Geomechanics and Engineering, *Vol. 37, No. 3 (2024) 253-262*
https://doi.org/10.12989/gae.2024.37.3.253

Prediction of dynamic soil properties coupled with machine learning algorithms

Dae-Hong Min[a] and Hyung-Koo Yoon*

Department of Construction and Disaster Prevention Engineering, Daejeon University, Daejeon 34520, Republic of Korea

(*Received March 9, 2024, Revised April 16, 2024, Accepted April 19, 2024*)

Abstract. Dynamic properties are pivotal in soil analysis, yet their experimental determination is hampered by complex methodologies and the need for costly equipment. This study aims to predict dynamic soil properties using static properties that are relatively easier to obtain, employing machine learning techniques. The static properties considered include soil cohesion, friction angle, water content, specific gravity, and compressional strength. In contrast, the dynamic properties of interest are the velocities of compressional and shear waves. Data for this study are sourced from 26 boreholes, as detailed in a geotechnical investigation report database, comprising a total of 130 data points. An importance analysis, grounded in the random forest algorithm, is conducted to evaluate the significance of each dynamic property. This analysis informs the prediction of dynamic properties, prioritizing those static properties identified as most influential. The efficacy of these predictions is quantified using the coefficient of determination, which indicated exceptionally high reliability, with values reaching 0.99 in both training and testing phases when all input properties are considered. The conventional method is used for predicting dynamic properties through Standard Penetration Test (SPT) and compared the outcomes with this technique. The error ratio has decreased by approximately 0.95, thereby validating its reliability. This research marks a significant advancement in the indirect estimation of the relationship between static and dynamic soil properties through the application of machine learning techniques.

Keywords: deep neural network; dynamic soil property; important value; random forest; static soil property

1. Introduction

Evaluating soil stability under dynamic loading conditions, such as those induced by earthquakes, wind, or machinery, is essential, requiring the estimation of the soil's dynamic properties (Gomaa *et al.* 2023, Nguyen *et al.* 2023, Rajabian 2023). The constants of soil determined under static loading conditions can differ significantly from those under dynamic loading, underscoring the necessity of obtaining results from experiments conducted under dynamic conditions (Yazdandoust *et al.* 2023, Zhang *et al.* 2023). The fundamental distinction between static and dynamic loading is illustrated in Figure 1, reflecting the variations in conditions due to the strain states of the soil medium (Benemaran and Esmaeili-Falak 2023). This figure demonstrates the normalized shear modulus plotted against shear strain values, with a clear demarcation between small strain and large strain regions, using a 0.1% shear strain threshold as a criterion. To grasp the dynamic behavior in the small strain domain, it's imperative to emulate conditions pertinent to the strain sector prior to result derivation. For inferring soil behavior in this domain, field experiments typically leverage seismic surveying techniques like refraction and reflection to obtain the velocities of compressional and shear waves within the soil (Park *et al.* 2022, Do *et al.* 2023). In the lab, the resonant column test is frequently employed, capable of determining varying elasticity moduli across shear strain zones. Nonetheless, this approach is limited by the substantial costs and time investment required for experimental setup. Field experiments, which primarily measure seismic wave velocities, face difficulties in translating these values into the soil's modulus of elasticity, a critical design parameter, without knowledge of the soil's density. Moreover, replicating field condition samples in lab settings poses challenges. In addition to seismic survey techniques, dynamic properties are commonly predicted using empirical formulas based on N-values measured by the Standard Penetration Test (Anbazhagan and Sitharam 2013, Ataee *et al.* 2019, Heo *et al.* 2023). However, this research relies solely on N-values from specific areas, which limits its applicability to other regions and diminishes its reliability due to the one-dimensional correlation between N-values and dynamic properties. Consequently, there is a significant emphasis on methods that predict dynamic property values from static conditions using empirical formulas.

Machine learning methods establish the mathematical relationship between input and output values, enabling the derivation of corresponding outputs for new inputs (Min *et al.* 2023, Park *et al.* 2023). These methods are utilized across various fields to predict outcomes from input values (Lee *et al.* 2022, Lee and Choi 2023). In geotechnical engineering, ML methods have been applied to derive safety factors from factors associated with landslide occurrences. Research has been conducted to assess the distribution of safety factors in new areas using established architectures and hyperparameters (Min and Yoon 2021).

*Corresponding author, Associate Professor
E-mail: hyungkoo@dju.ac.kr
[a]Ph.D. Candidate

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

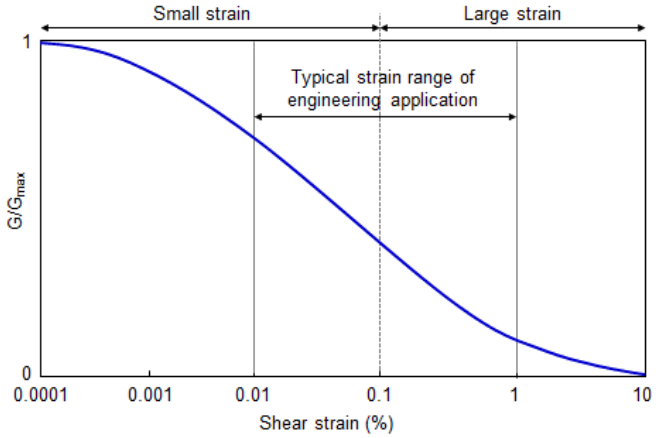


Fig. 1 Distribution of shear modulus according to shear strain

Additionally, Hong *et al.* (2022) employed machine learning to predict soil constants from input values obtained through time domain reflectometry, a technique frequently used in geotechnical engineering. The importance of this work lies in identifying the significance of each input factor, which is crucial for accurate predictions. Many studies in geotechnical engineering have also utilized machine learning techniques. This paper aims to leverage machine learning algorithms to estimate dynamic properties, drawing on the methodologies and applicability of previous research. Herein, static properties are used as input values, with the ultimate output properties defined as the velocities of compressional and shear waves, components of seismic wave speeds. The goal of this paper is to explore whether properties acquired under static loads can effectively predict dynamic loading factors through machine learning techniques.

This paper starts with an exploration of the background theories behind the techniques utilized in this research, followed by a detailed description of the methodologies adopted for data construction. Utilizing the assembled data, we categorized the input and output data and conducted an analysis on the significance of each input data to pre-validate its impact prior to making predictions on the output data. Ultimately, the study verified the reliability of our approach by comparing the dynamic properties predicted through machine learning techniques against actual values, with adjustments made to the number of influencing factors.

2. Background theory

In this research, we employed two machine learning techniques: Random Forest (RF) to assess the significance of input properties, and Deep Neural Network (DNN) to execute predictions using these input properties. The features of each algorithm are detailed as follows.

2.1 Important value

The Random Forest algorithm operates on a decision tree-based framework, deriving desired outcomes through ensemble learning (Kim and Yoon 2023). Although RF is commonly applied in classification tasks, it was employed in this study specifically to assess the significance of input property values (Yoon *et al.* 2022). The variables of input properties are incorporated into the decision trees at random, and the importance of these input properties is determined by comparing the difference between the output values generated from the selected input properties (Randomn) and the original data (Rawn). The importance score is calculated as shown in Eq. (1), where n denotes the count of data points (Yu *et al.* 2021).

$$IS_n = \frac{\frac{1}{n}\sum_1^n |Raw_n - Random_n|}{\sqrt{\frac{1}{n-1}\sum_1^n \left[\left(|Raw_n - Random_n| \right) - \left(\frac{1}{n}\sum_1^n |Raw_n - RN_n| \right) \right]^2}} \quad (1)$$

2.2 Artificial neural network-based prediction model

The Deep Neural Network (DNN) is a widely utilized algorithm in regression analysis among machine learning techniques, improving reliability by adding nodes between the input and output values (Fereidooni and Karimi 2023, Ihsan *et al.* 2023, Lee *et al.* 2023). The number of nodes is variably determined, taking into account the user's goals and data characteristics, with each node allocated appropriate weight and bias values for predicting the output from the input values. The precision of these weights and biases is critical to the reliability of the outcome. Should there be a substantial difference between the predicted and actual output values, the weights and biases are adjusted via the backpropagation method. This adjustment process continues until the predictions closely align with the user-defined error rate, incorporating various optimization methods to achieve values near this target rate. Given that DNN computes the output value (O) by applying the input value (X) along with weights (w) and bias (b) set at different nodes, the mathematical model is concisely represented as a linear equation, indicated in Eq. (2).

$$O_{DNN} = \sigma \left(\sum_n X \cdot w_n + b_n \right) \quad (2)$$

Fig. 2 Description of test sites

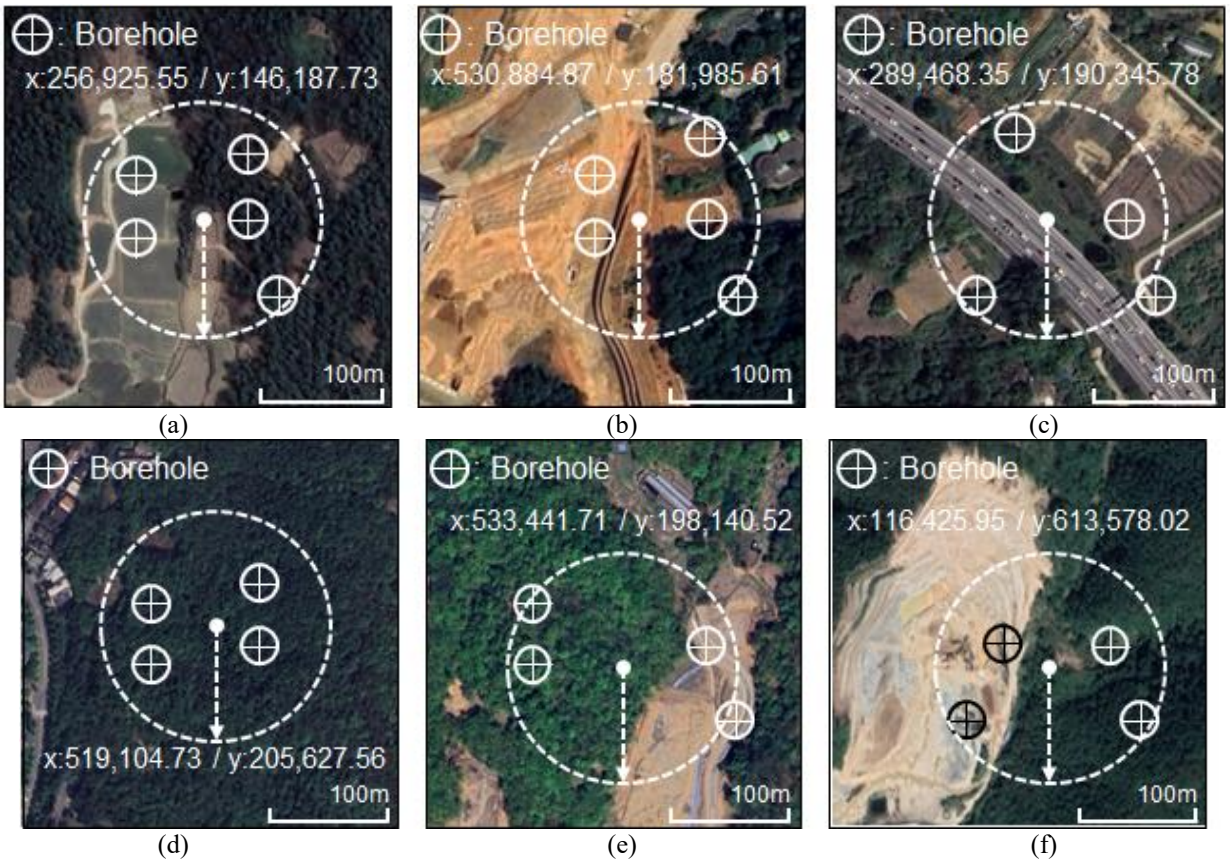


Fig. 3 Detail description of test sites: (a) Muan City in Jeollanam-do, (b) Siheung City in Gyeonggi-do, (c) Gwangju Metropolitan City, (d) Suwon City in Gyeonggi-do, (e) Anyang City in Gyeonggi-do and (f) Chuncheon in Gangwon-do. The x and y represent the latitude and longitude coordinates

3. Dataset

In this study, geotechnical investigation reports that surveyed the soil across South Korea were employed to build a database of dynamic and static properties. Data from six regions were utilized: Muan City in Jeollanam-do, Siheung City in Gyeonggi-do, Gwangju Metropolitan City, Suwon City in Gyeonggi-do, Anyang City in Gyeonggi-do, and Chuncheon in Gangwon-do as shown in Fig. 2. After evaluating the data quality in each report, 26 boreholes were selected, compiling a total of 130 datasets. Fig. 3 displays clear maps of each region at a 100-meter scale. Fig. 3(a) shows Muan City in Jeollanam-do, with central coordinates 256,925.55 / 146,187.73. Fig. 3(b)

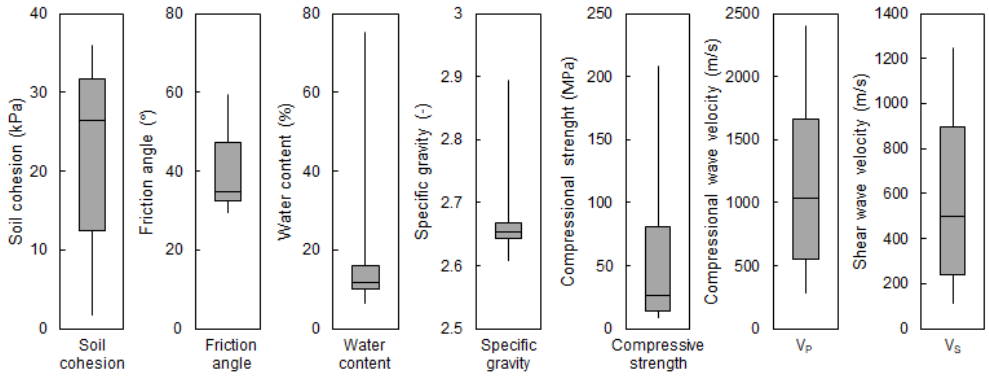
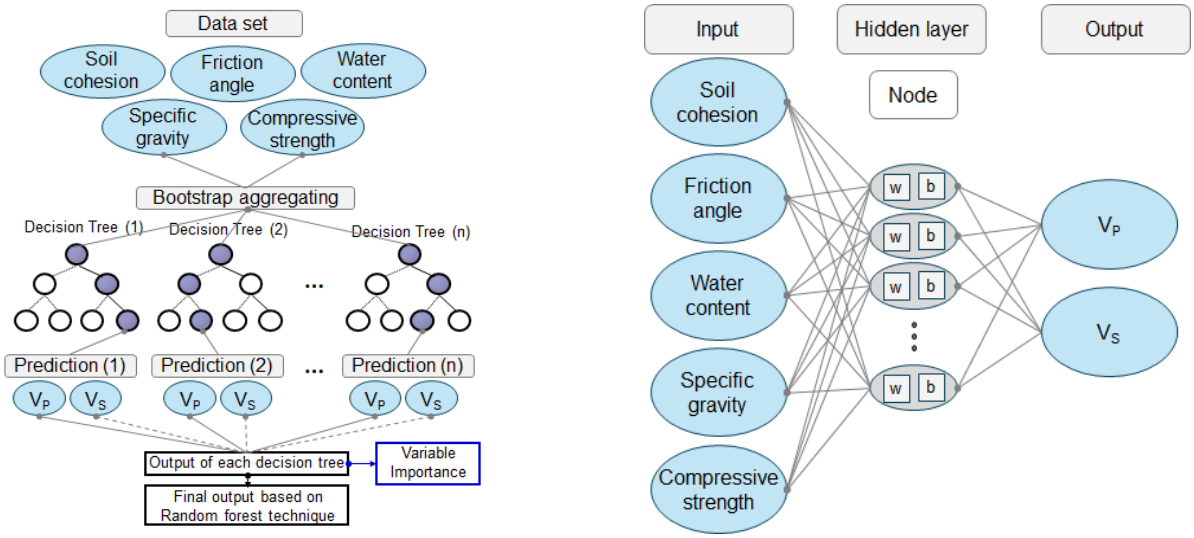


Fig. 4 Input and output variables captured by geotechnical investigation reports



(a) (b)

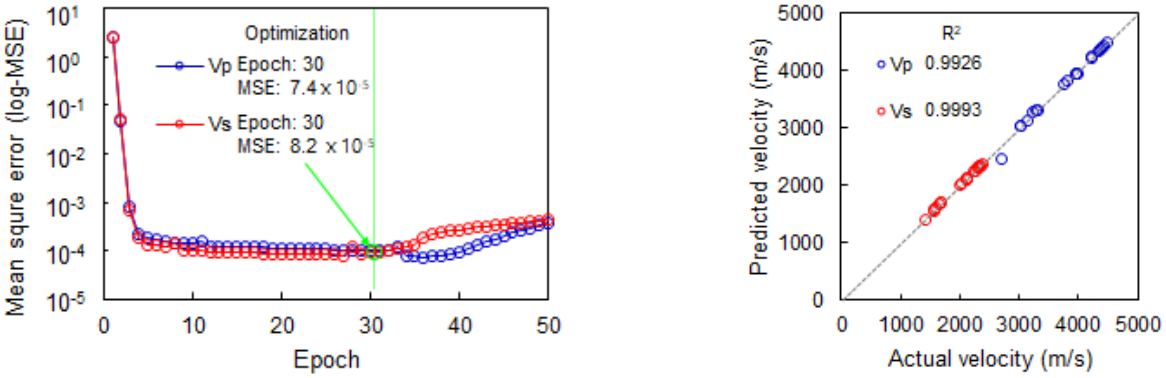
Fig. 5 Architecture of machine learning algorithm: (a) random forest (RF) and (b) deep neural network (DNN)

depicts Siheung City in Gyeonggi-do, with central coordinates 530,884.87 / 181,985.61. Figs. 3(c), 3(d), 3(e), and 3(f) respectively represent Gwangju Metropolitan City, Suwon City in Gyeonggi-do, Anyang City in Gyeonggi-do, and Chuncheon in Gangwon-do, with central coordinates 519,104.73 / 205,627.56, 533,441.71 / 198,140.52, and 116,425.95 / 613,578.02. The boreholes are located within a 200-meter diameter around the central coordinates of each region.

It marks the boreholes drilled in each area, with Muan City in Jeollanam-do, Siheung City in Gyeonggi-do, Gwangju Metropolitan City, Suwon City in Gyeonggi-do, Anyang City in Gyeonggi-do, and Chuncheon in Gangwon-do having conducted 5, 5, 4, 4, 4, and 4 boreholes respectively. The borehole investigations extended to a maximum depth of 20 meters, with data collection intervals set at 1-2 meters, focusing on depths where both input and output properties were documented. The input properties included soil cohesion, friction angle, water content, specific gravity, and compressional strength, whereas the output properties comprised velocities of compressional and shear waves. The distribution of these input and output properties is depicted in Fig. 4, where a box plot reveals the maximum, minimum, and median values of each factor. The ranges of maximum and minimum values for the input parameters soil cohesion, friction angle, water content, specific gravity, and compressional strength were 1.76-36 kPa, 29.6-59.3°, 6.5-75.1%, 2.61-2.895, and 8.6-207.9 MPa respectively. The median values for these input parameters were determined to be 26.35 kPa, 34.73°, 11.64 %, 2.65, and 26.34 MPa, respectively. The distribution for soil cohesion was found to be significantly wide, whereas specific gravity displayed a relatively narrow range. The ranges for the output values of compressional and shear wave velocities were 287-2402 m/s and 111-1249 m/s, respectively, with median values observed at 1038 m/s and 502 m/s. Fig. 4 indicates that the distributions of compressional and shear wave velocities are nearly similar.

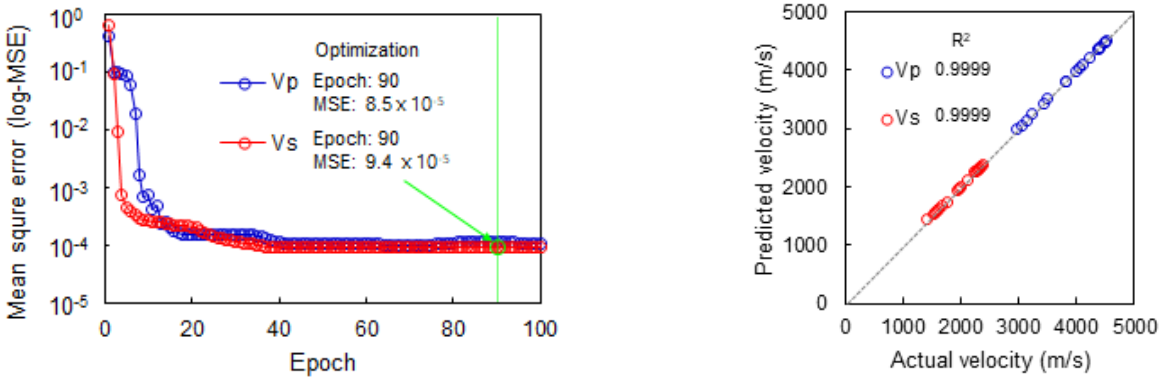
4. Methodology

As outlined earlier in this study, our objective was to establish the relationship between input and output properties through the utilization of the Random Forest (RF) and Deep Neural Network (DNN) algorithms. Prior to employing each algorithm, we performed normalization on the values of each



(a) (b)

Fig. 6 DNN results with Levenberg-Marquardt (LM) method: (a) mean square error and (b) comparison between actual and predicted velocities



(a) (b)

Fig. 7 DNN results with Bayesian Regularization (BR) method: (a) mean square error and (b) comparison between actual and predicted velocities

property using the min-max scaling method to eliminate inaccuracies associated with units and significant digits. In the case of Random Forest, the primary hyperparameters, namely the number of estimators and the depth, were configured to 500 and 1000, respectively. To assess the influence of the 5 input properties on the output parameters, we conducted an analysis on the importance values of each input as shown in Fig. 5(a).

The DNN algorithm was designed with five input properties and a single hidden layer as shown in Fig. 5(b). This layer was allocated 10 nodes, employing the ReLU function as the activation mechanism. Typically, more than two hidden layers are recommended, but due to the limited data volume and the goal of developing a dependable architecture via hyperparameter tuning, only one layer was opted for in this study. The training and testing datasets were split according to the widely adopted 7:3 ratios for validation purposes. The optimization technique selected was the Adam method, which, as mentioned earlier, is crucial for the accuracy of backpropagation. In addition to the Adam method, this study also incorporated the Levenberg-Marquardt (LM) and Bayesian Regularization (BR) methods concurrently. The LM approach merges gradient descent with Gauss-Newton's method, adeptly pinpointing the moment the loss function reaches its nadir. This strategy mitigates the gradient descent method's shortcoming of potentially bypassing the minimum point, facilitating more precise output value predictions. Furthermore, the BR method updates weights and biases via prior probability density functions and distributions, leveraging probability functions to foresee when the loss function will hit its lowest, thus contributing to improved reliability. Ultimately, the predicted output data were verified using the holdout method, and the mean squared error (MSE) served as the loss function.

5. Results

The DNN results for compressional wave velocity and shear wave velocity are presented in Figs. 5 and 6, demonstrating the performance graphs and prediction outcomes. The performance graphs depict the mean squared error (MSE) across the number of learning iterations (epochs) within the optimization algorithm. The epoch demonstrating the lowest error rate was chosen for further learning iterations. Prediction outcomes were obtained by comparing the actual values to the predicted ones, and prediction accuracy was evaluated based on the coefficient of determination.

Figs. 6(a) and 6(b) show the results obtained using the Levenberg-Marquardt (LM) algorithm, demonstrating a rapid decrease in the MSE values for both compressional and shear wave velocities up to the initial 3 epochs, followed by a trend

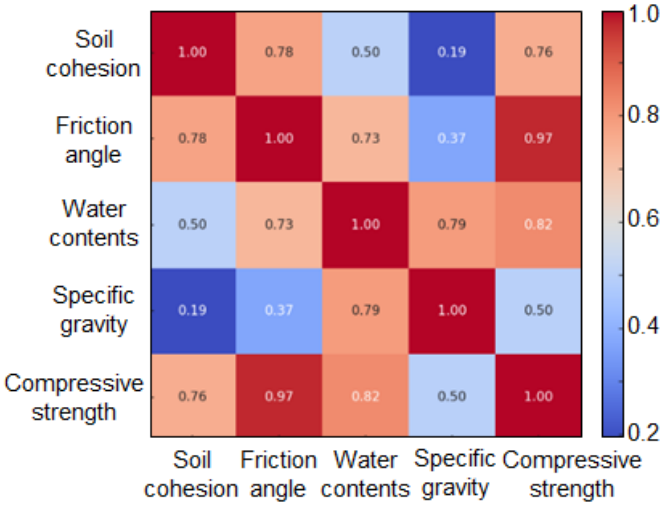


Fig. 8 Correlation matrix between each factor

towards convergence. At epoch 30, the MSE values reached their lowest at 7.4×10^{-5} and 8.2×10^{-5}, respectively, and thereafter displayed an increasing trend due to overfitting beyond 30 iterations. This number of epochs is deemed the most suitable for predictive learning, with the outcomes after 30 iterations depicted in Fig. 6(b). Fig. 6(b) relies on test data to ascertain the final prediction accuracy, employing 19 data points. The predicted compressional wave velocity ranged from a minimum of 2465 m/s to a maximum of 4500 m/s, which is within the actual value range of 2973 m/s to 4514 m/s. Similarly, the predicted shear wave velocity ranged from a minimum of 1393 m/s to a maximum of 2374 m/s, fitting within the actual value range of 1430 m/s to 2374 m/s. The predicted values for both compressional and shear wave velocities were within the ranges of actual values, and the coefficient of determination was analyzed for precise accuracy assessment. The coefficient of determination for compressional waves was calculated to be 0.9926, and for shear waves, it was 0.9993, both indicating a high level of accuracy.

Figs. 7(a) and 7(b) display the outcomes of the Bayesian Regularization (BR) algorithm, similarly showcasing performance graphs and prediction outcome graphs as with the LM algorithm. Fig. 7(a), the performance graph, indicated the lowest MSE values for both compressional and shear wave velocities at epoch 90, with respective values of 8.5×10^{-5} and 9.4×10^{-5}. Both the BR and LM algorithms exhibited similar error rates, though the BR algorithm required a higher number of epochs due to its algorithmic structure. Contrary to LM, the BR algorithm demonstrated a stable MSE convergence zone after epoch 5 without showing signs of overfitting. Fig. 7(b), which illustrates prediction outcomes, revealed predicted ranges for compressional and shear wave velocities of 2982 m/s – 4512 m/s and 1450 m/s – 2372 m/s, respectively, both encompassing the actual value ranges. The coefficient of determination was calculated as 0.9999 for both compressional and shear wave velocities, suggesting that the BR algorithm achieved slightly higher accuracy than the LM algorithm, by 0.0073 for compressional waves and 0.0006 for shear waves.

6. Discussion

6.1 Important value

Fig. 8 displays a matrix of Pearson correlation coefficients among various factors, indicating that values closer to 1 signify higher correlations between two factors. Soil cohesion has high correlation coefficients of 0.78 with friction angle and 0.76 with compressive strength, but shows lower coefficients of 0.5 with water contents and 0.19 with specific gravity. The friction angle is highly correlated with soil cohesion at 0.78, water contents at 0.73, and compressive strength at 0.97, with the lowest correlation being 0.5 with specific gravity. Water contents exhibit correlation coefficients of 0.79 with specific gravity and 0.82 with compressive strength. Specific gravity has a correlation coefficient of 0.5 with compressive strength. These correlation coefficients quantitatively describe the relationships between each pair of factors. The study aimed to ascertain the impact of each input variable on the output variable based on these correlations and to determine which types of input variables should be used. To assess the importance of input variables, algorithms such as Random Forest (RF), SHapley Additive Explanations (SHAP), and Local Interpretable Model-agnostic Explanations (LIME) were utilized (Menze *et al.* 2009; Saarela and Jauhiainen 2021, Meddage *et al*. 2022, Thisovithan *et al*. 2023, Madhushani *et al*. 2024). In this study, Random Forest (RF) was used to select key factors.

Fig. 9 displays the importance of input data derived from random forest (RF) results, quantitatively highlighting the key input factors for predicting compressional and shear wave velocities. The importance ranking for factors influencing both compressional and shear wave velocity predictions follows a descending order: specific gravity, water content, friction angle, compressive strength, and soil cohesion. The significance for predicting compressional wave velocity is represented as specific gravity 0.208, water content 0.204, friction angle 0.202, compressive strength 0.200, and soil cohesion 0.199, showing similar values within the range of

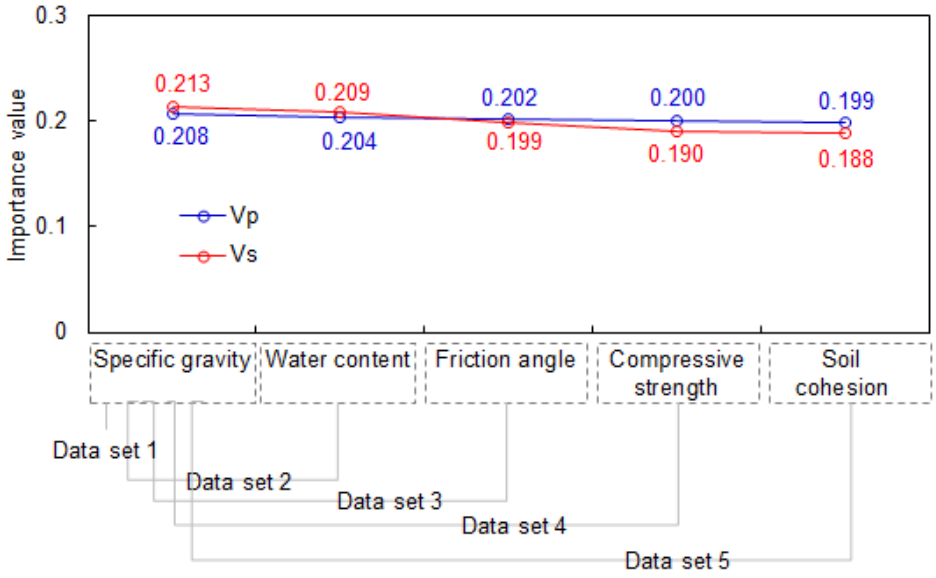


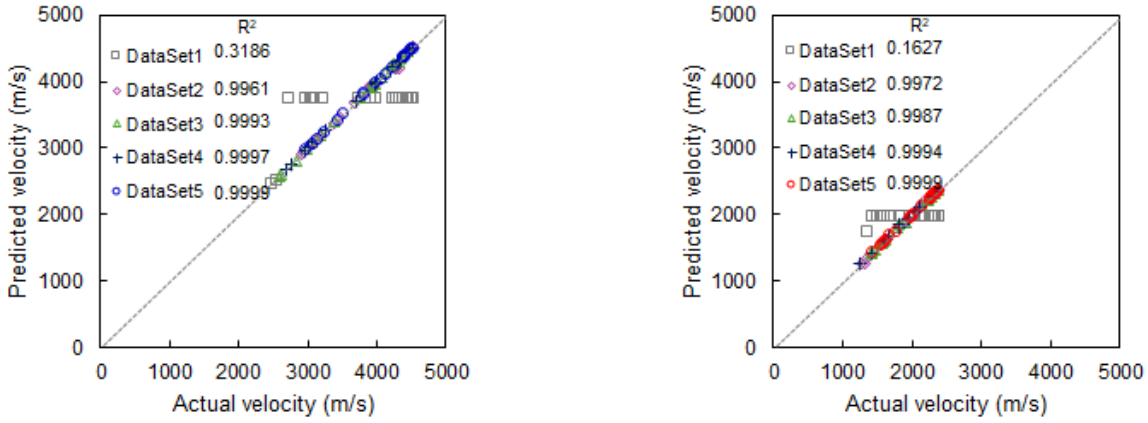
Fig. 9 Important values of input variable deduced by random forest algorithm

0.199–0.208. For shear wave prediction, the importance is specific gravity 0.213, water content 0.209, friction angle 0.199, compressive strength 0.190, and soil cohesion 0.188, indicating a similar level of importance despite a greater difference between the minimum and maximum importance values of 0.025 compared to compressional wave velocity. The importance value quantitatively demonstrates the significance of each factor in predicting the output variable. Therefore, a higher value indicates a greater influence on the prediction of the output variable. This importance reflects relative significance as determined by the random forest algorithm, emphasizing the order of importance over the magnitude of differences. Consequently, the study reorganized the input data based on the importance ranking and structured six data sets according to the importance order as illustrated in Fig. 9. The composition of these data sets took into account the minimum required input factors and their quantities for DNN prediction.

6.2 Prediction of compression and shear wave velocities

This study explores the multidimensional correlations between various static and dynamic properties using DNN to predict dynamic properties. Although five static properties (inputs) were utilized, acquiring all these factors in the field is subject to time and economic constraints, and employing multiple factors increases the likelihood of outliers, impacting the accuracy of predictions. To determine the appropriate input factors and their quantities, predictions for compressional and shear wave velocities were made based on the RF importance-based data set composition depicted in Fig. 9. Data set 1 consists solely of specific gravity/ data set 2 - specific gravity and water content/ data set 3 - specific gravity, water content and friction angle/ data set 4 includes specific gravity, water content, friction angle and compressive strength/ data set 5 incorporates all five factors as shown in Fig. 4. The DNN algorithm employed was the BR algorithm, which demonstrated higher prediction accuracy than the LM algorithm, with results illustrated in Fig. 10. Fig. 10 plots the predicted values against the actual values, comparing the accuracy of data sets based on the coefficient of determination. Fig. 10(a) presents the prediction accuracy for compressional wave velocity, with predicted compressional wave velocities ranging from a minimum of 2461 m/s to a maximum of 4531 m/s across all data sets. Data set 1, predicting compressional wave velocity using only density, yielded results ranging from a minimum of 2461 m/s to a maximum of 375 m/s, corresponding to 62.9 % of the actual value range of 2458 m/s – 4517 m/s, indicating low prediction outcomes. The coefficient of determination was calculated to be 0.3186, showing the lowest accuracy, suggesting that a data set utilizing only one factor, density, has limitations in prediction. Data set 2's predicted values ranged from 2914 m/s to 4531 m/s, with a coefficient of determination calculated as 0.9961. Data sets 2, 3, 4, and 5 exhibited predicted ranges for compressional wave velocity of 2914 m/s – 4531 m/s, 2578 m/s – 4438 m/s, 2683 m/s – 4514 m/s, and 2982 m/s–4512 m/s, respectively, with coefficients of determination calculated as 0.9961, 0.9993, 0.9997, and 0.9999, respectively. From data set 2 onwards, high accuracy with a coefficient of determination above 0.9 was observed, with an average difference between factors of 0.001, indicating a similar level of accuracy.

Fig. 10(b) shows the results for predicting shear wave velocity, with the overall range identified as 1255 m/s to 2372 m/s. Data set 1 revealed a predicted value range of 1750 m/s to 1969 m/s, accounting for 21.4% of the actual value range, thus indicating low prediction outcomes. The coefficient of determination was calculated to be 0.1627, mirroring the lowest value observed in the prediction of compressional wave velocity. Data sets 2, 3, 4, and 5 demonstrated predicted ranges for shear wave velocity of



(a) (b)

Fig. 10 Relationship between accrual and predicted velocities with various input variable groups. The input variable groups were selected with consideration of important values in Fig. 7

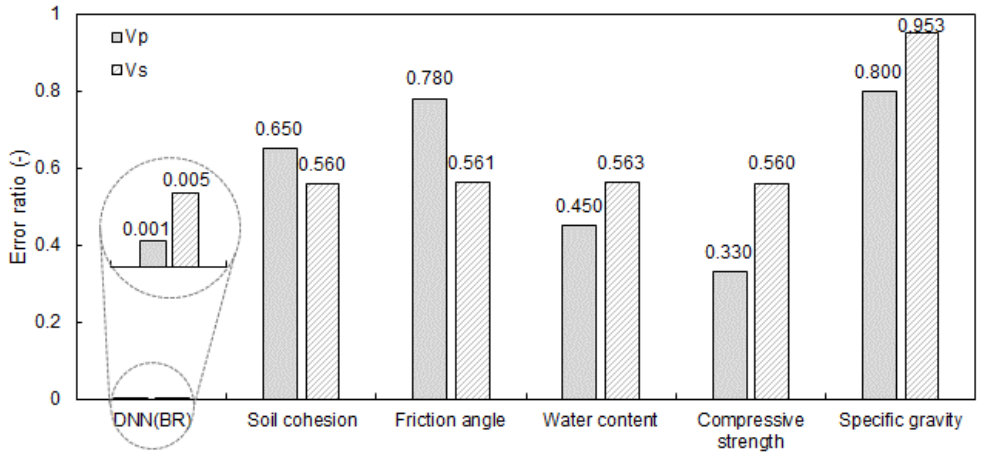


Fig. 11 Comparison of error ratio for deducing dynamic properties with each input variable

1255 m/s to 2366 m/s, 1409 m/s to 2369 m/s, 1257 m/s to 2356 m/s, and 1450 m/s to 2372 m/s, respectively, with coefficients of determination calculated as 0.9972, 0.9987, 0.9994, and 0.9999, respectively. From data set 2 onwards, a high level of accuracy was observed, with a coefficient of determination above 0.9 and an average difference between factors of 0.0009, indicating a consistent level of accuracy.

The coefficients of determination for both compressional and shear wave velocities indicate a high level of accuracy above 0.9 from data set 2 onwards, suggesting that a data set configuration using only one factor, such as specific gravity, is limited in predictive capability. Given the high accuracy levels, the optimal composition of data sets involves configurations with more than two static properties. It is believed that combining more than two factors and their quantities, considering the ease of data acquisition, will facilitate reliable predictions.

6.3 Comparison of error ratio for verification

In this study, the velocities of compressional and shear waves were predicted based on the relationship between static property values and SPT N values for verification, and these predictions were compared with those made using DNN. The static properties of soil cohesion, friction angle, water content, compressive strength, and specific gravity, along with SPT N values from the same borehole locations, had ranges of 3-35, 5-50, 8-50, 16-50, and 3-50, respectively. Compressional and shear wave velocities were derived using empirically correlated equation (Fauzi *et al.* 2014, Ghazi *et al.* 2015) at the same locations and depths where these five properties were obtained. The derived velocities for compressional and shear waves averaged 2800 m/s and 800 m/s, respectively, based on 130 data points. The error ratio was determined by comparing the difference between the empirical formula-predicted values of compressional and shear wave velocities and the actual values, and this was then compared with the DNN(BR) error ratio, as illustrated in Fig. 11. The error ratios for predicting compressional and shear wave velocities using empirical formulas were calculated to be 0.650 and 0.560 for soil cohesion, 0.780 and 0.561 for friction angle, 0.450 and 0.563 for water content, 0.330 and 0.560 for compressive strength, and 0.800 and 0.953 for specific

gravity. These ratios were found to be at least 330 times and up to 950 times higher than the DNN(BR) error ratios of 0.001 and 0.005, respectively. The empirical formulas, which establish relationships based on the characteristics of individual factors without considering various physical properties, tend to have relatively low reliability. However, the method proposed in this study utilizes all property values to establish relationships that interconnect the characteristics of each property, thereby demonstrating a high level of reliability. This validation confirms the reliability of the DNN(BR)-based dynamic property prediction model and its potential to replace traditional empirical formulas.

7. Conclusions

In this paper, we employed machine learning algorithms to establish and predict the relationship between static and dynamic properties, as well as to evaluate the reliability of these predicted dynamic properties. The detailed conclusions of the study are as follows:

Data were sourced from geotechnical investigation reports, from which 26 boreholes were selected based on surveys conducted in six different geological strata. From these boreholes, 130 instances of soil cohesion, friction angle, water content, specific gravity, and compressional strength were identified as input properties. Furthermore, the output parameters were chosen to be the dynamic properties, specifically the velocities of compressional and shear waves.

To evaluate the influence of five static properties on dynamic properties, the random forest algorithm was employed. It was determined that specific gravity had the most substantial effect on both compressional wave velocity and shear wave velocity, whereas soil cohesion exhibited a comparatively lesser impact.

The dynamic properties predicted from static properties demonstrated results that closely align with a 1:1 line when utilizing gradient descent, Levenberg-Marquardt (LM), and Bayesian regularization (BR) methods. This indicates that the coefficient of determination values exceeded 0.9, suggesting that reliable results can be obtained in the estimation of dynamic properties based on static properties.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1A2C2012113).

References

Anbazhagan, P., Kumar, A. and Sitharam, T.G. (2013), "Seismic site classification and correlation between standard penetration test N value and shear wave velocity for Lucknow City in Indo-Gangetic Basin", *Pure Appl. Geophys*., **170**, 299-318. https://doi.org/10.1007/s00024-012-0525-1.

Ataee, O., Moghaddas, N.H. and Lashkaripour, G.R. (2019), "Estimating shear wave velocity of soil using standard penetration test (SPT) blow counts in Mashhad city", *J. Earth Syst. Sci*., **128**(3), 66. https://doi.org/10.1007/s12040-019-1077-x.

Benemaran, R.S. and Esmaeili-Falak, M. (2023), "Predicting the Young's modulus of frozen sand using machine learning approaches: State-of-the-art review", *Geomech. Eng*., **34**(5), 507-527. https://doi.org/10.12989/gae.2023.34.5.507.

Do, T.M., Laue, J., Mattsson, H. and Jia, Q. (2023), "Excess pore water pressure generation in fine granular materials under undrained cyclic triaxial loading", *Int. J. Geo-Eng*., **14**(1), 8. https://doi.org/10.1186/s40703-023-00185-y.

Fauzi, A., Irsyam, M. and Fauzi, U.J. (2014), "Empirical correlation of shear wave velocity and N-SPT value for Jakarta", *GEOMATE J*., **7**(13), 980-984. https://doi.org/10.21660/2014.13.3263.

Fereidooni, D. and Karimi, Z. (2023), "Predicting rock brittleness indices from simple laboratory test results using some machine learning methods", *Geomech. Eng*., **34**(6), 697-726. https://doi.org/10.12989/gae.2023.34.6.697.

Ghazi, A., Moghadas, N.H., Sadeghi, H., Ghafoori, M. and Lashkaripur, G.R. (2015), "Empirical relationships of shear wave velocity, SPT-N value and vertical effective stress for different soils in Mashhad", *Iran. Annal. Geophys*., **58**(3), 2015. https://doi.org/10.4401/ag-6635.

Gomaa, A.E., Hasan, A.M., Mater, Y.M. and AbdelSalam, S.S. (2023), "Shell folded footings using different angles and EPS cavity filling: experimental study", *Int. J. Geo-Eng*., **14**(1), 10. https://doi.org/10.1186/s40703-023-00187-w.

Heo, G., Kim, J., Jeong, S. and Kwak, D. (2023), "Evaluation of shear wave velocity prediction models from standard penetration test N values depending on geologic attributes: A case study in Busan", *South Korea. Geotechnics*, **3**(4), 1004-1016. https://doi.org/10.3390/geotechnics3040054.

Hong, W.T., Lee, J.S., Lee, D. and Yoon, H.K. (2022), "Estimation of bulk electrical conductivity in saline medium with contaminated lead solution through TDR coupled with machine learning", *Process Saf. Environ. Protection*, **161**, 58-66. https://doi.org/10.1016/j.psep.2022.03.018.

Ihsan, S., Saqib, S., Rashid, H.M.A., Niazi, F.S. and Qureshi, M.U. (2023), "Predicting blast-induced ground vibrations at limestone quarry from artificial neural network optimized by randomized and grid search cross-validation, and comparative analyses with blast vibration predictor models", *Geomech. Eng*., **35**(2), 121-133. https://doi.org/10.12989/gae.2023.35.2.121.

Kim, S. and Yoon, H.K. (2023), "Application of classification coupled with PCA and SMOTE, for obtaining safety factor of landslide based on HRA", *Bull. Eng. Geol. Environ*., **82**(10), 381. https://doi.org/10.1007/s10064-023-03403-0.

Lee, D.G., Lee, S.Y. and Song, K.I. (2023), "Development of stability evaluation system for retaining walls: Differential evolution algorithm–artificial neural network", *Geomech. Eng*., **34**(3), 329-339. https://doi.org/10.12989/gae.2023.34.3.329.

Lee, J.S., Park, J., Kim, J. and Yoon, H.K. (2022), "Study of oversampling algorithms for soil classifications by field velocity resistivity probe", *Geomech. Eng*., **30**(3), 247-258. https://doi.org/10.12989/gae.2022.30.3.247.

Lee, S.J. and Choi, S.O. (2023), "Mean fragmentation size prediction in an open-pit mine using machine learning techniques and the Kuz-Ram model", *Geomech. Eng*., **34**(5), 547-559. https://doi.org/10.12989/gae.2023.34.5.547.

Madhushani, C., Dananjaya, K., Ekanayake, I.U., Meddage, D.P. P., Kantamaneni, K. and Rathnayake, U. (2024), "Modeling streamflow in non-gauged watersheds with sparse data considering physiographic, dynamic climate, and anthropogenic factors using explainable soft computing techniques", *J.*

Hydrology, 130846. https://doi.org/10.1016/j.jhydrol.2024.130846.
Meddage, P., Ekanayake, I., Perera, U.S., Azamathulla, H.M., Md Said, M.A. and Rathnayake, U. (2022). Interpretation of machine-learning-based (black-box) wind pressure predictions for low-rise gable-roofed buildings using Shapley additive explanations (SHAP)", *Buildings*, **12**(6), 734. https://doi.org/10.3390/buildings12060734.
Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W. and Hamprecht, F.A. (2009), "A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data", *BMC Bioinformatics*, **10**, 1-16. https://doi.org/10.1186/1471-2105-10-213.
Min, D.H. and Yoon, H.K. (2021), "Suggestion for a new deterministic model coupled with machine learning techniques for landslide susceptibility mapping", *Scientific Reports*, **11**(1), 1-24. https://doi.org/10.1038/s41598-021-86137-x.
Min, D.H., Kim, Y., Kim, S. and Yoon, H.K. (2023), "Strategy of oversampling geotechnical parameters through geostatistical, SMOTE, and CTGAN methods for assessing susceptibility of landslide", *Landslides*, 1-17. https://doi.org/10.1007/s10346-023-02166-9.
Nguyen, A.D., Nguyen, V.T. and Kim, Y.S. (2023), "Finite element analysis on dynamic behavior of sheet pile quay wall dredged and improved seaside subsoil using cement deep mixing", *Int. J. Geo-Eng*., **14**(1), 9. https://doi.org/10.1186/s40703-023-00186-x.
Park, J., Lee, J.S., Jang, B.S., Min, D.H. and Yoon, H.K. (2022), "A comprehensive laboratory compaction study: Geophysical assessment", *Geomech. Eng*., **30**(2), 211-218. https://doi.org/10.12989/gae.2022.30.2.211.
Park, J., Lee, J.S. and Yoon, H.K. (2023), "Geoacoustic and geophysical data-driven seafloor sediment classification through machine learning algorithms with property-centered oversampling techniques", Comput.-Aided Civil Infrastruct. Eng., https://doi.org/10.1111/mice.13126.
Rajabian, A. (2023), "Effect of initial failure geometry on the progress of a retrogressive seepage-induced landslide", *Int. J. Geo-Eng*., **14**(1), 11. https://doi.org/10.1186/s40703-023-00189-8.
Saarela, M. and Jauhiainen, S. (2021), "Comparison of feature importance measures as explanations for classification models", *SN Appl. Sci*., **3**(2), 272. https://doi.org/10.1007/s42452-021-04148-9.
Thisovithan, P., Aththanayake, H., Meddage, D.P.P., Ekanayake, I. U. and Rathnayake, U. (2023), "A novel explainable AI-based approach to estimate the natural period of vibration of masonry infill reinforced concrete frame structures using different machine learning techniques", *Results in Eng*., **19**, 101388. https://doi.org/10.1016/j.rineng.2023.101388.
Yazdandoust, M., Jamnani, A.R. and Sabermahani, M. (2023), "The role of wall configuration and reinforcement type in selecting the pseudo-static coefficients for reinforced soil walls", *Geomech. Eng*., **35**(5), 555-570. https://doi.org/10.12989/gae.2023.35.5.555.
Yoon, H.K., Lee, J.S. and Yu, J.D. (2022), "Correlation of granite rock properties with longitudinal wave velocity in rock bolt", Int. J. Rock Mech. Min. Sci., **159**, 105200. https://doi.org/10.1016/j.ijrmms.2022.105200.
Yu, J.D., Lee, J.S. and Yoon, H.K. (2021), "Effects of rock weathering on guided wave propagation in rock bolts", *Tunn. Undergr. Sp. Tech*., **115**, 104069. https://doi.org/10.1016/j.tust.2021.104069.
Zhang, J., Zhou, X., Huang, X., Liu, X., Yuan, J., Liang, X. and Li, J. (2023), "Study on the root interaction characteristics and nonlinear deformation prediction of root piles", *Geomech.* Eng., **35**(3), 221-239. https://doi.org/10.12989/gae.2023.35.3.221.

IC