Journal of Engineering, Project, and Production Management
2024, 14(1), 0010

Bridge Pier Displacement Prediction and Control in Subway Tunnel Construction

Zhaozhao Huang[1] and Jingwen Wang[2]
[1]Lecturer, Technical College, Zhengzhou Technology and Business University, Zhengzhou, 451400, China, E-mail: huangzhaozhao1988@163.com (corresponding author).
[2]Lecturer, Technical College, Zhengzhou Technology and Business University, Zhengzhou, 451400, China, E-mail: simple6475@163.com

Project Management
Received May 9, 2023; revised August 12, 2023; September 28, 2023; accepted November 11, 2023
Available online November 27, 2023

Abstract: As the scale of underground rail transit construction in urban areas continues to expand, the tunnel construction environment has become progressively more complex. In recent years, an emerging artificial intelligence (AI) method in the civil engineering field, called the Random Forest (RF) method, has been widely used. In the construction of Zhengzhou Metro Line 7, the RF method was used to predict and control the vertical displacement of the bridge pier pile foundation. Such displacement can indicate the deformation of the structure, particularly under long-term utilization and strenuous circumstances that could sink or lift the pier body. Moreover, the vertical displacement of the bridge pier can affect the stiffness and bearing capacity of the bridge, thus impacting driving safety and the bridge's service life. Therefore, the vertical displacement of bridge piers has become the main prediction and control indicator for research. In the Zhengzhou Metro Line 7 tunnel, the tunnel continuously passes through 78 bridge pier foundations, among which the pile foundations of 4 key bridge piers are less than 0.5 times the tunnel diameter with a clear distance planned for the tunnel. However, limitations such as surface traffic and environmental conditions prevent the reinforcement of the bridge pier foundation in advance. Therefore, determining and setting sensible shield construction parameters is crucial to effectively controlling the vertical displacement of these essential bridge piers. This project can serve as a model for future endeavors. The study combines Random Forest with Particle Swarm Optimization Algorithm (PSO) to upgrade the technology of shield tunneling through Pier 2, introduces the Bayesian principle for statistical analysis, and optimizes various main construction variables. Random Forest is an ensemble learning method based on decision trees, which has high flexibility and predictive performance. It can automatically filter out important features from a large number of input features, thereby establishing an effective prediction model. The primary research objective is to enhance tunnel construction by accurately predicting and controlling the vertical displacement of pier foundations. To achieve this objective, the study utilizes the PSO to optimize the parameters and structure of the RF model. By doing this, the model's ability to predict the pier's vertical displacement accuracy can be improved. By combining these two methods, the accuracy of the prediction model and the optimization effect of construction parameters can be improved. In addition, the reliability of the model is further improved by using the Bayesian principle for statistical analysis. The paper compares and evaluates the engineering data objectively, presenting the evaluation index and feature selection method. This approach is innovative and purposeful, aiming to enhance the predictive ability, construction efficiency, and quality. This method can provide support for decision-making and optimization of engineering projects and promote sustainable development of the project. After the construction was completed, the model was established, and the results were predicted. The actual engineering measurement data of Pier Two was taken for comparison with it. Two parameters, Root-Mean-Square Error (RMSE) and Linear Curvature (R^2), were introduced to evaluate the prediction results, and the results were subjected to Correlation-based Feature Selection (CFS). The test sets for the downstream and the upstream tunnel were extracted, in which R^2 for the three extracted comparisons of the downstream were 0.83, 0.82, and 0.89, respectively, while R^2 for the upstream was 0.88, 0.86, and 0.86, respectively. From this, it can be seen that the optimized model has good predictive performance. In the construction process of other projects, the model can be used to predict the vertical displacement of bridge piers, which has real-time performance in preventing accidents.

Keywords: Particle swarm optimization - Random forest (pso-rf) model, tunnel, bayesian, pier, displacement, tunnel boring machine (TBM).

DOI 10.32738/JEPPM-2024-0010

1. Introduction

As small and medium-sized cities continue to expand, urbanization accelerates, resulting in persistent population growth. To alleviate the issue of tight land resources and the increasing demand for transportation, cities are upgrading their underground railways to prevent traffic congestion. However, during this process, the construction of tunnels with substantial impact may affect the adjacent pile foundations and disrupt the normal lifestyles of nearby residents (Gao et al., 2019; Jung et al., 2019). The construction plan is developed by prioritizing the impact on the neighboring residents' lives and implementing measures to minimize it. If a piece of equipment can satisfy the requirements of both swift construction and minimal environmental disruption, such as a shield machine, it will be utilized. The tunnel construction industry commonly employs shields, and there will be further developments in tunnel boring machinery's mechanization. Tunnels constructed using the shield method differ from traditional construction in that the shield machine's construction specifications can affect the response of surface buildings (Hasanpour et al., 2020; Liu et al., 2022). The construction parameters should be carefully determined when using shield construction in the presence of both surface buildings and bridge piles in the vicinity of the tunnel (Gao et al., 2019). Special circumstances may arise during the shield construction, for example, if the pile perimeter is not reinforced with slurry, this cannot be addressed hastily. The piers' displacement includes both lateral and vertical directions, and vertical displacement can cause the bridge deck above to deform, affecting the bridge's aesthetics and causing serious problems such as personal danger. To address the issue of vertical displacement of piers and abutments, this study develops a Particle Swarm Optimization Algorithm Random Forest (PSO-RF) model that utilizes random forest and combines PSO and Bayesian optimization. This model is reputable in forecasting bridge pier displacement during tunnel shield construction. As a machine learning approach, random forest can effectively analyze several input features and exhibit robust predictive performance. By combining PSO and Bayesian methods, the parameters and structure of the random forest model can be optimized to consider seven construction parameters. The implementation of the PSO-RF model for bridge pier displacement prediction can facilitate informed decisions and targeted optimization measures during tunnel construction. This, in turn, improves the accuracy of predicting the displacement of bridge piers, thereby helping to avoid potential safety hazards caused by vertical displacement of piers and abutments. Moreover, this enhances the efficiency and quality of tunnel construction systems.

2. Related Works

The safe construction of the project has a rational approach to setting the correct parameters for the construction of the shield and even to reduce the environmental impact during construction. With the goal of creating a new predictor weighting technique, (Alaoui et al.,2022) created a novel scheme in which they combined linear regression with random forest (RF). The method, which includes basic and constrained configurations, not only models seasonal patterns but also replicates historical dates. Both forecasters allow for the selection of optimal weights to ensure the physical relevance of the utilized weather variables. They developed the AnEn system, which is not only statistically consistent but can also be applied to most airports, where time and space can improve the correct rate by 50% and the RMSE by 30%. Its performance is altitude-dependent, with a slight degradation at airports on top of mountains (Alaoui et al., 2022). With the use of evolutionary polynomial regression and random forests, (Yang et al.,2022) created two new performance prediction models. By contrasting the RF-based models with conventional numerical regression techniques, the effectiveness of the RF-based models was evaluated. The effectiveness of the RF-based models was further examined and contrasted, along with the models' resilience to unknown data sets and the significance of variables. The results showed that RF-based models have higher prediction accuracy, especially in identifying outliers (Yang et al., 2022). (Bouwmeester et al., 2019) developed and implemented 36 metabolomics datasets, came up with a novel idea with a variety of feature sets, and were able to assess the effectiveness of seven machine learning methods. It was discovered that there is no single learning algorithm that works for all analyses or procedures. Instead, different types of algorithms perform better in certain cases but worse in others. These findings demonstrate that there is no one-size-fits-all solution. Additionally, combining different models reduces outlier errors, showing that combining various strategies has significant potential for creating more versatile and high-performing algorithms (Bouwmeester et al., 2019).

In various experiments that ignored the outlier problem, (Shahjaman et al., 2020) employed a variety of machine learning techniques to complete the task. The process of generating data from samples for image analysis involves several steps, which can often result in contamination by outliers, thus reducing the accuracy of most algorithms. To evaluate performance, the authors utilized five well-known machine learning methods - SVM, RF, Naive Bayes, k-NN, and LDA. To evaluate the performance of the five machine learning algorithms, we considered three separate outlier rates: 5%, 10%, and 50%. The results show that RF is more suitable than the other four algorithms for randomly taken values (Shahjaman et al., 2020). The ability and viability of six machine learning techniques, including RF, to forecast vertical displacements brought on by tunnel excavation were investigated by (Chen et al.,2019). In order to create the model, field data sets were gathered from four portions of the 3.93 km-long tunnel. To demonstrate the computational models' level of performance, three metrics were used: mean absolute error, root mean square error (RMSE), and coefficient of determination. The results show that the RF algorithm outperforms traditional multiple linear regression methods and is able to accurately track the progression of subsidence induced by the tunnel (Chen et al., 2019). To identify prospective rock types in the operational data, (Zhang et al., 2019) introduced a predictive tunnel-based drilling machine that used the K-means++ algorithm. A support vector classifier exhibiting an average accuracy of 98.6% was selected after comparing classifiers. The cutter torque and thrust force more precisely represent the array of rock types in contrast to the propulsion speed and cutter speed. The generated prediction model is capable of producing test data with 84.4% accuracy and 88.8% recall performance. The proposed approach can effectively identify, characterize, and predict rock types using extensive operational data (Zhang et al., 2019).

According to numerous scholars' research, RF is prone to overfitting, while PSO requires excessive iterations. This study proposes an innovative integration of RF and PSO, utilizing RF for iteration and PSO for fitting, effectively eliminating their respective shortcomings. This approach aims to meet the accuracy requirements of engineering prediction.

3. Research based on Random Forest Algorithm in Tunnel Construction

In tunnel construction, the response of pile foundations to shield construction is complex. Many influencing factors may cause the vertical displacement of piers and abutments during the shield tunnel crossing pile foundation. These factors include the burial depth of piles, the distance between tunnels and piles, the tunnel geometry, ground conditions, and shield tunnel parameters. Nevertheless, excessive influencing factors hinder the progress of this study. Therefore, selecting the main construction variables through technical methods to analyze the vertical displacement of piers and abutments caused by shield tunneling through pile foundations is crucial in determining the principal control factors and controlling pile foundation deformation more effectively. The employed technical method is feature selection, which plays a critical role in the random forest method for feature selection. Feature selection refers to the process of selecting subsets of features from the original feature set to achieve optimal evaluation criteria. This ensures that classification and regression models based on the optimal feature subset can achieve greater prediction accuracy compared to before the feature selection process. Feature selection methods can be used to identify and remove unnecessary, irrelevant, and redundant attributes from data that do not contribute to the accuracy of the prediction model or may actually reduce the accuracy of the model. Efficient features are more suitable for the examination and understanding of outcomes.

3.1. Construction of Bridge Pier Displacement Prediction Algorithm based on RF and PSO

The Random Forest (RF) method is inseparable from the implementation of decision trees, a process that can be interpreted as integrated learning. There are two types of decision trees, including regression trees and classification trees. The objectives of regression trees and classification trees are different (Xiao, 2019; Li et al., 2022; Mengci et al., 2021). A regression tree predicts only continuous values, and its target variable must also be continuous. In contrast, a classification tree targets discrete values for both the target and predictor variables. The objective of this study is to predict the vertical displacement of bridge piers during tunneling due to shield construction. Therefore, the study requires a continuous target variable for building a random forest model. As shown in Fig. 1, only regression trees can be built based on a clear case in this study.

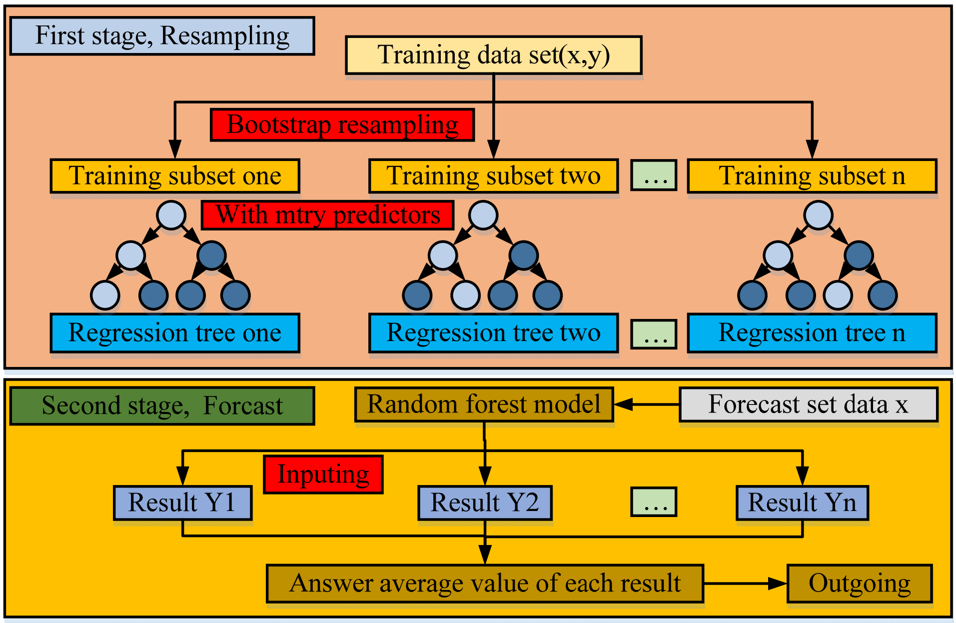

Fig. 1. Random forest regression tree

In Fig. 1, the model is initially trained. The study employs a statistical method in which the original data is randomly and centrally sampled, and the samples are replaced at the end of the sampling process (Williams et al., 2021; Tf et al., 2022). The original data set is assumed to have n samples so that each sampling probability is the same and is $\frac{1}{n}$. The probability of not being sampled is the $1-\frac{1}{n}$, and hence Eq. (1) is obtained.

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n=\frac{1}{e} \tag{1}$$

A regression tree is then constructed based on the sub-dataset to form a random forest, and the test set data can be fed into the RF model for prediction. The initial sample dataset D is input and split. Assuming that the feature a has a range of values and the range has V, then the gain obtained from D is is calculated by Eq. (2).

$$\mathrm{Gain}(D,a)=\mathrm{Ent}(D)-\sum_{v=1}^{V}\mathrm{Ent}\,(D^V) \tag{2}$$

Defining A as a characteristic attribute that will result in the corresponding a. Eq. (2) can divide D into two parts, i.e. D_1 and D_2, then under the condition of A, the Gini coefficient of D is shown in Eq. (3).

$$\mathrm{Gini}(D,A)=\frac{|D_1|}{|D|}\mathrm{Gini}(D_1)+\frac{|D_2|}{|D|}\mathrm{Gini}(D_2) \tag{3}$$

By utilizing Eq. (3), a CART decision tree can be generated with random completeness. Eq. (4) demonstrates the squared error and metric related to the complete decision tree, as defined.

$$\min_{A,S}\left[\min_{C1}\sum_{X\in D1(A,S)}(y_I-c_1)^2+\min_{C2}\sum_{X\in D2(A,S)}(y_I-c_2)^2\right] \tag{4}$$

In the above Eq. (4), y_I represents the true value. There are two samples of $D1, D2$, and the output values are c_1, c_2, respectively. The random forest is built by many randomnesses. The most important one is the use of the random sampling method. After obtaining the qualified training data, the decision tree is trained with this data. The tree possesses the splitting property, which generates various feature properties while splitting. However, the research has introduced PSO with the aim of improving the method's effectiveness. The study continues this process by merging the features and using only the optimal features. Such a method improves the noise immunity of the random forest and also reduces overfitting. The specific information of particle i can be expressed using a D-dimensional vector; the position vector is expressed as $x_i=(x_{i1},x_{i2},\dots,x_{iD})^T$; the velocity vector is expressed as $v=(v_{i1},v_{i2},\dots,v_{iD})^T$, then the updated equation of this velocity and position is expressed as Eq. (5).

$$v_{id}^{k+1}=v_{id}^{k}+C_1r_1\left(pbest_{id}^{k}-x_{id}^{k}\right)+C_2r_2\left(gbest_{id}^{k}-x_{id}^{k}\right) \tag{5}$$

In Eq. (5), v_{id}^{k} represents the flight speed. k represents the number of iterations. i represents the particle. d represents the dimension. x_{id}^{k} reserved space reserved space reserved space reserved space reserved space reserved space reserved space is a parameter of the current position. The coordinate position of the individual extremum is noted as $pbest_{id}^{k}$; The coordinate position of the global extremum in the d dimension can be represented by $gbest_{id}^{k}$, and the randomly taken values of all particles on [0,1] are r_1 and r_2, C_1 and C_2 are factors, which represent learning. The particles need to have their own experience and social experience in the motion, which can not only be used to regulate, but also to record the two motion parameters of each particle, pbest and gbest. The PSO algorithm contains six processes, and its flowchart is depicted in Fig. 2 (Chen et al., 2019).

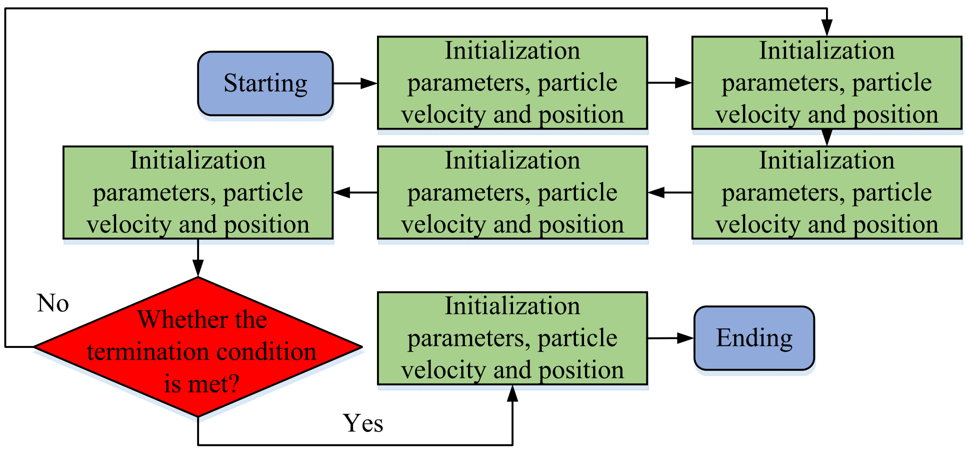

Fig. 2. Flow chart of particle swarm optimization algorithm

After completing the process outlined in Fig. 2, each particle follows its own optimal path. The optimal solution for the population is identified as the location with the highest fitness value. At the beginning of the process, the position and velocity of each particle are randomized. The velocity of each particle moves towards not only the global optimum but also towards the vicinity of the individual optimum. Throughout the motion, each particle reaches its own fitness optimum. The particle constantly updates its information, moving towards both the global and individual optimum solutions. The information includes position and search speed, which Eq. (6) updates.

$$V_i=\omega V_i+C_1\mathrm{random}(0,1)(P_i-X_i)+C_2\mathrm{random}(0,1)\left(P_{gi}-X_i\right) \tag{6}$$

In Eq. (6),$\omega(\omega)\geq 0$ is the inertia factor of search speed, which is the most important hyperparameter of PSO to balance the global and local search ability. When ω starts to increase, the global search ability increases with it, and the local search ability decreases relatively, and vice versa. C_1 and C_2 represent the acceleration constant in the optimization. The history of the algorithm in the direction of the individual optimal value is called C_1, and the number of times the algorithm learns in the direction of the global historical optimal value is C_2. When the set number of iterations is reached, the algorithm terminates. In Eq. (6), the velocity V_i of the particle is multiplied by the inertia weight coefficientω, andωadjusts the formula to Eq. (7).

$$\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{iter_{max}} \times iter \tag{7}$$

In Eq. (7), the maximum and minimum values of ω are denoted as ω_{max} and ω_{min}; The current and maximum number of generations are denoted by iter and $iter_{max}$. In order to achieve the global and local optimization, ω is set to 0.9 before the algorithm runs, and ω is decremented to 0.4 during the algorithm run by Eq. (6).

The above model was constructed from relevant references, and now the application of the model was studied to shield tunnel construction. The construction parameters of the four piles in the tunnel were determined by using the model constructed above. The influencing factors to be considered in the model mainly include geometrical parameters, geological parameters, and shield construction parameters. The geometric parameters include tunnel diameter, buried depth, buried depth of pile foundation, and horizontal distance between them. As a result, the tunnel diameter is constant. The distance between the pile foundation under the bridge pier and the outer edge of the tunnel structure is recorded as S. The lower tunnel of the shield tunnel is recorded as L_1, and the depth of the upper tunnel is recorded asL_2. L_1. and L_2 are important geometric factors in the study. The geological parameters include soil geological and mechanical properties, such as elastic modulus, Poisson's ratio, and cohesive force. The soil mass of the research project is mainly soft silty clay and hard silty clay. The soil thickness of the soil layer is T_1 and T_2, respectively.T_1 and T_2 are taken as the geological conditions parameters. In the project, the piers are sometimes reinforced with grouting around piles, making it a geological parameter to reinforce the pile foundation through grouting. The construction parameters of the shield consist of silo pressure, total thrust, grouting pressure, excavation speed, and grouting quantities. During the project's monitoring operation, the displacement value refers to the stable height difference prior to the cutter head's arrival and upon the departure of the shield tail. The corresponding value for each pier is determined by the geometric and geological parameters surrounding the pier, as well as the geometric relationship with the shield tunnel. Technical term abbreviations will be defined upon first use. However, the recorded EPBM construction parameters are frequently adapted and revised during tunneling. There is a set of EPBM construction parameters corresponding to each ring of shield tunneling. The whole fuselage of the shield may have a significant impact on the pile foundation when it passes through the pile foundation, but it is impossible to include all the EPBM construction parameters in the whole process. Therefore, how to reasonably determine the EPBM construction parameters when the shield tunneling side passes through the pile foundation under the pier becomes the key to reasonably setting up the training set. In order to minimize the error caused by sampling discretization, the average value of EPBM construction parameters monitored over a period of time is used as the shield construction parameter in the random forest model.

3.2. PSO-RF Algorithm based on Bayesian Network Optimization

To prove the displacement conjecture of the model and further optimize the construction parameters, the study verifies the prediction accuracy by the established random forest model and also introducesR^2 as an indicator for the evaluation of the RF model, which is calculated as Eq. (8) (Zhu et al., 2019).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(s_v^{obs} - s_v^{pred})^2}{\sum_{i=1}^{n}(s_v^{obs} - s_{vm}^{pred})^2} \in [0,1] \tag{8}$$

In Eq. (8), n is the sample size. $s_v^{obs}$and$s_v^{pred}$are the observed and predicted vertical displacement values, respectively. s_{vm}^{pred} is the average observed value. R^2reflects the degree of correlation between the measured data and the predicted results. The RF model predicts the vertical displacement of shield construction parameters, but optimization of the main control items is necessary for quality assurance. Bayesian optimization is employed to consider the project's overall construction situation during the optimization process. The Bayesian theorem constitutes a crucial machine learning algorithm component, encompassing several key concepts: prior probability, posterior probability, and conditional probability, among others. Before an event occurs, there is a pre-determined probability, i.e., the prior probability. After the event occurs, a reverse conditional probability also needs to be solved. By the prior probability, the result will be found. If the form is the same as the conditional probability, it is called the posterior probability. Sometimes an event occurs followed by another event; then, the above cases are referred to as conditional probability. The source of this prognosis is empirical summary or data statistics, and they are related as in Eq. (9).

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \tag{9}$$

$p(y|x)$ is the posterior probability, which is the result of Eq. (9). $p(x|y)$ can only be obtained statistically from historical data and is called conditional probability. $p(y)$ is generally known and is called the prior probability, which can generally be given by subjective opinion. In fact, $p(x)$ is also called Bayesian prior probability and can be calculated by the full probability formula, which is given in Eq. (10).

$$x_i = \frac{x_i - \min_{1\le j\le n}\{x_j\}}{\max_{1\le j\le n}\{x_j\} - \min_{1\le j\le n}\{x_j\}} \tag{10}$$

Due to the inconsistency, there may be serious differences between the data. The data normalization mainly uses the extreme value method (max-min), which maps all the original data to the range interval from 0 to 1 by a linear transformation, as shown in Eq. (10). In Eq. (10), $\min_{1\le j\le n}\{x_j\}$ denotes the minimum value of the sample and $\max_{1\le j\le n}\{x_j\}$ denotes the maximum value of the sample. The Bayesian optimization process around the research content and research method is shown in Fig. 3.

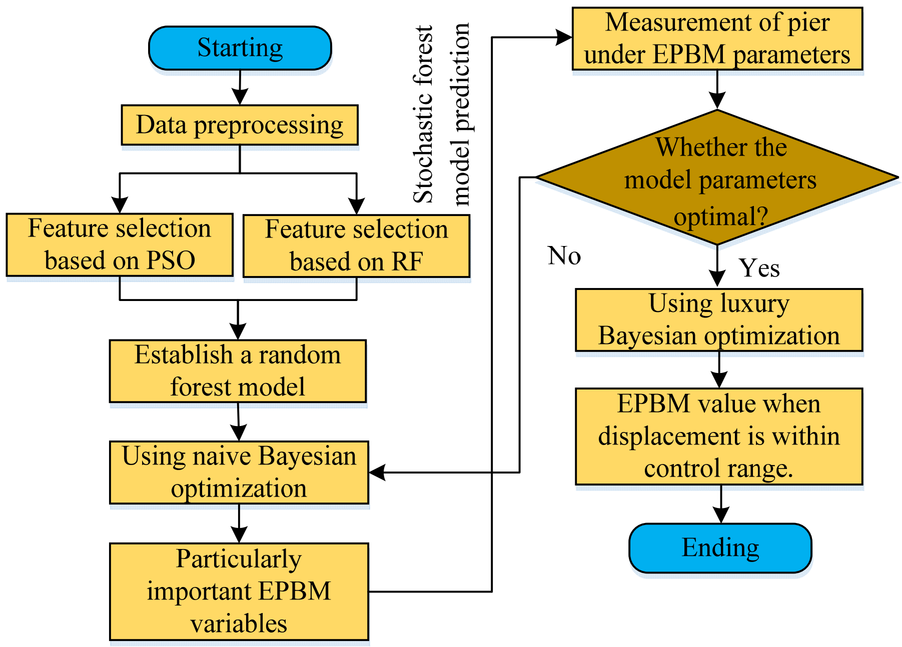

Fig. 3. Bayesian optimization flow chart

In Fig. 3, the Bayesian process first establishes the raw data and then undergoes feature selection through PSO and RF, followed by naive Bayesian optimization. The EPBM variables obtained undergo random forest prediction to determine their optimality. If deemed optimal, they undergo deluxe Bayesian optimization and output EPBM values within the control range. If not, the process is repeated from the previous step. The best prediction result is achieved through adjusting two parameters, Mtry and Ntree. Avoidance of process over is preferred. A new parameter out-of-bag error rate ERR_{OOB} is also introduced, which serves as an unbiased estimate of the generalization error, as in Eq. (11).

$$ERR_{OOB} = \left[\frac{1}{Ntree}\right] \sum_{i=1}^{Ntree} [Y_i - g_{OOB}(X_i)]^2 \tag{11}$$

In the above Eq. (11), X_i (number 1 of i to Ntree) is the sub-sample group drawn from X. The elements not contained in the sub-sample group X_i are called Out-Of-Bag Data (OOB). Y_i is the out-of-bag data, and $g_{OOB}(X_i)$ is the predicted value of the regression tree for OOB. To quantitatively evaluate the accuracy of this algorithm in predicting the displacement of the surrounding piers due to the shield tunnel, a parameter is also introduced to assist in understanding the relationship between the features and the response variables. When considering a more realistic problem, the RMSE is usually chosen, as shown in Eq. (12).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(s_v^{obs} - s_v^{pred})^2}{N}} \in (0, +\infty) \tag{12}$$

In Eq. (12), n is the sample size. There are two values of vertical displacement, i.e., observed and predicted, denoted by s_v^{obs} and s_v^{pred}, respectively. RMSE is a square root, which is calculated by the ratio of the square of the observed deviation from the true value to the number of observations n. The deviation between the observed and predicted values of the vertical displacement of the bridge pier is measured by RMSE. The j -th construction variable PV_j takes values in the i subinterval and is represented by ij. c is another instance of an occurrence. It shows that the vertical displacement of the bridge pier is kept within safe bounds.

It can be operated by the Bayesian principle and the probability that the j important primary control construction variable takes value in the PV_ja subset of decisions is shown in Eq. (13) (Xu et al., 2019; Chen et al., 2022; Hou et al.,2022).

$$p(ij|c) = \frac{p(c|ij)p(ij)}{p(c)} \tag{13}$$

Assuming two variables X, Y, measuring the linear relationship between the variables. The correlation coefficient can be obtained by Eq. (14).

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\varsigma_X)(Y-\varsigma_Y)]}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X)^2-E^2X}\sqrt{E(Y)^2-E^2Y}} \tag{14}$$

This time includes two distinct processes: feature selection and learner training. The embedded approach incorporates both and is designed to achieve feature selection and model training simultaneously. Embedded feature selection represents a new trend that necessitates the identification of a model prior to learning. The deviation of the observed value from the theoretical value is to be calculated and the estimation of the correctness of the theoretical value is to be performed. Based on χ^2, the basic formula is (15).

$$\chi^2 = \sum \frac{(A-E)^2}{E} = \sum_{i=1}^{k} \frac{(A_i-E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{(A_i-np_i)^2}{np_i} \tag{15}$$

According to the current setting, select an appropriate threshold and the best data mining technique based on the particular data characteristics to test the model. To achieve higher accuracy in the final outcomes, continual adjustment and enhancement are applied. The connection between PSO and RF is established, and optimization is accomplished using Bayesian, as depicted in Fig. 4 (Ramosaj and Pauly, 2019; Yang et al., 2022).

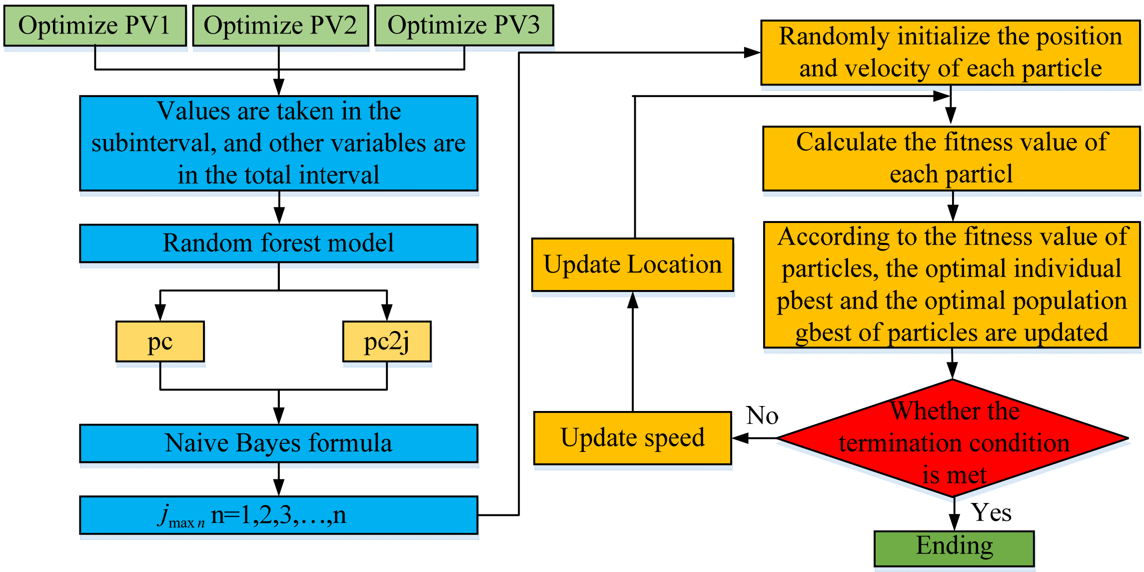

Fig. 4. Bayesian optimization flow chart based on improved RF and PSO fusion

In Fig. 4, by using randomly generated samples PV_n (n=1,2, 3,...,n), values are taken and optimized in sub-intervals or full intervals. After optimization, a random forest model is built to obtain the optimized value pc or pc2j. This is fed into the PSO algorithm, which can output the best method for predicting the bridge pier displacement in tunnel construction as well as for controlling it.

4. PSO-RF Model based on Bridge Pier Displacement Prediction in Tunnel Construction

4.1. Determination of Model Parameters for PSO-RF

The study used Python for simulation experiments and then used the data collected during construction as the experimental dataset to train and test the model. In the experiment, a dual-core CPU was used, with a memory size of 8GB and a solid-state hard drive of 256GB. The GPU used an Intel-integrated graphics card, and the operating system was Windows 10. The above experimental configuration can help the random forest algorithm achieve a prediction model for the vertical displacement of bridge piers while combining it with the particle swarm optimization algorithm to find the optimal model parameters. In the experiment, the parameters of the random forest model and particle swarm optimization algorithm can be adjusted to obtain good prediction results. The study optimizes the bi-directional feature selection for random forests using a particle swarm technique. The resulting data can be used as model input. As a comparative experiment, the study employs Correlation-based Feature Selection (CFS), which uses the best-first search strategy for feature subset selection. The respective feature subset is also selected by CFS.

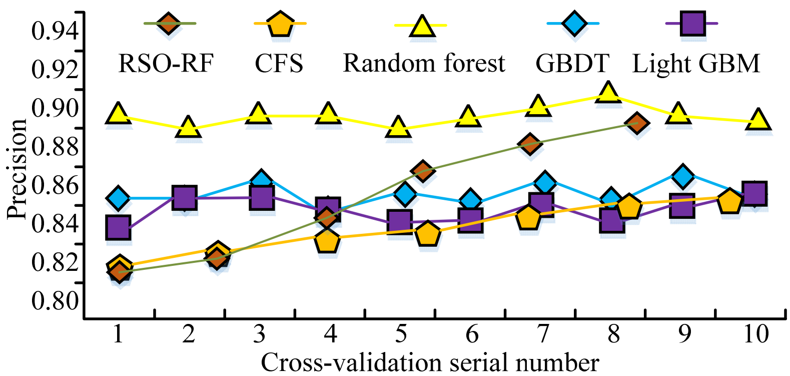

Fig. 5. Comparison of classification results of different models

The accuracy of each cross-validation using bidirectional feature selection algorithms generated by Random Forest, GBDT, LightGBM, CFS, and RSO-RF is shown in Fig. 5. The cross-validation includes ordinal numbers, denoted as the horizontal coordinates in Fig. 5. It also yields a classification accuracy, expressed as a percentage and represented by the vertical coordinate in Fig. 5. As shown in the figure, the PSO-RF two-way feature selection method achieves an F1 score of 90.66% and a classification accuracy of 88.73%. When compared to using the CFS feature selection approach, it outperforms both in terms of accuracy and F1 value, and the algorithm also saves time. In order to take the values of Mtry and Ntree, the calculation of different combinations of ERR_{OOB} is plotted according to Eq. (11), as shown in Fig. 6.

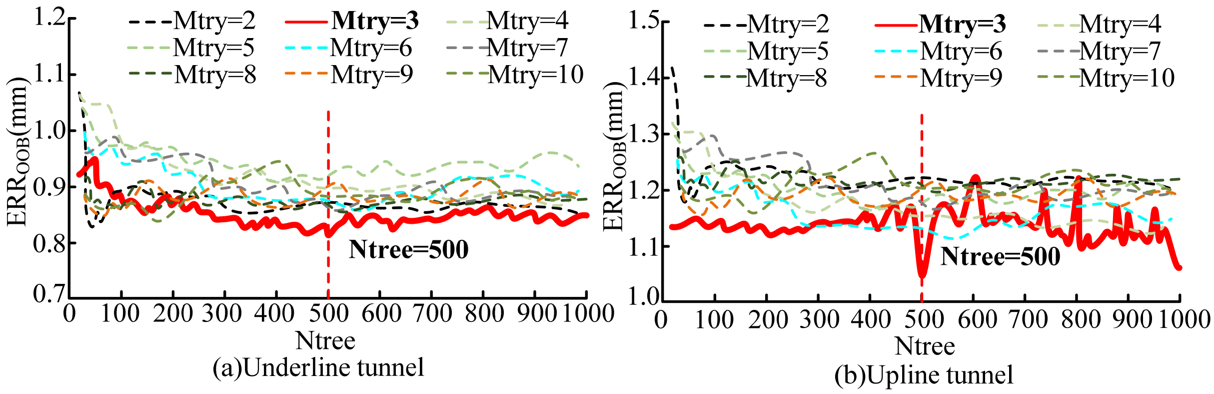

Fig. 6. Performance of model under different combinations of mtry and ntree values

It can be clearly seen from Fig.6 that when Mtry=3, ERR_{OOB} reaches the minimum value in both plots. It can also be seen from Fig.6 that when Ntree is greater than 500, ERR_{OOB} increases with the value of Ntree. Although it still increases, but the increase is small and does not have a large impact on the results, and its performance is very stable when Ntree=500. Considering the goals of performance maximization and time efficiency, the parameters Mtry and Ntree were determined to be optimal at 3 and 500, respectively. Based on the features chosen from the training set, Fig. 7 illustrates the random forest model's training set for the lower and upper tunnels.

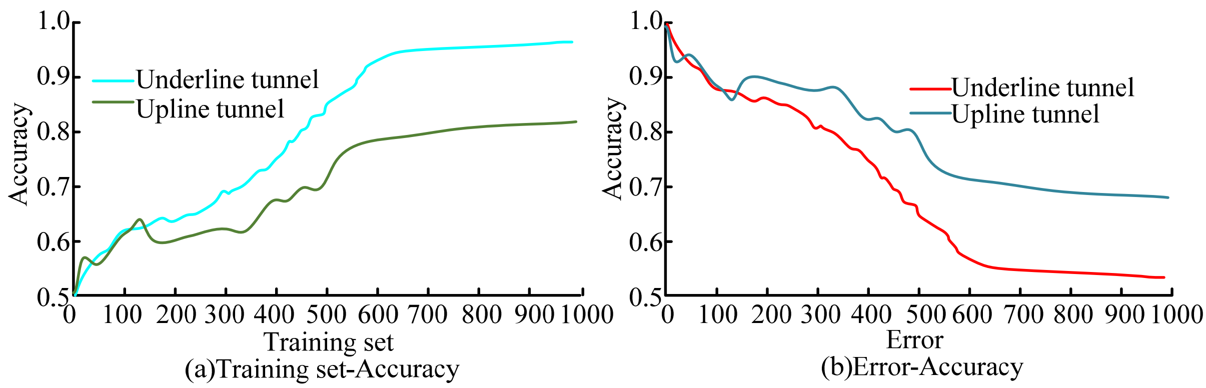

Fig. 7. Error, Training set-accuracy image

After obtaining the training set experimentally, a random forest model can be built, and feature selection can be performed, which is the cornerstone for optimization. Zhengzhou Metro Line 7 is a double-track through tunnel with a total length of 49.5km. This section of the tunnel is laid entirely underground, with one EPBM (Herrenknecht shield machine) being used for its operation. Given the geological and site conditions of the interval tunnel, a field profile has been created, as shown in Fig. 8.

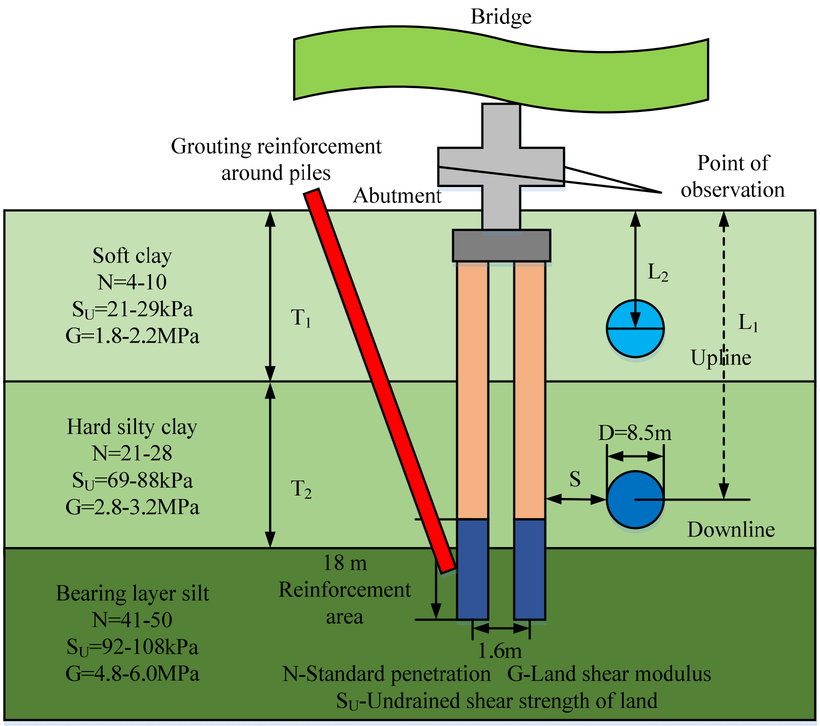

Fig.8. Tunnel-soil layer-pile foundation profile

According to Fig. 8, the top thickness T1 of the soft soil layer varies between 18.2 and 20.6 m. Below the soft soil layer, there is a hard soil layer with a thickness of 10.6 to 18.6 m, known as the middle layer or T2. The bearing layer of the piers can be found at the bottom. Each pier is supported by two piles that are 48 m long with a 1.6 m pile spacing. The Fig. 8 includes an explanation for N, G, and other symbols.

4.2. Field Validation based on PSO-RF Model

Taking the pier of Zhengzhou Metro Line 7 as an example, Fig. 9 shows the change in vertical displacement of the representative pier (pier 2) with the forward excavation of EPBM, and the distance between the tunnel face and the monitoring pier.

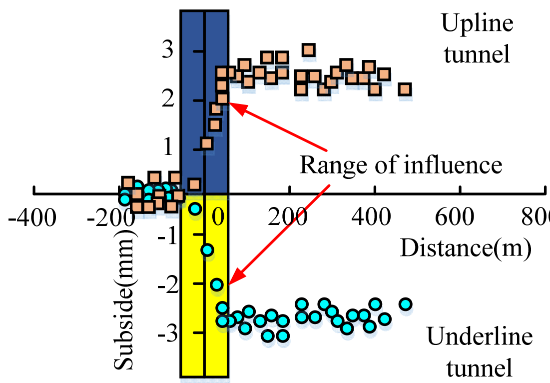

Fig.9. Changes of vertical displacement of pier No.2 with EPBM's advancement

As displayed in Fig. 9, the monitored pier is impacted by the EPBM once the tunnel palm face is 30 meters away from it, but this impact becomes insignificant once the tunnel palm face exceeds it by approximately 60 meters. Hence, the random forest model is fed with EPBM data within the range of [-30m, 60m] as input. In practical applications, the EPBM tunnelling process involves various modules working together to propel the shield head forward. Fig. 10 illustrates a schematic diagram of the shield operation in an actual project.

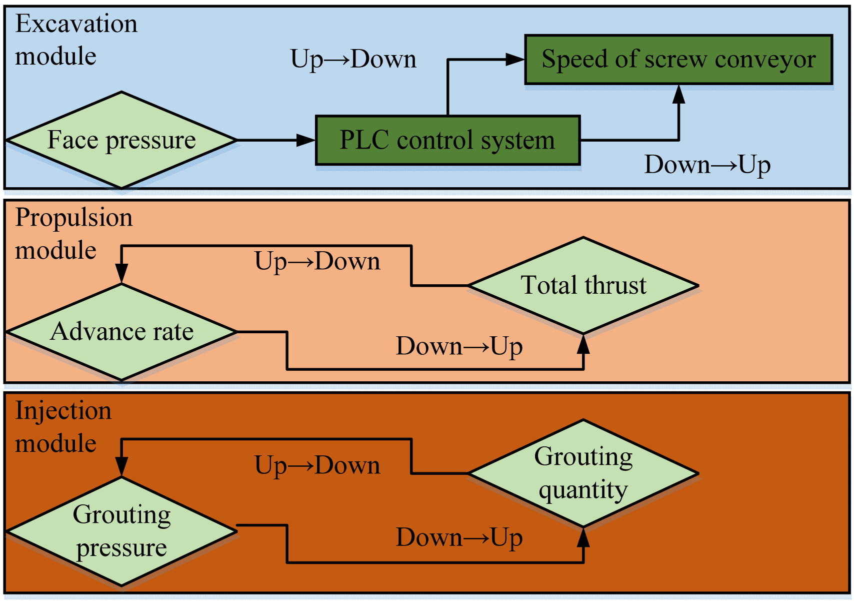

Fig.10. Logic diagram of EPBM operation module

In actual projects, the excavation, propulsion, and tail grouting modules are utilized to control the shield machine's excavation. Technical term abbreviations will be defined upon first use. The excavation module removes soil in front of the shield's face, while the propulsion module allows the equipment to advance through the soil layer. The pressure of the grouting pump and the grouting volume in the grouting module are interrelated control variables. After construction is complete, the engineering data for Pier 2 will be used to verify the random forest model. The measured engineering data of Pier 2 are shown in Table 1.

Table 1. Engineering measured data of pier No. 2

S	L	T_1	T_2	V	ΔP	Vg	Sv
24.56	26.08	21.36	13.22	28.20	201.25	5.01	-0.20
21.39	27.88	21.39	14.52	25.87	207.52	5.20	-0.88
1.74	25.83	22.01	12.56	26.85	215.24	4.11	-0.54
25.82	24.05	20.86	13.15	25.97	205.13	4.85	-0.47
13.14	25.44	24.54	15.74	25.45	204.45	5.45	-0.83
1.22	20.57	22.74	12.44	25.44	214.55	4.25	-0.65
28.45	20.44	28.23	11.57	24.32	222.36	5.11	-0.26
22.93	29.11	26.41	15.34	34.39	207.53	4.44	-0.55
1.85	22.93	29.04	12.15	22.66	222.62	4.60	-0.50

The study was optimized based on the predicted results using the Bayesian principle, and the target variable S_v was obtained by changing the equivalents, such as the standard penetration degree. The results of the experiments are shown in Table 1 for a total of nine groups. The sixth and ninth groups exhibit values of S that deviate from the norm, owing to the volume limitation of the bridge pier preventing a larger value. As bridge piers differ in real-world scenarios, outliers were maintained to widen the scope of the results. Eq. (12) was employed to obtain the RMSE for the nine data sets' target variables, which are featured in Fig. 11.

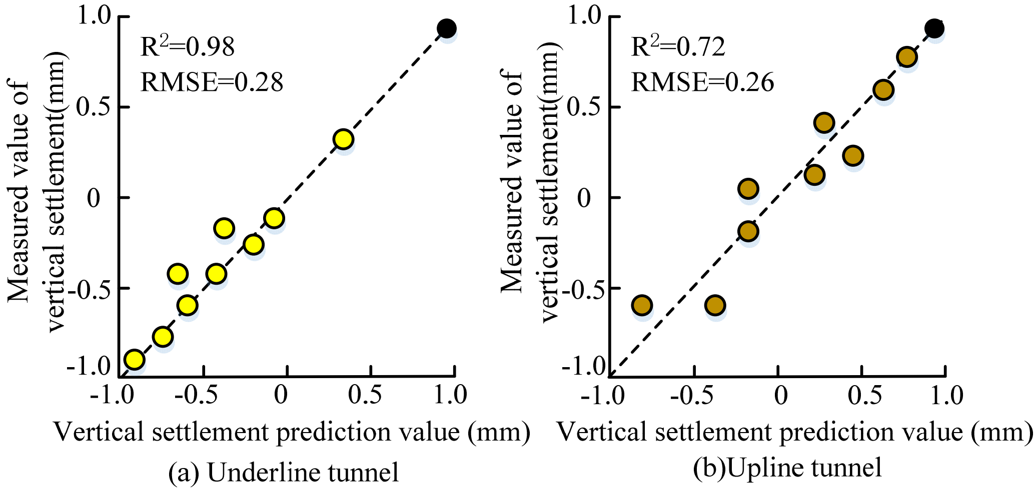

Fig.11. Verification of RSO-RF algorithm with no.2 pier

The RSO-RF algorithm is used to verify the No. 2 pier of Zhengzhou Metro Line 7 and to plot the RMSE images of its upper and lower tunnels. As can be seen in Fig. 11, the R2 of Fig. 11(a) is 0.98 and the RMSE value is 0.28; The R2 of Fig. 11(b) is 0.72 and the RMSE value is 0.26. The linear fit of the down-line tunnel is very close to the predicted value, but the linear fit of the up-line tunnel is not good, and the deviation of the second and fifth sets of data is relatively large from the average value. In the experiment, the data collection was conducted with high air humidity, so the collected data will have some influence on the bridge pier pile foundation, but the results can still meet the actual demand.

5. Discussion

Based on the experimental results, the bidirectional feature selection method utilizing particle swarm optimization was employed to forecast the vertical displacement of bridge piers. The actual project of Qinghai Metro Line 8 experienced advantageous results. By processing and selecting features from the collected construction data, a training set was acquired, and the prediction and verification were accomplished through the implementation of a random forest model. In the experiment, the study was compared with the CFS feature selection method, and the results showed that the PSO-RF method achieved significant improvements in both accuracy and F1 value. The model has high classification accuracy and low error, which shows its effectiveness in predicting the vertical displacement of bridge piers. In addition, the optimization of the random forest model parameters resulted in the discovery that the model performs most effectively when Mtry is three and Ntree is 500, according to experiments conducted. The chosen parameters serve as a foundation for case-specific adjustments, enhancing prediction accuracy and minimizing experimental duration. Despite this, certain limitations must be acknowledged. Primarily, the study's experimental data solely entails Zhengzhou Metro Line 7's actual project, which may yield distinctive conclusions in alternate projects or tunnels. Different geological conditions, tunnel structures, and construction processes may impact the anticipated outcomes. Potential recipients of the research include researchers, engineers, and practitioners who need to accurately predict the vertical displacement of piers in order to design, construct, and maintain tunnel and bridge structures and assess their stability and safety. The practical applications of the research results include the utilization of the random forest model, optimized through the particle swarm optimization algorithm, to predict the vertical displacement of bridge piers. This allows engineers to evaluate the stability of tunnel and bridge structures, facilitating the incorporation of appropriate measures during design, construction, and maintenance. The bidirectional feature selection method and tuning of random forest parameters can be used to address similar prediction problems. These techniques aid researchers in selecting the most pertinent features and optimal model parameters for improved prediction performance and accuracy in various engineering situations. There are limitations to these findings. The experimental data are solely from the Qinghai Metro Line 8 project; Therefore, further verification is necessary for its applicability in other tunnel and engineering cases. Different geological conditions, engineering environments, and construction processes may result in different forecast outcomes. Additionally, the selection methods used in this study regarding parameters and features may require customization to adapt to the needs of other engineering cases.

6. Conclusion

In tunnel engineering, the implementation of shield tunneling machines is expected to affect pile foundations. The influence of a reduced count of shield tunneling machines on bridge pier pile foundations poses a more intricate problem. Hence, the principal research objective pertains to the prediction and regulation of bridge pier pile foundations. The study utilized the random forest algorithm to strengthen the base of the three piers of Qinghai Metro Line 8, while the PSO-RF method was employed to horizontally connect the shield tunneling of the second pier. Statistical analysis of the data was conducted using the Bayesian principle, and the parameters of the primary control structure were optimized. After the construction was completed, the measured engineering data of pier two were compared with the predicted results of the established model, and the predicted results were evaluated using two parameters, RMSE and R2. The results were CFS.9 groups of experimental data target displacements. SV were measured, which were -0.20, -0.88, -0.54, -0.47, -0.83, -0.65, -0.26, -0.55, -0.50, mostly around -0.50. The average value taken is -0.54, which shows that the optimized values are closer to the predicted values. The RMSEs of the lower and upper tunnels are 0.28 and 0.26, respectively, both of which are very small;

R^2 is 0.75 and 0.73, respectively. Based on the RMSEs plotted on the data, $R^2 = 0.98$ for the lower tunnel, which fits almost exactly with the predicted value, and it can be seen that the optimized construction parameters can effectively control the vertical displacement of the bridge pier. However, the linear fit of the up-line tunnel is not very good, with R2 being only 0.72. The study successfully predicted the settlement of bridge piers and other elements using the proposed method during tunnel construction using a shield machine. In brief, this study's advantage lies in utilizing PSO-RF to optimize parameters. The particle swarm optimization algorithm-based optimization method can efficiently explore the parameter space, identify the optimal combination of building parameters, and enhance predictive performance. Employing the Bayesian principle to statistically analyze data, this method can help gain better insights into and process measured data, leading to more dependable prediction outcomes. However, there are still shortcomings in the experiment. The model's applicability can be verified by collecting measured data from various projects and regions. Additionally, increasing the number of samples can enhance the stability and generality of the model. External factors such as soil disturbance and the impact of rainwater on the soil affect the shield tunneling of the tunnel. Therefore, subsequent research can conduct simulation experiments in an external environment. The study introduces the random forest and particle swarm optimization algorithm and applies these methods to predict the displacement of bridge pier foundations. The study demonstrates that these techniques can enhance the accuracy of prediction and optimize the construction parameters to some degree. The aim of this study is to furnish a practical approach and structure for anticipating and enhancing the uplift of foundations, with the additional aim of providing advice and direction for practical engineering applications.

Author Contributions

The random forest combined with the Particle Swarm Optimization algorithm (PSO) method (PSO-RF) was used to upgrade the technology of the No.2 pier, which was laterally penetrated by the shield, and the Bayesian principle was introduced for statistical analysis, and each main control construction variable was optimized. Zhao Huang conducted experiments, recorded data, analyzed the results, and Jingwen Wang wrote a manuscript. All authors agreed to the published version of the manuscript.

Institutional Review Board Statement

Not applicable.

Funding

This research received no specific financial support from any funding agency.

References

Alaoui, B., Bari, D., and Ghabbar, Y. (2022). Surface weather parameters forecasting using analog ensemble method over the main airports of morocco. *Journal of Meteorological Research*, 36(6), 866-881.doi:10.1007/s13351-022-2019-0.

Bouwmeester, R., Martens, L., and Degroeve, S. (2019). Comprehensive and empirical evaluation of machine learning algorithms for small molecule LC retention time prediction. *Analytical Chemistry*, 91(5), 3694-3703.doi:10.1021/acs.analchem.8b05820.

Chen, R., Zhang, P., Wu, H. N., Wang,Z. T.,and Zhong,Z. Q. (2019). Prediction of shield tunneling-induced ground settlement using machine learning techniques. *Frontiers of Structural and Civil Engineering*, 13(6), 1363-1378.doi:10.1007/s11709-019-0561-3.

Chen, R., Zhang, P., Wu, H. N., Wang, Z.,and Zhong,Z. (2019). Prediction of shield tunneling-induced ground settlement using machine learning techniques. *Frontiers of Structural and Civil Engineering*, 13(6), 1363-1378.doi:10.1007/s11709-019-0561-3.

Chen, T., Zhang, T., Niu, C., Feng, T., Tang, H., Cheng, X., and Li, H. (2022). Multi-element quantitative analysis of single micro-sized suspended particles in air with high accuracy based on random forest and variable selection strategies. *Analytical Chemistry*, 94(50), 17595-17605.doi:10.1021/acs.analchem.2c04163.

Gao, X. J., Shi, M. L., Song, X. G., Zhang, C., and Zhang, H. W.(2019). Recurrent neural networks for real-time prediction of TBM operating parameters.*Automation in Construction*, 98(5), 225-235.doi:10.1016/j.autcon.2018.11.013.

Gao, X., Shi, M., Song, X., C, Zhang., and Zhang, H.(2019). Recurrent neural networks for real-time prediction of TBM operating parameters. *Automation in Construction*, 98(5), 225-235.doi: 10.1016/j.autcon.2018.11.013.

Hasanpour, R., Rostami, J., Schmitt, J., Ozcelik, Y., and Sohrabian, B.(2020). Prediction of TBM jamming risk in squeezing grounds using Bayesian and artificial neural networks.*Journal of Rock Mechanics and Geotechnical Engineering*, 12(1), 21-31.doi: 10.1016/j.jrmge.2019.04.006.

Hou, S., Liu, Y., and Yang, Q. (2022). Real-time prediction of rock mass classification based on TBM operation big data and stacking technique of ensemble learning.*Journal of Rock Mechanics and Geotechnical Engineering*,14(1), 123-143.doi:10.1016/j.jrmge.2021.05.004.

Jung, J. H., Chung, H., Kwon, Y. S., and Lee, I. M. (2019). An ANN to predict ground condition ahead of tunnel face using TBM operational data.*KSCE Journal of Civil Engineering*, 23(320), 5-6.doi:10.1007/s12205-019-1460-9.

Li, J., Zhang, W., Diao, W., Feng, Y., Sun, X., and Fu, K.(2022). CSF-Net: Color Spectrum Fusion Network for Semantic Labeling of Airborne Laser Scanning Point Cloud. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15(12), 339-352.doi:10.1109/JSTARS.2021.3133602.

Liu, L., Zhou, W., and Gutierrez, M. (2022). Effectiveness of predicting tunneling-induced ground settlements using machine learning methods with small datasets. *Journal of Rock Mechanics and Geotechnical Engineering*, 14(4), 1028-1041.doi:10.1016/j.jrmge.2021.08.018.

Mengci, L., Zhenbin, G., Bo, Q., Jiannan, Z., Yonghuan, M., Guanjie, X., and Yuxin, Z. (2021). Photometric redshifts estimation for galaxies by using FOABP-RF. *Monthly Notices of the Royal Astronomical Society*, 506(4), 5924-5934.doi:10.1093/mnras/stab2040.

Ramosaj, B. and Pauly, M.(2019). Consistent estimation of residual variance with random forest Out-Of-Bag errors Stat. *Statistics & Probability Letters*, 151(8),49-57.doi:10.1016/j.spl.2019.03.017.

Shahjaman, M., Rashid, M. M., Asifuzzaman, M. I., Akter, H.,Islam, S.,and Mollah,M. (2020). Performance evaluation of different machine learning algorithms in presence of outliers using gene expression data. *Journal of Bio-Science*, 28(12), 69-80.doi:10.3329/jbs.v28i0.44712.

Tf, A., Cw, A., Jian, Z. A., Bw, B., and Yfj, C. (2022). An improved artificial bee colony-random forest (IABC-RF) model for predicting the tunnel deformation due to an adjacent foundation pit excavation. *Underground Space*, 7(4), 514-527.doi:10.1016/j.undsp.2021.11.004

Williams, P. T., Wynne, R. H., Thomas, V. A., and Defries, R. (2021). Mapping smallholder forest plantations in Andhra Pradesh, India using multitemporal Harmonized Landsat Sentinel-2 S10 data. *Land Degradation & Development*, 32(15), 4212-4226.doi:10.1002/ldr.4027.

Xiao, G. (2019). Problems of railway tunnel construction under some special geological conditions in China and their countermeasures. *Tunnel Construction*, 39(11), 1748-1758.doi:10.3973/J.ISSN.2096-4498.2019.11.002.

Xu, G. X., Xu, X. Y., Wang, L., Guiquan,F. U.,Zhao,P., and Ding, A. (2019). Sand-fix effects of Haloxylon ammodendron forests under the different densities and patterns under wind tunnel test. *Journal of Arid Land Resources and Environment*, 33(9), 189-195.doi:10.13448/j.cnki.jalre.2019.255.

Yang, J., Yagiz, S., Liu, Y. J., and Laouafa,F. (2022). Comprehensive evaluation of machine learning algorithms applied to TBM performance prediction. *Underground Space*, 7(1), 37-49. doi:10.1016/j.undsp.2021.04.003.

Yang, J., Yagiz, S., Liu, Y. J., and Laouafa, F. (2022). Comprehensive evaluation of machine learning algorithms applied to TBM performance prediction. *Underground Space*, 7(1), 37-49.doi:10.1016/j.undsp.2021.04.003.

Zhang, Q., Liu, Z., and Tan, J. (2019). Prediction of geological conditions for a tunnel boring machine using big operational data. *Automation in Construction*, 100(4), 73-83.doi:10.1016/j.autcon.2018.12.022.

Zhu, R., Fang, J., Li, S., Wang,Q., and Yul,H. (2019). Vehicle re-identification in tunnel scenes via synergistically cascade forests. *Neurocomputing*, 381(14),227-239.doi:10.1016/j.neucom.2019.11.069.

Zhaozhao Huang was born in 1988 in Henan, China. She graduated from Xi'an Technological University, majoring in Architecture and Civil Engineering, and received her bachelor's degree in 2012 and master's degree in 2014. Since 2015, she has been working at Zhengzhou Technology and Business University. She has published three papers and presided over or participated in several research projects. Her research interest is geotechnical engineering technology.

Jingwen Wang was born in 1989 in Henan, China. She studied Environmental Engineering at Zhengzhou University. She received her bachelor's and master's degrees in 2011 and 2014. Since 2015, she has been working at Zhengzhou Technology and Business University. She has published five papers and conducted or participated in several research projects. Her research interests are environmental science and urban and rural planning.