




Article

Physics-Informed Ensemble Machine Learning Framework for Improved Prediction of Tunneling-Induced Short- and Long-Term Ground Settlement

Linan Liu ^{1,†} , Wendy Zhou ^{1,*}  and Marte Gutierrez ² 

¹ Department of Geology & Geological Engineering, Colorado School of Mines, Golden, CO 80401, USA; linanliu@cugb.edu.cn

² Department of Civil & Environmental Engineering, Colorado School of Mines, Golden, CO 80401, USA; mgutierr@mines.edu

* Correspondence: wzhou@mines.edu

† Current address: School of Engineering and Technology, China University of Geosciences (Beijing), Beijing 100083, China.

Abstract: Machine learning (ML), one of the AI techniques, has been used in geotechnical engineering for over three decades, resulting in more than 600 peer-reviewed papers. However, AI applications in geotechnical engineering are significantly lagging compared with other fields. One of the reasons for the lagging is that hyperparameters used in many AI techniques need physical meaning in geotechnical applications. This paper focuses on widening the applications of ML in predicting tunneling-induced short- and long-term ground settlement and optimizing ML architectures considering their interpretability and ability to provide physically consistent results. Informed by the underlying physics knowledge, tunneling-induced ground settlement is divided into long-term and short-term settlements since different mechanisms and influencing parameters contribute to these two deformation types. Based on the above considerations, this paper introduces a physics-informed ensemble machine learning (PIML) framework to strengthen the connection between ML techniques and physics theories, followed by identifying/utilizing different sets of parameters for effectively predicting short- and long-term tunneling-induced settlements, respectively. Together with in situ observations and experimental lab results, parameters obtained from physics equations are set as inputs for the ML models. Results show that the proposed PIML framework effectively predicts tunneling-induced ground movements, with a predicting accuracy above 0.8. Additionally, parametric studies of variable significance and comparisons among different ML designs reveal that in situ observed dynamic parameters, for instance tunnel face and monitoring points (DTM), gap parameter, and tunnel depth, are essential in predicting tunneling-induced short-term settlement, while predicting long-term settlements largely depends on features, such as tunnel depth, volume compressibility, and excess pore pressure, derived from physics theories.

Keywords: ground settlement; tunneling; ensemble machine learning; physics-based model; feature importance; physics-informed machine learning



Citation: Liu, L.; Zhou, W.; Gutierrez, M. Physics-Informed Ensemble Machine Learning Framework for Improved Prediction of Tunneling-Induced Short- and Long-Term Ground Settlement. *Sustainability* **2023**, *15*, 11074. <https://doi.org/10.3390/su151411074>

Academic Editor: Antonio Miguel Martínez-Graña

Received: 28 May 2023

Revised: 29 June 2023

Accepted: 6 July 2023

Published: 15 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Settlement is of great concern with soft ground tunneling in urban settings where above-ground structures and facilities may be negatively impacted. Shield tunnel excavation inevitably leads to surface settlement since the ground is disturbed, particularly for shallow tunnels. Shield tunnel excavation provides an opening into which the soil can deform [1]. The soil movements are typically classified into ground loss (short-term) and consolidation (long-term) settlements. In other words, ground loss induces short-term settlement, while soil consolidation results in long-term settlement.

A great deal of research effort has been devoted to the study of settlements caused by tunneling through soft ground. Machine learning (ML) techniques are becoming a promising alternative to predict tunneling-induced ground deformation based on ground observations and measurements besides remote sensing monitoring [2–6]. However, ML applications in predicting tunneling-induced ground behavior are less common due to the limitation of available data size, ML interpretability, and physics understanding of the mechanisms of ground deformation.

Some ML methods have been applied to predict tunneling-induced ground settlements, such as neural networks [7], support vector machines [8], random forest [9], and other tree-based ML methods [10]. When applying these ML techniques, parameters related to geological conditions, tunnel geometry, and Tunnel Boring Machine (TBM) are typically collected before or during tunnel excavation [8]. Furthermore, it is recommended to collect as many monitored parameters related to the tunneling process as possible [11]. Immediate short-term ground deformation induced by tunneling can, therefore, be predicted by inputting these parameters into various ML models. However, the generalization of ML application from tunneling-induced ground movements might not be robust since input parameter selection largely depends on the data availability.

With available data of adequate quality and quantity, ML models are typically promising for scientific problems with physics processes that are even not fully understood. Although the quantitative relationship between contributing parameters and ground movements is not well defined in tunneling-induced ground settlements, ML algorithms may be successfully applied in some cases, contingent on data quantity and quality. However, the available dataset sizes of in-situ measured parameters, typically ranging from 10s to 100s for tunnel projects, are generally insufficient in using ML models [9,12–14]. Consequently, a growing consensus is that complex engineering problems require novel methodologies to integrate advanced ML algorithms with physics-based modeling approaches [15,16]. Physics-informed model initialization and framework design are two widely recognized approaches combining ML models and physics laws to predict surface settlement [13]. Analytical and numerical methods account for tunneling-induced ground movements using simplified physics theories, such as consolidation, constitutive, and ground loss models [17–19]. Physics-based models are widely applied, accepted, and demonstrated in practice [20–23], even though these models have their limitations or pre-assumptions [24]. Widening ML applications motivate us to provide physically consistent results and address the interpretability of such statistical inference approaches.

While ML has revolutionized modeling and decision-making in many fields, its challenge lies in the interpretability of the ML model and in providing physically consistent results. Additionally, the generalization of applying ML models to tunneling-induced ground deformation is not robust since inputs are determined mainly on or limited by the data availability, even though ML might have good generalization in terms of algorithms themselves. This paper proposes a Physics-Informed Ensemble Machine Learning (PIML) framework to seek answers of these questions.

The proposed PIML framework introduces domain-specific knowledge to ML models, aims to increase the generalization of ML-based methods to predict tunneling-induced ground movements, improves the interpretability of the ML models, and bridges the gap between physics and ML methods. The proposed PIML framework uses simplified theoretical models to inform the learning module. Inputs of the PIML framework are determined by considering physics-based models, and observed or measured factors. Short-term and long-term settlements are separately learned and predicted due to their different deformation mechanisms. The ML algorithm in the PIML framework is the ensemble voting regression technique, which combines multiple learners to make an optimal collective prediction. Finally, the model performance is evaluated by the coefficient of determination (R^2). Parametric studies for short- and long-term settlements are conducted using the permutation importance feature method [14]. Different ML model designs with and

without physics considerations are compared and contrasted to highlight the importance of informing the ML models with physical theories.

2. Methodology

2.1. Physics-Based Models

Ground movement is related to ground loss in the proximity of a tunnel, i.e., the amount of soil excavated in excess of a designed excavation during the process. Based on field case studies, the ground loss typically comes from four parts along with the tunnel advances in soil mass, including (i) face loss, (ii) shield loss, (iii) losses due to the erection of the shield, and (iv) time-dependent losses. Correspondingly, the excavation process can be divided into four stages, i.e., face approach, shield passage, tail departure, and soil consolidation [25,26]. It is challenging to ascertain during which stage the TBM is accurately passing unless settlements are monitored hourly. The shield usually takes only a few hours to pass a section during excavation [27]. In practice, ground deformations induced by the first three stages are typically combined as “ground loss” settlements, providing a convenient way to analyze a highly complex phenomenon. Therefore, tunneling-induced ground settlements can be classified into ground loss (short-term) and consolidation (long-term) settlements, as illustrated in Figure 1 [26,28,29]. Short-term settlement is caused by the ground losses in the undrained state immediately after the passing of the tunnel. In contrast, long-term settlement is induced by soil consolidation due to time-dependent behavior, such as volume compressibility and excess pore pressure [18].

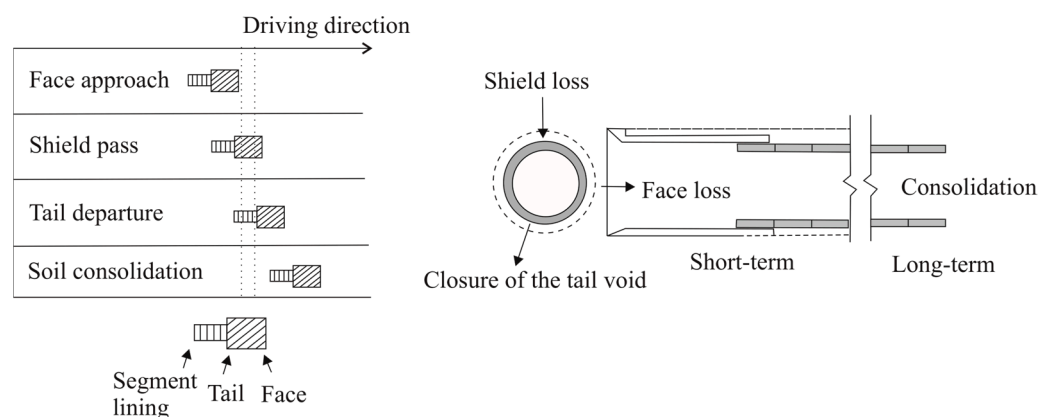


Figure 1. Illustration of short-term and long-term displacements caused by shield tunneling (modified from [26,28,29]).

Two physics-based methods, i.e., equivalent ground loss and Terzaghi’s consolidation theory, are applied to inform ML design. These two physics-based models have been demonstrated and validated in historical shield tunnel cases [30,31]. Other commonly-used approaches, such as the empirical formula model, stochastic medium theory, and analytical model, are available to analyze tunnel-induced ground settlements. However, these additional methods are data-driven and statistical. Hence parameters lacked physics-based meanings and were not adopted in this study.

2.1.1. Calculating Short-Term Settlement Using Equivalent Ground Loss Model

As many case studies [18,20,32] demonstrated, the equivalent ground loss model is a proving-working analytical method for estimating tunneling-induced short-term ground movements. In this model, the ground loss is defined in terms of the gap parameter (mm) as the magnitude of the equivalent two-dimensional void formed around the tunnel [33]. The gap parameter is the sum of deformation (mm) due to the tail void of the shield machine and radial deformation (mm) at the tunnel face. It is considered the maximum settlement at the tunnel crown and expressed as Equation (1), neglecting the consolidation of clayey soil. Specifically, three combined effects are considered in the gap model, including

(1) the physical gap related to the tunneling machine, shield, and lining geometry, (2) the over-excavation of soil around the periphery of the tunnel shield, and (3) three-dimensional elastoplastic ground deformation at the tunnel face [20].

$$g = G'_p + U_{3D}^* + \omega \quad (1)$$

where G'_p represents deformation (mm) due to the tail void of the shield machine, U_{3D}^* represents equivalent 3D elastoplastic deformation (mm) at the tunnel face, and the term ω represents radial deformation (mm) at the tunnel face. More detailed calculations of the gap parameter can be found in publications, such as [20,34,35].

The equivalent ground loss model incorporates the gap parameter into the analytical solution for a tunnel in a homogeneous elastic half-space derived by Verruijt and Booker [17]. As a result, the vertical short-term ground movement (S_1) can be given as:

$$S_1 = R^2 \left\{ -\frac{y-H}{x^2 + (y-H)^2} + (3-4\nu) \frac{y+H}{x^2 + (y+H)^2} - \frac{2y[x^2 - (y+H)^2]}{[x^2 + (y+H)^2]^2} \right\} \times \frac{4gR + g^2}{2R^2} \exp \left\{ -\left[\frac{3.12x^2}{(R + H \tan \beta)^2} + \frac{0.69y^2}{H^2} \right] \right\} \quad (2)$$

where parameter g represents the gap parameter, R represents the radius of the tunnel, ν represents soil Poisson's ratio, H represents the depth of the tunnel from the ground surface, and β represents the horizontal angle of the influence zone of ground settlement. Figure 2 provides more intuitive explanations of the equivalent ground loss model, showing the vertical ground movement influence zone where most soil displacements occur. The relationship between the settlement trough width and the tunnel depth is expressed as a horizontal angle (β). The β value is $\sim 45^\circ$ for soft to stiff clays, indicating that the ground movement occurs predominantly within the 45° wedge between the ground surface and the tunnel [20].

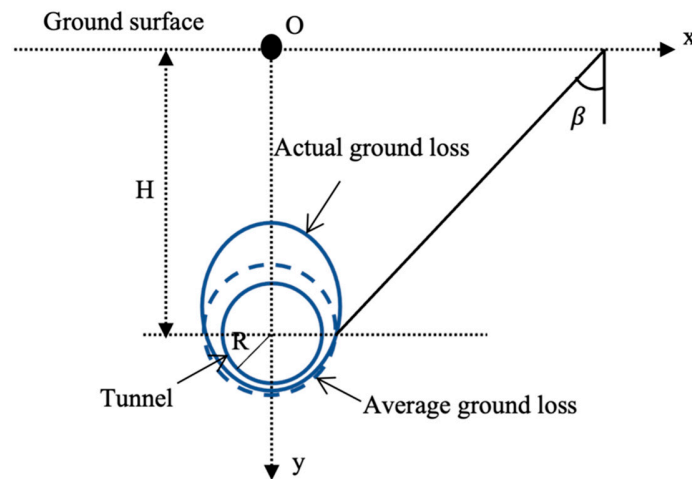


Figure 2. Schematic of tunneling-induced short-term ground loss (modified from [18]).

In practice, it is believed that the maximum ground deformation occurs at the ground surface above the tunnel crown, as indicated by the letter O in Figure 2; the parameters x and y in Equation (2) are originated from location O, i.e., $x = 0$ and $y = 0$ at location O. In such a case, Equation (2) can be simplified to Equation (3). Intuitively by Equation (3), three important parameters contributing to the ground settlements can be identified, including the tunnel depth (H), soil Poisson's ratio (ν), and gap parameter (g).

$$S_1 = \frac{1}{2H} \left[4(1-\nu) - \frac{1}{H^3} \right] \cdot (4gR + g^2) \quad (3)$$

2.1.2. Calculating Long-Term Settlement Using Terzaghi's Consolidation Theory

Measurements of the maximum ground settlement are continually recorded and collected for investigating long-term settlements. Long-term ground movements above a tunnel may increase with time in clayey soil as the tunneling-induced excess pore pressure dissipates [36]. Because of the high excess pore pressures induced in the soil mass surrounding the tunnel and the slow dissipation rate of these pressures, consolidation of soft clays may be a predominant source of settlements. This process may drag on for weeks, if not months [27]. However, consolidation will not be a problem for cohesionless materials in which excess pore pressure dissipates rather quickly, and consolidation settlements finish in a very short time.

The problem of soil consolidation was first introduced by Terzaghi [37]. Terzaghi's consolidation theory states that all quantifiable changes in soil stress directly result from a change in effective stress, such as compression, deformation, and shear resistance. Furthermore, effective stress is related to total stress and pore pressure.

Terzaghi's consolidation theory has successfully solved geotechnical problems, including foundations, tunnels, and embankments [31]. Terzaghi's one-dimensional consolidation theory is applied in this study since only ground settlements measured right above the tunnel alignment are considered, as indicated by the letter O in Figures 2 and 3. In such a case, the following assumptions and corresponding explanations should be provided besides Terzaghi's theory's assumptions.

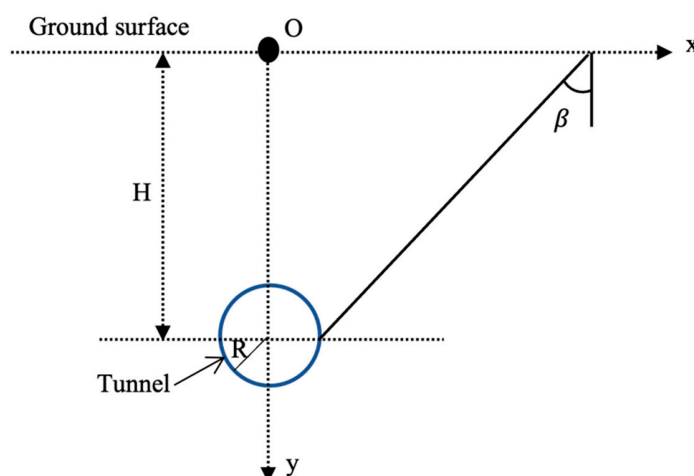


Figure 3. Schematic of tunneling-induced long-term consolidation settlement.

1. The thickness of the compressed layer is equal to the tunnel depth, and consolidation settlement at the surface is an accumulation of the consolidation settlements of all the underlying layers;
2. The tunnel lining is assumed to be impermeable because its permeability coefficient is far smaller than that of soil;
3. The initial excess pore water pressure (u) induced by excavation is equivalent to soil weight above the tunnel crown multiplied by the stress release ratio, expressed in Equation (4). Empirical values, finite element modeling, and back-analysis of observations can determine the stress release ratio for clayey soils. Excess pore pressure is considered 0 for sands due to the high permeability.

$$u = \alpha \bar{\gamma} H \quad (4)$$

where $\bar{\gamma}$ is the average weighted soil weight, H is the tunnel depth, and α is the stress release ratio;

4. The coefficient of soil volume compressibility remains constant with time. The slight variance of volume compressibility induced by excavation disturbance is not considered;

5. Tunnel elastic deformation is neglected, which is much smaller than soil deformation.

Terzaghi's one-dimensional consolidation equation can be represented by Equations (5)–(7). The parameter U_z represents the degree of consolidation, also known as the consolidation ratio, indicating consolidation progress at any time. Given the time factor T_v , the degree of consolidation (U_z) can be obtained using the U-T (U_z vs. T_v) curve [38].

$$S(t) = U_z \frac{a_v u}{1 + e} H \quad (5)$$

$$T_v = \frac{k(1 + e)t}{y_w a_v H^2} \quad (6)$$

$$m_v = \frac{a_v}{1 + e} \quad (7)$$

where a_v represents the coefficient of compressibility, e represents the void ratio, m_v represents the coefficient of volume compressibility, k represents the coefficient of soil permeability, t represents time, and y_w represents water weight. Intuitively according to Equations (5)–(7), important parameters contributing to long-term consolidation can be identified, including the degree of consolidation (U_z), tunnel depth (H), the coefficient of volume compressibility (m_v), and excess pore pressure (u). The weighted average coefficient of volume compressibility ($\overline{m_v}$) is adopted to represent the overall behavior of variable soil layers.

2.2. Ensemble ML Models

An ensemble approach combines predictions from multiple ML algorithms to make more accurate predictions than individual models [39]. Other models can compensate for errors of a single model by combining multiple algorithms [40], leading to a better overall model performance than a single ML model.

Voting is an ensemble algorithm combining various ML techniques and returns the average predicted values [41]. Specifically, multiple base models with good performances on par are ensembled to provide an average prediction.

Six models are most widely applied in predicting tunneling-induced ground deformation, based on a review of ML performances in historical cases [8,10,13]. These models include multiple linear regression (MLR), support vector machine (SVM), random forest (RF), decision tree (DT), gradient boosting (GB), and artificial neural network (ANN). Therefore, a pilot experiment evaluating model performances has been conducted on these six algorithms to select the base algorithms of the VR model. Sample datasets are collected and randomly divided into training and testing sets to test the prediction accuracy and stability of these six models. More detailed information can be found in Liu et al. [14], where the effectiveness of tree-based ML methods on small dataset predicting has been demonstrated and compared. As a result, three tree-based models with good performances on par are ensembled to form a VR model, including RF, GB, and DT (Figure 4), respectively. The VR validates a set of equally well-performing models to balance their weaknesses [42,43].

DT builds regression in the form of a tree structure, splitting a dataset into smaller and smaller subsets while it is incrementally developed. The tree's depth determines how deep the tree can grow. The dataset is split at several split points for each independent variable. The DT algorithm calculates the error between the predicted value and observations concerning the pre-defined fitness function at each split point. Then, computed errors of split points across the variables are compared. A variable yielding the lowest error values is chosen as the split point. Such a process recursively continues. The deeper the tree, the more information it obtains about the dataset [44].

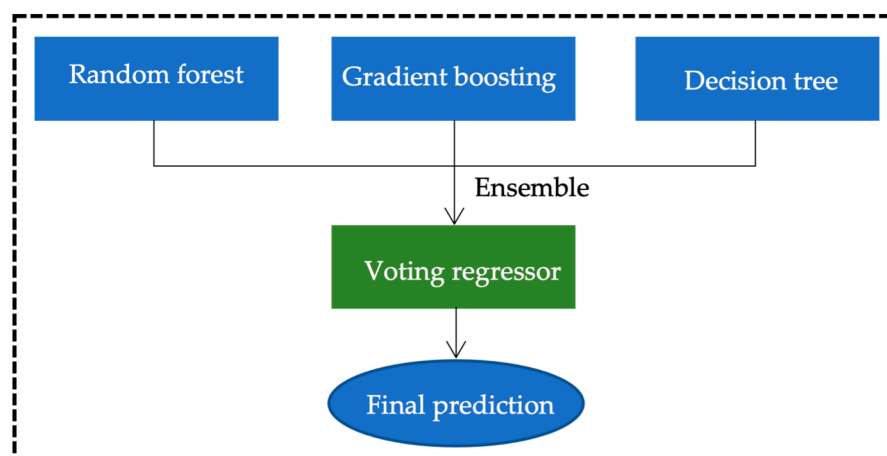


Figure 4. The schematic diagram of the ensemble VR model.

RF algorithm illustrates the power of combining many DTs into one model. Tree structures in the RF model are run in parallel, with no interaction between these trees. Two parameters need to be optimized in applying the RF algorithm, i.e., the number of regression trees and input variables per node [45].

GB is one of the most powerful techniques for building predictive models. “Boosting” in ML combines multiple simple models into a single composite model, also known as an additive model [46]. Since simple models (also known as weaker learners) are added one at a time while keeping existing learners in the model unchanged, the final model becomes a more robust predictor. Specifically, the GB algorithm involves a weak learner to make predictions, a loss function to be optimized, and an additive model to add weak learners to minimize the loss function.

The VR architecture is implemented by Scikit-learn (Sklearn) library in Python [42]. Several important hyperparameters and corresponding values involved in three selected ML algorithms are listed in Table 1.

Table 1. Important hyperparameters and corresponding values involved in ML algorithms.

DT		RF		GB	
Criterion	Mean squared error	Criterion	Mean squared error	Loss function	Mean squared error
splitter	best	number of trees	10	number of boosting stages to perform	500
max_depth	40	max_depth	no limit	max_depth	4
min_samples_split	2	min_samples_split	2	min_samples_split	5
min_samples_leaf	1	min_samples_leaf	1	learning rate	0.01
random state	0	random state	0	subsample	1

Notes: (1) Criterion is the function to measure the quality of a split. (2) Splitter is the strategy to choose the split at each node. Two options are available in the Sklearn package, including best and random. (3) The parameter max_depth represents the maximum depth of the tree. No limit means nodes are expanded until all leaves are pure or contain less than min_samples_split samples. (4) The parameter min_samples_split represents the minimum number of samples required to split an internal node. (5) The parameter min_samples_leaf represents the minimum number of samples required at a leaf node. (6) Random state controls the randomness of the estimator. (7) Subsample is the fraction of samples to be used for fitting the individual base learners.

2.3. Physics-Informed Ensemble ML Model (PIML)

The PIML framework introduces physics knowledge into ML models for better performance. The physics theories that calculate short-term and long-term settlements have been mentioned in Section 2.1. Several important physics-related parameters can be identified accordingly, including the tunnel depth (H), soil Poisson’s ratio (ν) and gap parameter (g),

the degree of consolidation (U_z), the weighted average coefficient of volume compressibility (\bar{m}_v), and excess pore pressure (u). These physics-derived features, together with in situ observations before or along with the tunnel excavation process, are set as PIML inputs. In situ observations include the distance between the tunnel face and monitoring points (DTM), groundwater level (GL), and TBM operational parameters. Specifically, the PIML framework is implemented via the following steps, with a schematic flowchart given in Figure 5.

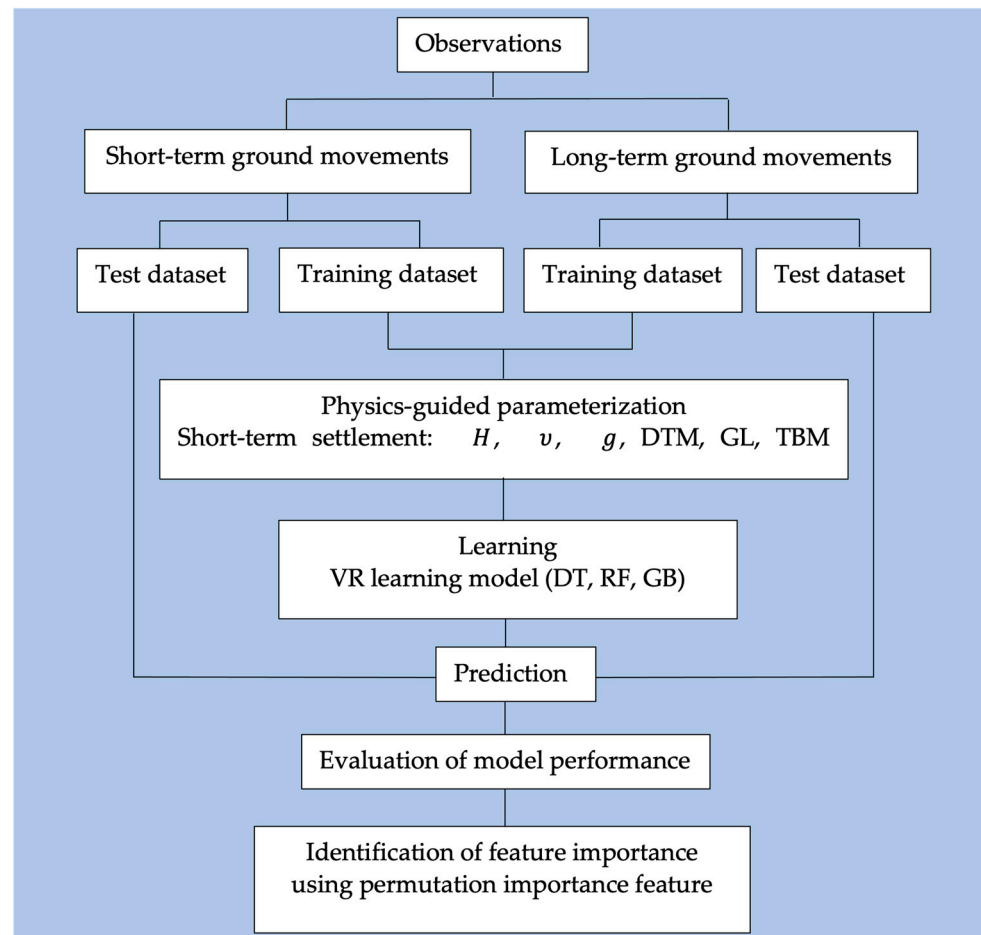


Figure 5. Flowchart of the PIML framework.

1. Step 1: Pre-classification of observations informed by physics mechanism. Since the short-term and long-term settlements are induced by different physics mechanisms and affected by different factors, the observations of short-term and long-term settlements are collected separately;
2. Step 2: Split data for ML learning and predicting. Short-term and long-term datasets need to be split into training and test subsets to achieve better ML learning and prediction. According to the tunneling construction timeline, the first 70% of data are selected for the training set, and the remaining 30% is sampled into the test data set. Such a data split is aimed to simulate prediction in an actual tunneling process, i.e., the past collected data are used to train the model and predict ground behavior in the future;
3. Step 3: Physics-informed parameterization. Input selections are physically informed. Considering the physics mechanism and possibly influenced factors, selected inputs for short-term settlements are H , v , g , DTM, GL, and TBM operations. Variables selected for long-term settlements include H , U_z , \bar{m}_v , u , DTM, GL, and TBM operations;

4. Step 4: Learn and predict tunneling-induced ground settlements using the VR model. Physics-informed variables are fed into the VR algorithm involving DT, RF, and GB to learn and predict tunneling-induced ground settlements;
5. Step 5: Evaluate model performance and identify important features. The performance of the PIML framework can be evaluated using the coefficient of determination (R^2). In addition, the permutation importance feature technique can be applied to identify important features. Identified important features will enhance our understanding of tunneling-induced ground movements.

2.4. Procedure for Feature Importance Analysis

One way to provide a model-agnostic global explanation is to analyze feature (variable) importance, which indicates how much each feature contributes to the model prediction. For example, the permutation feature importance method provides insights into an ML model's behavior by estimating and ranking feature importance based on the impact each feature has on the ML model's predictions.

Specifically, every feature is shuffled many times (K) and returns the feature importance. The calculation of feature importance can be generally implemented via (i) computing the reference score (s) of the tree-structure model on a collected dataset; (ii) shuffling every feature randomly; and (iii) returning feature importance.

(i) Compute the reference score s of the decision tree model on a collected dataset, which typically is R^2 value.

For each feature j :

(ii) For each repetition k in $1, \dots, K$:

Randomly shuffle column j to generate a corrupted version of the data.

Compute the score $s_{k,j}$ of decision tree model on computed corrupted data.

(iii) Compute importance i_j for feature f_j defined as $i_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j}$.

3. Case Study

3.1. Tunnel Project Case

This study investigates an urban tunnel section with a total length of 1.1 km in China. Figure 6 provides the location of the tunnel alignment. A slurry shield TBM excavates the tunnel. Tunnels in urban areas are generally constructed by the shield method due to their advantages of high speeds and minor disturbance to the ground surface. Before the tunneling, 75 boreholes were drilled to obtain the geological profile presented in Figure 7. The tunnel is generally constructed in a mixed silty clay–sand layer. The geological profile below the surface indicates that the ground condition encountered includes artificial fills, silty sands, silty clays, silts, and sands. All these soil types are Quaternary deposits, such as middle Pleistocene alluvial, upper Pleistocene alluvial, and Holocene deposits.

A set of settlement markers is installed along the tunnel alignment to measure real-time ground response during the tunneling process. A dataset of 187 observations has been recorded and collected from these ten surface monitoring points. Vertical ground surface displacement was monitored by precise leveling of transverse profiles.

3.2. Data Collection

3.2.1. Ground Movements

Ground movements are measured along the tunnel excavation process. The frequency of the monitored ground movements is displayed with a histogram. Their distribution is plotted using a kernel density technique, as presented in Figure 8. Positive values represent ground uplift, and negative values represent a ground settlement. Surface displacements before the arrival of TBM might be uplifted, mainly controlled by face pressure. During the passage of TBM, the dragging effect of shield friction leads to significant surface settlements. The closure of the annular gap at the shield tail further increases the surface settlements. Post-construction settlements tend to be stable gradually due to soil consolidation.

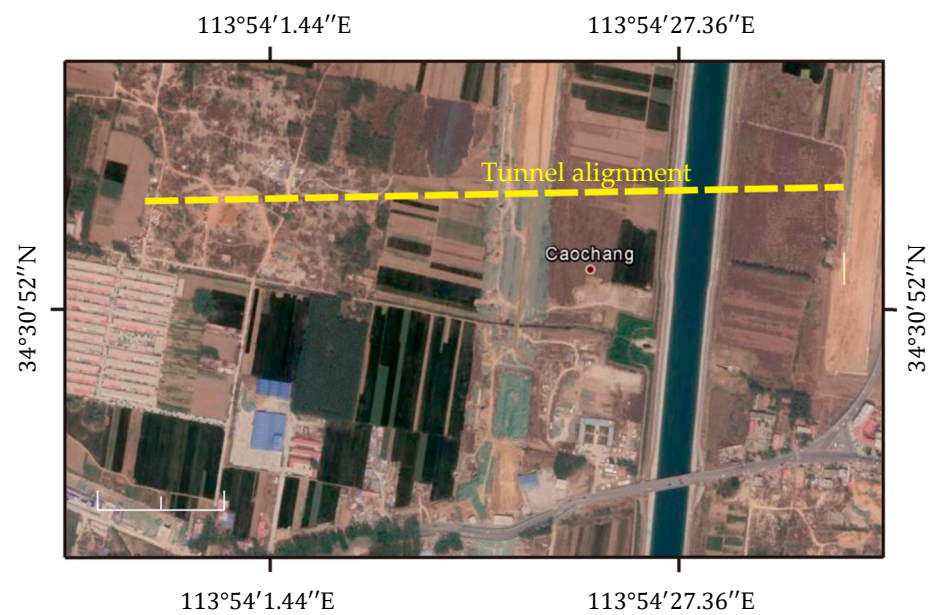


Figure 6. Location of the studied tunnel.

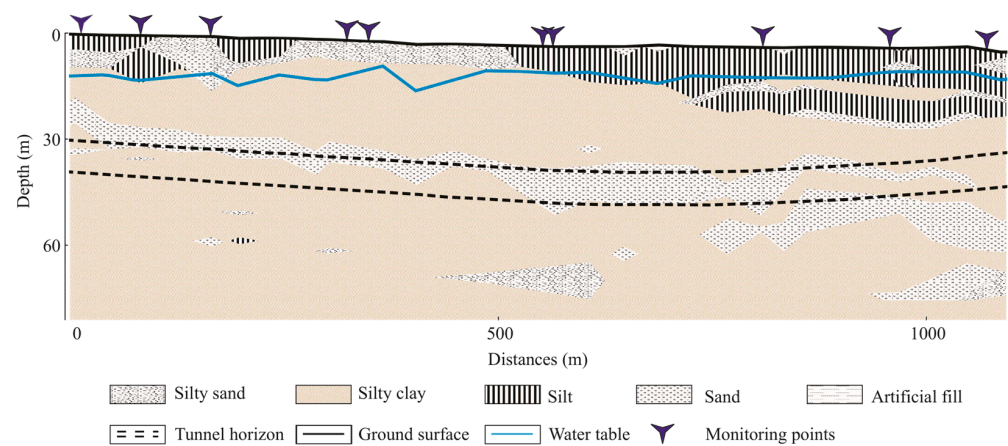


Figure 7. Geological profile of the study site.

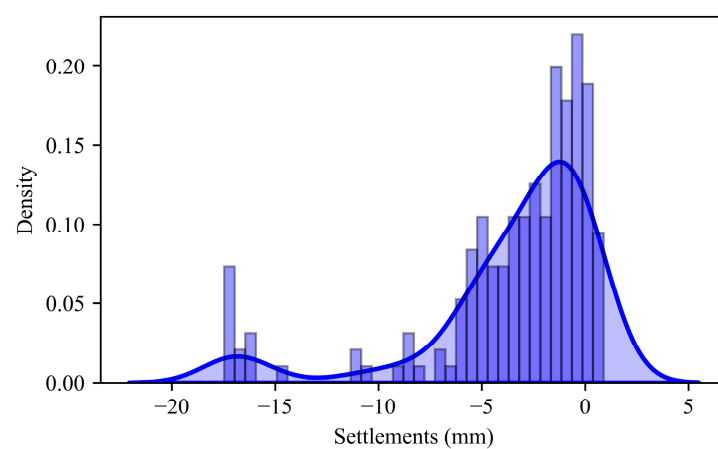


Figure 8. Data distribution of tunneling-induced ground deformation with positive values representing uplifts and negative values representing settlements.

3.2.2. Dataset for Physics-Based Model

According to equations of physics-based models, several parameters are required, including tunnel depth (H), soil Poisson's ratio (ν), gap parameter (g), excess pore pressure (u), the weighted average coefficient of volume compressibility (\bar{m}_v), and the degree of consolidation (U_z). The parameter U_z is obtained using the $U_z - T_v$ curve, indicating that the determination of parameter U_z depends on the calculation of the time factor (T_v). More detailed information of the $U_z - T_v$ curve can be found in Lovisa et al. [38].

The tunnel depth can be directly obtained from the geological profile within approximately 30~36 m. The Poisson's ratio for various soil classifications ranges between 0.25 to 0.3, which does not change significantly. The physics-based model can obtain the gap parameter and excess pore pressure.

From the soil mechanics perspective, the ground surface experience downward settlements since the soil is compressible. The coefficient of volume compressibility is an important parameter for calculating long-term settlement. Five hundred seventy-six soil samples collected from 75 drilled boreholes have been used for laboratory tests. As a result, Figure 9 plots the distribution of the coefficient of volume compressibility. The coefficient of soil permeability (k) for multiple soil layers is illustrated in Figure 10. Given the tunnel depth (H), the coefficient of volume compressibility (m_v), the coefficient of soil permeability (k), and the time factor (T_v) can be calculated using Equations (6) and (7).

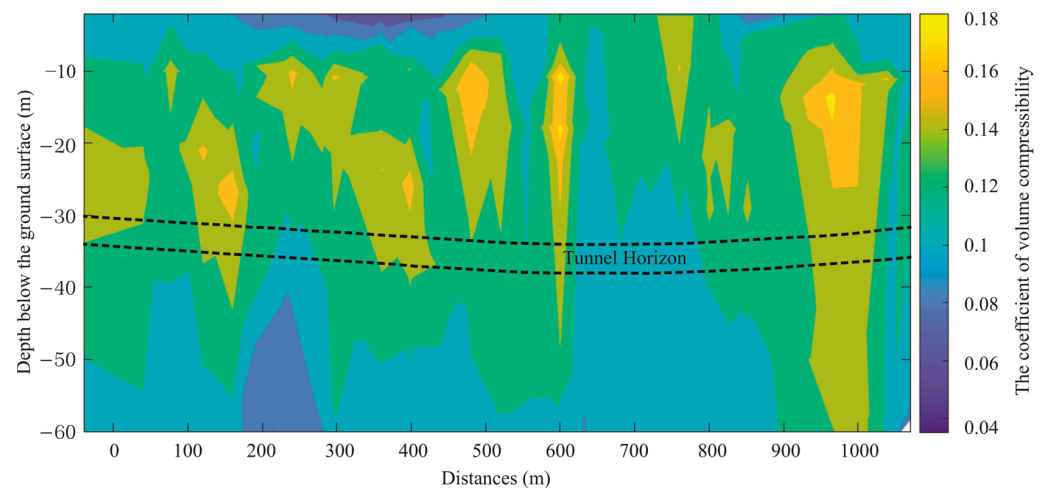


Figure 9. Distribution of the coefficient of volume compressibility.

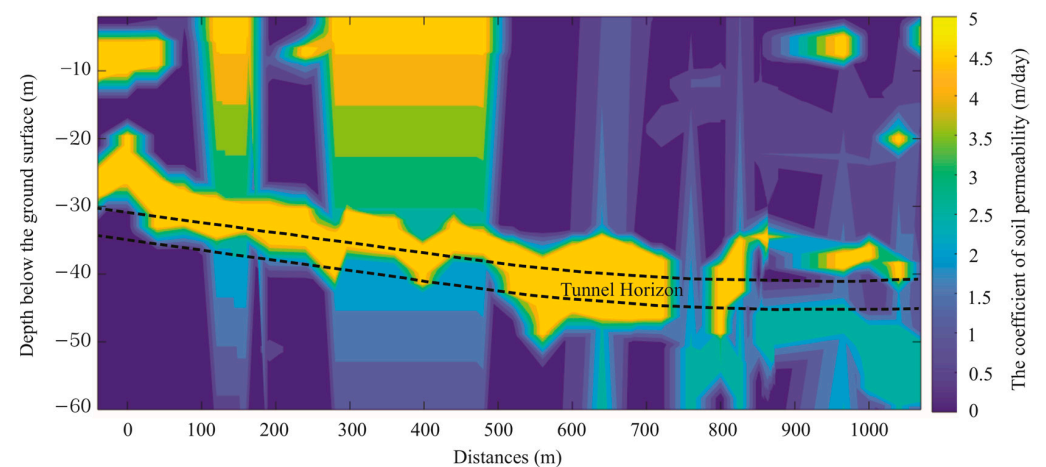


Figure 10. Distribution of the coefficient of soil permeability.

3.2.3. Other Observations or Measurements

In addition to parameters related to the physics-based model, other observations collected along the excavation process have been included in inputs, including the DTM, GL, and TBM operational parameters. TBM operations include cutter head rotational speed, cutter head torque, roller displacement, penetration rate, thrust force, grouting pressure, injection rate, and discharge flow rate. These parameters and their corresponding range of values are summarized in Table 2.

Table 2. Characteristic parameters obtained from in situ observations or measurements and range of values.

Variables	Minimum Values	Maximum Values	Mean	Standard Deviation	Unit
DTM	0.00	234.00	71.44	57.89	m
Groundwater level	7.1	13.89	9.56	2.22	m
Cutterhead rotational speed	1.20	1.60	1.42	0.07	rpm
Cutterhead torque	1800.00	18361.20	11,266.38	2718.10	kNm
Roller displacement	0.22	0.8	0.46	0.09	mm
Penetrating rate	3.00	26.60	11.62	4.36	mm/min
Thrust force	52,626.00	100,883.00	70,217.44	10,260.88	kN
Grouting pressure	270	530	391	63	kPa
Grouting injection rate	1600.40	7072.90	2297.43	383.35	m ³ /h
Discharge flow rate	1376.50	8248.00	2364.11	633.82	m ³ /h

4. Results

4.1. Model Performances

Model performance can assess and compare the difference between predictions and observations. The coefficient of determination denoted as R^2 (Equation (8)) is applied to evaluate the model performance. The higher the R^2 value, the better the model performance. The best possible R^2 score is 1.00. A constant model that always predicts the expected value would get a R^2 value of 0.00, disregarding the input features. The R -square value (R^2) can sometimes be negative since a model can be arbitrarily worse.

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

where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ is the mean value, \hat{y}_i represents the predicted value of the i -th sample, y_i represents the corresponding measured value, and n represents the number of samples.

Model performance evaluated by R^2 for short-term and long-term settlement predictions are presented in Figure 11. The PIML model shows a solid performance in both short-term and long-term cases, with the test accuracy equal to 0.84 and 0.97, respectively. This result demonstrates that the proposed PIML framework effectively learns and predicts tunneling-induced ground settlement, with test accuracy above 80%.

4.2. Feature Importance

Permutation feature importance is typically applied to analyze the feature importance of non-linear or opaque models, defined as the decrease in a model score when a single feature value is randomly shuffled [47]. Identified feature importance for both short-term and long-term settlements is presented in Figure 12.

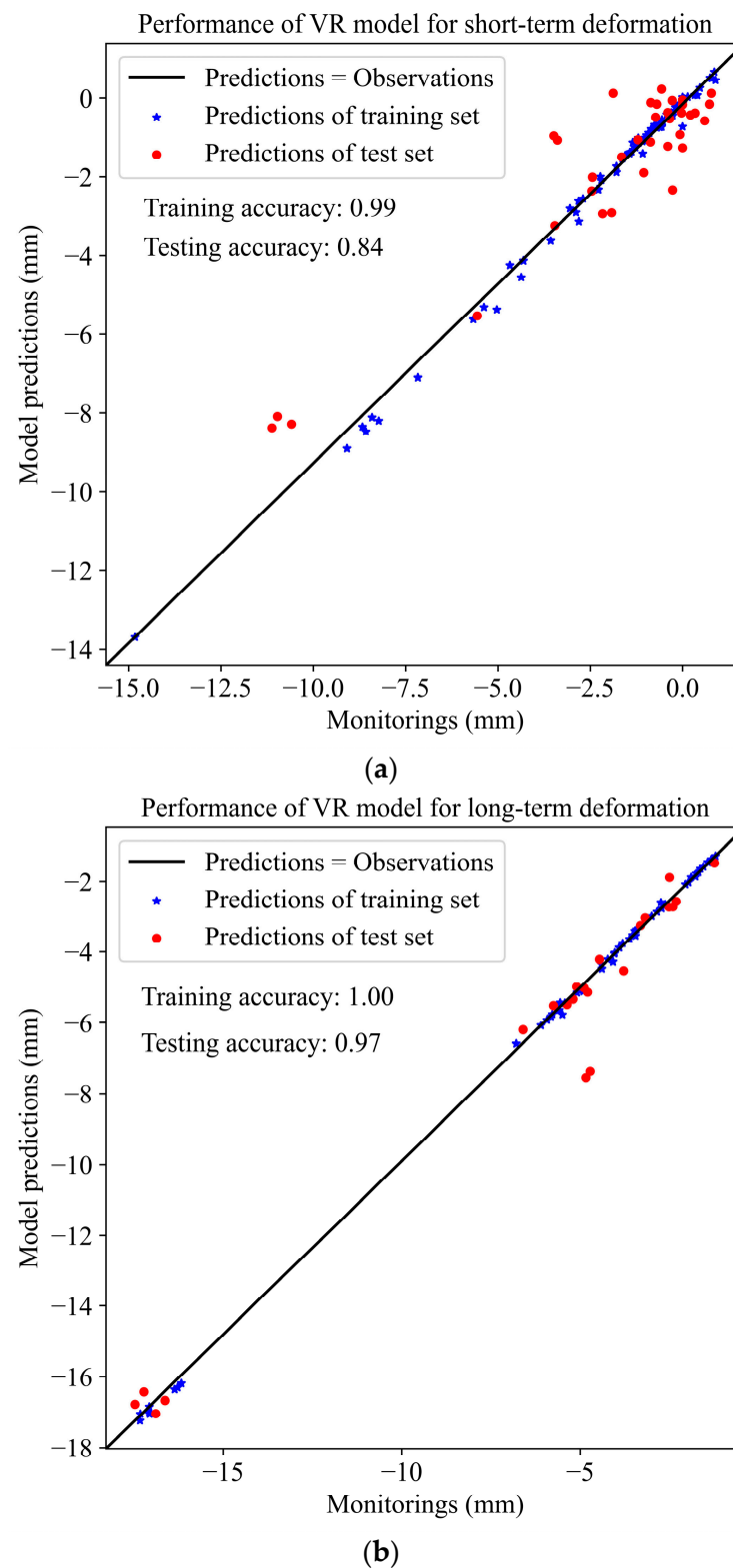


Figure 11. Performances of the PIML model. (a) Performances of the PIML model for short-term settlement. (b) Performances of the PIML model for long-term settlement.

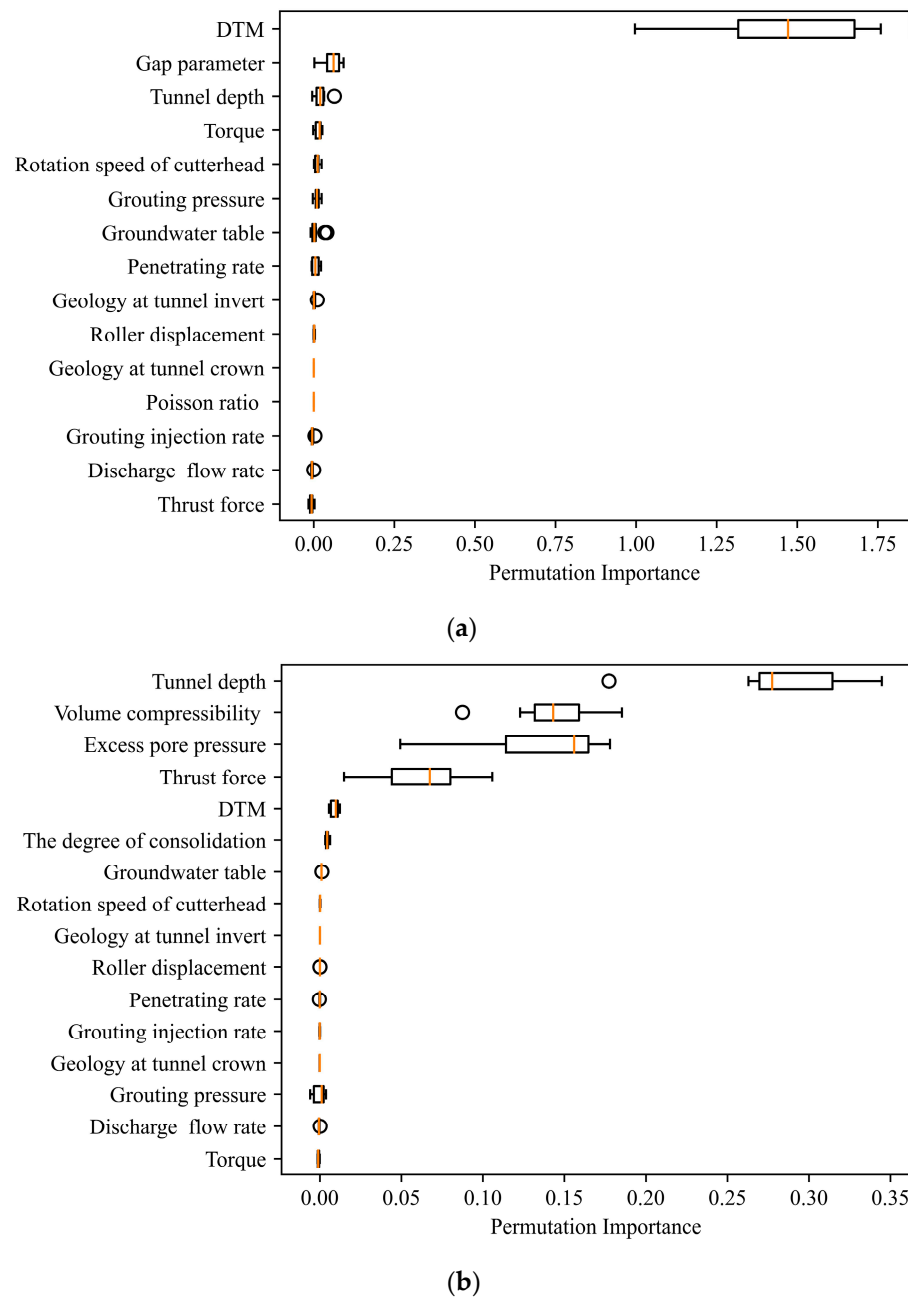


Figure 12. Feature importance for short-term and long-term settlement: (a) feature importance of the PIML model for calculating short-term settlement; (b) feature importance of the PIML model for calculating long-term settlement.

(1) Feature importance for the short-term settlement

Consideration is given to the effects of fifteen parameters. Identified three top important features contributing to short-term settlement are the DTM, gap parameter (g), and tunnel depth (H). The observing parameter DTM is the first important feature, indicating the excavation process and capturing the dynamic ground response. Two of three parameters (g and H) informed by the physics-based model have been identified as the second and third important features. The third parameter informed by the physics-based model is the soil Poisson's ratio (ν), but not considered a very important parameter by the PIML framework. The reason might be that the Poisson's ratio varies very small for different soil types, within the range of 0.25~0.3 in this study case. Low variance features are believed to not significantly affect the changing of output [48].

(2) Feature importance for long-term settlement

For the long-term settlement, the six most important features among the sixteen variables are tunnel depth (H), the weighted average coefficient of volume compressibility ($\overline{m_v}$), excess pore pressure (u), thrust force, the DTM and the degree of consolidation (U_z). All four parameters informed by the physics-based model (H , $\overline{m_v}$, u , U_z) have been identified. Additionally, the PIML model shows an interest in learning new parameters directly from observations. Two variables, i.e., the DTM and thrust force, that represent the dynamic excavation process play an essential role.

An identified important feature indicates that the model relies on the feature for the prediction. In contrast, a model ignores a feature for predicting if shuffling its values leaves model error unchanged, i.e., the value of permutation importance is approximately equal to zero. According to the results shown in Figure 12, the prediction of ground movements implemented by the PIML framework is contributed by both physics-informed features and in situ observations. Furthermore, dynamic observations play the most role in predicting short-term settlements, while physics-informed features contribute significantly to long-term settlements.

5. Discussion

The proposed PIML framework explored the continuum between physics-based and ML models by synergistically integrating scientific knowledge and data. A few identified parameters defining physics-based models and observed variables indicating the dynamic tunneling process are set as inputs for the PIML procedure. ML techniques are powerful at capturing statistical relationships between input and output variables. Regardless of the underlying physics-based mechanisms, blind application of ML methods might complicate the problem, affect model performances, and increase unnecessary data collection workloads. Thus, the proposed PIML framework filled the gap between physics theories and ML architectures, taking advantage of the generalizability of physics-based models and the robustness of data-driven ML models.

Another two models are comparably studied with the PIML to discuss their physical importance. The one is ML without physics-informed inputs, where all the physics-related features are removed from inputs. The other is ML with physics-informed inputs alone, where only several features derived from physics theories are set as inputs. The ML technique applied in these two models is the VR algorithm with the same architectural design as the PIML framework. The Taylor diagram compares the predictive results of three models for both short-term and long-term settlements (Figure 13).

The Taylor diagram, invented by Taylor [49], is a mathematical diagram designed to indicate which approximation best agrees with observations graphically. The diagram facilitates the comparative assessment of different models, quantifying the degree of correspondence between the predicted and observed behavior in terms of three statistics: the standard deviation, the Pearson correlation coefficient, and predictive accuracy measured by the root-mean-square error (RMSE).

Six models are compared using the Taylor diagram in this study, each represented by a different letter on the diagram. The distance between each model and the observation point on the x-axis measures the agreement between predictions and observations. Gray contours represent model accuracy evaluated by the RMSE. The dashed arc indicates the standard deviation of the observed field.

It is evident from the diagram that predictions obtained from the PIML for both short-term and long-term cases agree best with observations, lying nearest the observation point. The ML model with physics-informed inputs alone (represented by letters C and F in the diagram) fails to predict short-term settlement with low correlation (<0.3) and predictive accuracy (<0.5) but provides a relatively good prediction for long-term settlement with a predictive accuracy of more than 0.95. Similar things happen to the ML model without physics-informed inputs (represented by letters B and E). It provides predictions on par with PIML for the short-term case, but the prediction of long-term settlements is not ideal enough.

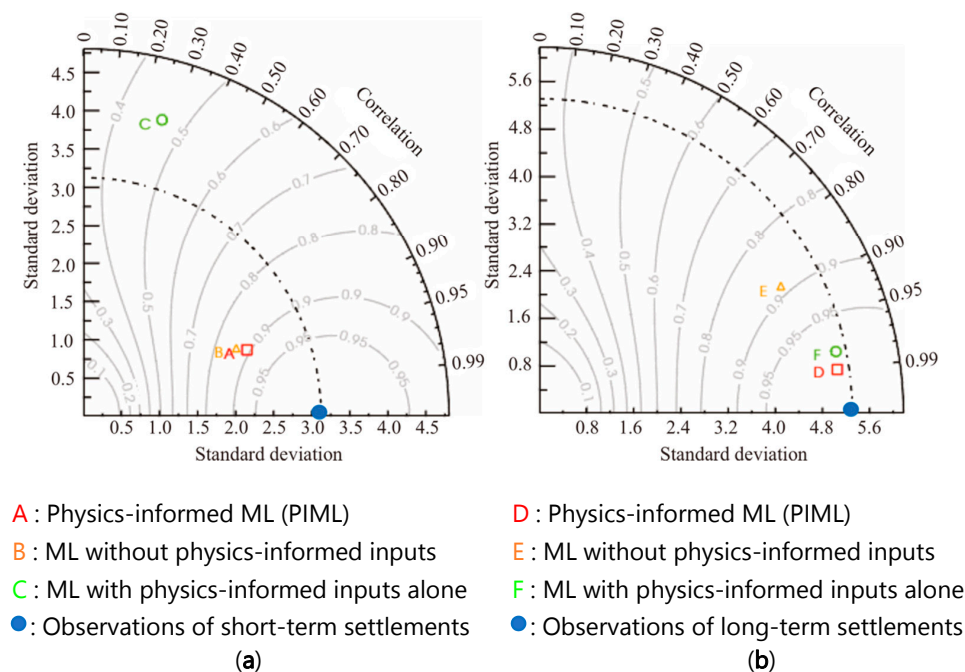


Figure 13. Comparisons among different models: (a) Predictions of short-term settlements; (b) Predictions of long-term settlements.

This finding, to some degree, agrees well with the results illustrated in Figure 11. These results recommend that observed dynamic features for the short-term case and physics-based features for long-term settlement be paid special attention in practice.

6. Conclusions

The PIML framework used simplified physics theories to inform ML model initialization and framework design. Ground deformations were classified into short-term and long-term settlements instead of predicting ground deformation using observed parameters directly and blindly. Physics-informed parameterization was applied to determine inputs. Parameters related to physics theories are available for any tunnel case. Such parameterization can increase the generalization of applying ML techniques.

As a result, the proposed PIML framework achieved good performance accuracy above 80%. The analyses of feature importance and comparison among different model designs indicated that the prediction of short-term and long-term cases relied on different sets of parameters. Observed dynamic parameters (e.g., DTM, gap parameter, and tunnel depth) contributed significantly to predicting short-term settlements, while predicting long-term settlements largely depends on features (e.g., tunnel depth, volume compressibility, and excess pore pressure) derived from physics theories.

Additionally, a comparative study among PIML, ML without physics-informed inputs, and ML with physics-informed inputs alone models demonstrated that the PIML model for both short-term and long-term cases agrees best with observations. Observed dynamic features for the short-term case and physics-based features for long-term settlement should be paid special attention in practice.

Author Contributions: Conceptualization, L.L. and W.Z.; methodology, L.L. and W.Z.; software, L.L. and W.Z.; validation, L.L.; formal analysis, L.L., W.Z., and M.G.; investigation, L.L.; resources, L.L., W.Z., and M.G.; data curation, L.L. and W.Z.; writing—original draft preparation, L.L.; writing—review and editing, W.Z. and M.G.; visualization, L.L.; supervision, W.Z. and M.G.; project administration, W.Z. and M.G.; funding acquisition, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the University Transportation Center for Underground Transportation Infrastructure (UTC-UTI) at the Colorado School of Mines under Grant No. 69A3551747118 from the U.S. Department of Transportation (DOT). The opinions expressed in this paper are those of the authors and not the DOT.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, K.M.; Kerry, R.; Lo, K.Y. Subsidence owing to tunnelling. I. *Estimating the gap parameter*. *Can. Geotech. J.* **1992**, *29*, 929–940.
2. Milillo, P.; Giardina, G.; DeJong, M.J.; Perissin, D.; Milillo, G. Multi-Temporal InSAR Structural Damage Assessment: The London Crossrail Case Study. *Remote Sens.* **2018**, *10*, 287. [\[CrossRef\]](#)
3. Wnuk, K.; Walton, G.; Zhou, W. Four-dimensional filtering of InSAR persistent scatterers elucidates subsidence induced by tunnel excavation in the Sri Lankan highlands. *J. Appl. Remote Sens.* **2019**, *13*, 34508. [\[CrossRef\]](#)
4. Wnuk, K.; Zhou, W.; Gutierrez, M. Mapping Urban Excavation Induced Deformation in 3D via Multiplatform InSAR Time-Series. *Remote Sens.* **2021**, *13*, 4748. [\[CrossRef\]](#)
5. Liu, L.; Zhou, W.; Gutierrez, M. Mapping Tunneling-Induced Uneven Ground Subsidence Using Sentinel-1 SAR Interferometry: A Twin-Tunnel Case Study of Downtown Los Angeles, USA. *Remote Sens.* **2023**, *15*, 202. [\[CrossRef\]](#)
6. Li, X.; Wang, H. Adaptive Principal Component Analysis. In Proceedings of the 2022 SIAM International Conference on Data Mining (SDM) Society for Industrial and Applied Mathematics, Alexandria, VA, USA, 28–30 April 2022; pp. 486–494.
7. Hasanipanah, M.; Noorian-Bidgoli, M.; Jahed, A.D.; Khamesi, H. Feasibility of PSO-ANN model for predicting surface settlement caused by tunneling. *Eng. Comput.* **2016**, *32*, 705–715. [\[CrossRef\]](#)
8. Chen, R.; Zhang, P.; Wu, H.; Wang, Z.; Zhong, Z. Prediction of shield tunneling-induced ground settlement using machine learning techniques. *Front. Struct. Civ. Engineering* **2019**, *13*, 1363–1378. [\[CrossRef\]](#)
9. Kohestani, V.R.; Bazargan-Lari, M.R.; Asgari-Marnani, J. Prediction of maximum surface settlement caused by earth pressure balance shield tunneling using random forest. *Artif. Intell. Data Min.* **2017**, *5*, 127–135.
10. Zhang, W.; Zhang, R.; Wu, C.; Goh, A.T.C.; Lacasse, S.; Liu, Z.; Liu, H. State-of-the-art review of soft computing applications in underground excavations. *Geosci. Front.* **2020**, *11*, 1095–1106. [\[CrossRef\]](#)
11. Suwansawat, S.; Einstein, H.H. Artificial neural networks for predicting the maximum surface settlement caused by EPB shield tunneling. *Tunneling Undergr. Space Technol.* **2006**, *21*, 133–150. [\[CrossRef\]](#)
12. Ahangari, K.; Moenossadat, S.R.; Behnia, D. Estimation of tunnelling-induced settlement by modern intelligent methods. *Soils Found.* **2015**, *55*, 737–748. [\[CrossRef\]](#)
13. Mahmoodzadeh, A.; Mohammadi, M.; Daraei, A.; Farid, H.A.H.; Kameran, A.S.N.; Mohammed, D.O.R. Forecasting maximum surface settlement caused by urban tunneling. *Autom. Constr.* **2020**, *120*, 103375. [\[CrossRef\]](#)
14. Liu, L.; Zhou, W.; Gutierrez, M. Effectiveness of predicting tunneling-induced ground settlements using machine learning methods with small datasets. *J. Rock Mech. Geotech. Eng.* **2022**, *14*, 1028–1041. [\[CrossRef\]](#)
15. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating physics-based modeling with machine learning: A Survey. *arXiv* **2020**, arXiv:2003.04919.
16. Pawar, S.; San, O.; Aksoylu, B.; Rasheed, A.; Kvamsdal, T. Physics guided machine learning using simplified theories. *Phys. Fluids* **2021**, *33*, 011701. [\[CrossRef\]](#)
17. Verruijt, A.; Booker, J.R. Surface settlements due to deformation of a tunnel in an elastic half plane. *Geotechnique* **1998**, *48*, 709–713. [\[CrossRef\]](#)
18. Loganatha, N.; Poulos, H.G. Analytical prediction for tunnelling-induced ground movements in clays. *J. Geotech. Geoenvironmental Eng.* **1998**, *124*, 846–856. [\[CrossRef\]](#)
19. Maji, V.B.; Adugna, A. Numerical modelling of tunnelling induced ground deformation and its control. *Int. J. Min. Geo-Eng.* **2016**, *50*, 183–188.
20. Chi, S.Y.; Chern, J.C.; Lin, C.C. Optimized back-analysis for tunneling-induced ground movement using equivalent ground loss model. *Tunn. Undergr. Space Technol.* **2001**, *16*, 159–165. [\[CrossRef\]](#)
21. Karpatne, A.; Atluri, G.; Faghmous, J.H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2318–2331. [\[CrossRef\]](#)
22. Fattah, M.Y.; Shlash, K.T.; Salim, N.M. Effect of Reduced Ko Zone on Time Dependent Analysis of Tunnels. *Adv. Civ. Eng.* **2011**, *2011*, 963502.
23. Fattah, M.Y.; Shlash, K.T.; Salim, N.M. Settlement Trough Due to Tunneling in Cohesive Ground. *Indian Geotech. J.* **2011**, *41*, 64–75.

24. Zhang, Z.; Zhang, M.; Jiang, Y.; Bai, Q.; Zhao, Q. Analytical prediction for ground movements and liner internal forces induced by shallow tunnels considering non-uniform convergence pattern and ground-liner interaction mechanism. *Soils Found.* **2017**, *57*, 211–226. [\[CrossRef\]](#)
25. Kasper, T.; Meschke, G. On the influence of face pressure, grouting pressure and TBM design in soft ground tunnelling. *Tunneling Undergr. Space Technol.* **2006**, *21*, 160–171. [\[CrossRef\]](#)
26. Zheng, G.; Fan, Q.; Zhang, T.; Zheng, W.; Sun, J.; Zhou, H.; Diao, Y. Multistage regulation strategy as a tool to control the vertical displacement of railway tracks placed over the building site of two overlapped shield tunnels. *Tunneling Undergr. Space Technol.* **2019**, *83*, 282–290. [\[CrossRef\]](#)
27. Hwang, R.N.; Fan, C.B.; Yang, G.R. *Consolidation Settlements Due to Tunnelling*; Proceedings of South East Asian Symposium on Tunnelling and Underground Space Development; Japan Tunnelling Association: Bangkok, Thailand, 1995; pp. 79–86.
28. Cording, E.J. Control of ground movements around a tunnel. General report. In Proceedings of the 9th Pan-American Conference on Soil Mechanics and Foundation Engineering, Santiago, Chile, 26–30 August 1991; Volume 4, pp. 2195–2244.
29. Hunt, D.V.L. Predicting the Ground Movements above Twin Tunnels Constructed in London Clay. Ph.D. Thesis, University of Birmingham, Birmingham, UK, 2005.
30. Li, S.H.; Li, C.; Yao, D. Functional relationship of parameters in different theoretical models for ground settlement caused by shield tunneling. *KSCE J. Civ. Eng.* **2020**, *24*, 310–321. [\[CrossRef\]](#)
31. Komiya, K.; Takiyama, K.; Akagi, H. Settlement behavior of a shield tunnel constructed in subsiding reclaimed area. In Proceedings of the 5th International Conference on Geotechnical Aspects of Underground Construction in Soft Ground, Amsterdam, The Netherlands, 15–17 June 2006; pp. 239–244.
32. Park, H.; Oh, J.Y.; Kim, D.; Chang, S. Monitoring and Analysis of Ground Settlement Induced by Tunnelling with Slurry Pressure-Balanced Tunnel Boring Machine. *Adv. Civ. Eng.* **2018**, *2018*, 1–10. [\[CrossRef\]](#)
33. Rowe, R.K.; Lo, K.Y.; Kack, G.J. A method of estimating surface settlement above tunnels constructed in soft ground. *Can. Geotech. J.* **1983**, *20*, 11–22. [\[CrossRef\]](#)
34. Lo, K.Y.; Ng, R.M.C.; Rowe, R.K. Predicting settlement due to tunnelling in clays. In *Tunnelling in Soil and Rock*; American Society of Civil Engineers: Atlanta, GA, USA, 1984; pp. 48–76.
35. Lee, K.M. Prediction of Ground Deformations Resulting from Shield Tunnelling in Soft Clays. Ph.D. Thesis, Faculty of Engineering Science, The University of Western Ontario, London, ON, Canada, 1989.
36. Soga, K.; Laver, R.G.; Li, Z. Long-term tunnel behaviour and ground movements after tunnelling in clayey soils. *Undergr. Space* **2017**, *2*, 149–167. [\[CrossRef\]](#)
37. Terzaghi, K. *Theoretical Soil Mechanics*; John Wiley and Sons: Hoboken, NJ, USA, 1943; p. 42.
38. Lovisa, J.; Read, W.; Sivakugan, N. Time factor in consolidation: Critical review. *Int. J. Geomech.* **2013**, *13*, 83–86. [\[CrossRef\]](#)
39. Elish, M.O.; Helmy, T.; Hussain, M.I. Empirical study of homogeneous and heterogeneous ensemble models for software development effort estimation. *Math. Probl. Eng.* **2013**, *2013*, 1–21. [\[CrossRef\]](#)
40. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [\[CrossRef\]](#)
41. An, K.; Meng, J. Voting-averaged combination method for regressor ensemble. In Proceedings of the International Conference on Intelligent Computing, Changsha, China, 18–21 August 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 540–546.
42. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
43. Ma, X.Y.; Lyu, H.Y.; Dong, X.J.; Zhang, Z.; Hao, K.R.; Yan, Q.B.; Su, G. Voting data-driven regression learning for accelerating discovery of advanced functional materials and applications to two-dimensional ferroelectric materials. *J. Phys. Chem. Lett.* **2021**, *12*, 973–981. [\[CrossRef\]](#)
44. Pekel, E. Estimation of soil moisture using decision tree regression. *Theor. Appl. Climatology.* **2020**, *139*, 1111–1119. [\[CrossRef\]](#)
45. Wang, Z.; Wang, Y.; Zeng, R.; Srinivasan, R.S.; Ahrentzen, S. Random Forest based hourly building energy prediction. *Energy Build.* **2018**, *171*, 11–25. [\[CrossRef\]](#)
46. Duffy, N.; Helmbold, D.P. Boosting methods for regression. *Mach. Learn.* **2002**, *47*, 153–200. [\[CrossRef\]](#)
47. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
48. Lavanya, P.G.; Mallapps, S. Automatic summarization and visualization of healthcare tweets. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Manipal, Karnataka, India, 13–16 September 2017; pp. 1557–1563.
49. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **2001**, *106*, 7183–7192. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.



Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

Effectiveness of predicting tunneling-induced ground settlements using machine learning methods with small datasets

Linan Liu^a, Wendy Zhou^{a,*}, Marte Gutierrez^b^a Department of Geology and Geological Engineering, Colorado School of Mines, Golden, CO, 80401, USA^b Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, CO, 80401, USA

ARTICLE INFO

Article history:

Received 31 March 2021

Received in revised form

7 July 2021

Accepted 20 August 2021

Available online 11 December 2021

Keywords:

Ground settlements

Tunneling

Machine learning

Small dataset

Model accuracy

Model stability

Feature importance

ABSTRACT

Prediction of tunneling-induced ground settlements is an essential task, particularly for tunneling in urban settings. Ground settlements should be limited within a tolerable threshold to avoid damages to above-ground structures. Machine learning (ML) methods are becoming popular in many fields, including tunneling and underground excavations, as a powerful learning and predicting technique. However, the available datasets collected from a tunneling project are usually small from the perspective of applying ML methods. Can ML algorithms effectively predict tunneling-induced ground settlements when the available datasets are small? In this study, seven ML methods are utilized to predict tunneling-induced ground settlement using 14 contributing factors measured before or during tunnel excavation. These methods include multiple linear regression (MLR), decision tree (DT), random forest (RF), gradient boosting (GB), support vector regression (SVR), back-propagation neural network (BPNN), and permutation importance-based BPNN (PI-BPNN) models. All methods except BPNN and PI-BPNN are shallow-structure ML methods. The effectiveness of these seven ML approaches on small datasets is evaluated using model accuracy and stability. The model accuracy is measured by the coefficient of determination (R^2) of training and testing datasets, and the stability of a learning algorithm indicates robust predictive performance. Also, the quantile error (QE) criterion is introduced to assess model predictive performance considering underpredictions and overpredictions. Our study reveals that the RF algorithm outperforms all the other models with the highest model prediction accuracy (0.9) and stability (3.02×10^{-27}). Deep-structure ML models do not perform well for small datasets with relatively low model accuracy (0.59) and stability (5.76). The PI-BPNN architecture is proposed and designed for small datasets, showing better performance than typical BPNN. Six important contributing factors of ground settlements are identified, including tunnel depth, the distance between tunnel face and surface monitoring points (DTM), weighted average soil compressibility modulus (ACM), grouting pressure, penetrating rate and thrust force.

2022 Institute of Rock and Soil Mechanics, Chinese Academy of Sciences. Production and hosting by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

During tunnel excavation, changes of in situ stress state and ground loss occur in the mass around the excavation. These changes are often manifested in the ground settlement, particularly for shallow tunnels in urban settings. For urban shield tunnel constructions, the ground settlement must be limited within a tolerable threshold to prevent damage to aboveground structures.

Therefore, predictions of tunneling-induced ground movement are of vital importance.

Machine learning (ML) techniques have recently experienced immense growth driven by advanced computational performance and developed for many tunneling applications, including tunnel boring machine (TBM) performance prediction, tunnel condition assessment, predictions of tunnel face stability and tunnel-induced settlement, etc. (Sheil et al., 2020; Zhu et al., 2021; Feng et al., 2021). Typically, ML techniques are implemented on the assumption of big data. However, one of the prevalent topics, prediction of tunneling-induced ground deformation, usually has access to limited monitoring datasets ranging from 10 s to 100 s, as summarized in Table 1. Thus, the effectiveness of applying ML

* Corresponding author.

E-mail address: wzhou@mines.edu (W. Zhou).

Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

algorithms in such a research area needs to be further discussed instead of using them directly and blindly.

Big data are characterized as being large in volume, produced continuously, and varied in nature, even though they are often a by-product of a system rather than designed to investigate particular phenomena or processes. Big data provide rich inputs for ML algorithms to extract underlying patterns and build predictive models (Zhou et al., 2017). As Kitchin (2013) described, big data are enormous in volume consisting of terabytes or petabytes of data. In contrast, small data are generally characterized by limited volume, non-continuous collection, and little variety, and are generated to answer specific questions (Kitchin and Lauriault, 2015). Small datasets can also be identified when the ratio of the number of training samples to the Vapnik–Chervonenkis (VC) dimensions of a learning function is less than 20 (Chao et al., 2011; Vladimir, 2020). The VC dimension is defined as the cardinality of the largest set of points that the algorithm can shatter, largely determined by input space. Since a tunnel project is typically constructed within a limited period, the average dataset size and inputs in the publications are approximately 148 and 10, respectively (Table 1). Therefore, the prediction of tunneling-induced ground settlements is a small dataset question.

The prediction of ground surface settlements induced by shallow tunnel excavation with TBM is a complex problem associated with TBM operation parameters, encountered geological conditions, tunnel geometry, etc. By applying MLs, these possible contributing factors (input data) collected from monitoring instrumentations can directly serve as the “textbook” for learning and understanding the tunneling-induced ground settlement (output data) without pre-assumption of data distribution.

The most popular ML algorithm is the ANN (Zhang et al., 2020c), a deep structure model. ANN has also been integrated with other theories such as wavelet theory and fuzzy logic (Pourtaghi and Lotfollahi-Yaghin, 2012; Ahangari et al., 2015) to improve the prediction performance. In most applications of the ANN to ground settlement prediction, the input layer has the number of nodes equal to the number of influencing variables, and the output layer has one node

representing predicted ground settlements (Suwansawat and Einstein, 2006). However, the neural network (NN) architecture design for applying small datasets is less considered, easily leading to overfitting performances. Apart from ANNs, shallow architecture models have also been utilized, such as support vector machine (SVM), decision tree (DT) and random forest (RF). SVM is designed to analyze large amounts of data and capable of handling high dimensionality (>1000) very well (Yang and Trewin, 2004). SVM used in regression problems is known as support vector regression (SVR). Tree-based methods (TMs) such as DT and RF are recently introduced to predict tunneling-induced ground settlements (Chen et al., 2019). These applied ML methods are typically evaluated by the difference between predicted and target outputs. Predictive results could be either underprediction or overprediction. Calculated errors for model assessment typically consider underprediction and overprediction equally. However, underprediction is more severe than overprediction since underprediction may cause project failure. Additionally, the model stability is less considered. Model stability indicates if model performance is affected by variabilities in the model architecture and if the obtained results can be considered robust.

Motivated by the above considerations, this paper aims to answer the question – “If ML algorithms can be applied effectively in small tunneling-related datasets.” Based on the literature review and summarization (Table 1), seven ML algorithms widely considered in publications will be applied, including multiple linear regression (MLR), DT, RF, gradient boosting (GB), SVR, an ANN trained by the back-propagation (BPNN) and permutation importance-based BPNN (PI-BPNN) algorithms. It should be noted that the PI-BPNN architecture is first-time proposed based on the consideration of applying NNs to small datasets. These seven algorithms are used and compared in an urban shield tunnel in China. Both model accuracy and stability are analyzed for seven learning algorithms. In addition, the quantile error (QE) is proposed to assess the model performance based on the consideration of underprediction and overprediction. Besides identifying the most suitable model, this paper also provides the feature importance using the permutation feature selection method. Several variables are identified as important factors affecting tunneling-induced ground settlements.

Table 1

Published research for tunneling-induced ground settlements via ML algorithms.

Dataset size	Number of inputs	Applied ML method	Reference
49	10	RF	Kohestani et al. (2017)
49	10	ANN	Suwansawat and Einstein (2006)
49	10	Wavelet-ANN	Pourtaghi and Lotfollahi-Yaghin (2012)
50	5	ANN	Darabi et al. (2012)
53	5	ANN	Ahangari et al. (2015)
55	—	SVM	Zhang et al. (2017)
81	14	ANN	Santos and Celestino (2008)
95	—	ANN	Boubou et al. (2010)
95	15	Fuzzy-ANN	Bouayad and Emeriault (2017)
113	19	ANN	Kim et al. (2001)
115	8	ANN	Goh and Hefney (2010)
143	3	ANN	Hasanipanah et al. (2016)
143	—	MLR, ANN	Moghaddasi and Noorian-Bidgoli (2018)
147	10	RVM	Wang et al. (2013)
148	8	ANN, SVM, TM	Zhang et al. (2020a)
160	9	ANN	Marto et al. (2012)
200	11	NNs, RF, SVM	Chen et al. (2019)
230	13	ANN, SVM	Ocak and Seker (2013)
294	12	RF, SVM, NNs	Zhang et al. (2020b)
300	7	SVM, DT, LR, NNs	Mahmoodzadeh et al. (2020)
328	17	ANN	Zhang et al. (2020c)
356	11	ANN	Shi et al. (1998)

Note: ANN – Artificial neural network; NNs – Neural networks; RF – Random forest; SVM – Support vector machine; TM – Tree-based method; RVM – Relevance vector machine; LR – Linear regression; MLR – Multiple linear regression.

2. Employed ML algorithms

For ML models, parameters connecting inputs and outputs are initiated randomly and optimized based on a cost function. The cost function indicates the difference between the actual and the predicted outputs. An ML algorithm aims to minimize the cost function, “learn” the model parameters, and eventually achieve the prediction. The relationship between inputs and outputs could be linear or nonlinear. A linear ML model (MLR) and six nonlinear models (DT, RF, GB, SVR, BPNN and PI-BPNN) will be discussed.

The successful application of ML techniques largely depends on the architecture design and hyperparameter training, such as the kernel selection of SVM, tree depth and tree number of tree-based algorithms, hidden layer design of NN models, etc. Without these considerations and detailed model design, it easily leads to the failed application of ML techniques. Additionally, model selection or improvement should be based on the consideration of the problem to be solved. For example, since the NN is powerful at capturing complex relationships, it might not provide ideal performance. Consequently, an improved BPNN model named PI-BPNN is proposed in this paper, which fully considers the relationship between a small dataset and network architecture.

All the algorithms and analyses are programmed in MATLAB and Python. The Python code is implemented using Scikit-learn (Pedregosa et al., 2011) Python library.

2.1. MLR

MLR is a method of finding a linear model for the relationship between the output parameter (dependent variable) and a set of input parameters (independent variables). The mathematical form of the model is expressed as

$$y = b_1x_1 + b_2x_2 + \cdots + b_nx_n + c \quad (1)$$

where b_i is the regression coefficient representing the change of output (y) when the inputs (x_i) change by 1 unit, and c is a bias term. Parameters b_i and c are initiated randomly in the beginning, and then optimized using the gradient descent method so that this formula can predict the output closely.

2.2. DT model

DT model is a tree structure that consists of an arbitrary number of nodes and branches at each node. The depth of the tree determines how deep the tree can grow. The deeper the tree, the more it obtains information about the data. DT regression uses a fast divide-and-conquer-greedy algorithm that recursively splits the data into smaller parts (Pekel, 2020). The dataset is split at several split points for each input variable. At each split point, the errors of the cost function are calculated and compared. The split point is eventually determined by the variable yielding the lowest cost function values.

The general procedure for decision tree regression is:

- (1) Start with a single node.
- (2) Determine the split point that offers the minimum value of the cost function.
- (3) For each new node, go back to step (2). The procedure exists when the stopping criterion is reached.

2.3. RF model

RF is an ensemble learning technique developed by Breiman (2001) to improve the regression tree method by combining multiple DT models. In RF regression, each tree is built using a deterministic algorithm by selecting a random set of variables and a random sample from the training dataset (Mutanga et al., 2012). The RF algorithm is not biased since there are multiple trees, and each tree is trained on a subset of data. Thus, even if a new data point is introduced in the dataset, the overall algorithm is not affected much since new data may impact one tree, but it is hard to impact all the trees.

Three parameters need to be optimized in RF:

- (1) The number of regression trees, which is the most critical parameter.
- (2) The number of different predictors tested at each node. The default value is 1/3 of the total number of the variables.
- (3) The minimal size of the trees' terminal nodes, the default value of which is 1.

2.4. GB model

GB regression is an ensemble method that employs a collection of simple additive regression model predictions and averages them to estimate the response (Lyashevskaya et al., 2020). GB regression trees consider additive models of the following form:

$$F_m(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (2)$$

where γ_m is the learning rate, and $h_m(x)$ is the weak learner. DTs of fixed size typically serve as weak learners. GB regression trees build the additive model in a forward stepwise fashion (Eq. (3)). At each stage, the weak learner $h_m(x)$ is chosen to minimize the cost function. Specifically, GB splits each tree and uses the most contributing variable to minimize the cost function.

$$F_m(x) = F_{m-1}(x) + \sum_{m=1}^M \gamma_m h_m(x) \quad (3)$$

Five essential parameters in the GB regression model need to be specified:

- (1) The number of boosting stages that will be performed.
- (2) The maximum depth, which limits the number of nodes in the tree. The optimum value depends on the interaction of the input variables.
- (3) The minimum number of samples required to split an internal node.
- (4) The learning rate, which describes how much the contribution of each tree will shrink.
- (5) The cost function to be optimized. The least squares function is used in this study.

2.5. SVR model

A set of training patterns $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ are given, where $x_i \in \mathbf{R}^n$ and $y_i \in \mathbf{R}$ ($i = 1, 2, \dots, l$). Each y_i value is the desired output for the input vector x_i . The SVR model is learnt from these patterns and used to predict the target values of unseen input vectors (Yeh et al., 2011).

SVR is a popular choice for linear and nonlinear regression types, depending on the kernel selection. Transforming data is achieved by kernel functions. Kernels can be linear, nonlinear, polynomial, radial basis function (RBF), sigmoid, etc. Kernel selection is dataset-dependent. The SVR model with RBF kernel (RBF-SVR) is used in this study, chosen by trial-and-error experiments. Two associated hyperparameters, γ and C , are needed to be specified. The parameter γ defines how far the influence of a single training example reaches. It can also be considered as the inverse of the radius of influence of samples selected by the model as support vectors. The parameter C trades off the correct domain of training examples against the maximum decision function margin.

2.6. BPNN model

As a specific example of deep learning methods, NN models can predict the outcome from a developed network. A feedforward NN is a parametric function that takes a set of input values and maps them to a corresponding setting of output values (Darabi et al., 2012). The ANN employed in this study is a feedforward network trained by a back-propagation algorithm named BPNN. Ideally, the network becomes more knowledgeable about the relationships after each iteration or epoch of the learning process with the error back-propagation algorithm being used for this purpose. Furthermore, BPNN provides a meaningful nonlinear relationship between input and output neurons by iterative learning, with the predictive ability for dynamic nonlinear behaviors. Such inherent properties

offer a clear edge over sophisticated nonlinear regression techniques (Mukherjee and Routroy, 2012). The general architecture of a NN applied to tunnel engineering consists of an input layer, a hidden layer, and an output layer. It has been demonstrated that a NN model with one hidden layer can sufficiently model highly complex nonlinear functions with enough number of hidden neurons (Hegazy et al., 1994). Specifically, a three-layer BPNN architecture shown in Fig. 1 includes the following components:

- (1) A general network model consisting of simple processing elements named neurons. Neurons are arranged in layers and are combined through excessive connectivity. The number of neurons in the input, hidden, and output layers is denoted as m , n , and s , respectively. Similarly, i , j , and k represent an individual neuron of the input, hidden, and output layers.
- (2) The input layer plays no computational role but merely serves to pass the input vector to the network. The input neurons transmit these values across the links to the hidden layers (Eq. (4)). Each node computes the weighted sum of its input (x_i). The result is then put through an activation function (f_1) to generate a signal result for the neuron (y_i). The activation functions of tan-sigmoid and log-sigmoid are commonly applied, which are differentiable and bounded. The parameter (b_j) represents a bias associated with a neuron.
- (3) The results of the hidden neurons (y_i) are then transmitted across their activation function f_2 to the neurons in the output layer (Eq. (5)). The output signal of neuron k is denoted by y_k , where the parameter $w_{k,j}$ represents the weight on the connection between hidden nodes (j) and output nodes (k), and b_k is the bias.
- (4) A back-propagation signal (an error signal) propagates backward through the network. The desired response or target output can be denoted by d_k . Consequently, an error signal denoted by E is produced by Eq. (6). The error signal is then propagated back to adjust weights and biases of each layer using a gradient descent method, namely, the weights and biases can be updated and revised in the direction of the loss function's negative gradient (E). Specifically, parameters related to each neuron are adjusted according to the delta rule shown in Eqs. (7)–(10), where η denotes the learning rate.

During the training process, weight matrices and bias vectors in the network model are adjusted until the desired input-output mapping occurs. Thus, the network model can learn and predict through training. The following are the equations involved in the BPNN algorithm:

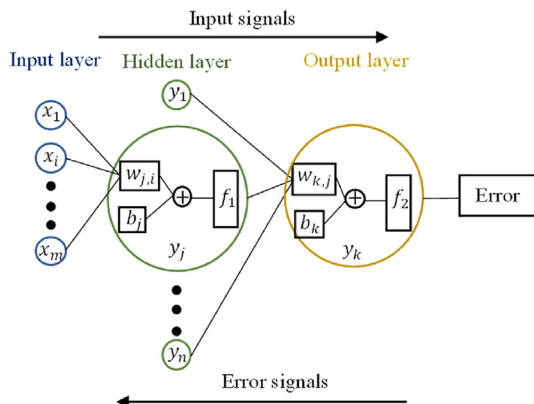


Fig. 1. ANN and signal propagation.

$$y_j = f_1 \left(\sum_{i=1}^m (w_{j,i} x_i + b_j) \right) \quad (4)$$

$$y_k = f_2 \left(\sum_{j=1}^n (w_{k,j} y_j + b_k) \right) \quad (5)$$

$$E = \frac{1}{2} \sum_{k=1}^s (d_k - y_k)^2 \quad (6)$$

$$w_{j,i}(j+1) = w_{j,i}(j) + \Delta w_{j,i} \quad (7)$$

$$w_{k,j}(k+1) = w_{k,j}(k) + \Delta w_{k,j} \quad (8)$$

$$\Delta w_{j,i} = -\eta \frac{\partial E}{\partial w_{j,i}} \quad (9)$$

$$\Delta w_{k,j} = -\eta \frac{\partial E}{\partial w_{k,j}} \quad (10)$$

2.7. PI-BPNN

The complexity of a network is related to the network size, as measured by the number of weight matrices and bias vectors. Network architecture plays a vital role in model performance since it determines the number of parameters to be trained. Take the model structure shown in Fig. 1 as an example. The dimension of the weight matrices $w_{j,i}$ and $w_{k,j}$ is $m \times n$ and $n \times s$, respectively. The size of the bias vector is $n + s$. Therefore, the total number of parameters needed to be trained is $(m + s + 1) \times n + s$. In applying ground surface settlement predictions, one neuron ($s = 1$) in the output layer represents settlement. Thus, the total number of parameters needed to be trained becomes $(m + 2) \times n + 1$. The number of hidden neurons (n) can be identified via a trial-and-error experiment, which also can be expressed in the form of the number of input nodes (m) suggested by Heaton (2015). The appropriate number of hidden neurons is estimated by $(2m)/3 + s$, which should not exceed $1.5m$. Consequently, the model complexity largely depends on the number of input nodes.

Many researchers directly set the influencing variables as the input neurons and ground settlement as the output neuron in the review of related publications. It is widely believed that as many parameters related to the ground settlement as possible should be collected and fed into the input layer (Suwansawat and Einstein, 2006). However, increasing the number of input neurons leads to a more complex model so that much more weights and biases are needed to be learnt. For many NN applications in settlement predictions, the size of both training and testing data is 10 s or around 100 s (Santos and Celestino, 2008; Darabi et al., 2012; Ahangari et al., 2015; Bouayad and Emeriault, 2017). When the number of parameters needed to be learnt is larger than the size of learnt samples, these parameters may not be well trained. In addition, since a network with a sufficient number of neurons in the hidden layer can accurately implement an arbitrary training set, it can learn both signals and noises (Tetko and Villa, 1997). With limited data points, some of these trained relationships are possibly the results of sampling noise. The PI-BPNN model is proposed considering the network architecture design for small datasets. In the proposed PI-BPNN model, input initialization is achieved by the permutation feature importance selection technique, which is a model

inspection technique and especially useful for nonlinear or opaque estimators. The permutation feature importance is defined as the decrease in a model score when a single feature value is randomly shuffled (Breiman, 2001), indicating how much the model depends on the feature. Thus, important features will be filtered, selected, and set as network inputs to decrease architecture complexity.

3. Description of the study site and collected dataset

3.1. Geological conditions

Data used in this study are collected from the Yuji tunnel project constructed in China, which is excavated by a slurry shield TBM. Many boreholes have been drilled to obtain information about the geological condition below the surface. The soil profile of the tunnel horizon is plotted in Fig. 2 based on the Unified Soil Classification System (USCS). The tunnel section with a total length of 1.1 km is investigated. The ground condition encountered by the shield tunnel includes silty sands, silty clays, silts, sands, and artificial fills. The complexity of geological conditions varies. In general, the tunnel is constructed in a mixed silty clay-sand layer. Most areas were found with sands in the crown and silty clays in the invert. Some parts are entirely excavated in sand or silty clay layer. Ten surface settlement markers or settlement arrays are installed along the selected tunnel section (>1 km) to measure surface settlements during excavation. Vertical ground surface displacement is monitored by precise leveling of transverse profiles. Each profile consists of at least 13 monitoring points. A dataset of 187 observed maximum surface displacements is recorded, collected, and used in this study.

3.2. Collected dataset for model construction

The ground surface displacements could result from multiple influencing variables fed into ML models as inputs. Based upon case history reviews of urban shield tunnels (Clough and Leca, 1993; Suwansawat and Einstein, 2006; Zhang et al., 2020b) and extensive observations in this tunnel project, the influencing variables can be grouped into three major categories: tunnel geometry, geological conditions and TBM operational parameters. Such classification proposed by Suwansawat and Einstein (2006) is widely accepted by researchers (Ding et al., 2013; Ocak et al., 2013; Wang et al., 2013; Zhou et al., 2017; Chen et al., 2019). Parameters included in every category are mainly derived from measurements before tunneling or observations along with tunnel excavation. Mahmoodzadeh

et al. (2020) also mentioned that urban tunneling projects do not always have access to a vast number of parameters, therefore, it is preferable to consider parameters that can be easily achieved.

3.2.1. Input variables

(1) Tunnel geometry

Tunnel depth and diameter are considered in the modeling. The tunnel depth of the selected section varies from around 31 m–36 m. The tunnel project has a constant diameter of 12.4 m. A baseline feature selection approach in ML techniques is to remove features with low variance. Zero-variance features, i.e. features that have the same value in all samples, will not affect the changes of output. Thus, the tunnel diameter is not used as an input.

Additionally, the horizontal distance between tunnel face and monitoring points (DTM) is also considered a parameter that might influence the monitored ground settlements, which indicates various excavation stages from before shield approaching till after shielding pass.

(2) Geological conditions

Geological conditions considered in ML models include soil types at tunnel invert and crown, groundwater level, and weighted average soil compressible modulus (ACM). All these parameters related to geological conditions are obtained from many drilled boreholes along the tunnel centerline.

Intuitively by the geological profile shown in Fig. 3, there are two geological conditions of tunnel crown and invert, i.e. silty clays or sands. Binary code symbolized by 0 and 1 typically represents two states, “off” and “on”, respectively. Thus, the binary code can be adopted to indicate if a certain soil type is encountered or not (Suwansawat and Einstein, 2006). In this study area, four geological conditions are summarized, as listed in Table 2.

Although geological complexity can be roughly determined intuitively by the geological profile, quantitative characteristics of soil consolidation are unknown, which may serve as a vital factor affecting the ground settlements. From the soil mechanics perspective, since the soil is a compressible material, the ground surface experiences downward settlements. Five hundred and seventy-six soil samples have been collected from 75 drilled boreholes along the tunnel alignment. These collected soil samples have been used to conduct laboratory soil consolidation tests to

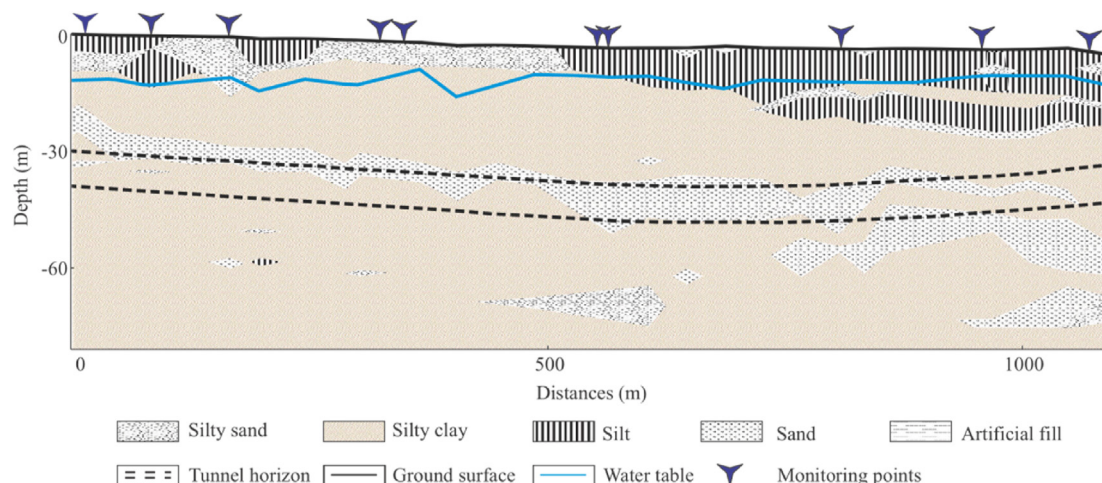


Fig. 2. The geological profile of the study area.

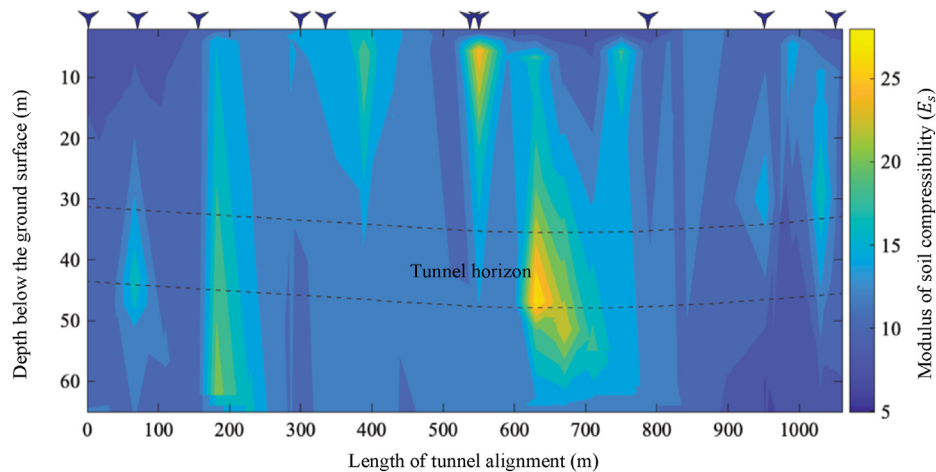


Fig. 3. Distribution of weighted average compressible modulus (ACM).

study compression properties. Plots of soil void ratio e vs. effective pressure p can be drawn according to results of laboratory compression test. Based on the e - p curve, the coefficient of compressibility (a_v) expressed as Eq. (11) can be calculated, representing the change of void ratio concerning the applied effective pressure during compression.

Given the initial void ratio of the consolidation layer (e_0), the coefficient of volume compressibility (m_v) is obtained according to Eq. (12). The parameter m_v represents the compression of soil per unit due to a unit increase of pressure. The modulus of soil compressibility (E_s) expressed in Eq. (13) is eventually selected as the parameter to represent the soil compressibility. Results of the calculated modulus of compressibility are plotted in Fig. 3. The ACM along drilled borehole can be calculated to represent the overall soil compressibility for every monitoring point using Eq. (14), where E_{si} is the modulus of soil compressibility of the i th soil layer, and h_i represents the height of the i th soil layer.

$$a_v = -\Delta e / \Delta p \quad (11)$$

$$m_v = a_v / (1 + e_0) \quad (12)$$

$$E_s = 1 / m_v \quad (13)$$

$$ACM = \frac{\sum E_{si} h_i}{\sum h_i} \quad (14)$$

(3) TBM operational parameters

In soft ground with high water content, the slurry shield TBM is suitable. This TBM is equipped with a slurry system that controls the pressure in the excavation face by injecting pressurized slurry into the cutter chamber. Given a TBM, its specifications depend on and reflect soil conditions. The criterion of selecting TBM-related parameters is that parameters directly affect face stability, ground volume loss, etc., and further affect ground settlement (Monney et al., 2016; Chen et al., 2019; Zhang et al., 2020c). Additionally, real-time collection in the shield system is suggested to ensure the accuracy of these parameters. Based on the TBM parameters related to ground settlements suggested by Zhang et al. (2020b), eight features are selected, including rotational speed of cutterhead, penetrating rate, thrust force, torque, roller displacement, grouting pressure, grouting injection rate, and discharge flow rate. Controlling

Table 2

Geological conditions of tunnel crown and invert represented by binary code.

Type	Geology at tunnel crown		Geology at tunnel invert	
	Silty clay	Sand	Silty clay	Sand
1	1	0	1	0
2	1	0	0	1
3	0	1	1	0
4	0	1	0	1

these parameters well may effectively reduce the surface settlement caused by shield construction.

TBM parameters, such as cutterhead rotational speed and penetrating rate, significantly influence TBM vibration. The thrust force of soft ground TBMs is controlled by earth pressure. Torque factor times the cube of tunnel diameter describes the torque of soft ground TBMs. TBM torque is governed by the friction between shield cutters and the ground or by the shear strength. This friction is caused by a few factors, such as earth pressure, chamber pressure acting on a bulkhead, driving force generated by direction changes in curved alignments, frictional force acting between the segments and tail seals, and hauling force of trailing units (Ates et al., 2014). In slurry shield TBM operations, the slurry circuit is a pipeline loop that connects the TBM to the slurry treatment plant (STP) at the surface. Pumps in the slurry discharge pipeline transport the loaded slurry to the STP. In the slurry feed line and discharge line, sensors are installed to determine the flow rate and density. Densities of slurry injection flow and discharge flow are almost constant. Thus, three parameters are selected to represent the slurry circuit condition: grouting pressure, grouting injection rate, and discharge flow rate.

3.2.2. Output variables

Ground settlement induced by tunneling excavations is set as the output variable in ML models, which is influenced by multiple input factors. Given that data are collected during the excavation process, ground settlements within eight times of the tunnel diameter at the excavation face are considered, with a measurement accuracy of 0.3 mm/km.

Table 3
Characteristic parameters and ranges of values for settlement predictions.

Category	Variables	Minimum	Maximum	Mean	Standard deviation
Tunnel geometry	Depth (m)	31.2	36	34.26	1.97
	DTM (m)	0	234	71.44	57.89
Geological conditions	Groundwater level (m)	7.1	13.89	9.56	2.22
	ACM (MPa)	7.92	13.44	10.04	2.2
TBM operations	Cutterhead rotational speed (rpm)	1.2	1.6	1.42	0.07
	Cutterhead torque (kN m)	1800	18,361.2	11,266.38	2718.1
	Roller displacement (mm)	0.22	0.8	0.46	0.09
	Penetrating rate (mm/min)	3	26.6	11.62	4.36
	Thrust force (kN)	52,626	100,883	70,217.44	10,260.88
	Grouting pressure (bar)	2.7	5.3	3.91	0.63
	Grouting injection rate (m ³ /h)	1600.4	7072.9	2297.43	383.35
	Discharge flow rate (m ³ /h)	1376.5	8248	2364.11	633.82

Note: 1 bar = 100 kPa.

3.3. Data distribution

The dataset used to design ML models for predicting the ground surface settlements consists of 14 input variables and 187 observations. Twelve parameters and their corresponding ranges of values are summarized in Table 3. The other two parameters are geology at tunnel crown and invert, represented by values 0 and 1, respectively, in ML models (Table 3).

Fig. 4 presents the variable correlation matrix illustrating the relationship between pairwise parameters to provide a descriptive overview of the data distribution. Most parameters show a relatively low correlation with Pearson correlation coefficient ($r < 0.8$), except for two pairs: grouting pressure vs. groundwater level ($r = 0.82$) and geology at tunnel invert vs. ACM ($r = -0.81$). The larger the r value, the stronger the linear relationship. However, there are no direct physical links between these two pairwise datasets, so that all the data will be kept for learners.

The data distribution also reveals that the ground settlements have inverse relationships with a few parameters. These parameters are the DTM, tunnel depth, ACM, thrust force, grouting pressure, grouting injection rate and discharge flow rate. Thus, the increase in these parameters can control the magnitude of ground settlements to some extent. Similarly, the decrease in positively related parameters might help to decrease the ground surface settlements.

3.4. Training and testing sets

ML methods ingest data to learn from it. Specifically, ML algorithms aim to find the optimal hyperparameters that yield the best model. Corresponding hyperparameters are initiated randomly and optimized using a cost function. The process to find the optimal hyperparameters is achieved by learning from the training set. Using trained hyperparameters, ML models have the ability to predict outputs.

The dataset in this study is split randomly into two sets: a training set and a testing set, as plotted in Fig. 5. The training set consists of 70% of the data, while the remainder (30%) is assigned to the testing set. A two-part split is a general practice in the applications of ML algorithms, aiming at testing the predictive performances of the models.

4. Results

4.1. Comparison of model performances

Typically, statistical performance evaluations can be conducted to assess model performance quantitatively. The coefficient of determination (R^2) defined as Eq. (15) is adopted to assess and compare model performances. It measures how well predictions a model provides. The best possible score is 1. The higher the R^2 value, the better the model performance. Additionally, the R^2 value can be negative because a model can be arbitrarily worse.

$$R^2 = 1 - \frac{\sum (\text{monitoring predictions})^2}{\sum (\text{monitoring mean of monitoring})^2} \quad (15)$$

Training and testing accuracies for seven models have been calculated and are presented in Fig. 6. According to the calculated testing accuracy, the predictive performance ranging from best to worst is RF (0.9) > GB (0.82) > DT (0.72) > PI-BPNN (0.79) > BPNN (0.59) > SVR (0.52) > MLR (3.89 10¹⁷).

Among these applied models, the MLR has neither learning ability nor predicting ability, which means that the relationship between induced ground settlements and influencing variables is nonlinear. The nonlinear relationship can also be indicated by the variable coefficient matrix shown in Fig. 4. The surface settlements weakly correlate to the 14 influencing variables with correlation coefficients (R) smaller than 0.3.

In the applications of nonlinear models, the ensemble methods like RF and GB have been successfully used with both relatively high learning and testing accuracies. The RF model achieves the best performance. Also, the ensemble models (RF and GB) outperform the single DT method in terms of model accuracy.

In contrast to shallow-structure learning architectures, the NN model adopts deep architecture to learn hierarchical representations. Deep-structure ML methods can often capture more complicated, hierarchically launched statistical patterns of inputs to adapt to new areas than traditional learning methods. Furthermore, as the data continuously get bigger, deep learning is essential in providing predictive analytic solutions for large-scale datasets (Pasini, 2015; Qiu et al., 2016). However, for the application of tunneling-induced ground settlements where the available dataset is small, the predictive performance of BPNN is worse than tree-based ML algorithms. Additionally, the proposed PI-BPNN model outperforms typical BPNN, indicating that the NN architecture and input selection are vital to the model performance when applying a small dataset.

4.2. Feature importance evaluation considering underpredictions and overpredictions

Besides the R^2 used in this study, model performances can be assessed by other performance indicators such as mean absolute error (MAE), root mean square (RMSE), relative RMSE (RRMSE), standard deviation (σ), and multiple-objective error (MOE), which are defined as

$$MAE = \frac{1}{n} \sum_{i=1}^n |d_i - y_i| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - y_i)^2} \quad (17)$$

$$RRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{d_i - y_i}{d_i}^2} \quad (18)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2} \quad (19)$$

$$MOE = RRMSE / (1 + R) \quad (20)$$

where d is the monitoring value; y is the predicted value; \bar{d} and \bar{y} are the means of actual and predicted values, respectively; e is the absolute error, and $e = |d_i - y_i|$; \bar{e} is the mean of absolute error; and n is the number of datasets.

In the above parameters, MAE reflects the average magnitude of error between predicted and measured data, and RMSE describes the standard deviation of differences. RRMSE is the ratio of RMSE to the mean of monitoring values (Bsaibes et al., 2009). RRMSE, RMSE, MAE, and σ merely illustrate the error without correlations. MOE function combines the above-mentioned metrics.

These equations are commonly used in ML applications to indicate the overall deviation between predictions and observations. The higher the errors, the less likely the model generalized correctly from the training data. Although these equations have slight differences, one common aspect is that underpredictions and overpredictions are equally considered. However, for ground

settlements induced by tunnel excavations, underpredictions may cause project failure and even loss of lives compared with overpredictions. Consequently, a QE criterion is proposed to assess the model performance. Different penalties for overestimation and underestimation are given in the QE criterion. In the QE calculation, different penalties can be given to overestimation and underestimation. These penalties are based on the value of chosen quantile (γ), as shown in Eq. (21). Quantile error can be considered an extension of MAE when the quantile is 50% ($\gamma = 0.5$), and the equation for QE is reduced to the equation for MAE. A quantile loss function of $\gamma > 0.5$ gives more penalty to underestimation than overestimation. In the model performance evaluation, the parameter γ in the QE calculation has been set as 0.7. Predictions (y) for each model are the mean of 1000 predictive results.

$$QE = \frac{1}{n} \left(\sum_{i: d < y} (\gamma - 1) |d - y| + \sum_{i: d > y} (\gamma) |d - y| \right) \quad (21)$$

Using the QE criterion, the predicted model performances ranging from best to worst are RF (0.36) > GB (0.38) > DT (0.41) > SVR (1.02) > PI-BPNN (1.17) > BPNN (1.22) > MLR (1.81 $\times 10^9$). The RF algorithm has the best model performance with the QE errors, even if more penalties are given to underpredictions. Compared with the model performance evaluated by R^2 (RF > GB > DT > PI-BPNN > BPNN > SVR > MLR), the SVR outperforms NNs using QE errors, indicating that NNs are prone to providing underpredictions.

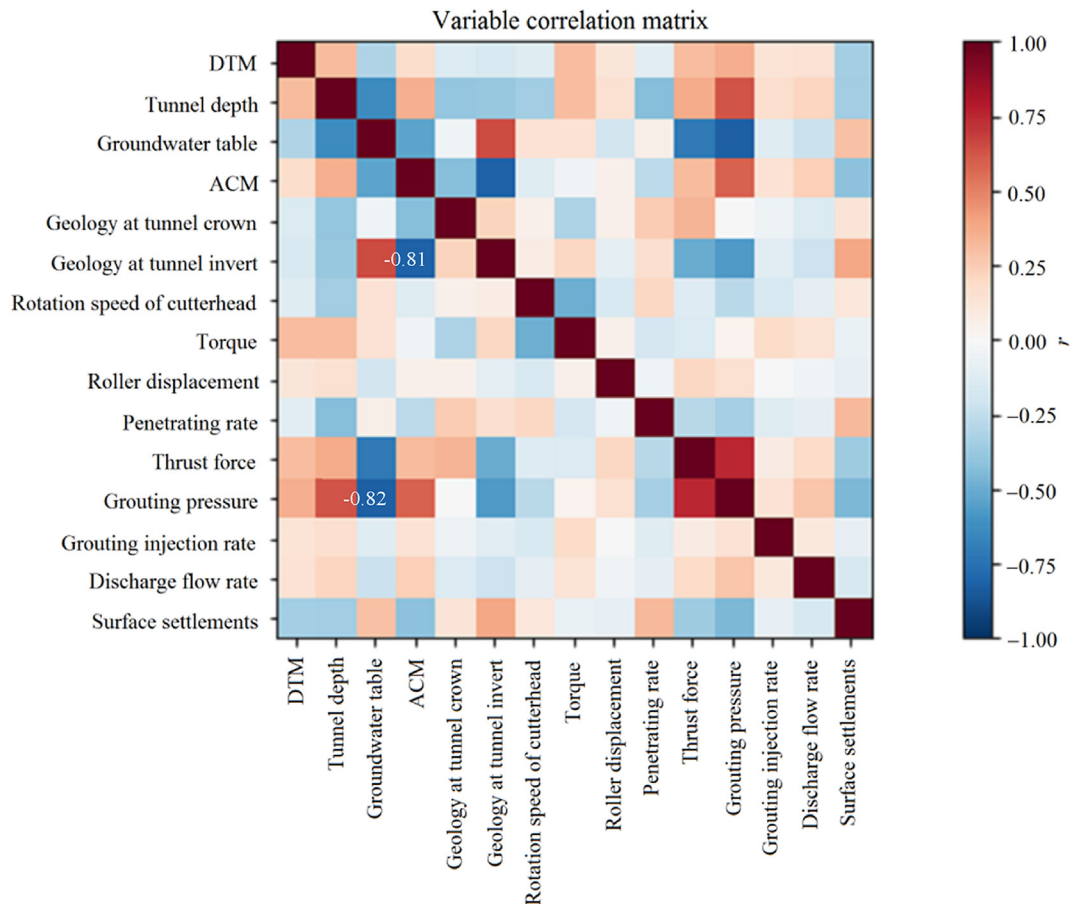


Fig. 4. Variable correlation matrix of influencing variables.

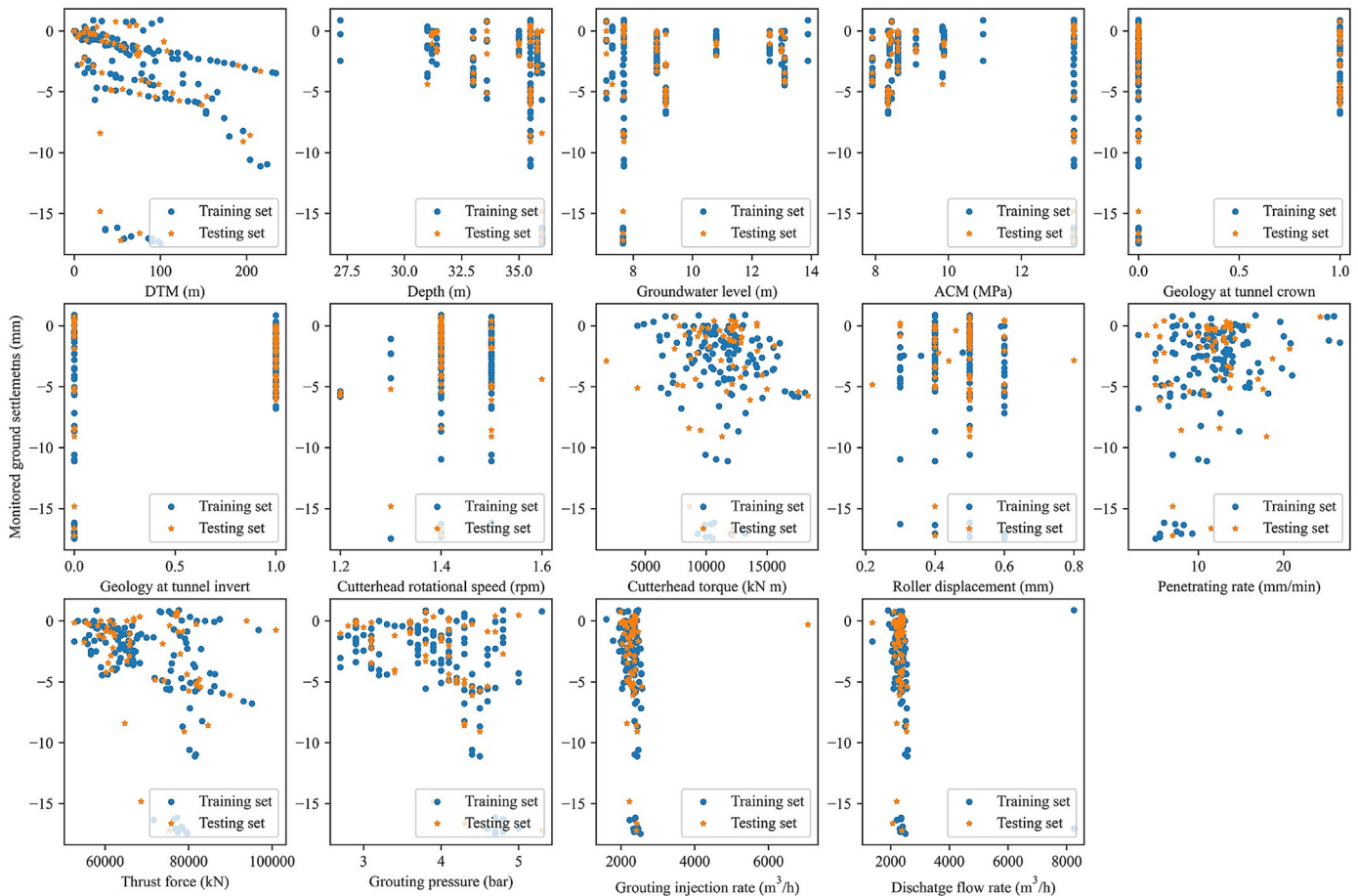


Fig. 5. Data distribution of training and testing sets.

4.3. Feature importance

ML models are used to identify the relationship between ground settlements and various influencing variables. One way to explain why a given model behaves the way it does is to analyze feature importance. Feature importance provides a highly compressed, global insight into the model's behavior.

The model inspection technique of permutation feature importance tests the model performance by removing every feature. In this way, the importance of individual variables can be directly compared. Methods like this can be used to dispel the notion of ML algorithms as irreproducible “black-box” and can be used to help to gain new insights. Specifically, permutation feature importance measures the increase in the prediction error of a model after we permute a feature's value, i.e. the decrease in a model score when a single feature value is randomly shuffled. A feature is “important” if shuffling its values increases the model error because it relies on the prediction feature. Conversely, a feature is “unimportant” if shuffling its values leaves model error unchanged because the model ignores the feature for the prediction in this case.

However, not all learning methods can identify important features because they are too complex to analyze contributions of single covariates to the overall results, for instance, ANNs and SVR (Altmann et al., 2010). Additionally, the failed application of MLR (Fig. 6a) demonstrates that the nonlinear relationship between inputs and output was not captured, indicating that no features are considered “important” in the MLR model. Thus, the permutation

feature importance will be evaluated based on the predictive results of the DT, RF and GB methods.

According to the results presented in Fig. 7, six important parameters contributing to ground settlements have been identified, including tunnel depth, DTM, ACM, grouting pressure, penetrating rate and thrust force. Typically, parameters of tunnel depth, DTM and ACM are given by tunnel design and geological conditions. Thus, the tunneling-induced ground settlements can be further controlled by TBM operations such as grouting pressure, penetrating rate and thrust force.

4.4. Model stability

In applying ML algorithms, model accuracy is typically considered the primary criterion for evaluating and selecting a learning algorithm. Algorithms with high accuracy can predict the outcome with a high degree of confidence. However, the stability of the learning algorithm is also a vital issue, which is often less considered. Therefore, it should be paid more attention, particularly for the ML applications of a small dataset. The stability of a learning algorithm, also called learning variance, refers to the output change when the training dataset is changed. A learning algorithm can be considered stable if the output does not change much when the training set is modified. Also, the algorithm stability can be considered a representation of how well hyperparameters have been trained. Models with high variance pay a lot of attention to training data and do not generalize on the unseen data.

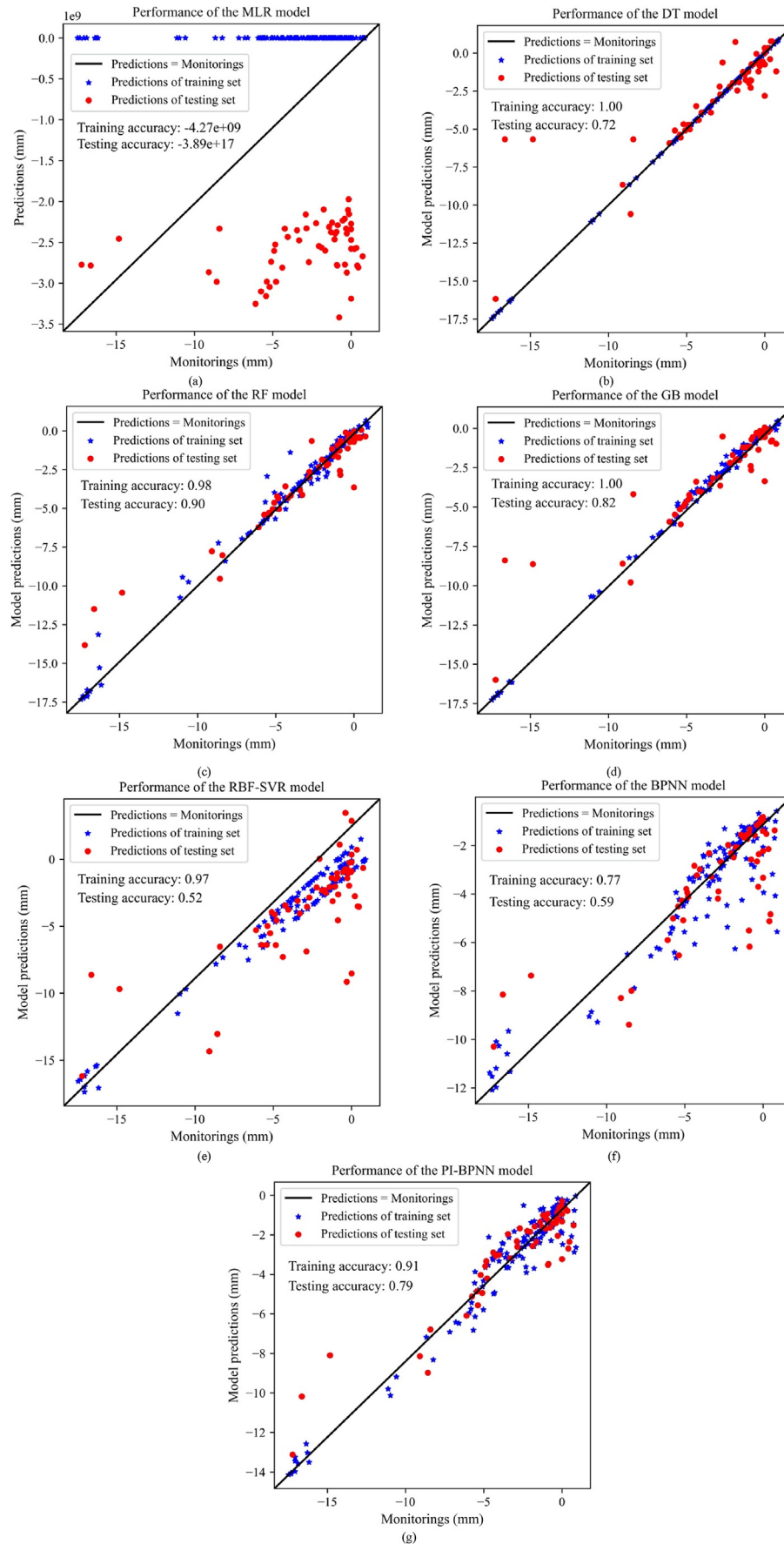


Fig. 6. Accuracy of applied ML algorithms: (a) MLR, (b) DT, (c) RF, (d) GB, (e) SVR, (f) BPNN, (g) PI-BPNN models.

Each ML model in this study will conduct multiple predictions to study the model stability on a small dataset. The variance s^2 (Eq. (22)) is applied to assess the model stability. It is calculated as the average squared deviation of each prediction from the averaged prediction. The more spread the data, the larger the variance concerning the mean.

$$s^2 = \frac{\sum (X - \bar{X})^2}{n-1} \quad (22)$$

where X is the predictions, \bar{X} is the average predictions, and n is the repeated time of prediction. The lower the variance, the more stable the ML model.

Evaluation of model stability is conducted in two steps here. First, an extreme case where the training set does not change is tested. Each single test sample has been predicted 1000 times. The variances of 1000 predictive results (Fig. 8) are calculated to assess model stabilities represented by dash lines. Variance trend represented by solid lines is obtained using the moving average technique. According to the results presented in Fig. 8, the prediction results vary significantly given the same training set and model architecture. The average variance of stability of BPNN and PI-BPNN are 5.76 and 3.58, respectively. BPNN and PI-BPNN show poor model stability, indicating that a small dataset fails to train hyperparameters in deep architecture fully. Compared with typical BPNN, the proposed PI-BPNN outperforms both in model accuracy and model stability. For shallow-structure ML algorithms (DT, RF, GB, SVR and MLR), the DT model shows high variance (1.06), while RF,

GB, SVR, and MLR models show good stability with variance almost equal to zero. Since the MLR model entirely failed to learn and predict tunneling-induced ground settlements, it will not be further discussed to save computing efforts. The second step is to further investigate the stability of RF, GB, and SVR models using a changeable dataset. The training and testing sets are randomly split 1000 times to train models repeatedly. Three algorithms are trained 1000 times to provide 1000 predictions. Results of model performance presented in Fig. 9 indicate that RF, GB and SVR can be considered stable models since the output performances do not change much, with variance less than 0.2. The average variance of GB, SVR, and RF are 5.5×10^{-3} , 5.39×10^{-27} and 3.02×10^{-27} , respectively. Thus, it can be concluded that the model stability ranging from best to worst is RF > SVR > GB > DT > PI-BPNN > BPNN.

Among all these applied learning algorithms, the RF model shows the best accuracy and stability.

5. Discussion

Applying ML techniques results in deriving patterns from existing datasets and approximating future behavior (i.e., predictions). It is important to evaluate ML performances, pros and cons to identify their ability of accurate prediction. Also, method selection should be appropriate based on the problem to be solved. Table 4 compares the seven ML algorithms applied for tunneling-induced ground settlement prediction with a small dataset.

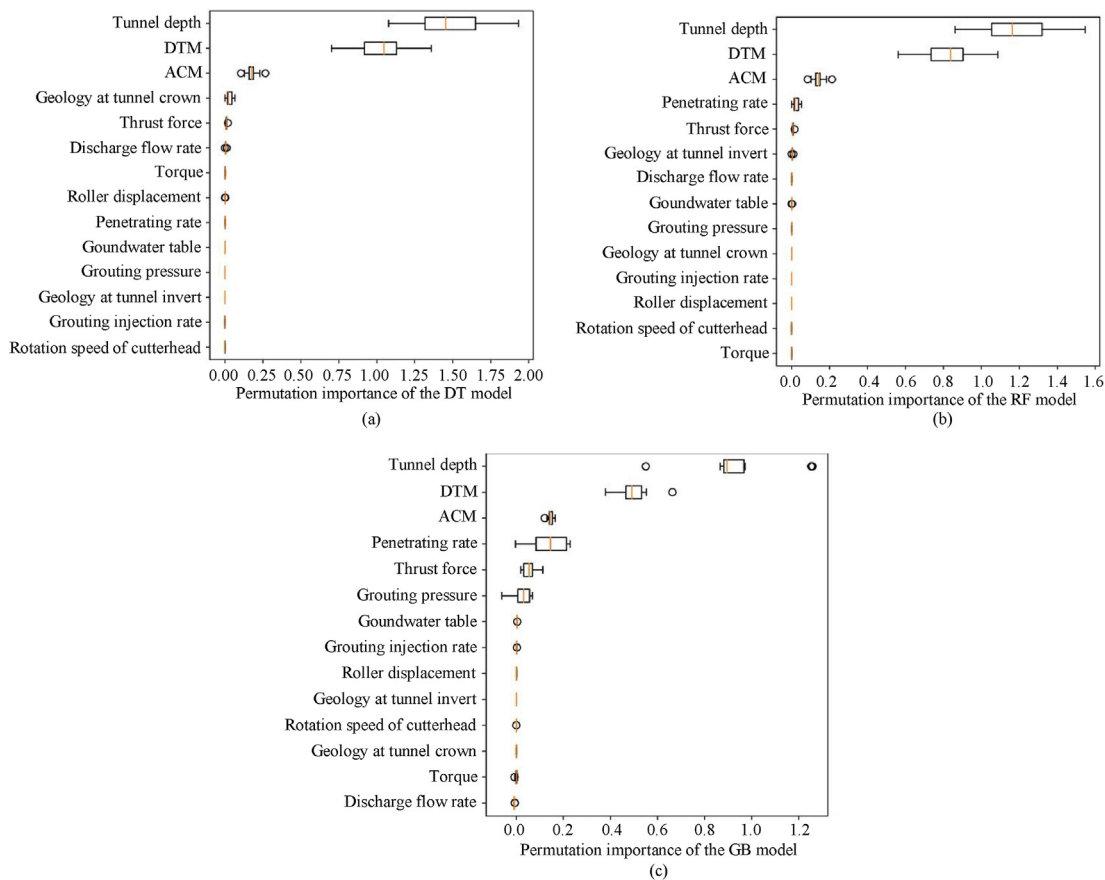


Fig. 7. Feature importance identified by ML algorithms: (a) DT, (b) RF, and (c) GB models.

A general advantage of ML models is the ability to handle multi-variate data and extract implicit relationships in a complex, dynamic and even chaotic environment. Moreover, they make no distributional assumptions about the predictor variables.

DT, RF and GB are three tree-based algorithms. The DT is a basic tree-structure algorithm, serving the base learner for both RF and GB. The RF and GB are state-of-the-art ensemble learning techniques, a method that combines the predictions from multiple ML algorithms (base learners) to provide more accurate predictions. A single DT might be relatively weaker compared with a set of trees. Theoretically, ensemble methods with more trees usually yield better results and serve as a more robust model (Elish et al., 2013). This study reveals that RF and GB outperform the DT model in terms of prediction accuracy and model stability. However, there is a trade-off between prediction accuracy and processing cost. The more the trees, the slower the process, since every tree has to be generated, processed, and analyzed. Also, the processing will be slower given more features. The difference between RF and GB is how trees are built. The RF builds each tree independently (in parallel), while the GB builds one tree at a time (sequentially) (Callens et al., 2020).

MLR models the linear relationship between independent and dependent variables, while the other six MLs work on processing nonlinear relationships. As its mathematic formula indicated (Eq. (1)), its main advantage is to provide the relative influence of one or multiple predictor variables intuitively. However, the MLR fails to approximate the nonlinear relationship between the influencing variables and ground settlements.

Data transforming is required to construct an SVR model, achieved by kernel functions. Kernel selection is dataset-dependent. It can be selected through the tried-and-error method. The kernel with the highest testing accuracy is chosen. Once the kernel is selected, SVR has a few hyperparameters needed to be determined, easy for implementation. The SVR shows good model stability but poor prediction accuracy. Given more penalties to underprediction and using the QE criterion, SVR has higher prediction accuracy than NN models.

In a multi-layer NN, each neuron is connected to other neurons with certain hyperparameters. Given many hyperparameters such as weight matrixes and bias vectors, NNs have a remarkable capability to learn and model nonlinear and complex relationships. However, many hyperparameters involved in the network may not

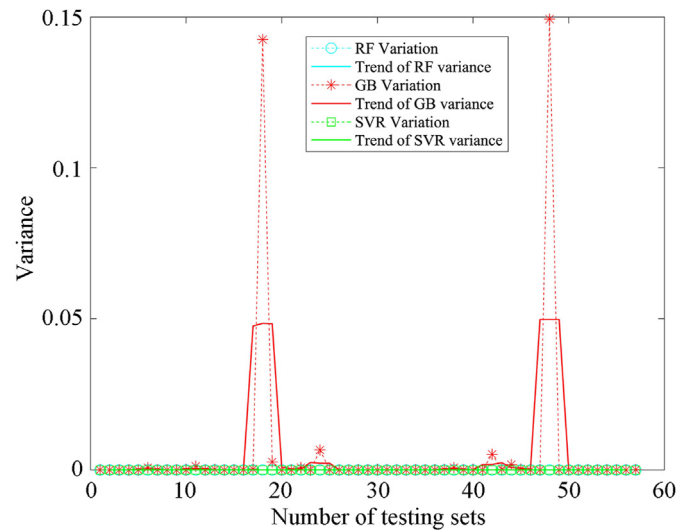


Fig. 9. Model stability with changeable training sets.

be well trained with the limited dataset to provide ideal predicting performance. The PI-BPNN proposed in this paper is designed for small datasets considering the relationship between model architecture and dataset size, showing better performance than typical BPNN. Although the NNs are powerful in solving complex problems, they are not effective enough to apply small datasets with both poor prediction accuracy and model stability.

6. Concluding remarks

Due to the limited construction period of any given tunnel project, the available monitoring datasets of ground deformation are small. Can ML methods effectively predict tunneling-induced ground settlement when the datasets are small? With both shallow- and deep-structure, seven ML algorithms have been applied for predicting tunneling-induced ground settlements. These ML methods include MLR, DT, RF, GB, SVR, BPNN, and PI-BPNN.

The seven ML algorithms are assessed from the perspective of model accuracy and stability. Our results showed that the MLR model failed to learn and predict. Furthermore, the nonlinear relationship between tunneling-induced ground settlements and possible influencing variables, including TBM operation parameters, encountered geological conditions, and tunnel geometry, was demonstrated. In the applications of nonlinear learning algorithms, the RF model showed the best performance with high prediction accuracy reaching 0.9 and low stability variance of 3.02×10^{-27} . While the BPNN, which is believed powerful in approximating complicated functions, shows worse performance than conventional shallow-structure learning algorithms. The proposed PI-BPNN method applied a permutation feature selection technique to decrease model complexity, achieving a better performance than typical BPNN for small datasets. Thus, no matter how attractive a model is, model selection should base on the full consideration of the problem to be solved. In conclusion, ML algorithms can be successfully used to predict tunneling-induced ground settlements even if the available dataset size is small. Furthermore, ensemble methods, such as RF and GB regression, outperform other learning algorithms.

Additionally, six variables that contributed most to the ground settlements occurring during the excavation process have been identified, including tunnel depth, DTM, ACM, grouting pressure,

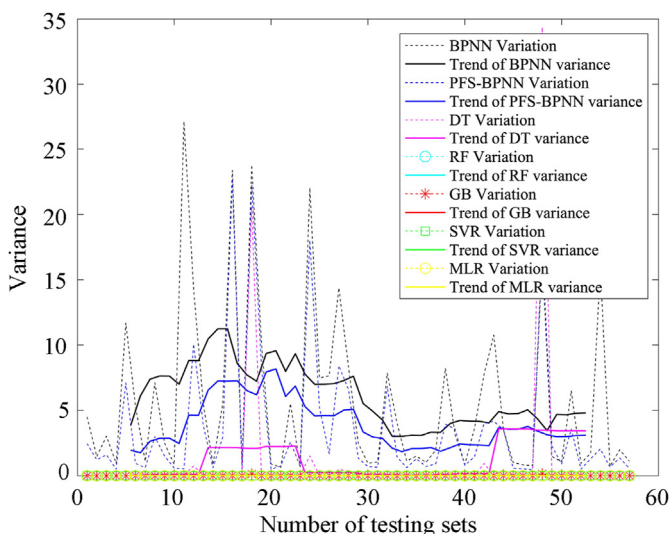


Fig. 8. Model stability with the same training set.

Table 4

Comparisons of models for predicting ML-based tunneling-induced ground settlement with small datasets.

Item	MLR	BPNN	PI-BPNN	SVR	DT	RF	GB
Approximation	Linear	Nonlinear	Nonlinear	Nonlinear with RBF kernel	Nonlinear	Nonlinear	Nonlinear
Learner	Single	Single	Single	Single	Single	Ensemble	Ensemble
Effectiveness	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Applicability	No	Yes	Yes	Yes	Yes	Yes	Yes
Prediction	Failed	Poor (<60%)	Medium (60%–80%)	Poor (<60%)	Medium (60%–80%)	Good (>80%)	Good (>80%)
accuracy	QE Failed	Poor (>1)	Poor (>1)	Poor (>1)	Medium (0.5–1)	Good (<0.5)	Good (<0.5)
Model stability (learning variance)	Good (≈0)	Poor (5.76)	Poor (3.58)	Good (≈0)	Medium (1.06)	Good (≈0)	Good (≈0)
Pros (+) and cons (–) in small dataset application	– Unable to solve complex problems	+ Good at processing complex nonlinear problems – Complicated architecture design with many hyperparameters	+ Work better than BPNN for small dataset	+ Few parameters needed to be determined – Tricky kernel selection	+ Good balance between prediction accuracy and calculation speed – Weaker compared with ensemble tree-based learners	+ Good prediction accuracy and stability – Relatively low computation speed	

penetrating rate and thrust force. Variables such as tunnel depth and ACM are determined by tunnel design and geological conditions. Thus, the ground settlements can be further controlled by TBM operations. Three TBM operations, i.e., grouting pressure, penetrating rate, and thrust force, should be paid much attention to during the tunnel excavation.

It should be noticed that our conclusions are only suitable and applicable for small dataset cases, such as those tunneling cases listed in Table 1. They are not valid for tunneling cases with big datasets which are out of the scope of this research.

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.

Acknowledgments

This research was funded by the University Transportation Center for Underground Transportation Infrastructure (UTC-UTI) at the Colorado School of Mines under Grant No. 69A3551747118 from the US Department of Transportation (DOT). The opinions expressed in this paper are those of the authors and not of the DOT.

References

- Ahangari, K., Moeinossadat, S.R., Behnia, D., 2015. Estimation of tunneling-induced settlement by modern intelligent methods. *Soils Found.* 55 (4), 737–748.
- Altmann, A., Tološi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics* 26 (10), 1340–1347.
- Ates, U., Bilgin, N., Copur, H., 2014. Estimating torque, thrust and other design parameters of different type TBMs with some criticism to TBMs used in Turkish tunneling projects. *Tunn. Undergr. Space Technol.* 40, 46–63.
- Bouayad, D., Emeriault, F., 2017. Modeling the relationship between ground surface settlements induced by shield tunneling and the operational and geological parameters based on the hybrid PCA/ANFIS method. *Tunn. Undergr. Space Technol.* 68, 142–152.
- Boubou, R., Emeriault, F., Kastner, R., 2010. Artificial neural network application for the prediction of ground surface movements induced by shield tunneling. *Can. Geotech. J.* 47 (11), 1214–1233.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Bsaibes, A., Courault, D., Baret, F., et al., 2009. Albedo and LAI estimates from FORMOSAT-2 data for crop monitoring. *Remote Sens. Environ.* 113 (4), 716–729.
- Callens, A., Morichon, D., Abadie, S., Delpey, M., Lique, B., 2020. Using random forest and gradient boosting trees to improve wave forecast at a specific location. *Appl. Ocean Res.* 104, 102339.
- Chao, G.Y., Tsai, T.J., Lu, T.J., et al., 2011. A new approach to prediction of radiotherapy of bladder cancer cells in small dataset analysis. *Expert Syst. Appl.* 38 (7), 7963–7969.

- Chen, R., Zhang, P., Wu, H., Wang, Z., Zhong, Z., 2019. Prediction of shield tunneling-induced ground settlement using machine learning techniques. *Front. Struct. Civ. Eng.* 13, 1363–1378.
- Clough, G.W., Leca, E., 1993. EPB shield tunneling in mixed face conditions. *J. Geotech. Eng.* 119 (10), 1640–1656.
- Darabi, A., Ahangari, K., Noorzad, A., Arab, A., 2012. Subsidence estimation utilizing various approaches – a case study: tehran No. 3 subway line. *Tunn. Undergr. Space Technol.* 31, 117–127.
- Ding, L., Wang, F., Luo, H., Yu, M., Wu, X., 2013. Feedforward analysis for shield-ground system. *J. Comput. Civ. Eng.* 27 (3), 231–242.
- Elish, M., Helmy, T., Hussain, M., 2013. Empirical study of homogeneous and heterogeneous ensemble models for software development effort estimation. *Math. Probl. Eng.* 2013, 312067.
- Feng, S., Chen, Z., Luo, H., et al., 2021. Tunnel boring machines (TBM) performance prediction: a case study using big data and deep learning. *Tunn. Undergr. Space Technol.* 110, 103636.
- Goh, A.T.C., Hefney, A.M., 2010. Reliability assessment of EPB tunnel-related settlement. *Geomech. Eng.* 2 (1), 57–69.
- Hasanipanah, M., Noorian-Bidgoli, M., Jahed Armaghani, D., Khamesi, H., 2016. Feasibility of PSO-ANN model for predicting surface settlement caused by tunneling. *Eng. Comput.* 32, 705–715.
- Heaton, J., 2015. Artificial Intelligence for Humans. In: *Deep Learning and Neural Networks*, ume 1. Heaton Research, Inc., St. Louis, MO, USA.
- Hegazy, T., Fazio, P., Moselhi, O., 1994. Developing practical neural network applications using back-propagation. *Comput. Aided Civ. Infrastruct. Eng.* 9 (2), 145–159.
- Kim, C.Y., Bae, G.J., Hong, S.W., Park, C.H., Moon, H.K., Shin, H.S., 2001. Neural network based prediction of ground surface settlements due to tunneling. *Comput. Geotech.* 28 (6–7), 517–547.
- Kitchin, R., Lauriault, T.P., 2015. Small data in the era of big data. *Geojournal* 80, 463–475.
- Kitchin, R., 2013. Big data and human geography: opportunities, challenges and risks. *Dialogues Hum. Geogr.* 3 (3), 262–267.
- Kohestani, V.R., Bazargan-Lari, M.R., Asgari-Marnani, J., 2017. Prediction of maximum surface settlement caused by earth pressure balance shield tunneling using random forest. *J. AI Data Min.* 5 (1), 127–135.
- Lyashevskaya, O., Harma, C., Minto, C., Clarke, M., Brophy, D., 2020. Long-term trends in herring growth primarily linked to temperature by gradient boosting regression trees. *Ecol. Inf.* 60, 101154.
- Mahmoodzadeh, A., Mohammadi, M., Daraei, A., Farid Hama Ali, H., Kameran Al-Salihi, N., Mohammed Dler Omer, R., 2020. Forecasting maximum surface settlement caused by urban tunneling. *Autom. Construct.* 120, 103375.
- Marto, A., Hajjassani, M., Kalatehjari, R., Namazi, E., Sohaei, H., 2012. Simulation of longitudinal surface settlement due to tunneling using artificial neural network. *Int. Rev. Mod. Simul.* 5 (2), 1024–1031.
- Moghaddasi, M.R., Noorian-Bidgoli, M., 2018. ICA-ANN, ANN and multiple regression models for prediction of surface settlement caused by tunneling. *Tunn. Undergr. Space Technol.* 79, 197–209.
- Mooney, M., Grasmick, J., Kenneally, B., Fang, Y., 2016. The role of slurry TBM parameters on ground deformation: field results and computational modeling. *Tunn. Undergr. Space Technol.* 57, 257–264.
- Mukherjee, I., Routroy, S., 2012. Comparing the performance of neural networks developed by using Levenberg–Marquardt and quasi–Newton with the gradient descent algorithm for modeling a multiple response grinding process. *Expert Syst. Appl.* 39 (3), 2397–2407.
- Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Obs. Geoinf.* 18, 399–406.

- Ocak, I., Seker, S.E., 2013. Calculation of surface settlements caused by EPBM tunneling using artificial neural network, SVM, and Gaussian processes. *Environ. Earth Sci.* 70, 1263–1276.
- Pasini, A., 2015. Artificial neural networks for small dataset analysis. *J. Thorac. Dis.* 7 (5), 953–960.
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pekel, E., 2020. Estimation of soil moisture using decision tree regression. *Theor. Appl. Climatol.* 139, 1111–1119.
- Pourtaghi, A., Lotfollahi-Yaghin, M.A., 2012. Wavenet ability assessment in comparison to ANN for predicting the maximum surface settlement caused by tunneling. *Tunn. Undergr. Space Technol.* 28, 257–271.
- Qiu, J., Wu, Q., Ding, G., Xu, Y., Feng, S., 2016. A survey of machine learning for big data processing. *EURASIP J. Adv. Signal Process.* 2016 (67).
- Santos, O.J., Celestino, T.B., 2008. Artificial neural networks analysis of São Paulo subway tunnel settlement data. *Tunn. Undergr. Space Technol.* 23 (5), 481–491.
- Sheil, B., Suryasentana, S.K., Mooney, M.A., Zhu, H., 2020. Machine learning to inform tunnelling operations: recent advances and future trends. *Proc. Inst. Civ. Eng. – Smart Infrastruct. Constr.* <https://doi.org/10.1680/jsmic.20.00011>.
- Shi, J., Ortigao, J.A.R., Bai, J., 1998. Modular neural networks for predicting settlements during tunneling. *J. Geotech. Geoenviron. Eng.* 124 (5), 389–395.
- Suwansawat, S., Einstein, H.H., 2006. Artificial neural networks for predicting the maximum surface settlement caused by EPB shield tunneling. *Tunn. Undergr. Space Technol.* 21 (2), 133–150.
- Tetko, I.V., Villa, A.E.P., 1997. Efficient partition of learning data sets for neural network training. *Neural Network.* 10 (8), 1361–1374.
- Vladimir, N.V., 2000. *The Nature of Statistical Learning Theory*. Springer, New York, USA.
- Wang, F., Gou, B., Qin, Y., 2013. Modeling tunneling-induced ground surface settlement development using a wavelet smooth relevance vector machine. *Comput. Geotech.* 54, 25–32.
- Yang, K., Trewn, J., 2004. *Multivariate Statistical Methods in Quality Management*. McGraw-Hill, New York, USA.
- Yeh, C.Y., Huang, C.W., Lee, S.J., 2011. A multiple-kernel support vector regression approach for stock market price forecasting. *Expert Syst. Appl.* 38 (3), 2177–2186.
- Zhang, K., Lyu, H.M., Shen, S.L., Zhou, A., Yin, Z.Y., 2020c. Evolutionary hybrid neural network approach to predict shield tunneling-induced ground settlements. *Tunn. Undergr. Space Technol.* 106, 103594.
- Zhang, L., Wu, X., Ji, W., AbouRizk, S.M., 2017. Intelligent approach to estimation of tunnel-induced ground settlement using wavelet packet and support vector machines. *J. Comput. Civ. Eng.* 31 (2), 04016053.
- Zhang, P., Wu, H.N., Chen, R.P., Chan, T.H.T., 2020b. Hybrid meta-heuristic and machine learning algorithms for tunneling-induced settlement prediction: a comparative study. *Tunn. Undergr. Space Technol.* 99, 103383.
- Zhang, W., Zhang, R., Wu, C., et al., 2020a. State-of-the-art review of soft computing applications in underground excavations. *Geosci. Front.* 11 (4), 1095–1106.
- Zhou, L., Pan, S., Wang, J., Vasilakos, A.V., 2017. Machine learning on big data: opportunities and challenges. *Neurocomputing* 237, 350–361.
- Zhu, M., Zhu, H., Guo, F., Chen, X., Ju, J., 2021. Tunnel condition assessment via cloud model-based random forests and self-training approach. *Comput. Aided Civ. Infrastruct. Eng.* 36 (2), 164–179.



Dr. Wendy Zhou is a Professor at the Department of Geology and Geological Engineering and the Director of the GIS & GeoInformatics Program, Colorado School of Mines in Golden, Colorado, USA. Wendy Zhou received her Ph.D. degree in geological engineering from Missouri University of Science & Technology, USA, in 2001. Her research interests include: (1) geographic information system (GIS) and remote sensing applications in geohazard analysis, environmental impact study, and natural resource management, (2) interferometric synthetic aperture radar (InSAR) and machine learning applications in ground movement monitoring and landslide assessment, (3) rock mass characterization, and (4) numerical modeling for the ground thermal regime and stress-strain distribution in the rock or soil mass. She has supervised more than 30 masters and 9 Ph.D. students. She has published more than 60 papers in peer-reviewed journals or peer-reviewed conference proceedings. She has also contributed to 7 book chapters. She has been serving as an Associate Editor of the *Journal of Applied Remote Sensing* since 2014 and a Topic Editor of *Remote Sensing* since 2020.