KSCE Journal of Civil Engineering
DOI 10.1007/s12205-024-0066-z

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Tunnel Engineering

A Hybrid Numerical-ML Model for Predicting Geological Risks in Tunneling with Electrical Methods

Minkyu Kang[a], **Khanh Pham**[b], **Kibeom Kwon**[c], **Seunghun Yang**[d], **and Hangseok Choi**[e]

[a]*Member, Center for Defense Resource Management, Korea Institute for Defense Analyses, Seoul 02455, Korea*
[b]*Dept. of Civil Engineering, International University, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam*
[c]*Member, Dept. of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Korea*
[d]*Dept. of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Korea*
[e]*Member, School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Korea*

ARTICLE HISTORY

Received 11 January 2024
Revised 11 June 2024
Accepted 17 July 2024
Published Online 12 September 2024

KEYWORDS

Tunnel
Electrical resistivity
Geological risks
Numerical analysis
Machine learning algorithm

ABSTRACT

In order to ensure construction efficiency and stability during tunnel excavation, it is essential to predict geological risks ahead of tunnel faces. In this study, a geological risk prediction model was developed based on a machine learning (ML) algorithm. The database used to implement the ML model was synthetically acquired from a series of finite-element (FE) numerical analyses, which could simulate electrical resistivity surveys during tunnel excavation. The developed FE model helped obtain resistivity data representing various risky ground conditions (such as typical fault zones, water intrusion, mixed ground, geological transitions, and cavities) encountered during tunnel advancement. Four ML algorithms (support vector machine, k-nearest neighbors, random forest, and extreme gradient boosting) were used to develop the prediction model. The evaluation results showed that the proposed ML prediction models produced highly accurate results. Among the ML algorithms, the prediction model based on the random forest (RF) algorithm exhibited superior performance, with an accuracy of 97.33%. Given the feasibility and efficiency of recognizing hazardous ground conditions, the proposed model is expected to serve as a reliable approach for risk management. Finally, an engineering flowchart was proposed to assist in the application of the study results to actual tunneling sites.

1. Introduction

With the increasing demand for underground infrastructure in urban areas, tunnel construction has become a popular method for developing underground structures such as utilities and traffic tunnels (Sharafat et al., 2021). Moreover, the rising density of urban populations has led to high social costs and traffic congestion (Broere, 2016), further emphasizing the need for tunnel excavation in recent decades.

Despite the benefits of tunnel construction, unexpected geological risks often arise during tunneling, significantly reducing construction safety and efficiency (Rafie and Namin, 2015; Bai et al., 2021). Fault fracture zones, for example, are among the most dangerous geological risks in tunnel construction, and can cause ground settlement, face collapse, and water intrusion (Zhao et al., 2007; Farrokh and Rostami, 2009; Zhao et al., 2014; Xu et al., 2022). Mixed ground conditions are also highly hazardous, and can cause face collapse, and ground surface settlement (Toth et al., 2013). In summary, geological risks encountered in actual tunneling sites can make tunnel construction challenging. Therefore, accurately predicting hazardous ground conditions ahead of a tunnel face is critical for avoiding potential problems in advance (Yazdani-Chamzini, 2014).

However, geotechnical investigations in the design stage of tunnel construction focus on identifying overall geological structures and cannot precisely predict ground conditions along the tunnel route. To overcome these limitations, a series of nondestructive surveys have been developed to predict geological risks ahead of tunnel faces, including tunnel seismic prospecting (TSP), ground-penetrating radar (GPR), transient electromagnetic method (TEM), and bore-tunneling electrical ahead monitoring (BEAM) (Dickmann and Sander, 1996; Grodner, 2001; McDowell et al., 2002; Kaus

CORRESPONDENCE Hangseok Choi hchoi2@korea.ac.kr, School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Korea

and Boening, 2008). In particular, electrical resistivity exploration has gained significant attention for predicting geological risks due to its low cost and rapid data analysis (Carrière et al., 2013; Park et al., 2016, 2017, 2018; Liu et al., 2019; Mifkovic et al., 2021; Kang et al., 2023).

With the remarkable development of computer technology, numerical analyses have been gradually adopted for simulating electrical resistivity surveys to predict risky ground conditions ahead of a tunnel face. Schaeffer and Mooney (2016) proposed a numerical analysis model for electrical resistivity methods applied to a tunneling environment, and Mifkovic et al. (2021) numerically simulated imaging technique ahead of a tunnel face with DC resistivity. Lee et al. (2019) conducted a 2D numerical simulation of the electrical resistivity tomography method to determine the optimal electrode array for predicting anomalies ahead of a tunnel face. However, these studies primarily focused on assessing the influence of equipment (e.g., tunnel boring machine) on resistivity measurements, rather than the impacts of geological hazards ahead of a tunnel. Kang et al. (2022) developed a 3D finite element (FE) numerical model to simulate an electrical resistivity survey during tunnel excavation toward mixed ground conditions. However, this model could only predict the existence of mixed ground but could not identify other risky conditions, such as fault fracture zones, cavities, and water ingress.

In recent decades, machine learning (ML) algorithms have emerged as robust tools for addressing engineering challenges in various fields (Arumugam et al., 2023; Kafy et al., 2023; Méndez et al., 2023; Pallathadka et al., 2023). Furthermore, enhancements in communication networks and data transmission capabilities have led to the collection of more extensive and higher-quality data from TBM tunneling projects. As a result, ML approaches have become increasingly prevalent in TBM tunneling, leveraging the abundant data gathered during these projects, including site-specific features (Eftekhari et al., 2018; Lee et al., 2022; Mahmoodzadeh et al., 2022a; Li et al., 2023). Specifically, ML algorithms have demonstrated significant potential in accurately predicting ground conditions ahead of a TBM tunnel face (Jung et al., 2019; Shi et al., 2019; Liu et al., 2020; Sebbeh-Newton et al., 2021). Nevertheless, there are few ML approaches for predicting various geological risks based on electrical resistivity surveys due to the lack of resistivity measurement data during TBM tunneling.

Meanwhile, numerical analysis serves as a valuable tool for generating databases to implement ML algorithms. For instance, Hasanpour et al. (2020) and Hou et al. (2023) employed numerical analysis of TBM tunneling under adverse geological conditions to supplement data samples, aiming to predict shield jamming probability based on an ML algorithm. Additionally, Mahmoodzadeh et al. (2022b) conducted numerical simulations of hard rock failure induced by structural planes around deep tunnels to generate datasets for ML algorithms.

Building upon these findings, this study proposes a new research paradigm for electrical resistivity surveys. In this paradigm, numerical analysis allows to overcome the lack of resistivity measurement data corresponding to each geological risk, which is essential for implementing the ML algorithm. In this study, FE numerical models were developed to predict various geological risks ahead of a tunnel face based on an electrical resistivity survey. These simulated geological risks include faults, water intrusion, mixed ground, geological transition, and cavities, which are commonly encountered in actual tunneling sites and may compromise construction safety (Shang et al., 2004; Farrokh and Rostami, 2009; Hyun et al., 2015; Gong et al., 2016; Bayati and Hamidi, 2017; Lee and Moon, 2020; Chung et al., 2021). Based on ML algorithms, a geological risk prediction model was developed using the resistivity measurement data obtained from the numerical FE models. Furthermore, an engineering flowchart for predicting geological risks using the ML algorithms based on the electrical resistivity database obtained from the numerical analysis was provided to validate the practical applicability of the developed prediction model.

2. Methodology

2.1 Electrical Resistivity Survey

The electrical resistivity of geological formations is a physical property that reflects the ability of the ground to resist the flow of injected electrical current. It is a measure of the ground's ability to conduct or transmit electrical currents, which is influenced by various factors. Therefore, the electrical resistivity of rocks and soils can vary widely, depending on their physical properties and composition. Typically, rocks and soils with high porosity or high clay mineral content exhibit lower electrical resistivity, while those with low porosity or low clay mineral content tend to have higher resistivity. Fig. 1 provides a summary of the general ranges of the electrical resistivity for various types of rocks and soils under saturated conditions. However, the actual electrical resistivity of a particular rock or soil can vary significantly depending on factors such as mineral composition, porosity, degree of saturation, temperature, and pressure. Note that Fig. 1 shows only the general range of electrical resistivity.

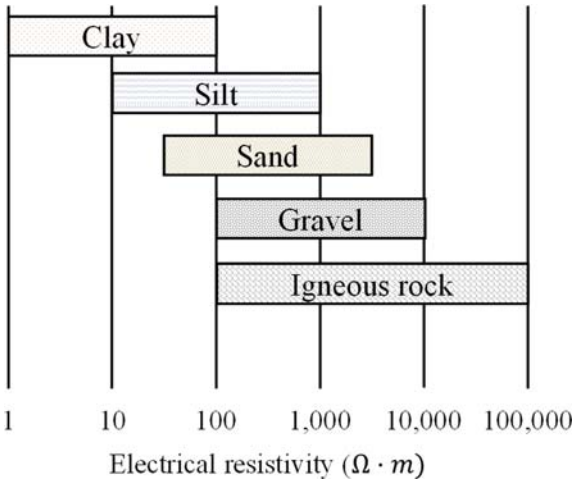

Fig. 1. Typical Range of the Electrical Resistivity of Saturated Rocks and Soils

The electrical conductivity (i.e., the inverse of the electrical resistivity) of a saturated ground formation (σ_{mix}) can be expressed as Eq. (1) (Santamarina et al., 2001).

$$\sigma_{mix} = (1-n)\sigma_p + n\sigma_{el} + (1-n)\frac{r_p}{g}\lambda_{ddl}S_a, \quad (1)$$

where σ_p is the electrical conductivity of soil particles, n is the porosity, σ_{el} is the electrical conductivity of pore water, r_p is the unit weight of soil particles, g is the gravitational acceleration, λ_{ddl} is the electrical conductivity of soil surfaces, and S_a is the specific area of soil particles. The electrical conductivity of soil particles is much lower than that of pore water ($\sigma_{el} \gg \sigma_p$). Furthermore, the double-layer effect is not significant in granular soils ($\lambda_{ddl} \approx 0$). Consequently, Eq. (1) can be simply represented as Eq. (2).

$$\rho_{mix} = \frac{\rho_{el}}{n}, \quad (2)$$

where ρ_{mix} is the electrical resistivity of a saturated ground formation, and ρ_{el} is the electrical resistivity of pore water. Referring to Eq. (2), the electrical resistivity of saturated rock, sand, and silt is mainly influenced by the electrical resistivity of pore water and its porosity.

To conduct electrical resistivity surveys, a four-electrode system was adopted to measure the electrical resistivity of ground formations. The electrode system consisted of two current electrodes and two potential electrodes. One current electrode (A) supplied the electrical current (I) to the ground, and the other current electrode (B) received the electrical current to form a potential electrical distribution in the ground formations, as illustrated in Fig. 2. The voltage (V), which is the electrical potential difference, was then measured using the two potential electrodes (M, N). The electrical resistivity (ρ) is calculated using Eq. (3) (Reynolds, 2011).

$$\rho = 2\pi\left(\frac{1}{AM} - \frac{1}{MB} - \frac{1}{AN} + \frac{1}{NB}\right)^{-1}\frac{V}{I}, \quad (3)$$

where AM is the distance between the electrodes A and B, and MB, AN, and NM are the distances between each electrode, as shown in Fig. 3(a).

Different electrode array methods have distinct characteristics, such as investigation depth, signal amplitude, and resolution, that must be considered when selecting an optimal method to achieve high performance. For predicting geological risks ahead of a tunnel face, the Wenner electrode array is a suitable choice due to its high signal amplitude and vertical resolution, as depicted in Fig. 3(b). The Wenner electrode array calculates the electrical resistivity (ρ_a) using Eq. (4) (Telford et al., 1990).

$$\rho_a = 2\pi a\left(\frac{\Delta V}{I}\right), \quad (4)$$

where a is the distance between the electrodes, ΔV is the measured voltage, and I is the supplied electric current.

2.2 Machine Learning Model

This study applied well-known machine learning (ML) algorithms, including the k-nearest neighbor (KNN), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGB), to predict geological risks ahead of a tunnel face. The KNN algorithm operates on the basic idea that similar data points are likely to have similar outcomes or labels. The algorithm identifies the k-nearest neighbors to a given testing data point and assigns the label or outcome of the majority of these neighbors to the testing data point (as shown in Fig. 4(a)). To implement the KNN algorithm, the first step is to choose the number of neighbors (k) to consider. The algorithm then computes the distance between the testing data point and all the other data points in the dataset using a distance metric, such as the Euclidean or Manhattan distance. Next, the k-nearest neighbors to the testing data point are identified by finding the k data points with the smallest distances to the testing data point. Finally, the algorithm assigns the label or outcome of the majority of

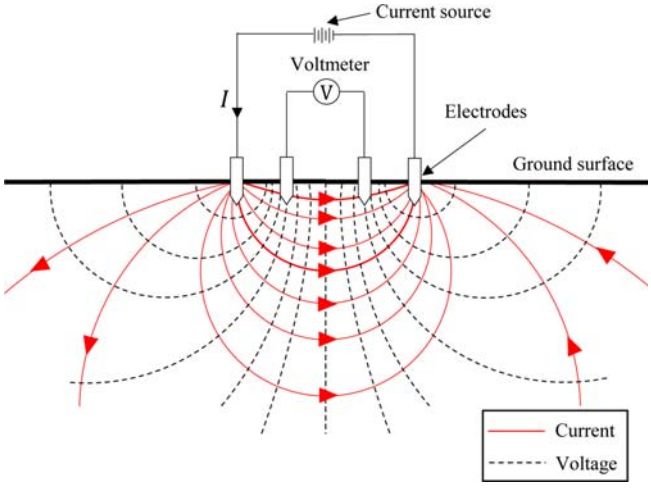

Fig. 2. Schematic of an Electrical Resistivity Survey

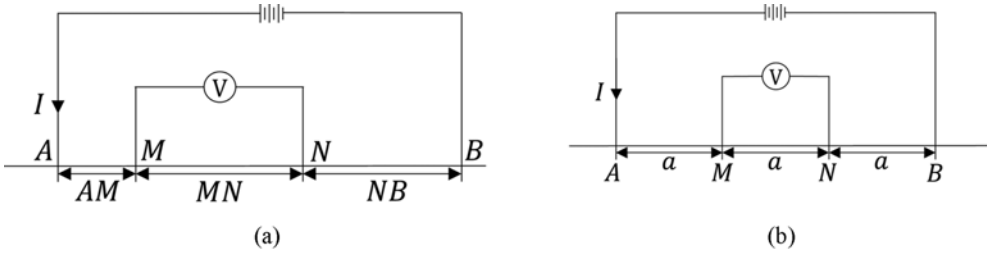

Fig. 3. Schematics of Electrode Arrays: (a) Typical Electrode Array, (b) Wenner Electrode Array

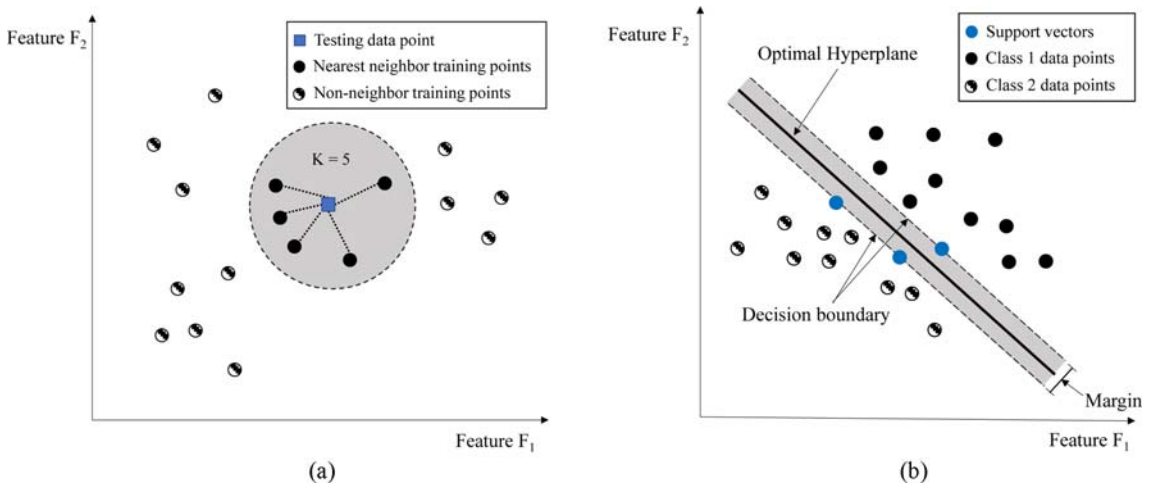

Fig. 4. Graphical Descriptions of Machine Learning Algorithms: (a) KNN Algorithm, (b) SVM Algorithm

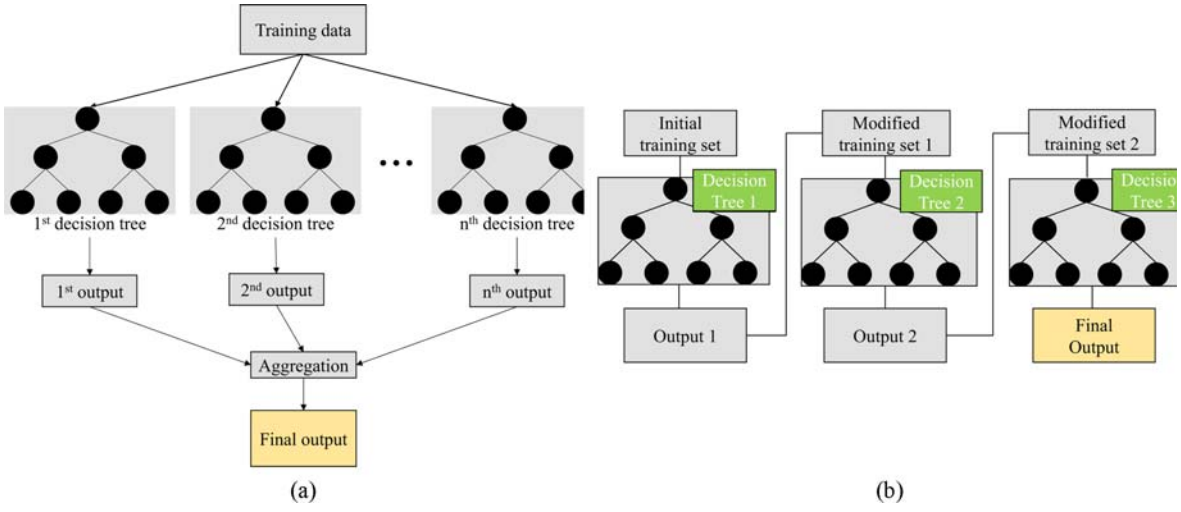

Fig. 5. Workflow of RF and XGB Algorithms: (a) RF Algorithm, (b) XGB Algorithm

k-nearest neighbors to the testing data point.

The SVM algorithm is based on the concept of finding the optimal hyperplane that best separates two classes of data points as shown in Fig. 4(b). The hyperplane is formed in a way that maximizes the margin, which is the distance between the hyperplane and the nearest data points of each class. In other words, the SVM algorithm aims to find the decision boundary that maximizes the distance between two classes while minimizing the chance of misclassification. As a result, new data points can be classified using the obtained decision boundaries.

The RF algorithm is based on the concept of building many decision trees using different subsets of the training data (see Fig. 5(a)). Each decision tree in the forest is trained on a random subset of the training data. The final prediction is obtained by aggregating the predictions of all the individual trees. In classification problems, aggregation is typically performed by taking the majority vote for prediction.

The XGB algorithm is a state-of-the-art sequential learning algorithm that sequentially adds decision trees into the ensemble in an additive manner with the objective of correcting the previous decision trees, as shown in Fig. 5(b). The XGB algorithm is an ensemble learning method that combines multiple decision trees to form an accurate and robust prediction model. The algorithm minimizes the loss function by adding new trees to the ensemble, with each new tree trying to correct the errors made by the previous trees.

3. FE Numerical Simulation

3.1 Verification of FE Model

In this study, a 3D finite-element (FE) numerical model was developed using the commercial software COMSOL Multiphysics to simulate electrical resistivity surveys and predict geological risks during tunnel excavation. The FE model allowed for the construction of a measurement database for the ML algorithms. The governing equations used in the FE model were based on the equations of continuity and Gaussian's law, as expressed in Eq. (5).

$$-\nabla\cdot\left(\frac{1}{\rho}\nabla V - J^{e}\right) = Q_{j}, \tag{5}$$

where ρ is the electrical resistivity of a ground formation, $V$is the voltage, J^{e} is the externally injected electrical current, and Q_{j} is the electrical current source. The voltage (V) of each element, is obtained by Eq. (6).

$$\vec{J} = -\left(\frac{1}{\rho}\right)\vec{\nabla}V \tag{6}$$

No-flux boundary conditions (i.e., electrical insulation) were adopted at the external boundaries of the polycarbonate soil chamber in contact with air (referring to Eq. (7)).

$$n\cdot\vec{J} = 0, \tag{7}$$

where n is a normal vector.

To verify the developed FE model, laboratory-scale chamber experiments were conducted by simulating ground formations with faults in a soil chamber. The experimental setup consisted of a tunnel model, a soil chamber, a measuring device, electrodes, and a power supply, as shown in Fig. 6. The tunnel model and soil chamber were made of polycarbonate, which is a nonconductive material. The electrical resistivity of the synthetic faults was measured using the Supersting R8/IP (Advanced Geosciences, Inc.). The arrangement for tunnel model advancement with electrodes was designed to resemble a tunnel excavation and electrical resistivity survey on a tunnel face (Fig. 6). The dimensions of the FE model were equivalent to those of the experimental setup, as shown in Fig. 7. Table 1 presents the material properties of the rock and soil specimens used in the laboratory-scale test, including granite (rocky ground) and gravel (fault gouges) to simulate ground conditions with faults. The rock and soil specimens were saturated with tap water for more than 24 hours, with the tap water having an electrical resistivity of 30.25 Ω·m and a salinity of 0.4‰ at a temperature of 8°C. However, only the electrical properties (i.e., electrical resistivity, relative permittivity) were utilized in the software (i.e., COMSOL Multiphysics) for the FE method. The AC/DC module within the software is designed specifically for electrical and electromagnetic simulations and does not support the direct input of macroscopic mechanical parameters. Consequently, the developed FE model incorporated only the electrical properties. Nevertheless, this approach is justified because electrical resistivity can, to some extent, reflect mechanical properties, a relationship supported by numerous previous experimental studies (Banton et al., 1997; Bryson, 2005;

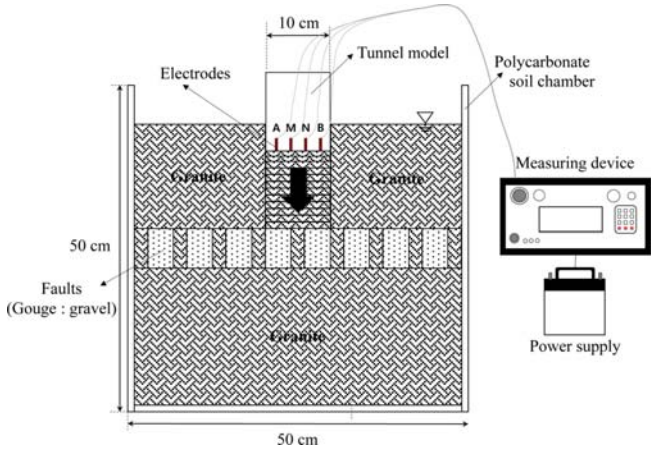

Fig. 6. Schematics of Experimental Setup

Table 1. Electrical and Mechanical Properties of Testing Materials

Type	Property	Value
Gravel	Water content [%]	3.03
	Coefficient of uniformity (C_u)	0.6696
	Coefficient of curvature (C_c)	0.8673
	Dry unit weight (r_d) [kN/m^3]	14.396
	Soil type (USCS)	GP
	Electrical resistivity (ρ) [Ω·m]	119.665
	Relative permittivity	20
Granite	Dry unit weight (r_d) [kN/m^3]	25.064
	Electrical resistivity (ρ) [Ω·m]	303.89
	Relative permittivity	7
Polycarbonate	Electrical resistivity (ρ) [Ω·m]	1.15×10^{13}
	Relative permittivity	0.866

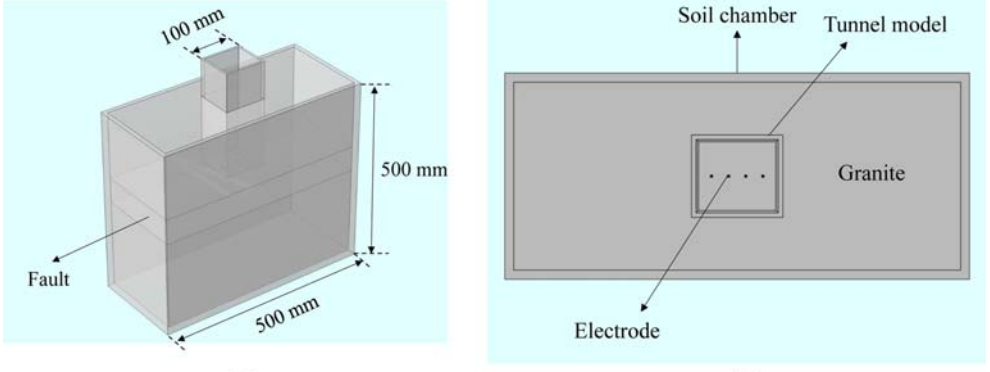

Fig. 7. FE Model for Simulating Laboratory-Scale Experiment of Electrical Resistivity Survey: (a) Perspective View, (b) Plan View

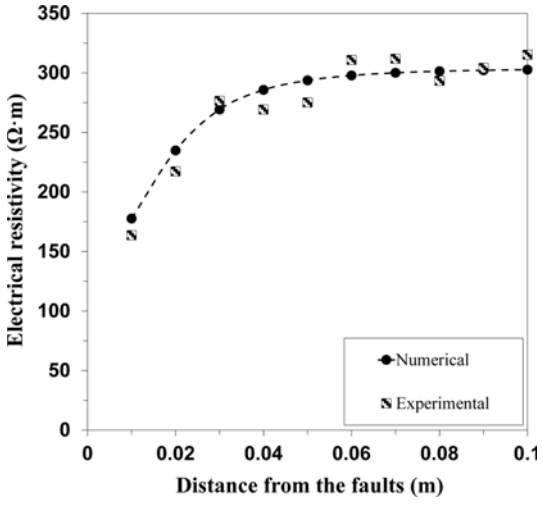

Fig. 8. Comparison between the Numerical and Experimental Results

Abidin et al., 2013; Hazreek et al., 2015; Pandey et al., 2015; Alsharari et al., 2020).

The numerical analysis results were compared with the experimental results, as shown in Fig. 8. The average error and root-mean-square error (RMSE) of the numerical results were 4.59% and 13.05 Ω·m, respectively, when compared to the experimental results. Thus, the developed FE model was able to accurately simulate the electrical resistivity surveys during tunnel excavation.

3.2 Simulation of Geological Risks Ahead of a Tunnel Face

The developed numerical model was used to simulate several geological risks that can significantly affect tunnel construction safety and efficiency, such as fault fracture zones, water intrusion, mixed ground, geological transition, and cavities (Tóth et al., 2013; Hyun et al., 2015; Gong et al., 2016), as shown in Fig. 9. Therefore, these hazardous conditions were incorporated in the FE numerical model to simulate the prediction of such conditions ahead of a tunnel face.

In this study, five hazardous ground conditions were simulated by modifying the geometry of the developed FE numerical model. It was assumed that the tunneling site was located on a granite ground formation and the geological risks were composed of granular soil (i.e., faults and cavities), tap water (i.e., water intrusion), and relatively weak rock formations (i.e., mixed ground and geological transition). The tunnel diameter of the modified FE model was set to 10 m, which is a typical value for transportation tunnels (as shown in Fig. 10(a)). The tunnel diameter (D) and distance (L) between the geological risks and the tunnel face were selected as variables. A tetrahedral configuration was used to generate the FE mesh. The element size in the FE numerical

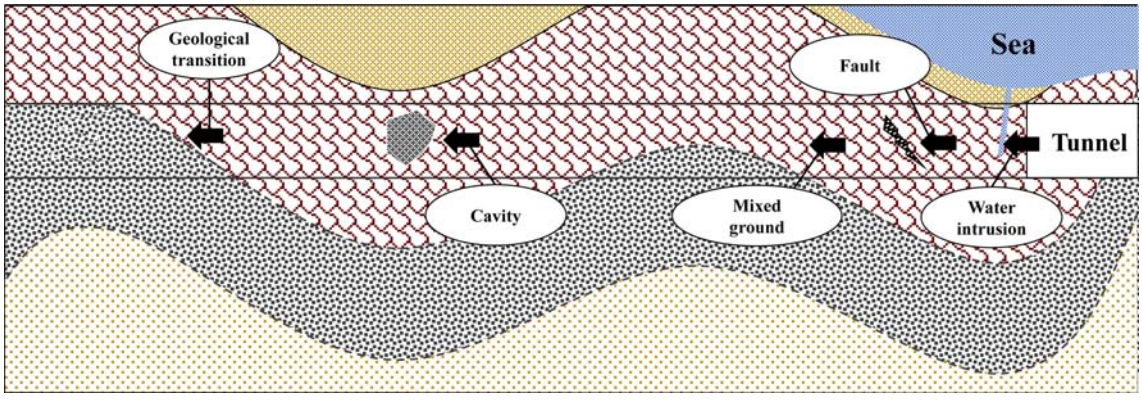

Fig. 9. Hazardous Ground Conditions during Tunnel Construction

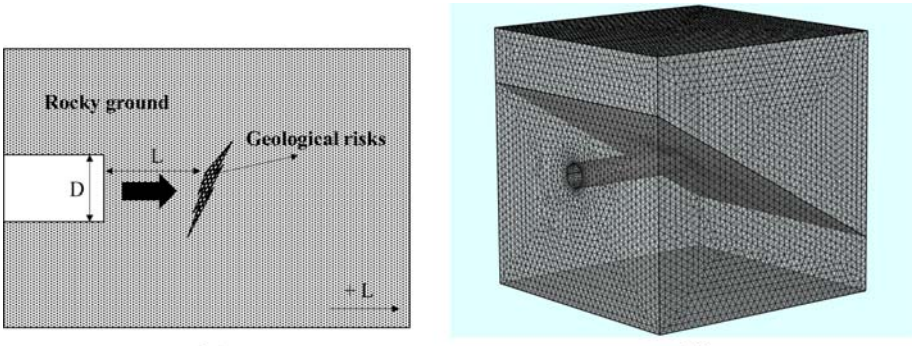

Fig. 10. Numerical Modeling for Simulating Geological Risks Ahead of a Tunnel Face: (a) FE Numerical Model Outline, (b) Mesh Configurations of FE Model

model ranged from 0.03 m to 3 m, as depicted in Fig. 10(b).

During tunnel excavation, fault zones can cause ground subsidence and water ingress into the tunnel face, resulting in reduced construction safety and efficiency. In this study, faults were simulated by placing granular soil (with an electrical resistivity range of 100 – 150 Ω·m) representing fault gouge between the rock ground formations (i.e., granite), as shown in Fig. 11(a). In addition, water intrusion was simulated by filling the rock formations with tap water in the FE numerical model (Fig. 11(a)).

Mixed ground conditions refer to geological strata that contain two or more materials adjacent to the tunnel excavation. Such unexpected geological situations can significantly influence the performance of the excavation apparatus, leading to additional costs, low penetration rates, long downtimes, and indefinite delays. In this study, mixed ground conditions were simulated by combining granite and relatively weak rock specimens. The interface slope between the two different ground formations was set to 30°, as shown in Fig. 11(b).

Accurately predicting geological transitions ahead of a tunnel face is crucial for managing risks such as face collapse, ground settlement, and sinkhole. In this study, geological transitions were simulated by arranging adjacent ground formations (i.e.,

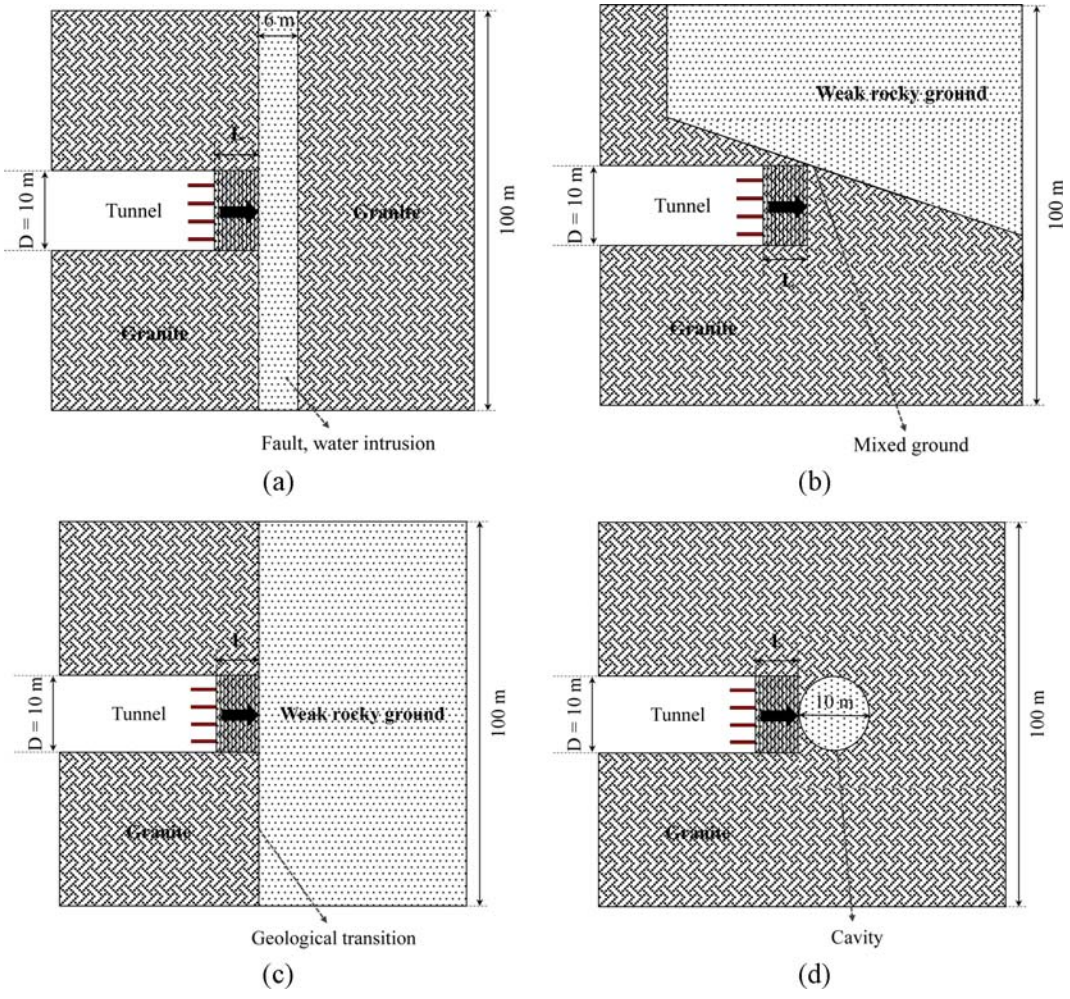

Fig. 11. Illustration of Numerical Analysis Cases for Predicting Geological Risks: (a) Fault, Water Intrusion, (b) Mixed Ground, (c) Geological Transition, (d) Cavity

Table 2. Numerical Analysis Cases for Predicting Geological Risks

	Class 1	Class 2	Class 3	Class 4	Class 5
Type	Fault	Water intrusion	Mixed ground (Slope = 30°)	Geological transition	Cavity
Material	Granular soil	Tap water	Relatively weak rock formations	Relatively weak rock formations	Granular soil
Electrical resistivity of adjacent rock	375 Ω·m				
Electrical resistivity of risks	100 – 150 Ω·m	5 – 30 Ω·m	150 – 200 Ω·m	150 – 200 Ω·m	100 – 150 Ω·m

granite) and changing ground formations with relatively weak strength. The changing ground formation was applied to the rock specimens with an electrical resistivity range of 150 – 200 Ω·m. In the developed model, geological transitions were simulated by arranging granite and weaker rock specimens in the adjacent ground and changing the ground, respectively, as shown in Fig. 11(c).

Cavities in the rocky ground may also be encountered during tunnel excavation, leading to water ingress and ground settlement that can halt tunnel construction. Therefore, it is crucial to manage the potential risks caused by cavities. In the numerical analysis, the cavity was simulated as a spherical space filled with granular soil (similar to faults) in granite (Fig. 11(d)). The cavity had a diameter of 10 m. In summary, Table 2 presents the five types of geological risks that were simulated.

4. Model Implementation

4.1 Analysis of the Input Parameters

In this study, an ML model was developed to predict five geological risks encountered during tunnel excavation in rocky ground composed of granite, including typical fault zones, water intrusion, mixed ground, geological transition, and cavities. The granite material considered has an electrical resistivity of 375 Ω·m.

An electrical resistivity measurement dataset was constructed for each ground condition corresponding to the tunnel advancement using the developed FE model. In the datasets, the distance from the tunnel face to the geological risks (L) was normalized using the tunnel diameter (D). Tunnel excavation was simulated from L/D = 0.5 to 0, with L/D decreasing by 0.05 increments, where L/D = 0 represents the tunnel face reaching the geological risks (refer to Fig. 11).

For geological risks excluding water intrusion, five hundred measurement datasets were collected for each condition by varying the electrical resistivity of the risks in increments of 0.1 Ω·m (referring to Table 2). However, for water intrusion scenarios, data were collected with adjustments in increments of 0.05 Ω·m. Note that this variation indicated that the electrical resistivity of the ground formations at an actual tunneling site had different values or ranges depending on various factors, even for the same material.

As a result, a database of 2500 datasets was compiled. Each dataset consists of the electrical resistivity values measured at each L/D during tunnel excavation under specific ground conditions, along with the corresponding geological risk condition (i.e., class), as illustrated in Table 3. The distribution of the database is represented with respect to *L*/*D* and is summarized in Fig. 12. In each histogram within Fig. 12, the vertical and horizontal coordinates represent the number of data points and measured electrical resistivity (Ω·m), respectively. As the tunnel face approaches the geological risk, the distribution shape of the measurement database becomes similar to that of a multimodal distribution. However, the Gaussian distribution was represented by the long distance

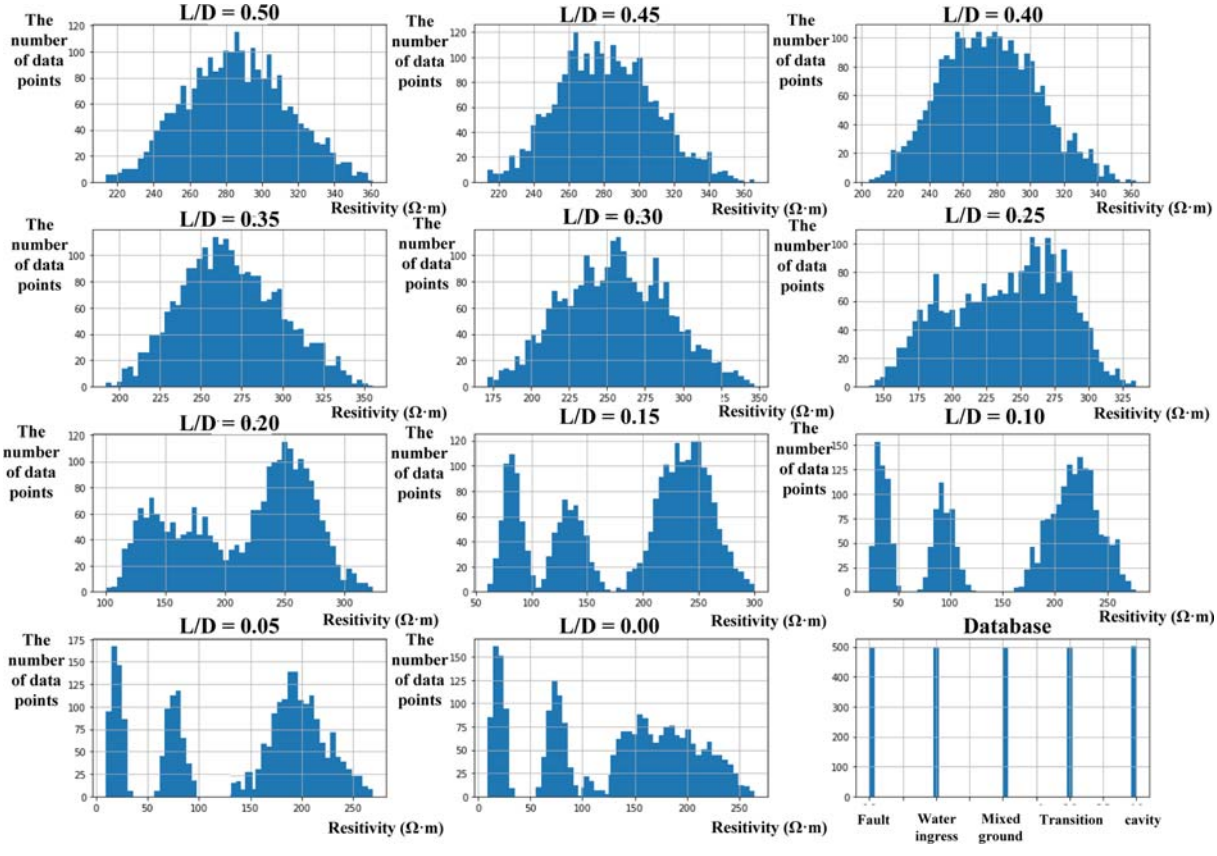

Fig. 12. Histograms of the Database Applied to the ML Algorithm

Table 3. Statistical Descriptions of Electrical Resistivity in the Numerical Database

	L/D	Average (Ω·m)	Minimum (Ω·m)	Maximum (Ω·m)	Standard deviation (Ω·m)
Class 1 (Fault)	0.5	286.9722	223.596	352.1907	41.94198
	0.45	282.628	217.7086	345.6706	40.32048
	0.4	269.9267	209.5919	334.3043	39.10779
	0.35	257.969	199.4546	318.5273	36.25306
	0.3	244.4501	185.0296	296.076	29.73532
	0.25	235.053	170.9927	295.2258	27.78344
	0.2	215.2903	166.6109	289.12	26.01672
	0.15	205.2578	164.2493	267.085	27.00119
	0.1	212.0286	138.9595	305.4346	30.23661
	0.05	191.6532	138.3559	254.0201	27.04116
	0.00	179.0104	134.15	223.8835	25.20512
Class 2 (Water intrusion)	0.5	279.4866	215.7991	342.2063	40.59312
	0.45	271.5505	209.6786	333.7445	37.83347
	0.4	258.4895	198.8835	317.0344	36.7599
	0.35	240.8797	184.0847	297.5619	33.08703
	0.3	215.3123	164.402	268.7418	30.83074
	0.25	182.5997	137.2594	228.8185	25.6353
	0.2	133.9804	99.77077	173.8713	19.64693
	0.15	82.75644	56.77636	112.7648	13.30907
	0.1	35.34405	20.81047	53.90161	7.384876
	0.05	20.15323	8.556905	35.05197	5.733576
	0.00	19.7097	7.955793	34.65867	5.528376
Class 3 (Mixed ground)	0.5	302.0047	236.6901	372.7687	44.09175
	0.45	297.457	235.1557	368.4915	42.00475
	0.4	297.9642	233.8494	365.3976	40.75258
	0.35	294.9797	231.0201	361.5834	42.00982
	0.3	289.6498	226.6319	354.0285	40.03056
	0.25	281.1511	221.9375	343.092	40.56087
	0.2	271.5348	213.5108	329.9162	38.18108
	0.15	253.7349	200.3849	305.8547	35.31018
	0.1	228.5713	181.2196	272.2478	31.91506
	0.05	189.2418	149.4852	231.4989	27.32425
	0.00	143.8522	109.9909	179.8803	20.99784
Class 4 (Geological transition)	0.5	298.224	233.2321	366.8467	42.51971
	0.45	294.9015	230.6846	363.5852	41.88469
	0.4	290.6116	227.5292	359.6608	41.26868
	0.35	285.5348	222.4549	351.6535	40.22642
	0.3	273.4369	215.5187	340.1034	39.72768
	0.25	265.9789	205.847	325.2792	37.51284
	0.2	246.9731	191.9106	304.5647	34.84964
	0.15	226.3523	173.2689	278.7849	32.12198
	0.1	198.2925	154.0159	248.6614	29.8746
	0.05	185.451	142.0172	230.6026	26.72848
	0.00	179.5961	139.3156	226.1048	26.05643
Class 5 (Cavity)	0.5	259.5971	207.707	316.0247	36.98412
	0.45	264.085	208.5646	316.8517	37.35396
	0.4	260.1759	208.5967	316.9708	37.34936
	0.35	263.4504	207.6316	315.9628	37.68098
	0.3	261.7859	207.0586	314.4558	37.04668
	0.25	254.6629	202.0502	305.8869	36.10362
	0.2	248.412	196.7086	296.5491	34.687
	0.15	239.6565	191.5156	287.3131	33.36255
	0.1	231.1566	185.4132	279.4352	33.2637
	0.05	228.7498	180.6503	273.838	31.96999
	0.00	223.2701	175.1832	267.3564	32.16927

between the tunnel and risky ground conditions. Table 3 presents the statistical variables of the numerical database including the minimum, average, maximum, and standard deviation.

4.2 Implementation Procedure

The implementation procedure of the developed ML model was divided into two steps: creating the dataset from the numerical analysis and applying the dataset to the selected ML algorithms, such as SVM, KNN, RF, and XGB. Firstly, the input dataset collected using the FE model was standardized by adjusting the average and variance to zero and one, respectively, to represent a normal distribution for each input dataset. The standardized dataset was divided into training and testing datasets, with the percentage of 80% and 20%, respectively.

To execute the learning process, a grid search and a five-fold cross-validation were conducted to create a geological risk prediction model using each ML algorithm. The grid search is an optimization technique that searches for the optimal combination of hyperparameters to obtain the maximum prediction performance. Furthermore, the five-fold cross-validation was employed to assess the performance of the machine learning model by dividing the available data into five equal folds. The process involved training the model on four folds and testing it on the remaining fold. This was repeated five times, with each fold used exactly once as the test set. The final evaluation of the model performance was based on the average of the five test set results (shown in Fig. 13).

In this study, the prediction performance of the classification models was evaluated using a confusion matrix, and average accuracy. The average accuracy was calculated by dividing the number of correct prediction results in the main diagonal of the confusion matrix by the total number of results (refer to Eq. (8)).

$$Accuracy = \frac{Correction\ predictions}{All\ prediction} \tag{8}$$

Figure 14 shows the detailed development procedure of the ML model.

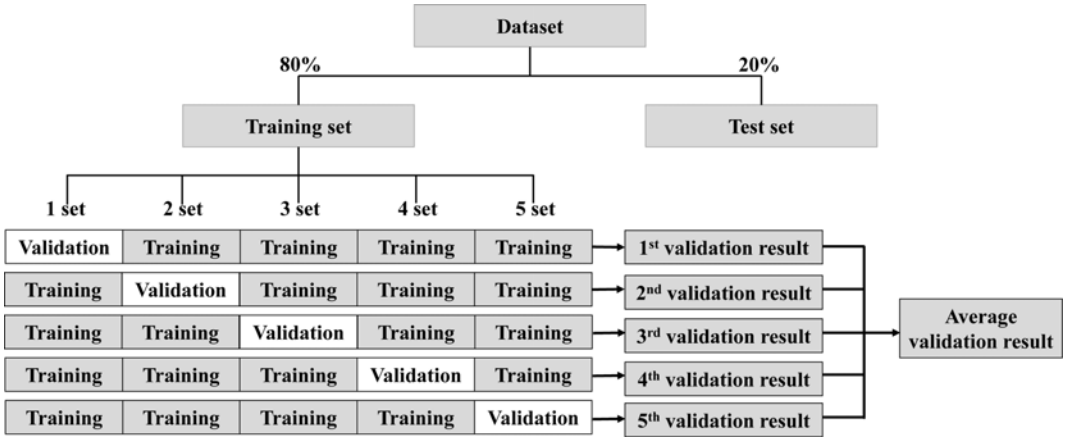

Fig. 13. Five-Fold Cross-Validation Procedure

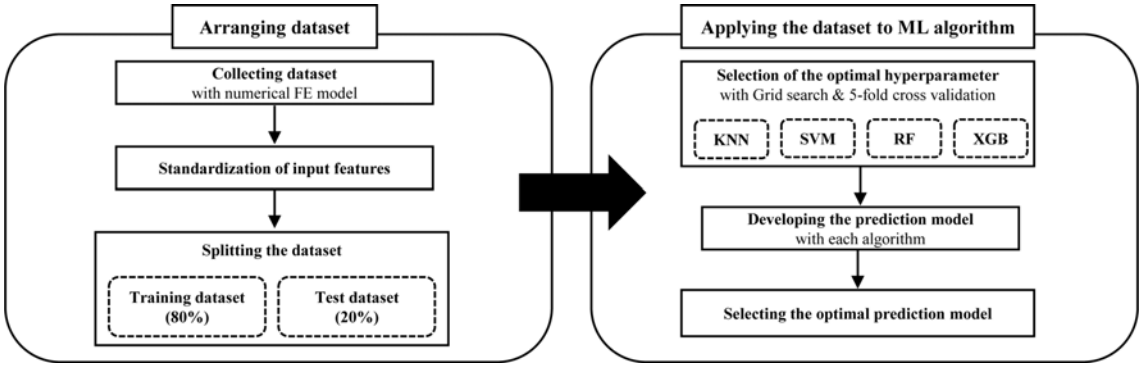

Fig. 14. Flowchart of Development Procedure for Optimal Model

Table 4. Prediction Performance of Geological Risks and Optimal Hyperparameters

Algorithm	Prediction error (%)	Optimal hyperparameter		
		Type	Optimal values or types	Description
KNN	8	Metric	Euclidean	Distance used to compute the similarity between data points
		N neighbors	14	Number of neighbors
		Weights	Distance	Importance given to each of k-nearest neighbors when making a prediction
SVM	8.93	C	60	Penalty for misclassified data points
		Gamma	0.05	Distance of influenced data points
		Kernel	rbf	Similarity between data points in the feature space
RF	2.67	Max depth	20	Maximum depth of tree models
		Min samples split	3	Minimum number of samples required to split a node
		N estimators	100	Number of tree models
XGB	4.27	Learning rate	0.05	Step size shrinkage used in the update to prevent overfitting
		Max depth	29	Maximum depth of tree models
		Min child weight	3	Minimum sum of weights in a leaf
		N estimators	3000	Number of tree models
		Subsample	0.1	Subsample ratio of features in each tree

4.3 Results and Discussions

The prediction performance of the ML models based on the KNN, SVM, RF, and XGB algorithms were evaluated using the five-fold cross-validation. The prediction errors of the KNN, SVM, RF, and XGB models were found to be 8%, 8.93%, 2.67%, and 4.27%, respectively. Table 4 summarizes the prediction performance and optimal hyperparameters of the prediction model for each ML algorithm.

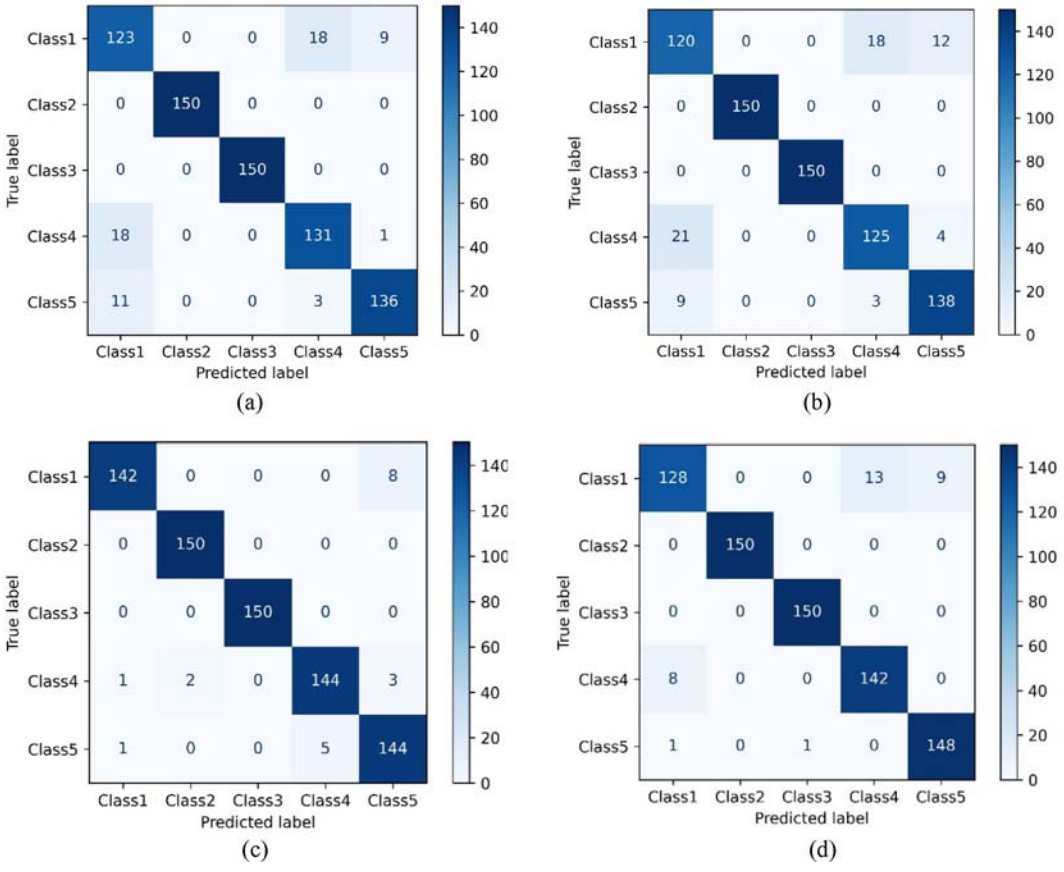

Fig. 15. Confusion Matrices of the Adopted ML Models: (a) KNN Algorithm, (b) SVM Algorithm, (c) RF Algorithm, (d) XGB Algorithm

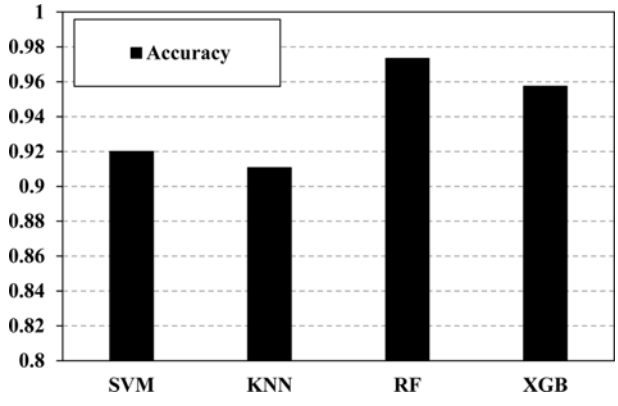

Fig. 16. Evaluation of the Accuracy as the Prediction Performance

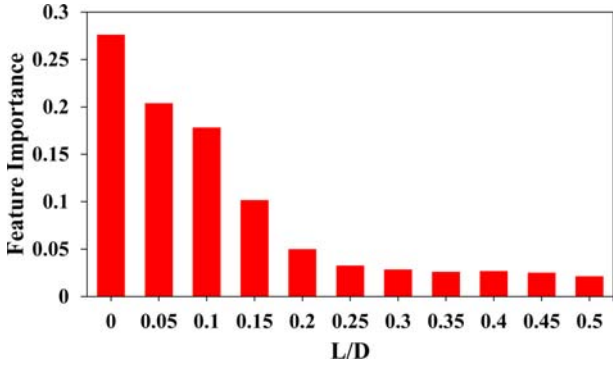

Fig. 17. Feature Importance of the RF Algorithm

Confusion matrices of the four ML models demonstrate the training and test results (Fig. 15). Classes 1, 2, 3, 4, and 5 in these matrices represent faults, water intrusion, mixed ground, geological transition, and cavities, respectively. The average accuracies obtained from the four ML models are shown in Fig. 16. Among the adopted ML algorithms, the RF model exhibited the highest prediction performance, with an average accuracy of 0.9733, making it the most suitable model for identifying geological risks. In conclusion, the developed ML model with the RF algorithm accurately predicts geological risks using the measurement database obtained from the numerical analysis, providing a powerful tool for managing potential risks during tunneling.

The feature importance was utilized to identify the most relevant variables in the input dataset applied to the developed model (Fig. 17). The feature importance increased when the tunnel face approached the geological risks (i.e., L/D decreased), particularly when the distance between the tunnel face and the geological risk (L) was less than 0.15 times the tunnel diameter (D). This feature importance can be useful in understanding the relative importance of adopted features and improving the performance of the ML model.

4.4 Application of Prediction Results in Practice

To verify the practical applicability of the developed prediction

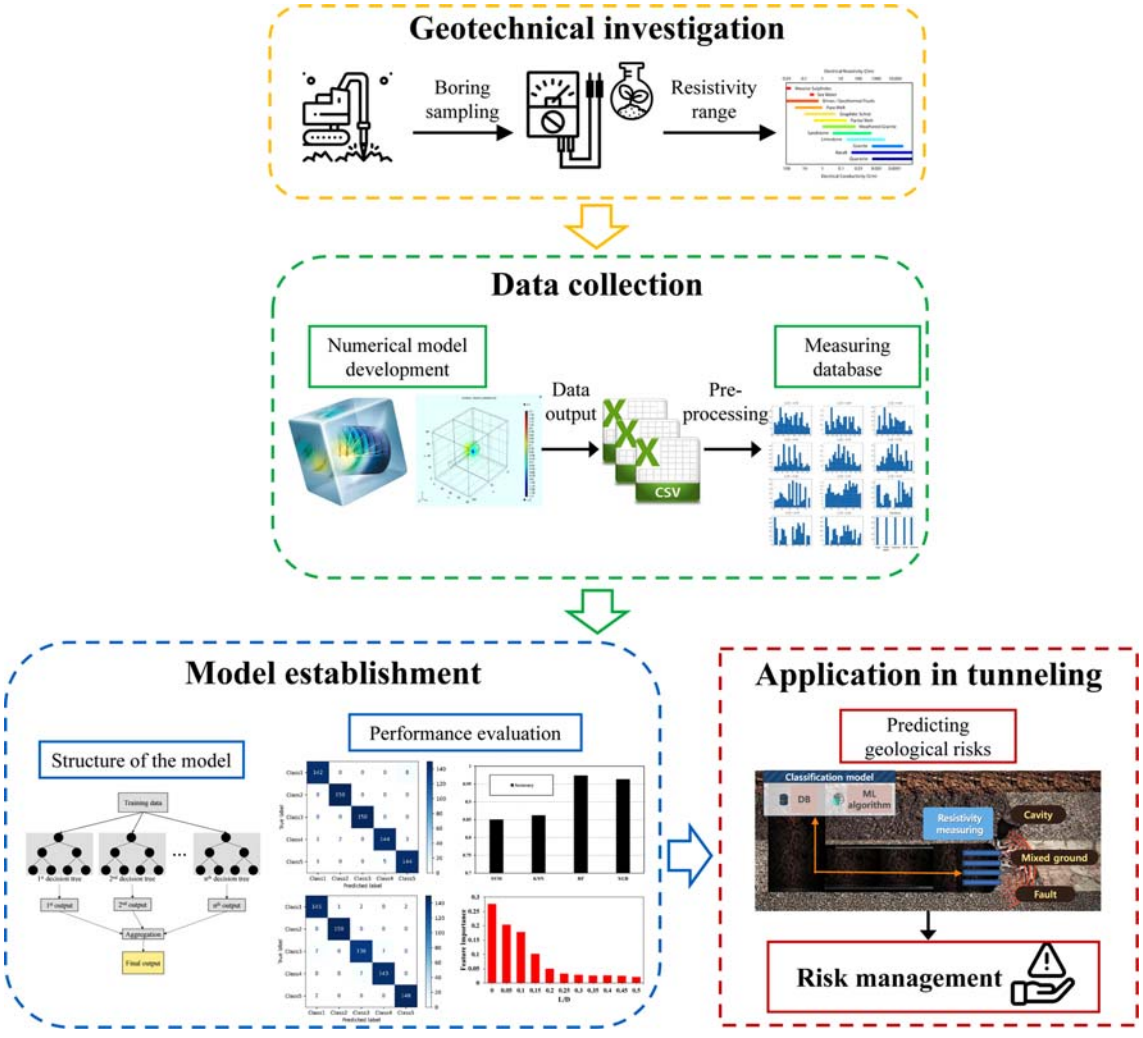

Fig. 18. Flowchart for Application of Geological Risk Prediction Model in Practice

model, an engineering flowchart for predicting geological risks using ML algorithms based on the electrical resistivity database obtained from the numerical analysis is shown in Fig. 18. Firstly, a geotechnical investigation should be conducted along the tunnel alignment to obtain the electrical resistivity range of the soil and rock specimens in the ground formations. It is recommended to measure the electrical resistivity of undisturbed specimens obtained from a subsurface exploration. Tunnel engineers can identify several anticipated geological risks during tunnel excavation from the geological profiles along the tunnel routes.

Based on the measured electrical resistivity of the rock and soil specimens, an FE numerical model should be developed to simulate the geological risks that can be encountered during construction at the tunneling site. Subsequently, an electrical resistivity database can be constructed based on the developed numerical FE model corresponding to the potential geological risks.

Finally, a geological risk prediction model should be developed with the measurement database constructed using the FE model. It is recommended in this study to employ the RF algorithm in developing a prediction model owing to its high performance. During tunnel excavation, an appropriate measuring apparatus may be installed to obtain the electrical resistivity data corresponding to the tunnel advancement. Using the developed model and in-situ electrical resistivity measurements, the tunnel engineer can manage potential risks beforehand by predicting geological risks ahead of a tunnel face.

It's important to note that the developed prediction model is specifically designed for tunnel excavation in rocky ground primarily composed of granite. Furthermore, the model currently predicts five geological risks: faults, water intrusion, mixed ground, geological transitions, and cavities. This limitation necessitates further research to extend the model's applicability to various geological conditions and additional risks. Future studies should focus on developing more versatile models that can accommodate diverse geological scenarios and a broader range of risks.

5. Conclusions

This study aimed to develop an ML-based electrical resistivity survey framework for predicting geological risks ahead of tunnel

faces. To build the database for the ML algorithms, an FE numerical model was developed to simulate the electrical resistivity survey during tunnel advancement toward potential geological risks. Four ML classification algorithms were considered with the resistivity measurement database, which includes five hazardous ground conditions: typical faults, water intrusion, mixed ground, geological transition, and cavity zones. Overall, the prediction error for all the algorithms was less than 8.93%, indicating a sufficient prediction performance. Based on the study results, the following conclusions can be made.

1. The RF algorithm was found to be the most suitable for predicting the geological risks among the commonly used ML classification algorithms (KNN, SVM, RF, and XGB), due to its low prediction error of 2.67%.
2. The prediction results of all the ML algorithms showed that faults, water intrusions, geological transitions, mixed ground, and cavities could be classified. However, compared to other geological risks, faults and cavities were classified with relatively high prediction errors. Therefore, a detailed examination is necessary to appropriately manage the risks when applying the developed model to predict a fault or cavities ahead of a tunnel face.
3. The study results lead to the proposal of an engineering flowchart for accurately predicting the various geological risks in front of tunnel faces. The flowchart outlines several processes, including geotechnical surveys, numerical analysis, and ML algorithms, that can be applied during tunnel construction.
4. Despite the promising prediction performance of the developed ML model, additional research is required to validate and refine the proposed method and engineering flowchart in real-world tunneling projects. Moreover, expanding the database to include a wider range of operational parameters and geological conditions will be essential for improving the model's prediction accuracy and robustness. This will facilitate the adaptation of the model to a variety of field conditions, thereby enhancing its practical applicability.

Acknowledgments

This research was supported by the National R&D Project for Smart Construction Technology (No. RS-2020-KA157074) funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure, and Transport, and managed by the Korea Expressway Corporation.

ORCID

Minkyu Kang https://orcid.org/ 0000-0001-8205-5424
Khanh Pham https://orcid.org/0000-0001-6566-6776
Kibeom Kwon https://orcid.org/0000-0001-8889-1386
Seunghun Yang https://orcid.org/0009-0004-1752-9717
Hangseok Choi https://orcid.org/0000-0002-2040-8850

References

Abidin MZ, Wijeyesekera D, Saad R, Ahmad F (2013) The influence of soil moisture content and grain size characteristics on its field electrical resistivity. *Electronic Journal of Geotechnical Engineering* 18:699-705, DOI: 10.1088/1742-6596/495/1/012014

Alsharari B, Olenko A, Abuel-Naga H (2020) Modeling of electrical resistivity of soil based on geotechnical properties. *Expert Systems with Applications* 141:112966, DOI: 10.1016/j.eswa.2019.112966

Arumugam K, Naved M, Shinde P, Leiva-Chauca O, Huaman-Osorio A, Gonzales-Yanac T (2023) Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings 80:3682-3685, DOI: 10.1016/j.matpr.2021.07.361

Bai XD, Cheng WC, Ong DE, Li G (2021) Evaluation of geological conditions and clogging of tunneling using machine learning. *Geomechanics and Engineering* 25(1):59-73, DOI: 10.12989/gae.2021.25.1.059

Banton O, Cimon M, Seguin M (1997) Mapping field-scale physical properties of soil with electrical resistivity. *Soil Science Society of America Journal* 61(4):1010-1017, DOI: 10.2136/sssaj1997.03615995006100040003x

Bayati M, Hamidi JK (2017) A case study on TBM tunnelling in fault zones and lessons learned from ground improvement. *Tunnelling and Underground Space Technology* 63:162-170, DOI: 10.1016/j.tust.2016.12.006

Broere W (2016) Urban underground space: Solving the problems of today's cities. *Tunnelling and Underground Space Technology* 55: 245-248, DOI: 10.1016/j.tust.2015.11.012

Bryson LS (2005) Evaluation of geotechnical parameters using electrical resistivity measurements. *In Earthquake Engineering and Soil Dynamics*, 1-12, DOI: 10.1061/40779(158)10

Carrière SD, Chalikakis K, Sénéchal G, Danquigny C, Emblanch C (2013) Combining electrical resistivity tomography and ground penetrating radar to study geological structuring of karst unsaturated zone. *Journal of Applied Geophysics* 94:31-41, DOI: 10.1016/j.jappgeo.2013.03.014

Chung H, Park J, Kim BK, Kwon K, Lee IM, Choi H (2021) A causal network-based risk matrix model applicable to shield TBM tunneling projects. *Sustainability* 13(9):4846, DOI: 10.3390/su13094846

Dickmann T, Sander BK (1996) Drivage concurrent tunnel seismic prediction (TSP). *Felsbau* 14(6):406-411

Eftekhari A, Aalianvari A, Rostami J (2018) Influence of TBM operational parameters on optimized penetration rate in schistose rocks, a case study: Golab tunnel Lot-1, Iran. *Computers and Concrete* 22(2):239-248, DOI: 10.12989/CAC.2018.22.2.239

Farrokh E, Rostami J (2009) Effect of adverse geological condition on TBM operation in Ghomroud tunnel conveyance project. *Tunnelling and Underground Space Technology* 24(4):436-446, DOI: 10.1016/j.tust.2008.12.006

Gong Q, Yin L, Ma H, Zhao J (2016) TBM tunnelling under adverse geological conditions: An overview. *Tunnelling and Underground Space Technology* 57:4-17, DOI: 10.1016/j.tust.2016.04.002

Grodner M (2001) Delineation of rockburst fractures with ground penetrating radar in the Witwatersrand Basin, South Africa. *International Journal of Rock Mechanics and Mining Sciences* 38(6):885-891, DOI: 10.1016/S1365-1609(01)00054-5

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

2019.04.006

Hazreek Z, Aziman M, Azhar A, Chitral W, Fauziah A, Rosli S (2015) The behaviour of laboratory soil electrical resistivity value under basic soil properties influences. IOP Conference Series: Earth and Environmental Science 23(1):012002. IOP Publishing, DOI: 10.1088/1755-1315/23/1/012002

Hou S, Liu Y, Zhuang W, Zhang K, Zhang R, Yang Q (2023) Prediction of shield jamming risk for double-shield TBM tunnels based on numerical samples and random forest classifier. *Acta Geotechnica* 18(1):495-517, DOI: 10.1007/s11440-022-01567-9

Hyun KC, Min S, Choi H, Park J, Lee IM (2015) Risk analysis using fault-tree analysis (FTA) and analytic hierarchy process (AHP) applicable to shield TBM tunnels. *Tunnelling and Underground Space Technology* 49:121-129, DOI: 10.1016/j.tust.2015.04.007

Jung JH, Chung H, Kwon Y, Lee IM (2019) An ANN to predict ground condition ahead of tunnel face using TBM operational data. *KSCE Journal of Civil Engineering* 23(7):3200-3206, DOI: 10.1007/s12205-019-1460-9

Kafy A, Bakshi A, Saha M, Al Faisal A, Almulhim AI, Rahaman ZA, Mohammad P (2023) Assessment and prediction of index based agricultural drought vulnerability using machine learning algorithms. *Science of The Total Environment* 867:161394, DOI: 10.1016/j.scitotenv.2023.161394

Kang M, Kim S, Lee J, Choi H (2022) FE model of electrical resistivity survey for mixed ground prediction ahead of a TBM tunnel face. *Geomechanics and Engineering* 29(3):301-310, DOI: 10.12989/gae.2022.29.3.301

Kang M, Lee J, Kwon K, Park S, Choi H (2023) Laboratory simulations on hybrid non-destructive survey of electrical resistivity and induced polarization to predict geological risks ahead of a TBM tunnel. *Tunnelling and Underground Space Technology* 135:105066, DOI: 10.1016/j.tust.2023.105066

Kaus A, Boening W (2008) BEAM–Geoelectrical ahead monitoring for TBM-Drives. *Geomechanik und Tunnelbau: Geomechanik Und Tunnelbau* 1(5):442-449, DOI: 10.1002/geot.200800048

Lee KH, Park J, Park J, Lee IM, Lee SW (2019) Electrical resistivity tomography survey for prediction of anomaly in mechanized tunneling. *Geomechanics and Engineering* 19(1):93-104, DOI: 10.12989/gae.2019.19.1.093

Lee HL, Song K, Qi C, Kim KY (2022) Sequential prediction of TBM penetration rate using a gradient boosted regression tree during tunneling. *Geomechanics and Engineering* 29(5):523-533, DOI: 10.12989/GAE.2022.29.5.523

Lee S, Moon JS (2020) Excessive groundwater inflow during TBM tunneling in limestone formation. *Tunnelling and Underground Space Technology* 96:103217, DOI: 10.1016/j.tust.2019.103217

Li JB, Chen ZY, Li X, Jing LJ, Zhang YP, Xiao HH, Wang SJ, Yang WK, Wu LJ, Li PY, Li HB, Yao M, Fan LT (2023) Feedback on a shared big dataset for intelligent TBM Part I: Feature extraction and machine learning methods. *Underground Space* 11:1-25, DOI: 10.1016/j.undsp.2023.01.001

Liu B, Guo Q, Liu Z, Wang C, Nie L, Xu X, Chen L (2019) Comprehensive ahead prospecting for hard rock TBM tunneling in complex limestone geology: A case study in Jilin, China. *Tunnelling and Underground Space Technology* 93:103045, DOI: 10.1016/j.tust.2019.103045

Liu Q, Wang X, Huang X, Yin X (2020) Prediction model of rock mass class using classification and regression tree integrated AdaBoost algorithm based on TBM driving data, *Tunnelling and Underground Space Technology* 106:103595, DOI: 10.1016/j.tust.2020.103595

Mahmoodzadeh A, Nejati H, Ibrahim H, Ali H, Mohammed A, Rashidi S, Majeed M (2022a) Several models for tunnel boring machine performance prediction based on machine learning, *Geomechanics and Engineering* 30(1):75-91, DOI: 10.12989/GAE.2022.30.1.075

Mahmoodzadeh A, Nejati H, Mohammadi M, Mohammed A, Ibrahim H, Rashidi S (2022b) Numerical and Machine learning modeling of hard rock failure induced by structural planes around deep tunnels. *Engineering Fracture Mechanics* 271:108648, DOI: 10.1016/j.engfracmech.2022.108648

McDowell PW, Barker RD, Butcher AP, Culshaw MG, Jackson PD, McCann DM, Arthur JCR (2002) Geophysics in engineering investigations. Ciria, London, UK, 61-68

Méndez M, Merayo MG, Núñez M (2023) Machine learning algorithms to forecast air quality: A survey. *Artificial Intelligence Review*, 1-36, DOI: 10.1007/s10462-023-10424-4

Mifkovic M, Swidinsky A, Mooney M (2021) Imaging ahead of a tunnel boring machine with DC resistivity: A laboratory and numerical study. *Tunnelling and Underground Space Technology* 108:103703, DOI: 10.1016/j.tust.2020.103703

Pallathadka H, Wenda A, Ramirez-Asís E, Asís-López M, Flores-Albornoz J, Phasinam K (2023) Classification and prediction of student performance data using various machine learning algorithms. *Materials Today: Proceedings* 80:3782-3785, DOI: 10.1016/j.matpr.2021.07.382

Pandey L, Shukla S, Habibi D (2015) Electrical resistivity of sandy soil. *Géotechnique Letters* 5(3):178-185, DOI: 10.1680/jgele.15.00066

Park J, Lee KH, Kim BK, Choi H, Lee IM (2017) Predicting anomalous zone ahead of tunnel face utilizing electrical resistivity: II. Field tests. *Tunnelling and Underground Space Technology* 68:1-10, DOI: 10.1016/j.tust.2017.05.017

Park J, Lee KH, Park J, Choi H, Lee IM (2016) Predicting anomalous zone ahead of tunnel face utilizing electrical resistivity: I. Algorithm and measuring system development. *Tunnelling and Underground Space Technology* 60:141-150, DOI: 10.1016/j.tust.2016.08.007

Park J, Ryu J, Choi H, Lee IM (2018) Risky ground prediction ahead of mechanized tunnel face using electrical methods: Laboratory tests. *KSCE Journal of Civil Engineering* 22(9):3663-3675, DOI: 10.1007/s12205-018-1357-z

Rafie M, Namin FS (2015) Prediction of subsidence risk by FMEA using artificial neural network and fuzzy inference system. *International Journal of Mining Science and Technology* 25(4):655-663, DOI: 10.1016/j.ijmst.2015.05.021

Reynolds JM (2011) sAn introduction to applied and environmental geophysics – 2nd edition. John Wiley & Sons, Inc., Hoboken, NJ, USA, 61-68

Santamarina JC, Klein KA, Fam MA (2001) Soils and waves: Particulate materials behavior, characterization and process monitoring. *Journal of Soils and Sediments* 1(2):130-130, DOI: 10.1007/BF02986486

Schaeffer K, Mooney MA (2016) Examining the influence of TBM-ground interaction on electrical resistivity imaging ahead of the TBM. *Tunnelling and Underground Space Technology* 58:82-98, DOI: 10.1016/j.tust.2016.04.003

Sebbeh-Newton S, Ayawah P, Azure W, Kaba A, Ahmad F, Zainol Z, Zabidi H (2021) Towards TBM automation: On-the-fly characterization and classification of ground conditions ahead of a TBM using data-driven approach, *Applied Sciences* 11(3):1060, DOI: 10.3390/app11031060

Shang Y, Xue J, Wang S, Yang Z, Yang J (2004) A case history of Tunnel Boring Machine jamming in an inter-layer shear zone at the yellow river diversion project in China. *Engineering Geology* 71(3-4):199-211, DOI: 10.1016/S0013-7952(03)00134-0

Sharafat A, Latif K, Seo J (2021) Risk analysis of TBM tunneling projects based on generic bow-tie risk analysis approach in difficult ground conditions. *Tunnelling and Underground Space Technology* 111: 103860, DOI: 10.1016/j.tust.2021.103860

Shi M, Sun W, Zhang T, Liu Y, Wang S, Song X (2019) Geology prediction based on operation data of TBM: Comparison between deep neural network and soft computing methods. In 2019 1st International Conference on Industrial Artificial Intelligence (IAI), IEEE, DOI: 10.1109/ICIAI.2019.8850794

Telford WM, Geldart LP, Sheriff RE (1990) Applied geophysics – 2nd edition. Cambridge University Press, Cambridge, UK, 103-107

Tóth Á, Gong Q, Zhao J (2013) Case studies of TBM tunneling performance in rock–soil interface mixed ground. *Tunnelling and Underground Space Technology* 38:140-150, DOI: 10.1016/j.tust.2013.06.001

Xu ZH, Yu TF, Lin P, Wang WY, Shao RQ (2022) Integrated geochemical, mineralogical, and microstructural identification of faults in tunnels and its application to TBM jamming analysis. *Tunnelling and Underground Space Technology* 128:104650, DOI: 10.1016/j.tust.2022.104650

Yazdani-Chamzini A (2014) Proposing a new methodology based on fuzzy logic for tunnelling risk assessment. *Journal of Civil Engineering and Management* 20(1):82-94, DOI: 10.3846/13923730.2013.843583

Zhao J, Gong QM, Eisensten Z (2007) Tunnelling through a frequently changing and mixed ground: A case history in Singapore. *Tunnelling and Underground Space Technology* 22(4):388-400, DOI: 10.1016/j.tust.2006.10.002

Zhao K, Janutolo M, Barla G, Chen G (2014) 3D simulation of TBM excavation in brittle rock associated with fault zones: The brenner exploratory tunnel case. *Engineering Geology* 181:93-111, DOI: 10.1016/j.enggeo.2014.07.002