Geomechanics and Engineering, *Vol. 42, No. 3 (2025) 179-189*
https://doi.org/10.12989/gae.2025.42.3.179

Optimized ground settlement classification during TBM tunneling by combining machine learning with statistical analysis

Kibeom Kwon[1a], Minkyu Kang[2b], Dongku Kim[3c], Khanh Pham[4,5d] and Hangseok Choi*[6]

[1]*Future and Fusion Lab of Architectural, Civil and Environmental Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea*
[2]*Center for Defense Acquisition and Requirement Analysis, Korea Institute for Defense Analyses, 37 Hoegi-ro, Dongdaemun-gu, Seoul 130-871, Republic of Korea*
[3]*Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology (KICT), 283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do, Republic of Korea*
[4]*School of Civil Engineering and Management, International University, Ho Chi Minh City, Vietnam*
[5]*Vietnam National University, Ho Chi Minh City, Vietnam*
[6]*School of Civil, Environmental and Architectural Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea*

(*Received October 18, 2024, Revised May 28, 2025, Accepted July 8, 2025*)

Abstract. Ground settlement management is crucial in tunnel boring machine (TBM) operations. Previous attempts to predict ground settlement have required substantial assumptions or information, complicating the explicit determination of their predictive criteria. This study developed an optimized system with simplicity and transparency for predicting ground settlements. By selecting three key features through correlation analysis and literature reviews, the optimized system was constructed to predict three settlement classes (heaving, normal, and large settlement) using a combination of machine learning and statistical analysis. The optimized system achieved an accuracy of 0.846, with recall values of 0.667 for heaving, 0.895 for normal, and 0.750 for large settlement. These results surpassed those of two comparison models that employed eight features and ensemble learning algorithms. Notably, the comparison models failed to correctly predict any instances of large settlement, highlighting the effectiveness of the optimized system in handling imbalanced datasets. Unlike conventional black-box models, the optimized system explicitly defined the predictive criteria. Moreover, among the four instances misclassified by the optimized system, three involved minor settlements within ±3 mm. The consistent decrease in accuracy when excluding each feature from the optimized system highlighted the importance of incorporating these features to accurately identify patterns in settlement predictions.

Keywords: ground settlement; machine learning; optimized system; statistical analysis; tunnel boring machine

1. Introduction

The increasing population density and rising social costs associated with traffic congestion have driven a growing demand for tunnel construction (Kwon *et al*. 2023). Among various tunnel construction methods, tunnel boring machines (TBMs) have been extensively adopted owing to their superior eco-friendliness, stability, and constructability. However, the inherent geological, mechanical, and combined uncertainties encountered during TBM tunneling can result in adverse accidents, such as collapse and water/mud inflow (Sousa and Einstein 2021). Among these accidents, managing ground settlement within acceptable limits is crucial to prevent severe damage to existing surface and underground structures (Ercelebi *et al*. 2011, Chung *et al*. 2021). Therefore, an optimized system for predicting ground settlement that accounts for geological and TBM operational parameters is essential to ensure the safety and efficiency of TBM tunneling projects.

Previous studies on predicting ground settlement induced by tunneling have employed three primary approaches: empirical-statistical (Peck 1969, Atkinson and Potts 1977), analytical (Sagaseta 1987, Verruijt 1997, Loganathan and Poulos 1998), and numerical (Franzius 2005, Chakeri 2013, Huang *et al*. 2015) methods. The empirical-statistical approach employs simple mathematical formulas based on prior tunnel excavation experience and data collected from extensive field observations. This method has been widely used in practical applications (Moghaddasi and Noorian-Bidgoli 2018). However, applying this approach to TBM tunneling is challenging because it doesn't easily account for TBM operation features like thrust force and torque. The analytical approach provides closed-form solutions for predicting induced ground settlement through theoretical deductions. Despite its theoretical foundation, deriving suitable solutions using this approach is inherently complex due to

*Corresponding author, Professor
E-mail: hchoi2@korea.ac.kr
[a]Ph.D.
[b]Ph.D.
[c]Ph.D.
[d]Professor

ISSN: 2005-307X (Print), 2092-6219 (Online)

necessary assumptions (e.g., isotropic, homogeneous, and incompressible soil) and uncertain relationships between settlement-inducing features, such as soil-structure interactions (Zhou *et al*. 2017, Kim *et al*. 2020). Furthermore, most analytical results cannot capture TBM operation features, similar to empirical results (Chen *et al*. 2022). With the rapid advancement of computer technology and numerical algorithms, several studies have utilized numerical approaches to predict ground settlement (Kim *et al*. 2018, Zhou *et al*. 2023). Unlike previous methods, numerical approaches can account for TBM operation and site-specific features (e.g., geology and geometry). However, they have limitations stemming from the use of constitutive models with substantial assumptions about uncertain material properties, operational features, and boundary conditions (Ye *et al*. 2022). Moreover, considerable computational resources are needed to ensure the reliability of obtained results.

Recently, machine learning approaches have been introduced in various studies to predict TBM tunneling-induced ground settlement (Suwansawat and Einstein 2006, Kohestani *et al*. 2017, Chen *et al*. 2019, Mahmoodzadeh *et al*. 2020, Kim *et al*. 2022a, Kim *et al*. 2022b, Liu *et al.* 2022, Ye *et al.* 2022, Zhou *et al.* 2023). This approach can consider TBM operation and site-specific features using extensive data and effectively analyze complex relationships between settlement-inducing features. However, most machine learning-based studies have relied on black box-type algorithms, complicating the provision of explicit predictive criteria. Moreover, previous studies may have limitations due to their requirement for extensive or localized feature information.

This study aims to develop an optimized system with simplicity and transparency for classifying ground settlements. First, correlation analysis is conducted to identify key features utilized in the optimized system. Then, the predictive criteria for multiple settlement classes are established through a combination of machine learning and statistical analysis. To validate the practical applicability of the developed system, comparative analysis is conducted with ensemble learning models. Furthermore, error and sensitivity analysis are performed to evaluate the predictive performance regarding ground settlement values and to assess the influence of each feature on the predictions.

2. Overview of machine learning algorithms

2.1 Decision tree (DT)

Decision tree (DT) divides the predictor space into multiple regions, resulting in a tree-like structure developed based on a set of splitting rules (Jong *et al*. 2021). Generally, the Gini index is used as the splitting criteria, as expressed in Eq. (1)

$$Gini(D) = 1 - \sum_{K=1}^{K} (\frac{|C_K|}{|D|})^2 \quad (1)$$

where $|D|$ represents the number of instances in the dataset D and $|C_k|$ denotes the number of instances in class C_K, with k denoting the different classes in the dataset. When a specific feature is chosen to split the dataset into two subsets (D_1 and D_2), the Gini index of set D under the condition of the feature is calculated using Eq. (2).

$$Gini(D) = \frac{|D_1|}{D} Gini(D_1) + \frac{|D_2|}{D} Gini(D_2) \quad (2)$$

DT recursively selects the optimal feature that minimizes the Gini index at each split, continuing this process until the stop conditions, such as reaching a predefined maximum depth of DT, are satisfied. As DT explicitly provides the predictive criteria unlike black box-type algorithms (Salimi *et al*. 2018), DT was employed to develop an initial model in this study.

2.2 Random forest (RF)

Random forest (RF) is a bagging-type ensemble learning algorithm introduced by Breiman (2001) based on a set of decision trees. For each decision tree construction, a bootstrap sample (i.e., a randomly selected sample from the original training set with replacement) is drawn, with the number of data instances and features in the sample being randomly determined. The decision tree then evolves based on the selected bootstrap sample. In this manner, n bootstrap samples and their corresponding n decision trees are repeatedly constructed. By aggregating all the decision tree outputs, the final prediction is determined as the mode (in classification) or the average value (in regression). Because of its high predictive performance and low risk of overfitting, RF has been widely applied in various fields.

2.3 Extreme gradient boosting (XGB)

Extreme gradient boosting (XGB) is a gradient boosting-type ensemble learning algorithm extensively applied in diverse domains, initially proposed by Chen and Guestrin (2016). The core of gradient boosting-type algorithms lies in the sequential creation of numerous weak learners, with each learner focused on rectifying the residual (i.e., the gradient of a loss function) left by its predecessor. Finally, a robust model emerges as an ensemble of these weak learners. Specifically, the aggregated prediction of these weak learners forms the outcome of the robust model.

Nevertheless, gradient boosting-type algorithms run the risk of constructing an excessively complex model by exclusively minimizing the residual, potentially leading to overfitting. To mitigate this risk, XGB incorporates an advanced objective function that combines the loss function with a regularization term, optimally balancing predictive performance and model complexity. In this study, RF and XGB were used to develop comparison models for validating the optimized system.

3. Dataset construction

3.1 Project overview

Table 1 Slurry shield TBM specifications

Description	Specification
TBM excavation diameter	7.4 m
TBM length	11.45 m
Max. thrust force	47,897 kN
Max. torque	5 MN·m
Opening ratio	29 %
Segment diameter	7.1 m (OD), 6.5 m (ID)
Segment width	1.5 m

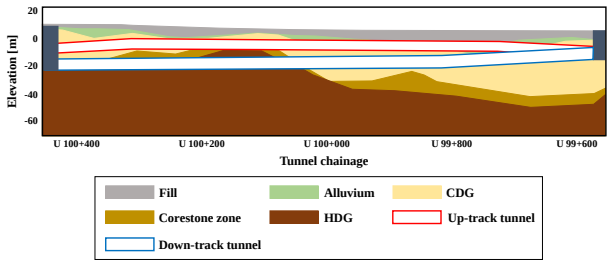


Fig. 1 Longitudinal geological profile

This study used a dataset obtained from a slurry shield TBM tunneling project in Hong Kong. Twin-bored slurry shield TBM subway tunnels with a total length of 850 m, comprising both up-track and down-track tunnels, were constructed at the tunneling site. The up-track tunnel was excavated five months after the down-track tunnel to minimize ground disturbance. These tunnels were categorized as shallow tunnels because their cover depth was less than twice the TBM excavation diameter (Park *et al*. 2018). The detailed specifications of the slurry shield TBM are summarized in Table 1.

After the TBM passed through the second tunnel (i.e., up-track tunnel), daily settlement measurements were made at 253 locations along the tunnel alignment, and the settlement value finally measured at each location was used. Four different ground types were encountered in the up-track alignment: fill, alluvium, completely decomposed granite (CDG), and highly decomposed granite (HDG). HDG and CDG can be classified as highly fractured normal to weak rock and very weak rock to weathered soil, respectively (Park *et al*. 2018). Representative geotechnical properties of these ground types are summarized in Table 2.

The longitudinal geological profile of the tunnel alignment is illustrated in Fig. 1. More detailed information on data acquisition procedures and ground types can be referred to (Kim *et al*. 2022a, Kim *et al*. 2022b).

3.2 Data pre-processing

Generally, features that induce settlement are categorized into three categories: geometry, geology, and TBM operation characteristics (Suwansawat and Einstein 2006). In this study, the features commonly addressed and readily obtainable in various TBM tunneling projects were selected to ensure the practical applicability of the developed optimized system (Mahmoodzadeh *et al*. 2020, Liu *et al*. 2022).

Table 2 Geotechnical properties of the encountered ground types (Kim *et al*. 2022b)

Ground type	N-value	Cohesion [kPa]	Friction angle [°]	Permeability [m/s]
Fill	10	0	35	3.15×10^{-5}
Alluvium	D*<4: 10 D≥4: 1.67D+3.33	0	35	2.32×10^{-5}
CDG	D<5: 15 D≥5: 2.71D+1.43	8	38	3.34×10^{-6}
HDG	200	12	40	4.60×10^{-6}

* D: Depth [m]

Previous studies have considered the cover depth and the horizontal distance between the tunnel centerline and settlement monitoring points (HTM) as geometry features (Ahangari *et al*. 2015, Mahmoodzadeh *et al*. 2020, Ding *et al*. 2022, Kim *et al*. 2022a, Kim *et al*. 2022b, Liu *et al*. 2022). These features become particularly important for predicting settlement in shallow tunnels, similar to the site of this study. Geology features adopted here include the standard penetration test (SPT) N-value and the distance between the ground surface and the groundwater level (DSG). These features account for the potential influence of ground strength and groundwater on settlement. Notably, the SPT N-value, being more subdivided than other geotechnical properties (refer to Table 2), is likely to influence machine learning predictions (Liu *et al*. 2022). Concurrently, this study focused on five TBM operation features commonly measured in tunneling projects: thrust force, torque, face pressure, penetration rate, and grouting volume. These features have demonstrated their contributions to ground settlements in several previous studies (Fargnoli *et al*. 2013, Mooney *et al*. 2016, Kwong *et al*. 2019, Lee *et al*. 2021, Samadi *et al*. 2021, Kim *et al*. 2022b).

Since settlement measurements were conducted after the excavation of the up-track tunnel, this study used the features in each category, corresponding to the up-track tunnel excavation. Although, in certain chainages, settlements were measured at various locations with different horizontal distances. In this study, the settlements closest to the tunnel centerline were exclusively considered.

The raw data for geometric, geological, and operational features were collected at different frequencies. This study synchronized all these features on a per-ring basis. The geometric and geological data corresponding to the midpoint of each ring were used, while the mean value of the non-zero operational data recorded during each segment ring excavation was considered. Subsequently, data from segment rings close to settlement measurement points were adopted. Statistical descriptions of the settlement-inducing features and the settlements are presented in Table 3.

This study categorized the measured settlements (S) into three classes based on their magnitude: heaving (S<0 mm), normal (0 mm<S≤10 mm), and large settlement (S>10 mm), as summarized in Table 4. A total of 129 instances

Table 3 Statistical description of the selected features and settlements

Category	Feature	Min	Q1	Median	Q3	Max	COV*	Unit
Geometry	Cover depth	6.70	7.10	7.50	8.90	12.80	0.17	m
	HTM	0.002	4.206	10.118	16.413	39.892	0.82	m
Geology	N-value	10.00	15.00	16.00	22.00	36.00	0.36	-
	DSG	3.02	3.88	4.31	4.68	7.23	0.23	m
Operation	Thrust force	8,676	12,850	13,700	15,400	26,000	0.24	kN
	Torque	0.20	0.50	0.70	1.00	1.80	0.42	MN·m
	Face pressure	1.20	1.45	1.59	1.84	2.45	0.16	bar
	Penetration rate	13.00	27.00	35.00	40.00	48.00	0.25	mm/min
	Grouting volume	5.30	6.30	6.43	6.60	7.60	0.05	m^3
-	Settlement	-6.90	2.50	4.40	8.00	16.00	0.63	mm

*COV: Coefficient of variation denoted as the standard deviation normalized by the mean

Table 4 Categorized settlement classes defined in this study

No	Class	Count	Proportion
1	Heaving (S<0 mm)	16	12.4 %
2	Normal (0 mm<S≤10 mm)	92	71.3 %
3	Large settlement (S>10 mm)	21	16.3 %

comprise nine features (including geometry, geology, and operation features) and one target variable (settlement class).

3.3 Data analysis

Identifying highly correlated features with the categorized settlement class among all the collected features is important. To simplify the developed optimized system for practical use by TBM operators, this study selectively adopted a limited number of features based on correlation analysis and literature reviews. Given the combined influences of site-specific and TBM operational features on ground settlements, the study chose not to select features exclusively from one category. Through Pearson correlation analysis, the relationships between these features and the target class (i.e., the settlement class) were identified. The results of this correlation analysis, presented by Pearson's coefficient, are shown in Fig. 2.

Among TBM operational features, the correlation analysis indicates that thrust force, torque, and face pressure were the most relevant to the settlement class. Especially, thrust force and torque exhibited strong negative correlations with the settlement class, represented by Pearson's coefficients of -0.49 and -0.48, respectively. In other words, increases and decreases in these features (i.e., thrust force and torque) were closely associated with the heaving and large settlement classes, respectively. These trends were also evident in the results of statistical analysis, as later presented in Figs. 3(a) and 3(b). In addition, thrust force and torque are widely recognized as influential factors affecting ground settlement (Zhang *et al*. 2019, Kim *et al*. 2020, Hussaine and Mu 2022). Face pressure is also a critical parameter for controlling induced ground

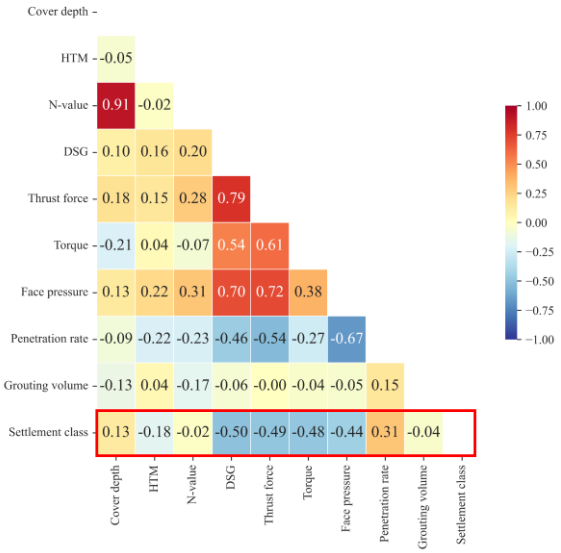


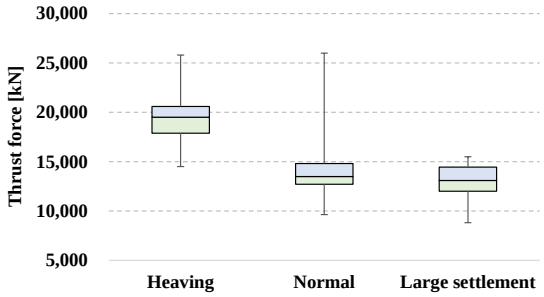
Fig. 2 Results of Pearson correlation analysis

settlements due to its significant influence on face stability during TBM tunneling (Zhou *et al*. 2023). Fig. 2 reinforces this perspective by demonstrating a strong correlation between face pressure and the settlement class, indicated by a Pearson's coefficient of -0.44. Nevertheless, it is important to note that face pressure also exhibited a strong correlation with thrust force (i.e., Pearson's coefficient ≥ 0.7) and had a low coefficient of variation (COV), as shown in Table 3. Consequently, to simplify the optimized system, this study replaced face pressure with thrust force. Consequently, thrust force and torque were selected as key features for the optimized prediction system.

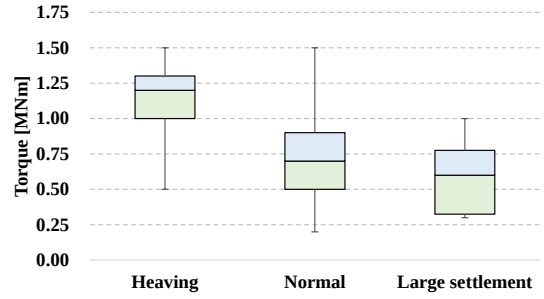
Next, while most site-specific features showed weak correlations with the settlement class, DSG exhibited a high Pearson's coefficient of -0.50. However, to avoid redundancy in feature composition, thrust force was used as a substitute for DSG, given its strong correlation with DSG (as shown in Fig. 2) and its well-established impact on ground settlement. Instead, this study adopted the N-value as a site-specific feature to account for the significance of weak ground (Sarna *et al*. 2022), despite its relatively low

Table 5 Statistical description of the selected features corresponding to each class

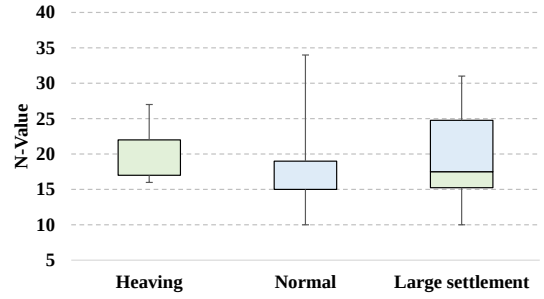
Indices	Thrust force [kN]			Torque [MN·m]			N-value		
	Heaving	Normal	Large settlement	Heaving	Normal	Large settlement	Heaving	Normal	Large settlement
Min	14,500	9,628	8,822	0.5	0.2	0.3	16	10	10
Q1	17,890	12,700	12,000	1.0	0.5	0.325	17	15	15.25
Median	19,500	13,500	13,100	1.2	0.7	0.6	22	15	17.5
Q3	20,600	14,800	14,444	1.3	0.9	0.775	22	19	24.75
Max	25,800	26,000	15,500	1.5	1.5	1.0	27	34	31



(a) Thrust force



(b) Torque



(c) N-value

Fig. 3 Box plots of each feature corresponding to each class

correlation with the settlement class. Given the proportional relationship between the N-value and elastic modulus (Wagh and Bambole 2024), a low N-value in weak ground could result in increased radial displacements. Additionally, the low undrained shear strength of soils associated with a low N-value has the potential to cause severe settlements (Chou and Bobet 2002).

As a result, the dataset consisted of three features: thrust force, torque, and N-value, along with the target class, which is the settlement class. In the statistical analysis of dataset distributions, five key indices were considered: the minimum and maximum values, the lower and upper quartiles (Q1 and Q3), and the median. These indices of the three features (thrust force, torque, and N-value) are summarized for each settlement class in Table 5 and visually depicted in a box plot shown in Fig. 3.

4. System development

4.1 Criterion for heaving class

The primary objective of this study is to develop an optimized system for ground settlement prediction. Specifically, this system aims to classify the three classes, as outlined in Table 4, using three key features: thrust force, torque, and N-value. However, establishing predictive criteria for this system from the beginning, especially when dealing with four-dimensional data encompassing thrust force, torque, N-value, and settlement class, is a complex challenge. Therefore, this study first constructed an initial model employing a machine learning approach to provide preliminary predictive criteria.

The initial model was produced using the DT algorithm, chosen for its capacity to offer clear visualization and an intuitive interpretation of predictive criteria, in contrast to black box-type algorithms such as the RF and XGB algorithms. Subsequently, the dataset was divided into training and test sets, with 80% of the instances allocated for training and the remaining 20% for testing. Throughout this process, equal proportions of each settlement class were maintained in both datasets. It is important to note that the same test set was used for evaluating prediction accuracies corresponding to the initial model, the optimized system, and comparison models. In addition, this study utilized Bayesian optimization (BO) combined with 5-fold cross-validation for hyperparameter tuning.

BO is an efficient strategy for finding optimal solutions, such as hyperparameter combinations that maximize an unknown objective function (e.g., accuracy and coefficient of determination). It relies on two core components: surrogate model and acquisition function. The surrogate model estimates the posterior distribution of the objective function using prior knowledge derived from previously evaluated samples (e.g., hyperparameter combinations) based on Bayes' theorem (Su *et al*. 2022). The acquisition function then leverages this estimated distribution to determine the next sample for evaluation (Zhou *et al*. 2021).

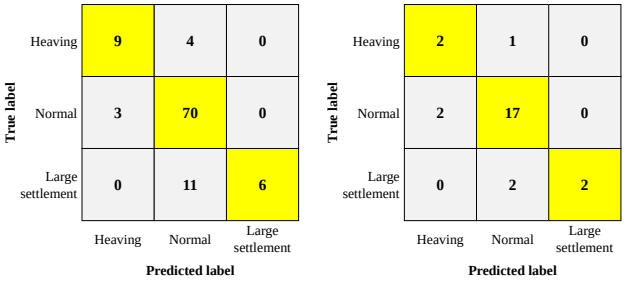
In this study, the Gaussian process (GP) was employed

Table 6 Search spaces for the explored DT hyperparameters

Hyperparameter	Description	Search space
max_depth	Maximum depth of tree model	2–4
min_samples_split	Minimum number of samples required to split a node	3–10
min_samples_leaf	Minimum number of samples required in a leaf	3–10

Table 7 Predictive performance for each class obtained by the initial model

Performance metric	Training	Test
Accuracy	0.825	0.808
Recall_H	0.692	0.667
Recall_N	0.959	0.895
Recall_L	0.353	0.500



(a) Training (b) Test

Fig. 4 Confusion matrix for the initial model

as the surrogate model, while expected improvement (EI) were adopted as the acquisition function. Detailed information on BO based on GP and EI can be found in Zhang *et al*. (2020). To control the training period, early stopping was employed by monitoring performance metric improvements and training time, thereby mitigating overfitting and conserving time and resources.

Table 6 presents the explored hyperparameters and their corresponding search spaces. It can be noted that the *max_depth* hyperparameter was set to less than five to mitigate overfitting and reduce complexity.

The initial model was developed with the following optimal hyperparameters: *max_depth* =4, *min_samples_split* =10, and *min_samples_leaf* =3. Table 7 and Fig. 4 show predictive performance in the training and test phases. In this study, four performance metrics were used: accuracy, recall for heaving (Recall_H), recall for normal (Recall_N), and recall for large settlement (Recall_L). Since failing to identify the heaving and large settlement classes incurs a higher cost than misclassifying the normal class in terms of anomaly detection (Liang *et al*. 2024), this study incorporated the recall of each class along with accuracy.

Although over two-thirds of the heaving class instances in both training and test sets were correctly predicted, significant prediction errors were observed from the large settlement class in the test phase, with the recall for that class of 0.500. Throughout the training and test phases, only eight out of the 21 instances for the large settlement class

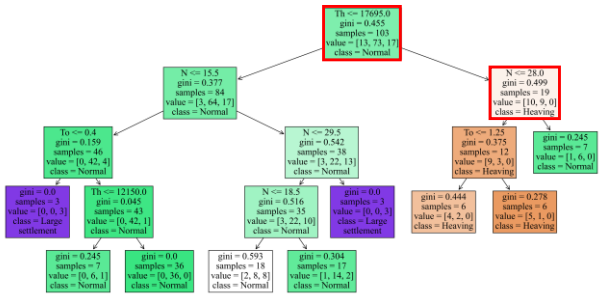


Fig. 5 Structure of the DT-based initial model

were accurately predicted, while the rest were misclassified as the normal class. This misclassification issue can be attributed to the initial model's lack of emphasis on predicting the large settlement class, which is a common challenge when applying machine learning algorithms to imbalanced datasets.

Consequently, according to the DT structure of the initial model (Fig. 5), the predictive criterion for the heaving class (Eq. (3)) was adopted for the optimized system developed in this study. However, additional analyses were required to address the misclassifications between the normal and large settlement classes, thereby deriving their predictive criteria.

$$N \leq 28 \; and \; Th > 17{,}695 \; kN \tag{3}$$

where N: N-value and Th: Thrust force.

4.2 Criterion for large settlement class

In this study, statistical analysis was employed to establish the criteria for predicting the normal and large settlement classes, as summarized in Table 5. Fig. 3 revealed that both thrust force and torque decrease as settlement magnitude increases, following the order of the heaving, normal, and large settlement classes. While the decline in thrust force is less pronounced, the trend in torque is more noticeable. Hence, this study defined the criterion for predicting the large settlement class as follows: instances that do not meet the heaving class criterion while concurrently exhibiting torque values below the lower quartile for the normal class. Based on Eq. (3) and Table 5, the criterion for predicting the large settlement class is summarized in Eq. (4).

$$[N > 28 \; or \; Th \leq 17{,}695 \; kN] \; and \; [To < 0.5 \; MN \cdot m] \tag{4}$$

where N: N-value, Th: Thrust force, and To: Torque.

4.3 Criterion for normal class

The criterion for predicting the normal class can be determined as follows: instances that do not meet the conditions specified in Eqs. (3) and (4). Therefore, an optimized system comprising three predictive criteria was formulated by combining machine learning with statistical analysis. Fig. 6 summarizes the established predictive criteria. The predictive criteria indicate that high thrust force in the ground where the N-value is not high is

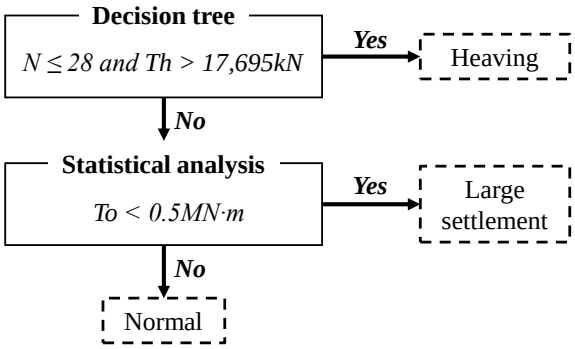


Fig. 6 Optimized system comprising the established predictive criteria

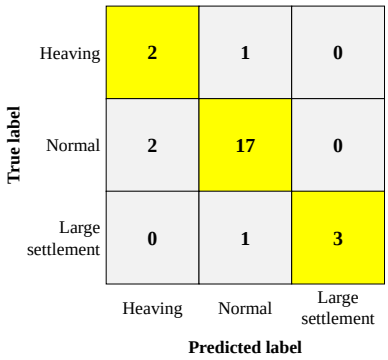


Fig. 7 Confusion matrix for the optimized system

associated with heaving, while low torque is related to large settlement. These trends are consistent with findings from previous studies (Chen *et al*. 2019, Kim *et al*. 2020, Zhong *et al*. 2021, Guo *et al*. 2022, Zhou *et al*. 2023).

5. System application

5.1 Predictive performance

Table 8 presents predictive performances of both the initial model and the optimized system when evaluated on the same test set, as described in Section 4. Fig. 7 illustrates a confusion matrix for the optimized system.

According to Table 8, the optimized system improved the predictive performance compared to the initial model, achieving an accuracy of 0.846 and recall values of 0.667 for heaving, 0.895 for normal, and 0.750 for large settlement. Notably, in the large settlement class, the optimized system correctly predicted three out of four instances, whereas the initial model correctly predicted only two. Meanwhile, the number of correctly predicted instances for the heaving and normal classes remained consistent. Importantly, all incorrectly predicted instances were classified into adjacent classes, reducing the risk of applying conflicting countermeasures. Consequently, the optimized system, which combines machine learning with statistical analysis, outperformed the DT-based initial model, while both systems can explicitly provide predictive criteria.

Table 8 Predictive performances of the initial model and the optimized system

Performance metric	Initial model	Optimized system
Accuracy	0.808	0.846
$Recall_H$	0.667	0.667
$Recall_N$	0.895	0.895
$Recall_L$	0.500	0.750

Table 9 Search spaces for RF and XGB hyperparameters

Algorithm	Description	Search space
RF	*n_estimators*	100–500
	max_depth	1–20
	min_samples_split	5–10
	min_samples_leaf	5–10
XGB	*n_estimators*	100–500
	max_depth	1–20
	min_child_weight	5–10
	gamma	0–1
	subsample	0–1
	learning_rate	0.01–1

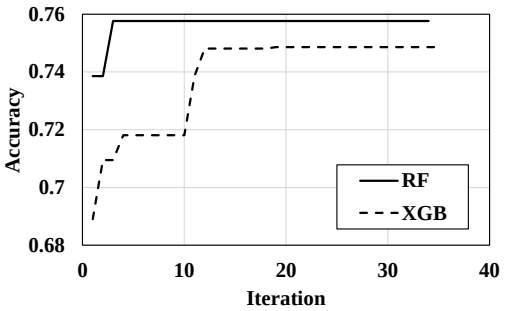


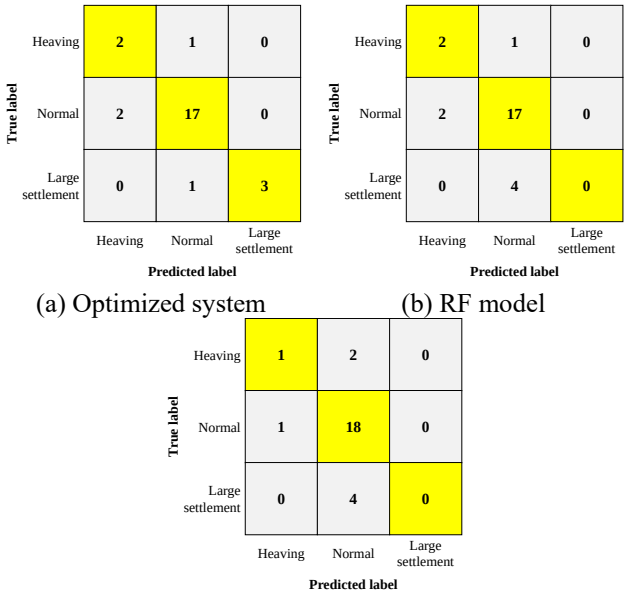
Fig. 8 Iteration curves for comparison models

5.2 Comparative analysis

The purpose of this section is to validate the practical applicability of the developed optimized system through comparative analysis with different comparison models. Typically, bagging and boosting-based ensemble learning algorithms have been widely employed to address imbalanced datasets (Sun *et al*. 2015, Zhou *et al*. 2019). As shown in Table 4, the high proportion of the dataset corresponds to the normal class, signifying its imbalance. Consequently, this study developed two comparison models using representative bagging and boosting-based ensemble learning algorithms, namely RF and XGB. Among the nine input features listed in Table 3, the cover depth was excluded from these comparison models to eliminate redundancy, due to its high correlation with the N-value, as shown in Fig. 2 (Xue *et al*. 2023). This feature selection highlights the simplicity of the optimized system, which utilizes only three features. For hyperparameter tuning of these comparison models, BO and 5-fold cross-validation were employed. The search spaces for RF and XGB hyperparameters are described in Table 9.

Table 10 Predictive performances of the optimized system and comparison models

Performance metric	Optimized system	RF model	XGB model
Accuracy	0.846	0.731	0.731
$Recall_H$	0.667	0.667	0.333
$Recall_N$	0.895	0.895	0.947
$Recall_L$	0.750	0.000	0.000



(a) Optimized system (b) RF model

(c) XGB model

Fig. 9 Confusion matrices for the optimized system and comparison models

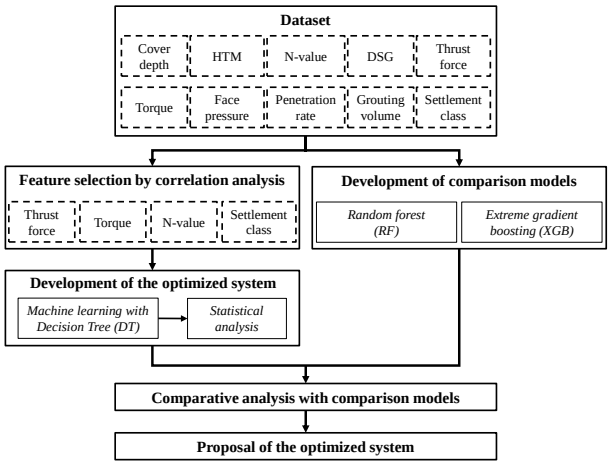


Fig. 10 Flow chart outlining the optimized system

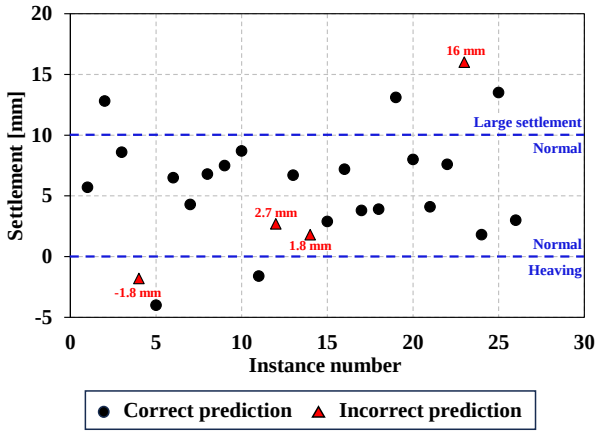


Fig. 11 Results of error analysis

In machine learning, an iteration curve based on a performance metric can display the optimization process (Qiu and Zhou 2023). Fig. 8 illustrates the iteration curves based on the average accuracy from 5-fold cross-validation for each comparison model, indicating the impact of early stopping on both models. Table 10 presents the predictive performances of the optimized system and the comparison models when evaluated on the same test set, while Fig. 9 illustrates the confusion matrix for each model.

The comparative analysis revealed that the optimized system outperformed the RF and XGB models in most performance metrics, except for recall of the normal class. Notably, both comparison models failed to detect any instances of the large settlement class. The imbalanced dataset (refer to Table 4) can cause the comparison models to be biased toward predicting the normal class. In contrast, the optimized system, which combines machine learning with statistical analysis, effectively mitigated this bias and provided more balanced predictive performance.

Moreover, this study compared the predictive performance of the optimized system with that of a settlement classification model from Kim *et al*. (2022b). Despite differences in the number of settlement classes addressed, the optimized system of this study achieved an accuracy of 0.846, surpassing the accuracy of 0.724 reported in the previous study.

In addition, the optimized system was validated with its simplicity and transparency, as it was developed using only three features (i.e., thrust force, torque, and N-value) and provided explicit predictive criteria. The overall flow chart outlining the optimized system developed in this study is presented in Fig. 10.

5.3 Error and sensitivity analysis

This study performed error analysis examining the prediction results for the instances within the test set, with the consideration of their settlement values. Fig. 11 displays the settlement values of these instances and indicates whether the optimized system predicted them correctly or incorrectly. It was evident that three out of four incorrectly predicted instances had settlement values within ±3 mm. These inaccuracies, closely associated with subtle ground deformations, were expected to have a relatively minor impact on the safety and efficiency of TBM tunneling projects. In contrast, it should be noted that the instance characterized by a substantial settlement value of 16 mm was incorrectly predicted, which could have significant adverse consequences. As shown in Table 3, this instance represented the maximum settlement value in the entire dataset. It can be inferred that the developed optimized system faced challenges in accurately predicting this

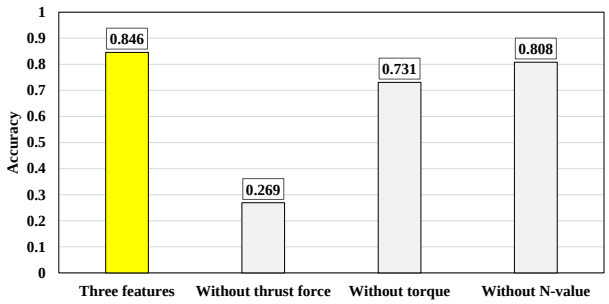


Fig. 12 Results of sensitivity analysis

instance due to the scarcity of large settlement data, particularly those exceeding 15 mm. This observation highlights the need for an extensive dataset encompassing a substantial number of instances with large settlement values. Such a dataset would not only enhance the prediction accuracy of the developed optimized system but also ensure the safety and efficiency of TBM tunneling projects.

Furthermore, this study conducted a sensitivity analysis to evaluate the importance of the three features within the optimized system. The analysis involved comparing the reduction in accuracy resulting from the exclusion of each feature from the predictive criteria. Fig. 12 presents the results of the sensitivity analysis.

While an accuracy of 0.846 was achieved using all three features, excluding each of thrust force, torque, and N-value resulted in accuracies of 0.269, 0.731, and 0.808, respectively. These consistent reductions suggest that incorporating all three features is essential for accurately identifying the patterns involved in settlement predictions. Notably, the substantial drop in accuracy from 0.846 to 0.269 upon excluding thrust force highlights the critical role of thrust force in settlement predictions.

5.4 Discussion

The primary contribution of the optimized system is its ability to predict three settlement classes with greater transparency and simplicity. Unlike conventional black-box models, the optimized system explicitly provided three predictive criteria. Furthermore, it outperformed the comparison models by utilizing only three key features (thrust force, torque, and N-value), thereby reducing the need for extensive and localized feature information. This simplicity and transparency enhance the practical applicability of the optimized system.

Despite the validated practical applicability of the optimized system, this study has limitations that may affect its reliability and robustness. The optimized system was developed using a relatively small dataset of only 129 instances collected per segment ring. Moreover, the optimized system faces challenges in updating as excavation progresses. Future studies could address these limitations by constructing a more extensive dataset through data generation or augmentation techniques, such as generative adversarial networks and synthetic minority oversampling technique (Kuntalp and Düzyel 2024, Kwon *et al*. 2024, Kwon *et al*. 2025). Enhancing the optimized system with time-series data could also improve its effectiveness as a warning system for detecting heaving and significant settlements. Furthermore, it would be more practical to predict ground settlements in the remaining sections by continuously updating the optimized system using data from sections already excavated by TBM.

6. Conclusions

This study proposed an optimized system for ground settlement management by combining machine learning with statistical analysis. The practical applicability of the optimized system was validated by comparison with ensemble learning models. The principal findings of this study are summarized as follows.

- With three key features (thrust force, torque, and N-value), the optimized system incorporating three predictive criteria demonstrated superior performance in predicting three settlement classes. It achieved an accuracy of 0.846, with recall values of 0.667 for heaving, 0.895 for normal, and 0.750 for large settlement. Moreover, these predictive criteria are consistent with previous studies regarding the relationships between these features and settlements. Unlike the black-box comparison models, the optimized system can provide explicit predictive criteria.
- The optimized system, utilizing only three key features, outperformed the comparison models that used eight features. Notably, the comparison models failed to predict any instances of the large settlement class, as they could not address the imbalanced dataset. In contrast, the optimized system successfully predicted three out of four instances of this class. These findings highlight the optimized system's ability to mitigate the effects of imbalanced datasets while maintaining a simple structure based on only three features.
- Out of the four instances incorrectly predicted by the optimized system, three were associated with settlement values within ±3 mm, which generally have a minor impact on TBM tunneling. In addition, the observed reductions in accuracy when excluding each feature from the optimized system demonstrated the importance of incorporating these features to accurately identify patterns involved in settlement predictions.

Acknowledgments

The research described in this paper was supported by the National R&D Project for Smart Construction Technology (No. RS-2020-KA157074) and for Consecutive Excavation Technological Development Project of Tunnel Boring Machine (No. RS-2022-00144188) funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure, and Transport.

References

Ahangari, K., Moeinossadat, S.R. and Behnia, D. (2015), "Estimation of tunnelling-induced settlement by modern intelligent methods", *Soils. Found.*, **55**(4), 737-748. https://doi.org/10.1016/j.sandf.2015.06.006.

Atkinson, J.H. and Potts, D.M. (1977), "Subsidence above shallow tunnels in soft ground", *J. Geotech. Engrg. Div.*, **103**(4), 307-

325. https://doi.org/10.1061/AJGEB6.0000402.
Breiman, L. (2001), "Random forests", *Mach. Learn.*, **45**, 5-32. https://doi.org/10.1023/A:1010933404324.
Chakeri, H., Ozcelik, Y. and Unver, B. (2013), "Effects of important factors on surface settlement prediction for metro tunnel excavated by EPB", *Tunn. Undergr. Sp. Tech.*, **36**, 14-23. https://doi.org/10.1016/j.tust.2013.02.002.
Chen, R., Zhang, P., Wu, H., Wang, Z. and Zhong, Z. (2019), "Prediction of shield tunneling-induced ground settlement using machine learning techniques", *Front. Struct. Civ. Eng.*, **13**, 1363-1378. https://doi.org/10.1007/s11709-019-0561-3.
Chen, R.P., Song, X., Meng, F.Y., Wu, H.N. and Lin, X.T. (2022), "Analytical approach to predict tunneling-induced subsurface settlement in sand considering soil arching effect", *Comput. Geotech.*, **141**, 104492. https://doi.org/10.1016/j.compgeo.2021.104492.
Chen, T. and Guestrin, C. (2016), "Xgboost: A scalable tree boosting system", *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794. https://doi.org/10.1145/2939672.2939785.
Chen, Y., Xu, Y., Jamhiri, B., Wang, L. and Li, T. (2022), "Predicting uniaxial tensile strength of expansive soil with ensemble learning methods", *Comput. Geotech.*, **150**, 104904. https://doi.org/10.1016/j.compgeo.2022.104904.
Chou, W.I. and Bobet, A. (2002), "Predictions of ground deformations in shallow tunnels in clay", *Tunn. Undergr. Sp. Tech.*, **17**(1), 3-19. https://doi.org/10.1016/S0886-7798(01)00068-2.
Chung, H., Park, J., Kim, B.K., Kwon, K., Lee, I.M. and Choi, H. (2021), "A causal network-based risk matrix model applicable to shield TBM tunneling projects", *Sustainability*, **13**(9), 4846. https://doi.org/10.3390/su13094846.
Ding, Z., Zhao, L.S., Zhou, W.H. and Bezuijen, A. (2022), "Intelligent prediction of multi-factor-oriented ground settlement during TBM tunneling in soft soil", *Front. Built. Environ.*, **8**, 848158. https://doi.org/10.3389/fbuil.2022.848158.
Ercelebi, S.G., Çopur, H. and Ocak, I. (2011), "Surface settlement predictions for Istanbul Metro tunnels excavated by EPB-TBM", *Environ. Earth. Sci.*, **62**, 357-365. https://doi.org/10.1007/s12665-010-0530-6.
Fargnoli, V., Boldini, D. and Amorosi, A. (2013), "TBM tunnelling-induced settlements in coarse-grained soils: The case of the new Milan underground line 5", *Tunn. Undergr. Sp. Tech.*, **38**, 336-347. https://doi.org/10.1016/j.tust.2013.07.015.
Franzius, J.N., Potts, D.M. and Burland, J.B. (2005), "The influence of soil anisotropy and K0 on ground surface movements resulting from tunnel excavation", *Géotechnique*, **55**(3), 189-199. https://doi.org/10.1680/geot.2005.55.3.189.
Guo, D., Li, J., Li, X., Li, Z., Li, P. and Chen, Z. (2022), "Advance prediction of collapse for TBM tunneling using deep learning method", *Eng. Geol.*, **299**, 106556. https://doi.org/10.1016/j.enggeo.2022.106556.
Huang, H., Gong, W., Khoshnevisan, S., Juang, C.H., Zhang, D., and Wang, L. (2015), "Simplified procedure for finite element analysis of the longitudinal performance of shield tunnels considering spatial soil variability in longitudinal direction", *Comput. Geotech.*, **64**, 132-145. https://doi.org/10.1016/j.compgeo.2014.11.010.
Hussaine, S.M. and Mu, L. (2022), "Intelligent prediction of maximum ground settlement induced by EPB shield tunneling using automated machine learning techniques", *Mathematics*, **10**(24), 4637. https://doi.org/10.3390/math10244637.
Jong, S.C., Ong, D.E.L. and Oh, E. (2022), "State-of-the-art review of geotechnical-driven artificial intelligence techniques in underground soil-structure interaction", *Tunn. Undergr. Sp. Tech.*, **113**, 103946. https://doi.org/10.1016/j.tust.2021.103946.
Kim, D., Kwon, K., Pham, K., Oh, J.Y. and Choi, H. (2022a), "Surface settlement prediction for urban tunneling using machine learning algorithms with Bayesian optimization", *Automat. Constr.*, **140**, 104331. https://doi.org/10.1016/j.autcon.2022.104331.
Kim, D., Pham, K., Oh, J.Y., Lee, S.J. and Choi, H. (2022b), "Classification of surface settlement levels induced by TBM driving in urban areas using random forest with data-driven feature selection", *Automat. Constr.*, **135**, 104109. https://doi.org/10.1016/j.autcon.2021.104109.
Kim, D., Pham, K., Park, S., Oh, J.Y. and Choi, H. (2020), "Determination of effective parameters on surface settlement during shield TBM", *Geomech. Eng.*, **21**(2), 153-164. https://doi.org/10.12989/gae.2020.21.2.153.
Kim, K., Oh, J.Y., Lee, H., Kim, D. and Choi, H. (2018), "Critical face pressure and backfill pressure in shield TBM tunneling on soft ground", *Geomech. Eng.*, **15**(3), 823-831. https://doi.org/10.12989/gae.2018.15.3.823.
Kohestani, V.R., Bazarganlari, M.R. and Asgari Marnani, J. (2017), "Prediction of maximum surface settlement caused by earth pressure balance shield tunneling using random forest", *J. Artif. Intell. Data Min.*, **5**(1), 127-135. https://doi.org/10.22044/jadm.2016.748.
Kuntalp, M. and Düzyel, O. (2024), "A new method for GAN-based data augmentation for classes with distinct clusters", *Exp. Syst. Appl.*, **235**, 121199. https://doi.org/10.1016/j.eswa.2023.121199.
Kwon, K., Kang, M., Kim, D. and Choi, H. (2023), "Prioritization of hazardous zones using an advanced risk management model combining the analytic hierarchy process and fuzzy set theory", *Sustainability*, **15**(15), 12018. https://doi.org/10.3390/su151512018.
Kwon, K., Hwang, B., Park, H., Oh, J.Y. and Choi, H. (2024), "Enhancing machine learning-based anomaly detection for TBM penetration rate with imbalanced data manipulation", *J. Korean Tunn Undergr Sp. Assoc.*, **26**(5), 519-532. https://doi.org/10.9711/KTAJ.2024.26.5.519.
Kwon, K., Choi, H., Jung, J., Kim, D. and Shin, Y.J. (2025), "Prediction of abnormal TBM disc cutter wear in mixed ground condition using interpretable machine learning with data augmentation", *J. Rock Mech. Geotech. Eng.*, **17**(4), 2059-2071. https://doi.org/10.1016/j.jrmge.2024.05.027.
Kwong, A.K.L., Ng, C.C.W. and Schwob, A. (2019), "Control of settlement and volume loss induced by tunneling under recently reclaimed land", *Undergr. Space*, **4**(4), 289-301. https://doi.org/10.1016/j.undsp.2019.03.005.
Lee, H.K., Song, M.K. and Lee, S.S. (2021), "Prediction of Subsidence during TBM operation in mixed-face ground conditions from realtime monitoring data", *Appl. Sci.*, **11**(24), 12130. https://doi.org/10.3390/app112412130.
Liang, J., Liang, S., Ma, L., Zhang, H., Dai, J. and Zhou, H. (2024), "Leak detection for natural gas gathering pipeline using spatio-temporal fusion of practical operation data", *Eng. Appl. Artif. Intell.*, **133**, 108360. https://doi.org/10.1016/j.engappai.2024.108360.
Liu, L., Zhou, W. and Gutierrez, M. (2022), "Effectiveness of predicting tunneling-induced ground settlements using machine learning methods with small datasets", *J. Rock. Mech. Geotech. Eng.*, **14**(4), 1028-1041.
Loganathan, N. and Poulos, H.G. (1998), "Analytical prediction for tunneling-induced ground movements in clays", *J. Geotech. Geoenviron. Eng.*, **124**(9), 846-856. https://doi.org/10.1061/(ASCE)1090-0241(1998)124:9(846).
Mahmoodzadeh, A., Mohammadi, M., Daraei, A., Ali, H.F.H., Al-Salihi, N.K. and Omer, R.M.D. (2020), "Forecasting maximum surface settlement caused by urban tunneling", *Automat. Constr.*, **120**, 103375. https://doi.org/10.1016/j.autcon.2020.103375.

Moghaddasi, M.R. and Noorian-Bidgoli, M. (2018), "ICA-ANN, ANN and multiple regression models for prediction of surface settlement caused by tunneling", *Tunn. Undergr. Sp. Tech.*, **79**, 197-209. https://doi.org/10.1016/j.tust.2018.04.016.

Mooney, M.A., Grasmick, J., Kenneally, B. and Fang, Y. (2016), "The role of slurry TBM parameters on ground deformation: Field results and computational modelling", *Tunn. Undergr. Sp. Tech.*, **57**, 257-264. https://doi.org/10.1016/j.tust.2016.01.007.

Park, H., Oh, J.Y., Kim, D. and Chang, S. (2018), "Monitoring and analysis of ground settlement induced by tunnelling with slurry pressure-balanced tunnel boring machine", *Adv. Civ. Eng.*, **2018**(1), 5879402. https://doi.org/10.1155/2018/5879402.

Peck, R.B. (1969), "Deep excavations and tunnelling in soft ground", *Proceedings of the 7th International Conference on Soil Mechanics and Foundation Engineering*, Mexico City, Mexico, 225-290.

Qiu, Y. and Zhou, J. (2023), "Short-term rockburst damage assessment in burst-prone mines: an explainable XGBOOST hybrid model with SCSO algorithm", *Rock Mech. Rock Eng.*, **56**(12), 8745-8770. https://doi.org/10.1007/s00603-023-03522-w.

Sagaseta, C. (1987), "Analysis of undrained soil deformation due to ground loss", *Géotechnique*, **37**(3), 301-320. https://doi.org/10.1680/geot.1987.37.3.301.

Salimi, A., Faradonbeh, R.S., Monjezi, M. and Moormann, C. (2018), "TBM performance estimation using a classification and regression tree (CART) technique", *Bull. Eng. Geol. Environ.*, **77**, 429-440. https://doi.org/10.1007/s10064-016-0969-0.

Samadi, H., Hassanpour, J. and Farrokh, E. (2021), "Maximum surface settlement prediction in EPB TBM tunneling using soft computing techniques", *J. Phys: Conf. Ser.*, **1973**(1). https://doi.org/10.1088/1742-6596/1973/1/012195.

Sarna, S., Gutierrez, M., Mooney, M. and Zhu, M. (2022), "Predicting upcoming collapse incidents during tunneling in rocks with continuation length based on influence zone", *Rock. Mech. Rock. Eng.*, **55**(10), 5905-5931. https://doi.org/10.1007/s00603-022-02971-z.

Sousa, R.L. and Einstein, H.H. (2021), "Lessons from accidents during tunnel construction", *Tunn. Undergr. Sp. Tech.*, **113**, 103916. https://doi.org/10.1016/j.tust.2021.103916.

Su, J., Wang, Y., Niu, X., Sha, S. and Yu, J. (2022), "Prediction of ground surface settlement by shield tunneling using XGBoost and Bayesian Optimization", *Eng. Appl. Artif. Intell.*, **114**, 105020. https://doi.org/10.1016/j.engappai.2022.105020.

Sun, Z., Song, Q., Zhu, X., Sun, H., Xu, B. and Zhou, Y. (2015), "A novel ensemble method for classifying imbalanced data", *Pattern. Recogn.*, **48**(5), 1623-1637. https://doi.org/10.1016/j.patcog.2014.11.014.

Suwansawat, S. and Einstein, H.H. (2006), "Artificial neural networks for predicting the maximum surface settlement caused by EPB shield tunneling", *Tunn. Undergr. Sp. Tech.*, **21**(2), 133-150. https://doi.org/10.1016/j.tust.2005.06.007.

Verruijt, A. (1997), "A complex variable solution for a deforming circular tunnel in an elastic half-plane", *Int. J. Numer. Anal. Method. Geomech.*, **21**(2), 77-89. https://doi.org/10.1002/(SICI)10969853(199702)21:2<77::AID-NAG857>3.0.CO;2-M

Wagh, J.D. and Bambole, A.N. (2024), "Improved correlation of soil modulus with SPT N values", *Open Eng*, **14**(1), 20240046. https://doi.org/10.1515/eng-2024-0046.

Xue, Y.D., Luo, W., Chen, L., Dong, H.X., Shu, L.S. and Zhao, L. (2023), "An intelligent method for TBM surrounding rock classification based on time series segmentation of rock-machine interaction data", *Tunn. Undergr. Sp. Tech.*, **140**, 105317. https://doi.org/10.1016/j.tust.2023.105317.

Ye, X.W., Jin, T. and Chen, Y.M. (2022), "Machine learning-based forecasting of soil settlement induced by shield tunneling construction", *Tunn. Undergr. Sp. Tech.*, **124**, 104452. https://doi.org/10.1016/j.tust.2022.104452.

Zhang, P., Chen, R.P.. and Wu, H.N. (2019), "Real-time analysis and regulation of EPB shield steering using Random Forest", *Automat. Constr.*, **106**, 102860. https://doi.org/10.1016/j.autcon.2019.102860.

Zhang, Q., Hu, W., Liu, Z. and Tan, J. (2020), "TBM performance prediction with Bayesian optimization and automated machine learning", *Tunn. Undergr. Sp. Tech.*, **103**, 103493. https://doi.org/10.1016/j.tust.2020.103493.

Zhong, Z., Li, C., Liu, X., Fan, Y. and Liang, N. (2021), "Analysis of ground surface settlement induced by the construction of mechanized twin tunnels in soil-rock mass mixed ground", *Tunn. Undergr. Sp. Tech.*, **110**, 103746. https://doi.org/10.1016/j.tust.2020.103746.

Zhou, J., Qiu, Y., Zhu, S., Armaghani, D.J., Khandelwal, M. and Mohamad, E.T. (2021), "Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization",\ *Undergr. Space*, **6**(5), 506-515. https://doi.org/10.1016/j.undsp.2020.05.008.

Zhou, J., Shi, X., Du, K., Qiu, X., Li, X. and Mitri, H.S. (2017), "Feasibility of random-forest approach for prediction of ground settlements induced by the construction of a shield-driven tunnel", *Int. J. Geomech.*, **17**(6), 04016129. https://doi.org/10.1061/(ASCE)GM.1943-5622.0000817.

Zhou, X., Zhao, C. and Bian, X. (2023), "Prediction of maximum ground surface settlement induced by shield tunneling using XGBoost algorithm with golden-sine seagull optimization", *Comput. Geotech.*, **154**, 105156. https://doi.org/10.1016/j.compgeo.2022.105156.

Zhou, Y., Li, S., Zhou, C. and Luo, H. (2019), "Intelligent approach based on random forest for safety risk prediction of deep foundation pit in subway stations", *J. Comput. Civ. Eng.*, **33**(1), 05018004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000796.

GC