Journal of Mineral Resources Engineering, 9(3): 77-94, (2024)

Imam Khomeini International University
Vol. 9, No. 3, Autumn 2024

نشریه مهندسی منابع معدنی
Journal of Mineral Resources Engineering (JMRE)

Research Paper

Prediction of the Maximum Surface Settlement Induced by Urban Area Tunneling Using Multiple Linear Regression Model and Random Forest Algorithm

Dehghani H.[1*], Matinpour F.[2], Mohammadi Bolbanabad Sh.[3], Monjezi M.[4]

1- Associate Professor, Mining Engineering Faculty, Hamedan University of Technology, Hamedan, Iran
2- Ph.D Student, School of Mining Engineering, University College of Engineering, University of Tehran, Tehran, Iran
3- Ph.D Student, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran
4- Professor, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran

Received: 27 Jul. 2023 **Accepted: 04 Nov. 2023**

Abstract: The occurrence of settlement induced by tunneling operations in urban environments is an inevitable phenomenon. The risk of damage to nearby infrastructures and surface structures can be greatly reduced by predicting and controlling this event. This paper uses multiple linear regression (MLR) model and random forest (RF) algorithm to predict the maximum surface settlement (Smax) due to shallow tunnel excavation. Nine input parameters, including the distance of the tunnel center from the ground surface (H), height of the underground water level above the tunnel (W.T), tunnel diameter (D), elastic modulus of soil (E), undrained shear strength of soil (Cu), earth pressure coefficient (K_0), unit weight of soil (γ), gap parameter (g), and stability number (N) were selected from 24 data sets related to 14 tunneling projects. The MLR and RF techniques were then implemented for predicting Smax. Three performance indicators of coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were employed for the training and test phases to evaluate the efficiency of the suggested models. The coefficient of determination values for MLR and RF for training data were 0.814 and 0.957, respectively, while for test data were 0.793 and 0.96, indicating that the RF approach is more efficient than MLR. Moreover, the findings reveal that the RF algorithm exhibits lower RMSE and MAE values in both the training and testing phases compared to the MLR method. This suggests that the RF algorithm exhibits reduced error and higher reliability and accuracy when compared to the MLR model. Also, the performance study demonstrates that among the input parameters, the gap parameter (g) and the undrained shear strength of soil (Cu) have the greatest and least influence on Smax, respectively.

Keywords: Tunneling operations, Maximum surface settlement (Smax), Multiple linear regression (MLR) model, Random forest (RF) algorithm.

How to cite this article
Dehghani, H., Matinpour, F., Mohammadi Bolbanabad, Sh., and Monjezi, M. (2024). *"Prediction of the maximum surface settlement induced by urban area tunneling using multiple linear regression model and random forest algorithm"*. Journal of Mineral Resources Engineering, 9(3): 77-94.
DOI: 10.30479/JMRE.2023.19111.1654

*Corresponding Author Email: *dehghani@hut.ac.ir*

INTRODUCTION

Ground settlement resulting from tunnelling operations is an inevitable occurrence that poses a significant concern in urban underground projects. The ability to predict and control this phenomenon is essential to reduce possible damage to surface structures and nearby infrastructures. Due to the importance of the settlement and its direct impact on the project costs, so far many studies have been conducted based on different methods as well as presenting relationships to predict it. Various methods, including empirical, analytical, numerical, and artificial intelligence approaches are available. In practice, ground movement prediction is mainly based on empirical relationships that do not consider the intrinsic parameters of the soil [1-4]. Although analytical methods are useful tools for predicting S_{max} and present a closed-form solution, they cannot include all the parameters affecting the settlement [5-8]. In numerical methods, mathematical equations are solved using the finite element [9] or discrete element [10] method to predict the movement of the overburden layers and, finally, the settlement of the ground surface is calculated.

In contrast to the three aforesaid methods, artificial intelligence methods provide the possibility of analysing a larger amount of data by integrating complex nonlinear combinations to predict the target parameter. In this paper, using the two methods of multiple linear regression (MLR) model and random forest (RF) algorithm, the amount of S_{max} resulting from tunneling operations is predicted. The results showed that RF algorithm can predict S_{max} more accurately than MLR model.

METHODS

In this study, multiple linear regression (MLR) model and random forest (RF) algorithm were used to predict S_{max} induced by tunneling operations. MLR is a statistical methodology that is widely employed to estimate the result of a variable predicted by the value of two or more variables. Random Forest (RF) is a type of ensemble learning technique, originally introduced by Breiman in 2001 [11], designed to enhance the accuracy of classification and regression tree (CART) models by integrating a vast collection of decision trees. Decision trees can be classified into two categories, namely, classification and regression trees [11]. Given that the primary goal of this research is to predict S_{max} caused by tunneling operations, we only focus on the regression mode. The RF regression technique is classified as a nonparametric regression methodology.

FINDINGS AND ARGUMENT

In order to attain the optimal version of MLR and RF models for predicting S_{max}, their performances are assessed throughout the training and testing phases by utilizing three performance indicators: determination coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE). Through the utilization of the trial and error methodology, the optimal MLR and RF models are achieved. The values of the aforesaid indicators for the training and test phase data are shown in Table 1. Figure 1 displays the acquired relationship between the measured and predicted values of S_{max} from both the RF and MLR models within the training and testing stages. As illustrated in Figure 1, the MLR and RF models exhibit R^2 values of 0.814 and 0.957, respectively, when predicting S_{max} during the training phase. In the testing phase, these models yield R^2 values of 0.793 and 0.96, respectively. Moreover, the findings reveal that the RF algorithm exhibits lower RMSE and MAE values in both the training and testing phases compared to the MLR method. This suggests that the RF algorithm exhibits reduced error and higher reliability and accuracy when compared to the MLR model. From the results obtained above, it can be inferred that the RF algorithm demonstrates greater reliability and heightened accuracy when compared to the MLR model in its ability to predict S_{max} induced by tunneling operations. Furthermore, the RF results reveal a strong correlation with the measured

Table 1. Values of statistical indicators in training and testing phases of MLR and RF models

index	Train phase		Test phase	
	RF	MLR	RF	MLR
R^2	0.9569	0.8139	0.9605	0.7934
RMSE	9.36	15.93	9.94	38.42
MAE	6.84	12.64	8.38	29.09

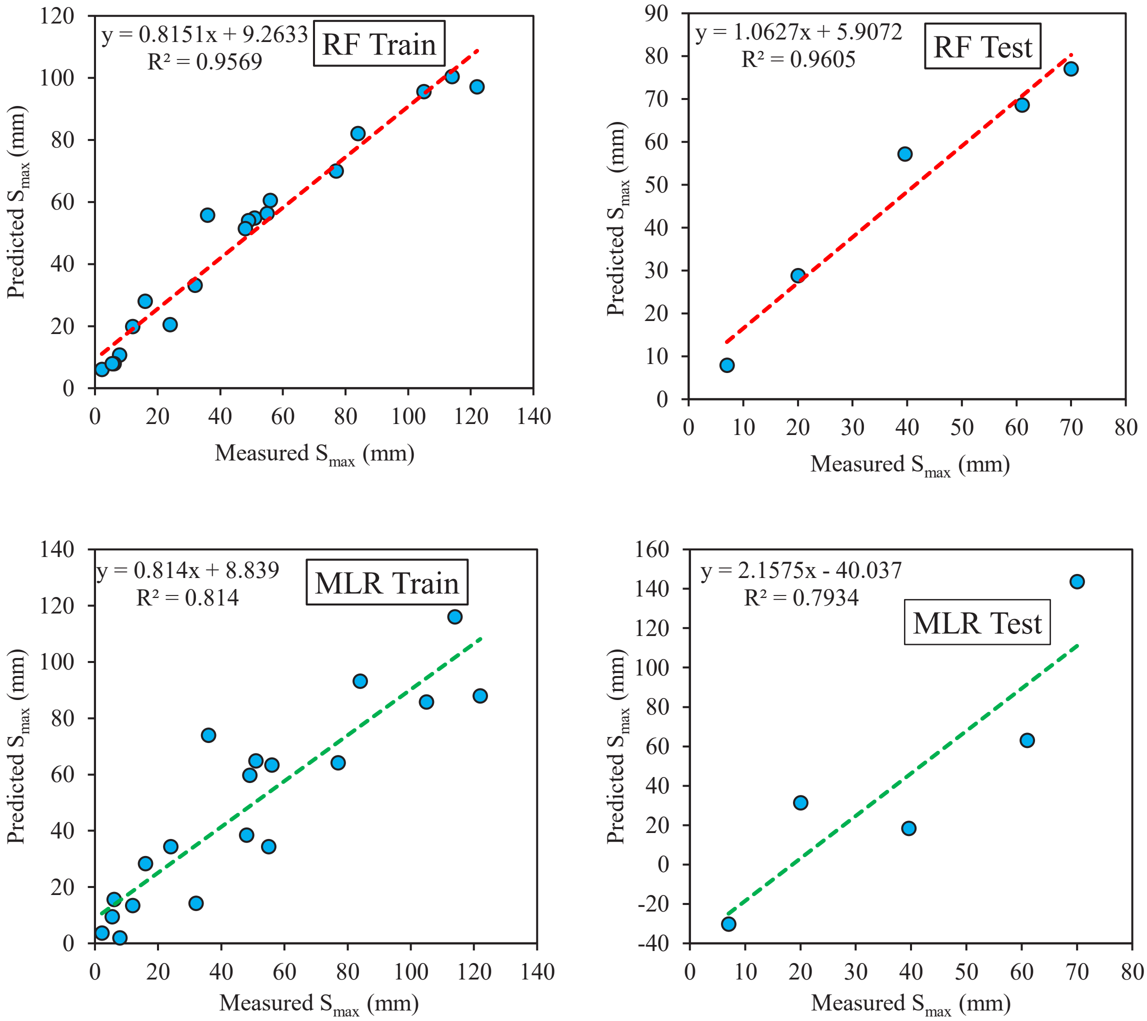

Figure 1. The relationship of measured and predicted values of the S_{max} in training and testing phases of RF and MLR models

data and align closely with the actual values. Consequently, it can be deduced that the RF algorithm serves as a suitable tool for the prediction of S_{max} resulting from tunneling operations conducted in urban environments.

Based on the methodology employed to assess importance, the impact values of 9 input variables on the resulting S_{max} value derived from the RF algorithm are computed and presented in Figure 2. It is evident that the parameter g holds the highest importance in determining S_{max}. Conversely, the parameter C_u is of the least importance in modeling the S_{max} parameter.

CONCLUSIONS

In this paper, using two methods of multiple linear regression (MLR) and random forest (RF) algorithm, the prediction of S_{max} caused by tunneling operations in urban environments was studied. MLR and RF methods were implemented based on 24 data sets related to different tunneling projects and selection of 9 variables as influencing parameters on S_{max}. In the first stage, 80% of the data were used as training data and 20% as test data to implement the suggested methods. Then, in order to evaluate the efficiency of the used methods, three performance indicators of the coefficient of determination parameter (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used for the training and test data. The results of the models showed that the R^2 values of MLR and RF methods for training data are 0.814 and 0.957, respectively, and for test data are 0.793 and 0.96, respectively. Also, the findings demonstrate that the RF algorithm exhibits lower RMSE and MAE values in both the training and testing phases compared to the

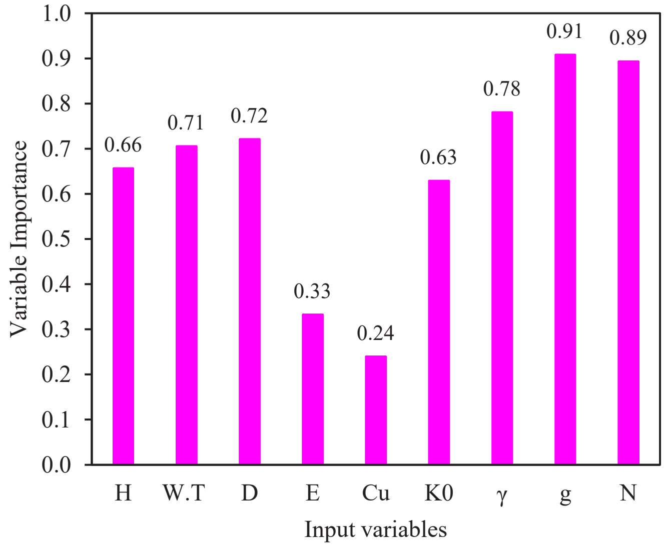

Figure 2. The result of importance study for S_{max} modeling

MLR method. The obtained results indicate the higher efficiency and accuracy of RF method compared to MLR. Finally, variable importance (VI) index was utilized to determine the most and least important parameters affecting S_{max}. The obtained results showed that among the input parameters, the gap parameter (g) and the undrained shear strength of soil (C_u) have the greatest and least influence on S_{max}, respectively.

REFERENCES

[1] Herzog, M. (1985). *"Surface subsidence above shallow tunnels (in German)"*. Bautechnik, 62: 375-377.

[2] O'Reilly, M. P., and New, B. M. (1982). *"Settlement above tunnels in the United Kingdom-their magnitude and prediction"*. In: Proceedings of the tunnelling conference, Brighton, 173-181.

[3] Peck, R. B. (1969). *"Deep excavations and tunnelling in soft ground"*. In: 7th International conference on soil mechanics and foundation engineering, State of the Art Volume, 225-290.

[4] Schmidt, B. (1969). *"A method of estimating surface settlement above tunnels constructed in soft ground"*. Canadian Geotechnical Journal, 20: 11-22.

[5] Chou, W. I., and Bobet, A. (2002). *"Predictions of ground deformations in shallow tunnels in clay"*. Tunnelling and Underground Space Technology, 17: 3-19.

[6] Loganathan, N., and Poulos, H. G. (1998). *"Analytical prediction for tunneling-induced ground movements in clays"*. Journal of Geotechnical and Geoenvironmental Engineering, 124(9): 846-856.

[7] Park, K. H. (2005). *"Analytical solution for tunneling-induced ground movement in clays"*. Tunnelling and Underground Space Technology, 20(3): 249-261.

[8] Verruijt, A., and Booker, J. R. (1996). *"Surface settlement due to deformation of a tunnel in an elastic half plane"*. Geotechnique, 46(4): 753-756.

[9] Ercelebi, S. G., Copur, H., and Ocak, I. (2011). *"Surface settlement predictions for Istanbul Metro tunnels excavated by EPB-TBM"*. Environmental Earth Sciences, 62(2): 357-365. DOI: 10.1007/s12665-010-0530-6.

[10] Chakeri, H., and Unver, B. (2014). *"A new equation for estimating the maximum surface settlement above tunnels excavated in soft ground"*. Environmental Earth Sciences, 71: 3195-3210.

[11] Breiman, L. (2001). "*Random forest*". Machine Learning, 45(1): 5-32.