Hindawi
Geofluids
Volume 2022, Article ID 1983303, 12 pages
https://doi.org/10.1155/2022/1983303

Research Article

Data-Driven Method for Predicting Soil Pressure of Foot Blades within a Large Underwater Caisson

Can Huang,[1,2] Hao Zhu,[1] Kunyao Li,[1] Jianxin Zheng,[1] Hao Li,[1] Jiaming Li,[3] and Yao Xiao[1]

[1]*Research and Development Center of Transport Industry of Intelligent Manufacturing Technologies of Transport Infrastructure, CCCC Second Harbor Engineering Company Ltd., Wuhan 430000, China*
[2]*CCCC Highway Bridge National Engineering Research Centre Co. Ltd., Beijing 100120, China*
[3]*State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116024, China*

Correspondence should be addressed to Jiaming Li; jiaming_li@mail.dlut.edu.cn

Received 24 October 2021; Revised 21 December 2021; Accepted 4 March 2022; Published 19 March 2022

Academic Editor: Keliu Wu

The soil pressure on the bottom surface of the foot blades is an important monitoring point during the sinking process of large underwater caissons. Complex soil-structure interactions occur during the sinking process, making it difficult to accurately predict the soil pressure of foot blades. Accurate construction processes often rely on data from the soil pressure of foot blades in the field. In this study, a data-driven approach is used to establish the relationship between the amount of sinking of the caisson and the soil pressure of foot blades. Furthermore, by improving the splitting method of the original Classification and Regression Tree (CART) algorithm, a single model's numerical prediction of 80-foot blades soil pressures is realized. The improved CART model, multilayer perceptron (MLP), long short-term memory (LSTM), and a linear regression model are compared through a comprehensive multiparameter evaluation method. Finally, this article discusses the deployment scheme of the model by comparing and analyzing the data in the time period of 10 : 00 on July 29, 2020, and 23 : 00 on August 7, 2020. The experimental results can satisfy the engineering demands and provide a basis for further data-driven intelligent control of large caisson sinking.

1. Introduction

As the main bridge engineering deep foundation, a caisson has the advantages of excellent integrity, high load-bearing capacity, superior structural stiffness, small floor area, and good seismic performance [1–3]. By extracting soil from the well, the caisson uses its gravity and sinking aid to overcome buoyancy and soil resistance in order to sink. The sinking process can be analyzed using the soil pressure of foot blades to calculate the end resistance, thus, providing a basis for instructions during construction [4]. However, with the gradual increase in the number of large caissons being built, the theories developed for small foundations may not predict the soil pressure of large caissons accurately [5].

At present, some achievements have been made in studying caisson foot blades. For example, Jiang et al. [6] found that the sand migration during the sinking of the Hutong Yangtze River Bridge has a relatively significant effect on the soil pressure of foot blades through traditional model experiments. Yan et al. [7] studied the caisson of the Oujiang River North Estuary Bridge as an example to demonstrate the effective reduction of the soil pressure of foot blades by layered excavation during the construction of the caisson. Zhang et al. [8] and Baogang et al. [9] found that the sudden sinking of the caisson during construction was related to the short time decrease of the soil pressure of foot blades. Yea and Kim [10] investigated the three-dimensional distribution pattern of the soil pressure of foot blades during sinking for the caisson of Youngjong Grand Bridge by field experiments. The characteristics of soil pressure of caisson in different engineering backgrounds have been well studied. However, due to complex soil-structure interactions, mechanical properties have significant nonlinearity and plasticity during the sinking of the caisson [11]. Existing descriptive design approaches do not readily

capture the accurate prediction of the soil pressure of foot blades. In recent years, during the construction of a caisson, a series of sensors are usually installed to obtain information on the dynamics of the sinking foot blades of the caisson [12]. However, the data analysis capability is not robust; therefore, it cannot predict the soil pressure state of the caisson's foot blades, which increases the uncertainty of caissons construction.

With the development of artificial intelligence and machine learning technology, data-driven applications are being researched and applied in many fields. Most of the current data-driven bridge engineering applications are focused on bridge health monitoring [13]. For example, a data-driven and computer vision-based approach to automatically identify pitting corrosion [14], crack recognition based on a convolutional neural network [15], and a hybrid artificial neural network-based imperial competitive algorithm used to predict damage of slab-on-girder bridge structures [16]. However, there are minimal studies on data-driven methods for bridge construction processes, especially for large caisson construction. In terms of algorithm analysis, most data-driven bridge engineering-based applications focus on the output of a single target. Prediction of multiclass damage [17] and crack width [18] by algorithms such as convolutional neural networks. Output values such as these have only one category or a single regression objective, making it difficult to solve the current problem of the multilabel regression of the caisson soil pressure of foot blades. There are two main solutions to the multilabel problem. One type of processing is through problem transformation, which focuses on manipulating and processing the dataset in order to transform the multilabel learning problem into one or more single-label problems [19–21]. The other is via algorithmic adaptation methods, by improving existing methods that are needed to be directly suitable for learning multilabel datasets [22–24]. In the caisson sinking dataset, the values of the multiple soil pressures of foot blades are predicted simultaneously, and this kind of multilabel processing during the bridge construction process needs to be further studied. Regarding model evaluation indexes, regression problems are generally evaluated by a single index such as mean square error or a fitting coefficient. However, during multimodel evaluation, it is easy to have inconsistent index tendencies of different models, making it difficult for multiple models to conduct comprehensive and accurate quantitative evaluations. Therefore, the comprehensive judgment of multiple models is conducive to comparing and selecting models during the experimental process.

An algorithm based on an improved Classification and Regression Tree (CART) implemented for multilabel prediction will be investigated in detail within this study. Specifically, the primary contributions of this study are as follows: (1) the performance of neural network methods and improved CART models for multilabel foot blades regression prediction of soil pressure is investigated by comparison. Among them, multilayer perceptron (MLP) and long short-term memory (LSTM) are chosen as typical neural network representatives. (2) A multilabel comprehensive evaluation method is improved, and a model comparison and a parameter optimization are performed through comprehensive evaluation indexes. (3) The results of the impact of the model on different learning approaches during field tests are discussed. Based on the natural advantages of the data-driven approach, the present method can be easily extended to other scenarios of caisson construction after sufficient data is collected.

The framework of this study is shown in Figure 1. The second section will explain the engineering background as well as the data acquisition and preprocessing of the GPS data and the soil pressure of foot blades. The third section will investigate the improved CART algorithm. The fourth section focuses on parameter optimization and model comparison based on the improved multiparameter integrated evaluation index. In the fifth part, the deployment scheme of the model and the field prediction results is discussed.

2. Engineering Background and Data Preparation

2.1. Engineering Background. The main channel bridge of the Changtai Yangtze River Bridge is a double-layer cable-stayed bridge. The upper layer of the bridge is a highway, and the lower layer is an intercity railway and ordinary highway. The two pylons of the main channel bridge adopt a large-scale steel caisson foundation, as shown in Figure 2. The foundation plane of the caisson at pier #5 of the main bridge is round-end, the elevation is stepped, and the width of the step is 9.0 m. The bottom surface of the caisson is 95.0 m long, 57.8 m wide, and the radius of the round end is 28.9 m. The top surface of the caisson is 77.0 m in length, 39.8 m in width, and 19.9 m in radius at the round end. The outer wall of the caisson is 1.8 m thick and 43 m high, the inner wall thickness is 2.0 m, and the height is 64 m. The inner and outer ring partition walls are both 1.4 m thick, the outer ring partition wall is 64 m high, and the inner ring partition wall is 39 m high. The standard size of the inner wellbore is 11 m in length and 11 m in width. The partition and inner shaft wall are inverted by 1.5 m in length and 1.5 m in width. The steel shell structure has 28 compartments. The river section where the steel caisson is located is a tidal section of the lower reaches of the Yangtze River. The tidal level is affected by both the Yangtze River runoff and the tide. The 20-year encounter bridge cross-section vertical average maximum velocity ranges from 1.93 m/s to 2.1 m/s. The dry period vertical average maximum velocity is less than 1.05 m/s. The steel caisson is located on the north side of the main channel area. The topography of the pier is relatively stable. The surface layer of the riverbed is loose silt, with an uneven layer thickness that ranges from 11.6 to 4.8 m thick and has poor engineering properties. Most of the sandy soil layers drilled into pier #5 reveals a sandy gravel cement layer, which is a nonlayered structure and is distributed sporadically, revealing that the depth primarily ranges from -35 to -45 m under the river bed.

The location hydrology and geological conditions of Changtai Yangtze River Bridge are complex, the volume of the caisson is large, and the structure form is particularly novel. The specific characteristics and construction

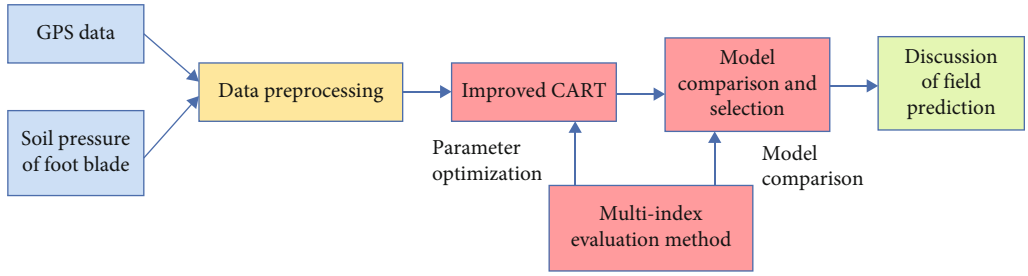

Figure 1: The flow chart of the soil pressure of foot blades prediction.

Figure 2: Caisson under construction.

difficulties are mainly concentrated in that the safety risk and attitude control of initial caisson are difficult. As well as during construction, the caisson has been in a state of dynamic balance, requiring many monitoring items and a large number of measuring points. Traditional monitoring systems have lagging information feedback, which cannot predict the soil pressure of foot blades accurately, thus, increasing the uncertainty of caisson construction.

2.2. Data Preparation. In order to monitor the sinking posture of the caisson and adjust the construction plan in time, Beidou GPS sensors are arranged on the top of the caisson, and the results of the manual high-precision measurement are checked every day to ensure that the monitoring data is accurate and reliable. The installation position of the Beidou control points is shown in Figure 3(a). Figure 3(b) shows the installation positions of the soil pressure of foot blades sensors.

In Figure 3(a), h_1, h_2, h_3, h_4, and h_5 are the measured GPS values, and h_5 is the average value of the four measured points. The sinking amount (SA) is calculated by the following equation:

$$\mathrm{SA}_i = h_{i,t} - h_{i,t-1}, \tag{1}$$

where i is 1-5, representing the sinking amount of 5 locations, $h_{i,t}$ is the vertical value of the i-th monitoring position at time t, and $h_{i,t-1}$ is the monitoring data of the vertical direction at time $t-1$ of the i-th monitoring position.

The sinking data from July 19, 2020, to July 29, 2020, and the soil pressure of foot blades data for the corresponding times formed the data set, as shown in Tables 1 and 2.

The sinking data is recorded every 10 minutes, and the soil pressure of foot blades is recorded every 30 minutes. During the complex construction of the caisson, some of the foot blades sensors are damaged despite the many protections made to protect the sensors; therefore, 80 of them with normal sensors were selected for learning and prediction. Data aggregation of the sinking amount data and the 80 foot blades soil pressure data is based on a 1 h period; thus, the raw dataset was constructed. The raw dataset contains a total of 249 samples and 85 features (5 features for the sinking amount and 80 features for the soil pressure of foot blades). The model was tested on 5 out of the 249 samples from July 29 at 4 : 00 to July 29 at 8 : 00. The remaining 244 were tested for model training and validation. The training and validation sets were split according to 80% (195 samples) and 20% (49 samples), respectively. The model was trained via the training set and the validation set evaluated the model. Mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the fitted coefficient (R^2) were calculated for each model.

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

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|(y_i - \hat{y}_i)|, \tag{3}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|, \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{5}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the true value of the sample, and n is the number of samples.

3. Improved CART Algorithm

A decision tree is an example-based inductive learning approach that constructs a tree-like regression model from the given samples. It is a relatively simple algorithm with excellent robustness compared to other regressions.

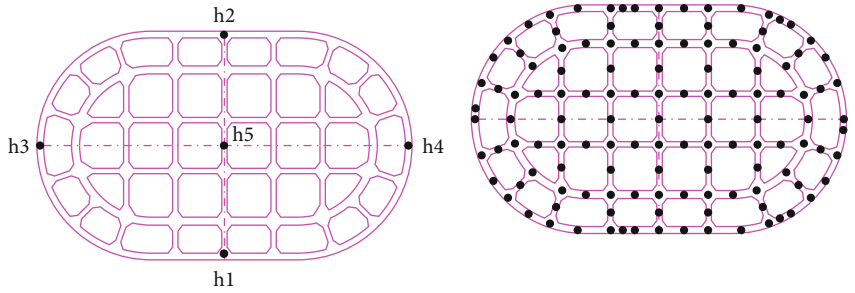

Figure 3: Sensor locations. (a) Beidou GPS sensor locations. (b) Bottom soil pressure of foot blades sensor locations.

Table 1: Sample sinking amount data.

Time	SA-1	SA-2	SA-3	SA-4	SA-5
2020-07-19 00 : 00	5.701	5.413	5.554	5.534	5.568
2020-07-19 00 : 10	5.734	5.429	5.582	5.534	5.632
2020-07-19 00 : 20	5.715	5.386	5.553	5.511	5.601
2020-07-19 00 : 30	5.749	5.442	5.582	5.531	5.605
2020-07-19 00 : 40	5.738	5.435	5.577	5.536	5.599
2020-07-19 00 : 50	5.767	5.456	5.604	5.562	5.633

Table 2: Sample soil pressure of foot blades data.

Time	RF-1	RF-2	RF-16	RF-124	RF-125
2020-07-19 00 : 00	0.100	0.100	0.775	2.968	2.204
2020-07-19 00 : 30	0.100	0.100	0.806	2.964	2.201
2020-07-19 01 : 00	0.100	0.100	0.929	2.959	2.195
2020-07-19 01 : 30	0.100	0.100	0.796	2.953	2.189
2020-07-19 02 : 00	0.100	0.100	0.638	2.941	2.180
2020-07-19 02 : 30	0.100	0.100	1.008	2.936	2.172

Research on the soil pressure of foot blades prediction based on CART [25] is carried out in this study and introduces how to use the decision trees in order to solve the regression problem; lets us define the input matrix x as

$$x = \left(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}\right) = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & x_1^{(4)} & x_1^{(5)} \\ x_2^{(1)} & \ddots & & \vdots & \\ \vdots & & \ddots & & \vdots \\ x_n^{(1)} & & & \ddots & \vdots \\ \vdots & & & & \vdots \\ x_N^{(1)} & x_N^{(2)} & x_N^{(3)} & x_N^{(4)} & x_N^{(5)} \end{pmatrix}, \tag{6}$$

where $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}$ represent the input monitoring SA_1, SA_2, SA_3, SA_4, and SA_5 sinking amount data, also called features, and N represents the number of samples. Each sample strip corresponds to 80 foot blades soil pressure points in the following equation.

$$y_n = \left\{y_{n,1}, \cdots, y_{n,q}, \cdots, y_{n,80}\right\}. \tag{7}$$

Figure 4: Training methods.

It is not possible to directly output 80 foot blades soil pressure values with conventional CART. For the multiobjective regression problem, there are two general solutions: the first one is to convert the multiple regression problem into a single-objective regression, such that if 80 foot blades soil pressure values need to be predicted, then 80 single regression models are constructed, and the soil pressure values at different locations are output. However, this method is too cumbersome, and the training complexity is large and time consuming. The alternative way is to build a multiobjective CART regression model [26]:

$$\min_{p,a}\left[\min_{c_1}\sum_{x_n\in R_1(p,a)}(y_n - c_1)^2 + \min_{c_2}\sum_{x_n\in R_2(p,a)}(y_n - c_2)^2\right], \tag{8}$$

$$\min_{p,a}\left[\min_{c_1}\sum_{x_n\in R_1(p,a)}\left(y_{n,q} - c_{1,q}\right)^2 + \min_{c_2}\sum_{x_n\in R_2(p,a)}\left(y_{n,q} - c_{2,q}\right)^2\right]. \tag{9}$$

Equation (8) is the traditional CART regression node division method, where yn is the target value, $c1$ and $c2$ are the predicted values within the intervals of $R1$ and $R2$, respectively. The minimum point a is chosen as the splitting point by calculating the mean square error of $R1$ and $R2$. The target values in this method are not multidimensional, and the splitting process cannot calculate the loss under the multidimensional data, so the traditional CART cannot be carried out for multiobjective regression. With the improved splitting methods of Eq. (8), Eq. (9) integrates

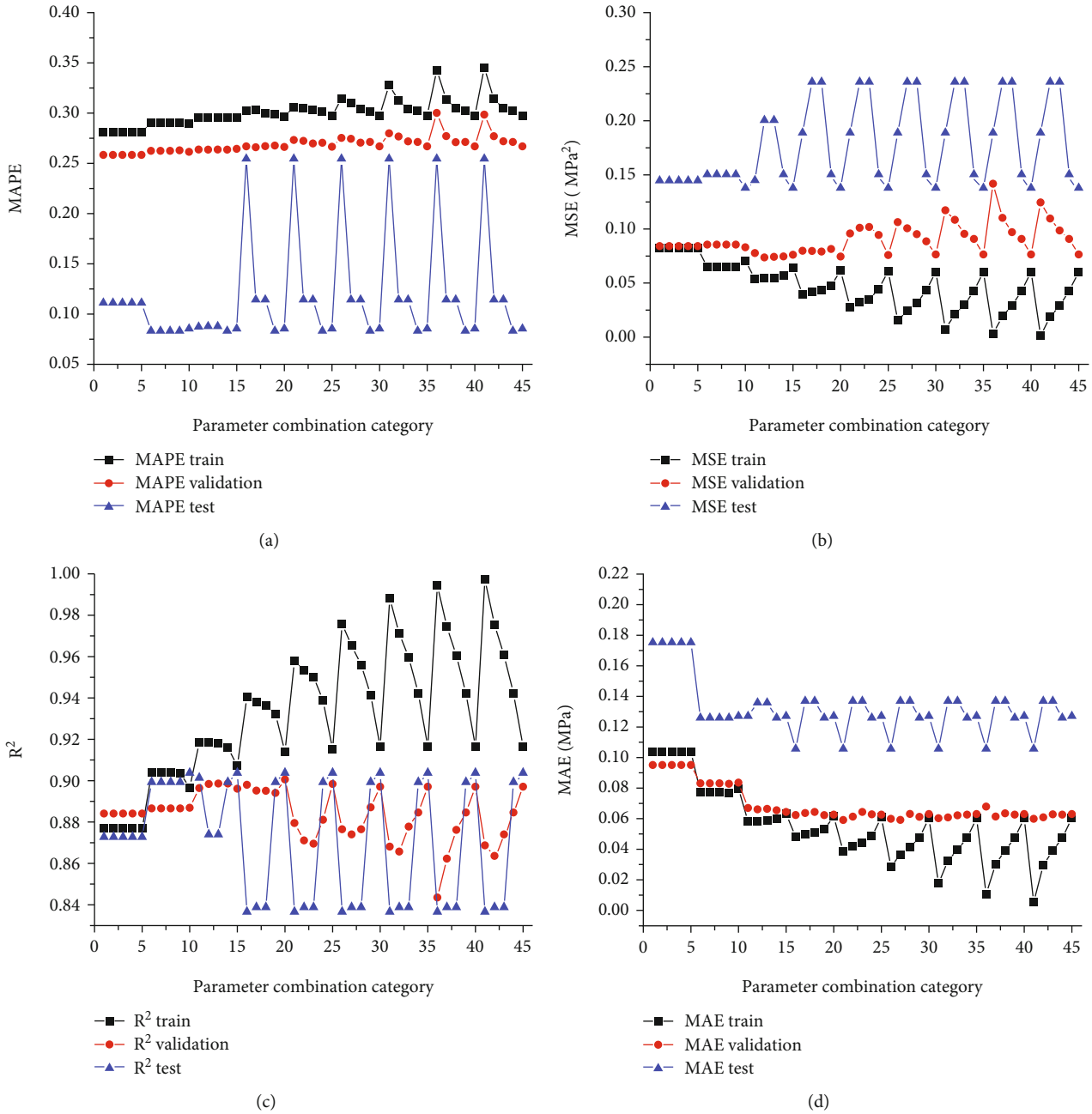

Figure 5: Continued.

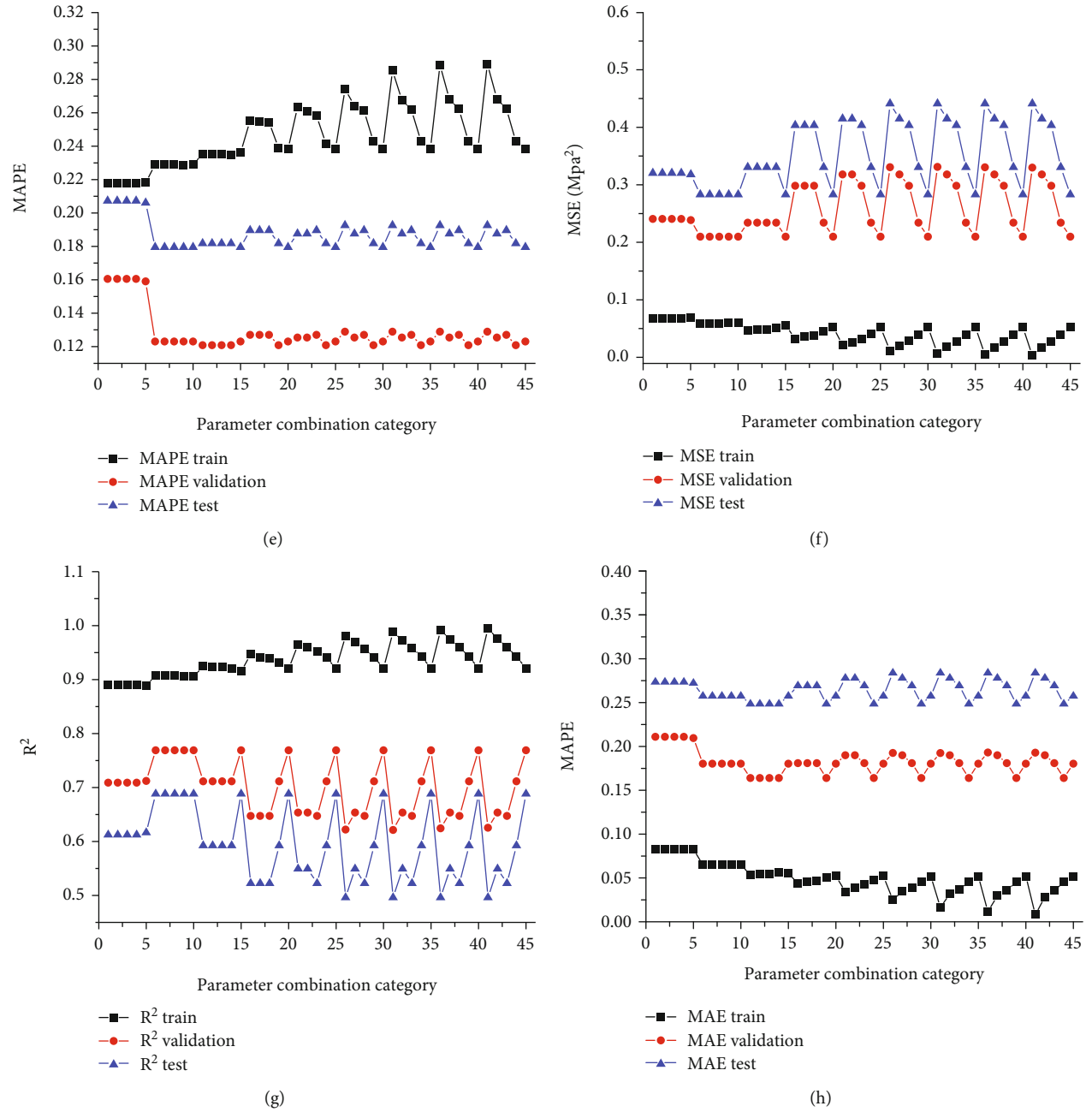

Figure 5: Results of different training methods. (a) Disordered training method MAPE value. (b) Disordered training method MSE value. (c) Disordered training method R^2 value. (d) Disordered training method MAE value. (e) Ordered training method MAPE value. (f) Ordered training method MSE value. (g) Ordered training method R^2 value. (h) Ordered training method MAE value.

the minimum loss of the q-dimensional output value as the splitting point when calculating the loss. The leaf nodes of CART can be written explicitly as

$$\hat{c}_1 = \frac{1}{N_m} \sum_{x_n \in R_1(p,a)} y_n \text{ and } \hat{c}_2 = \frac{1}{N_m} \sum_{x_n \in R_2(p,a)} y_n, \tag{10}$$

$$\hat{c}_{1,q} = \frac{1}{N_m} \sum_{x_n \in R_1(p,a)} y_{n,q} \text{ and } \hat{c}_{2,q} = \frac{1}{N_m} \sum_{x_n \in R_2(p,a)} y_{n,q}. \tag{11}$$

Equation (10) is the output form of the traditional model that calculates the average value of the $R1$ and $R2$ regions as the output of the target values. Improving the above method in order to obtain Eq. (11), the average value of the output target in q dimensions was calculated as the output of multiple targets within the $R1$ and $R2$ regions, respectively.

4. Analysis of the Experimental Results

4.1. Improved CART Training Methods. We split the data according to the temporal order for the prediction of the soil

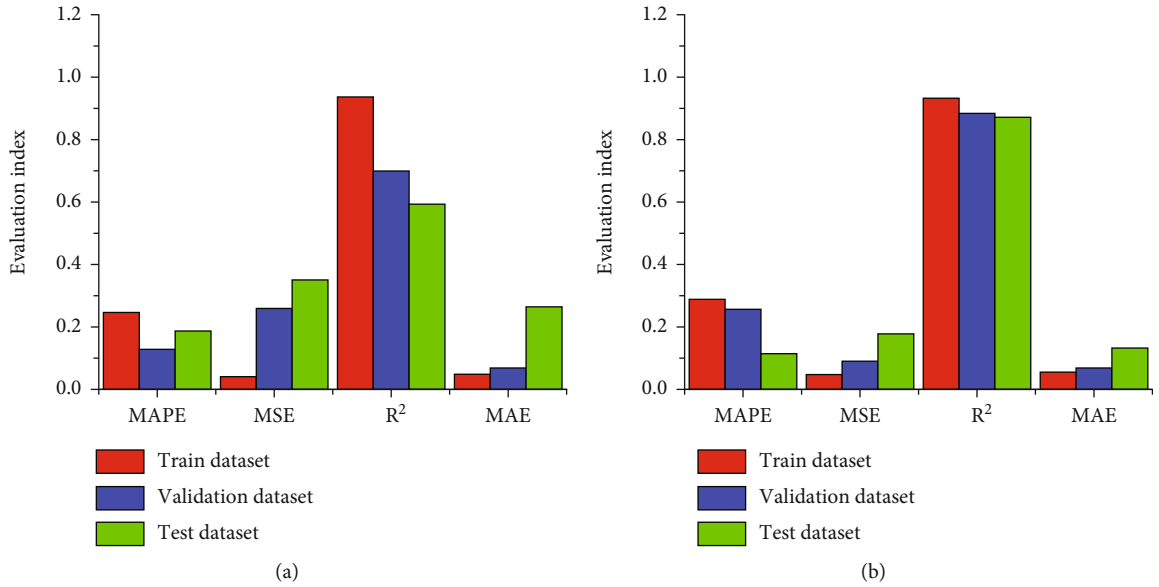

Figure 6: The average value of the index under different training methods. (a) Ordered training of the mean value of each index. (b) Disordered training of the mean value of each index.

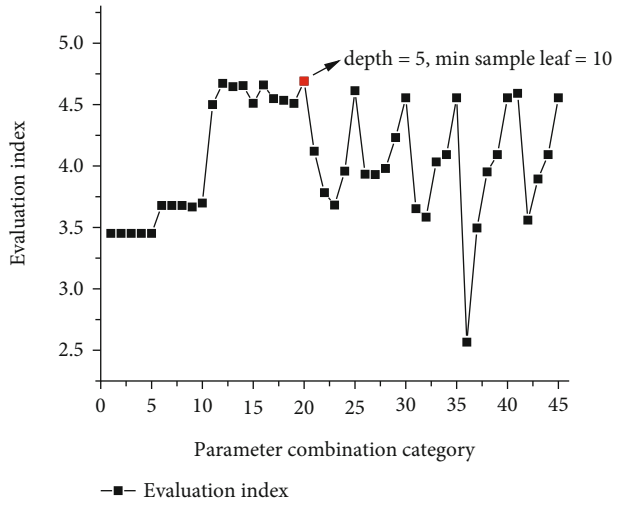

Figure 7: Integrated evaluation index.

pressure of foot blades, where the training was performed separately according to whether the data set was shuffled or not. Examples are shown in Figure 4(a) below, blue 1-10 are the 10 samples on the timeline, and red 1-5 are the test samples. We used two training methods, the first one is shown in Figure 4(b), according to the time order, the first 80% of the data is used as the training set (data samples 1-8 in the yellow area) and 20% as the validation set (samples 9-10 in the green area). Alternatively, as shown in Figure 4(c) ignoring, the temporal order, the overall random selection divides the training and validation sets, i.e., 1-10 in the original data are shuffled, 80% of the data is randomly selected as the training set and 20% as the validation set.

A grid search method is used to select the parameters for the maximum depth of the decision tree and the minimum number of leaf node samples, where the maximum depth is 2, 3, 4, 5, 6, 7, 8, 9, and 10, and the minimum number of leaf node samples is 1, 2, 3, 5, and 10. There are 45 combinations of the maximum depth and the minimum number of leaf node samples, such as <maximum depth is 2 and the minimum number of leaf node samples is 1>, and <maximum depth is 2 and the minimum number of leaf node samples is 2>. In Figure 5, (a), (b), (c), and (d) are the results of training method 2 (disordered method) in Figure 5, and (e), (f), (g), and (h) are the results of training method 1 (ordered method) in Figure 5, respectively.

As the depth increases, the nonlinear representation of the tree model is gradually strengthened. The metrics of both ordered and disordered training on the training set show a step-up trend. The accuracy of the ordered training, however, gradually decreases on the validation and test sets. For the regression coefficient, for example, when the depth is 2, and the minimum number of leaf node samples 1, the fitting coefficient of the training set is 0.8900, and when the depth is 10, and the minimum number of leaf node samples is 1, the fitting coefficient of the training set is 0.9947. The accuracy improved by 11.76%; however, the validation and test sets decreased by 37.16%, and the accuracy of the test set decreased by 50.17%. In the disordered training, the fit coefficient of the training set is 0.8768 when the depth is 2, and the number of minimum leaf samples is 1. It increased to 0.9973 when the depth is increased to 10, while the maximum decrease in the fit coefficient of the validation set and the test set is approximately 16%. As the depth increases, the performance of the training set gradually increases, and the performance of both the validation and test sets decreases, especially within the ordered training.

Increases in the minimum number of leaf node samples at the same depth show an improvement in the robustness of the model, in the model at a depth of 10, the number of minimum leaf samples in the validation set ranges from 1 to 10 in the disordered training, and the fit coefficient increases

Table 3: Cross-validation results.

Validation set / Model	MAPE	MSE (MPa^2)	R^2	MAE (MPa)	Test set MAPE	Test set MSE (MPa^2)	Test set R^2	Test set MAE (MPa)
MLP	0.1127	0.0959	0.8545	0.1273	0.1852	0.2186	0.8634	0.2584
LSTM	0.1988	0.1482	0.7754	0.2109	0.6275	0.9932	-4.3973	0.6421
Linear regression	0.0900	0.0868	0.8764	0.0931	0.9598	0.2793	0.8509	0.3126
Improved CART	0.0487	0.0487	0.8862	0.0671	0.1114	0.1516	0.8632	0.1658

Table 4: Multiparameter model evaluation results.

Model	Validation set $\overline{MAPE}$	Validation set $\overline{MSE}$	Validation set $\overline{R^2}$	Validation set $\overline{MAE}$	Test set $\overline{MAPE}$	Test set $\overline{MSE}$	Test set $\overline{R^2}$	Test set $\overline{MAE}$	Integrated evaluation index value	Ranking
BP	0.2479	0.2669	0.7139	0.3064	0.5492	0.6383	1.0000	0.5169	4.2395	2
LSTM	0.0000	0.0000	0.0000	0.0000	0.0701	0.0000	0.0000	0.0000	0.0701	4
Linear regression	0.3922	0.3462	0.9116	0.5904	0.0000	0.4604	0.9976	0.3669	4.0653	3
Improved CART	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	8.0000	1

Figure 8: Prediction scheme.

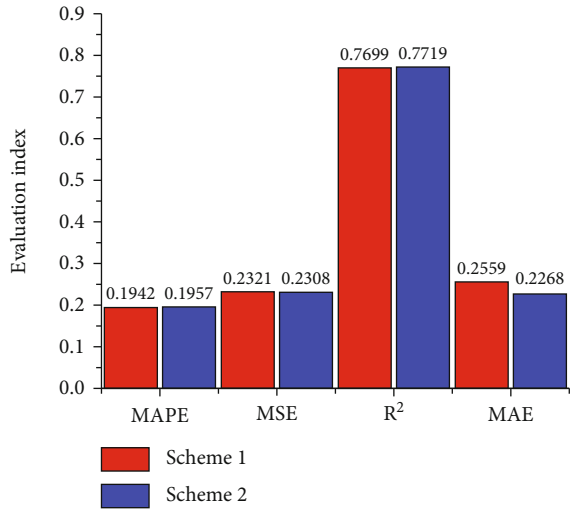

Figure 9: Results of different forecasting schemes.

from 0.8687 to 0.9038. It indicates that in the depth determined case, the accuracy of the validation model and test sets can be improved by increasing the number of minimum leaf samples.

In the ordered training method, since the data set is split according to the temporal order, the construction conditions change more significantly over time, and the data learning in the preceding time period may not be sufficient to support the later predictions. Prediction ability is significantly reduced in the training, validation, and test sets in Figure 6(a), realizing that MSE, R^2, and MAE indicators show the best results in the training set, followed by the validation set, and the worst was observed in the test set. The gap between the training and test set indicators is relatively large; in terms of the fit coefficient, the test set is reduced by 36.65% compared to the training set. In Figure 6(b), the average value of the disorder training is shown. Because the model learns the data characteristics of extended time periods and multiple working conditions during training, the model realizes a better robustness. The average fitting coefficient decreases from 0.9325 in the training set to 0.8716, which is 6.53% lower, and the stability of the model is better compared with the ordered training method.

In the evaluation metrics, the closer the R^2 is to 1, the closer the MSE, MAE, and MAPE are to 0, the better the model performance is. Different indicators do not have the same tendency to evaluate the predictive capability of the model. Furthermore, in this study, the MSE, MAE, and MAPE are transformed as follows:

$$\mathrm{MSE}' = \frac{1}{100 \times \mathrm{MSE}}, \tag{12}$$

$$\mathrm{MAE}' = \frac{1}{100 \times \mathrm{MAE}}, \tag{13}$$

$$\mathrm{MAPE}' = \frac{1}{100 \times \mathrm{MAPE}}. \tag{14}$$

In the regression models, different evaluation indicators are focused on different aspects. The evaluation of different models via a single indicator lacks comprehensiveness. Accordingly, a comprehensive evaluation of different models with multiple indicators is needed. The ranking method proposed by Zorlu et al. [27] in 2008 is a commonly used multi-index comprehensive evaluation method. Zhang et al. [28] carried out optimization in terms of the tendency uniformity and normalization based on Zorlu. However, there is no comprehensive consideration of the different data sets, and the following steps further improve the above study.

Step 1. Calculate the evaluation indexes of MSE, MAE, MAPE, and R^2 for the training set and validation set, respectively.

Step 2. MSE′, MAE′, and MAPE′ by tendency uniform conversion of MSE, MAE, and MAPE.

Step 3. MSE′, MAE′, MAPE′, and R^2 are normalized by the formula:

$\bar{I}_M = I_M - \max(I)/\max(I) - \min(I)$ to get $\overline{\mathrm{MSE}}$, $\overline{\mathrm{MAE}}$, $\overline{\mathrm{MAPE}}$, and $\overline{R^2}$.

Step 4. $\overline{\mathrm{MSE}}$, $\overline{\mathrm{MAE}}$, $\overline{\mathrm{MAPE}}$, and $\overline{R^2}$ of the training and validation set are summed to get the comprehensive evaluation index.

The evaluation metrics for different parameter combinations of the disordered training under the training and validation sets are calculated according to the above comprehensive evaluation algorithm to obtain Figure 7, showing that the model has the highest comprehensive evaluation metrics at a depth of 5 and a minimum number of leaf node samples of 10.

4.2. Comparison of Model Performances. A comparison of the test results of MLP [29], LSTM [30], and the linear regression [31] models are analyzed in this study. The 5-fold cross-validation method is used to divide the sample into 5 equal parts according to 20%, and 4 parts are taken for training and the other 1 part for validation. The average of the five results is used as the evaluation result of the model validation set. The test set evaluation results are obtained by the last five hours of data.

Average values of the 5-fold cross-validation of the MLP, LSTM, linear regression model, and the improved CART model are shown in Table 3. The linear regression model, which requires the modeling of each soil pressure of foot blade feature; thus, constructing 80 linear regression models to meet the engineering requirements, is more complex and computationally intensive compared to the improved CART model. Table 4 is obtained through tendency transformation, and after normalization, the multiparameter model

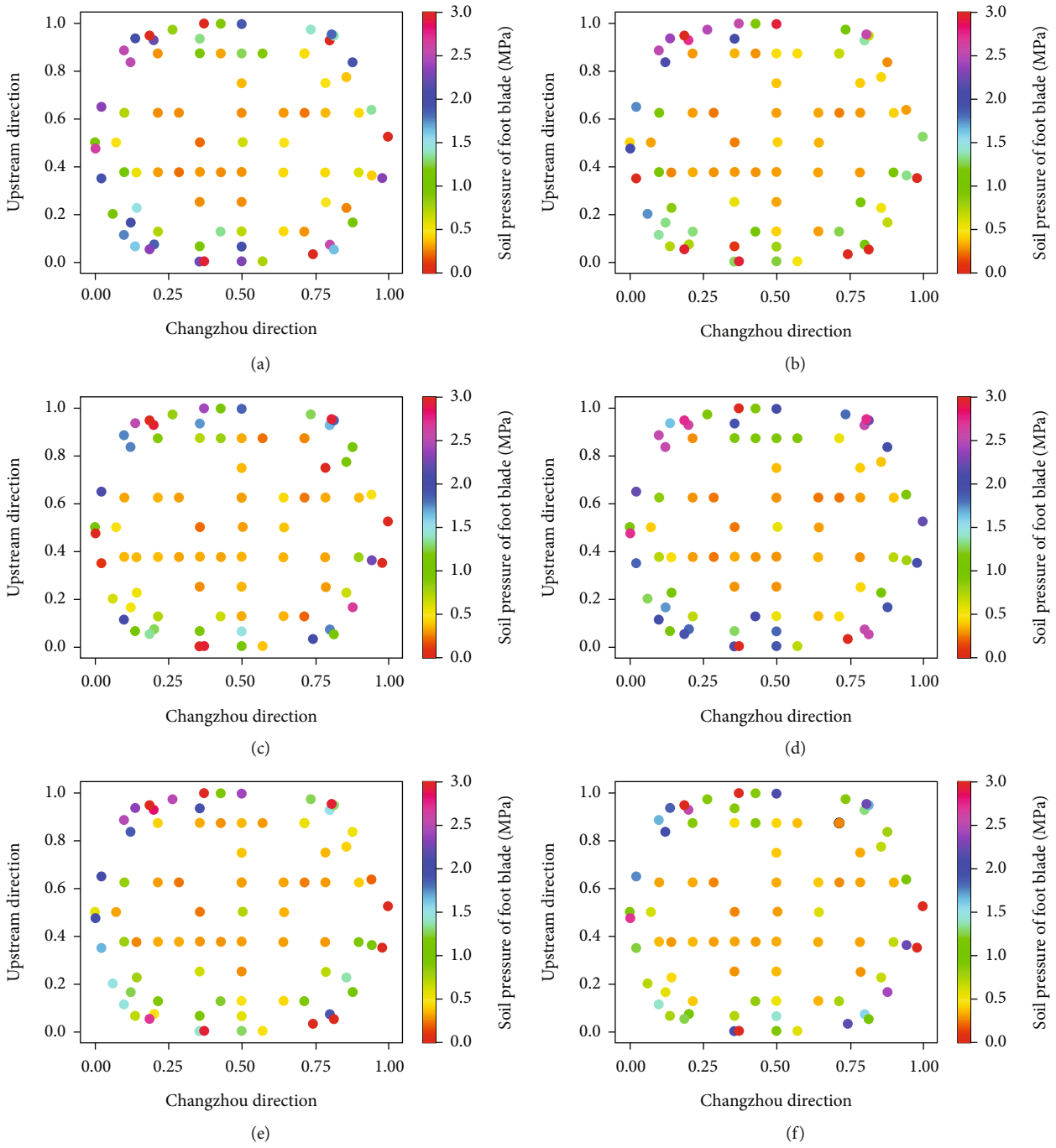

Figure 10: Real and predicted values of the model at different moments in time. (a) July 30, 2020 20 : 00 real value; (b) August 3, 2020 20 : 00 real value; (c) August 7, 2020 18 : 00 real value; (d) July 30, 2020 20 : 00 predicted value; (e) August 3, 2020 20 : 00 predicted value; (f) August 7, 2020 18 : 00 predicted value.

evaluation results are shown in Table 3. Various index results of the validation and the test sets were added to combine and evaluate the robustness of the model simultaneously. Finally, the results were ranked. Based on the evaluation results of the multi-index model in Table 4, it is clear that the improved CART model performs the best in all indexes compared with other models in the validation and test sets in terms of the soil pressure of foot blades prediction it was ranked first overall.

5. Discussion

Previous analytical studies confirm the forecasting schemes and the forecasting algorithms, but whether the model has

sufficient capacity for field forecasting is subject to further discussion. A total of 230 data points from July 29, 2020, 10 : 00 to August 7, 2020, 23 : 00, are used in the discussion as the evaluation data set for the long-term prediction of the model. Due to the differences in the sinking conditions of the caissons, two schemes were selected for the field prediction. In Figure 8(a) scheme one, forgetting is used for the first scheme, where the model learns 43 hours of data, predicts the next five hours of data, and updates the model every five hours. Figure 8(b) scheme two, continuous accumulation of the data set, each time is predicting the next five hours of data, and then updating the model once.

The prediction results of schemes 1 and 2 are shown in Figure 9. It is found that there is no significant difference between the two schemes with a fitting coefficient of 0.7699 for scheme 1 and 0.7719 for scheme 2, under the detection of 230 data points. It shows that in the field forecasting process, it is only necessary to update the model at regular intervals in order to achieve a relatively good forecasting effect. We extracted the true and predicted values of the data at 20 : 00 on July 30, 2020, 20 : 00 on August 3, 2020, and 18 : 00 on August 7, 2020, for visualization. The model's predicted values for the soil pressure of foot blades match the real values, as shown in Figure 10.

6. Conclusion

During the sinking of large caissons, monitoring the soil pressure of foot blades plays a key role during its bridge engineering-based applications safe and smooth sinking. In this study, research based on the data-driven prediction of the soil pressure of foot blades during sinking was developed using the Changtai Yangtze River Bridge caisson as an example, several conclusions were reached as follows:

(1) The multilabel task of a single model for the soil pressure of foot blades of a caisson was achieved by improving the splitting rule of CART. A multiparameter model evaluation algorithm was implemented to select parameters for the maximum depth and a minimum number of samples of the minimum leaf node of the improved CART. The optimal combination of parameters with a maximum depth of 5 and a minimum number of leaf node samples of 10 was selected. In the test set, MAPE is 0.1114, MSE is 0.1516 MPa^2, R^2 is 0.8632, and MAE is 0.1658 MPa

(2) The improved multiparameter model evaluation algorithm compared and analyzed the improved CART, MLP, LSTM, and the linear regression models. It is concluded that the CART model is more suitable for predicting the soil pressure of foot blades during the sinking of the caisson

(3) A total of 230 samples from July 29, 2020, 10 : 00 to August 7, 2020, 23 : 00, were used to continue the evaluation of the improved CART model. Comparison of the two prediction schemes for whether the data were forgotten or not revealed no significant differences, with the average fit coefficient being approximately 0.77. The results of the experiments can be satisfied with the engineering requirements

(4) A shift from empirical decision making to a data-driven based approach needs to be further investigated. Data-driven predictions of the soil pressure of foot blades are part of the overall intelligent construction of the caisson

Data Availability

The data in this study are available from the corresponding author upon reasonable request.

Additional Points

Computer Code Availability. The code in this study is available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Kunyao Li and Jiaming Li wrote the manuscript. Can Huang provided suggestions for the research method and revised the article completely. Hao Li and Yao Xiao calculated and analyzed data. Hao Zhu, Jianxin Zheng, Hao Li, and Yao Xiao processed the basic data.

Acknowledgments

This research was funded by the National Development Plan (2021YFF0500900 and 2021YFF0500902).

References

[1] F. E. Lefler and V. D. R. Romero, "Reinforced concrete caissons for port structures in Spain," *Proceedings of the Institution of Civil Engineers: Maritime Engineering*, vol. 162, no. 2, pp. 73–81, 2009.

[2] S. Chakrabarti, P. Chakrabarti, and M. Krishna, "Design, construction, and installation of a floating caisson used as a bridge pier," *Journal of Waterway, Port, Coastal, and Ocean Engineering*, vol. 132, no. 3, pp. 143–156, 2006.

[3] T. Matsuda, K. Maeda, M. Miyake, J. Miyamoto, H. Sumida, and K. Tsurugasaki, "Instability of a caisson-type breakwater induced by an earthquake-tsunami event," *International Journal of Geomechanics*, vol. 16, no. 5, article C4016003, 2016.

[4] M. Guo, X. Dong, S. Wang, and G. Lan, "Study on the change law of resistance at the sinking end of soil extraction of super-large caisson foundation," *Journal of Rock Mechanics and Engineering*, vol. 40, no. S1, pp. 2976–2985, 2021.

[5] M. Guo, X. Dong, and J. Li, "Study on the earth pressure during sinking stage of super large caisson foundation," *Applied Sciences*, vol. 11, no. 21, article 10488, 2021.

[6] B. N. Jiang, M. T. Wang, T. Chen, L. L. Zhang, and J. L. Ma, "Experimental study on the migration regularity of sand outside a large, deep- water, open caisson during sinking," *Ocean Engineering*, vol. 193, p. 106601, 2019.

[7] X. Yan, W. Zhan, Z. Hu, D. Xiao, Y. Yu, and J. Wang, "Field study on deformation and stress characteristics of large open caisson during excavation in deep marine soft clay," *Advances in Civil Engineering*, vol. 2021, Article ID 7656068, 2021.

[8] D. Y. Zhang Zhicheng, F. Zheng, and J. Wang, "Analysis on sudden sinking behaviors of massive open caisson in deep-thick soft clay area," *Chinese Journal of Underground Space and Engineering*, vol. 16, no. 3, pp. 933–943, 2020.

[9] M. U. Baogang, W. Yan, Z. H. U. Jianmin, W. Gong, and O. Zuliang, "Analysis of large caisson sinking measured resistance," *Journal of Civil and Environmental Engineering*, vol. 34, no. S1, pp. 107–115, 2012.

[10] G. G. Yea and T. H. Kim, "Vertical cutting edge forces measured during the sinking of pneumatic caisson," *Marine Georesources & Geotechnology*, vol. 30, no. 2, pp. 103–121, 2012.

[11] K. Sobhan, "Principles of Geotechnical Engineering: Eighth Edition," in *Pws*, pp. 1–13, Cengage Learning, USA, 2012.

[12] R. Royston, B. B. Sheil, and B. W. Byrne, "Monitoring the construction of a large-diameter caisson in sand," *Proceedings of the Institution of Civil Engineers: Geotechnical Engineering*, vol. 12, pp. 1–48, 2020.

[13] X. H. Zhou and X. G. Zhang, "Thoughts on the development of bridge technology in China," *Engineering*, vol. 5, no. 6, pp. 1120–1130, 2019.

[14] N. D. Hoang, "Image processing-based pitting corrosion detection using metaheuristic optimized multilevel image thresholding and machine-learning approaches," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6765274, 2020.

[15] X. Kong and J. Li, "Vision-based fatigue crack detection of steel structures using video feature tracking," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 783–799, 2018.

[16] M. Gordan, H. A. Razak, Z. B. Ismail, K. Ghaedi, Z. X. Tan, and H. H. Ghayeb, "A hybrid ANN-based imperial competitive algorithm methodology for structural damage identification of slab-on-girder bridge using data mining," *Applied Soft Computing*, vol. 88, p. 106013, 2020.

[17] S. Sony, S. Gamage, A. Sadhu, and J. Samarabandu, "Multiclass damage identification in a full-scale bridge using optimally tuned one-dimensional convolutional neural network," *Journal of Computing in Civil Engineering*, vol. 36, no. 2, article 04021035, 2022.

[18] K. Yang, Y. Ding, P. Sun, H. Jiang, and Z. Wang, "Computer vision-based crack width identification using F-CNN model and pixel nonlinear calibration," *Structure and Infrastructure Engineering*, pp. 1–12, 2021.

[19] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.

[20] E. Hüllermeier, J. Fürnkranz, W. Cheng, and K. Brinker, "Label ranking by learning pairwise preferences," *Artificial Intelligence*, vol. 172, no. 16-17, pp. 1897–1916, 2008.

[21] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-labelsets for multilabel classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 7, pp. 1079–1089, 2011.

[22] A. Clare and R. D. King, *Knowledge Discovery in Multi-Label Phenotype Data*, Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery, 2001.

[23] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in pp. 681–687, Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, 2001.

[24] M. L. Zhang and Z. H. Zhou, "ML-KNN: a lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[25] X. Liang, L. Yuetian, X. Yifei, L. Yanli, C. Xuehui, and L. Gang, "A data-driven shale gas production forecasting method based on the multi-objective random forest regression," *Journal of Petroleum Science and Engineering*, vol. 196, article 107801, 2021.

[26] D. Kocev, C. Vens, J. Struyf, and S. Dzeroski, "Ensembles of multi-objective decision trees," *European Conference on Machine Learning*, pp. 624–631, 2007.

[27] K. Zorlu, C. Gokceoglu, F. Ocakoglu, H. A. Nefeslioglu, and S. Acikalin, "Prediction of uniaxial compressive strength of sandstones using petrography- based models," *Engineering Geology*, vol. 96, no. 3-4, pp. 141–158, 2008.

[28] T. Zhang, Q. Liu, and X. Huang, "Study on TBM penetration rate prediction model and multi-index evaluation method," *Coal engineering*, vol. 53, no. 5, pp. 107–113, 2021.

[29] R. Y. M. Li, B. Tang, and K. W. Chau, "Sustainable construction safety knowledge sharing: a partial least square-structural equation modeling and a feedforward neural network approach," *Sustainability*, vol. 11, no. 20, p. 5831, 2019.

[30] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[31] L. S. Aiken, S. G. West, and S. C. Pitts, "Multiple linear regression," *Handbook of Psychology*, pp. 481–507, 2003.