Proceedings 10th NUMGE 2023
10th European Conference on Numerical Methods in Geotechnical Engineering
Zdravkovic L, Kontoe S, Taborda DMG, Tsiampousi A (eds)

https://doi.org/10.53243/NUMGE2023-94

Improving Soil Liquefaction Potential Evaluation through AI-Based Prediction of SPT Values: A Comparative Study

A. Baghbani[1], F. Daghistani[2], M.M. Shalchiyan[3], S. Costa[1], H. Baghbani[4], R. Shirani Faradonbeh[5], J. Bolouri Bazaz[4]

[1]*School of Engineering, Deakin University, Melbourne, VIC, Australia*
[2] *Engineering Department, La Trobe University, Melbourne, VIC, Australia*
[3] *School of Engineering, Guilan University, Rasht, Iran*
[4]*Ferdowsi University, Mashhad, Iran*
[5]*WA School of Mines: Minerals, Energy and Chemical Engineering, Curtin University, Kalgoorlie, WA, Australia*

ABSTRACT: Liquefaction of soil is one of the most dangerous phenomena that can significantly impact civil structures and cause financial and human losses. The Standard Penetration Test (SPT) is one of the methods for evaluating soil liquefaction potential. However, despite the importance of this test, its uncertainty and the diversity of factors influencing it have made it difficult to predict. To address these challenges, this study attempts to predict the SPT value based on mechanical tests and artificial intelligence (AI) methods, namely support vector regression (SVR), and classification and regression random forests (CRRF). Models were developed using four inputs, including liquid limit, plasticity index, cohesion and friction angle, and the SPT as an output parameter. While it is true that SPT correction factors play a role in determining the final SPT value, the study specifically aimed to investigate the relationship between geotechnical parameters and SPT values. A statistical method, namely multiple linear regression (MLR). The researchers sought to gain insights into the complex nature of soil liquefaction and assess the effectiveness of AI techniques in predicting the SPT values, which are indicators of liquefaction potential. The comparison of the AI models allowed for an evaluation of their respective prediction performances and an understanding of how different techniques could potentially improve accuracy in assessing liquefaction potential. According to the results, the MLR model predicted the SPT value with coefficient of determination (R^2) of 0.254 and the mean absolute error (MAE) of 3.353, whereas the CRRF and SVR models predicted SPT with R^2 0.922 and 0.978, and MAE of 0.971 and 0.476. Moreover, the sensitivity analysis conducted on the best CRRF model suggested that the friction angle parameter has the greatest impact on all three models.

Keywords: Standard penetration test; Liquefaction; classification and regression random forest; support vector regression; multiple linear regression

1 INTRODUCTION

Liquefication is a phenomenon that can occur in loose soils subjected to cyclic loading, resulting in a loss of strength and stiffness that can damage civil structures such as bridges, buildings, and pipelines (Sahebzadeh et al., 2017; Baghbani et al., 2022a; Nguyen et al., 2023). In order to assess the liquefaction potential of soil, the Standard Penetration Test (SPT) is widely used. SPT test results can, however, be affected by a variety of factors, such as soil type, sampling and testing procedures, and the operator's experience, leading to uncertainty (Baghbani et al., 2022b).

Predicting SPT values using soil mechanical properties and machine learning algorithms can help overcome the challenges associated with traditional geotechnical testing methods, including time and cost constraints as well as uncertainty in results. The use of machine learning algorithms has demonstrated promising results in the prediction of soil properties, including the SPT value (Baghbani et al., 2023a).

In spite of the importance of predicting the SPT number, a comprehensive model for its prediction has not yet been developed. This issue can be attributed to the number of effective parameters on the SPT number (Chen and Willson, 1999). Furthermore, the influence of effective parameters is nonlinear. In the past two decades, engineers and geotechnical researchers have used artificial intelligence methods to solve engineering and geotechnical. Numerous fields, including soil mechanics, soil dynamics (Baghbani et al., 2023b) and tunnelling (Lin et al., 2021), have successfully applied artificial intelligence methods. Several studies have used artificial intelligence methods to predict the SPT number, but artificial neural networks have been used in most studies.

In a study conducted by Kim et al. (2022), they analyzed a database of 134 items with 5 input variables: effective stress, standard penetration test result, liquid limit, plastic limit, and plasticity index. Different artificial intelligence models were used to predict the outputs, and the artificial neural network model was found to be the most accurate. In the study of Kim et al. (2022), the artificial neural network has been able to predict SPT values with a coefficient of determination, $R^2 = 0.83$ and mean absolute error, MAE = 14.64 and root mean square error, RMSE = 22.74.

A neural network was developed by Fernando et al. (2021) based on a database that contained input data including: the tip resistance, the sleeve resistance, the effective pressure on the soil overburden, the liquid limit, the plastic limit, and the percentage of sand, silt, and clay. ANNs proved to be effective in both networks with and without normalization of data. According to the study of Fernando et al. (2021), ANNs without data normalization exhibit a lower error value than ANNs with data normalization. As a result of the network model without data normalization, the RMSE values in the training data were 3.024, MAE 1.822, R^2 0.952, and in the test data, the RMSE values were 2.163, MAE 1.233, and R^2 0.976 (Fernando et al., 2021).

In this study, two artificial intelligence models that were not used in the literature, for the first time, were employed to fill the gap in history, namely support vector regression (SVR) and classification and regression random forest (CRRF), and a statistical method, namely multiple linear regression (MLR). Four inputs were used in this study, including liquid limit (LL), plasticity index (PI), cohesion (c) and friction angle (φ). After modelling, the results of the models were compared, and the importance of input parameters was discussed.

2 DATABASE PROCESSING

This study utilized 32-sets database (refer Yusof and Zabidi, 2018) to perform statistical and artificial intelligence modelling. The database contains four inputs and one output, which is a SPT number. The inputs of the database included the liquid limit (LL), plasticity index (PI), cohesion (c) and friction angle (φ) of soil. The statistical parameters derived from the collected data are shown in Table 1.

Table 1. Statistical information of database

Variable	Obs	Min	Max	Mean	Std. deviation
SPT	32	2.50	19.50	9.24	4.70
LL	32	36.00	71.00	51.34	10.28
PI	32	10.00	45.00	19.16	6.90
C	32	0.00	338.00	119.59	69.00
ϕ	32	30.00	40.00	35.97	2.01

Table 2 shows the correlation coefficients between the five independent variables - LL, PI, *c*, *φ*, and SPT. According to the table, the LL and PI variables are strongly positively correlated, with a correlation coefficient of 0.802. Other parameters do not have any strong linear correlation.

Table 2. Correlation coefficients between parameters

	LL	PI	c	ϕ	SPT
LL	1	0.802	-0.347	-0.291	-0.248
PI	0.802	1	-0.331	-0.271	-0.190
C	-0.347	-0.331	1	0.730	0.372
ϕ	-0.291	-0.271	0.730	1	0.039
SPT	-0.248	-0.190	0.372	0.039	1

The linear normalization process involves scaling and transforming data so that it has a specific range of values. The data is first scaled so that the minimum value is mapped to 0 and the maximum value is mapped to 1. Linear normalization is calculated as follows:

$$X_{norm}=\frac{X-X_{min}}{X_{max}-X_{min}} \quad (1)$$

where X_{max}, X_{min}, X, and X_{norm} are maximum, minimum, actual, and normalized values, respectively.

As part of the process of developing an AI model, it is common practice to split the available data into training and testing sets in order to evaluate the performance of the model. It has been decided to randomly divide the main database into two sets, with one set containing 80% of the training data and the other 20% of the testing data. There are 26 observations in the training database, while there are only six observations in the testing database. In both databases, the same six variables are included - SPT, LL, PI, c, and φ. Due to the random selection of values for each variable, the minimum, maximum, mean, and standard deviation differ slightly between the two databases, but they are close. In order to ensure that the AI model can accurately generalize to new, unknown data, the model will be trained on the larger training database and evaluated on the smaller testing database (Tables 3 and 4).

Table 3. Statistical information of training database

Variable	Obs	Min	Max	Mean	Std. deviation
SPT	26	3.90	19.50	9.46	4.80
LL	26	36.00	68.00	50.38	9.75
PI	26	10.00	45.00	19.04	7.17
C	26	0.00	338.00	119.50	72.18
ϕ	26	30.00	40.00	35.96	2.16

Table 4. Statistical information of training database

Variable	Obs	Min	Max	Mean	Std. deviation
SPT	6	2.50	15.00	8.27	4.53
LL	6	40.00	71.00	55.50	12.41
PI	6	45.00	205.00	120.00	58.92
C	6	34.00	38.00	36.00	1.26
ϕ	6	2.50	15.00	8.27	4.53

In order to evaluate the accuracy of machine learning models, it is important to use metrics that provide insights into the model's performance. In this study, four metrics are used: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Root Mean Squared Logarithmic Error (RMSLE), and R-squared (R^2). Following equations define all four metrics:

$$MAE = \frac{\sum_N \left(X_m - X_p\right)}{N} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_N \left(X_m - X_p\right)^2}{N}} \quad (3)$$

$$RMSLE = \sqrt{\frac{\sum_N \left(\log(X_m + 1) - \log(X_p + 1)\right)^2}{N}} \quad (4)$$

$$R^2 = \left[\frac{\sum_{i=1}^{N} (X_m - \overline{X_m})(X_p - \overline{X_p})}{\sum_{i=1}^{N} (X_m - \overline{X_m})^2 \sum_{i=1}^{n} (X_p - \overline{X_p})^2}\right]^2 \quad (5)$$

where N is the number of datasets, X_m and X_p are actual and predicted values, and $\overline{X_m}$, $\overline{X_p}$ are the average of actual and predicted values, respectively. Ideally, the model should have a R^2 value of 1 and a MAE, RMSE, RMSLE value of 0.

3 DATA-DRIVEN MODELING

3.1 Multiple linear regression (MLR)

In statistics, multiple linear regression (MLR) is a method used to model the relationship between a dependent variable and two or more independent variables. It is a type of linear regression that predicts a continuous dependent variable by using two or more predictor variables. MLR is calculated as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon \quad (6)$$

where Y is the dependent variable, X_1, X_2, ... X_n are the independent variables, β_0, β_1, β_2, ... β_n are the coefficients that measure the effect of each independent variable on the dependent variable, and ε is the error term that accounts for the variability in the dependent variable that is not explained by the independent variables.

3.2 Support vector regression (SVR)

Support Vector Regression (SVR) is a machine learning algorithm that is used for regression tasks. It is based on the principles of Support Vector Machines (SVM), which is primarily used for classification. SVR extends the concept of SVR to solve regression problems by formulating the task as a constrained optimization problem.

The goal of SVR is to find a function that best represents the mapping between the input variables and the corresponding continuous target variables. Unlike traditional regression techniques, SVR does not focus on minimizing the error between predicted and actual values. Instead, it aims to find a hyperplane that lies within a specified epsilon tube around the true values, while maximizing the margin of the tube.

SVR has several advantages, including its ability to handle both linear and nonlinear regression problems, robustness against outliers, and versatility in handling various kernel functions. However, SVR can be computationally expensive and sensitive to the choice of hyperparameters, such as the regularization parameter (C) and the kernel function parameters.

3.3 Classification and Regression Random Forests (CRRF)

A Random Forest (RF) is a type of ensemble learning algorithm that utilizes multiple decision trees to improve the accuracy and robustness of a prediction model. In classification and regression tasks, Classification and Regression Random Forests (CRRF) are specific variations of Random Forest.

As a result of CRRF, multiple decision trees are combined into a forest, where each tree is trained on a random subset of the training data and a random subset of the features. Each tree's output is then combined to produce a final prediction. The final prediction is determined by the mean of the individual predictions made by the trees in classification tasks, while in regression tasks, the final prediction is determined by the mode of the individual predictions made by the trees in regression tasks. It has the advantage of being less prone to overfitting than a single decision tree, since each tree only sees a random subset of the data and features. As a result, the CRRF is capable of generalizing to new data more easily. Furthermore, CRRF is capable of handling non-linear relationships between features and the target variable since the trees can capture these relationships piecewise.

4 RESULTS

4.1 Multiple linear regression (MLR)

Table 5 illustrates the performance of MLR for the prediction of SPT using a training and testing databases. The MLR model performed moderately in predicting SPT values, with a relatively high RMSE for the training dataset and a slightly lower RMSE for the testing dataset. Furthermore, the R^2 values indicate that the model was only able to explain a small portion of the variance in SPT values.

Table 5. The performance of the MLR model to predict SPT

Statistic	Training database	Testing database
Observations	26	6
MAE	3.168	3.353
RMSE	4.000	3.573
RMSLE	0.380	0.504
R^2	0.278	0.254

Further, Figure 1 shows the results of the MLR model based on predicted values versus actual SPT. The results indicate that the MLR model does not have good accuracy for predicting SPT and that a more complex model such as artificial intelligence is required.

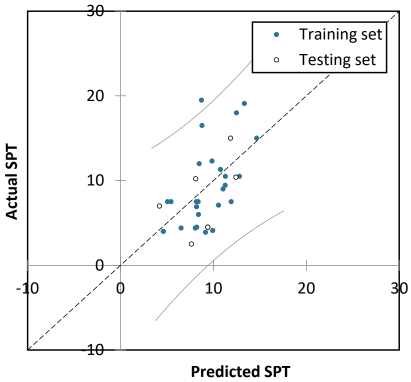


Figure 1. The results of MLR model

4.2 Support vector regression (SVR)

Table 6 shows the performance of a SVR model in predicting SPT values using a training and testing database. Both the training and testing databases indicate that the SVR model performed well. Both the training and testing datasets showed good performance for predicting SPT values using the SVR model, with low MAE and RMSE values, with high R^2 values for both datasets.

Table 6. The performance of the SVR model to predict SPT

Statistic	Training database	Testing database
Observations	26	6
MAE	0.842	0.476
RMSE	0.965	0.615
RMSLE	0.108	0.122
R^2	0.958	0.978

Figure 2 displays the results of the best SVR model based on the best prediction of the model against the actual SPT values. According to the results, SVR model has been able to accurately predict SPT.

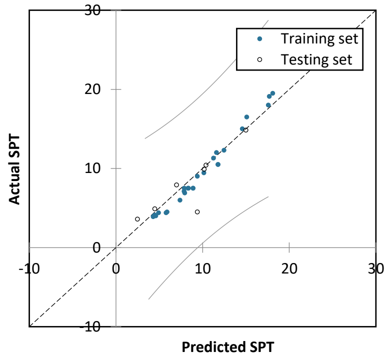


Figure 2. The results of SVR model

4.3 Classification and Regression Random Forests (CRRF)

The specifications of the best CRRF model were determined by using a certain set of parameters. Two sets of parameters are considered: trees parameters and forest parameters. Parameters of the tree include the minimum size of the nodes, the minimum size of the sons, the maximum depth, the number of variables used at each split (Mtry), and the complexity parameter (CP). Minimum node size refers to the minimum number of observations required to form a terminal node. It is the smallest size of the node before the tree-growing process ceases. In order to prevent overfitting, the maximum depth limits the depth of the tree. The number of variables randomly selected as candidates for each split is determined by M_{try}. By adding a penalty term to each split, the complexity parameter controls the size of the tree and prevents overfitting.

In contrast, forest parameters include the sampling method, sample size, and number of trees. A sampling method describes how samples are selected from a dataset. The method used in this case was random with replacement. In this case, the sample size was set at 24, which corresponds to the number of observations in the dataset used to build each tree. There are 300 trees in the forest, which indicates the number of decision trees.

Table 7 presents the performance of the CRRF in predicting the SPT in both the training and testing databases. The low MAE, RMSE, and RMSLE values and high values of R^2 in the testing and training databases suggest that the CRRF model is an effective predictor of SPT.

Table 7. The performance of the CRRF model to predict SPT

Statistic	Training database	Testing database
Observations	26	6
MAE	1.078	0.971
RMSE	1.347	1.153
RMSLE	0.125	0.141
R^2	0.918	0.922

Figure 3 compares the predicted values of the CRRF model with the actual SPT values. The CRRF model has been successful in predicting SPT values based on the results presented in Figure 3 and Table 8.

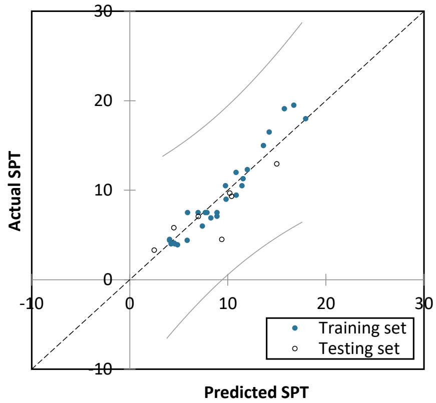


Figure 3. The results of CRRF model

5 DISCUSSION

5.1 *Comparison of different models*

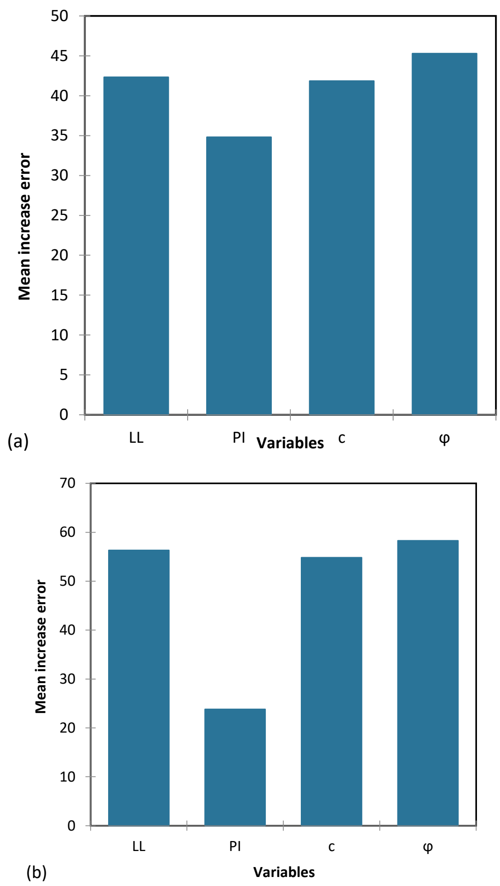
Table 8 presents the results of three mathematical models - MLR, SVR, and CRRF - for predicting SPT values in both training and testing databases. According to the training database, the MLR model showed the highest MAE values, while the SVR model showed the lowest MAE, RMSE, and RMSLE values, indicating better performance. In terms of R^2, the SVR model outperformed both the MLR and CRRF models. In particular, the R^2 value of the CRRF model was 0.958, which was higher than the R^2 value of the MLR model, which was 0.278, and the R^2 value of the CRRF model, which was 0.918. According to all the performance metrics measured in the testing database, the SVR model performed the best. As compared to other models, the SVR model had the lowest values for MAE, RMSE, and RMSLE, indicating superior predictive power. Among the three models tested in the testing database, the MLR model had the highest MAE, and RMSE values, while the CRRF model performed better than the MLR model, but not as well as the SVR model. The SVR model had the highest R^2 value in the testing database, followed by the CRRF model and then the MLR model.

Table 8. The results of four mathematical models for training and testing databases

Statistic	Training database			Testing database		
	MLR	SVR	CRRF	MLR	SVR	CRRF
MAE	3.17	0.84	1.08	3.35	0.48	0.97
RMSE	4.00	0.96	1.35	3.57	0.61	1.15
RMSLE	0.38	0.11	0.12	0.50	0.12	0.14
R^2	0.28	0.96	0.92	0.25	0.98	0.92

5.2 *The variable importance of input parameters*

An important aspect of artificial intelligence modelling is the input parameters. This issue is tested by increasing and decreasing each input parameter individually by 100%, while keeping the other parameters constant. This error is then calculated, and the maximum error is taken into account. Therefore, when the error created for a parameter is greater than that created for other input parameters, it indicates the model is more sensitive to change in this parameter, therefore the parameter is more important. The results of parameter importance error are illustrated in Figure 4 for all three mathematical models used in this study. Table 9 shows the importance of input parameters for each model based on Figure 4. Friction angle was most important in all three models, while plasticity index was least important. Furthermore, cohesion was the second most important parameter in MLR and SVR, and the third most important parameter in CRRF. Liquid limit was the third most important parameter in models MLR and SVR, and the second most important parameter in CRRF.



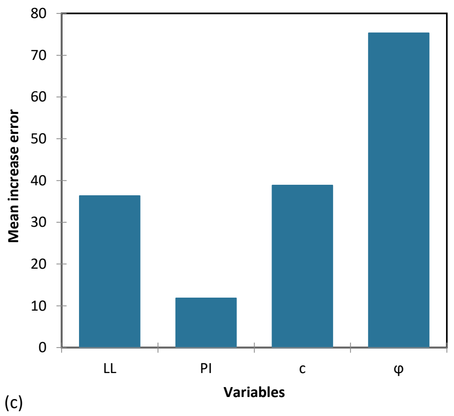


Figure 4. The importance of parameters in (a) MLR, (b) SVR and (c) CRRF

Table 9. The results of variable importance for all models

Models	Input parameters			
	LL	PI	c	ϕ
MLR	2	4	3	1
SVR	2	4	3	1
CRRF	3	4	2	1
Total score	7	12	8	4
Ranking	2	4	3	1

6 CONCLUSIONS

This study developed three mathematical models, including a statistical model, multiple linear regression (MLR), and two artificial intelligence models, classification and regression random forest (CRRF) and support vector regression (SVR). In the database, there are 32 datasets with four inputs each, namely liquid limit, plasticity index, cohesion, and friction angle of soils. Below is a summary of the results:

- In the testing database, the MLR showed even weaker performance, with higher MAE and RMSE of 3.35 and 3.57, respectively, and a lower R^2 of 0.25. Overall, the MLR performed the poorest among three models tested.
- The SVR exhibited strong performance in predicting SPT. In the training database, SVR had MAE of 0.84 and RMSE of 0.96, while in the testing database, it had MAE of 0.48 and RMSE of 0.61. SVR also showed low RMSLE of 0.11 in the training database and 0.12 in the testing database, and high R^2 of 0.96 and 0.98 in the training and testing databases, respectively. Overall, the SVR outperformed the CRRF and MLR in all the performance metrics.
- CRRF had MAE of 1.08 and RMSE of 1.35 in the training database, and MAE of 0.97 and RMSE of 1.15 in the testing database. The RMSLE was 0.12 in the training database and 0.14 in the testing database. R^2 was 0.92 in the training database and 0.92 in the testing database, indicating a good fit of the CRRF model in predicting SPT values.
- Friction angle was the most important parameter while plasticity index was the least important in all three models.

These outcomes demonstrate the effectiveness of AI techniques in predicting SPT values and provide insights into the relationship between geotechnical parameters and liquefaction potential.

7 REFERENCES

Baghbani, A., Costa, S., O'Kelly, B.C., Soltani, A., Barzegar, M. 2022a. Experimental study on cyclic simple shear behaviour of predominantly dilative silica sand. *International Journal of Geotechnical Engineering*, **17**(1), 91-105.

Baghbani, A., Choudhury, T., Costa, S., Reiner, J. 2022b. Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Science Reviews*, **228**, 103991.

Baghbani, A., Choudhury, T., Samui, P., Costa, S. 2023a. Prediction of secant shear modulus and damping ratio for an extremely dilative silica sand based on machine learning techniques. *Soil Dynamics and Earthquake Engineering*, **165**, 107708.

Baghbani, A., Abuel-Naga, H., Shirani Faradonbeh, R., Costa, S., Almasoudi, R. 2023b. Ultrasonic Characterization of Compacted Salty Kaolin–Sand Mixtures Un-der Nearly Zero Vertical Stress Using Experimental Study and Machine Learning. *Geotechnical and Geological Engineering*, **41**, 2987–3012.

Chen, C.L. and Willson, A.N. 1999. A trellis search algorithm for the design of FIR filters with signed-powers-of-two coefficients. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, **46**(1), 29-39.

Fernando, H., Nugroho, S.A., Suryanita, R., Kikumoto, M. 2021. Prediction of SPT value based on CPT data and soil properties using ANN with and without normalization. *International Journal of Artificial Intelligence Research*, **5**(2), 123-131.

Kim, M., Okuyucu, O., Ordu, E., Ordu, S., Arslan, Ö., Ko, J. 2022. Prediction of Undrained Shear Strength by the GMDH-Type Neural Network Using SPT-Value and Soil Physical Properties. *Materials*, **15**(18), 6385.

Lin, S.S., Shen, S.L., Zhang, N. and Zhou, A. 2021. Modelling the performance of EPB shield tunnelling using machine and deep learning algorithms. *Geoscience Frontiers*, **12**(5), 101177.

Nguyen, M.D., Baghbani, A., Alnedawi, A., Ullah, S., Kafle, B., Thomas, M., Moon, E.M., Milne, N.A. 2023. Investigation on the suitability of aluminium-based water treatment sludge as a sustainable soil replacement for road construction. *Transportation Engineering*, **12**, 100175.

Sahebzadeh, S., Heidari, A., Kamelnia, H., Baghbani, A. 2017. Sustainability features of Iran's vernacular architecture: A comparative study between the architecture of hot–arid and hot–arid–windy regions. *Sustainability*, **9**(5), 749

Yusof, N.Q.A.M., Zabidi, H. 2018, August. Reliability of using standard penetration test (SPT) in predicting properties of soil. *In Journal of Physics: Conference Series*, **1082**, 1, 012094.