3	25.5	3	5.6	347.2	37
9	86.9	15.0	2.6	21.3	15.3	25.5	3	5.6	12.8	76
10	86.9	15.0	2.6	21.3	15.3	25.5	3	8.3	270.0	70
11	86.9	15.0	2.6	21.3	15.3	25.5	3	7.1	349.9	33

Table 2. Partial samples of surface horizontal displacements of the ground surface around piles.

(°). Feature7 is the resting time (day). Feature8 is the horizontal distance (m) between the center of the bearing platform and the monitoring point. Feature9 represents the angle between the line connecting the center of the bearing platform and the center of the pile and the positive direction of the X-axis (°, counterclockwise is positive). Feature10 is the pile diameter (mm); Feature11 is the pile length (m); Feature12 is the pile bending stiffness EI (N·mm²); Feature13 is the number of piles per row in the X direction; Feature14 is the number of piles per row in the Y direction; Feature15 is pile spacing (m); The output value is the horizontal displacements (mm) of the ground surface monitoring point. Part of the sample data in this paper are shown in Table 2. The values of Feature10 ~ Feature15 are unchanged (1000, 30, 1.29E21, 3, 4 and 2.5 respectively). For the same reason as above, these variables were removed in the modeling of this case.

Through the displacement monitoring, 512 monitoring data of the horizontal displacements of the pile tops and 459 monitoring data of the horizontal displacements of the ground surface around piles were obtained. At the same time, the soil parameters of each piles group site are collected from the geological survey report of the project.

In order to analyze the influence of sample size on the model prediction results, the horizontal displacements predictions of the pile tops and ground surface around piles were carried out based on large samples and small samples, respectively. For the dataset of horizontal displacements of the pile tops, the number of large samples and small samples are 512 and 103, respectively. For the dataset of horizontal displacements of the ground surface around piles, the number of large samples and small samples are 459 and 84, respectively.

The *K*-Fold cross-validation method can solve the problem of a small number of samples in a dataset and also address the issue of hyperparameter tuning. Taking the prediction of the horizontal displacements of the pile tops as an example, the specific method of the fivefold cross-validation used in this paper is introduced as follows: i) Randomly select the horizontal displacements data at the pile tops corresponding to a certain bearing platform from all datasets as the test set (the random sampling tool in Excel was used to extract 12 samples). ii) Divide the remaining dataset into five equal parts. iii) Use one of them as the validation set and the remaining four as the training set each time. iv) Train the model and calculate the validation error. v) Repeat the above process five times. vi) Take the average of the verification errors of the five times as the final evaluation index.

The mathematical formula for *K*-fold cross-validation:

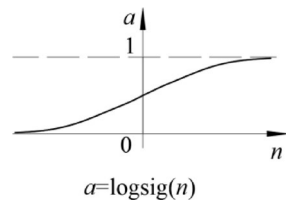


Fig. 6. Log-Sigmoid transfer function.

Type	BP			AdaBoost	
	Learning rate	Maximum training number	Number of hidden layer nodes	Number of weak predictors	Initial weight $D_1(i)$
Hyperparameters ranges	0.002, 0.005, 0.01, 0.015, 0.02	1000, 2000, 3000, 4000, 5000	8, 9, 10, 11, 12, 13, 14	8, 10, 12, 14, 16, 18	1/130, 1/160, 1/190, 1/220, 1/250
Optimized values	0.01	3000	11	10	1/190

Table 3. Superparameters ranges and optimized values of BP model and AdaBoost model.

$$CV_k = \frac{1}{K} \sum_{i=1}^K Loss_i \tag{12}$$

where K is the number of folds. $Loss_i$ is the loss value on the i -th fold validation set (the mean square error (MSE) is adopted). $CV_{(k)}$ is the average loss value of K -fold cross-validation.

Parameter tuning was performed using the grid search method. The steps are as follows: by traversing the parameter grid (such as the Learning rate of 0.002, 0.005, 0.01, 0.015, 0.02), calculate the K -fold cross-validation mean score (MSE) for each combination. Select the parameter combination with the lowest MSE value. Then, the model is retrained on the full training set using the optimal parameters, and the final performance is evaluated through the test set.

The machine learning algorithm runs in the Windows 10 operating system, with an Intel Core i5-7300HQ processor and 32G of memory. The computation time for a single model ranges from 3 to 15 min.

The BP network established consists of 3 layers, in which there are 9 neurons in the input layer (Feature1, Feature2, Feature3.....Feature15) and 1 neuron in the output layer (Y). The hidden layer of BP model uses S-type transfer function to calculate the number of hidden layer neurons according to formula (3)~(5). From Eqs. (3) to (5), the number of neurons in the hidden layer can be taken from 8 to 14.

The AdaBoost-BP prediction model was established by using MATLAB platform. Using Log-Sigmoid transfer function (Fig. 6), the return value of neural network is between (0,1).

The data of each sample is normalized by using the method of Min-Max Normalization (Eq. 13) to make it fall between (0,1).

$$x_{\text{normalized}} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{13}$$

Trainlm function is used for training, and the MSE target of the AdaBoost-BP model is set as $1e-5$. The superparameters ranges and optimized values of BP model and AdaBoost model are shown in Table 3.

Weak predictor parameters are shown in Table 4.

Taking the corresponding strong predictor for predicting the displacements of the pile tops (Small sample) as an example: $H(x) = 0.1064 \times h(1) + 0.1005 \times h(2) + 0.0989 \times h(3) + 0.0938 \times h(4) + 0.1044 \times h(5) + 0.0987 \times h(6) + 0.0919 \times h(7) + 0.1076 \times h(8) + 0.1073 \times h(9) + 0.0904 \times h(10)$.

The DNN, RF, and Xgboost models all use the Min-Max normalization method (see Eq. 13) to normalize the data of each sample to fall between (0,1).

For DNN model, appropriately increasing the number of hidden layers can effectively reduce the calculation error and improve the prediction accuracy to a certain extent, but it also complicates the calculation process of the model, which increases the training time of the model and may lead to “over fitting” phenomenon³⁸. The range of the number of hidden layers in DNN refers to the conclusion of Mohammad et al.³⁹.

Hyperparameter tuning is carried out using a fivefold cross-check. The superparameters ranges and optimized values of DNN model, RF model and XGboost model are shown in Tables 5, 6, 7.

The BP, AdaBoost-BP, DNN, RF, and Xgboost algorithms are all calculated based on MATLAB R2023b.

Evaluation indexes

There are various evaluation indexes for the model, and each has its own advantages and disadvantages. Among them, the coefficient of determination (R^2 , see Eq. 14) is easy to understand and is widely used in regression

No	Prediction displacements of the pile tops		Prediction displacements of the ground surface around piles	
	Small sample	Large sample	Small sample	Large sample
$h(1)$	0.1064	0.1008	0.0971	0.1014
$h(2)$	0.1005	0.0982	0.1016	0.1005
$h(3)$	0.0989	0.1003	0.1001	0.0983
$h(4)$	0.0938	0.0977	0.1032	0.1057
$h(5)$	0.1044	0.1004	0.0986	0.1001
$h(6)$	0.0987	0.0986	0.0971	0.0998
$h(7)$	0.0919	0.1013	0.1017	0.1015
$h(8)$	0.1076	0.0986	0.0971	0.0978
$h(9)$	0.1073	0.1028	0.0971	0.0976
$h(10)$	0.0904	0.1013	0.1064	0.0974

Table 4. Weak predictor parameter table.

Type	Learning rate	Regularization coefficient	Activation function	Number of hidden layers	Number of nodes in each hidden layer
Hyperparameters	0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.8	0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.8	ReLU, Sigmoid, tanh	3, 4, 5, 6	16, 32, 64, 128
Optimized parameters	0.06	0.05	ReLU	4	64

Table 5. Superparameters and optimized parameters of DNN model.

Type	Number of decision trees	Maximum depth	The minimum number of samples required for internal node splitting	Minimum number of samples required for leaf nodes
Hyperparameters	100, 150, 200, 250, 300, 350	4, 6, 8, 10, 12	2, 4, 6, 8, 10	1, 2, 3
Optimized parameters	150	6	4	1

Table 6. Superparameters and optimized parameters of RF model.

Type	Learning rate	Maximum depth	Min child weight	Subsample	Colsample bytree	Number of parallel tree
Hyperparameters	0.05, 0.1, 0.15, 0.2, 0.25, 0.3	2, 3, 4, 5, 6, 7	2, 4, 6, 8, 10	0.3, 0.4, 0.5, 0.6, 0.7, 0.8	0.2, 0.4, 0.6, 0.8, 1.0	1, 2, 3, 4
Optimized parameters	0.1	5	6	0.5	0.8	1

Table 7. Superparameters and optimized parameters of XGboost model.

analysis. However, outliers can affect the R^2 value, leading to distorted evaluation. Mean absolute error (MAE, see Eq. 15) is insensitive to outliers and is more robust, but MAE is an absolute indicator and needs to be combined with other indicators to evaluate different models. Mean square error (MSE, see Eq. 16) is suitable for scenarios that are sensitive to large errors, but outliers have a significant impact on MSE. Mean absolute percentage error (MAPE, see Eq. 17) represents error in percentage form and is not affected by the data dimension, facilitating comparison between different datasets. However, MAPE's asymmetric penalties for overestimation and underestimation may lead the model to lean towards underestimation. In this paper, the evaluation of the model uses R^2 , MAE, MSE, MAPE to judge the quality of the model. The predicted and experimental values of BP model, AdaBoost-BP model, DNN model, RF model and xGBoost are shown in Tables 8, 9, 10, 11. It can be seen from this Tables that for small sample datasets, the prediction accuracy of the AdaBoost-BP model, RF model, and DNN model is higher than that of the BP model. The prediction accuracy of xGBoost is slightly lower than that of AdaBoost-BP model, DNN model and RF model, but higher than that of BP model. For large sample datasets, the prediction accuracy of various models has improved, but the prediction accuracy of the BP model is lower than that of other models

Pile no	Measured value	Predicted value				
		BP	AdaBoost-BP	DNN	RF	Xgboost
1	46	28	58	55	62	61
2	72	83	83	67	59	48
3	32	26	35	37	34	32
4	87	64	67	72	81	58
5	63	46	69	55	56	59
6	77	87	71	83	62	65
7	53	72	63	64	71	69
8	39	68	41	35	49	42
9	59	56	69	51	52	53
10	61	80	71	67	77	72
11	90	100	88	81	87	96
12	39	55	45	27	47	41
R ²	/	0.16	0.73	0.77	0.61	0.43
MAE	/	15.1	8.2	8.2	10.1	10.7
MSE	/	277.3	90.8	76.5	128.4	188.7
MAPE	/	28.2	13.8	14.6	18.2	16.6

Table 8. Predicted values and measured values of the displacements of the pile tops (Small sample).

Pile no	Measured value	Predicted value				
		BP	AdaBoost-BP	DNN	RF	Xgboost
1	46	49	42	57	57	52
2	72	69	64	70	66	66
3	32	34	33	33	31	27
4	87	85	80	90	78	82
5	63	68	62	62	61	65
6	77	62	78	76	74	79
7	53	66	65	62	67	61
8	39	53	44	41	44	43
9	59	69	64	70	68	66
10	61	69	60	67	62	65
11	90	82	84	90	82	80
12	39	53	44	41	44	48
R ²	/	0.75	0.90	0.91	0.85	0.89
MAE	/	7.8	4.7	3.9	5.9	5.5
MSE	/	81.8	31.6	30.3	49.3	36.5
MAPE	/	14.7	8.2	7.3	10.2	10.3

Table 9. Predicted and measured values of the displacements of the pile tops (Large sample).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \frac{1}{n} \sum_{i=1}^n y_i)^2} \quad (14)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (16)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (17)$$

Observation no	Measured value	Predicted value				
		BP	AdaBoost-BP	DNN	RF	XGboost
1	32	22	34	31	26	31
2	38	28	36	38	31	28
3	35	30	28	29	30	51
4	37	25	42	44	31	39
5	31	47	34	41	38	31
6	40	40	34	43	39	41
7	30	54	41	32	40	35
8	31	34	26	38	35	41
9	37	54	38	44	39	31
R ²	/	-13.03	-1.46	-1.69	-1.89	-4.23
MAE	/	10.7	4.6	4.7	5.2	5.8
MSE	/	165.6	29.0	31.8	34.1	61.8
MAPE	/	32.1	13.6	13.9	15.8	16.9

Table 10. Predicted values and measured values of the displacements of the ground surface around piles (Small sample).

Observation no	Measured value	Predicted value				
		BP	AdaBoost-BP	DNN	RF	XGboost
1	32	33	32	31	31	31
2	38	35	35	36	35	36
3	35	33	34	33	34	34
4	37	34	35	36	34	35
5	31	30	31	30	33	32
6	40	35	36	37	35	35
7	30	32	31	30	30	29
8	31	30	31	31	33	31
9	37	42	38	37	36	36
R ²	/	0.22	0.69	0.82	0.52	0.69
MAE	/	2.5	1.4	1.1	2	1.5
MSE	/	9.1	3.7	2.1	5.7	3.6
MAPE	/	7.1	3.9	3.1	5.6	4.1

Table 11. Predicted and measured values of the displacements of the ground surface around piles (Small sample (Large sample)).

where n is the total number of samples. y_i is the true value of the i -th sample. \hat{y}_i is the predicted value for sample i .

To evaluate whether there are significant differences among the above-mentioned different prediction models, the Wilcoxon signed-rank test was used to assess the large sample prediction set based on SPSS26 software. The original assumption of all tests was that "there was no significant difference in the overall distribution of the two different models." The result of the Wilcoxon symbolic rank test is the p-value. The obtained P-value is compared with the confidence value α (such as 0.05). If the P-value is lower than α , the null hypothesis is rejected. Table 12 shows the test results. It can be found that there is no significant difference among the models at the 0.05 level. At the 0.1 level, BP and AdaBoost-BP, BP and Xgboost, AdaBoost-BP & DNN all show significant differences in predicting the horizontal displacements of the pile tops, which is caused by the structure of the models themselves. Therefore, the adoption of ensemble learning models (such as AdaBoost-BP, Xgboost) or deep learning models is more effective in solving the problem of displacements prediction at the pile tops.

Analysis the mechanism of characteristic variable driven displacement

In order to analyze the mechanism of displacement driven by each characteristic variable, the horizontal displacements of the pile tops and ground surface around piles were predicted based on the AdaBoost-BP model. The measured samples of the horizontal displacement of the pile tops are 512, and the measured samples of the horizontal displacement of the ground surface around piles are 459. A fivefold cross-validation is adopted, and 20% of the samples are randomly selected as the test set. Randomly divide the remaining samples into 5 subsets and traverse these 5 subsets in sequence. Each time, the current subset is used as the validation set, and all the remaining samples are used as the training set for the training and evaluation of the model. The mean square

Type	P-value	
	Horizontal displacements of the pile tops	Horizontal displacements of the ground surface around piles
BP & AdaBoost-BP	0.059	0.763
BP & DNN	1.000	0.730
BP & Random Forest	0.265	0.916
BP & Xgboost	0.069	0.723
AdaBoost-BP & DNN	0.073	0.480
AdaBoost-BP & Random Forest	0.639	0.729
AdaBoost-BP & Xgboost	0.408	0.234
DNN & Random Forest	0.229	1.000
Random Forest & Xgboost	1.000	0.480

Table 12. Wilcoxon signed rank test results.

error (MSE) is used as the loss function. The superparameters ranges and optimized values of AdaBoost-BP model are shown in Table 3. The calculation of a single model takes approximately 20 min.

Figure 7a and b are scatter plots of the horizontal displacements of the pile tops and ground surface around piles, respectively. It can be seen from the figures that the linear fitting between the predicted values and the measured values of the horizontal displacements of the pile tops and ground surface around piles is good, with R^2 values of 0.82 and 0.67 respectively. It indicates that the AdaBoost-BP model is applicable to the prediction of the horizontal displacements of the pile tops and ground surface around piles, and the prediction results have a certain degree of credibility.

Residual analysis was conducted on the predicted and measured values of the horizontal displacements of the pile tops and ground surface around piles. The results are shown in Fig. 8a and b. As can be seen from the figures, the residual distributions of the horizontal displacements are both close to the normal distribution, indicating that the model assumption is reasonable and the prediction error is statistically stable.

Figure 9a and b are the kernel density diagrams of the horizontal displacements of the pile tops and ground surface around piles, respectively. The figure shows that both the horizontal displacements data roughly follow a normal distribution, which is similar to the actual horizontal displacements. The fitting degree between the predicted displacements value and the measured displacements value is both good, but there are still certain errors. The causes of errors are related to factors such as the quality of the sample, the number of samples, the representativeness of the sample, and the values of model parameters.

The SHAP (shapley additive explanations) value ranking of each feature was given through SHAP analysis. The variable weight ranking of AdaBoost-BP model for horizontal displacements of the pile tops is Feature8、Feature9、Feature7、Feature1、Feature3、Feature6、Feature5、Feature4、Feature2 (shown in Fig. 10a). The variable weight ranking of AdaBoost-BP model for horizontal displacements of the ground surface around piles is Feature8、Feature9、Feature7、Feature3、Feature1、Feature6、Feature5、Feature4、Feature2 (as shown in Fig. 10b). This reflects the distance (Feature8) between the center of the bearing platform and the pile (or the monitoring point) is the main influencing factor. The impact of the angle between the line connecting the center of the bearing platform and the pile (or the monitoring point) and the positive X-axis direction (Feature9) ranks second. This is mainly because the number of rows of the piles group in the X and Y directions is different. The resting time (Feature7) ranks third. This reflects that the time effect (such as soil rheology) has a significant impact on the displacements of the pile tops or monitoring points. In soil properties indicator, the influence of moisture content (Feature1), relative density (Feature3) and internal friction angle (Feature6) on the displacements of the pile tops or the ground surface around piles is more significant than other indexes (cohesion force (Feature5), compression modulus (Feature4) and natural weight (Feature2)).

Figure 11 presents the contribution intensity and action direction of all input features to the displacements through the visualization of SHAP values. Each point in the Figure represents a characteristic value of a sample, and the depth of the color corresponds to the size of the characteristic value (the larger the value, the lighter the color; conversely, the darker the value). As can be seen from Fig. 11a, samples with larger or smaller feature values in Feature8 show a significant negative effect. This is because the pile bodies corresponding to these samples are located in the central area of the bearing platform (pile 5 and pile 8 in Fig. 5) or at the corners (pile 1, pile 3, pile 10 and pile 12 in Fig. 5). The pile 5 and pile 8 have relatively small displacements at pile tops due to the strong constraint effect of nearby piles. The pile 1, pile 3, pile 10 and pile 12 are the last four piles to be pressed in, and their displacements at pile tops is also relatively small.

The smaller the feature value of Feature8 in Fig. 11b, the greater its positive contribution to the displacements of ground surface around the piles. It indicates that the closer the distance between the ground surface monitoring points around piles and the center of the bearing platform, the greater the horizontal displacements. The pattern presented by Feature9 is similar to that of Feature8. The larger the feature value of Feature7 in Fig. 11a and b, the greater the positive effect on the displacements. It indicates that the longer the intermittent rest time, the greater the horizontal displacements of the pile tops and ground surface around piles. The smaller the feature value of Feature1 (moisture content) is, the greater the negative effect on displacements will be. It indicates that the lower the moisture content of the soil, the smaller the horizontal displacements. The greater the Feature3 (relative density) is, the denser the soil mass, and the stronger the interlocking and friction between soil particles during soil squeezing, making it more difficult for the soil to be compressed and flow laterally. The larger the internal

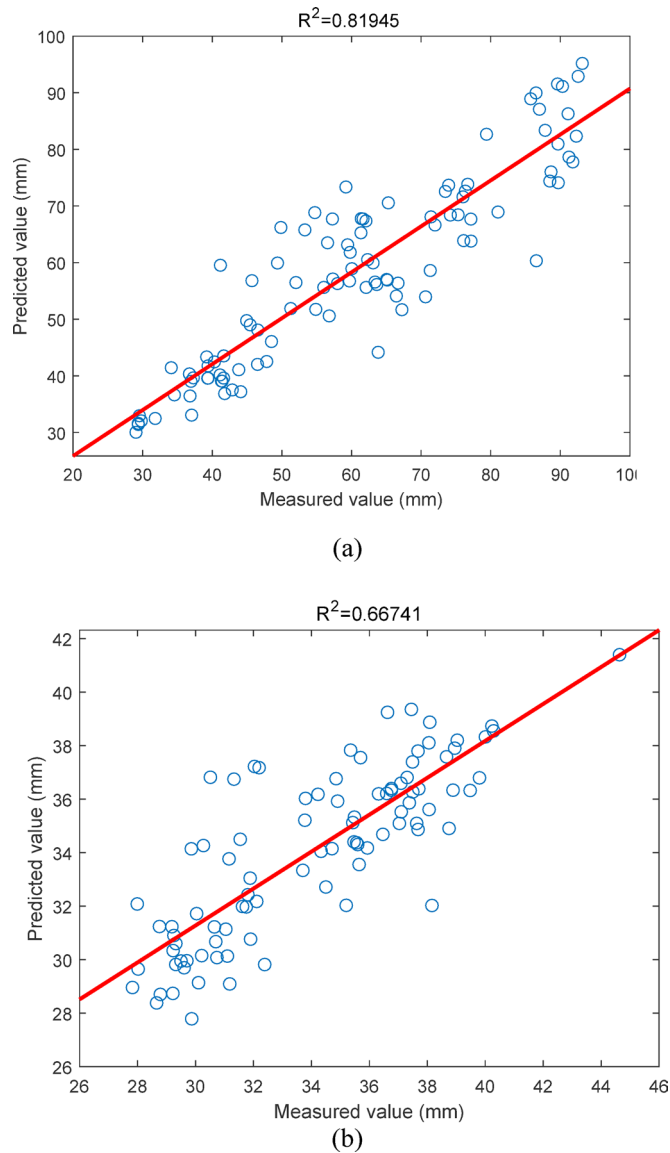


Fig. 7. Scatter plot of predicted and measured values in the test set. **(a)** Horizontal displacements of the pile tops; **(b)** Horizontal displacements of the ground surface around piles.

friction angle (Feature6), the higher the shear strength of the soil, and the less likely the soil is to undergo shear failure during soil squeezing. That is, the smaller the soil deformation. The SHAP values of other features are relatively small, indicating that the influence laws of these features on the displacements are not obvious.

Quantile regression analysis

Quantile regression analysis was conducted with the horizontal displacements of the pile tops and ground surface around piles as dependent variables, and Feature1 to Feature9 as independent variables (large samples were used). The regression results are shown in Tables 13 and 14, respectively. As shown in Table 13, Feature1 has a significant impact at both the 0.1 and 0.9 percentiles ($P < 0.05$), Feature4 and Feature7 have a significant impact at the 0.9 percentiles ($P < 0.05$), Feature8 has a significant impact at the 0.1, 0.25, and 0.9 percentiles ($P < 0.05$), and Feature9 has a significant impact at all percentiles ($P < 0.05$). As shown in Table 14, Feature1 has a significant impact at the 0.1, 0.25, 0.75, and 0.9 quantiles ($P < 0.05$), Feature3 has a significant impact at the 0.9 quantile ($P < 0.05$), Feature7 and Feature8 have a significant impact at all quantiles ($P < 0.05$), and Feature9 has a significant impact at the 0.75 and 0.9 quantiles ($P < 0.05$). From these tables, it can be seen that Feature1, Feature7, Feature8, and Feature9 are the main influencing factors. From the 0.1 percentile to the 0.9 percentile, the regression coefficients of Feature1 and Feature7 show a general trend of increasing from small to large. It can be seen that with the increase of moisture content and interval time, the displacements of the pile tops and ground surface around piles both show an increasing trend. The regression coefficients of Feature8 are all negative, indicating a negative correlation between the horizontal displacements and the distance r (Fig. 5). The regression coefficient of Feature9 does not monotonically increase (or decrease) with the increase of quantiles,

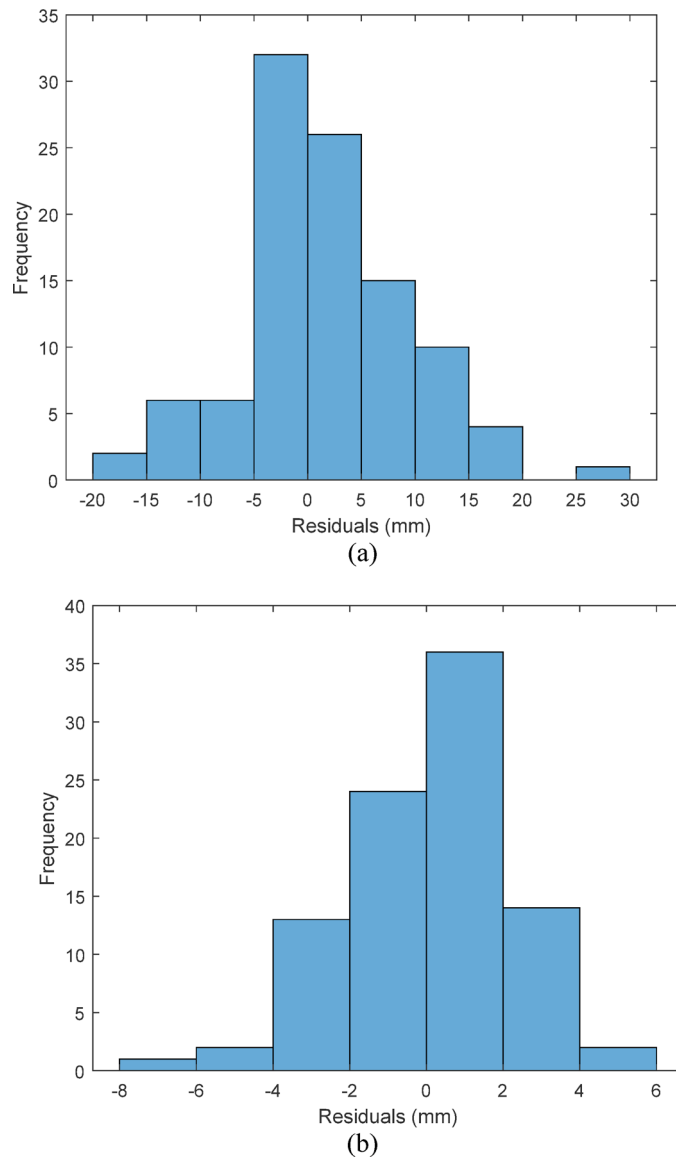


Fig. 8. Residual plot of predicted and measured values in the test set. **(a)** Horizontal displacements of the pile tops; **(b)** Horizontal displacements of the ground surface around piles.

indicating that the relationship between the horizontal displacements and the angle θ (Fig. 5) is complex and needs to be analyzed in conjunction with other indexes.

Calculation of soil displacements by cylindrical hole expansion theory

The theoretical calculation equation for cylindrical hole expansion does not take into account the influence factor of layered soil. Referring to the method proposed by Liu et al.⁷, the average Poisson's ratios, the average elastic modulus, the average cohesion, and the average internal friction angle of soil within the pile length range were obtained by weighted averaging of soil layer thickness. The Poisson's ratio of each layer of soil are referenced in relevant literature^{7,40}. Elastic modulus is calculated according to Formula (2). The value of Δ is 0.015⁷. The parameter values for the group of 12 piles are shown in Table 15. The radial displacements of the monitoring points of ground surface are calculated according to Eq. (1). The calculated values of the cylindrical hole expansion method and the measured values are shown in Table 16. According to Table 16, there is a certain error between the calculated values of the cylindrical hole expansion method and the measured values. As shown in Table 16, the prediction accuracy of machine learning model (such as DNN) is higher than that of the cylindrical hole expansion method.

Conclusion

The main influencing factors were analyzed on the horizontal displacements of the pile tops and ground surface around piles caused by the pile driving. Based on the measured data of the Bogota Metro project and combined

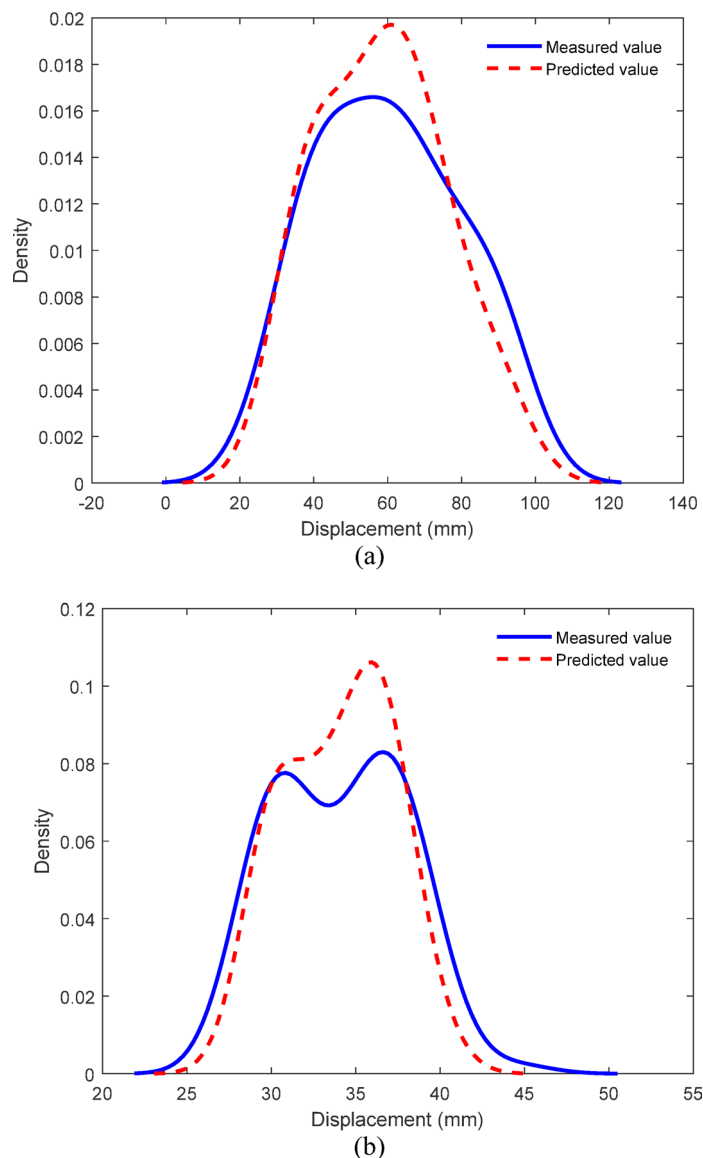


Fig. 9. Comparison plots of kernel density. **(a)** Horizontal displacements of the pile tops; **(b)** Horizontal displacements of the ground surface around piles.

with machine learning algorithms, prediction models for the horizontal displacements of the pile tops and ground surface around the piles was established. The conclusion is as follows:

- (1) The AdaBoost algorithm was applied to the BP neural network model to form the Adaboost-BP model, which improved the learning ability of the BP neural network. Hyperparameter tuning was carried out using a fivefold cross-check. For small sample datasets, the prediction accuracy of AdaBoost-BP model, RF model and DNN model is higher than that of BP model. For large sample datasets, the prediction accuracy of various models has been improved, but the prediction accuracy of the BP model is lower than that of other models.
- (2) SHAP analysis was conducted on 9 input features. Analysis shows that the horizontal distance (Feature8) and angle (Feature9) between the center of the bearing platform and the center of the pile tops (or ground surface monitoring point) are the two most important influencing factors. The resting time (Feature7) is also an important influencing factor. Among soil property indexes, moisture content (Feature1), relative density (Feature 3), and internal friction angle (Feature6) has a more significant influence on the horizontal displacements than other indexes (cohesion (Feature5), compressive modulus (Feature4), and natural weight (Feature2)).
- (3) Quantile regression analysis shows that the horizontal displacement of the piles top and ground surface around piles is negatively correlated with the horizontal distance (Feature8), and positively correlated with the rest time (Feature7) and moisture content (Feature1).

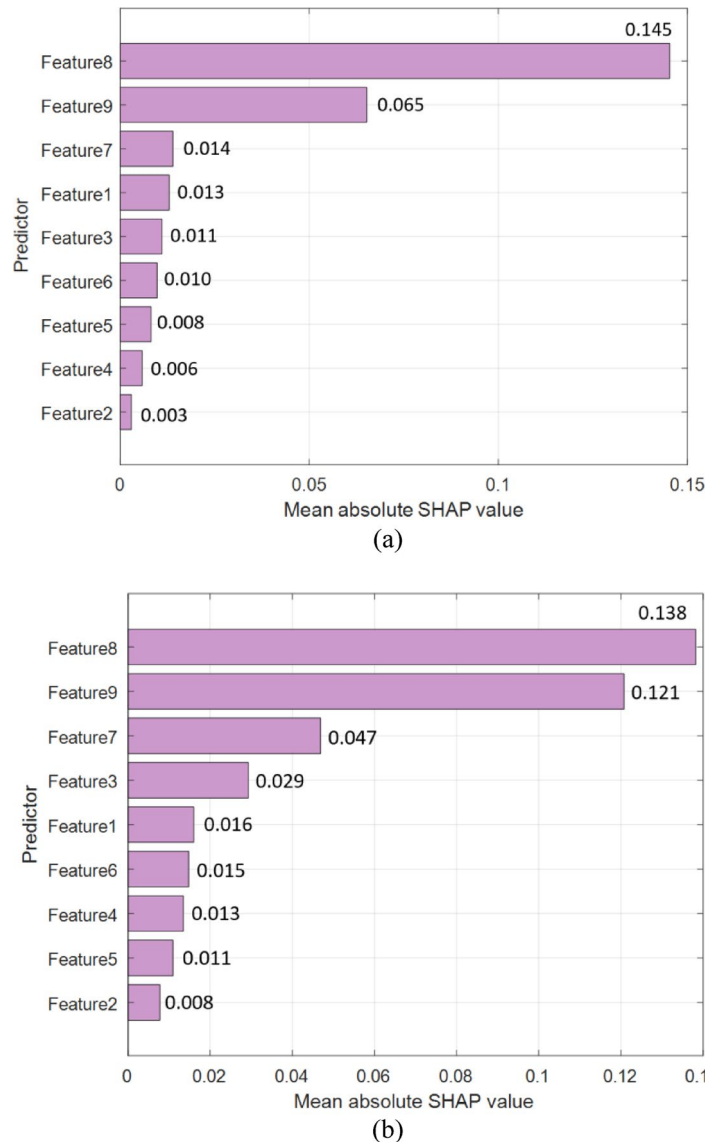


Fig. 10. SHAP bar plot. **(a)** Horizontal displacements of the pile tops; **(b)** Horizontal displacements of the ground surface around piles.

- (4) The horizontal displacement of the ground surface monitoring points was calculated by using the cylindrical hole expansion method. It was found through comparison that the prediction accuracy of machine learning algorithms (such as DNN) is higher than that of the cylindrical hole expansion method.
- (5) The machine learning prediction model established considers the influence of geological conditions (such as soil moisture content, compression modulus), geometric conditions (such as pile number, pile length), and other factors, and can be used to predict the displacement caused by soil squeezing effect of pile groups to be built (such as Bogota Metro Line 2). The method is to collect geological conditions and pile group design schemes during the design phase, and machine learning models are used to predict the pile tops and ground surface around piles. Based on the prediction results, optimize the pile diameter, pile length, pile spacing, and construction sequence to reduce the impact of soil squeezing effect. During the construction phase, collect data on soil and pile deformation caused by soil squeezing effects. Optimize the soil squeezing effect prediction model, improve prediction accuracy, and dynamically adjust construction parameters (such as pile driving speed and sequence) based on the prediction results.
- (6) The number of samples has a significant impact on the prediction accuracy of the model. How to improve the prediction accuracy of machine learning algorithms in small sample situations for projects that are difficult to obtain large samples is a key issue that needs to be addressed in the future.

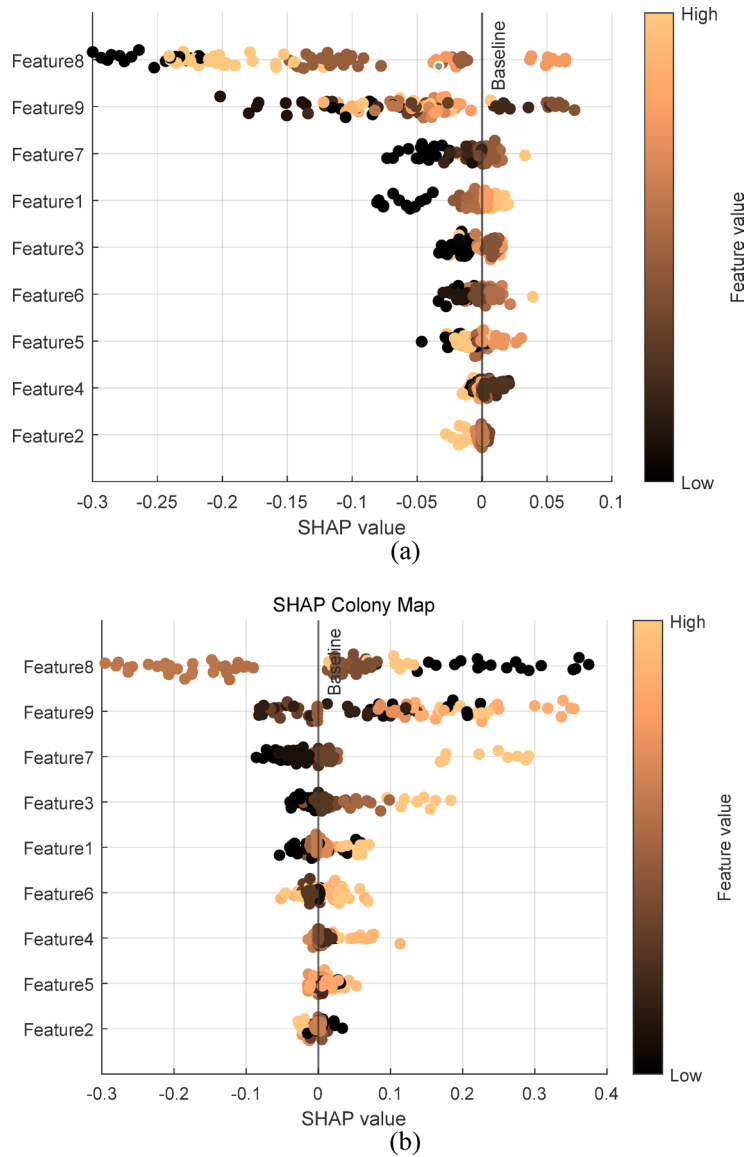


Fig. 11. SHAP analysis plots. (a) Horizontal displacements of the pile tops; (b) Horizontal displacements of the ground surface around piles.

Variables	0.10		0.25		0.5		0.75		0.90	
	RC	P	RC	P	RC	P	RC	P	RC	P
Feature1	0.095	0.017	0.099	0.309	0.126	0.181	0.220	0.069	0.223	0.000
Feature2	-0.043	0.935	-0.500	0.698	-0.328	0.793	0.408	0.799	-0.351	0.330
Feature3	-13.007	0.700	-17.652	0.832	-24.063	0.765	8.837	0.932	-8.169	0.725
Feature4	-0.023	0.931	-0.208	0.746	-0.204	0.744	-0.441	0.581	-0.907	0.000
Feature5	0.025	0.857	0.043	0.898	0.059	0.856	-0.025	0.953	0.085	0.362
Feature6	-0.113	0.789	0.205	0.844	-0.351	0.728	0.510	0.693	-0.220	0.449
Feature7	0.117	0.462	0.205	0.600	0.614	0.105	0.594	0.221	0.710	0.000
Feature8	-1.036	0.017	-3.497	0.001	-0.328	0.750	-1.381	0.295	-7.177	0.000
Feature9	0.019	0.000	-0.029	0.015	-0.038	0.001	-0.048	0.001	0.074	0.000

Table 13. Quantile regression analysis of displacements of pile tops. RC- regression coefficient.

Variables	0.10		0.25		0.5		0.75		0.90	
	RC	P	RC	P	RC	P	RC	P	RC	P
Feature1	0.018	0.021	0.019	0.010	0.026	0.225	0.041	0.000	0.039	0.001
Feature2	-0.176	0.251	-0.107	0.463	0.335	0.418	-0.404	0.056	-0.109	0.624
Feature3	13.313	0.140	-7.951	0.353	-37.914	0.120	-6.944	0.576	-28.107	0.032
Feature4	0.077	0.147	0.043	0.398	0.144	0.316	0.059	0.418	0.113	0.140
Feature5	0.053	0.091	0.034	0.254	-0.038	0.647	0.010	0.810	-0.027	0.550
Feature6	-0.183	0.082	-0.110	0.272	0.042	0.882	0.038	0.794	0.081	0.597
Feature7	0.273	0.000	0.271	0.000	0.230	0.027	0.264	0.000	0.318	0.000
Feature8	-2.477	0.000	-2.277	0.000	-1.347	0.003	-0.514	0.025	-0.771	0.001
Feature9	-0.001	0.790	-0.003	0.077	-0.002	0.740	0.007	0.012	0.013	0.000

Table 14. Quantile regression analysis of displacements of ground surface around piles.

Number of piles in the group	μ	c	φ	Δ	E_s	E
	/	kPa	°	/	kPa	kPa
12	0.35	20.0	23.7	0.015	13.3	3.506

Table 15. Parameter value table.

Sample no	r/m	U_{2D}/mm		
		Measured results	Equation calculation results	DNN
1	7.2	32	57	31
2	6.8	38	61	36
3	7.2	35	57	33
4	7.2	37	57	36
5	6.8	31	61	30
6	7.2	40	57	37
7	6.7	30	62	30
8	5.6	31	74	31
9	5.6	37	74	37
R^2	/	/	-70.97	0.82

Table 16. Table of observation points displacements. r -the distance between the calculated point and the center of the piles group. Relative error = $100 \times (\text{Equation calculation results} - \text{Measured results}) / \text{Measured results}$.

Data availability

The sequence data supporting the results of this study can be obtained from the corresponding author.

Received: 14 May 2025; Accepted: 13 January 2026

Published online: 23 January 2026

References

- Tan, N. & Bengt, H. F. Bidirectional static loading tests on barrette piles. A case history from Ho Chi Minh City, Vietnam. *Canadian Geotechnical Journal* **61**(5), 872–884 (2024).
- Pang, L., Jiang, C., Zeng, F. & Zhang, C. Cyclic response of long flexible piles in sands incorporating the cavity expansion/contraction theory. *Ocean Eng.* **310**, 13 (2024).
- Bellet, M., Keumo, T. J. & Zhang, Y. The inherent strain method for simulation of additive manufacturing—A critical assessment based on a new variant of the method. *Int. J. Numer. Meth. Eng.* **125**(2), 7378 (2024).
- Chen, S. L. & Abousleiman, Y. N. A graphical analysis-based method for undrained cylindrical cavity expansion in modified cam clay soil. *Geotech.: Int. J. Soil Mech.* **73**(8), 736–746 (2022).
- Gao, Z. & Shi, J. Theoretical solutions of soil-squeezing effect due to pile jacking considering geometrical characteristics of a pile. *J. Geotech. Eng.* **32**(6), 956–962 (2010).
- Liu, Y. H., Chen, Z. Z., Peng, Z. J., Gao, Y. S. & Gao, P. Analysis of pile driving effect of precast tubular pile using cylindrical cavity expansion theory. *Rock and Soil Mechanics* **28**(10), 2167–2172 (2007).
- Hight, D. W. & Bishop, A. W. The value of poisson's ratio in saturated soils and rocks stressed under undrained conditions. *Géotechnique* **27**(3), 369–384 (2015).
- Lu, Q., Gong, X. N., Cui, W. W., Zhang, K. P. & Xu, M. H. Finite element analysis of compacting displacements of single jacked pile. *Rock & Soil Mech.* **28**(11), 2426–2430. <https://doi.org/10.1007/s11747-006-0011-3> (2007).

9. Luo, Z., Tao, Y., Gong, X. & Zou, B. Soil compacting displacements for two jacked piles considering shielding effects. *Acta Geotech.* **15**(8), 2367–2377 (2020).
10. Shao, Y., Wang, S. & Guan, Y. Numerical simulation of soil squeezing effects of a jacked pipe pile in soft foundation soil and in foundation soil with an underlying gravel layer. *Geotech. Geological Eng.* **34**(2), 493–499. <https://doi.org/10.1007/s10706-015-9960-y> (2016).
11. Zhou, H. & Shi, J. Test research on soil compacting effect of full scale jacked-in pile in saturated soft clay. *Rock Soil Mech.* **30**(11), 3291–3296. [https://doi.org/10.1016/S1874-8651\(10\)60073-7](https://doi.org/10.1016/S1874-8651(10)60073-7) (2009).
12. Zhang, S. & Zhang, X. Study on influencing factors of soil compaction effect of pipe pile in soft soil area. *Front. Earth Sci.* **12**, 1495866 (2025).
13. Yuan, B. X., Li, Z., Zhao, Z., Ni, H. & Li, Z. Experimental study of displacement field of layered soils surrounding laterally loaded pile based on transparent soil. *J. Soils Sediments* **4**, 1–12 (2021).
14. Yuan, B. X., Chen, R. R., Teng, J., Peng, T. & Feng, Z. W. Effect of passive pile on 3d ground deformation and on active pile response. *Sci. World J.* **2014**, 1–6 (2014).
15. Mustafa, R. & Ahmad, M. T. Reliability analysis of pile foundation in cohesionless soil using machine learning techniques. *Transportation Infrastructure Geotechnology* **11**(4), 2671–2699 (2024).
16. Al-Haddad, L. A., Fattah, M. Y., Al-Soudani, W. H. S., Al-Haddad, S. A. & Jaber, A. A. Enhanced load-settlement curve forecasts for open-ended pipe piles incorporating soil plug constraints using shallow and deep neural net-works. *China Ocean Eng.* **2025**(3), 562–572 (2025).
17. Tran, T. H., Nguyen, B. P. & Tran, T. D. Machine learning applications in pile load capacity prediction: advanced analysis of pile driving forces and depths in urban ho chi minh city construction sites. *Indian Geotech. J.* **55**(3), 1795–1800 (2025).
18. Ren, J. & Sun, X. Prediction of ultimate bearing capacity of pile foundation based on two optimization algorithm models. *Buildings* **13**(5), 2075–5309 (2023).
19. Honarjoo, A. & Ghiasi, V. Analyzing analytical and software methods for deep foundation analysis and presenting a new solution for determining pile capacity using the pda test. *Int. J. Geo-Eng.* **16**, 14 (2025).
20. Kordjazi, A., Nejad, F. P. & Jaksa, M. B. Prediction of ultimate axial load-carrying capacity of piles using a support vector machine based on CPT data. *Comput. Geotech.* **55**, 91–102. <https://doi.org/10.1016/j.compgeo.2013.08.001> (2014).
21. Shahin, M. A. Load-settlement modeling of axially loaded steel driven piles using CPT-based recurrent neural networks. *Soils Found.* **54**(3), 515–522. <https://doi.org/10.1016/j.sandf.2014.04.015> (2014).
22. Moayed, H. & Hayati, S. Applicability of a cpt-based neural network solution in predicting load-settlement responses of bored pile. *Int. J. Geomech.* **18**(6), 1943–1954 (2018).
23. Tan, N., Duy-Khuong, L., Jim, S. & Phi, N. D. Optimizing load-displacement prediction for bored piles with the 3mSOS algorithm and neural networks. *Ocean Eng.* **304**, 117758. <https://doi.org/10.1016/j.oceaneng.2024.117758> (2024).
24. Tan, N., Duy-Khuong, L., Thien, Q. H. & Thanh, T. N. Soft computing for determining base resistance of super-long piles in soft soil A coupled SPBO-XGBoost approach. *Comput. Geotech.* **162**, 105707. <https://doi.org/10.1016/j.compgeo.2023.105707> (2023).
25. Tram, B. N., Duy-Khuong, L., Tan, N. & Nguyen-Thoi, T. Sustainable foundation design, Hybrid TLBO-XGB model with confidence interval enhanced load-displacement prediction for PGPN piles. *Adv. Eng. Inform.* **65**, 103288. <https://doi.org/10.1016/j.aei.2025.103288> (2025).
26. Tram, B. N., Tan, N., Minh-The, N. Q. & Jim, S. Predicting load-displacement of driven PHC pipe piles using stacking ensemble with Pareto optimization. *Eng. Struct.* **316**, 118574. <https://doi.org/10.1016/j.engstruct.2024.118574> (2024).
27. Yuan, B. X. et al. Study on the interaction between pile and soil under lateral load in coral sand. *Geomech. Energy Environ.* **42**, 100674 (2025).
28. Tiwari, M. K. & Chatterjee, C. Uncertainty assessment and en-semble flood forecasting using Bootstrap based Artificial Neural Networks. *J. Hydrol.* **382**, 20–33 (2010).
29. Guo, S. L., Zheng, D. J., Zhao, L. H. & Liu, X. K. ANN-AdaBoost model for the strength-weakening coefficient of soft clay in port engineering. *Sadhana, Academy Proceed. Eng. Sci.* **48**(4), 234. <https://doi.org/10.1007/s12046-023-02276-z> (2023).
30. Lin, E., Lin, C. & Lane, H. Y. Prediction of functional outcomes of schizophrenia with genetic biomarkers using a bagging ensemble machine learning method with feature selection. *Sci. Rep.* **11**, 10179 (2021).
31. Huang, G., Liu, Z., Maaten, L. V. D. & Weinberger, K. Q. Densely connected convolutional networks. *IEEE Comput. Soc.* **7**, 4700–4708. <https://doi.org/10.1109/CVPR.2017.243> (2016).
32. Zhang, C., Lv, W. C., Guo, Z. C., Liu, Y. & Xie, S. C. An optimized combined prediction model for surface subsidence based on GA-KF and BP-Adaboost. *J. Geodesy Geodynamics* **43**(2), 203–208 (2023).
33. He, Q. P., Si, Y. B. & Li S. Y. Settlement prediction of high-speed railway subgrade based on MIDBO-BP-Adaboost. *Journal of beijing jiaotong university* **49**(3) (2025).
34. Wang, Z. Parameter optimization and state evaluation of basketball teaching based on BPNN. *Mobile Inform. Syst.* **2022**, 1. <https://doi.org/10.1155/2022/4327356> (2022).
35. Deng, J., Gu, D., Li, X. & Zhong, Q. Structural reliability analysis for implicit performance functions using artificial neural network. *Struct. Saf.* **27**(1), 25–48. <https://doi.org/10.1016/j.strusafe.2004.03.004> (2014).
36. Murmu, S. et al. Identification of potent phytochemicals against magnaporthe oryzae through machine learning aided-virtual screening and molecular dynamics simulation approach. *Comput. Biol. Med.* **188**, 109862 (2025).
37. Chen, X., Ding, H., Fang, S. & Chen, W. Predicting the success of internet social welfare crowdfunding based on text information. *Appl. Sci.* **12**(3), 1572. <https://doi.org/10.3390/app12031572> (2022).
38. Khandel, O. & Soliman, M. Integrated framework for assessment of time-variant flood fragility of bridges using deep learning neural networks. *J. Infrastruct. Syst.* **27**(1), 1943–1955. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000587](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000587) (2021).
39. Mohammad, A. et al. Hybrid deep neural network optimization with particle swarm and grey wolf algorithms for sunburst attack detection. *Computers* **14**(3), 107–107. <https://doi.org/10.3390/COMPUTERS14030107> (2025).
40. Li, J., Zhang, Y., Chen, H. & Liang, F. Analytical solutions of spherical cavity expansion near a slope due to pile installation. *J. Appl. Math.* **2013**(4), 1–11 (2013).

Author contributions

(Corresponding Author)Shaolong Guo: Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing Original Draft; Penglin Li: Data Curation, Writing—Original Draft; Manman Liang: Visualization, Investigation, Software, Validation; Qun Lu: Visualization, Writing—Review & Editing.

Funding

This work was financially supported by the Major Science and Technology Research and Development Project of China Harbour Engineering Co., Ltd. in 2023 (METRO1-CS-E-230407), and Tianjin Technology Innovation Guidance Special Fund (23YDTPJC00110).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026