Research Article

Development of Machine Learning-based Methods to Reduce the Uncertainty of Tunneling Projects

Arsalan Mahmoodzadeh [1, a, *], Adil Hussein Mohammed [2, b], Hawkar Hashim Ibrahim [3, c], Shima Rashidi [4, d], Yusra Ahmed Salih [5, e]

[1] Rock Mechanics Division, School of Engineering, Tarbiat Modares University, Tehran, Iran
[2] Department of Communication and Computer Engineering, Faculty of Engineering, Cihan University-Erbil, Erbil, Kurdistan Region, Iraq
[3] Civil Engineering Department, College of Engineering, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq
[4] Department of Computer Science, College of Science and Technology, University of Human Development, Sulaymaniyah, Kurdistan Region, Iraq
[5] Department of Database Technology, College of Informatics, Sulaimani Polytechnic University, Sulaymaniyah Kurdistan Region, Iraq

[a] m.arsalan@modares.ac.ir, [b] adil.mohammed@cihanuniversity.edu.iq, [c] hawkar.ibrahim@su.edu.krd, [d] shima.rashid@uhd.edu.iq, [e] yusra.salih@spu.edu.iq

Access this article online		
Received on: 19 February 2022	**Accepted on:** 26 October 2022	**Published on: 27** December 2022
DOI: 10.25079/ukhjse.v6n2y2022.pp8-14	**E-ISSN:** 2520-7792	

Abstract
Having a good knowledge of the time and cost required to build a tunnel can be very important in reducing uncertainties related to the management of its construction. In this paper, using data obtained from the constructed parts of a tunnel, Gaussian process regression (GPR) method is developed to predict the time and cost of the non-constructed parts. Finally, by comparing the results predicted by the GPR model with the actual ones, it was concluded that the developed GPR model has a high potential to reduce uncertainties related to the time and cost of tunnel construction. Also, the ability of GPR model to predict time and cost of tunnel construction was compared with two other methods of support vector regression (SVR) and artificial neural networks (ANN). Finally, the GPR model was superior to the SVR and ANN methods in terms of prediction accuracy.

Keywords: Machine Learning, Gaussian Process Regression, GRP, Tunnel Construction, Time and Cost.

1. Introduction
Time and cost overruns usually encounter tunneling projects. Delays may affect the scope of tunneling projects negatively, leading to high costs (Ali & Wali, 2019; Xu et al., 2021). The use of contingencies and risk estimation at the project level usually do not show the numerous ambiguities in the tunnel project construction process (Li et al., 2021). Forecasting the construction duration and construction cost required for a tunneling project is a crucial task in determining the usefulness of a decision-making system (Liu et al., 2021).

There is always a lot of uncertainty in tunnel constructions due to unknown subsurface conditions (Ritter et al., 2013). Therefore, it cannot be expected that the uncertainties related to the tunnels, including the uncertainties related to construction time and construction costs, will be completely eliminated. However, it can be possible to determine the extent to which they are different. This can be made by any type of frequency diagram and probability distribution (Mahmoodzadeh & Zare, 2016; Mahmoodzadeh et al., 2022).

Several studies are conducted to minimize the uncertainties related to tunnels (Guan et al., 2014) such as soft computing tools (Rashid et al., 2019; Bezdan et al., 2021; Cuk et al., 2021), exploratory, or hard methods.

In recent years, good improvements in the management of time and cost of tunnels have been seen. Sousa and Einstein (2012) developed an approach based on Dynamic Bayesian Networks (DBN) to evaluate the expected costs and a tunnel collapse risk. Decision Aids for Tunneling (DAT) was developed by Einstein et al. (1999) to reduce the uncertainties of construction time and cost in tunneling projects.

Recently, artificial intelligence (AI) and machine learning (ML) techniques have shown their potential ability in the different engineering problems (Shamsaldin et al., 2019; Khishe & Mosavi, 2020). Mahmoodzadeh et al. (2021) presented an innovative methodology to predict the construction time and costs of tunneling projects based on ML techniques. They have shown the great ability of ML methods to reduce uncertainties regarding construction time and construction costs in tunneling.

In this paper, the Gaussian process regression (GPR) model is utilized to estimate the time and costs for a road tunnel. We chose GPR because:

- GPR directly captures the model uncertainty. As an example, in regression, GPR directly gives a distribution for the prediction value, rather than just one value as the prediction. This uncertainty is not directly captured in neural networks.
- When using GPR, we are able to add prior knowledge and specifications about the shape of the model by selecting different kernel functions. For example, based on the answers to the following questions we may choose different priors. Is the model smooth? Is it sparse? Should it be able to change drastically? Should it be differentiable? This capability gives researchers flexible models, which can be fit to various kind of datasets.

A comparison between the predicted results and the actual values was made to evaluate the performance prediction of the GPR model. The performance prediction of the GPR method has been compared by the support vector regression (SVR) and artificial neural networks (ANN). Also, three input parameters affecting the time and costs of tunnel construction: rock quality designation (RQD), groundwater, and Rock Mass Rating (RMR). The sensitivity analysis of these parameters on the time and cost of tunneling projects is investigated through mutual information test (MIT).

2. Methodology: GPR

GP is a nonparametric and Bayesian approach to regression that creates waves in ML. GP has several advantages, it works well on small data sets, and it has the ability to provide measurement of forecast uncertainty.

GP directly records the uncertainty of the model. In regression, for example, GP directly gives us a distribution of the forecasting value, not just a value as a prediction. This uncertainty is not recorded directly in neural networks.

When using the GPR method, we can add previous knowledge and specifications on the model shape by choosing various kernel functions. For example, we may choose different backgrounds according to the answers to the following questions. Is the model flat? Scattered? Should it be distinguishable? Should it be able to change drastically? This capability gives us flexibility models that can be adapted to different types of data sets.

GPR method has shown a good ability to solve various problems with a small amount of data. Since we often face data shortages in the forecasting tunnel time and costs and tunnel route geology, this method can be effective.

Concerning the parameter setting for the GPR model, we considered a squared exponential covariance function as it is presented in Equation (1). The kernel is very simple; it takes only two parameters: the length between the points and the average distance of them with the mean. We supposed to have a zero mean vector, so our distribution is around zero. Likewise, to match a signal with noise, we calculate their covariances as explained in Equation (1). We can identify the high correlations through the nearby points, leading to high covariances. On the other hand, while using the squared exponential, it is impossible to extrapolate more than ℓ units between any two points.

$$k(x_p, x_q) = \sigma_f^2 exp\left(-\frac{1}{2}(x_p - x_q)^T \, M(x_p - x_q)\right) = \sigma_f^2 exp\left(-\frac{1}{2}\sum_{i=1}^{D}\frac{(x_{p,i} - x_{q,i})^2}{l_i^2}\right) \tag{1}$$

where $M = diag(l)^{-2}$,$l = [l_1, l_2, \dots, l_D]^T$ $\theta = (\ln l_1, \ln l_2, \dots, \ln l_D, \ln\sigma_f]^T$ are the hyper-parameters.

The kernel typically contains hyper-parameters such as length-scale, signal variance, and noise variance. These are usually assumed to be unknown but rather are learned from the data. Hence, generally, the posterior distribution over the hyper-parameters is difficult to be obtained, the full Bayesian inference of the hyperparameters usually is not used. As an alternative, a point estimate of the hyperparameters is generally computed by maximizing the log-marginal likelihood. This is similar to the parameter estimation by maximum likelihood and is referred to as type-II maximum likelihood. Here, we tuned the hyperactive parameters using the automatic hyperactive parameter optimization method given in MATLAB 2018 with the 'fitrgp' function in which the five-fold cross-validation loss is minimized.

In the MATLAB software, four model types of rational quadratic, squared exponential, Matern 5/2, and exponential exist for the GPR method. That model type, which presented the most accurate predictions, is considered as the GPR predictions. Here, the model type of Matern 5/2 was the most accurate one. The other parameters considered in the GPR model through the MATLAB software's optimization mode are presented in Table (1).

Table 1. An overview of the GPR model parameters.

Parameter	Value or Type
KernelFunction (Form of the covariance function)	'Matern 5/2'
BasisFunction (Explicit basis in the GPR model)	'Constant'
Beta (Initial value of the coefficients for the explicit basis)	57.4223
Sigma (Initial value for the noise standard deviation of the Gaussian process model)	0.90568
FitMethod (Method to estimate parameters of the GPR model)	Exact Gaussian process regression

3. Data Preparation

To investigate the performance prediction of the GPR model in the tunnel construction time and construction cost, a road tunnel of 1900 m length in with a cross-section of 97 m² was considered. The tunnel's route Lithology is mainly composed of Sand, Shales, and Limestone. The top heading and benching method has been used to excavate the tunnel. The following support system is applied in the tunnel.

IPE 180 - Spacing 0.75-1.5 m

Rock bolts: Fully grouted, ɸ25 mm, L: 3-7 m

Shotcrete: 22 cm-Reinforced by 2 layer mesh ɸ6@100×100 mm

4. Results

In order to predict the construction time and cost of a tunnel route, we must first select the model inputs. These inputs must be selected from the parameters that affect time and cost. In this paper, three parameters of RMR, RQD, and groundwater are considered that have a great impact on the time and cost of tunnel construction. Then 350 data were obtained in the tunnel route. 300 data are used as training data and another 50 data as test data. Figure (1) shows the correlation matrix of these parameters.

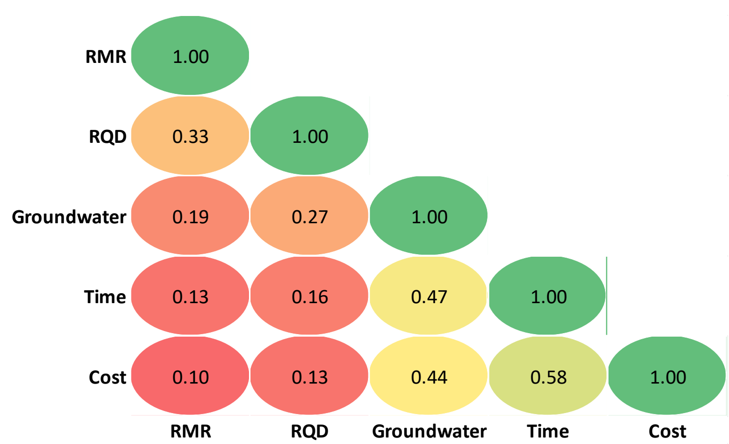


Figure 1. The correlation matrix of the datasets.

It should be noted that, a strong relationship exists between the cost and time of construction projects. But our goal in this article is that, by applying the pre-construction data such as type of machinery, type of support system, geotechnical data, geological data and other data that can be addressed before starting to construct a tunnel, both construction time and construction cost of the tunnel can be predicted with a good accuracy. But if we want to use the time parameter as one of the inputs of the cost forecasting model, for example, before we start constructing a tunnel, as there is high degree of uncertainty about construction costs, we have also a lot of uncertainty about the construction time of the tunnel. Using a parameter as the input of a model about which there is a lot of uncertainty and at the same time has a great impact on the output, can reduce the accuracy of the prediction model. So, in this paper, since we have assumed for the test tunnel that only the data before its construction are available and the construction time parameter is associated with a lot of uncertainties, the time parameter has not been considered as an input in the forecast model of tunnel construction cost. Of course, although the construction time parameter is not considered directly as an input parameter in the forecasting model, its effect is nonetheless indirectly applied. For example, one of the input parameters of the model is the groundwater status. More groundwater in the tunnel route means more time to construct and therefore more cost. Here, construction time has indirectly affected the cost of construction. In conclusion, the construction time parameter is one of the most effective parameters on the cost of tunnel construction, and if there is a little uncertainty about it, it is better to be considered in the cost forecasting model. But using it with high uncertainty can greatly reduce the accuracy of the tunnel construction cost forecasting model.

Figure (2) shows the estimated time and cost for the 50 test data and compared with the actual case. As shown in Figure (2), there is very little difference between the predicted results and the actual results. Therefore, it is possible to prove the high ability of the GPR model in predicting the time and costs of tunnel route construction. In Figure (3), the Taylor diagram shows the results predicted by the GPR model the other two models of SVR and ANN. As can be seen, all three models offer high prediction accuracy. But overall, the accuracy of the GPR model is higher.

The mutual information test (MIT) is applied for sensitivity analysis of the input parameters. As in Figure (4), all the input parameters have a great impact on the time and cost of tunnels. But the groundwater parameter has the higher impact on the construction time and construction cost than the RMR and RQD parameters.

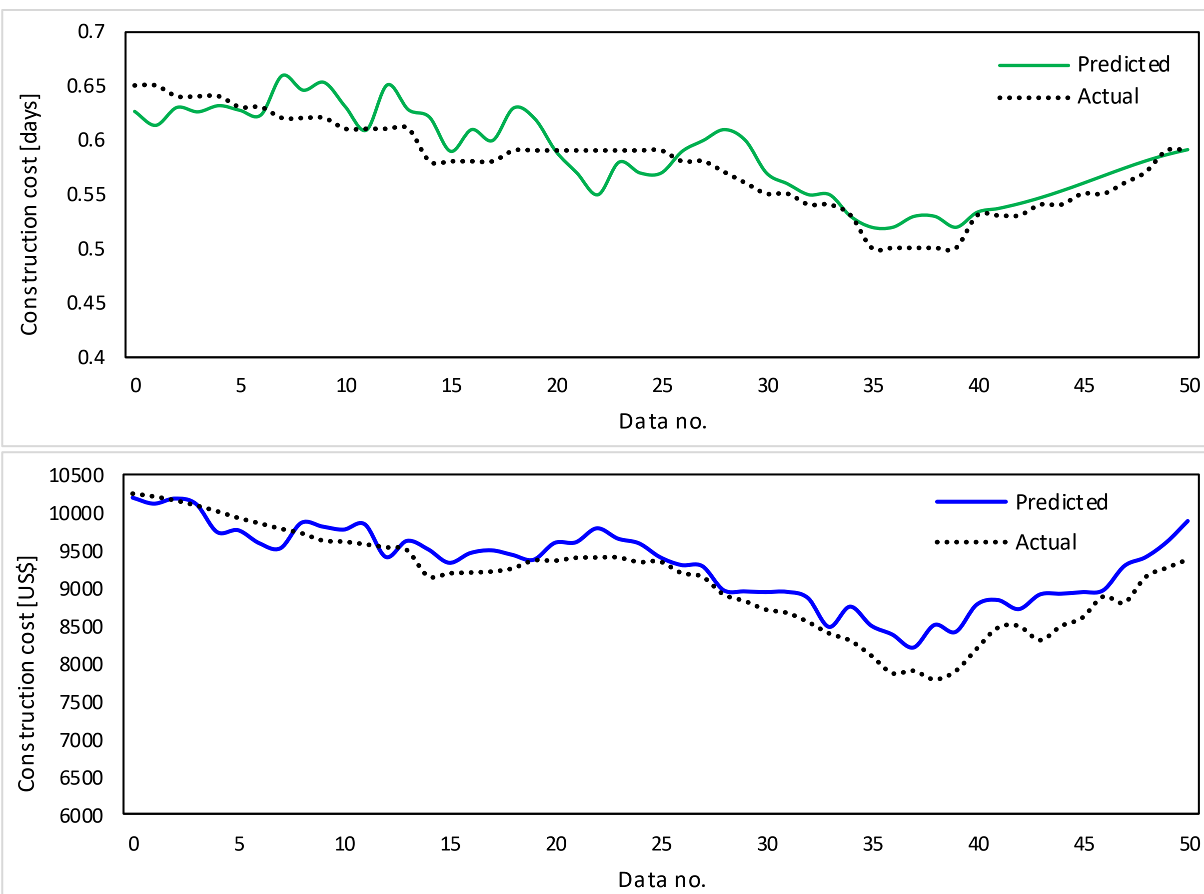
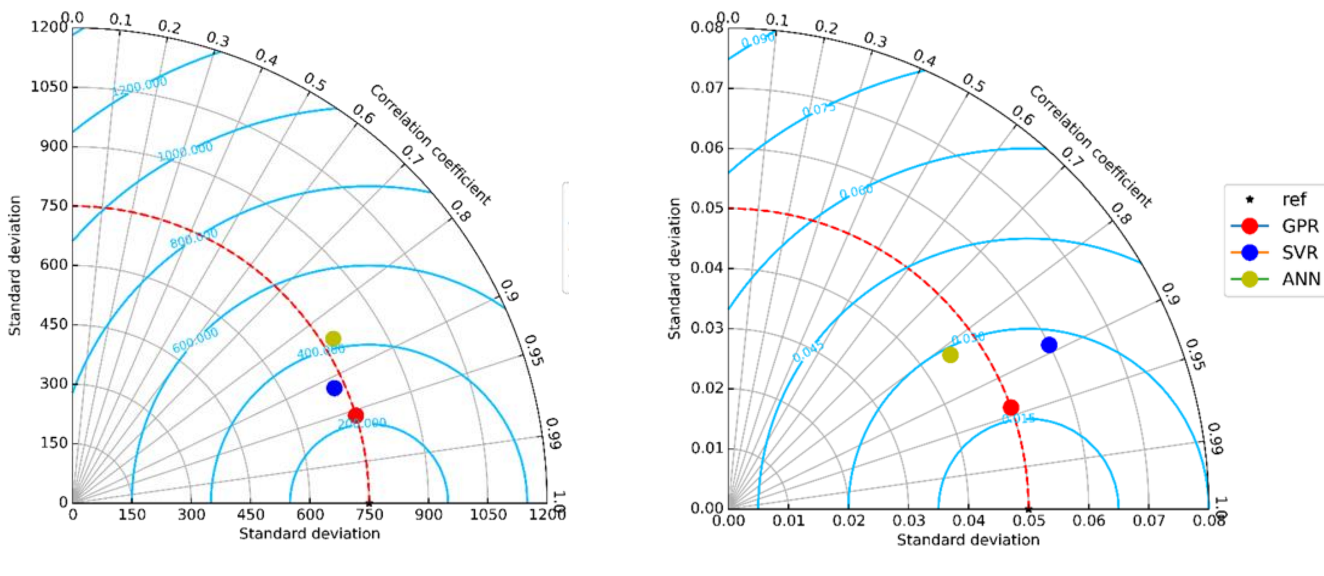


Figure 2. Comparison of AR model predictions with the actual time and cost of tunnel construction.



(a) (b)

Figure 3. Comparison of the ML performance prediction through Taylor diagram: (a) Construction cost, (b) Construction time.

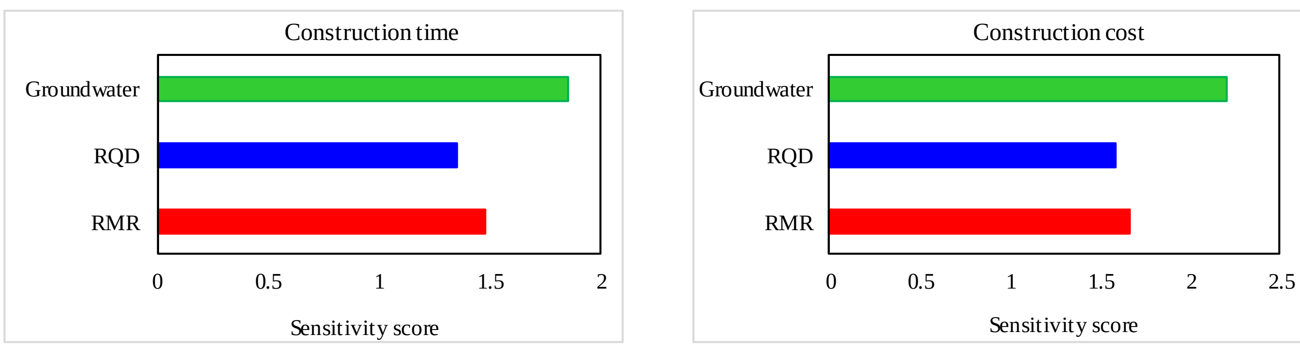


Figure 4. Sensitivity analysis of the input parameters on the time and cost of tunnel construction.

The construction performance measurement is a vexed issue. Despite much research effort, there is little agreement on how to measure and what to measure it. This issue is compounded by the tendency to evaluate the performance of national industry against the performance of other countries. Since the cost of construction is clearly part of the analysis, simply adjusting the cost data with an "international currency" has undermined past efforts to reach meaningful conclusions. The parameter that is important here is the construction efficiency. Construction efficiency is equal to the ratio of construction costs per one month and is used to comment on the relative performance of the procurement process at various locations.

In this paper, a model for predicting the construction cost of road tunnel projects was introduced. But it should be noted that the construction efficiency is also a very important parameter in the construction of tunnels. If we want that the proposed model by this study be suitable to comment on the relative performance of the tunnel construction process in different locations, as a suggestion in future works, it is better to provide a model for predicting the construction efficiency of tunneling projects.

5. Conclusions

The unknown subsurface conditions in tunneling projects has led to their management with many uncertainties. In this study, a GPR model was used to reduce the uncertainties related to construction time and cost in tunnelling projects. 350 datasets including three input parameters of RMR, RQD, and groundwater, and two output parameters of construction time and construction cost were applied in the GPR model. 300 datasets were used for training and 50 datasets for test.

The MIT method was used to investigate the sensitivity analysis of the parameters. The obtained results in this study led to the following conclusions:

- The performance prediction of the proposed GPR model in the prediction of construction time and cost of tunnels is very high.
- The performance prediction of the SVR and ANN to predict the tunnels' time and cost was also high but less than the GPR model.
- Three input parameters of RMR, RQD, and groundwater considered in the datasets had the significant effect on the construction time and construction cost.
- The sensitivity analysis using the MIT method showed that, among three parameters of RMR, RQD, and groundwater, the groundwater parameter has the most impact on the time and cost of tunnel construction.
- This work's significance is that it allows geotechnical engineers to determine the usefulness of a decision-making system in a tunneling project.
- Due to the great importance of construction efficiency in the engineering projects, especially in tunnelling, it was suggested to develop the model proposed by this article to predict the construction efficiency of tunnels in future studies.

Given that there are different types of tunnels, and depending on the type of tunnel, the parameters affecting the construction time and costs can be different, the investigation of the GPR model capability presented in this article on the different types of tunnels such as urban tunnels is suggested.

References

Mahmoodzadeh, A. & Zare, S. (2016). Probabilistic prediction of the expected ground conditions and construction time and costs in road tunnels. *Journal of Rock Mechanics and Geotechnical Engineering*, 8(5), 734-745.

Mahmoodzadeh, A., Mohammadi, M., Abdulhamid, S., Ibrahim, H., Hama-Ali, H. & Salim, S. (2021). Dynamic reduction of time and cost uncertainties in tunneling projects. *Tunnelling and Underground Space Technology*, 109, 103774.

Mahmoodzadeh, A., Rashidi, S., Mohammed, A., Hama Ali, H. & Ibrahim, H. (March 2022). *Machine Learning Approaches to Enable Resource Forecasting Process of Road Tunnels Construction*. 4TH INTERNATIONAL CONFERENCE ON COMMUNICATION ENGINEERING AND COMPUTER SCIENCE (CIC-COCOS'22). CIHAN UNIVERSITY-ERBIL, Erbil, Iraq. pp. 86-89. URL: https://conferences.cihanuniversity.edu.iq/index.php/COCOS/22/paper/view/718/203.

Shamsaldin, A., Fattah, P., Rashid, T. & Al-Salihi, N. (2019). A Study of The Convolutional Neural Networks Applications. *UKH Journal of Science and Engineering*, 3(2), 31-40.

Cuk, A.B., Bezdan, T., Bacanin, N., Zivkovic, M., Venkatachalam, K., Rashid, T. & Devi, V.K. (2021). Feedforward Multi-Layer Perceptron Training by Hybridized Method between Genetic Algorithm and Artificial Bee Colony. *Data Science and Data Analytics* (1st ed.), ImprintChapman and Hall/CRC, pp. 14, eBook ISBN: 9781003111290.

Li, B., Fu, Y., Hong, Y. & Cao, Z. (2021). Deterministic and probabilistic analysis of tunnel face stability using support vector machine. *Geomechanics and Engineering*, 25(1), 17-30.

Einstein, H.H., Indermitte, C., Sinfield, J., Descoeudres, F.P. & Dudt, J.P. (1999). Decision aids for tunneling. *Transportation Research Record: Journal of Transportation Research Board*, 1656(1), 6–13. DOI: https://doi.org/10.3141/1656-02.

Liu, J., Jiang, Y., Zhang, Y. & Sakaguchi, O. (2021). Influence of different combinations of measurement while drilling parameters by artificial neural network on estimation of tunnel support patterns. *Geomechanics and Engineering*, 25(6), 439-454.

Khishe, M. & Mosavi, M.R. (2020). Chimp optimization algorithm. *Expert Systems with Applications*, 149, 113338. DOI: https://doi.org/10.1016/j.eswa.2020.113338.

Ali, N. & Wali, K.I. (2019). Evaluation of Construction Delay of Public Projects in Erbil Governorate. *Cihan University-Erbil Scientific Journal*, 3(1), 18-26. URL: http://eprints.cihanuniversity.edu.iq/33/1/Article_CUESJ_13-05-2019.pdf.

Sousa, R.L. & Einstein, H.H. (2012). Risk analysis during tunnel construction using Bayesian Networks: Porto Metro case study. *Tunnelling and Underground Space Technology*, 27(1), 86–100.

Ritter, S., Einstein, H.H. & Galler, R. (2013). Planning the handling of tunnel excavation material – A process of decision making under uncertainty. *Tunnelling and Underground Space Technology*, 33, 193-201.

Bezdan, T., Stoean, C., Naamany, A.A., Bacanin, N., Rashid, T., Zivkovic, M. & Venkatachalam, K. (2021). Hybrid Fruit-Fly Optimization Algorithm with K-Means for Text Document Clustering. *Mathematics*, 9(16).

Rashid, T., Hassan, M.K., Mohammadi, M. & Fraser, K. (2019). Improvement of Variant Adaptable LSTM Trained With Metaheuristic Algorithms for Healthcare Analysis. *Advanced Classification Techniques for Healthcare Analysis*, 111-131.

Guan, Z., Deng, T., Jiang, Y., Zhao, C. & Huang, H. (2014). Probabilistic estimation of ground conditions and construction cost for mountain tunnels. *Tunnelling and Underground Space Technology*, 42, 175–183.

Xu, Z., Xian, M., Li, X., Zhou, W., Wang, J., Wang, Y. & Chai, J. (2021). Risk assessment of water inrush in karst shallow tunnel with stable surface water supply: Case study. *Geomechanics and Engineering*, 25(6), 495-508.