International Journal of GEOMATE, Nov., 2025 Vol.29, Issue 135, pp.1-15
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2025.135.5065
Geotechnique, Construction Materials and Environment

THE GATHERING OF A DATASET FOR TUNNELS AND THE ASSESSMENT OF PREDICTED CONSTRUCTION DELAY MODELS UTILIZING REGRESSION AND ADAPTIVE BOOSTING TECHNIQUES

* Tanawoot Kongsung [1], Nobuharu Isago [1] and Teppei Tomita [2]

[1] Faculty of Urban Environmental Sciences, Tokyo Metropolitan University, Japan; [2] Eight-Japan Engineering Consultants Inc., Japan

*Corresponding Author, Received: 04 June 2025, Revised: 02 Aug. 2025, Accepted: 04 Aug. 2025

ABSTRACT: Construction delays in tunnel projects have persisted over several decades, often resulting in significant financial and scheduling impacts. Despite extensive efforts, the root causes of these delays and effective predictive modeling approaches remain insufficiently resolved. This study aims to identify the key factors contributing to construction delays and to develop predictive models based on empirical data from tunnel projects in Japan constructed using the New Austrian Tunneling Method (NATM). The dataset includes initial and final displacements, displacement rate, categorical geological classifications, and advance rate (dependent variable), compiled from detailed design and construction records. Descriptive statistical analysis revealed a high frequency of outliers and a non-normal distribution, suggesting underlying heterogeneity in ground conditions. Regression models—both standalone and integrated with K-means clustering—were developed and further refined using Adaptive Boosting (Adaboost) algorithms. Adaboost outperformed other models, achieving higher coefficients of determination (R^2) and lower prediction errors. Feature importance and SHAP analysis confirmed final displacement as the most influential predictor of tunneling performance. The principal causes of delay were identified as insufficient geotechnical investigations and unanticipated disaster-related ground instabilities, both of which contributed to design revisions and prolonged construction periods. The study underscores the critical role of comprehensive geological surveys conducted at early project stages and demonstrates the utility of machine learning in enhancing delay prediction. These findings provide actionable insights for improving schedule reliability and risk management in future tunnel infrastructure development.

Keywords: Construction delay, Predictive Modeling, Influential predictor, Adaptive Boosting (AdaBoost), Machine Learning.

1. INTRODUCTION

Over the past several decades, the demand for tunnels has significantly increased due to rapid infrastructure and other expansions. Despite advancements, tunnel projects utilizing the New Austrian Tunneling Method (NATM) often encounter construction delays, where the actual duration exceeds the planned schedule, even in the presence and absence of major hazard reports. These delays frequently result in substantial financial losses. Unfortunately, research specifically addressing the characteristics, causes, and consequences of these delays in civil and tunnel engineering remains limited and typically focuses on expert literature reviews and basic statistical analyses. The effectiveness of existing research in addressing these delays is therefore uncertain.

Tunnel engineers have typically assessed tunnel planning based on advance rates relative to geological conditions. Anticipated progress rates have relied on engineering judgment, prior experience, reference projects, or limited statistical data. However, accurate models to predict advance rates in the early stages of tunnel projects are still lacking. A comprehensive study identifying the specific characteristics, causes, consequences, contributing factors, and predictive tools would enable early detection and mitigation of construction delays.

Doloi et al. [1] used factor analysis and linear regression to identify causes of construction delays in India, based on surveys and interviews with construction specialists. They found that a lack of commitment, poor site management, and inadequate coordination were key contributors. Similarly, Marzouk and El-Rasas [2] used frequency and severity indices in Egypt, identifying financial issues as the main cause of delays. While these studies highlight contract management-related delays, they may not directly apply to excavation delays in NATM tunnel projects.

A significant challenge in studying tunnel delays is limited access to reliable data. A comprehensive database spanning the period from 1980 to 2019 has been developed, encompassing information on hazards, tunnel dimensions, excavation techniques, and terrain types. Utilizing an edit distance search tool and fundamental statistical analysis, the study highlights the high variability associated with total delays in tunneling projects [3]. Building on this, Kongsung et al. (in press) [4] examined four NATM tunnel projects in Japan and found that

discrepancies between preliminary and actual ground conditions during construction frequently required substantial design modifications. Their study identified initial displacement as a key indicator and developed predictive delay models using both univariate and multivariate regression analyses, achieving over 80% accuracy in estimating total delay periods.

Although few studies have focused directly on NATM tunnel delays, there has been substantial work on predictive models for Tunnel Boring Machine (TBM) performance. Yagiz and Karahan [5] developed a highly accurate ($R^2 \approx 0.80$) TBM performance model using rock strength and discontinuity properties, leveraging Particle Swarm Optimization. Other studies have employed various hybrid AI techniques—such as Differential Evolution (DE) and Grey Wolf Optimizer (GWO)—to develop rate-of-penetration (ROP) models for tunnel boring machines (TBMs), achieving coefficients of determination (R^2) of 0.80 or higher [6–8]. However, these models typically require substantial datasets, which are not readily available for NATM projects.

The recent predictive Rate of Penetration (ROP) study [9] employed various variables, including Unconfined Compressive Strength (UCS), rock type, the distance between planes of weakness (DPW), and thrust force (TF), to perform descriptive statistical analysis and construct the ROP model. This was achieved by comparing robust machine learning techniques such as Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Adaptive Boosting (AdaBoost), and CatBoost (which incorporates categorical features into GB). These models demonstrated R^2 values approaching unity, indicating excellent predictive performance. The study identified UCS as the most significant feature, as confirmed by the SHAP (SHapley Additive exPlanations) methodology. This artificial intelligence (AI) and machine learning (ML) techniques have demonstrated their utility in assisting tunnel engineers in developing accurate TBM performance models. However, a key limitation remains: the lack of substantial datasets for constructing similarly precise models for NATM tunneling.

In civil engineering, AdaBoost has been successfully applied to various predictive modeling tasks. For instance, a concrete compressive strength prediction model achieved near-perfect R^2 values [10]. Simultaneously, the integrated simplicial homology global optimization method (SHGO), AdaBoost, and laboratory experiments were utilized to create a high-precision design model for the cement grouting industry [11]. Nguyen and Tran [12] also used AdaBoost and other tree-based models to predict asphaltic concrete rutting depth, confirming the method's strong predictive capability. Meanwhile, Uaisova et al. [13] utilized an artificial neural network (ANN) to predict road surface deterioration, demonstrating high precision with an R^2 value close to 1.

This study aims to bridge the gap in understanding the causes and consequences of construction delays in NATM tunnel projects and to refine predictive models by extending the dataset and analytical approach of Kongsung et al. Using data from five NATM tunnel projects in Japan, the study employs descriptive statistics, K-means clustering, regression analysis, and Adaptive Boosting (AdaBoost) to model advance rates, implemented using MATLAB (R2023b) and PyCharm. Hyperparameter tuning and SHAP (SHapley Additive exPlanations) analysis were also conducted to identify the most influential predictors of delay.

The remainder of this paper is structured as follows. Section 2 highlights the significance of the study. Section 3 presents descriptive statistics and explores various analytical approaches, including K-means clustering, regression analysis, and adaptive boosting. Section 4 explains the key variables used in the analysis. Section 5 outlines the study's methodology. Section 6 discusses the results of the analysis. Finally, Section 7 provides the conclusions and implications of the findings.

2. RESEARCH SIGNIFICANCE

This study investigates the causes and consequences of construction delays in NATM tunnels and develops predictive delay models using K-means clustering, regression analysis, and AdaBoost. Data from five NATM tunnel projects in Japan were analyzed, with descriptive statistics and hyperparameter tuning to optimize model performance. Seven statistical metrics and SHAP analysis identified key delay predictors. This data-driven approach enhances understanding and prediction of tunnel construction delays, contributing to improved planning, risk management, and future research in tunnel infrastructure.

3. DESCRIPTIVE STATISTICS, K-MEAN, REGRESSION, ADAPTIVE BOOSTING APPROACHES

This section describes the analytical approach employed in this study, incorporating descriptive statistics, K-means clustering, regression analysis, and adaptive boosting (AdaBoost) techniques.

3.1 Descriptive statistics

Descriptive statistics were used to evaluate the fundamental characteristics of the dataset, including measures of central tendency (mean, median), dispersion (standard deviation, quartiles), and distribution shape (skewness, kurtosis). Visual representations, such as histograms, violin plots, and pair-plots, were employed to investigate data distribution patterns and detect potential outliers [14–17]. Kernel density estimation (KDE) was applied to model nonparametric probability density functions for the dataset, providing insights into one-dimensional and multidimensional distribution structures [14]. Scatter plots and Pearson correlation coefficients were used to

identify potential multicollinearity issues among variables [17].

3.2 K-Means Clustering

K-means clustering, an unsupervised learning technique, was utilized to segment numeric variables based on Euclidean distances [18]. The optimal number of clusters was determined using the Elbow method, which plots the number of clusters against the within-cluster sum of squares (WCSS) to identify the point of diminishing returns [19]. The centroids of the identified clusters were subsequently used as input features in regression models, providing a hybrid approach that leverages data structure to enhance prediction.

3.3 Regression Analysis

Regression models, including univariate and multivariate forms, were developed to examine the relationship between the advance rate (AR) and independent variables [20–22]. Both linear and nonlinear (power) regression models were tested to evaluate model performance using coefficient of determination (R^2), adjusted R^2, root mean square error (RMSE), and p-values. Despite multicollinearity challenges observed in multivariate models, the optimal regression models balanced simplicity and predictive performance. Overarching mathematical expressions for multivariable linear and univariable power regressions can be articulated through the subsequent equations.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \quad (1)$$

$$y = \beta_0 (x)^{\beta_1} + \varepsilon \quad (2)$$

3.4 Adaptive Boosting (AdaBoost)

AdaBoost, an ensemble learning method, was implemented to improve predictive accuracy by combining multiple weak learners, specifically decision tree regressors, into a strong predictor [23–25]. Hyperparameter tuning was conducted using the RandomizedSearchCV function to optimize the number of weak learners, learning rates, and maximum tree depth [26–27]. Model performance was evaluated using seven statistical metrics: R^2, adjusted R^2, RMSE, mean absolute error (MAE), mean squared log error (MSLE), variance accounted for (VAF), and median absolute error (MedAE).

Feature importance was assessed using relative importance measures and SHAP (SHapley Additive exPlanations) analysis, which provides a reliable interpretation of variable contributions to model outputs [28, 29]. The SHAP results consistently identified final displacement as the most influential predictor of AR, followed by initial displacement, with other variables playing lesser roles.

The algorithm detail has been introduced as follows, proposed by Feng et al.:

The dataset has been defined as vector:

$$(\Theta) = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_m, Y_m)\} \quad (3)$$

Where X_i, Y_i (i = 1,…, m) represent the i-th sample in the training dataset, and m denotes the total number of samples. The subsequent stage involves using the training dataset to develop the weak learner G(X), after which the relative error (e_i) for all samples may be assessed using the loss function L(·). There are various loss functions, including linear, exponential, and squared loss functions. The conventional loss function (linear) is presented as follows:

$$L(\cdot) = \frac{|Y_i - G(X_i)|}{E} \quad (4)$$

E represents the maximum of $|Y_i - G(X_i)|$. Notwithstanding the subpar performance of individual weak learners, Adaboost amalgamated a sequence of weak learners (G_k (X), k = 1,2,…,N) to construct a more robust learner H(X), as delineated in Equation (5). Note that the weak learner and the integration of weak learner have been performed with the adaptation of decision tree regression (DT), proposed by [24].

$$H(X) = \nu \sum_{k=1}^{N} \left(\ln \frac{1}{\alpha_k}\right) g(X) \quad (5)$$

Where α_k is the weight of weak learner G_k (X); g(X) is the median of all the $\alpha_k G_k$ (X); $\nu \in (0,1]$ is the regularization influence (Learning rate) to mitigate the overfitting. Adaboost addresses mismatched sampling by augmenting the weight, while the weight of corrected sampling is diminished in subsequent iterations. The overall error rate and the weight of the weak learner are established by Eqa. (6) and (7) while the update weight contribution $w_{k+1,i}$ in the next training step can be examined by Eqa. (8).

$$e_k = \sum_{i=1}^{m} e_{ki} \quad (6)$$

$$\alpha_k = \frac{e_k}{1 - e_k} \quad (7)$$

$$w_{k+1,i} = \frac{w_{k,i} {\alpha_k}^{1-e_{ki}}}{\sum_{i=1}^{m} w_{k,i} {\alpha_k}^{1-e_{ki}}} \quad (8)$$

The Adaboost methodology can be delineated into four primary steps [10]: (1) Acquisition of experimental data; (2) Development of the robust learner; (3) Evaluation or validation of the learner; (4) Implementation of the learner in engineering problems. Figure 1 illustrates the operational procedure for Adaboost, with the input parameters including the number of weak learners (estimators), learning rate, and maximum tree depth.

This study integrated approach, combining descriptive statistical analysis, clustering, regression, and adaptive boosting with hyperparameter tuning technique,

provides a comprehensive, data-driven framework for understanding the underlying factors contributing to delays in NATM tunnel projects. The insights gained from this analysis support the development of robust predictive models and contribute to improved planning and risk management strategies in tunnel engineering applications.

4. VARIABLE EXPLANATIONS

Variables in this study are classified into two primary types: dependent and independent. The dependent variable is the advance rate (AR), derived from tunnel face observation records. The independent variables consist of numeric and categorical variables, enumerated as follows:

Numeric variables:

- Initial displacement (Ini): Initial measured displacement on the first observation date.
- Final displacement (Fi): Convergent or maximum measured displacement.
- Displacement rate (Dir): Ratio between final displacement and duration.

Categorical variables:

- Rock mass strength (Rs): Classified in table 1.
- Weathering/Alteration (Wa): Classified in table 2.
- Spacing of discontinuity (Sdd): Classified in table 3.
- Conditions of discontinuity (Cd): Classified in table 4.
- Effect of discontinuity perpendicular to tunnel alignment (Edp): Classified in table 5.
- Effect of discontinuity parallel to tunnel alignment (Edpa): Classified in table 6.
- Mode of occurrence (Mo): Classified in table 7.
- Patterns of crack (Pc): Classified in table 8.

The numeric variables align with measurement programs detailed in the Standard Specifications for Tunneling (2016): Mountain Tunnels. In contrast, the categorical variables are aligned with the face observation records from the same standard.

Table 1 Rock mass strength (Rs) category.

Grade class	Description
1	Very hard (VH)
2	Hard (H)
3	Fair (F)
4	Weak (W)
5	Very Weak (VW)
6	Extremely weak (EXW)

Table 2 Weathering/Alteration (Wa) category.

Grade class	Description
1	Fresh
2	Weathered along discontinuities (WAD)
3	Weathered to the rock mass core (WRC)
4	Unconsolidated (US)

Table 3 Spacing of discontinuity (Sdd) category.

Grade class	Description
1	Very Widely, D > 1m. (VW)
2	Widely, 1 m.≥ D > 0.50 m., (WI)
3	Medium, 0.50 m.≥ D > 0.20 m., (ME)
4	Closely, 0.20 m. ≥ D ≥ 0.05 m., (CL)
5	Very Closely, 0.05 m. > D, (VCL)

Table 4 Conditions of discontinuity (Cd) category.

Grade class	Description
1	Close (CLO)
2	Partially Open (POP)
3	Mostly Open (MOP)
4	Slightly Infilled Clay (SIC)
5	Largely Infilled Clay (LIC)

Table 5 Effect of discontinuity perpendicular to tunnel alignment (Edp) category.

Grade class	Description
1	Very favorable (VF)
2	Favorable (FAV)
3	Normal (NOR)
4	Unfavorable (UFAV)
5	Fair (F)

Table 6 Effect of discontinuity parallel to tunnel alignment (Edpa) category.

Grade class	Description
1	Normal (NOR)
2	Unfavorable (UFAV)
3	Fair (F)

Table 7 Mode of occurrence (Mo) category.

Grade class	Description
1	Alternation (ALT)
2	Unconformity (UCF)
3	Intrusion (INT)
4	Micro-folding (MIF)
5	Fault (FAU)
6	Others (OTH)

Table 8 Patterns of crack (Pc) category.

Grade class	Description
1	Random Squares (RAS)
2	Columns (COL)
3	Layers (LAY)
4	Fragmented Unconsolidated (FRUS)
5	Others (OTH)

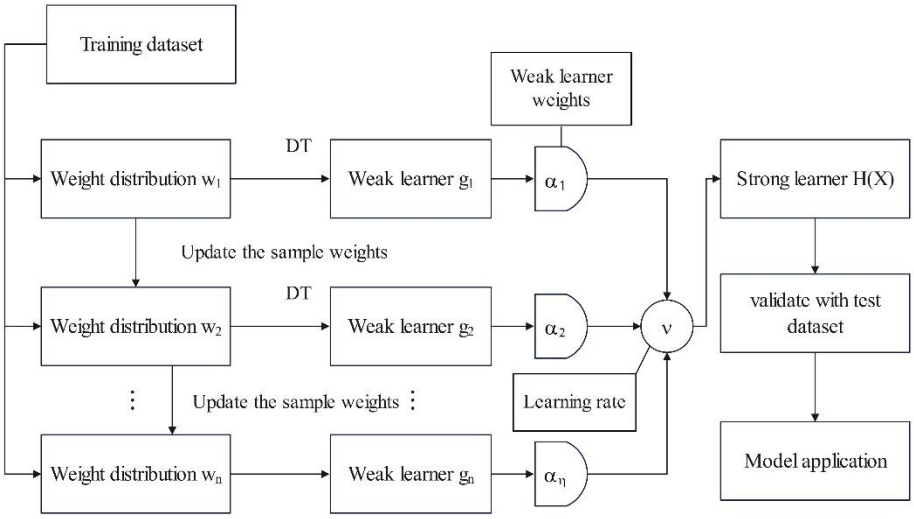

Fig. 1 Schematic of Adaptive boost the in context of regression

5. METHODOLOGY OF THIS STUDY

Figure 2 presents the comprehensive methodological framework of this study. Initially, data were gathered from design and construction records of five tunnel projects in Japan, resulting in a dataset comprising 509 entries. Descriptive statistical analysis, as detailed in Section 3, was conducted using Seaborn [30] in PyCharm to assess data quality, distributions, and outliers. The causes and consequences of construction delays were identified through examination of design and construction documents.

The dataset was partitioned into training and testing sets using the train-test split function from Scikit-learn. This process employed trial-and-error adjustments of test sizes (0.10–0.30) and pseudo-random seeds (0, 42, 100) to ensure robust model validation. Two baseline (regression approach) and Adaptive Boost models were developed: conventional univariate and multivariate regressions (linear and power forms), hybrid K-means regression, and an AdaBoost model. Factor Analysis for Mixed Data (FAMD) [31] was applied in selected scenarios to reduce dimensionality.

Hyperparameter optimization for the AdaBoost models was performed using RandomizedSearchCV, focusing on the number of weak learners, learning rate, and maximum tree depth. Final model performance was evaluated using seven statistical metrics: R^2, adjusted R^2, RMSE, MAE, MSLE, VAF, and MedAE. Feature importance and SHAP (SHapley Additive exPlanations) analyses were employed to interpret model outputs and identify the most influential variables.

This integrated methodological approach—combining descriptive statistics, hybrid clustering, regression analysis, and advanced machine learning—facilitates a comprehensive and data-driven understanding of the factors contributing to construction delays in NATM tunnel projects. The findings provide a foundation for developing predictive models to support improved planning, risk management, and decision-making in future infrastructure projects.

6. RESULTS OF ANALYSIS

6.1 Descriptive statistics

Table 9 summarizes the dataset for the tunnel projects analyzed in this study. The tunnel diameters range from 10 m to 14.65 m, with total project delays spanning from 40 to 94 days for projects A to D. Tunnel Project E experienced a substantial delay of approximately 11 years due to significant landslide issues.

6.1.1 Quantitative independent variables

The distribution of numeric variables (AR, Ini, Fi, Dir) is illustrated in the 4x4 pair plot in Figure 3. The descriptive statistics reveal that the standard deviation (SD) along the diagonal for these variables—except AR—equals or exceeds their respective means, indicating considerable variability. In particular, the SD for AR is approximately 50% of the mean. Skewness and kurtosis metrics indicate that Ini, Fi, and Dir exhibit moderate to strong positive skewness (right-skewed distributions), whereas AR shows a platykurtic distribution (negative excess kurtosis), suggesting light tails and fewer outliers. The excess kurtosis (Kurtosis-3.00) index for Ini, Fi, and Dir exhibits significantly positive values, indicating a leptokurtic distribution.

The lower diagonal of Figure 3 presents scatter plots between AR and independent variables, revealing mild to

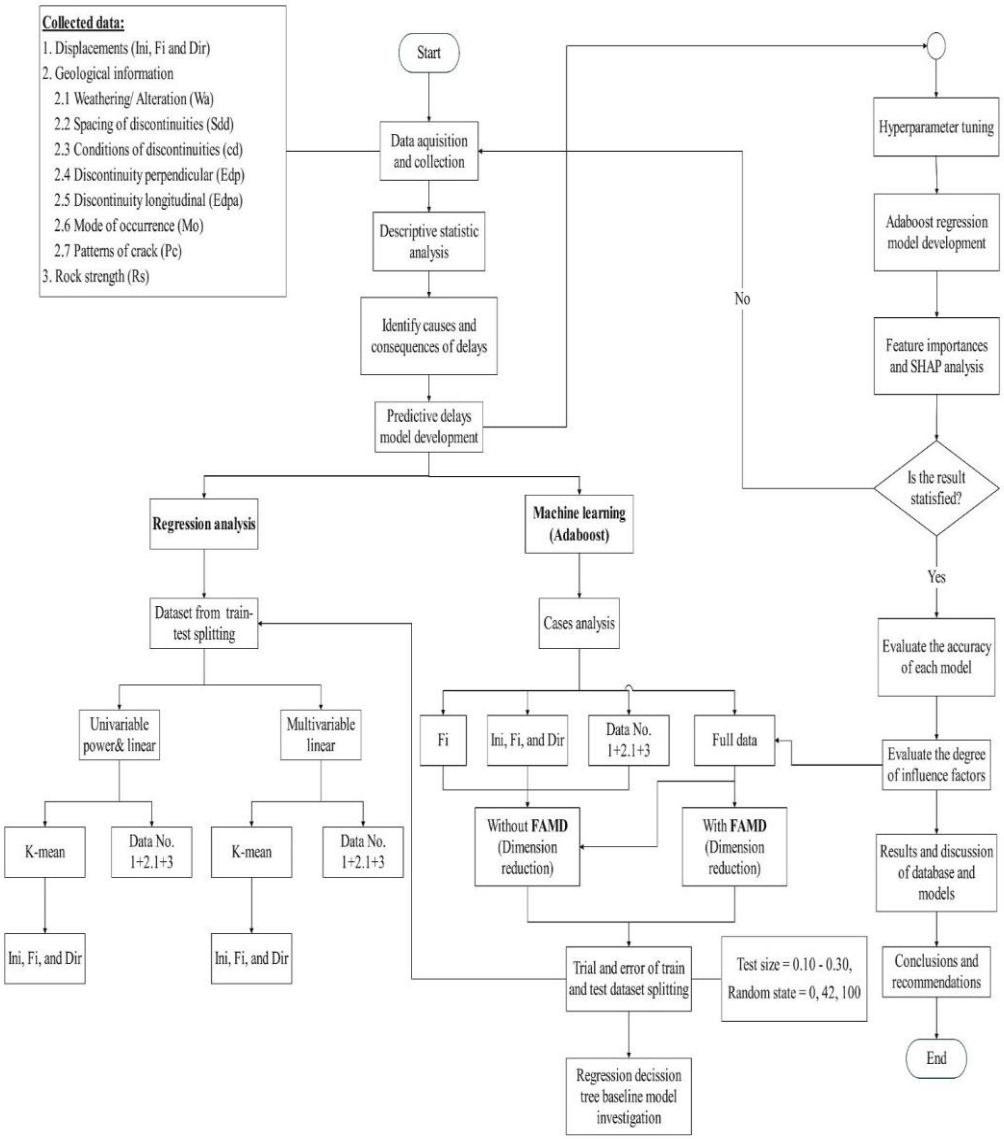

Fig. 2 Flowchart depicting the comprehensive methodology of the present study.

moderate negative Pearson correlation coefficients. Some points appear as potential outliers. Pearson correlations among independent variables, notably Ini and Dir, suggest moderate to strong positive relationships, indicating possible multicollinearity. The upper diagonal of Figure 3 (KDE contours) confirms the right-skewed nature of the data distributions, consistent with the skewness results. The densest zones for AR in relation to Ini, Fi, and Dir are approximately 1.50–2.50 m/day, corresponding to 0–10 for Ini, 0–30 for Fi, and 0–5 for Dir. The Fi-Dir plot displays a bimodal distribution, as does the Ini-Fi plot.

6.1.2 Categorical independent variables

Figure 4 displays violin plots correlating categorical variables with AR. In Figure 4a, the "H" categorization shows the highest median AR of ~3.50 m/day, though with wide variability and many outliers. The "F" and "W" categories have similar medians, while "VW" shows broader variability. In the "EXW" group, AR is lower with a narrow distribution and few outliers. There are extremely few points of a highly challenging category, which is why there is no trend in the violin plot. Figure 4b shows similar patterns in Wa categories,

with "WRC" and "WAD" showing comparable median AR values (~2.20 m/day), though "WAD" has broader dispersion. The "US" category displays lower AR with smaller distributions while the "Fresh" category shares the same issue and rationale as the very hard rock mass strength category.

Sdd violin plots (Figure 4c) reveal that the "ME" category has the highest AR (~3 m/day), with a broad distribution and numerous outliers due to varying rock types and conditions. The "CL" and "VCL" categories have similar median AR values (~1.80 m/day), but "CL" shows more outliers. Cd violin plots (Figure 4d) indicate that "MOP" has the highest AR and widest dispersion, followed by "POP". The geological explanation for these results is unclear, as factors like rock type and strata may play roles. "SIC" and "LIC" show lower AR with more outliers, especially "SIC," which has a bimodal distribution.

Edp plots (Figure 4e) show "F" with the highest AR (~2.40 m/day) and non-normal distribution, followed by "VF" with wide dispersion and extreme outliers. Other categories show similar AR medians and distribution shapes. The "Edpa" plot (Figure 4f) shows that "UFAV" has the highest AR (~2.10 m/day), while the "F" and "NOR" categories have similar, lower AR. Distributions approximate normal shapes but include significant outliers.

Mo plot (Figure 4g) reveals that "OTH" exhibits the highest median advance rate (AR) at approximately 2.60 m/day, characterized by a non-normal distribution with two distinct peaks. This is followed by "ALT", "INT", and "MIF", each with median ARs around 2.00 m/day. Among these, "INT" shows a more skewed distribution with a higher occurrence of outliers. In contrast, "UCF" and "FAU" are identified as critical modes with significantly lower ARs, narrower non-normal distributions, and moderate levels of outliers. Pc (Figure 4h) data is limited. The "OTH" category shows the highest AR (~3.30 m/day) with wide distribution and significant outliers. "FRUS" and "LAY" categories show similar abnormal distributions, with "LAY" having more outliers while the "COL" and "RAS" categories display the same issue as "VH" category of rock mass variable.

Overall, it is evident that uncertainties arising from variations in rock types, strata, and geological structures can lead to geological inconsistencies and significant variability. The dataset reflects this through substantial irregularities, including high variability, skewed distributions, and multicollinearity. These factors may adversely affect model fitting and reduce prediction accuracy in the subsequent analysis.

6.2 Identification of causes and effects of construction delay

The identification of construction delay causes and effects was primarily based on interviews with tunnel design engineers, supplemented by a detailed review of design and construction documents (Kongsung et al., in press) [4]. In tunnel project C, for example, delays were attributed to inconsistencies between the actual ground conditions and those classified during the design phase. The rock face was affected by multiple geological factors, including hydrothermal alteration and intersecting faults at approximately 45°, creating a complex structure prone to dislodgment. Although the design classification, based on drilling core results, identified the rock as tuff with high RQD and assigned it to class CII, field conditions during excavation revealed substantial fissures, prompting reclassification to class D for safety reasons. This reclassification led to significant additional labor and extended timelines for tunnel support installations.

In tunnel project E, characterized by extremely weak ground conditions and a high risk of landslides, the identification of delays was more challenging. Three principal causes of delays were identified:

1) The presence of deep, extremely weak argillaceous sandstone layers along bedding planes, compounded by low-angle reversal faults, extensive rock joints, and other discontinuities that created a high likelihood of landslides.

2) Inadequate precision in the initial geological survey programs, which failed to detect these problematic conditions, leading to major discrepancies between design-stage classifications (DI to CII) and the actual in-situ conditions (DII and E classes).

3) The need for comprehensive monitoring programs, the establishment of specialized technical committees, and the implementation of extensive temporary and

Table 9 Summary of tunnel database classified by tunnel project.

Tunnel project lists	Tunnel width (m.)	Total Delay (day)	Ground type
A	14.65	93.42	Mudstone
B	10.14	87.80	Granite
C	12.20	73.82	Mylonite
D	14.01	40.40	
E	14.64	4,130.25	Mixed sandstone and shale

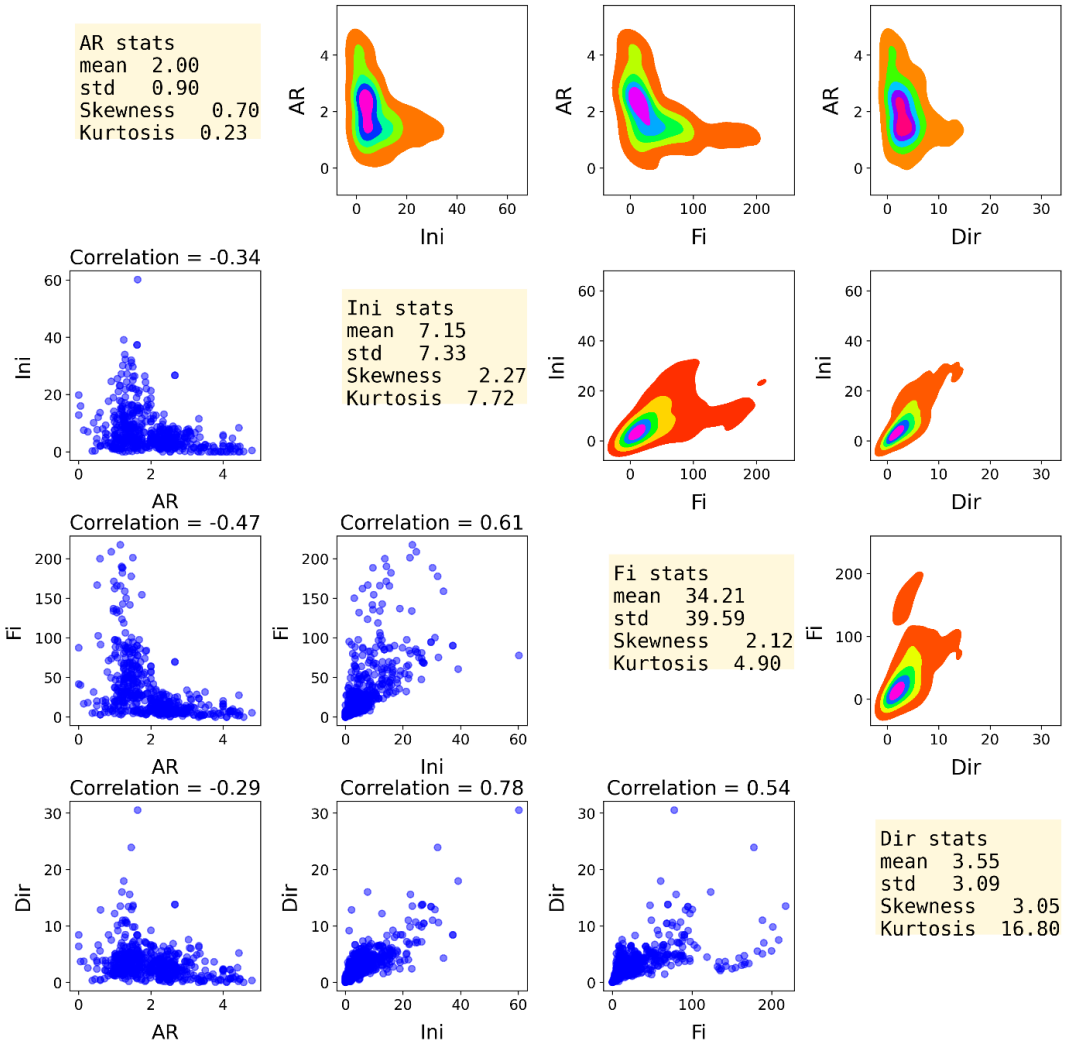

Fig. 3 Pairwise correlation, basic statistics, and distribution matrix of AR, Ini, Fi, and Dir variables.

permanent mitigation measures. These processes required significant time to ensure safety, resulting in substantial project delays.

Overall, the findings highlight the discrepancies between design-stage classifications and actual ground conditions, especially in complex geological settings, it can lead to extensive delays. Inadequate site investigations, the need for continual safety verifications, and the development of revised support designs further compounded these delays.

6.3 Conventional Regression, Hybrid Regression with K-means and Adaptive Boost

6.3.1 Dataset splitting

To mitigate data leakage, the dataset was partitioned into training and testing subsets. Following the methodology outlined in Section 5, Adaptive Boosting was employed to optimize data segmentation for model training. A trial-and-error approach was integrated with cross-validation to tune hyperparameters, including the maximum tree depth (1–10), the number of weak learners (500–2,000), and the learning rate (0.001–0.01), using the Random-Search method. These parameters are elaborated in subsequent sections.

Test sizes were systematically evaluated in increments of 0.05, and the optimal test size was determined based on seven performance metrics. For most scenarios, a test size of 0.20 and a pseudo-random parameter of 0 were identified as optimal, except for the full dataset without FAMD, where the test size was 0.25 and a pseudo-random parameter of 42 was used. Consequently, the final datasets comprised 381 to 407 training samples and 128 to 102 testing samples. These datasets were used to develop curve-fitting models employing both conventional and hybrid regression techniques. This structured partitioning ensures robust model validation while minimizing data leakage.

6.3.2 Conventional and hybrid regressions

A total of fifteen scenarios for conventional univariate and multivariate regression analyses were explored using independent variables Ini, Fi, Dir, Rs, and Wa, with 509 data points. Univariate regressions were evaluated using linear and power models, while multivariate regressions employed linear functions. Three statistical metrics—R^2,

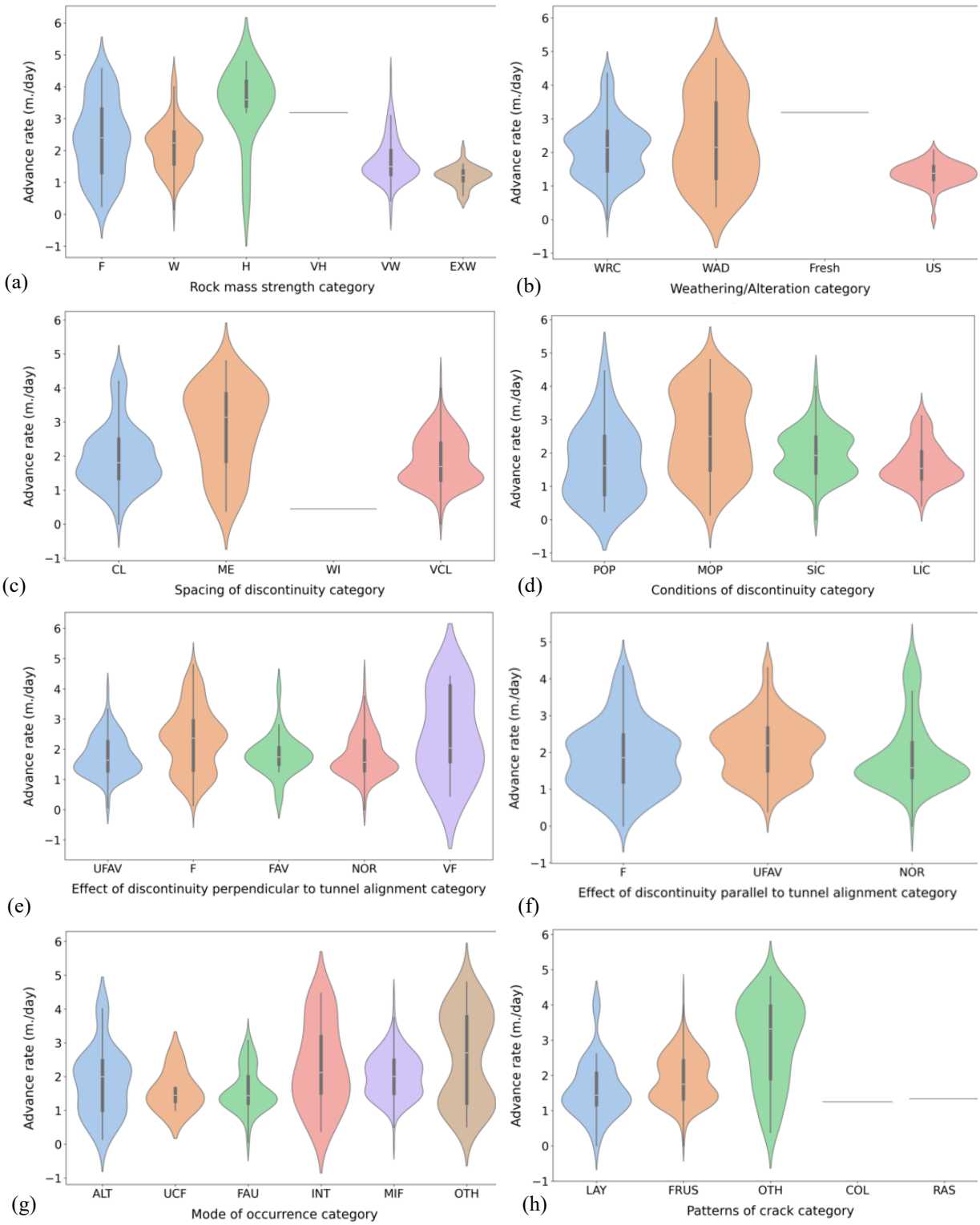

Fig. 4 Violin Plot Comparison of Advance Rate Across Different Geological and Geotechnical Categories: (a) Rock mass strength category; (b) Weathering/alteration category; (c) Spacing of discontinuity category; (d) Conditions of discontinuity category; (e) Effect of discontinuity perpendicular to tunnel axis category; (f) Effect of discontinuity parallel to tunnel axis category; (g) Mode of occurrence category; (h) Patterns of crack category.

adjusted R^2, and RMSE—alongside p-values, were used to evaluate model performance. Criteria for model acceptability included R^2 and adjusted $R^2 \geq 0.28$, RMSE $\leq$ 0.751, and p-value ≤ 0.05. While all models achieved significant p-values, R^2 and adjusted R^2 remained below.

The author enumerated the regression equations according to the subsequent criteria. $R^2 \geq 0.28$, and RMSE ≤ 0.751; the equations are provided below:

AR = 2.263 - 0.00693(Ini) - 0.01(Fi) +0.40(EXW) +0.1196(W)+0.172(F)+1.356(H)+0.963(VH), $\boldsymbol{R^2}$ = 0.285, $\boldsymbol{R^2}$(adj.) = 0.272, and RMSE= 0.751 (10)

AR = 2.232 - 0.00972(Ini) - 0.01(Fi) +0.012(DIR) +0.379(EXW)+0.132(W)+0.192(F)+1.382(H)+0.98(VH), $\boldsymbol{R^2}$ = 0.285, $\boldsymbol{R^2}$(adj.) = 0.271, and RMSE= 0.751 (11)

AR = 2.237- 0.00636 (Ini) - 0.009(Fi) +0.006(DIR) +0.315(EXW)+0.137(W)+0.35 (F)+1.88(H)+0.977(VH)-0.178(US)-0.509(WAD), $\boldsymbol{R^2}$ = 0.304, $\boldsymbol{R^2}$(adj.) = 0.286, and RMSE= 0.743 (12)

Although Equation (11) and (12) achieved slightly higher R^2 and adjusted R^2, it included six and eight additional variables compared to Equation (9), with negligible improvement in RMSE. Therefore, from a statistical standpoint, Equation (9) was deemed the most suitable for standard regression analysis. Validation of these baseline models using testing datasets (as described in Section 6.3.1) is discussed in subsequent sections.

For the hybrid regression approach, K-means clustering was applied to numeric variables (Ini, Fi, Dir), classifying the dataset into 10 clusters based on the Elbow method. The cluster centers were used as inputs for univariate and multivariate regressions. Results demonstrated that univariate power regression models outperformed linear and multivariate regressions, with R^2 values marginally below 0.56. In contrast, multivariate models showed only minor improvements and often had p-values exceeding 0.05, reflecting strong multicollinearity. The optimal fit for this methodology is seen in the subsequent equations:

AR = 3.662 $(\text{Ini})^{-0.361}$, $\boldsymbol{R^2}$ = 0.539, $\boldsymbol{R^2}$ (adj.) =0.481, RMSE= 0.606 (13)

AR = 5.263 $(\text{Fi})^{-0.3}$, $\boldsymbol{R^2}$ = 0.558, $\boldsymbol{R^2}$ (adj.) =0.503, RMSE= 0.593 (14)

AR = 3.123 $(\text{Dir})^{-0.423}$, $\boldsymbol{R^2}$ = 0.523, $\boldsymbol{R^2}$ (adj.) =0.463, RMSE= 0.616 (15)

Among these, Equation (14) was identified as the best fit for the hybrid regression approach, exhibiting the highest R^2 and lowest RMSE. These results highlight the superior performance of univariate power regressions within the hybrid framework and provide robust baseline models for comparative analysis with Adaptive Boosting in subsequent sections.

6.3.3 Adaptive boost

The Adaptive Boosting (AdaBoost) analysis encompassed five scenarios: Fi alone, displacements, displacements-Rs-Wa, all variables, and all variables with Factor Analysis of Mixed Data (FAMD). FAMD was employed to reduce dataset dimensionality, yielding 25 principal components and accounting for a 3% variance loss. Hyperparameter optimization was conducted using a trial-and-error approach integrated with 10-fold cross-validation and RandomizedSearchCV. The final configurations are summarized in Table 10, showing a consistent maximum tree depth of three across all scenarios, while the number of weak learners ranged from 500 to 2,000 and learning rates from 0.001 to 0.01.

Model fitting was assessed using seven performance metrics (Table 11). Figure 5 depicts the relationship between measured and predicted AR for all AdaBoost cases. Among these, the Fi and displacements-Rs-Wa scenarios achieved the highest R^2 values (~0.43). Models incorporating all variables—both with and without FAMD—exhibited marginal improvements in error terms but suffered from overfitting and multicollinearity,

Table 10 summary results of hyperparameter tuning for all adaptive boost cases.

Lists of control parameter	Adaptive boost cases analysis				
	Fi	Displacements	Displacements-Rs-Wa	All variables	All variables with FAMD
Maximum depth of tree	3	3	3	3	3
Maximum number of weak learners	1,982	1,982	500	2,000	1,982
Learning rate	0.001999	0.001999	0.001	0.01	0.001999

Table 11 summary results of hyperparameter tuning for all adaptive boost cases.

Lists of evaluated metrics	Adaptive boost cases analysis				
	Fi	Displacements	Displacements-Rs-Wa	All variables	All variables with FAMD
$\boldsymbol{R^2}$	0.426	0.385	0.420	0.438	0.374
$\boldsymbol{R^2}$ (adj.)	0.420	0.366	0.335	0.170	0.158
MSE	0.547	0.586	0.552	0.407	0.505
MAE	0.569	0.594	0.571	0.465	0.513
MSLE	0.0695	0.0690	0.0659	0.0464	0.0572
MedAE	0.427	0.424	0.445	0.354	0.327
VAF (%)	42.82	38.49	42.04	44.74	37.85

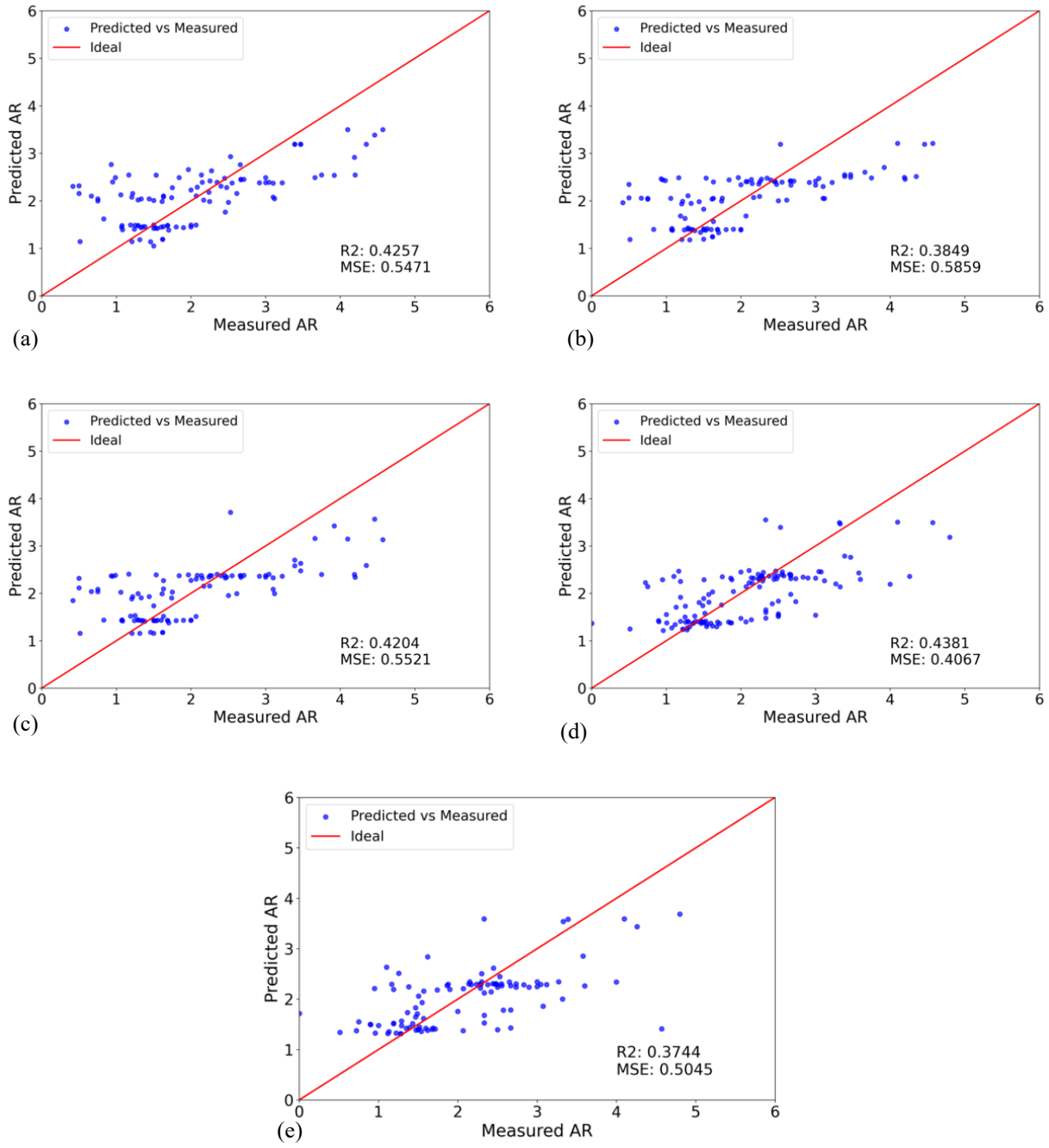

Fig. 5 Predicted vs measured of AR scatter plots with R^2 and MSE for Adaptive boost models: (a) Fi case; (b) Displacements case; (c) Displacements-Rs-Wa case; (d) All variable cases without FAMD; (e) All variable cases with FAMD.

rendering them less robust. The impact of FAMD on overall model performance was negligible, as the reduction in variable count did not substantially affect predictive accuracy.

Comparative analysis of optimal models indicates that the Fi and displacements–Rs–Wa scenarios consistently yielded superior performance compared to other combinations. However, the inclusion of additional variables in the latter scenario resulted in only marginal improvements in correlation and minimal reduction in fitting error (prone to overfitting) relative to the Fi-only model. This indicates that the additional variables do not significantly enhance the model's predictive capability. Therefore, the Fi scenario was identified as the optimal predictor configuration within the AdaBoost models.

6.4 Assessment and validation of all predictive models

The evaluation of baseline and AdaBoost models was based on the correlation between predicted and

measured AR, depicted in Figures 6 and 7. Adaptive Boost models generally outperformed conventional and hybrid baseline models, except for the Fi scenario with the baseline power regression, which exhibited comparable performance. This is likely due to multicollinearity challenges inherent in baseline models.

Despite the overall robustness of AdaBoost models, the correlation between observed and predicted AR values remained moderate. This can be attributed to dataset irregularities and several outliers as outlined in the descriptive statistics (Section 6.1). These issues highlight known limitations of adaptive boosting, particularly its sensitivity to noisy data, outliers, and class imbalance, as well as its tendency to overfit [32]. Additionally, the dataset exhibited fluctuations and instances of inaccurate sampling, for example, cases where similar independent variable values corresponded to medium to large differences in the dependent variable (AR). Such inconsistencies may have triggered multiple augmentation and other processes within the AdaBoost algorithm, ultimately reducing the accuracy of the model's fit.

The assessment also extended to overall delay periods, revealing that discrepancies in spatial stationing of predicted and measured AR can produce identical total delay sums despite underlying misalignments. Thus, correlation-based validation between measured and predicted AR is more reliable than total delay-based comparisons. In summary, Fi emerged as the most influential factor in predicting AR, with AdaBoost offering the highest overall predictive performance while recognizing its limitations in the presence of data noise and multicollinearity.

6.5 Investigation of main contributed variables of Adaptive boosting

The relative importance of variables in Adaptive Boosting models was assessed, excluding the Fi scenario. Figure 8 presents the relative importance of variables for the displacements-Rs-Wa scenario. Fi exhibited the highest weight (~0.70), followed by Ini (~0.15) and Dir (~0.10). Other variables had weights below 0.05, reflecting minor contributions. Similar patterns were observed in other Adaptive Boosting scenarios, albeit with slight variations in variable weights. These findings corroborate the observations from Section 6.3.3, where the addition of variables did not significantly improve model fitting, confirming Fi as the most influential predictor in construction delay models.

Additionally, SHAP (SHapley Additive exPlanations), based on game theory, was employed to provide a more detailed interpretation of the "black box" Adaptive Boosting models. Figure 9a presents a bar plot of the mean SHAP value contributions for each variable, revealing that the three primary contributing features are consistent with the results of the previous feature importance analysis. Notably, Fi demonstrates the highest contribution (~0.44), followed by Ini (~0.05), indicating that Fi is the dominant influencing factor in the model. However, the interpretability of the Adaptive Boosting model should be carefully examined and validated through engineering judgment to ensure the robustness and practical relevance of its predictions.

Figure 9b presents a comprehensive bee swarm plot for the displacements–Rs–Wa scenario, further confirming that Fi is the third most influential predictor of AR. The SHAP values on the x-axis indicate both the magnitude and direction of each variable's contribution to the predicted AR, with negative values suggesting a reduction in AR and positive values indicating an increase. The color gradient on the y-axis represents the feature value distribution, where red denotes higher values and blue indicates lower values.

For Fi, higher feature values are generally associated with negative SHAP values—implying a reduction in predicted

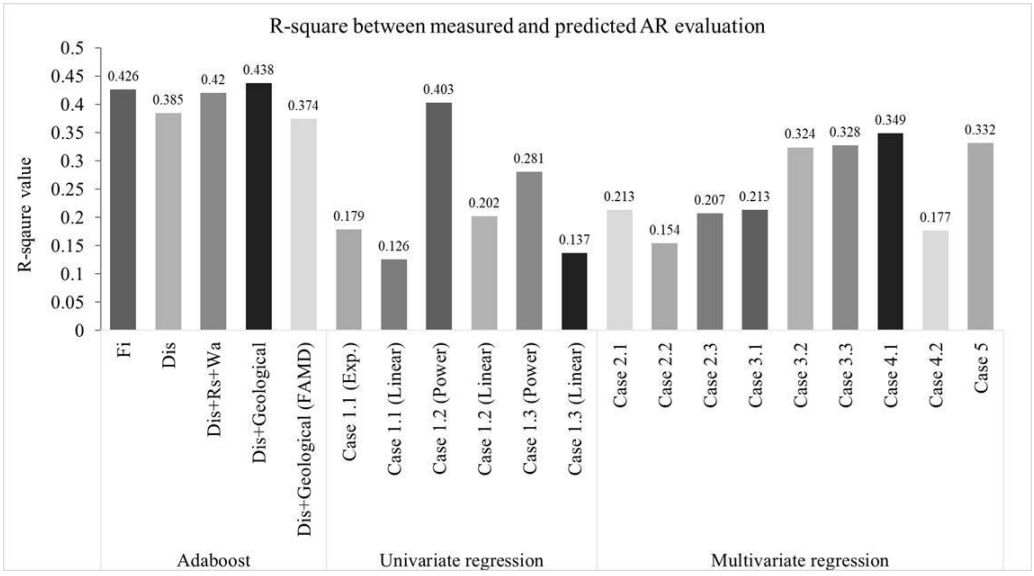

Fig. 6 Comparison of R-square values between measured and predicted AR evaluation for conventional baseline models and adaptive boost models.

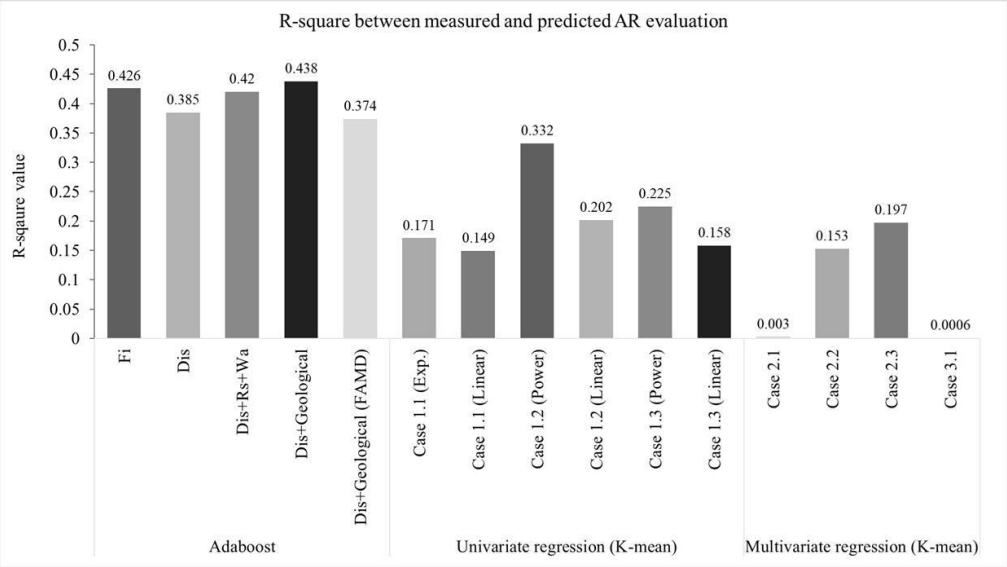

Fig. 7 Comparison of R-square values between measured and predicted AR evaluation for hybrid baseline models and adaptive boost models.

AR—whereas lower Fi values tend to correspond with an increase in AR. In contrast, Ini exhibits a predominantly positive influence on AR, although some lower values of Ini contribute insignificantly or inconsistently. These patterns suggest that Fi is a more reliable and interpretable predictor from an engineering standpoint, while the role of Ini remains less conclusive.

The lower variable Dir exhibits a negative effect on AR, which contrasts with the positive influence of Fi and aligns with practical engineering expectations. Some data points from the "H" and "F" categories of Rs exhibit atypical one-sided effects—either predominantly positive or negative. However, the underlying pattern remains unclear. In contrast, the overall contribution of geological features to AR remains limited and largely insignificant, as shown in Figure 9b. This scenario presents challenges for validation through engineering judgment, as the observed patterns are still ambiguous and lack clear interpretability.

In conclusion, both the relative importance and SHAP analyses consistently identify Fi as the dominant predictor in the Adaptive Boosting models for tunnel excavation delays, reinforcing its significance as a key variable in predictive delay modeling. Moreover, this finding aligns with the results presented in Section 6.3.3, further validating that the Fi-based Adaptive Boosting model provides the most reliable and engineering-consistent predictions.

7. CONCLUSION

This study comprehensively investigated the causes, consequences, and predictive modeling of construction delays in NATM tunnel projects in Japan. A robust dataset comprising 509 data points—including displacement geological factors, and the advance rate (AR) was analyzed. Statistical analysis revealed significant heterogeneity and outlier presence, suggesting complex ground conditions that complicate predictive efforts.

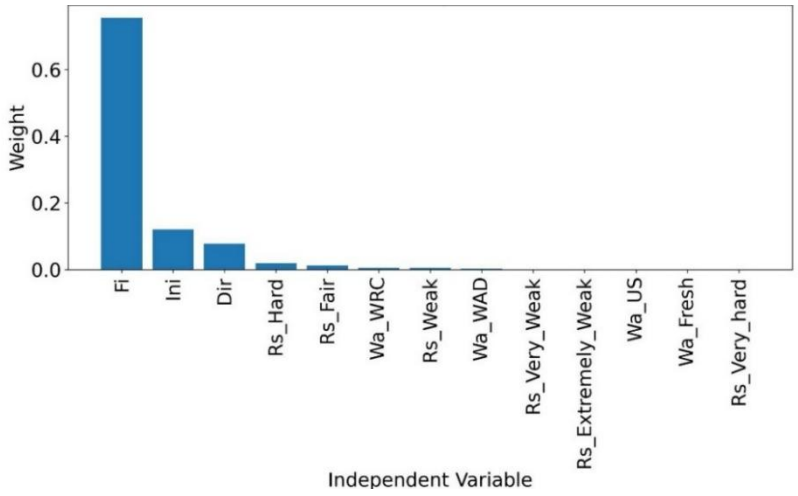

Fig. 8 Analysis of relative variable relevance for displacements in Ra-Wa scenarios (Adaboost).

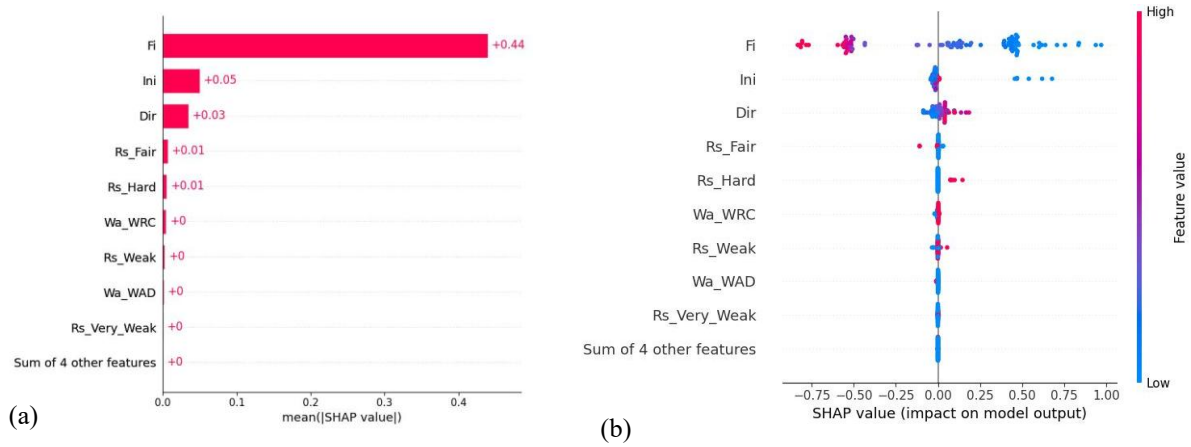

Fig. 9 Schematics developed from the SHAP methodology of displacements in Ra-Wa cases employing Adaboost. (a) Bar plot; (b) Bee swarm plot

Conventional and hybrid regression models were developed as baseline predictors for AR, while Adaptive Boosting was employed for enhanced predictive performance. Adaptive Boosting consistently outperformed other models, achieving higher R^2 values and lower error metrics, though it remains sensitive to outliers, variances and imbalanced datasets. Feature importance and SHAP analyses identified final displacement (Fi) as the most influential variable, underscoring its critical role in tunnel excavation performance.

The principal causes of construction delays were found to be discrepancies between actual and design stage ground classifications, especially in weak ground conditions complicated by complex geology and landslide hazards. In such cases, extensive mitigation efforts, including supplementary investigations and continuous monitoring, are necessary but contribute to extended project timelines.

This study acknowledges key limitations and suggests directions for practical implementation and future research. The variability in rock formations and mineral compositions can result in differing ground behaviors, which constrains the direct generalizability of the developed models. However, to address uncertainties related to geological conditions, practitioner experience, and standard practices, the proposed methodology should be re-applied to diverse and site-specific datasets. The application of machine learning in tunneling contexts must be conducted with caution—models should be interpreted and validated in close collaboration with tunnel or mining engineers to ensure practical reliability. Furthermore, enhancing geological investigations during the early stages of a project and incorporating final displacement considerations into standard tunnel design guidelines may substantially reduce associated risks.

Finally, the study outlines several directions for future research. Future applications should consider utilizing exclusive or expanded datasets or adapt models to account for local geological variations. Additionally, upcoming studies should focus on enhancing predictive accuracy by exploring advanced machine learning boosting techniques, such as Gradient Boosting or XGBoost, while also addressing issues related to dataset imbalances and the effects of outliers.

8. ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan, for providing access to the tunnel database.

9. REFERENCES

[1] Doloi H., Sawhney A., Iyer K.C., and Rentala S., Analysing factors affecting delays in Indian construction projects. International Journal of Project Management, Vol. 30, Issue 4, 2012, pp.479–489.

[2] Marzouk M.M., and El-Rasas T.I., Analyzing delay causes in Egyptian construction projects. Journal of Advanced Research, Vol. 5, Issue 1, 2014, pp.49–55.

[3] Paraskevopoulou C., Dallavalle M., Konstantis S., Spyridis P., and Benardos A., Assessing the failure potential of tunnels and the impacts on cost overruns and project delays. Tunnelling and Underground Space Technology, Vol. 123, 2022, pp.1-13.

[4] Kongsung T., Kawata K., Tomita T., and Isago N., The Utilization of Multivariable Regression Analysis for Construction Delay in Road Tunnel Projects. The Fourth International Conference on Sustainable Civil Engineering and Architecture 2025, Equatorial Ho Chi Minh City, Vietnam, Springer Nature, 2025, (in press).

[5] Yagiz S., and Karahan H., Prediction of hard rock TBM penetration rate using particle swarm optimization. International Journal of Rock Mechanics and Mining Sciences, Vol. 48, Issue 3, 2011, pp.427–433.

[6] Mahdevari S., Shahriar K., Yagiz S., and Akbarpour Shirazi M., A support vector regression model for predicting tunnel boring machine penetration rates. International Journal of Rock Mechanics and Mining Sciences, Vol. 72, 2014, pp. 214–229.

[7] Yagiz S., and Karahan H., Application of various optimization techniques and comparison of their performances for predicting TBM penetration rate

in rock mass. International Journal of Rock Mechanics and Mining Sciences, Vol. 80, 2015, pp.308–315.
[8] Bardhan A., Kardani N., GuhaRay A., Burman A., Samui P., and Zhang Y., Hybrid ensemble soft computing approach for predicting penetration rate of tunnel boring machine in a rock environment. Journal of Rock Mechanics and Geotechnical Engineering, Vol. 13, Issue 6, 2021, pp.1398–1412.
[9] Ghorbani E., and Yagiz S., Estimating the penetration rate of tunnel boring machines via gradient boosting algorithms. Engineering Applications of Artificial Intelligence, Vol. 136, 2024, pp.1-15.
[10] Feng D. C., Liu Z.T., Wang X. D., Chen Y., Chang J. Q., Wei D. F., and Jiang Z. M., Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Construction and Building Materials, Vol. 230, 2020, pp.1-11.
[11] Ren J., Zhao H., Zhang L., Zhao Z., Xu Y., Cheng Y., Wang M., Chen J., and Wang J., Design optimization of cement grouting material based on adaptive boosting algorithm and simplicial homology global optimization. Journal of Building Engineering, Vol. 49, 2022, pp.1-16.
[12] Nguyen H.L., and Tran V.Q., Data-driven approach for investigating and predicting rutting depth of asphalt concrete containing reclaimed asphalt pavement. Construction and Building Materials, Vol. 377, 2023, pp.1-19.
[13] Uaisova M., Zharlykassov B., Aldasheva D., Artykbayeva A., and Radchenko P., The Use of ANN and Machine Learning Algorithms to Predict Road Surface Deterioration. International Journal of GEOMATE, Vol. 27, Issue 121, 2024, pp.136-143.
[14] Fávero L.P., and Belfiore P., Univariate Descriptive Statistics. In: Data Science for Business and Decision Making, Elsevier, 2019, pp.21–91.
[15] Wand M.P., and Jones M.C., Kernel Smoothing. Springer, 1995, pp.1-212.
[16] Hintze J.L., and Nelson R.D., Violin Plots: A Box Plot-Density Trace Synergism. The American Statistician, Vol. 52, Issue 2, 1998, pp.181–184.
[17] Benesty J., Chen J., Huang Y., and Cohen I., Pearson Correlation Coefficient. In: Noise Reduction in Speech Processing, Springer, 2009, pp. 1–4.
[18] Hood S.B., Cracknell M.J., and Gazley M.F., Linking protolith rocks to altered equivalents by combining unsupervised and supervised machine learning. Journal of Geochemical Exploration, Vol. 186, 2018, pp. 270–280.
[19] Thorndike R.L., Who belongs in the family? Psychometrika, Vol. 18, Issue 4, 1953, pp. 267–276.
[20] El-Abbasy M.S., Senouci A., Zayed T., Mirahadi F., and Parvizsedghy L., Condition Prediction Models for Oil and Gas Pipelines Using Regression Analysis. Journal of Construction Engineering and Management, Vol. 140, Issue 6, 2014, pp. 1-17.
[21] Rong G., Wang X., Xu H., and Xiao B., Multifactor Regression Analysis for Predicting Embankment Dam Breaching Parameters. Journal of Hydraulic Engineering, Vol. 146, Issue 2, 2020, pp.1-15.
[22] Song X., Li Y., Mao J., Zhou K., Yang D., Deng Z., and Wang Q., A study of the thermal performance prediction of embedded pipes in protection engineering based on multiple regression analysis. Geothermics, Vol. 112, 2023, pp. 1-14.
[23] Freund Y., and Schapire R.E., A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, Vol. 55, Issue 1, 1997, pp.119–139.
[24] Drucker H., Improving Regressors using Boosting Techniques. Proceedings of the Fourteenth International Conference on Machine Learning, Morgan Kaufmann Inc., 1997, pp.1-9.
[25] Berk R.A., Statistical Learning from a Regression Perspective. Springer, 2008, pp.1-358.
[26] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Müller A., Nothman J., Louppe G., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., and Duchesnay É., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, Vol.12, 2011, pp.2825-2830.
[27] Bergstra J., and Bengio Y., Random search for hyper-parameter optimization. Journal Machine Learning Research, Vol. 13, 2012, pp.281–305.
[28] Friedman J.H., Greedy function approximation: A gradient boosting machine. The Annals of Statistics, Vol. 29, Issue 5, 2001, pp.1189-1232.
[29] Lundberg S., and Lee S.-I., A Unified Approach to Interpreting Model Predictions. The 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, Curran Associates, Inc., 2017, pp.1-10.
[30] Waskom M., seaborn: statistical data visualization. Journal of Open-Source Software, Vol. 6, Issue 60, 2021, pp. 1-4.
[31] Halford M., Prince. https://github.com/MaxHalford/prince. Accessed May 27, 2025.
[32] Chakraborty D., Ghosh A., and Saha S., A survey on Internet-of-Thing applications using electroencephalogram. Emergence of Pharmaceutical Industry Growth with Industrial IoT Approach, Elsevier, 2020, pp. 21–47.