applied sciences

Article

Prediction and Generalization Capability of Machine Learning Models for Shield TBM-Induced Settlement

Ji-Seok Yun [1,*], Wan-Kyu Yoo [1], Gi-Jun Lee [1], Je-Kyum Lee [1], Young-Suk Song [2], Chang-Yong Kim [1] and Han-Eol Kim [3]

[1] Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang-si 10223, Republic of Korea; lyu5553@kict.re.kr (W.-K.Y.); gijunlee@kict.re.kr (G.-J.L.); jklee@kict.re.kr (J.-K.L.); cykim@kict.re.kr (C.-Y.K.)
[2] Geological Safety Division, Korea Institute of Geoscience and Mineral Resources, Daejeon 34132, Republic of Korea; yssong@kigam.re.kr
[3] Underground Safety Assessment Center, Korea Expressway Corporation Research Institute, Hwaseong-si 18489, Republic of Korea; hekim0501@ex.co.kr
* Correspondence: jsyun0519@kict.re.kr

Abstract

Ground settlement induced by shield tunnel boring machine (TBM) excavation is a major geotechnical concern in urban tunneling because it may affect the safety of adjacent structures and underground infrastructure. In this study, machine learning models were developed to predict the maximum settlement induced by shield TBM excavation using a three-dimensional numerical analysis database comprising 320 simulation cases generated from combinations of tunnel diameter (*D*), ground elastic modulus (*E*), face pressure (*FP*), and backfill pressure (*BP*). Random forest (RF) and extreme gradient boosting (XGBoost) models were developed and compared with an existing regression-based settlement prediction equation. Predictive performance and generalization capability were evaluated using random split and GroupKFold validation techniques. Under random split validation, RF achieved the highest predictive performance, with a coefficient of determination of 0.997 and a root mean square error of 0.438 mm, followed by XGBoost. Both machine learning models outperformed the existing settlement prediction equation. However, model performance decreased substantially under GroupKFold validation, indicating limited generalization capability under unseen *D*–*E* grouped conditions. The results demonstrate that the developed machine learning models provide accurate predictions within the range of tunnel–ground conditions represented by the adopted numerical analysis database. The findings highlight the importance of evaluating both predictive performance and generalization capability, particularly when machine learning models developed from numerical analysis databases are applied beyond the conditions represented in the training database.

Keywords: shield TBM; ground settlement; machine learning; generalization capability

Academic Editor: Ricardo Castedo Ruiz

Received: 13 June 2026
Revised: 8 July 2026
Accepted: 9 July 2026
Published: 10 July 2026

1. Introduction

The growing concentration of population and traffic congestion in urban areas has increased the demand for underground space development [1,2]. In particular, the use of tunnel boring machine (TBM) methods has expanded alongside the rapid construction of underground expressways and urban rail transit systems. Compared with conventional blasting methods, TBM excavation offers several advantages—such as reduced noise and vibration, improved construction safety, and higher excavation efficiency—making it widely preferred in urban environments [3–5]. Nevertheless, the ground deformation and surface settlement induced by tunnel excavation remain major geotechnical issues. Excavation-induced settlement may cause damage to adjacent structures and reduce ground stability, potentially leading to severe ground failures in extreme cases [6]. Therefore, accurate prediction of ground settlement induced by TBM excavation is essential for ensuring construction safety and supporting long-term infrastructure management [7].

Various approaches have been proposed to predict ground settlement induced by tunnel excavation. Following the pioneering Gaussian-distribution-based empirical method proposed by Peck [8], numerous empirical prediction methods have been developed [9–12]. Theoretical approaches have also been proposed to estimate ground deformation associated with tunnel excavation [13–15]. However, both empirical and theoretical approaches often have limitations in representing the heterogeneity of real ground conditions and complex construction processes. While centrifuge model tests have also been conducted to investigate the mechanisms of ground response during tunneling [16–19], they are constrained by experimental scale and boundary conditions. To overcome these limitations, three-dimensional numerical analyses have been increasingly employed [20–23], enabling more realistic simulation of ground conditions and construction processes and providing more reliable settlement predictions.

In a previous study, a validated three-dimensional numerical model was used to investigate the effects of tunnel diameter (D), ground elastic modulus (E), face pressure (FP), and backfill pressure (BP) on ground settlement, and a regression-based prediction equation for maximum settlement was subsequently proposed [24]. The proposed equation was then validated through p-value analysis, variance inflation factor assessment, and comparison with field monitoring data. Although the regression-based model demonstrated satisfactory predictive capability, it may not adequately capture the complex nonlinear interactions among variables under diverse tunneling conditions. Therefore, more advanced approaches are required to improve the prediction of complex TBM-induced settlement behavior.

Recently, machine learning techniques have been widely applied in various geotechnical engineering fields because of their ability to effectively learn complex nonlinear relationships [25,26]. In particular, tree-based ensemble models such as random forest (RF) and extreme gradient boosting (XGBoost) have been recognized for their capability to capture nonlinear interactions among multiple variables [27,28]. These models have demonstrated excellent predictive performance in various geotechnical engineering applications [25,29,30]. However, prior machine learning studies on TBM-induced settlement prediction have used diverse datasets and modeling conditions depending on the availability of field observations and project-specific information. Consequently, the applicability of the developed models to conditions different from those represented in the training data remains a key concern. In addition, previous studies have primarily focused on evaluating model performance using random data partitioning approaches, while limited attention has been given to predictive capability under tunnel and ground conditions not represented in the training dataset. Therefore, further investigation of model generalization capability is required to support reliable practical applications.

In this study, machine learning models were developed using a numerical database generated from a previously validated three-dimensional numerical model for shield TBM tunneling. Tunnel diameter (D), ground elastic modulus (E), face pressure (FP), and backfill pressure (BP) were used as input variables, while maximum settlement was selected as the target output. RF and XGBoost models were developed and compared with an existing regression-based settlement prediction equation. In addition, random split and GroupKFold validation techniques were employed to evaluate both predictive performance and generalization capability. In particular, this study investigated whether the high prediction accuracy commonly reported using random data partitioning can be maintained under previously unseen tunnel–ground conditions. Through this approach, the predictive performance and generalization capability of machine learning techniques for estimating shield TBM-induced maximum settlement were systematically evaluated. The findings provide a basis for assessing the applicability of machine learning models developed from validated numerical analysis databases under previously unseen tunnel–ground conditions.

2. Materials and Methods

2.1. Numerical Model and Database

In this study, a three-dimensional numerical analysis database developed in a previous study was adopted [24]. The database was constructed using a numerical model validated against field monitoring data obtained from the Tehran Metro Line 7 project. The numerical model was validated through comparison with measured settlement data from the project [31]. Figure 1 presents the route of Tehran Metro Line 7 and the geological profile of the section used for model validation. Based on the validation results, the numerical model was established to reproduce the ground settlement behavior induced by shield TBM excavation.

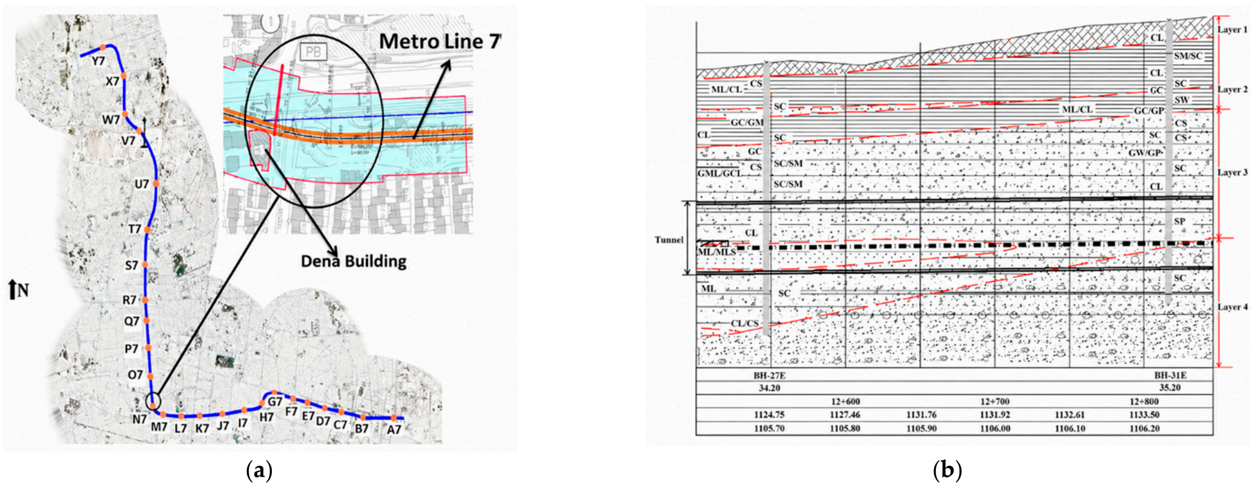


Figure 1. Tehran Metro Line 7 used for numerical model validation: (**a**) route and analyzed section; (**b**) geological profile of the analyzed section (adapted from [24,31]).

The ground was modeled using the Mohr–Coulomb constitutive model, whereas the shield, segment lining, and grout were modeled using elastic models. The numerical model adopted the same domain size, boundary conditions, and initial stress generation procedure as the previously validated numerical model. The shield, segment lining, and grout properties were identical to those used in the validation model. The sequential construction procedure included face pressure application, shield advancement, segment installation, tail void grouting, and grout hardening. The numerical database was generated using this numerical modeling procedure. Additional details are available in the previous study [24].

The numerical model was developed considering the sequential construction process of shield TBM excavation. In particular, the model simulated the *FP* application, shield passage, and tail void grouting processes to capture the ground deformation induced by shield TBM excavation [24]. A comparison between the numerical results and field monitoring data showed a settlement prediction error of 4.23%, indicating a high level of agreement. These results confirmed that the validated numerical model can reasonably reproduce the settlement behavior associated with shield TBM excavation.

To construct the numerical database, D, E, FP, and BP were selected as the primary input variables. The ranges of these variables were determined based on the parametric study conducted in the previous study and selected to represent practical shield TBM construction conditions [24].

In this study, D values of 3, 6, 9, and 12 m were considered, and E values of 50, 100, 200, and 400 MPa were adopted. The applied face pressure was defined using a normalized face pressure ratio, FP_{ref}, following the definition adopted in the previous numerical study [24]. In this definition, FP_{ref} represents the ratio of the applied face pressure to the reference face pressure, which was calculated based on the earth pressure at rest. Accordingly, an FP_{ref} value of 1.0 corresponds to the reference face pressure, whereas an FP_{ref} value greater than 1.0 represents an over-pressure condition considered to investigate the influence of higher face pressures on settlement behavior. The FP_{ref} values were assigned as 0.25, 0.5, 0.75, 1.0, and 1.25 (Equations (1) and (2)). In addition, BP was defined relative to the applied FP and was assigned four levels: FP + 20, FP + 80, FP + 140, and FP + 200 kPa. Because BP was defined as the applied FP plus one of four prescribed pressure increments, BP is not statistically independent of FP in the adopted numerical database. Nevertheless, FP and BP were retained as separate input variables because they represent different operational control parameters in shield TBM excavation. The selected input variables and their corresponding ranges are summarized in Table 1.

$$FP_{ref} = \frac{Applied\ face\ pressure}{Reference\ face\ pressure} \tag{1}$$

$$Reference\ face\ pressure = \left(1 - sin\varnothing'\right) \cdot \sigma_z' + u \tag{2}$$

where FP_{ref} is the face pressure ratio, $\varnothing'$ is the effective friction angle, σ_z' is the effective vertical stress, and u is the pore water pressure.

Table 1. Input variables and values considered in the numerical analyses.

Input Variable	Description	Value
D	Tunnel diameter (m)	3, 6, 9, 12
E	Elastic modulus (MPa)	50, 100, 200, 400
FP	Face pressure (kPa)	44.41, 88.83, 133.24, 177.66, 222.07
BP	Backfill pressure (kPa)	FP + 20, FP + 80, FP + 140, FP + 200 kPa

A total of 320 numerical analyses were performed based on the combinations of the selected variables, corresponding to a full-factorial design of D, E, FP, and BP ($4 \times 4 \times 5 \times 4$). The maximum settlement (S_{max}) was obtained from each numerical analysis. The resulting numerical database was subsequently used as the reference dataset for training and validating the machine learning models developed in this study.

2.2. Machine Learning Framework

The overall workflow of the present study is illustrated in Figure 2. Based on the developed numerical database, RF and XGBoost models were constructed to predict the maximum settlement induced by shield TBM excavation.

Figure 3 presents the general architectures of the RF and XGBoost algorithms. RF is an ensemble learning algorithm that combines multiple decision trees to generate a final prediction. By aggregating the predictions of individual trees, RF reduces overfitting and improves model generalization capability [27]. XGBoost is a gradient boosting algorithm that sequentially minimizes prediction errors from previous iterations and is widely recognized for its high predictive accuracy and computational efficiency [28,32].

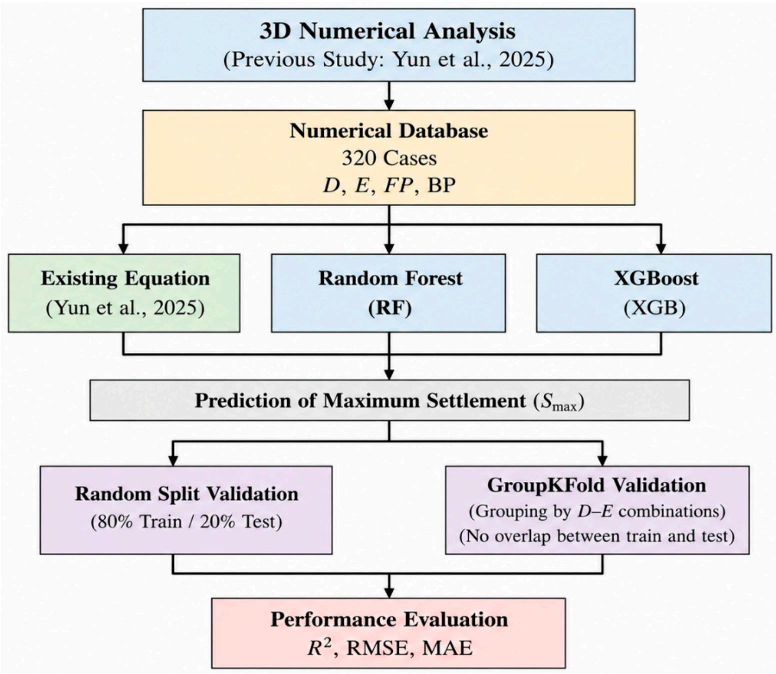
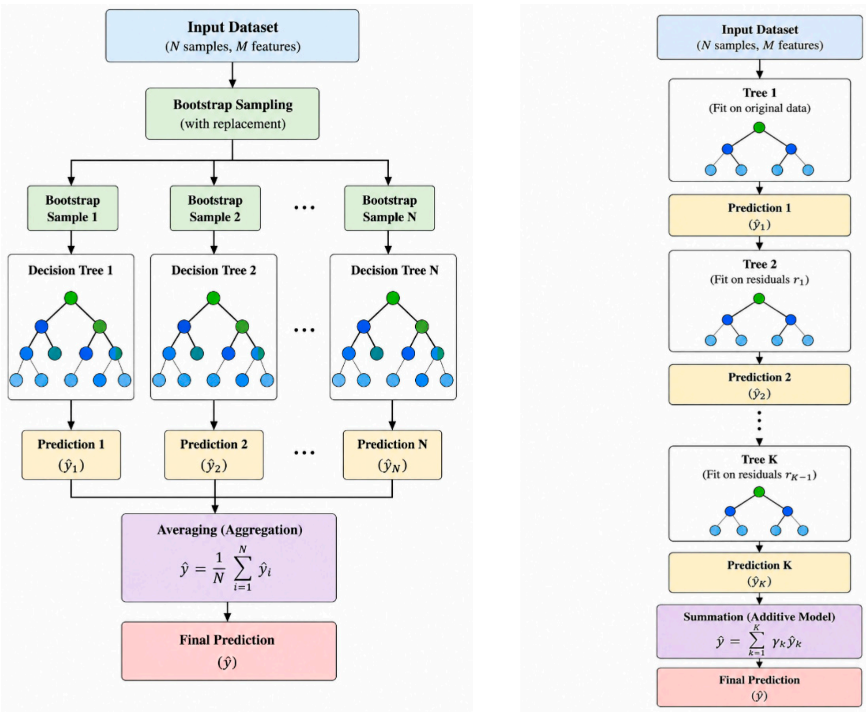


Figure 2. Overall framework of the proposed methodology for maximum settlement prediction [24].



(**a**) (**b**)

Figure 3. General architectures of (**a**) Random Forest and (**b**) Extreme Gradient Boosting (XGBoost) algorithms.

Table 2 summarizes the major hyperparameters adopted for the RF and XGBoost models. Hyperparameters not listed in Table 2 were left at their default values. The objective of this study was to evaluate the predictive performance and generalization capability of the machine learning models under different validation strategies rather than to optimize the hyperparameters of individual machine learning models. Therefore, additional hyperparameter tuning was not performed.

Table 2. Hyperparameters adopted for the Random Forest and XGBoost models.

Model	Parameter	Value
Random Forest	n_estimators	500
	max_depth	None
	min_samples_split	2
	min_samples_leaf	1
	random_state	42
XGBoost	n_estimators	500
	max_depth	5
	learning_rate	0.05
	subsample	0.9
	colsample_bytree	0.9
	random_state	42

To evaluate model performance and generalization capability, two validation strategies were employed: random split validation and GroupKFold validation. For random split validation, 80% of the dataset was randomly selected for training and the remaining 20% was used for validation. However, because the numerical database comprised a limited number of parameter combinations, random split validation alone may not adequately assess the generalization capability of the developed models. Therefore, GroupKFold validation was additionally conducted to provide a more rigorous evaluation of model generalization performance. In GroupKFold validation, samples sharing the same D and E combination were assigned to the same group. Consequently, identical D–E combinations were excluded from appearing simultaneously in the training and validation datasets [33]. Because D and E each had four levels, the 320 samples were divided into 16 D–E groups. Each D–E group contained 20 samples generated from the combinations of FP and BP. A four-fold GroupKFold validation was employed. In each fold, four D–E groups, corresponding to 80 samples, were used for validation, whereas the remaining 12 D–E groups, corresponding to 240 samples, were used for training. The GroupKFold performance reported in this study was calculated as the arithmetic mean of the metrics obtained from the four validation folds. For the R^2 comparison, shuffled random four-fold cross-validation was additionally performed using the same number of folds as GroupKFold validation to provide a fold-averaged comparison between conventional random partitioning and grouped partitioning.

Finally, model performance was quantitatively evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). R^2 was used to assess the explanatory power of the models, whereas RMSE and MAE were employed to quantify the prediction errors between the predicted and actual values. The corresponding equations are presented below (Equations (3)–(5)).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (5)$$

3. Results and Discussion

3.1. Prediction Performance Comparison

Random split validation was performed to compare the predictive performance of the machine learning models with the maximum settlement prediction equation proposed in a previous study [24]. The regression-based settlement prediction equation was developed in our previous study using the same 320-case numerical analysis database adopted in the present study. Therefore, both the regression model and the machine learning models were developed from the identical numerical database, providing a consistent basis for comparison. The previously proposed settlement prediction equation is presented in Equation (6).

$$S_{max} = -2.328 + 112.360\left(\frac{D}{E}\right) - 152.90\left(\frac{1}{E}\right) + 0.197D - 0.973\left(\frac{FP}{BP}\right) \quad (6)$$

where S_{max} is the maximum settlement (mm), D is the tunnel diameter (m), E is the ground elastic modulus (MPa), FP is the face pressure (kPa), and BP is the backfill pressure (kPa).

Equation (6) is an empirical regression equation calibrated using dimensional variables within the 320-case numerical database. Therefore, its coefficients are unit-dependent, and the equation should be applied only with S_{max} in mm, D in m, E in MPa, and FP and BP in kPa. The equation is used in this study as a database-consistent benchmark rather than as a dimensionally derived mechanistic formulation. A total of 80% of the dataset was used for training, while the remaining 20% was used for validation. Model performance was quantitatively evaluated using R^2, RMSE, and MAE. The results are summarized in Table 3.

Table 3. Prediction performance of the machine learning models and the existing settlement prediction equation under random split validation.

Model	R^2	RMSE (mm)	MAE (mm)
Random Forest	0.997	0.438	0.158
XGBoost	0.992	0.729	0.316
Existing Equation	0.893	2.671	2.263

The RF model achieved the highest predictive performance, yielding R^2, RMSE, and MAE values of 0.997, 0.438, and 0.158, respectively. XGBoost also demonstrated high predictive performance, with R^2, RMSE, and MAE values of 0.992, 0.729, and 0.316, respectively. By contrast, the existing settlement prediction equation gave R^2, RMSE, and MAE values of 0.893, 2.671, and 2.263, respectively, indicating lower predictive accuracy than the machine learning models.

Figure 4 presents the relationship between the maximum settlement obtained from numerical analyses and the predicted maximum settlement. For both RF and XGBoost, most data points were distributed close to the 1:1 line, indicating good agreement between the numerical results and model predictions. The RF model exhibited data points clustered most closely around the 1:1 line and achieved the highest R^2 and lowest prediction errors. Although XGBoost also demonstrated a similar trend, larger deviations were observed for several data points. By contrast, the existing settlement prediction equation exhibited greater dispersion from the 1:1 line than the machine learning models. In particular, a considerable number of data points corresponding to larger settlement values were located below the 1:1 line, indicating a tendency to underestimate maximum settlement.

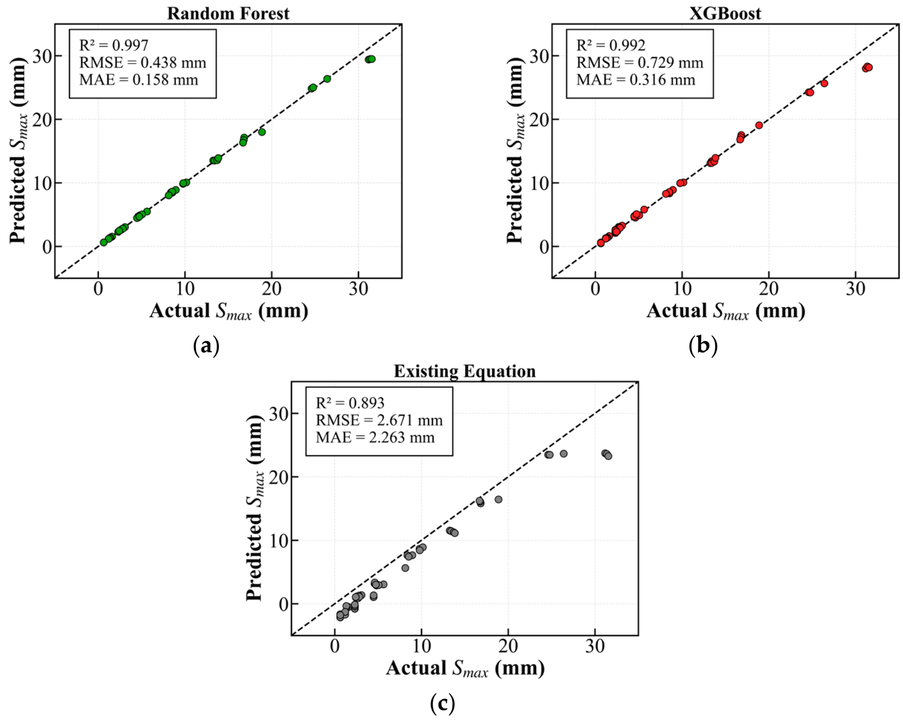


Figure 4. Relationship between actual and predicted maximum settlement (S_{max}): (**a**) Random Forest, (**b**) XGBoost, and (**c**) the existing settlement prediction equation.

Figure 5 shows the residual distributions of the evaluated models. The residuals of both the RF and XGBoost models were concentrated around zero, with the RF model exhibiting the narrowest residual range. This observation is consistent with the lowest RMSE and MAE values reported in Table 3. By contrast, the existing settlement prediction equation exhibited a wider residual distribution than the machine learning models, with residuals tending to be distributed in the positive direction. This result is also consistent with the underestimation tendency observed in Figure 4.

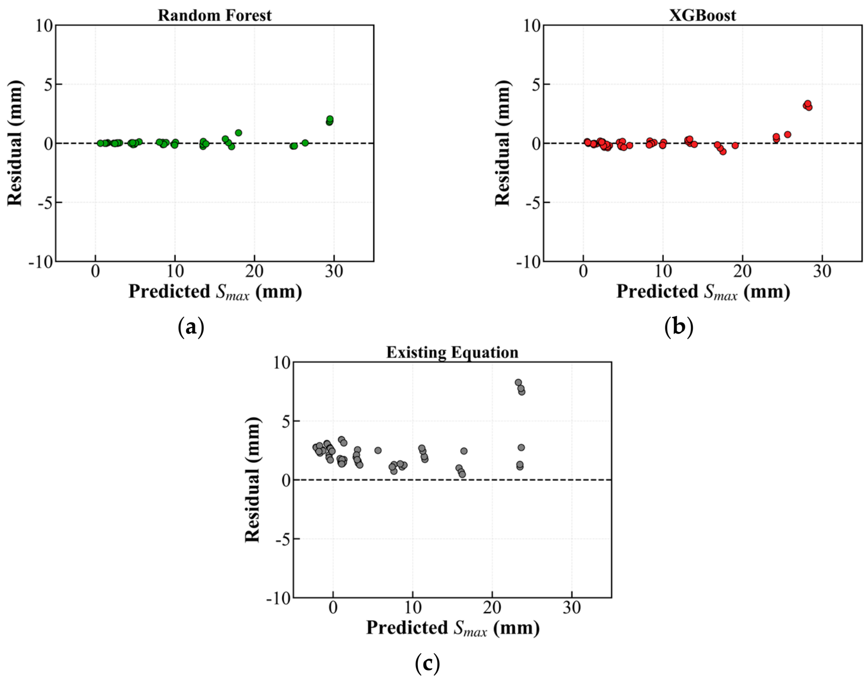


Figure 5. Residual distributions of prediction errors: (**a**) Random Forest, (**b**) XGBoost, and (**c**) the existing settlement prediction equation.

The superior predictive performance of the machine learning models may be attributed to their ability to capture nonlinear relationships among the input variables. The existing settlement prediction equation was developed using regression analysis and represents the relationships among variables in a simplified form. By contrast, the RF and XGBoost models can account for more complex interactions among *D*, *E*, *FP*, and *BP*. Consequently, the machine learning models demonstrated better predictive performance than the existing equation within the range of conditions represented in the numerical database developed in this study.

Overall, the results indicate that the machine learning models, particularly RF, provide excellent predictive performance within the range of conditions represented in the numerical database.

3.2. Generalization and Interpretation of Machine Learning Models

Random split validation randomly partitions the entire dataset into training and validation subsets, and therefore, with high probability, the validation data contain conditions similar to those included in the training data. Consequently, random split validation alone may not provide a sufficient assessment of model performance under previously unseen conditions. To address this limitation, GroupKFold validation was additionally performed. To evaluate model performance under different tunnel and ground conditions, groups were defined based on combinations of *D* and *E*, which represent the primary geometric and geotechnical parameters considered in the numerical database.

Figure 6 compares the fold-averaged R^2 values obtained using shuffled random 4-fold cross-validation and GroupKFold validation. Under random 4-fold cross-validation, RF and XGBoost achieved mean R^2 values of 0.999 and 0.995, respectively. By contrast, when GroupKFold validation was applied, the corresponding mean R^2 values decreased substantially to -0.935 and -0.943, respectively. These results indicate that model performance can vary considerably depending on the data partitioning strategy employed. To further clarify the GroupKFold validation procedure and the variability in model performance across folds, the fold-wise performance metrics are summarized in Table 4. Each validation fold consisted of four *D*–*E* groups, corresponding to 80 validation samples, and the S_{max} range differed considerably among folds. In the resulting fold assignment, each validation fold corresponded to one *E* level across all *D* values, thereby providing a stringent assessment of model performance under unseen ground stiffness conditions.

As shown in Table 4, most validation folds yielded negative R^2 values for both RF and XGBoost, indicating limited predictive reliability for *D*–*E* combinations not included in the training data.

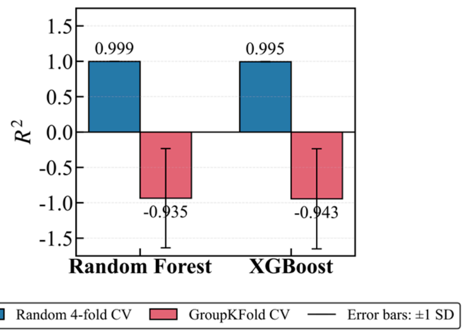


Figure 6. Comparison of fold-averaged R^2 values obtained using random 4-fold cross-validation and GroupKFold validation.

Table 4. Fold-wise GroupKFold validation results.

Fold	Validation *D*–*E* Groups	Validation Samples	S_{max} Range (mm)	Random Forest R^2/RMSE/MAE	XGBoost R^2/RMSE/MAE
1	*D* = 3, 6, 9, 12 m; *E* = 50 MPa	80	4.472–31.542	0.116/7.839/6.904	0.118/7.829/6.898
2	*D* = 3, 6, 9, 12 m; *E* = 100 MPa	80	2.302–17.060	−1.814/7.810/6.886	−1.828/7.829/6.898
3	*D* = 3, 6, 9, 12 m; *E* = 200 MPa	80	1.191–9.549	−1.218/3.940/3.394	−1.229/3.950/3.400
4	*D* = 3, 6, 9, 12 m; *E* = 400 MPa	80	0.623–5.533	−0.823/2.102/1.801	−0.835/2.109/1.805

Note: Values in the Random Forest and XGBoost columns are presented as R^2/RMSE/MAE. RMSE and MAE are expressed in mm.

The discrepancy between the two validation approaches can be attributed to differences in the similarity between the training and validation datasets. In random split validation, similar parameter combinations are likely to appear in both datasets, resulting in high predictive performance. By contrast, GroupKFold validation was designed such that the *D*–*E* combinations used for validation were not included in the training dataset. Consequently, the model was required to make predictions under tunnel and ground conditions that were not encountered during training. Therefore, random split validation primarily evaluates interpolation performance within the observed data range, whereas GroupKFold validation provides a more rigorous assessment of model performance under unseen *D*–*E* grouped conditions.

Both RF and XGBoost exhibited comparable reductions in performance under GroupKFold validation. This suggests that, under the present feature design and sparse discrete numerical database, the observed limitation was common to both RF and XGBoost rather than being specific to a single model. Although these models demonstrated excellent predictive capability within the represented tunnel–ground conditions, their performance deteriorated considerably when predicting previously unseen *D*–*E* combinations. For the two tree-based models evaluated in this study, this behavior is likely associated with their limited extrapolation capability under the present feature design and sparse discrete numerical database, whereby predictions are generated based on patterns learned from the training data distribution [34].

As the RF model exhibited the highest predictive performance, feature importance and partial dependence analyses were conducted using this model. Figure 7 presents the feature importance analysis obtained from the RF model. According to the impurity-based feature importance of the RF model, *D* and *E* exhibited substantially higher relative importance than *FP* and *BP* within the present numerical database. The corresponding relative importance values were approximately 49.5%, 49.5%, 0.9%, and 0.1%, respectively. The reported feature importance represents the relative contribution of the input variables within the present numerical database and should not be interpreted as a general ranking applicable to all TBM settlement problems.

The feature importance results also provide useful information for the interpretation of the GroupKFold validation results. Because GroupKFold validation was defined based on *D*–*E* combinations, the validation process required the models to predict responses for *D*–*E* combinations that were not included in the training data. Consequently, a more challenging prediction problem was formed than that under random split validation, which may explain the reduction in performance observed under GroupKFold validation.

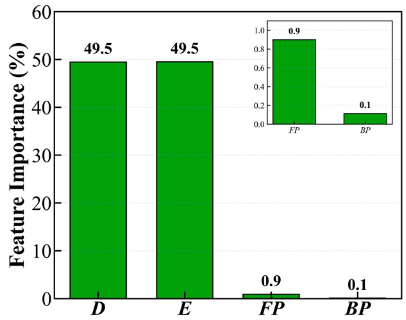


Figure 7. Relative importance of input variables in the Random Forest model for maximum settlement prediction.

Figure 8 presents partial dependence plots (PDPs) with individual conditional expectation (ICE) curves obtained from the RF model. The red lines represent the average partial dependence, while the gray lines show individual conditional responses. Blue rug marks along the *x*-axis indicate the distribution of the training samples. These PDPs visually illustrate the direction of the relationship between each input variable and maximum settlement. As shown in Figure 8a,b, the predicted maximum settlement tends to increase with increasing *D* and decrease with increasing *E*. These trends are generally consistent with previously reported settlement behavior.

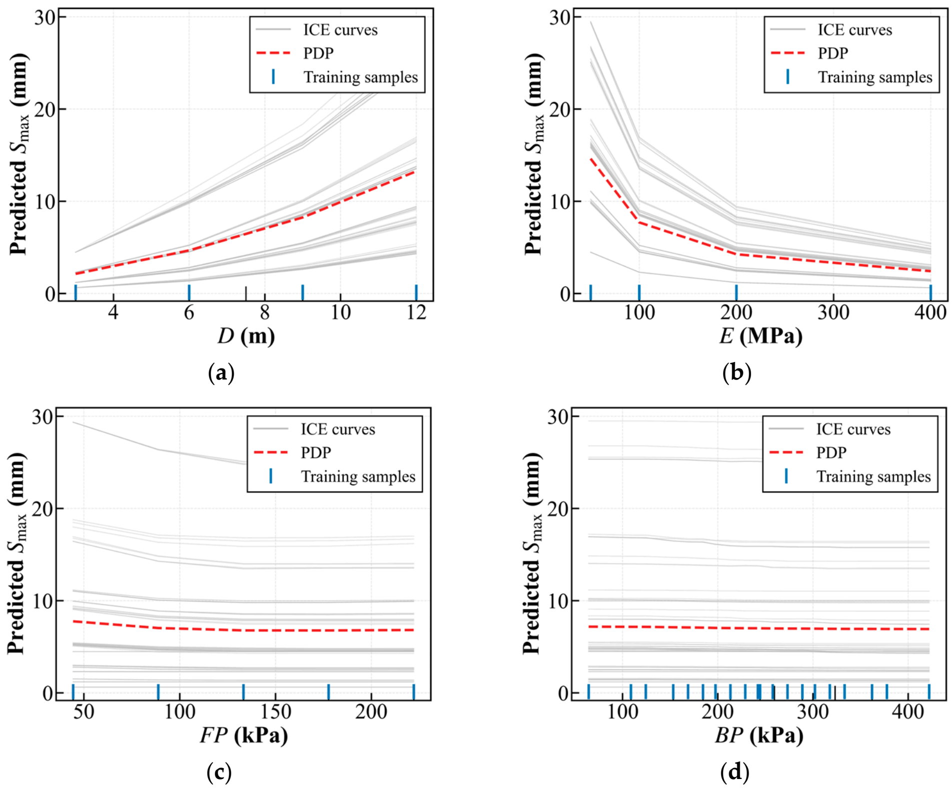


Figure 8. Partial dependence plots (PDPs) with individual conditional expectation (ICE) curves for (**a**) tunnel diameter (*D*), (**b**) ground elastic modulus (*E*), (**c**) face pressure (*FP*), and (**d**) backfill pressure (*BP*) obtained from the Random Forest model. Blue rug marks along the *x*-axis indicate the distribution of the training samples.

However, these relatively small variations should be interpreted in the context of the entire numerical database. Under specific *D* and ground conditions, *FP* and *BP* may still act as important construction control parameters. Previous studies have reported that *FP* and *BP* can influence settlement behavior and are important operational parameters for settlement control [24,31,35,36]. Therefore, the relatively low feature importance values observed in this study represent relative contributions calculated over the entire numerical database. These results should not be interpreted as indicating that *FP* and *BP* have limited engineering significance. Rather, they indicate that the influence of *FP* and *BP* is relatively small compared with *D* and *E* within the numerical database considered in this study, while under specific tunnel–ground conditions these operational parameters may still play an important role in settlement control.

Consequently, the high predictive performance obtained from random split validation may not fully represent model performance under previously unseen tunnel and ground conditions. For the database considered in this study, validation procedures that explicitly prevent identical *D*–*E* combinations from appearing in both the training and validation datasets may provide a more rigorous assessment of model generalization capability. Furthermore, when machine learning models developed from numerical analysis databases are applied to practical engineering problems, careful consideration should be given to whether the target tunnel–ground conditions are represented in the training database. Additional validation using field data or broader numerical databases is recommended before extending the models to conditions beyond those considered in this study.

4. Conclusions

In this study, machine learning models were developed to predict the maximum settlement induced by shield TBM excavation using a three-dimensional numerical analysis database. *D*, *E*, *FP*, and *BP* were used as input variables, and RF and XGBoost models were constructed and compared with an existing regression-based maximum settlement prediction equation. The main conclusions are summarized as follows:

(1) Under random split validation, RF and XGBoost demonstrated higher predictive performance than the existing maximum settlement prediction equation. In particular, RF achieved the highest predictive performance, with an R^2, RMSE, and MAE of 0.997, 0.438 mm, and 0.158 mm, respectively. These results indicate that machine learning techniques can effectively predict maximum settlement within the range of tunnel–ground conditions represented in the adopted numerical analysis database.

(2) High predictive performance obtained from random split validation does not necessarily indicate reliable performance under unseen *D*–*E* grouped conditions. The results of GroupKFold validation showed substantial reductions in the performance of both the RF and XGBoost models, indicating that the data partitioning strategy can significantly influence the evaluation and interpretation of machine learning model performance.

(3) Within the present numerical database, feature importance analysis indicated that *D* and *E* exhibited substantially higher relative importance than *FP* and *BP*. In addition, the PDP and ICE analyses showed that maximum settlement tended to increase with increasing *D* and decrease with increasing *E*. These results suggest that, within the adopted numerical analysis database, tunnel size and ground stiffness were the dominant factors affecting the model predictions.

(4) This study developed machine learning-based models for predicting the maximum settlement induced by shield TBM excavation by integrating a validated three-dimensional numerical analysis database with machine learning techniques. By comparing the developed models with an existing regression-based prediction equation, the proposed models demonstrated improved predictive performance for the tunnel–ground conditions represented in the adopted numerical analysis database.

(5) The comparison between random data partitioning and GroupKFold validation demonstrated that excellent prediction performance within the numerical database did not necessarily translate into good generalization performance for previously unseen *D*–*E*

combinations under the current data structure. Therefore, the principal finding of this study is that high prediction accuracy obtained using conventional random validation does not necessarily indicate good generalization capability. Accordingly, the developed models should be interpreted as database-specific prediction tools applicable mainly within the constructed numerical analysis database, while site-specific field monitoring data and engineering judgment remain necessary for practical applications.

For future research, an expanded database incorporating field monitoring data from actual TBM projects should be developed to further improve the machine learning models and enhance their applicability under diverse ground and construction conditions. In addition, quantitative analyses of the effects of *FP* and *BP* on settlement behavior under specific tunnel and ground conditions are expected to provide a clearer understanding of the relationship between TBM operational parameters and settlement behavior. These efforts may further improve the applicability and practical usefulness of settlement prediction models for shield TBM tunneling projects.

Author Contributions: Conceptualization, J.-S.Y. and C.-Y.K.; Methodology, J.-S.Y.; Software, J.-S.Y. and J.-K.L.; Validation, J.-S.Y., W.-K.Y., G.-J.L. and H.-E.K.; Formal analysis, J.-S.Y. and Y.-S.S.; Investigation, J.-S.Y.; Data curation, W.-K.Y. and G.-J.L.; Writing—original draft preparation, J.-S.Y.; Writing—review and editing, W.-K.Y., G.-J.L., J.-K.L., Y.-S.S., H.-E.K. and C.-Y.K.; Supervision, C.-Y.K. and H.-E.K.; Project administration, C.-Y.K.; Funding acquisition, C.-Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: Research for this paper was conducted under the KICT Strategic Research Program(SRP)(project no. 20260227, Development of Construction Technologies for Safe and Convenient Underground Roads (K-SCOUR)) funded by the Ministry of Science and ICT.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The numerical database supporting the findings of this study is available from the corresponding author upon reasonable request for research purposes.

Conflicts of Interest: Author Han-Eol Kim was employed by Korea Expressway Corporation Research Institute. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BP	Backfill pressure
FP	Face pressure
MAE	Mean absolute error
PDP	Partial dependence plot
ICE	Individual conditional expectation
RMSE	Root mean square error
RF	Random forest
TBM	Tunnel boring machine
XGBoost	Extreme gradient boosting

References

1. Ocak, I. Environmental effects of tunnel excavation in soft and shallow ground with EPBM: The case of Istanbul. *Environ. Earth Sci.* **2009**, *59*, 347–352. [CrossRef]
2. Lai, J.; Zhou, H.; Wang, K.; Qiu, J.; Wang, L.; Wang, J.; Feng, Z. Shield-driven induced ground surface and Ming Dynasty City Wall settlement of Xi'an Metro. *Tunn. Undergr. Space Technol.* **2020**, *97*, 103220. [CrossRef]

3. Huang, M.; Lu, Y.; Zhen, J.; Lan, X.; Xu, C.; Yu, W. Analysis of face stability at the launch stage of shield or TBM tunnelling using a concrete box in complex urban environments. *Tunn. Undergr. Space Technol.* **2023**, *135*, 105067. [CrossRef]
4. Weng, X.; Sun, Y.; Yan, B.; Niu, H.; Lin, R.; Zhou, S. Centrifuge testing and numerical modeling of tunnel face stability considering longitudinal slope angle and steady-state seepage in soft clay. *Tunn. Undergr. Space Technol.* **2020**, *101*, 103406. [CrossRef]
5. Liao, S.-M.; Liu, J.-H.; Wang, R.-L.; Li, Z.-M. Shield tunneling and environment protection in Shanghai soft ground. *Tunn. Undergr. Space Technol.* **2009**, *24*, 454–465. [CrossRef]
6. Ou, C.-Y.; Teng, F.-C.; Wang, I.-W. Analysis and design of partial ground improvement in deep excavations. *Comput. Geotech.* **2008**, *35*, 576–584. [CrossRef]
7. Yun, J.S.; Kim, H.E.; Yoo, H.K. Three-dimensional modeling and analysis of ground settlement due to twin tunneling using GIS. *Sustainability* **2024**, *16*, 5891. [CrossRef]
8. Peck, R.B. Deep Excavations and Tunnelling in Soft Ground. In Proceedings of the 7th International Conference on Soil Mechanics and Foundation Engineering (ICSMFE), Mexico City, Mexico, 25–29 August 1969; pp. 225–290.
9. Attewell, P.B.; Woodman, J.P. Predicting the dynamics of ground settlement and its derivitives caused by tunnelling in soil. *Ground Eng.* **1982**, *15*, 13–22+36.
10. O'Reilly, M.P.; New, B.M. Settlements Above Tunnels in the United Kingdom—Their Magnitude and Prediction. In *Proceedings of Tunnelling '82 Symposium*; Institution of Mining and Metallurgy: London, UK, 1982; pp. 173–181.
11. Mair, R.J.; Taylor, R.N.; Bracegirdle, A. Subsurface settlement profiles above tunnels in clays. *Géotechnique* **1993**, *43*, 315–320. [CrossRef]
12. Fang, Y.-S.; Wu, C.-T.; Chen, S.-F.; Liu, C. An estimation of subsurface settlement due to shield tunneling. *Tunn. Undergr. Space Technol.* **2014**, *44*, 121–129. [CrossRef]
13. Sagaseta, C. Analysis of undrained soil deformation due to ground loss. *Géotechnique* **1987**, *37*, 301–320. [CrossRef]
14. Verruijt, A.; Booker, J.R. Surface settlements due to deformation of a tunnel in an elastic half plane. *Géotechnique* **1996**, *46*, 753–756. [CrossRef]
15. Loganathan, N.; Poulos, H.G. Analytical prediction for tunneling-induced ground movements in clays. *J. Geotech. Geoenviron. Eng.* **1998**, *124*, 846–856. [CrossRef]
16. Ng, C.W.W.; Boonyarak, T.; Mašín, D. Three-dimensional centrifuge and numerical modeling of the interaction between perpendicularly crossing tunnels. *Can. Geotech. J.* **2013**, *50*, 935–946. [CrossRef]
17. Li, P.; Du, S.J.; Ma, X.F.; Yin, Z.Y.; Shen, S.L. Centrifuge investigation into the effect of new shield tunnelling on an existing underlying large-diameter tunnel. *Tunn. Undergr. Space Technol.* **2014**, *42*, 59–66. [CrossRef]
18. Shi, J.; Wang, Y.; Ng, C.W.W. Three-dimensional centrifuge modeling of ground and pipeline response to tunnel excavation. *J. Geotech. Geoenviron. Eng.* **2016**, *142*, 04016054. [CrossRef]
19. Lai, H.; Zhang, J.; Zhang, L.; Chen, R.; Yang, W. A new method based on centrifuge model test for evaluating ground settlement induced by tunneling. *KSCE J. Civ. Eng.* **2019**, *23*, 2426–2436. [CrossRef]
20. Comodromos, E.M.; Papadopoulou, M.C.; Konstantinidis, G.K. Numerical assessment of subsidence and adjacent building movements induced by TBM-EPB tunneling. *J. Geotech. Geoenviron. Eng.* **2014**, *140*, 04014061. [CrossRef]
21. Do, T.N.; Wu, J.H.; Lin, H.M. Investigation of sloped surface subsidence during inclined seam extraction in a jointed rock mass using discontinuous deformation analysis. *Int. J. Geomech.* **2017**, *17*, 04017021. [CrossRef]
22. Islam, M.S.; Iskander, M. Three-dimensional numerical investigation of ground settlement caused by side-by-side twin tunnels. *Tunn. Undergr. Space Technol.* **2025**, *159*, 106509. [CrossRef]
23. Wang, L.; Zheng, G. Numerical analysis of underground displacements stirred by compaction grouting during tunnel construction. *Tunn. Undergr. Space Technol.* **2024**, *152*, 105931. [CrossRef]
24. Yun, J.S.; Yoo, H.K.; Hwang, S.P.; Kim, W.S.; Kim, H.E. Prediction of settlement due to shield TBM tunneling based on three-dimensional numerical analysis. *Buildings* **2025**, *15*, 2235. [CrossRef]
25. Baghbani, A.; Choudhury, T.; Costa, S.; Reiner, J. Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Sci. Rev.* **2022**, *228*, 103991. [CrossRef]
26. 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]
27. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
28. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
29. Zhang, W.; Wang, Y.; Phoon, K.K. Comprehensive review of machine learning in geotechnical reliability analysis: Algorithms, applications and further challenges. *Appl. Soft Comput.* **2023**, *137*, 110066. [CrossRef]
30. Shin, Y.J.; Kwon, K.; Bae, A.; Choi, H.; Kim, D. Machine learning-based prediction model for disc cutter life in TBM excavation through hard rock formations. *Tunn. Undergr. Space Technol.* **2024**, *150*, 105826. [CrossRef]

31. Chakeri, H.; Ozcelik, Y.; Unver, B.; Copur, H. Effects of important factors on surface settlement prediction for metro tunnel excavated by EPB. *Tunn. Undergr. Space Technol.* **2013**, *36*, 14–23. [CrossRef]
32. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
33. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
34. Wenger, S.J.; Olden, J.D. Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods Ecol. Evol.* **2012**, *3*, 260–267. [CrossRef]
35. An, J.B.; Kang, S.J.; Kim, J.; Cho, G.C. Numerical evaluation of surface settlement induced by ground loss from the face and annular gap of EPB shield tunneling. *Geomech. Eng.* **2022**, *29*, 291–300. [CrossRef]
36. Kim, K.; Oh, J.; Lee, H.; Kim, D.; Choi, H. Critical face pressure and backfill pressure in shield TBM tunneling on soft ground. *Geomech. Eng.* **2018**, *15*, 823–831. [CrossRef]

https://doi.org/10.3390/app16146951