Geomechanics and Engineering, *Vol. 28, No. 1 (2022) 65-75*
https://doi.org/10.12989/gae.2022.28.1.065

Prediction of duration and construction cost of road tunnels using Gaussian process regression

Arsalan Mahmoodzadeh[1,2,a], Mokhtar Mohammadi[*3], Sazan Nariman Abdulhamid[4b], Hawkar Hashim Ibrahim[4c], Hunar Farid Hama Ali[2d], Hamid Reza Nejati[1e] and Shima Rashidi[5f]

[1]Rock Mechanics Division, School of Engineering, Tarbiat Modares University, Tehran, Iran
[2]Department of Civil Engineering, University of Halabja, Halabja, Kurdistan Region, Iraq
[3]Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Kurdistan Region, Iraq
[4]Department of Civil Engineering, College of Engineering, Salahaddin University-Erbil, 44002 Erbil, Kurdistan Region, Iraq
[5]Department of Computer Science, College of Science and Technology, University of Human Development, Sulaymaniyah, Kurdistan Region, Iraq

(Received August 17, 2021, Revised November 16, 2021, Accepted December 1, 2021)

Abstract. Time and cost of construction are key factors in decision-making during a tunnel project's planning and design phase. Estimations of time and cost of tunnel construction projects are subject to significant uncertainties caused by uncertain geotechnical and geological conditions. The Gaussian Process Regression (GPR) technique for predicting ground condition and construction time and cost of mountain tunnel projects is used in this work. The GPR model is trained with data from past mountain tunnel projects. The model is applied to a case study in which the predicted time and cost of tunnel construction using the GPR model are compared with the actual construction time and cost for model validation and reducing the uncertainty for the future projects. In addition, the results obtained from the GPR have been compared with to other models of artificial neural network (ANN) and support vector regression (SVR) that the GPR model provides more accurate results.

Keywords: construction cost; construction time; Gaussian process regression; ground conditions; tunneling

1. Introduction

Tunneling, similar but more so than other geotechnical endeavors, is characterized by the influence of uncertainty. Ground condition and time and cost of construction are key factors in decision making during the planning and design phase of a tunnel project. Owners, planners, designers, contractors and other parties involved in tunneling need to consider the uncertainties of ground conditions since they affect tunnel construction time and cost and the required and produced resources (Mahmoodzadeh *et al.* 2021a). Although most tunnel construction projects have been completed safely, several incidents in various tunneling projects have resulted in delays, cost overruns, and in a few cases more significant consequences such as injury and loss of life. Therefore, it is essential to systematically assess and manage the risks associated with tunnel construction (Sousa and Einstein 2012). In subsurface projects, the ground conditions are not known before construction and must be inferred based on general information describing the geologic formations near the project and on location-specific observations provided by sub-surface exploration programs (Leu and Adi 2011).

Many studies have been developed to predict and minimize ground conditions and tunnel alignment uncertainties (Guan *et al.* 2014). One of these methods is to use intelligent techniques such as Artificial Neural Network (ANN), time series and random process approaches which utilize mathematical models to predict the geological and geotechnical conditions along the tunnel alignment. Pilot drilling and advanced geophysical prospecting utilize in situ equipment to detect the geological and geotechnical information in some specific locations along tunnel path (Carrière *et al*. 2013). Hidden Markov Models (HMMs) are applied to analyze problems with uncertainty in transportation engineering (Mahmoodzadeh and Zare 2016, Mahmoodzadeh *et al*. 2019, Mahmoodzadeh *et al.* 2021b). Decision Aids for Tunneling (DAT) is a tool for probabilistic prediction of construction time and costs, and consumption of resources (Min and Einstein 2016), and other geological and geotechnical engineering problems.

The uncertainty in construction time and cost estimates

∗Corresponding author, Ph.D.
E-mail: mukhtar@lfu.edu.krd
[a]M.Sc.
E-mail: arsalan.mahmoodzadeh@uoh.edu.iq
[b]M.Sc.
E-mail: sazan.abdulhamid@su.edu.krd
[c]M.Sc.
E-mail: hawkar.ibrahim@su.edu.krd
[d]M.Sc.
E-mail: hunar.hamaali@uoh.edu.iq
[e]Ph.D.
E-mail: h.nejati@modares.ac.ir
[f]M.Sc.
E-mail: shima.rashid@uhd.edu.iq

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

results from the common variability of the construction performance and from the occurrence of extraordinary events (also denoted here as failures of the construction process) such as tunnel collapses (Mahmoodzadeh *et al.* 2021b). Risks resulting from construction failures are commonly analyzed separately using techniques such as fault tree or event tree analysis, decision trees or risk matrixes (Jurado *et al.* 2012). Sousa and Einstein (2012) present a Dynamic Bayesian Networks (DBN) model, which estimates the expected utility as a sum of the expected costs and the risk of a tunnel collapse. The full probability distribution of the construction costs, however, is not assessed. Some models allow one to probabilistically estimate the time or costs without taking into account the occurrence of extraordinary events. They typically use Monte Carlo (MC) simulation (Mahmoodzadeh *et al.* 2021a). Full probabilistic estimates of tunnel construction time or cost, considering both the common variability and the risk of extraordinary events, are presented (Mahmoodzadeh *et al*. 2019).

Recently, machine learning methods have shown a potential ability to solve various problems (Zhang and Xu 2021a, Zhang and Xu 2021b, Alade *et al*. 2021, Zhang and Xu 2020, Saffari *et al.* 2022, Taghavi and Kishe 2019, Kishe *et al.* 2018, Kishe and Mosavi 2019, Kishe and Mosavi 2020). These methods have been applied as a solution for different geotechnical problems such as prediction of geological conditions (Bai *et al.* 2021), settlement prediction of shallow foundations (Luat *et al*. 2020), tunnel surface settlement prediction (Mahmoodzadeh *et al.* 2020a), estimation of tunnel support patterns (Liu *et al.* 2021a), tunnel resources prediction (Mahmoodzadeh *et al.* 2020b), landslide susceptibility assessment (Liu *et al.* 2021b), soft soil flow prediction (Xiang *et al*. 2021), and many other geotechnical engineering problems (Mahmoodzadeh *et al.* 2021c,d,e,f,g,h).

The objective of this paper is to present the new intelligence approach to predicting ground condition and time and cost of construction that can be used as a basis for developing more effective decision support systems for tunneling design and reducing the uncertainties of construction time and costs during the planning and design phase of a tunnel project. For this purpose, the nonlinear Gaussian Process Regression (GPR), the nonlinear Support Vector Regression (SVR) and the Artificial Neural Network (ANN) are applied to develop a model for predicting ground conditions and construction time and costs. To verify its feasibility, these techniques as well as the other intelligence techniques, are applied to a tunnel in Iran to predict ground condition and time and costs of construction. Finally, for validation purposes, the predicted construction time and costs of the GPR, SVR and ANN techniques have been compared to the actual mode and together.

Despite the merits of various machine learning approaches, according to the No-Free-Lunch (NFL) theorem, there is no machine learning model to solve all engineering problems as the best method successfully. Therefore, researchers have tried to evaluate the efficiency of various machine learning approaches for solving various optimization. As the NFL theorem, the GPR model is used to predict the duration and construction time of a tunnel path. The main advantages and disadvantages of the GPR model are as follows:

Advantages:

- The prediction interpolates the observations (at least for regular kernels).
- The prediction is probabilistic (Gaussian). One can compute empirical confidence intervals and decide based on those if one should refit (online fitting, adaptive fitting) the prediction in some region of interest.
- Versatile: different kernels can be specified. Common kernels are provided, but it is also possible to specify custom kernels.

Disadvantages:

- They are not sparse, i.e., they use the whole samples/features information to perform the prediction.

They lose efficiency in high dimensional spaces – namely when the number of features exceeds a few dozens.

2. Machine learning models

2.1 GPR

GPRs are non-parametric. They belong to the family of probabilistic models (Zhang and Xu 2021c, Zhang and Xu 2021d). GPR models are based on the assumption that adjacent observations should convey information about each other. GPs are a way of specifying a prior directly over function space. This is a natural generalization of the Gaussian distribution, whose mean and covariance are vector and matrix. The Gaussian distribution is over vectors, whereas the Gaussian process is over functions. Thus, due to prior knowledge about the data and functional dependencies, no validation process is required for generalization. GPR models can understand the predictive distribution corresponding to the test input (Mahmoodzadeh *et al.* 2021c).

Probabilistic regression is usually formulated as follows: given a training set $\mathcal{D} = \{(x_i, y_i), i = 1, \ldots, n\}$ of n pairs of (vectorial) inputs x_i and noisy (real, scalar) outputs y_i, compute the predictive distribution of the function values f^* (or noisy y^*) at test locations x^*. In the simplest case (which we deal with here) we assume that the noise is additive, independent and Gaussian, such that the relationship between the (latent) function f(x) and the observed noisy targets y are given by (Mahmoodzadeh *et al*. 2021d)

$$y = f(x) + \epsilon \qquad , \qquad \epsilon \sim N(0, \sigma^2) \tag{1}$$

where, f(x) represents an arbitrary regression function while ϵ is the noise follows an independent, identically distributed Gaussian distribution with zero mean (usually, for notational simplicity we will take the mean function to be zero and the offsets and simple trends can be subtracted out before modeling) and variance of the noise σ^2. The symbol $\sim$ in statistics means sampling for.

A Gaussian process f(x) can be represented as: $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$, where $m(x)$ is the mean and $k(x, x')$ is the covariance (or kernel) function evaluated at x and x′.

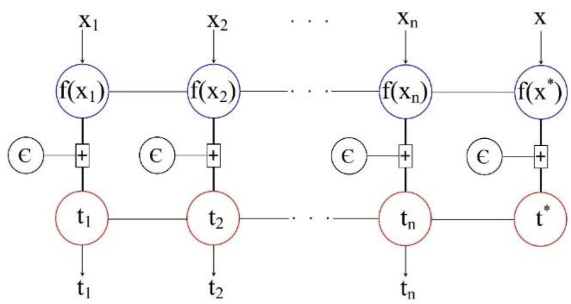

Fig. 1 Graphical model for GP

Furthermore, we assume that f = [f(x$_1$), f(x$_2$), …, f(x$_n$)]^T behaves according to a Gaussian process, that is P(f | X = *N*(0,K)), where K is the covariance matrix with element $K_{ij} = k(x_i,x_j)$ (Mahmoodzadeh *et al.* 2021e).

$$K(X,X) = \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n) \end{bmatrix} \quad (2)$$

The element K_{ij} is the covariance between values of the latent functions f(x$_i$) and f(x$_j$), and it encodes about the prior of our knowledge of nonlinear process among latent functions.

Gaussian process regression (GPR) is used to compute the predictive distribution of the function values f* at test points X* = [x$_1^*$, x$_2^*$, ..., x$_m^*$]. A graphical model representation of a GP is given in Fig. 1 (circled variables are random variables, non-circled ones are observed/given). In the figure, for given {x$_1$, x$_2$, …, x$_n$, x*}, the corresponding set of random variables {f(x$_1$), f(x$_2$), …, f(x$_n$), f(x*)} have a joint multivariate Gaussian distribution. With the additive independent white noise assumption, the corresponding circled {t$_1$, t$_2$, …, t$_n$, t*} are also jointly normally distributed. The marginal part of the joint gives us the probability of the data {t$_1$, t$_2$, …, t$_n$} and the posterior predictive distribution of f(x*) (or equivalently t*) corresponding to x* is obtained by conditioning on the data and x* (Mahmoodzadeh *et al.* 2021f).

Now, suppose we have collected observations $\mathcal{D}_t$ = {X$_t$,y$_t$} and we want to make predictions for new inputs X* by drawing f* from the posterior distribution p(f | $\mathcal{D}_t$). By definition, previous observations y$_t$ and function values f* follow a joint (multivariate) normal distribution. This distribution can be written as (Mahmoodzadeh *et al*. 2021g)

$$\begin{bmatrix} y_t \\ f^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X_t,X_t) + \sigma_\epsilon^2 I & K(X_t,X^*) \\ K(X^*,X_t) & K(X^*,X^*) \end{bmatrix}\right) \quad (3)$$

where I is an identity matrix (with 1's on the diagonal, and 0 elsewhere) and σ_ϵ^2 is the assumed noise level of observations (i.e., the variance of ∈). Deriving the conditional distribution corresponding to f(x) ~ $\mathcal{GP}(m(x), k(x,x'))$, we arrive at predictive distribution the key predictive equations for GPR in Eqs. (4) and (5). The output estimates are made not only with an expected prediction (mean) of latent function f* but also with a measure of uncertainty (variance) (Mahmoodzadeh *et al*. 2021h).

$$\bar{f}^* = K(X^*,X)[K(X,X) + \sigma^2 I]^{-1} y \quad (4)$$

$$cov(f^*) = K(X^*,X^*) - K(X^*,X)[K(X,X) + \sigma^2 I]^{-1} K(X,X^*) \quad (5)$$

The most commonly-used squared exponential covariance function with automatic relevance determination distance measure was adopted.

$$k(x_p,x_q) = \sigma_f^2 exp\left(-\frac{1}{2}(x_p - x_q)^T M (x_p - x_q)\right) = \sigma_f^2 exp\left(-\frac{1}{2}\sum_{i=1}^{D}\frac{(x_{p,i} - x_{q,i})^2}{l_i^2}\right) \quad (6)$$

where M = diag(l) $^{-2}$, $l = [l_1, l_2, …, l_D]^T$. θ = (ln l_1, …, ln l_D, ln σ$_f$) are the hyper-parameters.

The kernel contains hyper-parameters such as the length-scale, the signal variance, and the noise variance typically. These are usually not assumed to be known but rather are trained from the data. As the posterior distribution over the hyper-parameters is generally difficult to obtain, full Bayesian inference of the hyper-parameters is generally not used. Instead, a point estimate of the hyper-parameters is usually computed by maximizing the log-marginal likelihood. This is similar to parameter estimation by maximum likelihood and is also referred to as type-II maximum likelihood.

Given the data S = {X,y} and hyper-parameters θ, the log-marginal likelihood is

$$\text{logp}(y|X,\theta) = -\frac{1}{2}y^T(K_y)^{-1}y - \frac{1}{2}log|K_y| - \frac{n}{2}log2\pi \quad (7)$$

where $K_y = K(X,X) + \sigma_\epsilon^2 I$ is the covariance matrix of the noisy output values y. The log-marginal likelihood can be viewed as a penalized t measure, where the term $-1/2\ y^T K_y^{-1} y$ measures the data fit, that is how well the current kernel parametrization explains the dependent variable, and $-1/2 \log|K_y|$ is a complexity penalization term. The final term $-n/2 \log 2\pi$ is a normalization constant.

2.2 SVR

Vapnik (1995) modified his first version model (ε-support vector regression, SVR) by changing the ε-insensitive loss function. This modification permits the SVR model to use the margin idea in the regression process. Margin in the modified model can be described by the summation of the distances of the hyperplane from the closest points of two classes. Minimizing errors between the actual training data and the hyperplane are the main target of the SVR. Kernel function idea has been introduced by Vapnik (1995) for non-linear SVR. Readers are directed to Vapnik (1995) to understand more about SVR. To learn more about the SVR method, researchers can refer to other sources such as Alade *et al.* (2021).

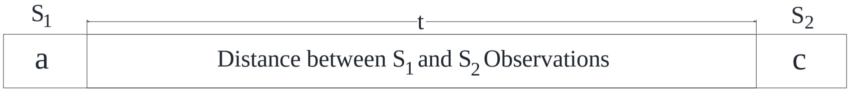

Fig. 2 The position of the two observations S_1 and S_2 at the distance t from each other, where the modes of the parameter *X* are a and c respectively

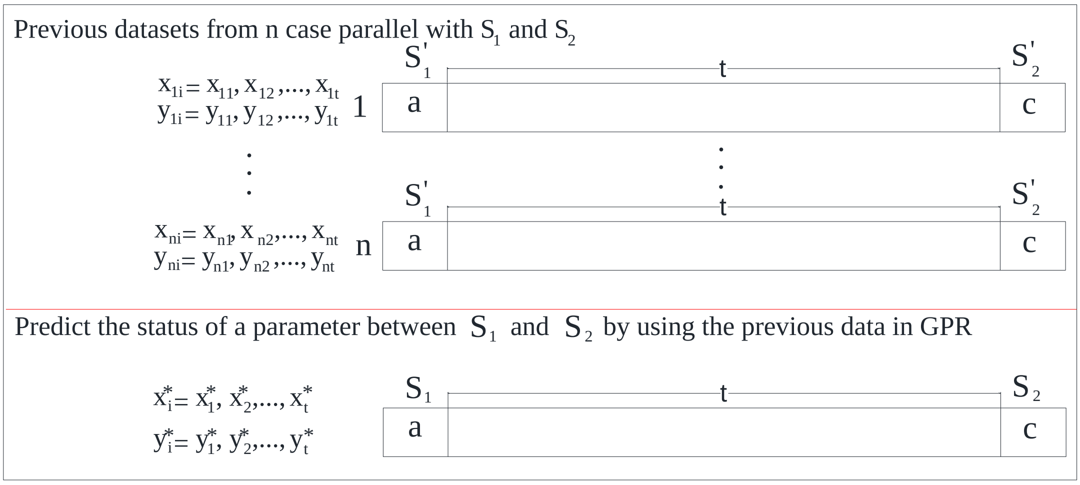

Fig. 3 Presentation of accessing the GPR input data from previously excavated tunnels

2.3 ANN

An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron receives a signal then processes it and can signal neurons connected to it. The "signal" at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer) to the last layer (the output layer), possibly after multiple layers traversing.

3. GPR model of ground condition prediction

As noted in the previous section, GPR requires inputs to train to help them make predictions. These inputs can vary depending on the type of project and the type of problem. Since the main focus of this paper is on the prediction of tunnel alignment, different geological or geotechnical parameters (e.g., RMR, RQD, lithology, underground water conditions, etc).

This paper considers the distances from the tunnel input as inputs *(x_i)* and the status of a geological or geotechnical parameter as the outputs *(y_i)*. For example, the Rock Quality Designation, can be varied from zero to 100 in terms of earth's conditions, that is, the same states for RQD.

The initial number of data *(x_i, y_i)* and the borehole in a tunnel may not be sufficient to be used as the initial inputs for the GPR. We need more inputs for training, so the more they are, GPR provides more accurate predictions. For this reason, the technique used in this paper is that in addition to using the initial data on the tunnel, it uses the information of the previous tunnels that are excavated and their data are available after drilling.

Stages of forecasting the status of a geological or geotechnical parameter are given below:

1) According to Fig. 2, two observational positions S_1 and S_2 are considered, which are essential to predict the state of the parameter *X* at the distance *t.*
2) Determine the status of the parameter *X* in the positions S_1 and S_2, as shown in Fig. 2, which are *a* and *c* respectively.
3) In other previously excavated tunnels, we look for two positions, such as, S_1 and S_2, spaced at or near distance *t*. S_1 is similar to S'_1 and S_2 is the same position as S'_2. For a similar situation, for example, if in position S_1 the condition of the parameter *X* is *a*, then it should be in the same position as S'_1.
4) In the interval between the positions S'_1 and S'_2, for each x_i we specify y_i values (or parameter status *X*). The data obtained between the two positions S'_1 and S'_2 is called a dataset. Similarly, we can look at the number of *n* sets of data from other positions, such as S'1 and S'2 in the previously excavated tunnels and as shown in Fig. 3, we obtain the values of each case input of the GPR model.
5) As shown in Fig. 3, we put all *n*s of S'_1 and S'_2 parallel to S_1 and S_2, and each of *n* sets of data obtained is considered as the input of the GPR model in the interval between S_1 and S_2.

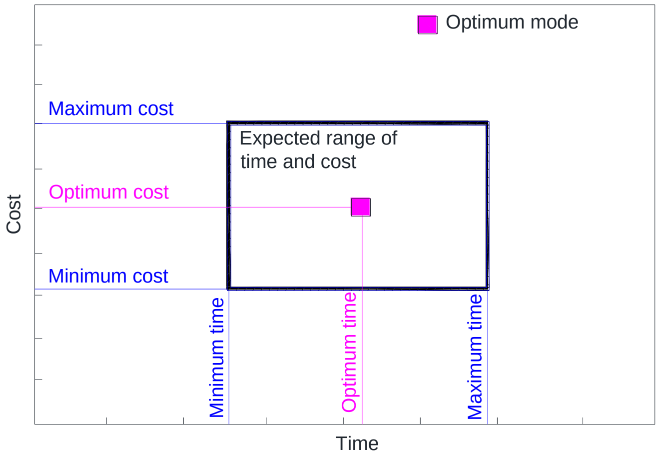

Fig. 4 Display the expected range and optimal mode of the construction time and cost

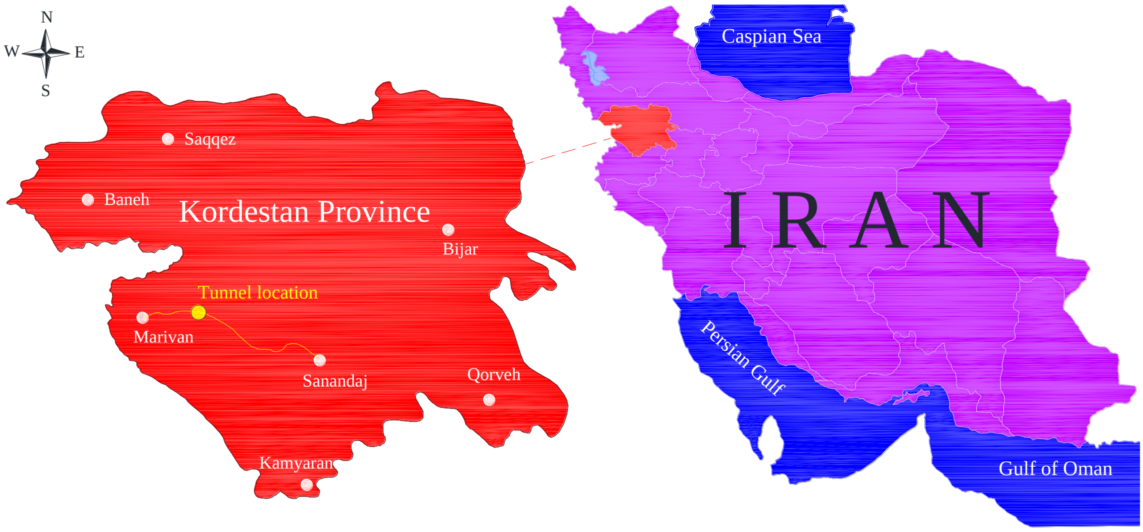

Fig. 5 Project location of the Garan tunnel

6) After obtaining input data and applying them to the GPR, for each x_i^* the corresponding y_i^* will be predicted in-between S_1 and S_2. In the following, for any distance in the tunnel route, we act in the manner described above to predict the probability of occurrence of the parameter states in other tunnel sections.

4. GPR model of time and cost of construction prediction

With the experience of previous tunneling projects in the same conditions, for each of the five ground classes specified in the previous section, we will determine the maximum and minimum time and cost of running each construction meter. Next, according to the last section, which indicates how many meters of the tunnel route is to be found in each ground class, can obtain the maximum and minimum time and cost of building the whole tunnel. According to the time and cost graph in Fig. 4, we determine the maximum and minimum time and cost of the total construction on the axes. In the maximum and minimum positions of the time, two lines parallel to the cost axis, as well as in the maximum and minimum positions of costs, we draw two lines parallel to the time axis. The interconnection of these four lines will be a quadrangle, which will be considered the expected range of time total costs for constructing the tunnel. Also, as an optimal mode, the center can be considered for the time and cost of the entire construction in the early stages of planning.

5. Engineering application

5.1 Engineering background

The present study is undertaken on Garan tunnel of 1900 m length and cross-section of 97 m^2 as part of the under-construction Sanandaj-Marivan road. In Fig. 5 the location of the Garan tunnel is shown.

The Lithology of the tunnel route is mainly composed of Sand Shales, Shale and Limestone. In Fig. 6, the geological profile of the Garan tunnel has been shown.

Garan tunnel has been excavated using top heading and benching method. The support system used in the Garan tunnel construction is as follows.

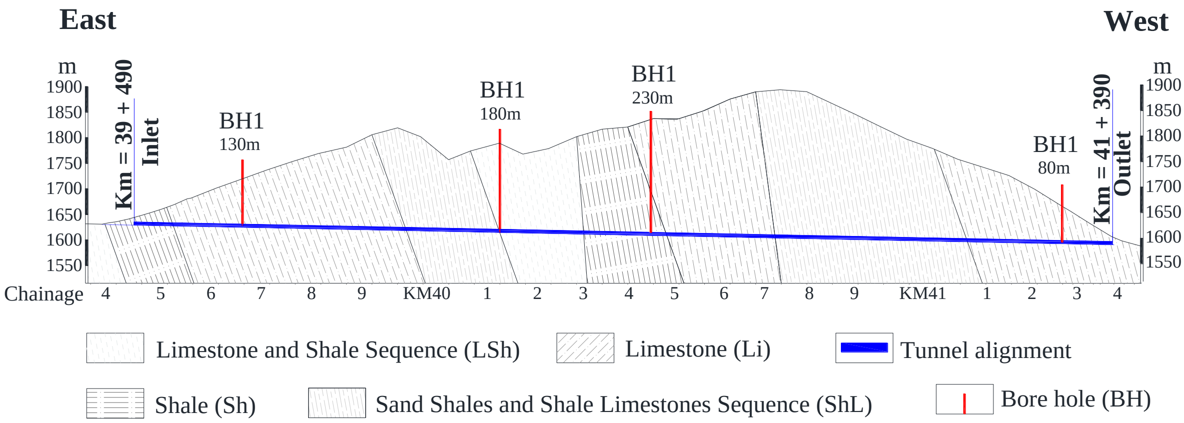

Fig. 6 Geological map of Garan mountain tunnel

Table 1 Input data to train the GPR model to predict the RMR parameter state between entrance and borehole 1

Tunnel chainage (m)	RMR							
SM39 + 490 (Entrance)	15-25	15-25	15-25	15-25	15-25	1 15-25	15-25	15-25
SM39 + 510	15-20	20-25	15-20	15-25	15-25	10-20	15-25	15-25
SM39 + 530	15-20	25-30	15-25	15-25	15-20	10-15	15-25	25-30
SM39 + 550	20-25	25-30	10-20	25-30	15-30	05-15	15-20	25-30
SM39 + 570	20-25	30-40	10-25	15-20	30-35	10-20	20-30	25-35
SM39 + 590	20-25	30-40	10-15	10-15	30-40	15-20	20-25	20-25
SM39 + 610	25-30	40-45	15-25	05-15	40-55	15-35	20-25	15-20
SM39 + 630	25-30	35-45	15-20	15-20	40-55	20-30	30-35	15-25
SM39 + 650	20-30	35-40	20-30	15-20	40-50	20-25	25-35	25-35
SM39 + 670	25-35	35-40	25-35	20-30	35-45	25-30	30-40	25-30
SM39 + 690	20-35	30-35	25-35	25-35	30-35	25-30	30-40	30-35
SM39 + 700 (BH1)	25-35	25-35	25-35	25-35	25-35	25-35	25-35	25-35

Table 2 Input data to train the GPR model to predict the RMR parameter state between entrance and borehole 1

Tunnel chainage (m)	RMR							
SM39 + 700 (BH1)	25-35	25-35	25-35	25-35	25-35	25-35	25-35	25-35
SM39 + 720	30-35	30-35	30-40	35-45	30-35	30-40	25-35	20-35
SM39 + 740	30-35	30-35	30-40	45-55	30-35	30-35	25-40	15-30
SM39 + 760	30-40	30-35	30-35	40-55	30-40	25-35	25-45	15-20
SM39 + 780	30-35	25-30	30-35	40-45	30-40	30-40	20-35	05-15
SM39 + 800	30-45	25-30	20-35	40-55	30-35	30-40	20-30	05-15
SM39 + 820	40-45	30-35	25-30	50-65	30-35	35-45	25-30	05-20
SM39 + 840	30-40	30-40	25-35	50-60	35-40	35-40	25-35	05-10
SM39 + 860	30-35	30-35	25-35	60-65	25-35	25-35	25-40	05-10
SM39 + 880	25-30	25-35	20-25	60-65	25-30	25-30	15-25	10-20
SM39 + 900	25-35	25-35	20-30	60-70	20-30	25-30	15-20	10-15
SM39 + 920	20-25	25-30	15-25	65-70	20-35	20-30	15-25	15-20
SM39 + 940	20-25	30-40	15-25	65-75	20-30	20-30	10-15	15-20
SM39 + 960	20-30	30-40	15-20	65-70	25-30	20-35	10-15	10-20
SM39 + 980	20-35	30-35	10-15	35-60	15-25	15-30	10-20	10-25
SM40 + 000	25-30	20-25	15-20	35-45	15-20	15-25	10-15	20-25
SM40 + 020	25-30	20-30	05-15	35-40	15-20	15-20	15-20	20-25
SM40 + 040	20-25	20-35	15-25	30-40	15-25	15-25	15-20	20-30
SM40 + 060	15-20	20-25	20-25	25-30	20-25	20-25	10-20	20-30
SM40 + 080	15-25	15-25	25-35	25-35	20-30	20-25	10-25	20-30
SM40 + 100	20-25	15-20	20-30	20-30	25-35	15-20	15-20	20-35
SM40 + 120	20-30	20-30	20-35	15-25	25-30	15-25	15-25	20-25
SM40 + 140	20-25	20-25	20-30	15-20	25-30	20-30	15-25	25-30
SM40 + 160	20-30	20-25	15-25	15-25	25-30	25-35	20-25	20-30
SM40 + 180	20-25	15-20	20-30	20-25	20-30	25-30	20-25	20-30
SM40 + 200 (BH2)	20-25	20-25	20-25	20-25	20-25	20-25	20-25	20-25

Table 3 Input data to train the GPR model to predict the RMR parameter state between borehole 2 and borehole 3

Tunnel chainage (m)	RMR							
SM40 + 200 (BH2)	20-25	20-25	20-25	20-25	20-25	20-25	20-25	20-25
SM40 + 220	20-35	15-20	20-25	20-30	25-35	10-25	20-30	15-25
SM40 + 240	20-30	15-25	20-30	30-45	25-35	10-20	20-30	15-30
SM40 + 260	20-25	15-25	20-35	35-45	30-35	10-15	25-35	15-20
SM40 + 280	20-25	20-30	15-30	35-40	30-40	10-15	25-30	20-25
SM40 + 300	25-35	20-30	15-25	30-40	20-35	05-15	25-35	20-35
SM40 + 320	35-50	20-25	20-25	35-45	20-30	05-15	20-35	20-30
SM40 + 340	45-65	15-25	20-30	35-50	20-30	10-20	20-35	25-30
SM40 + 360	45-55	20-35	25-35	40-45	25-30	10-25	30-40	35-45
SM40 + 380	40-45	25-40	30-40	40-50	25-35	20-25	25-40	45-50
SM40 + 400	40-50	35-45	35-40	40-50	25-30	20-35	25-35	45-50
SM39 + 420	40-55	30-40	35-45	45-55	30-40	25-35	25-35	40-50
SM39 + 440	35-45	30-35	35-40	45-50	30-35	30-40	35-40	40-45
SM40 + 460	35-40	35-40	30-40	45-50	35-45	30-45	35-45	40-45
SM40 + 480 (BH3)	40-45	40-45	40-45	40-45	40-45	40-45	40-45	40-45

IPE 180 - Spacing 0.75-1.5 m
Rock bolts: Fully grouted, φ25 mm, L: 4-6 m
Shotcrete: 22cm-Reinforced by 2 layer mesh φ6@100×100 mm.

Depending on the ground conditions, during the tunnel construction, the spacing of the IPEs has changed.

5.2 Sample selection and model establishment

In this paper, RMR parameter has been considered in order to predict the ground conditions with the help of GPR along the tunnel route. The reason for choosing the RMR parameter is that it can be used to estimate the status of several other parameters such as Rock Structure Rating (RSR), Q-value, maximum unsupported span, etc. Therefore, the RMR parameter is very important in the construction of underground spaces, especially tunnels.

Five boreholes and two entrance and exit portals are used as observational locations to predict the state of geological or geomechanical parameters along the tunnel route. To access the initial data of the GPR model for training, as in Section (3), the data between each of the two observations are obtained separately from historical tunnels. Totally, for the distance between each observation, 10 datasets were used to examine the information after constructing 27 tunnels for training the model. For example, to reach the initial data between the entrance of the tunnel and the borehole 1, spaced 373 meters apart, 10 datasets are obtained after studying the 27 historical tunnels (exactly it is not necessary to have exactly a distance of 373 meters, in this case 373± 20 meters are considered). In the Garan tunnel, the status of the RMR parameter in the entrance portal and borehole 1 locations, is 40 and 5, respectively. The 10 input datasets obtained from historical tunnels for the RMR parameter between the two observations are presented in Tables 1-5, respectively.

5.3 Results analysis and comparison

In thpresented in the previous section are used to train the three GPR, SVR,e next step, the parameter of RMR was predicted along the tunnel route using three GPR, SVR and ANN techniques, as shown in Fig. 7. In this section, all of the eight datasets presented in the previous section are used to train the three GPR, SVRpresented in the previous section is used to train the three GPR, SVR, and ANN techniques. Concerning the predicted values of the RMR in Fig. 7, it can be seen that the predicted results of the three methods are close together and mostly, the predicted status of the RMR is in fourth class.

In the next step, where the main focus of this paper is, with the help of these predicted values of RMR, the time and costs of the tunnel construction should be predicted. So, once again, each of the three GPR, SVR and ANN techniques should be used. In this way, with the help of constructed mountain tunnels, each meter's construction time and costs for different RMR values (from zero to 60) were obtained according to Table 6. These values are used to train the GPR, SVR and ANN models to predict the whole time and costs of the tunnel construction.

Since the predicted RMR values based on the three techniques are in the three classes of III, IV and V, the RMR values of the previously built tunnels are considered within these three classes (from zero to 60). Also, the selection of the previously constructed tunnels is such that the drilling method, the tunnel cross-section and the maintenance system are the same as the Garan mountain tunnel.

Finally, with the help of the predicted results for RMR according to Fig. 7, and also using Table 6 to train each of the three GPR, SVR and ANN techniques, the total construction time and costs of the Garan road tunnel were obtained as shown in Fig. 8. According to Fig. 8, since the Garan tunnel project was completed at the time of completion of this study and the total construction time and costs of this tunnel were fully available, for validation purposes, the predicted construction time and costs of the GPR, SVR and ANN techniques were compared to the actual mode. As shown in Fig. 8, the predicted results of the three GPR, SVR, and ANN techniques are accurate and close together, so that the predicted result for GPR is more accurate than SVR and for SVR is more accurate than ANN. The results of the GPR differ 12 days for construction time and 18450$ for construction costs from

Table 4 Input data to train the GPR model for predict the RMR parameter state between borehole 3 and borehole 4

Tunnel chainage (m)	RMR							
SM40 + 480 (BH3)	40-45	40-45	40-45	40-45	40-45	40-45	40-45	40-45
SM40 + 500	40-50	40-45	30-40	40-50	35-40	30-40	50-55	35-40
SM40 + 520	40-50	40-50	30-45	40-45	35-30	30-40	50-60	30-35
SM40 + 540	45-55	35-45	30-45	40-45	30-40	35-40	50-65	30-40
SM40 + 560	45-55	30-45	30-40	40-50	25-35	30-35	55-65	30-45
SM40 + 580	55-60	35-50	35-45	45-50	25-40	30-45	55-60	25-40
SM40 + 600	55-60	35-45	35-40	35-45	20-35	40-50	55-60	20-35
SM40 + 620	55-65	30-40	30-40	35-45	20-35	40-50	50-60	20-35
SM40 + 640	55-60	35-45	35-40	35-50	20-40	45-55	50-60	25-30
SM40 + 660	45-55	30-40	35-45	40-50	25-40	45-65	45-60	25-35
SM40 + 680	45-65	30-35	40-45	40-55	25-35	50-60	45-55	20-35
SM40 + 700	40-45	30-40	40-55	45-50	20-35	50-65	45-55	15-25
SM40 + 720	35-45	30-45	40-50	45-55	20-30	55-65	45-50	15-20
SM40 + 740	35-40	30-45	40-50	40-45	20-25	55-60	40-50	15-25
SM40 + 760	45-50	35-40	45-50	35-45	15-25	55-65	35-40	20-25
SM40 + 780	40-50	40-45	35-45	35-45	10-25	60-65	35-40	20-30
SM40 + 800	30-45	30-40	35-45	35-40	10-15	60-70	30-35	20-25
SM40 + 820	35-40	30-40	35-40	35-40	05-20	65-70	30-35	25-30
SM40 + 840	20-40	35-40	30-45	35-45	05-20	60-70	35-40	20-30
SM40 + 860	20-25	25-30	30-40	35-45	05-15	55-60	35-45	20-30
SM40 + 880	25-35	25-35	25-35	35-40	10-15	35-55	30-40	15-30
SM40 + 900	25-40	20-40	25-30	30-40	10-20	35-45	35-40	15-20
SM40 + 920	20-35	20-35	20-30	30-35	05-15	30-40	25-35	10-20
SM40 + 940	20-25	20-30	10-25	30-35	05-10	30-35	25-35	10-15
SM40 + 960	20-25	20-30	10-20	25-30	05-10	30-35	20-35	20-30
SM40 + 980	15-25	20-25	10-25	20-30	05-25	35-40	15-30	25-35
SM41 + 000	15-20	20-25	10-25	20-25	05-20	30-35	15-25	25-30
SM41 + 020	15-20	10-20	10-15	15-25	10-20	20-35	10-15	20-30
SM41 + 040	10-20	10-25	15-25	15-20	10-15	20-30	05-10	20-25
SM41 + 060	10-25	10-20	10-20	15-20	15-20	20-35	05-15	15-25
SM41 + 080	10-15	10-20	15-20	10-20	15-20	20-30	10-15	15-30
SM41 + 100	10-15	10-15	15-25	10-15	15-25	20-25	10-25	15-25
SM41 + 120	15-20	15-20	15-30	15-25	10-20	15-25	15-25	10-25
SM41 + 140	15-25	15-25	10-25	15-20	10-15	15-20	15-20	10-20
SM41 + 160	15-20	15-20	10-20	15-20	15-25	10-20	15-20	15-20
SM41 + 180 (BH4)	10-20	10-20	10-20	10-20	10-20	10-20	10-20	10-20

Table 5 Input data to train the GPR model for predict the RMR parameter state between borehole 4 and exit

Tunnel chainage (m)	RMR							
SM41 + 180 (BH4)	10-20	10-20	10-20	10-20	10-20	10-20	10-20	10-20
SM41 + 200	10-20	10-20	10-20	15-20	15-20	10-15	10-15	05-15
SM41 + 220	10-15	20-25	10-25	10-20	15-20	05-15	10-20	05-10
SM41 + 240	15-20	20-30	15-30	10-15	15-25	05-10	10-20	10-15
SM41 + 260	15-25	15-25	15-25	15-25	10-20	05-10	10-20	10-15
SM41 + 280	10-20	15-20	20-25	15-30	10-25	05-10	10-20	15-20
SM41 + 300	05-15	10-15	20-30	20-35	10-20	10-20	15-20	10-20
SM41 + 320	10-15	10-20	15-25	20-25	15-20	10-15	10-15	05-10
SM41 + 340	10-25	15-20	10-20	20-25	15-25	10-15	15-25	05-15
SM41 + 360	10-15	15-20	10-15	15-25	10-15	15-20	15-20	15-20
SM41 + 380	15-25	15-20	10-20	15-20	10-20	15-25	15-20	15-25
SM41 + 390 (Exit)	15-20	15-20	15-20	15-20	15-20	15-20	15-20	15-20

the actual mode. These differences for the SVR are 19 days and 26757$, and for the ANN are 38 days and 45215$. Therefore, the GPR, SVR and ANN techniques used in this paper, due to the proximity of the predicted results to the actual mode, can be very important during planning of the tunneling projects and can reduce the risk of decisions to an acceptable level.

6. Conclusions

The mean values of the construction time were obtained for GPR, SVR and ANN, 1418 days, 1387 days and 1368 days, respectively. These mean values of construction costs were obtained by GPR, SVR and ANN equal to 10892550$,

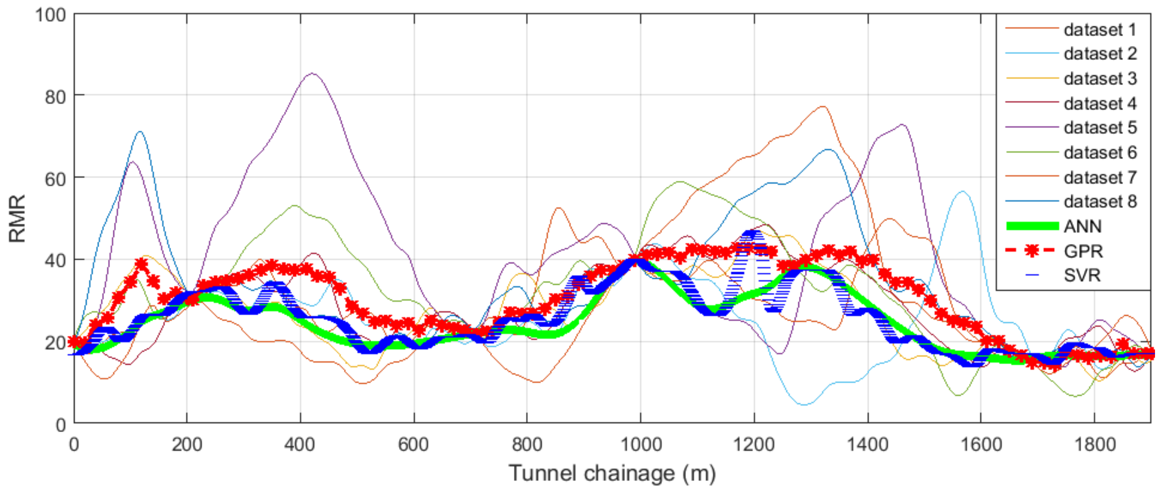

Fig. 7 The predicted results of GPR, SVR and ANN techniques for RMR during Garan tunnel route by using the eight datasets to training

Table 6 The construction time and costs per each meter of previous constructed tunnels in different RMR states (zero to 60). These selected tunnels are similar to the Garan mountain tunnel in terms of excavating method, maintenance system and tunnel cross-section

RMR	Time per each meter (day)	Cost per each meter ($)	RMR	Time per each meter (day)	Cost per each meter ($)	RMR	Time per each meter (day)	Cost per each meter ($)
0-5	0.92	6131	20-25	0.76	5874	40-45	0.41	5663
0-5	0.91	6126	25-30	0.72	5870	40-45	0.45	5648
0-5	0.88	6120	25-30	0.69	5865	40-45	0.41	5673
5-10	0.85	6115	25-30	0.61	5888	40-45	0.40	5684
5-10	0.81	6109	25-30	0.64	5892	40-45	0.39	5642
5-10	0.77	6113	25-30	0.67	5870	40-45	0.42	5619
5-10	0.75	6098	25-30	0.62	5864	40-45	0.40	5611
5-10	0.79	6057	25-30	0.63	5868	45-50	0.39	5586
5-10	0.83	6034	25-30	0.59	5862	45-50	0.37	5563
10-15	0.73	6046	30-35	0.61	5857	45-50	0.42	5538
10-15	0.75	6021	30-35	0.57	5859	45-50	0.44	5567
10-15	0.78	6035	30-35	0.58	5861	45-50	0.42	5546
10-15	0.74	6042	30-35	0.54	5845	45-50	0.38	5510
10-15	0.75	6031	30-35	0.55	5850	45-50	0.38	5467
10-15	0.75	5978	30-35	0.53	5843	50-55	0.39	5479
10-15	0.79	5965	30-35	0.55	5835	50-55	0.40	5487
10-15	0.80	5936	30-35	0.52	5804	50-55	0.39	5431
15-20	0.73	5953	30-35	0.57	5815	50-55	0.35	5421
15-20	0.71	5927	30-35	0.54	5812	50-55	0.37	5397
15-20	0.72	5903	30-35	0.54	5787	50-55	0.34	5366
15-20	0.69	5892	30-35	0.53	5768	50-55	0.33	5388
15-20	0.71	5882	30-35	0.51	5747	50-55	0.36	5344
15-20	0.74	5880	30-35	0.50	5743	50-55	0.38	5355
15-20	0.73	5903	35-40	0.43	5731	50-55	0.34	5365
15-20	0.69	5885	35-40	0.48	5776	55-60	0.35	5345
15-20	0.70	5876	35-40	0.47	5764	55-60	0.31	5331
20-25	0.73	5883	35-40	0.45	5743	55-60	0.30	5327
20-25	0.71	5885	35-40	0.52	5751	55-60	0.31	5289
20-25	0.68	5879	35-40	0.55	5737	55-60	0.35	5266
20-25	0.75	5870	35-40	0.47	5742	55-60	0.37	5255
20-25	0.78	5876	35-40	0.44	5712	55-60	0.34	5277
20-25	0.78	5878	40-45	0.46	5707	55-60	0.33	5269
20-25	0.74	5880	40-45	0.45	5687			

10884243$ and 10865785$, respectively. Also, the construction time and costs of the real mod were equal to1406 days and 10911000$. Therefore, the overall conclusion that can be drawn is that one can make very good predictions about the time and costs of construction with each of the GPR, SVR and ANN tools. But in these predictions, GPR results are more accurate than SVR, and SVR results are more accurate than ANN.

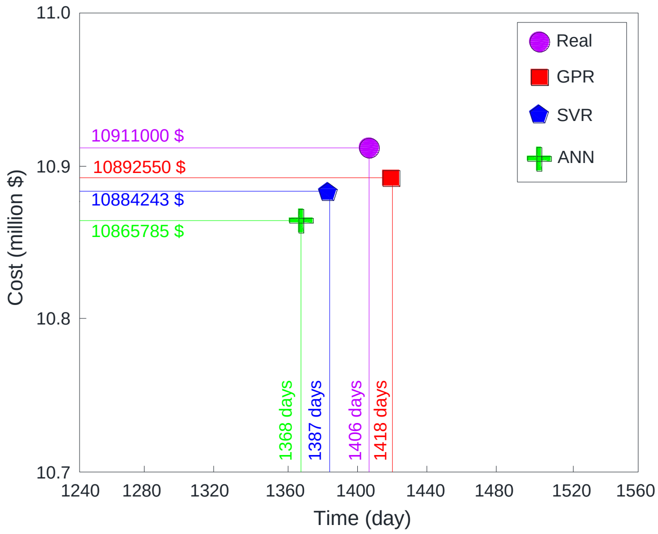

Fig. 8 The predicted construction time and costs by three GPR, SVR and ANN techniques and comparing them with the actual mode for the Garan mountain tunnel

References

Alade, I.O., Zhang, Y. and Xu, X. (2021), "Modeling and prediction of lattice parameters of binary spinel compounds (AM2X4) using support vector regression with Bayesian optimization", *New J. Chem.*, **34**. https://doi.org/10.1039/d1nj01523k

Bai, X.D., Cheng, W.C., Ong, D.E.L. and Ge Li. (2021), "Evaluation of geological conditions and clogging of tunneling using machine learning", *Geomech. Eng.*, **25**(1), 59-73. http://doi.org/10.12989/gae.2021.25.1.059.

Carrière, S.D., Chalikakis, K., Guy, S., Charles, D. and Christophe, E. (2013), "Combining electrical resistivity tomography and ground penetrating radar to study geological structuring of karst unsaturated zone", *J. Appl. Geophys.*, **94**, 31-41. https://doi.org/10.1016/j.jappgeo.2013.03.014.

Guan, Z., Deng, T., Jiang, Y., Zhao, C. and Huang, H. (2014), "Probabilistic estimation of ground condition and construction cost for mountain tunnels", *Tun. Undergr. Sp. Tech.*, **42**, 175-183. https://doi.org/10.1016/j.tust.2014.02.014.

Khishe, M., Mosavi, M.R. and Moridi, A. (2018), "Chaotic fractal walk trainer for sonar data set classification using multi-layer perceptron neural network and its hardware implementation", *Appl. Acoust.*, **137**, 121-139. https://doi.org/10.1016/j.apacoust.2018.03.012.

Khishe, M. and Mosavi, M.R. (2019), "Improved whale trainer for sonar datasets classification using neural network", *Appl. Acoust.*, **154**, 176-192. https://doi.org/10.1016/j.apacoust.2019.05.006.

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

Luat, N.V., Lee, K. and Duc-Kien, T. (2020), "Application of artificial neural networks in settlement prediction of shallow foundations on sandy soils.", *Geomech. Eng.*, **20**(5), http://doi.org/10.12989/gae.2020.20.5.385.

Liu, J., Jiang, Y., Zhang, Y. and Sakaguchi, O. (2021a), "Influence of different combinations of measurement while drilling parameters by artificial neural network on estimation of tunnel support patterns.", *Geomech. Eng.*, **25**(6), 439-454. http://doi.org/10.12989/gae.2021.25.6.439.

Liu, L.L., Yang, C. and Wang, X.M. (2021b), "Landslide susceptibility assessment using feature selection-based machine learning models.", *Geomech. Eng.*, **25**(1), 1-16. http://doi.org/10.12989/gae.2021.25.1.001.

Mahmoodzadeh, A. and Zare, S. (2016), "Probabilistic prediction of the expected ground conditions and construction time and costs in road tunnels", *J. Rock Mech. Geotech. Eng.*, **8**(5), 734-745. https://doi.org/10.1016/j.jrmge.2016.07.001.

Mahmoodzadeh, A., Mohammadi, M., Daraei, A., Rashid, T.A., Sherwani, A.F.H., Faraj, R.H. and Darwesh, A.M. (2019), "Updating ground conditions and time-cost scatter-gram in tunnels during excavation", *Autom. Constr.*, **105**, 102822. DOI: 10.1016/j.autcon.2019.04.017.

Mahmoodzadeh, A., Mohammadi, M., Abdulhamid, S.N., Ibrahim, H.H., Hama-Ali, H.F. and Salim, S.G. (2021a), "Dynamic reduction of time and cost uncertainties in tunneling projects", *Tun. Undergr. Sp. Tech.*, **109**, 103774. https://doi.org/10.1016/j.tust.2020.103774.

Mahmoodzadeh, A., Mohammadi, M., Abdulhamid, S.N., Nejati, H.R., Noori, K.M.G., Ibrahim, H.H. and Hama Ali, H.F. (2021b), "Predicting construction time and cost of tunnels using Markov chain model considering opinions of experts", *Tun. Undergr. Sp. Tech.*, **116**, 104109. https://doi.org/10.1016/j.tust.2021.104109.

Mahmoodzadeh, A., Mohammadi, M., Daraei, A., Hama Ali, H.F., Al-Salihi, N.K. and Omer, R.M.D. (2020a), "Forecasting maximum surface settlement caused by urban tunneling", *Autom. Constr.*, **120**, 103375. https://doi.org/10.1016/j.autcon.2020.103375.

Mahmoodzadeh, A., Mohammadi, M., Daraei, A., Faraj, R.H., Omer, R.M.D. and Sherwani, A.F.H. (2020b), "Decision-making in tunneling using artificial intelligence tools", *Tun. Undergr. Sp. Tech.*, **103**, 103514. https://doi.org/10.1016/j.tust.2020.103514.

Mahmoodzadeh, A., Mohammadi, M., Hama Ali, H.F., Abdulhamid, S.N., Ibrahim, H.H. and Noori, K.M.G. (2021c), "Dynamic prediction models of rock quality designation in tunneling projects", *Transportation Geotech.*, **27**, 100497. https://doi.org/10.1016/j.trgeo.2020.100497.

Mahmoodzadeh, A., Mohammadi, M., Daraei, A., Hama Ali, H.F., Abdullah, A.I. and Al-Salihi, N.K. (2021d), "Forecasting tunnel geology, construction time and costs using machine learning methods", *Neural Comput. Appl.*, **33**, 321–348. https://doi.org/10.1007/s00521-020-05006-2

Mahmoodzadeh, A., Mohammadi, M., Ibrahim, H.H., Abdulhamid, S.N., Hama Ali, H.F., Hasan, A.M., Khishe, M. and Mahmud, H.

(2021e), "Machine learning forecasting models of disc cutters life of tunnel boring machine", *Autom. Constr*,, **128**, 103779. https://doi.org/10.1016/j.autcon.2021.103779.

Mahmoodzadeh, A., Mohammadi, M., Noori, K.M.G., Khishe, M., Ibrahim, H.H., Hama Ali, H.F. and Abdulhamid, S.N. (2021f), "Presenting the best prediction model of water inflow into drill and blast tunnels among several machine learning techniques", *Autom. Constr*,, **127**, 103719. https://doi.org/10.1016/j.autcon.2021.103719.

Mahmoodzadeh, A., Mohammadi, M., Ibrahim, H.H., Noori, K.M.G., Abdulhamid, S.N. and Hama Ali, H.F. (2021g), "Forecasting sidewall displacement of underground caverns using machine learning techniques", *Autom. Constr*,, **123**, 103530. https://doi.org/10.1016/j.autcon.2020.103530.

Mahmoodzadeh, A., Mohammadi, M., Ibrahim, H.H., Rashid, T.A., Aldalwie, A.H.M., Hama Ali, H.F. and Daraei, A. (2021h), "Tunnel geomechanical parameters prediction using Gaussian process regression", *Mach. Learn. Appl*., **3**, 100020. https://doi.org/10.1016/j.mlwa.2021.100020.

Min, S. and Einstein, H.H. (2016), "Resource scheduling and planning for tunneling with a new resource model of the Decision Aids for Tunneling (DAT)", *Tun. Undergr. Sp. Tech.*, **51**, 212-225. https://doi.org/10.1016/j.tust.2015.10.038.

Sousa, R.L. and Einstein, H.H. (2012), "Risk analysis during tunnel construction using Bayesian Networks: Porto Metro case study", *Tun. Undergr. Sp. Tech*., 27(1), 86–100. https://doi.org/10.1016/j.tust.2011.07.003

Saffari, A., Zahiri, S.H and Khishe, M. (2022). "Fuzzy grasshopper optimization algorithm: A hybrid technique for tuning the control parameters of GOA using fuzzy system for big data sonar classification", *IJEEE*, **18**(1), 2131-2131. http://ijeee.iust.ac.ir/article-1-2131-en.html.

Taghavi, M. and Khishe, M. (2019). "A Modified Grey Wolf Optimizer by Individual Best Memory and Penalty Factor for Sonar and Radar Dataset Classification", *Iran J. Mar. Technol.*, **6**(1), 120-130. http://ijmt.iranjournals.ir/article_35433.html.

Vapnik, V.N. (1995), The nature of statistical learning theory. New York: Springer.

Xiang, G., Yin, D., Cao, C. and Yuan, L. (2021), "Application of artificial neural network for prediction of flow ability of soft soil subjected to vibrations.", *Geomech. Eng*., **25**(5), 395-403. http://doi.org/10.12989/gae.2021.25.5.395.

Zhang, Y. and Xu, X. (2021a), "Solid particle erosion rate predictions through LSBoost", *Powder Technol*., **388**, 517-525. https://doi.org/10.1016/j.powtec.2021.04.072.

Zhang, Y. and Xu, X (2021b), "Predictions of the total crack length in solidification cracking through LSBoost", *Metal. Mater. Trans. A*, **52**, 985-1005. https://doi.org/10.1007/s11661-020-06130-3.

Zhang, Y. and Xu, X. (2021c), "Predicting the material removal rate during electrical discharge diamond grinding using the Gaussian process regression: a comparison with the artificial neural network and response surface methodology", *Int. J. Adv. Manuf. Technol*., **113**, 1527-1533. https://doi.org/10.1007/s00170-021-06701-7.

Zhang, Y. and Xu, X. (2021d), "Lattice misfit predictions via the Gaussian process regression for ni-based single crystal superalloys", *Met. Mater. Int*., **27**, 235-253. https://doi.org/10.1007/s12540-020-00883-7.

Zhang, Y. and Xu, X. (2020), "Solubility predictions through LSBoost for supercritical carbon dioxide in ionic liquids", *New J. Chem*., **47**. https://doi.org/10.1039/d0nj03868g.

CC