Comprehensive evaluation of machine learning algorithms applied to TBM performance prediction

Jie Yang, Saffet Yagiz, Ying-Jing Liu, Farid Laouafa

Available online at www.sciencedirect.com

ScienceDirect

Underground Space 7 (2022) 37–49

www.keaipublishing.com/undsp

Comprehensive evaluation of machine learning algorithms applied to TBM performance prediction

Jie Yang [a,b,*], Saffet Yagiz [c], Ying-Jing Liu [a], Farid Laouafa [d]

[a] Zhongtian Construction Group Co. Ltd, Hangzhou, China
[b] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[c] School of Mining and Geosciences, Nazarbayev University, Nur-Sultan 010000, Kazakhstan
[d] National Institute for Industrial Environment and Risks (INERIS), Verneuil-en-Halatte, France

Received 20 February 2021; received in revised form 28 March 2021; accepted 7 April 2021
Available online 14 May 2021

Abstract

To date, the accurate prediction of tunnel boring machine (TBM) performance remains a considerable challenge owing to the complex interactions between the TBM and ground. Using evolutionary polynomial regression (EPR) and random forest (RF), this study develops two novel prediction models for TBM performance. Both models can predict the TBM penetration rate and field penetration index as outputs with four input parameters: the uniaxial compressive strength, intact rock brittleness index, distance between planes of weakness, and angle between the tunnel axis and planes of weakness (α). First, the performances of both EPR- and RF-based models are examined by comparison with the conventional numerical regression method (i.e., multivariate linear regression). Subsequently, the performances of the RF- and EPR-based models are further investigated and compared, including the model robustness for unknown datasets, interior relationships between input and output parameters, and variable importance. The results indicate that the RF-based model has greater prediction accuracy, particularly in identifying outliers, whereas the EPR-based model is more convenient to use by field engineers owing to its explicit expression. Both EPR- and RF-based models can accurately identify the relationships between the input and output parameters. This ensures their excellent generalization ability and high prediction accuracy on unknown datasets.

Keywords: Tunnel boring machine; Evolutionary polynomial regression; Random forest; Optimization; Regularization

1 Introduction

Tunnel boring machines (TBMs) have been extensively used for tunnel construction in rock and soil (Mahmoodzadeh et al., 2021; Zakhem & El Naggar, 2020). The accurate prediction of TBM performance is crucial for estimating project schedules and selecting machine types and specifications (Gong & Zhao, 2009; Hassanpour et al., 2009). Thus, many empirical and semi-empirical formulations have been proposed to predict TBM performance. Empirical formulations have been developed using linear and non-linear regression analysis of the studied performance and influential factors (Benato & Oreste, 2015; Kahraman et al., 2003). However, the datasets tend to be collected from a given project; Thus, the application scope of the empirical formulations is typically limited to a specific project. One of the most typical empirical models is the Norwegian University of Science and Technology model. Based on the rock properties and the cutting geometry, the semi-empirical models theoretically analyze the rock fragmentation process and predict the TBM performance. The most representative semi-empirical model is the family of the Colorado School of Mines model (Ozdemir, 1977; Rostami, 1997; Rostami & Ozdemir, 1993; Yagiz, 2002), although specific model parameters

* Corresponding author at: Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.
E-mail addresses: doc.jie.yang@gmail.com, yang-jie.yang@polyu.edu.hk (J. Yang).

https://doi.org/10.1016/j.undsp.2021.04.003

must be determined from the numerical regression of laboratory tests. Moreover, discrepancies always exist between the predicted performance of the empirical and semi-empirical formulations and the measured results owing to the complex interaction between the TBM and ground.

Machine learning (ML), characterized by strong mapping capability, has recently been applied to address sophisticated and uncertain problems that cannot be well addressed through conventional methodologies(Atangana Njock et al., 2020; Chen et al., 2019a; Elbaz et al., 2021; Shen et al., 2020; Zhang et al., 2021a, 2020a, 2020d). Hence, as an alternative method, ML algorithms have been gradually employed to predict TBM performance. These algorithms include the adaptive neuro-fuzzy inference system (Ghasemi et al., 2014; Grima et al., 2000), artificial neural network (ANN) (Armaghani et al., 2017; Benardos & Kaliampakos, 2004; Koopialipoor et al., 2019b; Yagiz et al., 2009), extreme learning machine (Shao et al., 2013), support vector machine (SVM) (Fattahi & Babanouri, 2017; Mahdevari et al., 2014), decision tree (Jakubowski et al., 2017), the group method of data handling (Koopialipoor et al., 2019a), and the deep neural network (DNN) (Koopialipoor et al., 2019b). Nevertheless, previous researches have tended to neglect three important objective factors: the applicability of ML algorithms, size of the datasets, and number of input variables. ML-based models require considerable domain expertise to extract the features that transform the raw data into a suitable internal representation.

Deep learning (DL) allows a machine to accept raw data and automatically discovers the representations required for detection or classification (LeCun et al., 2015). DL further improves ML algorithms making full use of the raw data and mitigating the dependence on domain expert knowledge. However, DL algorithms, such as DNN, require numerous data to extract the useful features and tune the internal hyperparameters. Moreover, they merely derive the modeling properties to map the prescribed parameters rather than search for the relationships between the parameters; therefore the physical meaning behind these models cannot be determined. To the best of the authors' knowledge regarding, the published research, the maximum database used for developing a TBM performance prediction model involved only 1286 datasets (Koopialipoor et al., 2019b), which was considerably smaller than the size of the model used in the ML domain. If the number of datasets is insufficient, it is possible that an advanced and complex ML algorithm is not the best option for the development of a model for TBM performance prediction. Moreover, the previous ML-based models for TBM performance prediction were developed based on the prescribed input parameters determined by a domain expert. However, a method for the selection of the input parameters was not emphasized, leading to different parameter combinations in these previous ML-based models. Increasing the number of input parameter would increases the model complexity; Moreover, it is unreasonable to adopt excessive input variables for a small number of datasets. However, if the number of input parameters is considerably small, the generalization ability of the model could be reduced. Therefore, the tradeoff between model complexity and generalization ability must be considered.

Finally, the meaning behind the current ML-based models has not been revealed, making it difficult to provide useful information to guide the TBM tunneling process. Considering the applicability of ML algorithms and the state-of-the-art, genetic programming (GP) and ensemble algorithm random forest (RF) are more appropriate for predicting TBM performance than other complex ML algorithms. GP is an evolutionary computing algorithm that derives its modeling properties by generating an explicit and structured representation of the system (Koza, 1992). The GP-based models are represented by an explicit expression. Therefore, they are practically useful in engineering practice, and the computational cost of determining a GP-based model is considerably less than that of an ML-based model. An enhanced GP algorithm integrated with the conventional numerical regression method and GP symbolic regression method, termed evolutionary polynomial regression (EPR), has been proposed (Giustolisi & Savic, 2006) and has been successfully used in geotechnical engineering (Jin et al., 2019; Nassr et al., 2018; Yin et al., 2016). RF is simple, quick, easily parallelized, and resistant to outliers and noise (Nadi & Moradi, 2019; Zhang, 2019). Breiman (2001) demonstrated that RF provides useful internal estimates of error, strength, correlation, and variable importance. Owing to these advantages, RF has been extensively used in the engineering domain and outperforms other ML algorithms such as ANN and SVM (Ao et al., 2019; Chen et al., 2019b; Zhang et al., 2020c, 2021b, 2020e; Zhou et al., 2016).

This study aims to develop two models with the minimum number of possible input parameters using EPR and RF alternatively for TBM performance prediction characterized by the penetration rate (PR) and field penetration index (FPI). The developed models are compared with a multivariate linear regression (MLR)-based model. The performance of both the RF- and EPR-based prediction models are comprehensively investigated, including the model robustness for unknown datasets, interior relationships between input and output parameters, and variable importance. Based on these results, recommendations for simple, reliable, and practical models for TBM performance prediction in engineering practices are presented.

2 Methodology for TBM performance prediction

2.1 EPR modeling process

EPR is a hybrid regression approach that integrates the conventional numerical regression and GP symbolic regression methods (Giustolisi & Savic, 2006). The EPR modeling process can be primarily categorized into two steps,

namely, structure identification and parameter estimation. In the first step, the optimization algorithm (e.g., genetic algorithm, GA), is employed to search for an exponent matrix that determines the expression of the new transformed variable (Zhang et al., 2021b). Further, it should be noted that other stochastic optimization algorithms, such as differential evolution, particle swarm optimization, and backtracking search algorithm, can be used, rather than GA. For a given input matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_j, \cdots, \boldsymbol{x}_n]$ and exponent matrix $\boldsymbol{E} = [\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_j, \cdots, \boldsymbol{e}_n]$, the expression of the ith transformed term is obtained by

$$\boldsymbol{x}_{ti} = \boldsymbol{x}_1^{e_{i1}} \boldsymbol{x}_2^{e_{i2}} \cdots \boldsymbol{x}_j^{e_{ij}} \cdots \boldsymbol{x}_n^{e_{in}}. \tag{1}$$

In the second step, the constant coefficient for each transformed term is estimated by solving the least-squares linear problem after determining the transformed term matrix $\boldsymbol{X}_{ti} = [\boldsymbol{x}_{t1}, \boldsymbol{x}_{t2}, \cdots, \boldsymbol{x}_{ti}, \cdots, \boldsymbol{x}_{tm}]$. Thus, a typical EPR expression can be formulated as

$$y = \sum\nolimits_{i=1}^{m} F(\boldsymbol{x}_{ti}, a_i) + a_0. \tag{2}$$

Here y is the predicted output, m is the number of transformed terms, F is the EPR function, a_i is the constant coefficient for the ith transformed term, and a_0 is the optional bias coefficient.

Subsequently, the value of the objective function for optimization is calculated using the temporary EPR-based formulation (Eq. (2)). If the results satisfy the termination condition, the optimum EPR formulation is determined. Otherwise, the values of the exponent matrix are updated by the optimization algorithm. In this study, the GA optimization algorithm is used as it is a standard algorithm in a class of evolutionary algorithms and has been extensively used in the engineering domain (Yin et al., 2016, 2017). The search process of the optimum matrix $\boldsymbol{E}$ is terminated when the value of the GA objective function remains constant, and the recursive step achieves the maximum generation.

The 10-fold cross validation (CV) and ridge (L2) regularization methods are adopted to enhance the model robustness and avoid overfitting. In particular, the 10-fold CV method divides the original training set into 10 subsets. In each training round, nine random subsets are employed to train the model, and the remaining subset is used to validate the model performance. A total of 10 EPR models with the same exponent matrix and different constants are established. Each model yields an error that is represented by the sum of squared errors on the respective validation subset. The performance of the exponent matrix is evaluated based on the average error of the 10 models, rather than a prescribed single training set. Consequently, the effect of the randomness of the training set on the model performance is reduced. This enhances the robustness of the model. L2 regularization penalizes model complexity by adding a term to the objective function. Thus, the overfitting problem can be avoided. The objective function of the GA is defined as

$$J(\boldsymbol{\Theta}) = \frac{1}{10l} \sum_{j=1}^{10} \sum_{i=1}^{l} \left[y_{ji}^{\mathrm{p}} - y_{ji}^{\mathrm{m}} \right]^2 + \frac{\lambda}{2} ||\boldsymbol{\omega}||_2^2 , \quad ||\boldsymbol{\omega}||_2^2 = \boldsymbol{\omega}^{\mathrm{T}} \boldsymbol{\omega} , \tag{3}$$

where l is the number of datasets in each validation subset, $\boldsymbol{\Theta}$ is the exponent matrix, y_{ji}^{p} is the predicted ith output in the jth validation set, y_{ji}^{m} is the measured ith output in the jth validation set, λ is the regularization parameter, and $\boldsymbol{\omega}$ is the vector consisting of the elements in the exponent matrix. The GA parameters used in this study are listed in Table 1.

2.2 Random forest modeling process

RF is an ensemble learning algorithm with a collection of decision trees (Zhang et al., 2020b). Two powerful ML techniques, namely, bootstrap aggregating (Breiman, 1996) and random subspace (Ho, 1998) are adopted in RF, thereby enhancing its generalization ability. In RF, n bootstrap sets are first created by sampling and replacing N training examples from the training set; the number of samples and features in each bootstrap set are arbitrary (Zhang et al., 2019). Then, n decision trees grow using the respective bootstrap set. Each node tests a particular feature, and the leaves of a tree represent the output labels. Each decision tree forms a regression or classification space, and the final result of RF aggregates the outputs from n decision trees (Liaw & Wiener, 2002).

$$y = \frac{1}{n} \sum\nolimits_{i=1}^{n} y_i(\boldsymbol{x}), \tag{4}$$

where n is the total number of decision trees, y is the average output of all trees, and $y_i(\boldsymbol{x})$ is the prediction of a single decision tree.

2.3 Evaluation indicators

Three indicators (i.e., mean absolute error (MAE), mean absolute percentage error (MAPE), and *Pearson* correlation coefficient (R); see Eqs. (5)–(7)) are used to evaluate the model performance. Small values of MAE, MAPE, and a large value of R indicate the excellent performance of the model.

$$\mathrm{MAE} = \frac{1}{n} \sum\nolimits_{i=1}^{n} \left| y_i^{\mathrm{p}} - y_i^{\mathrm{m}} \right|, \tag{5}$$

Table 1
Values of the parameters in GA.

p_{cross}	p_{mutation}	N_{maxgen}	N_{sizepop}	Lower	Upper
0.7	0.1	500	20	–1	5

Note: p_{cross} and p_{mutation} are the probability of crossover and mutation, respectively; N_{maxgen} is the prescribed maximum number of generations; N_{sizepop} is the population size; and lower and upper are the lower and upper bounds of the element value in the exponent matrix, respectively.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^{\mathrm{p}} - y_i^{\mathrm{m}}}{y_i^{\mathrm{m}}}\right| \times 100\%, \tag{6}$$

$$R = \frac{\sum_{i=1}^{n}\left(y_i^{\mathrm{m}} - \bar{y}^{\mathrm{m}}\right)\left(y_i^{\mathrm{p}} - \bar{y}^{\mathrm{p}}\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i^{\mathrm{m}} - \bar{y}^{\mathrm{m}}\right)^2 \sum_{i=1}^{n}\left(y_i^{\mathrm{p}} - \bar{y}^{\mathrm{p}}\right)^2}}, \tag{7}$$

here, y_i^{m} is the measured ith output, y_i^{p} is the predicted ith output, $\bar{y}^{\mathrm{m}}$ is the mean value of the measured output, $\bar{y}^{\mathrm{p}}$ is the mean value of the predicted output, and n is the total number of datasets.

3 Database

The database used in this study was obtained from the Queens Water Tunnel No. 3 Stage 2 (7.5 km in length and 7.06 m in diameter), which has been widely used in predicting the performance of TBMs (Yagiz, 2008). This database contains 151 datasets within five different types of rock: (1) mafic-to-mesocratic gneiss, amphibolite, and schist (60/151, 39.74%), (2) mafic-to-mesocratic orthogneiss (31/151, 20.53%), (3) massive garnet amphibolite and larger mafic dikes (13/151, 8.61%), (4) granitoid (felsic) gneiss and orthogneiss (45/151, 29.80%), and (5) rhyodacite dike (2/151, 1.32%). Seven parameters were recorded, including the uniaxial compressive strength (UCS), Brazilian tensile strength (BTS), intact rock brittleness index (BI) defined based on the punch penetration test (Yagiz, 2009), distance between the planes of weakness (DPW), the angle between the tunnel axis, and the planes of weakness (α), PR, and FPI. The output parameters PR and FPI were calculated using Eqs. (8)–(10) (Hassanpour et al., 2009). To achieve a tradeoff between model complexity and accuracy, according to the feature selection investigation displayed in Fig. 1, the four parameters UCS, BI, DPW, and α were proven to be closely related to the TBM performance, and they were selected as the input parameters of the ML-based models to predict PR and FPI. Moreover, the values of the objective function converged to a constant (i.e., no overfitting issue occurred), indicating the effectiveness of the overfitting prevention strategy used in this study. The statistical characteristics of these six parameters are summarized in Table 2.

$$\mathrm{PR(m/h)} = \frac{\text{Boring length(m)}}{\text{Boring time (h)}}, \tag{8}$$

$$\mathrm{PR(mm/r)} = \frac{\mathrm{PR(m/h)} \times 1000}{\mathrm{RPM} \times 60}, \tag{9}$$

$$\mathrm{FPI(kN \cdot r/mm)} = \frac{F_{\mathrm{n}}\mathrm{(kN)}}{\mathrm{PR\ (mm/r)}}, \tag{10}$$

here, RPM is the cutter head revolution speed and F_{n} is the cutter load or normal force.

The scatter plots in Fig. 2 indicate the relationships between the input and output parameters. Evidently, the relationships of the four input parameters with PR are opposite to those with FPI, ensuring consistency with the actual condition. The increase in BI induces a large increase in PR and a decrease in FPI, whereas the increase in DPW decreased PR and increased FPI. A quadratic relationship exists between α and PR and a similar quadratic relationship between α and FPI. The maximum PR and minimum FPI were observed when α was appropriately 45°. When α was close to 0° and 90°, PR and FPI demonstrated a decreasing and increasing trend, respectively. The relationships between the output parameters and three input parameters mentioned previously are commonly observed and reported in engineering practice (Gong & Zhao, 2009; Hassanpour et al., 2009). Regarding the remaining input parameter UCS, no clear relationships between UCS and the two output parameters PR and FPI were observed in this project. Considering that UCS is an important parameter for PR and FPI prediction in other models (Hassanpour et al., 2009; Hassanpour et al., 2011; Kahraman et al., 2003), it was adopted as an input parameter to ensure that the ML-based models could extend their application scope. In summary, in this project, the most relevant parameter for PR was BI, followed by DPW, UCS, and α. The most relevant parameter for FPI was DPW, followed by BI, α, and UCS.

All datasets were mapped to the interval (−1, 1) using Eq. (11) before training the model to reduce the computational cost and eliminate the scale difference of the parameters. For parameter x, the normalized value is obtained from

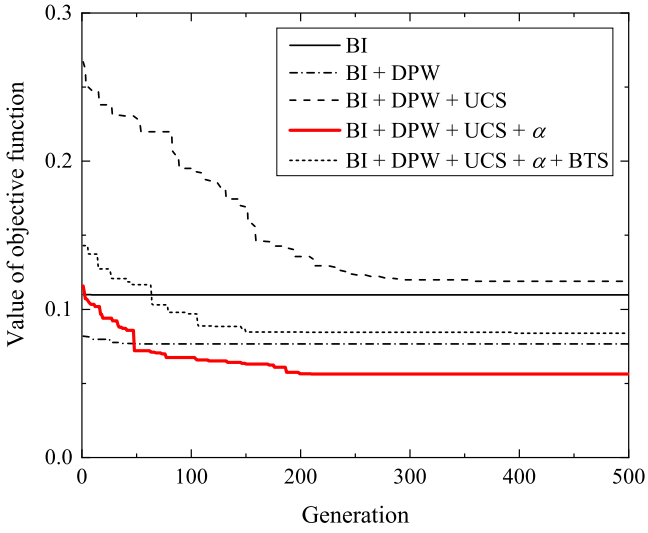


Fig. 1. Evolution of value of objective function on test set.

Table 2
Statistical characteristics of datasets.

Parameter	Max	Min	Mean	SD
UCS (MPa)	200	118	150.1	22.1
BI (kN/mm)	58.0	24.9	34.6	8.4
DPW (m)	2	0.05	1.02	0.64
α (°)	89	2	44.72	23.20
PR (m/h)	1.27	3.07	2.04	0.36
FPI (kN•r/mm)	147.7	45.2	81.8	19.5

Note: SD is the standard deviation.

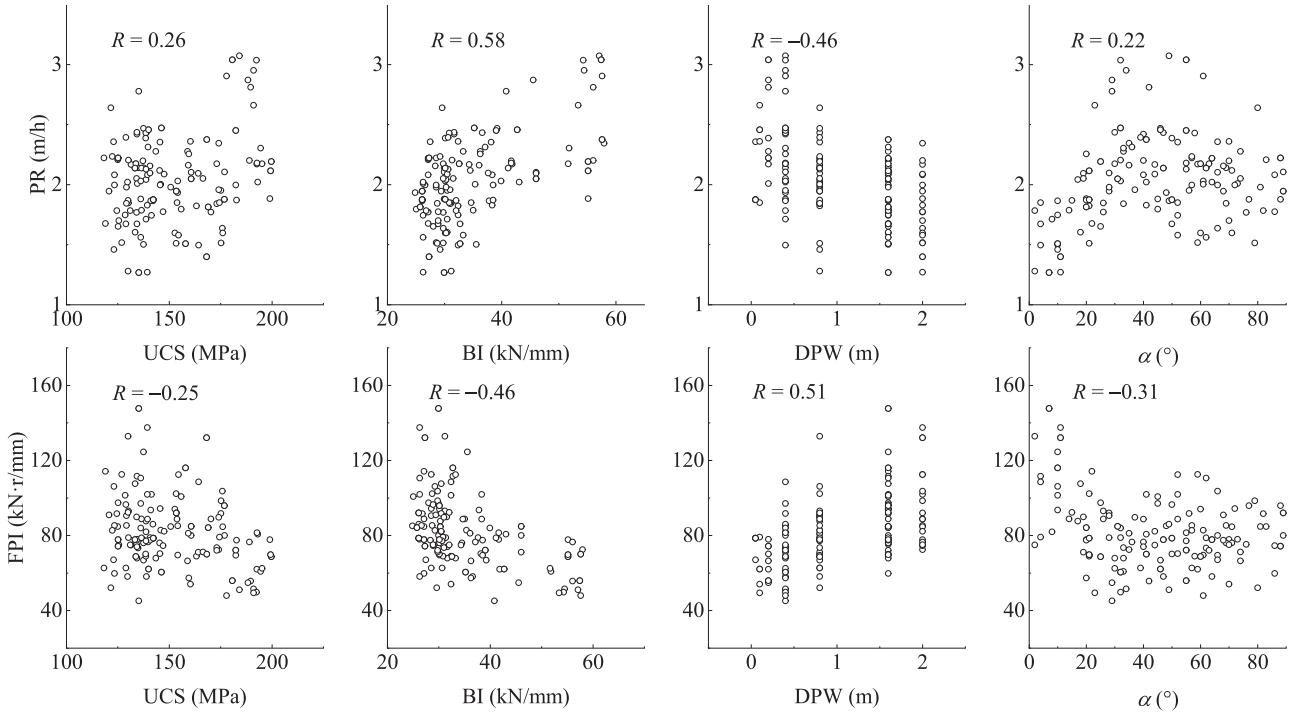


Fig. 2. Relationships between influential and output parameters.

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \left(\bar{x}_{\max} - \bar{x}_{\min}\right) + \bar{x}_{\min}, \tag{11}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the variable x and $\bar{x}_{\max}$ and $\bar{x}_{\min} = 1$ and -1, respectively. It was necessary to transform the final output into the original vector space.

4 Predicted results

4.1 MLR

MLR is a commonly used method to develop an empirical model for TBM performance prediction. In this study,

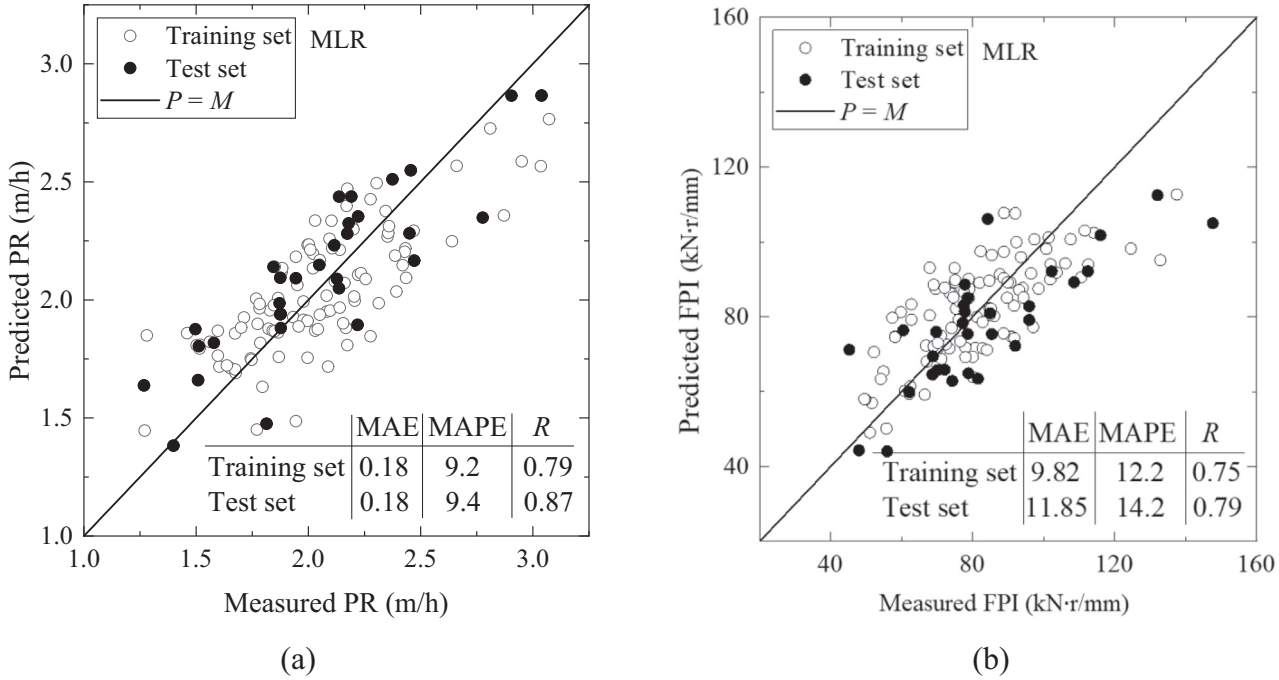
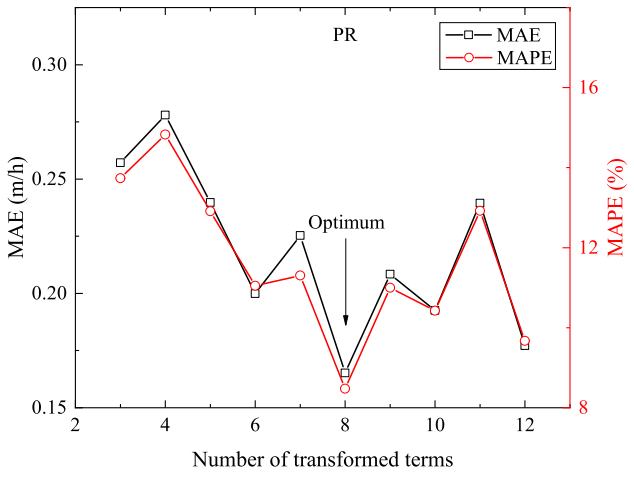
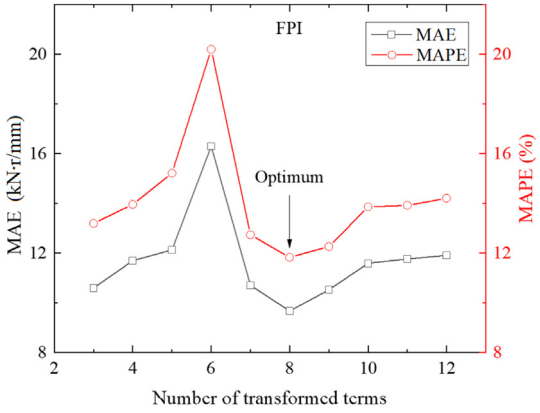


Fig. 3. Comparison between measured and predicted results using the MLR-based model for (a) PR and (b) FPI.



(a)



(b)

Fig. 4. Prediction performance of EPR-based models with different numbers of transformed terms for (a) PR and (b) FPI.

Table 3
Effect of λ value on the model performance on the test set.

λ	PR			FPI		
	MAE	MAPE	R	MAE	MAPE	R
0	0.171	9.238	0.89	9.666	11.827	0.88
0.0001	0.165	8.473	0.89	10.579	12.341	0.87
0.001	0.173	9.277	0.89	10.649	12.534	0.83
0.01	0.181	9.780	0.86	11.053	12.940	0.83
0.1	0.266	13.974	0.64	12.143	13.996	0.75

to compare with the two proposed ML-based models, the MLR-based model was employed to first predict PR and FPI. The predicted PR and FPI values obtained using the MLR-based model are presented in Fig. 3. Noticeably, the predicted PR and FPI were close to the measured results, and the prediction error for FPI was greater than that for PR. The discrepancy between the measured and predicted results of a large FPI, exceeding 140 kN·r/mm, for both the training and test sets is due to the scarcity of datasets. The corresponding empirical equations are expressed in Eqs. (12)–(13), which reveal the relationships between the input and output parameters.

$$\mathrm{PR} = -0.0033\mathrm{UCS} + 0.0289\mathrm{BI} - 0.2326\mathrm{DPW} + 0.0053\alpha + 1.5596 \quad (12)$$

$$\mathrm{FPI} = 0.0315\mathrm{UCS} - 1.0460\mathrm{BI} + 12.5863\mathrm{DPW} - 0.3329\alpha + 114.2132 \quad (13)$$

Noticeably, the input parameter relationships with PR and FPI were opposite. An increase in UCS and DPW led to a decrease in PR and increase in FPI, whereas an increase in BI and α resulted in an increase in PR and decrease in FPI. Except for the relationships of α with PR and FPI, the results were consistent with the commonly accepted relationships (Gong & Zhao, 2009; Hassanpour et al., 2009). The deficiency of the MLR method is that nonlinear relationships cannot be identified, thereby limiting its application scope and prediction accuracy.

4.2 EPR

The EPR modeling process must determine the optimum number of transformed terms and the value of the regularization parameter λ. Considering the tradeoff between model complexity and accuracy, 10 EPR-based models with the number of transformed terms in the range of 3–12 were developed to predict PR and FPI, respectively. The value of λ was in the range (0, 0.0001, 0.001, 0.01, 0.1), and the value of the element in the exponent matrix was in the range (–1, 0, 1, 2, 3, 4). The MAE and MAPE values obtained using these models are presented in Fig. 4. It should be noted that the value of the regularization parameter λ is optimum in these models. In summary, the values of MAE and MAPE decrease with an increase in the number of transformed terms. After MAE and MAPE reach a minimum, the rise in the number of

Table 4
Formulations for EPR-based PR and FPI prediction models.

	λ	NTT	Formulation
PR	0.0001	8	$\mathrm{PR} = 0.1164 - 0.2422\mathrm{UCS} - 0.2145\mathrm{BI}^3 + 1.5473\mathrm{BI}^3\mathrm{DPW} \times \mathrm{UCS} \times \alpha + 0.4558\mathrm{DPW} - 1.285\mathrm{BI}^3\mathrm{DPW} \times \mathrm{UCS}^3\alpha - 0.5008\mathrm{DPW}^2\alpha^2 + 0.4094\alpha^3 - 0.2129\frac{\mathrm{BI}^3\mathrm{DPW}\times\alpha}{\mathrm{UCS}}$
FPI	0	8	$\mathrm{FPI} = -0.3957 + 0.251\mathrm{DPW} + 0.0578\frac{\mathrm{UCS}^3\mathrm{BI}\times\alpha}{\mathrm{DPW}} - 0.151\mathrm{BI} \times \mathrm{DPW}^2 + 0.4448\mathrm{UCS}^2\mathrm{BI} \times \alpha - 0.4804\mathrm{UCS} \times \mathrm{BI} \times \mathrm{DPW} \times \alpha + 0.0577\frac{\mathrm{UCS}^3\mathrm{BI}\times\mathrm{DPW}}{\alpha} - 0.3577\alpha^3 - 0.8412\mathrm{BI}^3\alpha^2$

Note: NTT is the number of transformed terms.

transformed terms increases the MAE and MAPE. Table 3 displays the effect of λ on the model performance on the test set. By comparing the MAE and MAPE yielded by the models with different numbers of transformed terms, the optimum number of transformed terms was eight for both the PR and FPI prediction models. Further, the corresponding values of λ were 0.0001 and zero for the optimum PR and FPI prediction models, respectively. The formulations of the optimum EPR-based PR and FPI prediction models are presented in Table 4.

Figure 5 presents the predicted PR and FPI values obtained using the optimum EPR-based models. The EPR-based models outperformed the MLR-based models with reduced values of MAE and MAPE (approximately 10%) and greater values of R (approximately 0.9). By comparing Figs. 2(a) and (b), it can be observed that the EPR-based model improved the accuracy in predicting reduced values of PR (less than 2.5 m/h) and demonstrated superior performance in predicting greater values of FPI (greater than 120 kN•r/mm). The nonlinear combination of the input parameters in the EPR-based models enhanced the prediction accuracy. The explicit formulation is practically useful in engineering practice.

4.3 RF

In this study, the number of trees used in the RF ranged from 5 to 500 with an interval of five. The ranges of the four parameters were sufficiently large to determine the optimum combination of hyperparameters. In the RF-based models for PR and FPI prediction, the ultimate optimum hyperparameters were five. Figure 6 presents the predicted PR and FPI values obtained using the optimum RF-based models. Evidently, compared with MLR- and EPR-based models, the prediction performance of the RF-based models improved dramatically with the lowest values of MAE and MAPE, but the greatest value of R. The prediction accuracy was approximately double that of the MLR-based models. Based on Figs. 5(a) and (b), the RF-based models improved the accuracy in predicting large values of PR and FPI, and the results were close to the $P = M$ line (which denotes predictions equals to measurements). The results indicate that the RF performed well in identifying the nonlinear relationships of high-dimensional parameters, thereby improving the prediction accuracy.

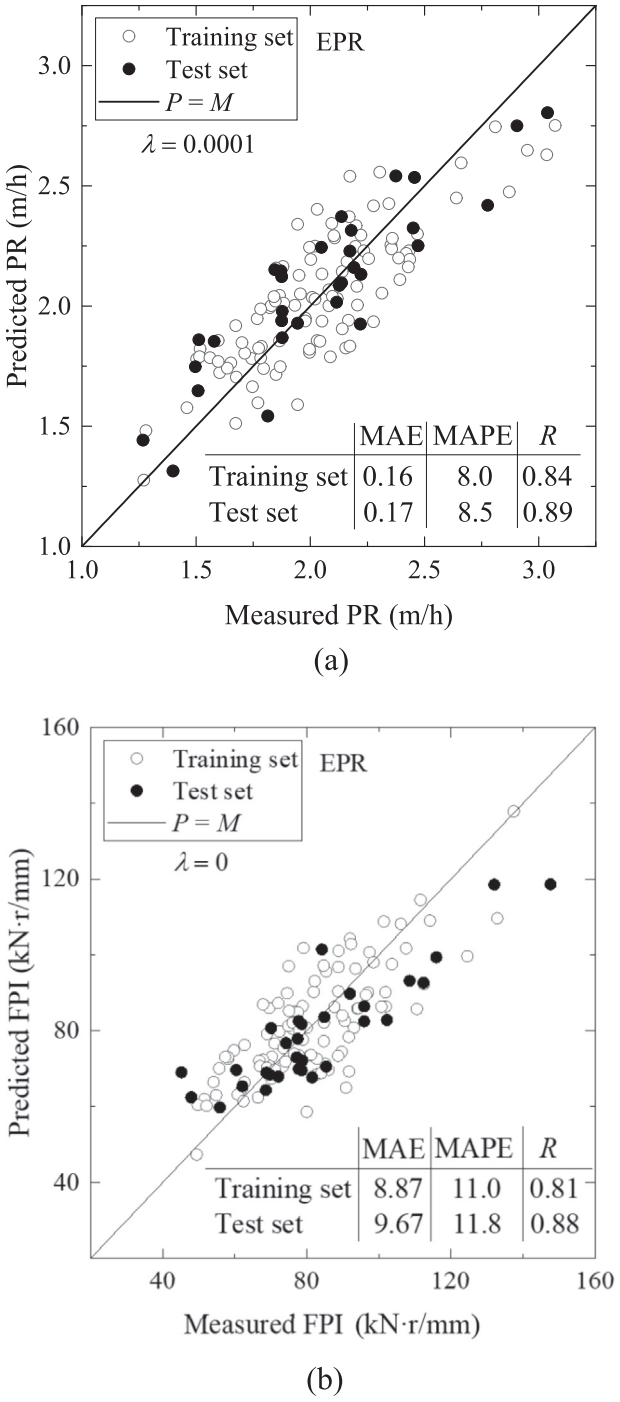


Fig. 5. Comparison between measured and predicted results using the EPR-based model for (a) PR and (b) FPI. ($P = M$ line denotes predictions equal to measurements.)

5 Performance comparison and further discussions

5.1 Performance comparison

Tables 5 and 6 summarize the values of the indicators for predicting PR and FPI using the optimum MLR-, EPR-, and RF-based models. It can be observed that the RF-based model demonstrated the best performance, followed by the EPR- and MLR-based models. Figure 7 presents the evolution of the predicted PR and FPI using the three methods. Compared with the MLR-based model, both EPR- and RF-based models demonstrated superior performance in capturing the evolution of PR and FPI. As observed in Fig. 7(b), the RF-based model could identify the large values of FPI (greater than 120 kN•r/mm) that occurred in Rock III, whereas both MLR- and EPR-based models underestimated the FPI in that range. All three methods produced the greatest prediction error

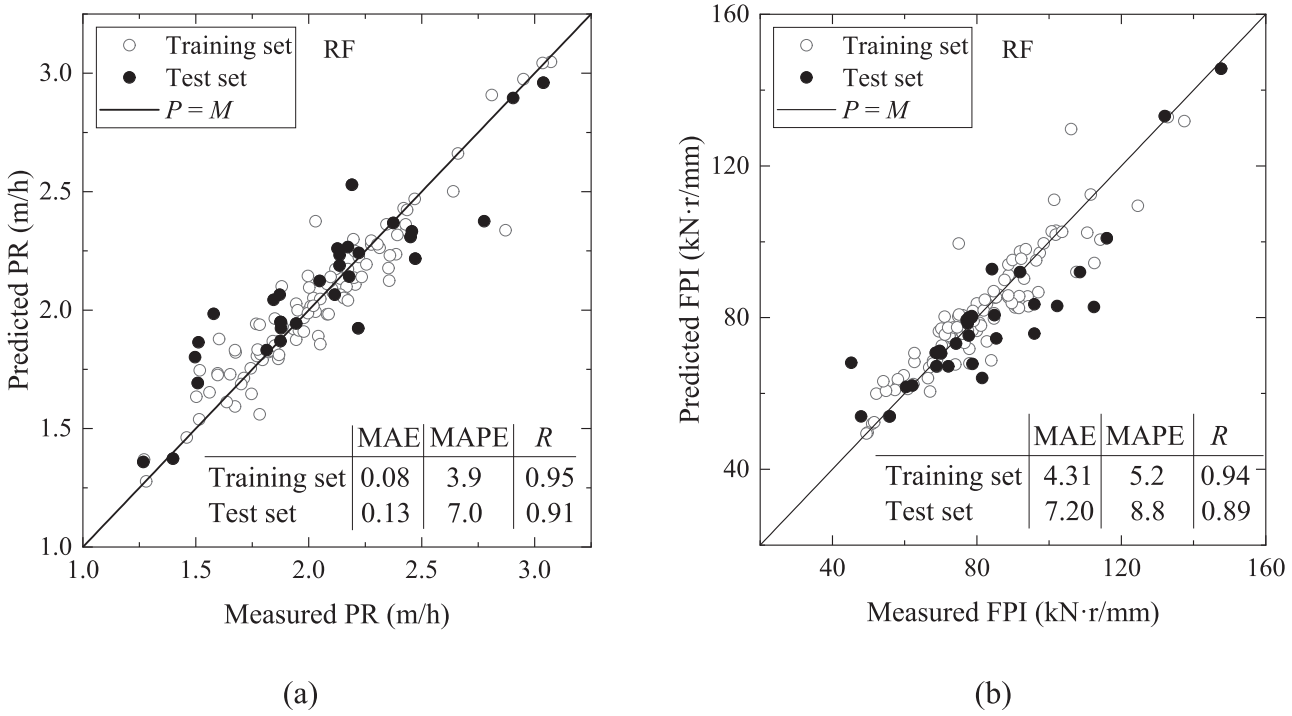


Fig. 6. Comparison between measured and predicted results using the RF-based model for (a) PR and (b) FPI.

in Rock I because Rock I consisted of mafic to mesocratic gneiss, amphibolite, and schist. The complexity of the rock components and insufficient datasets in this category made it difficult to predict PR and FPI. The results indicate that the simple ML algorithms (e.g., EPR and RF-based models with only four input parameters) are reliable and convenient for use in predicting PR and FPI in engineering practice.

5.2 Robustness investigation

Both EPR- and RF-based models can accurately predict PR and FPI, as mentioned previously, for a given tunnel project. The prediction performance of both models on unknown datasets is worthy of further exploration to examine their robustness. Hence, 10 000 synthetic datasets with four input parameters were generated wherein each input parameter complied with a normal distribution wherein the mean and SD values of each input parameter were identical to the values presented in Table 2. Based on these synthetic datasets, the distributions of the predicted PR and FPI using the EPR- and RF-based models are presented in Fig. 8. Noticeably, the predicted PR and FPI perfectly met the normal distribution with a coefficient of determination close to one. Moreover, the mean and SD values of the predicted PR and FPI using both ML-based models demonstrated acceptable agreement with the mean and SD values of the measured PR and FPI. To quantitatively evaluate robustness, Jin et al. (2019) proposed a robustness ratio index to define the model robustness.

$$x_R = \frac{N_r}{N_t} \times 100\%, \tag{14}$$

where x_R is robustness ratio, N_r is the number of datasets with the predicted values within a reasonable range, and N_t is the total number of datasets. The predicted output values should be within a reasonable range for a robust model. If the values of the input parameters are reasonable, a high value of the robustness ratio index indicates that the model was robust. As shown in Fig. 8, the robustness ratio attained 98.87% and 98.51% for the EPR-based PR and FPI prediction models, respectively. The robustness ratio of the RF-based PR and FPI models achieved 100% as the predicted results of RF are a linear combination of the output parameter values in the training set. Thus, the predicted results using the RF-based model were limited

Table 5
Indicator values for the PR prediction model.

Indicator	Training set			Test set		
	MAE	MAPE	R	MAE	MAPE	R
MLR	0.18	9.2	0.79	0.18	9.4	0.87
EPR	0.16	8.0	0.84	0.17	8.5	0.89
RF	0.08	3.9	0.95	0.13	7.0	0.91

Table 6
Indicator values for the FPI prediction model.

Indicator	Training set			Test set		
	MAE	MAPE	R	MAE	MAPE	R
MLR	9.82	12.2	0.75	11.85	14.2	0.79
EPR	8.87	11.0	0.81	9.67	11.8	0.88
RF	4.31	5.2	0.94	7.20	8.8	0.89

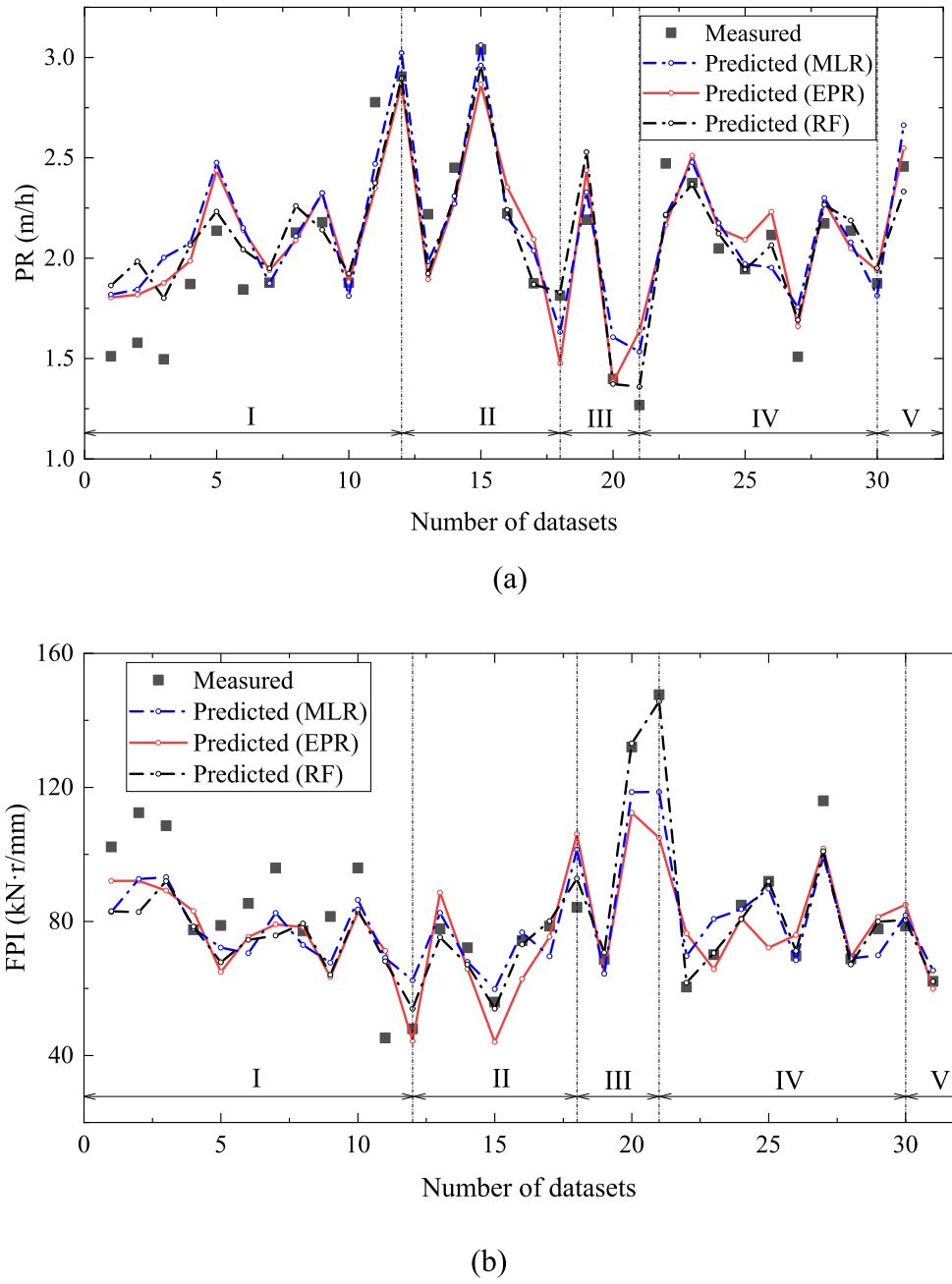


Fig. 7. Evolution of predicted results using three prediction models for (a) PR and (b) FPI.

to a fixed range. These results indicate that both EPR- and RF-based prediction models have significant predictive ability on unknown datasets. Therefore, they are sufficiently robust to extend their application scope.

5.3 Parametric analysis

To reveal the internal mechanism of the EPR- and RF-based models, a parametric analysis was performed. In each round, three input parameters were fixed to investigate the effect of each parameter on PR, and the parameter to study increased from its minimum to maximum. The relationships between the input and output parameters in the EPR-based PR and FPI prediction models are presented in Figs. 9(a) and 9(b), respectively. Noticeably, the four input parameters exhibited a smooth relationship with PR and FPI. Herein, the input parameters UCS, BI, and DPW exhibited an approximately monotonous relationship with PR and FPI, and α presented a quadratic relationship with PR and FPI. The parametric results in the EPR-based models complied with the relationships observed in the in situ datasets (Armetti et al., 2018; Gong & Zhao, 2009). Nevertheless, smooth relationships were not observed in the RF-based

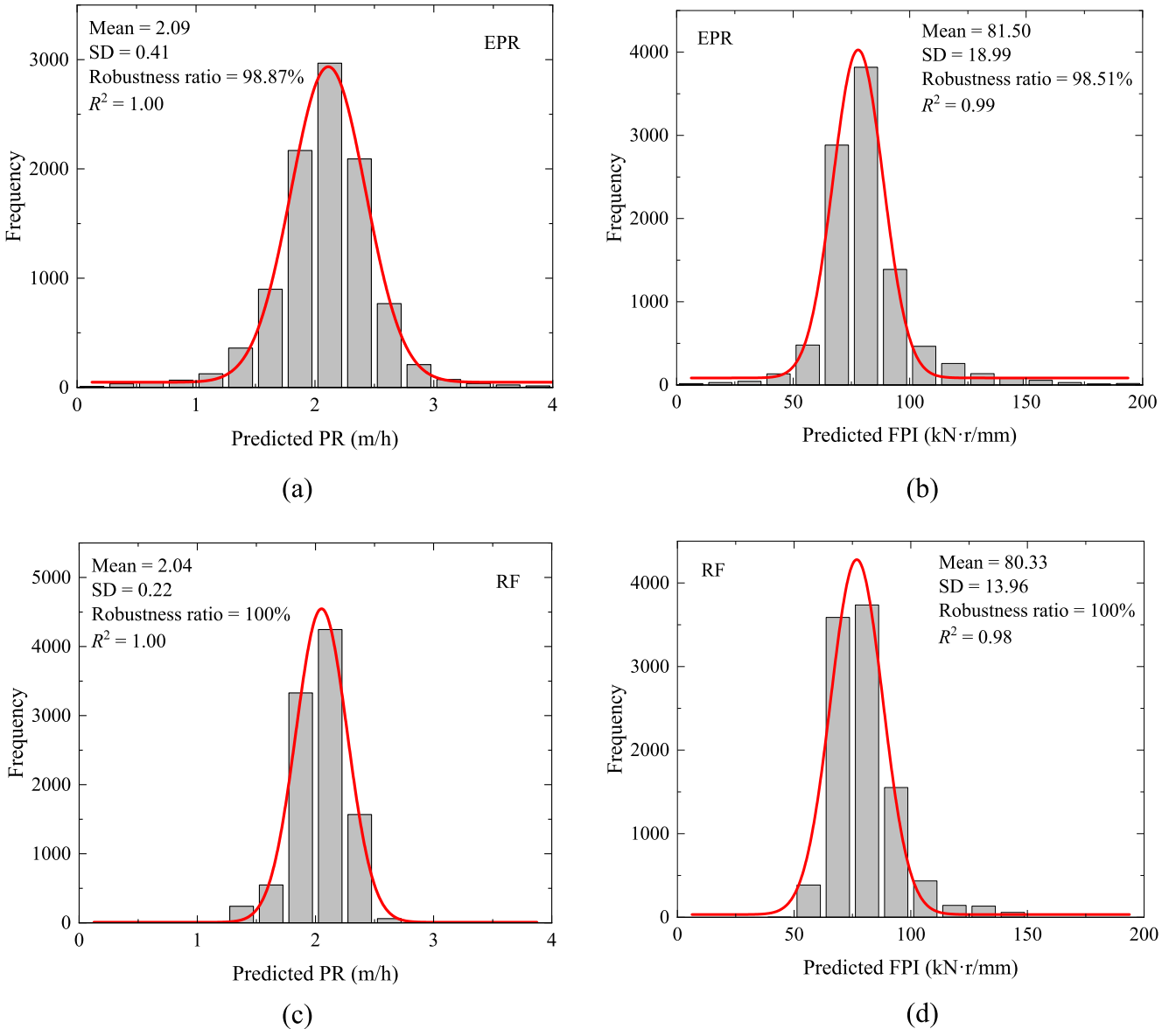


Fig. 8. Distribution of predicted (a) PR using EPR, (b) FPI using EPR, (c) PR using RF, and (d) FPI using RF.

models (see Figs. 9(c) and 9(d)) as the outputs of RF derived from the linear combination of the output parameters were discontinuous. Smoother relationships could be observed with the increasing size of the datasets. The general trends of PR and FPI with the variation of the four input parameters were consistent with the results in EPR, except for the relationships PR–UCS and FPI–α. In the RF-based model, the predicted PR decreased quickly to a minimum value with the increase in UCS. Thereafter, the predicted PR rebounded marginally. The predicted FPI decreased monotonously with an increase in α, rather than a quadratic relationship that appeared in the EPR-based model. Such relationships were different from the in situ investigation as the small size of datasets used in this study limited the capability of the RF-based models to accurately identify the relationships between the input and output parameters. Moreover, the relationships of BI and DPW with PR and FPI are nonlinear, whereas in the EPR-based models, these relationships are linear, which could explain why the RF-based models demonstrated greater accuracy in predicting PR and FPI. Overall, the accurate identification of the relationships between the input and output parameters indicates the excellent generalization ability of the EPR and RF-based models, thereby guaranteeing excellent prediction performance on both existing and unknown datasets.

5.4 Variable importance

The variable importance measure (VIM) in RF-based models can be determined internally. This section investi-

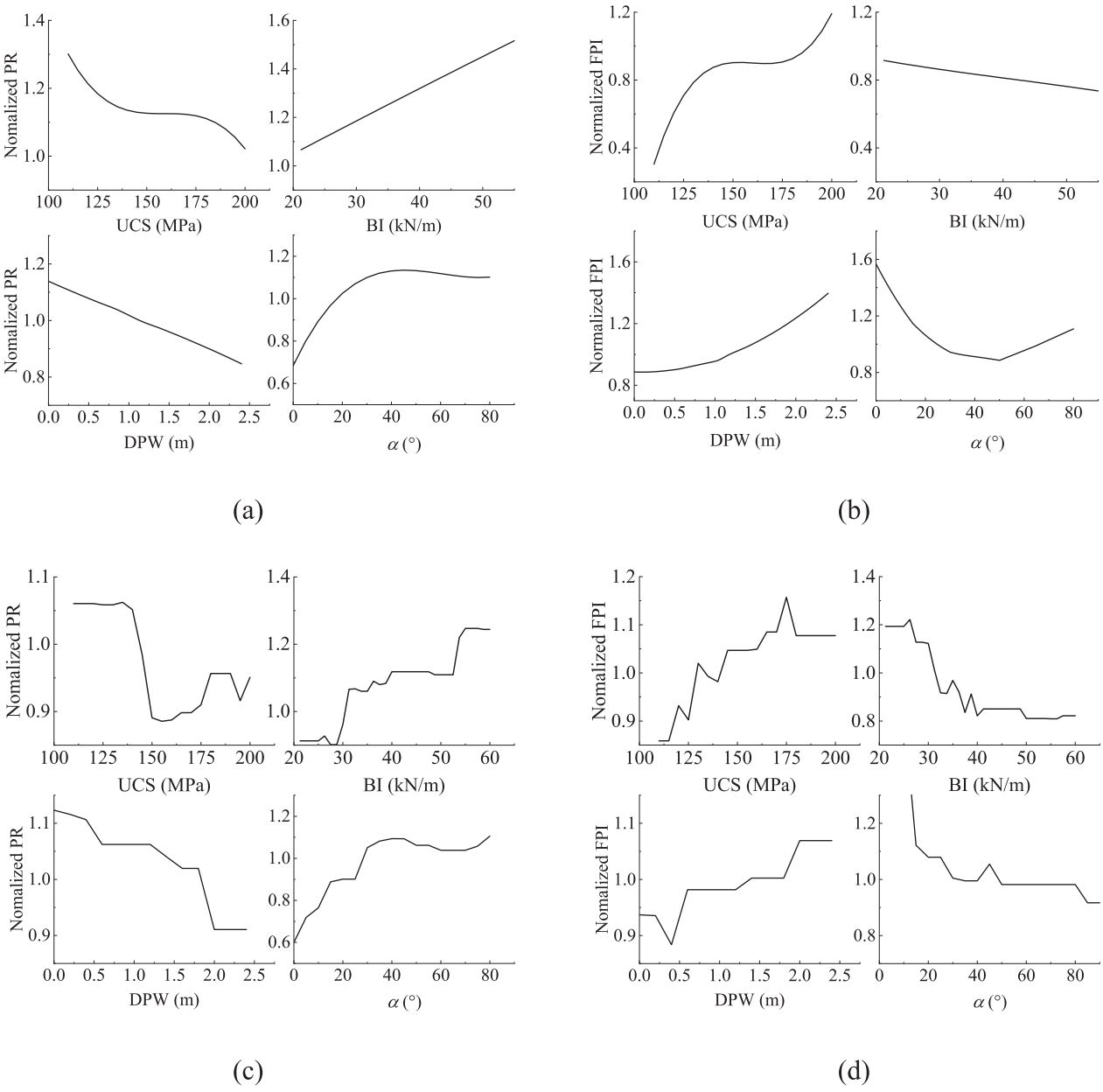


Fig. 9. Relationships between the input and (a) PR in the EPR-based model, (b) FPI in EPR-based model, (c) PR in RF-based model, and (d) FPI in RF-based model.

gates the VIM in the RF-based PR and FPI prediction models. The ultimate importance of the studied feature in the RF-based model is the average value of the VIM for all trees. On this basis, the mean and SD values of the variable importance in the RF-based PR and FPI prediction models are presented in Fig. 10. PR was considerably more sensitive to BI than to α, UCS, and DPW sequentially. The most important variables for predicting FPI was α, followed by BI, DPW, and UCS.

6 Conclusions

Using two ML algorithms (i.e., EPR and RF), this study developed two novel models for TBM performance prediction characterized by the PR and FPI. RF has recently exhibited excellent performance in modeling geotechnical issues. However, it has not been applied to predict PR and FPI. Thus, the performance of RF on this issue deserves further investigation. Unlike the RF-based model, which is an implicit model, EPR is an explicit formulation that integrates L2 regularization and *k*-fold CV to prevent the overfitting issue and enhance the model robustness. A comparison between EPR- and RF-based models is necessary to understand the tradeoff between accuracy and applicability. Only four input parameters were used: the UCS, intact rock BI, DPW, and α. The developed models were compared with an MLR-based model. The performance of both RF- and EPR-based pre-

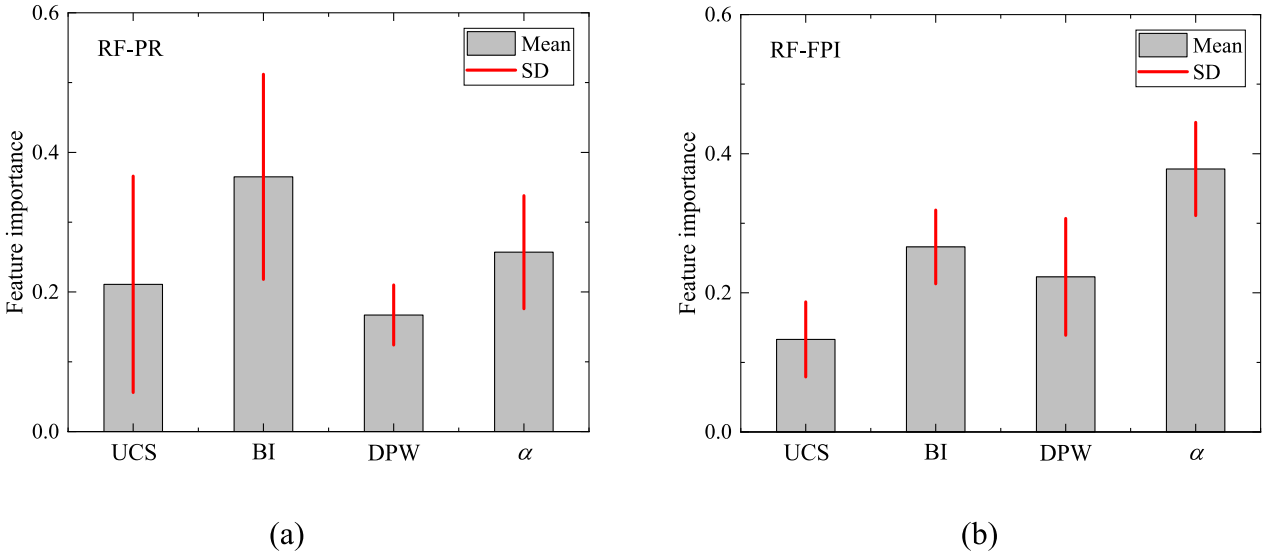


Fig. 10. Variable importance in the RF model for (a) PR and (b) FPI.

diction models were comprehensively investigated including the model robustness for unknown datasets, interior relationships between the input and output parameters, and variable importance. The following conclusions can be drawn:

(1) Unlike the conventional numerical regression method MLR, both EPR- and RF-based models demonstrated superior performance in predicting PR and FPI. The RF-based model had the greatest prediction accuracy, particularly in identifying outliers. However, its prediction range was more limited than that in the training set. The EPR-based model was more convenient to use by in situ engineers owing to its explicit expression.
(2) Both EPR- and RF-based models could accurately identify the relationships between the input and output parameters, thereby guaranteeing their excellent generalization ability and prediction performance on unknown datasets.
(3) Variable importance measured internally by the RF-based models indicated that PR is considerably more sensitive to BI than to α, UCS, and DPW sequentially, whereas the most important variable for predicting FPI is α, followed by BI, DPW, and UCS.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was financially supported by the research project of Zhongtian Construction Group Co. Ltd. (Grant No. ZTCG-GDJTYJS-JSFW-2020002). The authors also would like to thank Mr. Pin ZHANG from The Hong Kong Polytechnic University, who conducted a lot of numerical work for this study.

References

Ao, Y., Li, H., Zhu, L., Ali, S., & Yang, Z. (2019). Identifying channel sand-body from multiple seismic attributes with an improved random forest algorithm. *Journal of Petroleum Science and Engineering, 173*, 781–792.
Armaghani, D. J., Mohamad, E. T., Narayanasamy, M. S., Narita, N., & Yagiz, S. (2017). Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. *Tunnelling and Underground Space Technology, 63*, 29–43.
Armetti, G., Migliazza, M. R., Ferrari, F., Berti, A., & Padovese, P. (2018). Geological and mechanical rock mass conditions for TBM performance prediction. The case of "La Maddalena" exploratory tunnel, Chiomonte (Italy). *Tunnelling and Underground Space Technology, 77*, 115–126.
Atangana Njock, P. G., Shen, S.-L., Zhou, A., & Lyu, H.-M. (2020). Evaluation of soil liquefaction using AI technology incorporating a coupled ENN / t-SNE model. *Soil Dynamics and Earthquake Engineering, 130*, 105988.
Benardos, A. G., & Kaliampakos, D. C. (2004). Modelling TBM performance with artificial neural networks. *Tunnelling and Underground Space Technology, 19*(6), 597–605.
Benato, A., & Oreste, P. (2015). Prediction of penetration per revolution in TBM tunneling as a function of intact rock and rock mass characteristics. *International Journal of Rock Mechanics and Mining Sciences, 74*, 119–127.
Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.
Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
Chen, R. P., Zhang, P., Kang, X., Zhong, Z. Q., Liu, Y., & Wu, H. N. (2019a). Prediction of maximum surface settlement caused by EPB shield tunneling with ANN methods. *Soils and Foundations, 59*(2), 284–295.
Chen, R. P., Zhang, P., Wu, H. N., Wang, Z. T., & Zhong, Z. Q. (2019b). Prediction of shield tunneling-induced ground settlement using machine learning techniques. *Frontiers of Structural and Civil Engineering, 13*(6), 1363–1378.
Elbaz, K., Shen, S. L., Zhou, A. N., Yin, Z. Y., & Lyu, H. M. (2021). Prediction of disc cutter life during shield tunneling with AI via incorporation of a genetic algorithm into a GMDH-type neural network. *Engineering, 7*(2), 238–251.
Fattahi, H., & Babanouri, N. (2017). Applying optimized support vector regression models for prediction of tunnel boring machine

performance. *Geotechnical and Geological Engineering, 35*(5), 2205–2217.
Ghasemi, E., Yagiz, S., & Ataei, M. (2014). Predicting penetration rate of hard rock tunnel boring machine using fuzzy logic. *Bulletin of Engineering Geology and the Environment, 73*(1), 23–35.
Giustolisi, O., & Savic, D. A. (2006). A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics, 8*(4), 235–237.
Gong, Q. M., & Zhao, J. (2009). Development of a rock mass characteristics model for TBM penetration rate prediction. *International Journal of Rock Mechanics & Mining Sciences, 46*(1), 8–18.
Grima, M. A., Bruines, P. A., & Verhoef, P. N. W. (2000). Modeling tunnel boring machine performance by Neuro-Fuzzy methods. *Tunnelling and Underground Space Technology, 15*(3), 259–269.
Hassanpour, J., Rostami, J., Khamehchiyan, M., Bruland, A., & Tavakoli, H. R. (2009). TBM performance analysis in pyroclastic rocks: A case history of Karaj water conveyance tunnel. *Rock Mechanics and Rock Engineering, 43*(4), 427–445.
Hassanpour, J., Rostami, J., & Zhao, J. (2011). A new hard rock TBM performance prediction model for project planning. *Tunnelling and Underground Space Technology, 26*(5), 595–603.
Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(8), 832–844.
Jakubowski, J., Stypulkowski, J. B., & Bernardeau, F. G. (2017). Multivariate linear regression and CART regression analysis of TBM performance at Abu Hamour Phase-I tunnel. *Archives of Mining Sciences, 62*(4), 825–841.
Jin, Y. F., Yin, Z. Y., Zhou, W. H., Yin, J. H., & Shao, J. F. (2019). A single-objective EPR based model for creep index of soft clays considering L2 regularization. *Engineering Geology, 248*, 242–255.
Kahraman, S., Bilgin, N., & Feridunoglu, C. (2003). Dominant rock properties affecting the penetration rate of percussive drills. *International Journal of Rock Mechanics and Mining Sciences, 40*(5), 711–723.
Koopialipoor, M., Nikouei, S. S., Marto, A., Fahimifar, A., Jahed Armaghani, D., & Mohamad, E. T. (2019a). Predicting tunnel boring machine performance through a new model based on the group method of data handling. *Bulletin of Engineering Geology and the Environment, 78*(5), 3799–3813.
Koopialipoor, M., Tootoonchi, H., Jahed Armaghani, D., Tonnizam Mohamad, E., & Hedayat, A. (2019b). Application of deep neural networks in predicting the penetration rate of tunnel boring machines. *Bulletin of Engineering Geology and the Environment, 78*(8), 6347–6360.
Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. MIT Press.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.
Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News, 23*(23), 18–21.
Mahdevari, S., Shahriar, K., Yagiz, S., & Akbarpour Shirazi, M. (2014). A support vector regression model for predicting tunnel boring machine penetration rates. *International Journal of Rock Mechanics and Mining Sciences, 72*, 214–229.
Mahmoodzadeh, A., Mohammadi, M., Farid Hama Ali, H., Nariman Abdulhamid, S., Hashim Ibrahim, H., & M Gharrib Noori, K. (2021). Dynamic prediction models of rock quality designation in tunneling projects. Transportation Geotechnics, 27, 100497.
Nadi, A., & Moradi, H. (2019). Increasing the views and reducing the depth in random forest. *Expert Systems with Applications, 138*, 112801.
Nassr, A., Esmaeili-Falak, M., Katebi, H., & Javadi, A. (2018). A new approach to modeling the behavior of frozen soils. *Engineering Geology, 246*, 82–90.
Ozdemir, L. (1977). *Development of theoretical equations for predicting tunnelling borability* Colorado School of Mines Golden. USA: Colorado.
Rostami, J. (1997). *Development of a Force Estimation Model for Rock Fragmentation with Disc Cutters Through Theoretical Modeling and Physical Measurement of Crushed Zone Pressure*. Colorado, USA: Colorado School of Mines Golden.
Rostami, J., & Ozdemir, L. (1993). A new model for performance prediction of hard rock TBM. Proceedings of the Rapid Excavation and Tunneling Conference, Boston, 13–17 June 1993, 793–793.
Shao, C. J., Li, X. L., & Su, H. Y. (2013). Performance prediction of hard rock TBM based on extreme learning machine. International Conference on Intelligent Robotics and Applications, Berlin Heidelberg. pp 409–416.
Shen, S.-L., Atangana Njock, P. G., Zhou, A., & Lyu, H.-M. (2020). Dynamic prediction of jet grouted column diameter in soft soil using Bi-LSTM deep learning. *Acta Geotechnica, 16*(16), 303–315.
Yagiz, S. (2002). Development of rock fracture and brittleness indices to quantify the effects of rock mass features and toughness in the CSM Model basic penetration for hard rock tunneling machines Colorado School of Mines Golden, Colorado, USA.
Yagiz, S. (2008). Utilizing rock mass properties for predicting TBM performance in hard rock condition. *Tunnelling and Underground Space Technology, 23*(3), 326–339.
Yagiz, S. (2009). Assessment of brittleness using rock strength and density with punch penetration test. *Tunnelling and Underground Space Technology, 24*(1), 66–74.
Yagiz, S., Gokceoglu, C., Sezer, E., & Iplikci, S. (2009). Application of two non-linear prediction tools to the estimation of tunnel boring machine performance. *Engineering Applications of Artificial Intelligence, 22*(4/5), 808–814.
Yin, Z. Y., Jin, Y. F., Huang, H. W., & Shen, S. L. (2016). Evolutionary polynomial regression based modelling of clay compressibility using an enhanced hybrid real-coded genetic algorithm. *Engineering Geology, 210*, 158–167.
Yin, Z. Y., Jin, Y. F., Shen, S. L., & Huang, H. W. (2017). An efficient optimization method for identifying parameters of soft structured clay by an enhanced genetic algorithm and elastic–viscoplastic model. *Acta Geotechnica, 12*(4), 849–867.
Zakhem, A. M., & El Naggar, H. (2020). Three-dimensional investigation of how newly constructed buildings supported on raft foundations affect pre-existing tunnels. *Transportation Geotechnics, 22*, 100324.
Zhang, K., Lyu, H. M., Shen, S. L., Zhou, A., & Yin, Z. Y. (2020a). Evolutionary hybrid neural network approach to predict shield tunneling-induced ground settlements. *Tunnelling and Underground Space Technology, 106*, 103594.
Zhang, P. (2019). A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model. *Applied Soft Computing, 85*, 105859.
Zhang, P., Chen, R. P., & Wu, H. N. (2019). Real-time analysis and regulation of EPB shield steering using Random Forest. *Automation in Construction, 106*, 102860.
Zhang, P., Wu, H. N., Chen, R. P., Dai, T., Meng, F. Y., & Wang, H. B. (2020b). A critical evaluation of machine learning and deep learning in shield-ground interaction prediction. *Tunnelling and Underground Space Technology, 106*, 103593.
Zhang, P., Wu, H. N., Chen, R. P., & Chan, T. H. T. (2020c). Hybrid meta-heuristic and machine learning algorithms for tunneling-induced settlement prediction: A comparative study. *Tunnelling and Underground Space Technology, 99*, 103383.
Zhang, P., Yin, Z. Y., & Jin, Y. F. (2021a). State-of-the-art review of machine learning applications in constitutive modeling of soils. *Archives of Computational Methods in Engineering*, 1–26. https://doi.org/10.1007/s11831-020-09524-z, in press.
Zhang, P., Yin, Z. Y., Zheng, Y., & Gao, F. P. (2020d). A LSTM surrogate modelling approach for caisson foundations. *Ocean Engineering, 204*, 107263.
Zhang, P., Yin, Z. Y., Jin, Y. F., Chan, T., & Gao, F. P. (2021b). Intelligent modelling of clay compressibility using hybrid meta-heuristic and machine learning algorithms. *Geoscience Frontiers, 12*(1), 441–452.
Zhang, P., Yin, Z. Y., Jin, Y. F., & Chan, T. H. T. (2020e). A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. *Engineering Geology, 265*, 105328.
Zhou, J., Li, X., & Mitri, H. S. (2016). Classification of rockburst in underground projects: Comparison of ten supervised learning methods. *Journal of Computing in Civil Engineering, 30*(5), 04016003.