



# Forecasting tunnel geology, construction time and costs using machine learning methods

Arsalan Mahmoodzadeh<sup>1</sup> · Mokhtar Mohammadi<sup>2</sup> · Ako Daraei<sup>1</sup> · Hunar Farid Hama Ali<sup>3</sup> · Abdulqadir Ismail Abdullah<sup>4</sup> · Nawzad Kameran Al-Salihi<sup>5</sup>

Received: 11 October 2019 / Accepted: 2 May 2020 / Published online: 18 May 2020  
© Springer-Verlag London Ltd., part of Springer Nature 2020

## Abstract

This research intends to use machine learning approaches to predict tunnel geology and its construction time and costs. For this purpose, the Gaussian Process Regression (GPR), Support Vector Regression (SVR), and Decision Tree (DT) have been utilized. An estimation of the geological conditions of the Garan road tunnel and its construction time and cost has been conducted. In addition, after constructing about 200 m from the inlet and outlet sides of the tunnel, using the field-observed data of these sectors in the tools, all the previously forecasted results were updated for unconstructed parts. Fivefold cross-validation has been applied to assess the performance of each model. The obtained models are used to predict construction time and cost in real scenarios, and the accuracy of each model was investigated through different statistical evaluation criteria. Finally, it turns out that all the models provide relatively high performance and reduce the uncertainties of tunnel geology. However, the GPR provides more accurate results compared to the SVR and DT tools. Thus, we recommend the GPR for the prediction of geology and construction time and costs in future levels of a tunnel.

**Keywords** Gaussian Process Regression · Support Vector Regression · Decision Tree · Tunneling

## 1 Introduction

Cost overruns and delays often encounter tunnel construction projects. Delays may negatively affect the scope of tunneling projects, which leads to severe cost overruns. Applying contingencies and estimating risks at the project

level often do not capture the multiple uncertainties in the construction process of tunnel projects. An example of cost and duration underestimation in subsurface projects is the Channel Tunnel between England and France. The construction of the Channel Tunnel started in 1988, and the

---

✉ Arsalan Mahmoodzadeh  
arsalan.mahmoodzadeh@soran.edu.iq

Mokhtar Mohammadi  
mokhtar.mohammadi@uhd.edu.iq

Ako Daraei  
ako.daraei@soran.edu.iq

Hunar Farid Hama Ali  
hunar.hamaali@uoh.edu.iq

Abdulqadir Ismail Abdullah  
abdulqadir.abdullah@ue.edu.krd

Nawzad Kameran Al-Salihi  
n.al-salihi@ukh.edu.krd

<sup>2</sup> Department of Information Technology, University of Human Development, Sulaymaniyah, Kurdistan Region, Iraq

<sup>3</sup> Civil Engineering Department, University of Halabja, Halabja, Kurdistan Region, Iraq

<sup>4</sup> Department of Computer Science, College of Science, International University of Erbil, Erbil, Kurdistan Region, Iraq

<sup>5</sup> Computer Science and Engineering, University of Kurdistan Hewler, Erbil, Kurdistan Region, Iraq

<sup>1</sup> Department of Civil Engineering, Faculty of Engineering, Soran University, Soran, Kurdistan Region, Iraq

project took approximately 20% longer than planned (6 years instead of the scheduled 5) and 80% over the budget (costs increased £2 billion, from £2.6 billion to £4.6 billion). The Channel Tunnel is just one among many examples of subsurface projects completed with high cost and duration overruns. While these projects are very large, the phenomenon of cost overruns and delays in subsurface projects is widespread, as described by Flyvbjerg et al. [1]. Thus, there is a necessity for innovative methods and tools to eschew significant construction cost and duration overruns. In this paper, two artificial intelligence methods of Gaussian Process Regression (GPR) and Support Vector Regression (SVR) will be described with which cost and duration underestimation in road tunnel projects can be avoided. The models will be applied to a road tunnel in Iran.

Insurers have reported major losses and delays in tunnel construction projects. The necessity for analyzing the uncertainty and risks of tunnel construction has been recognized by the tunneling community [2–5]. At present, construction time and costs of engineering projects were commonly assessed on a deterministic basis. The deterministic tool, however, does not appropriately reflect the uncertain reality. Recently, many innovative studies have been developed to minimize uncertainties regarding construction time and costs [6, 7]. Flyvbjerg [8] proposed utilizing reference class forecasting for transportation construction projects. For example, this tool estimates the cost or duration performance of a project based on statistical analyses of the previously constructed projects rather than on the specifics of the project itself. Different risk factors causing project cost underestimation have been described in detail in the transportation construction literature [9–12].

A variety approaches and tools have been used to estimate the tunnel construction time and costs, such as Markov chains [7, 13], Monte Carlo simulation [8, 14], dynamic Bayesian networks [15], decision aids for tunneling (DAT) developed at MIT in collaboration with Ecole Polytechnique Fédérale de Lausanne—EPFL (Moret and Einstein, 2016), Bayesian analysis and artificial neural networks [16, 17], and analytical solution [18]. In the more recent researches, different machine learning techniques have been applied in different civil engineering problems [19], for example optimization of geotechnical problems [20], project control forecasting [21], construction management [22], construction injury prediction [23], slope collapse prediction [24], prediction of TBM operating parameters [25], tunnels' convergence rate forecasting [26], landslide displacement [27], rockburst prediction

[28], manufacturing cost estimation [29], uniaxial compressive strength prediction [30], ground surface settlement [31], determination of earth pressure [32], risk assessment and costs estimation [33, 34], rock fragmentation forecasting [35], and other engineering problems.

Although many studies have been conducted by researchers to solve various engineering problems by the machine learning tools, so far, these methods have not been provided to predict a tunnel path geology and time and costs required to its construction. Therefore, a novel approach has been implemented for this research which intends to use three machine learning tools widely used in solving complex engineering problems, Gaussian Process Regression (GPR), Support Vector Regression (SVR), and Decision Tree (DT) for prediction of geology conditions, construction time, and construction costs in road tunnels construction. The paper also aims to determine the relationship between the tunnel geology parameters with the time and costs required for construction, which in this paper, rock mass rating (RMR) parameter was considered. In order to train the GPR, SVR, and DT tools, datasets from the previously constructed road tunnels were gathered. Also, another interesting work in this paper is to update the previously predicted results by applying the more datasets obtained during the tunnel construction in the constructed parts. To verify the feasibility of these tools, as well as the other machine learning tools, they were applied to the Garan tunnel on Sanandaj–Marivan road in Iran. Next, to assess the performances of the applied machine learning tools, fivefold cross-validation was adopted using MATLAB 2018 tool through the regression learner app, and the accuracy of the GPR, SVR, and DT predictions was investigated through different statistical evaluation criteria. Also, in each step of the predictions, to predict the tunnel geology, construction time, and construction costs along the Garan tunnel route, the obtained models of the prediction tools through the regression learner app were exported. By applying the datasets of the Garan tunnel route as the test data in the exported models, all the parameters were predicted along the tunnel route. Finally, the accuracy of the predictions made by each tool, as well as the impact of the updates on the predicted results, was investigated.

The rest of this paper is organized as follows: In the next section, the work methodology is described step by step. In Sect. 3, the application of the methodology in a case study is briefly outlined. Results analysis and validations are presented in Sect. 4. Discussion and conclusions are described in Sects. 5 and 6, respectively.

## 2 Methodology

### 2.1 Geology parameter selection

In the first phase of the work, a geology parameter should be considered that can describe the tunnel’s ground conditions and affect tunnel construction time and costs. Rock mass classification systems are a crucial part of underground projects and can well describe tunnel geology conditions. RMR parameter is one of the most commonly applied classification systems in numerous civil and mining projects [36]. Therefore, in this paper, it is considered as an effective parameter on the tunnel geology and the tunnel construction time and costs.

RMR system is a geomechanical classification system for rocks, developed by Bieniawski [37]. It combines the most significant geologic parameters of influence and represents them with one overall comprehensive index of rock mass quality, which is used for the design and construction of excavations in rock, such as tunnels, mines, slopes, and foundations. Over the years, the RMR system has been successively refined as more case records have been examined, and the reader should be aware that Bieniawski has made significant changes in the rating assigned to different parameters. The discussion which follows is based upon the 1989 version of the classification [38]. The following six parameters are estimating the strength of rock mass using the RMR system:

- Uniaxial compressive strength of rock material
- Rock quality designation (RQD)
- Spacing of discontinuities
- Condition of discontinuities
- Groundwater conditions
- Orientation of discontinuities

Each of the six parameters is assigned a value corresponding to the characteristics of the rock. These values are derived from field surveys and laboratory tests. The sum of the six parameters is the RMR value, which lies between 0 and 100. The classification of the RMR system is provided in Table 1.

### 2.2 Machine learning tools of GPR, SVR, and DT

#### 2.2.1 GPR

GPR models as nonparametric kernel-based probabilistic models can be trained using a training set  $\{(x_i, y_i); i = 1, 2, \dots, n\}$ ,  $x_i \in R^d$ ,  $y_i \in R$ , drawn from an unknown distribution. A GPR model estimates the value of a response variable  $y_{new}$ , given the new input vector  $x_{new}$ , and the training data. A linear regressor can be modeled as

**Table 1** The classification for the RMR system [38]

RMR	0–20	21–40	41–60	61–80	81–100
Class no.	I	II	III	IV	V
Rock quality	Very poor	Poor	Fair	Good	Very good

$y = x^T\beta + \varepsilon$ , where  $\varepsilon \sim N(0, \sigma^2)$ . The error variance  $\sigma^2$  and the coefficients  $\beta$  are estimated from the data. A GPR model describes the response by introducing latent variables,  $f(x_i)$ ,  $i = 1, 2, \dots, n$ , from a Gaussian process (GP), and explicit basis functions,  $h$ . The covariance function of the latent variables captures the smoothness of the response, and basis functions map the inputs  $x$  into a  $p$ -dimensional feature space [39–41].

A GP is a set of random variables with a joint Gaussian distribution for any finite number of them. If  $\{f(x), x \in R^d\}$  is a GP, then given  $n$  observations of  $x_1, x_2, \dots, x_n$ , the joint distribution of the random variables  $f(x_1), f(x_2), \dots, f(x_n)$  is Gaussian. A GP is given by its mean function  $m(x)$  and kernel function,  $k(x, x')$ . That is, if  $\{f(x), x \in R^d\}$  is a GP, then:

$$E(f(x)) = m(x) \tag{1}$$

$$\begin{aligned} \text{Cov}[f(x), f(x')] &= E\{[f(x) - m(x)][f(x') - m(x')]\} \\ &= k(x, x') \end{aligned} \tag{2}$$

Let us consider the model of  $h(x)^T\beta + f(x)$ , where  $f(x) \sim \text{GP}(0, k(x, x'))$ , that is,  $f(x)$  are from a zero-mean GP with covariance function,  $k(x, x')$ .  $h(x)$  are a set of basis functions that transform the original feature vector  $x$  in  $R^d$  into a new feature vector  $h(x)$  in  $R^p$ .  $\beta$  is a  $p$ -by-1 vector of basis function coefficients. This model indicates a GPR model. An instance of response  $y$  can be modeled as [39–41]:

$$P(y_i|f(x_i), x_i) \sim N(y_i|h(x_i)^T\beta + f(x_i), \sigma^2) \tag{3}$$

Thus, a GPR model is a probabilistic model. There is a latent variable  $f(x_i)$  introduced for each observation  $x_i$ , which makes the GPR model nonparametric. In vector form, this model is equivalent to:

$$P(y|f, X) \sim N(y|H\beta + f, \sigma^2I) \tag{4}$$

$$\begin{aligned} \text{where } X &= \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad H = \begin{pmatrix} h(x_1^T) \\ h(x_2^T) \\ \vdots \\ h(x_n^T) \end{pmatrix}, \\ f &= \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}. \end{aligned}$$

The joint distribution of latent variables  $f(x_1), f(x_2), \dots, f(x_n)$  in the GPR model is as follows:

$$P(f|X) \sim N(f|0, K(X, X)) \quad (5)$$

close to a linear regression model, where  $K(X, X)$  is given as follows:

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

The covariance function  $k(x, x')$  is usually parameterized by a set of kernel parameters or hyperparameters,  $\theta$ . Often,  $k(x, x')$  is represented as  $k(x, x'|\theta)$  to explicitly show the dependence on  $\theta$  [39–41].

### 2.2.2 SVR

In SVR,  $\{x_i, y_i\}_{i=1}^l$  is considered as a training set, in which  $x_i \in R^p$  represents  $p$ -dimensional input vector and  $y_i \in R$  is a scalar measured output that indicates the system output. The goal is to construct a function  $y = f(x)$  that shows the dependency of the output  $y_i$  on the input  $x_i$  [42]. The function is expressed as:

$$y = w \cdot \phi(x) + b \quad (6)$$

where  $w$  is the weight vector, and  $b$  is the bias.

The following convex optimization problem (Eq. 7) [42, 43] can be used to express the regression problem. Figure 1 presents the concept of nonlinear SVR, corresponding to Eq. 7.

Minimize  $(w, b, \xi_i, \xi_i^*)$ :

$$\frac{1}{2} w^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (7)$$

Subject to:

$$y_i - w \cdot \phi(x) + b \leq \varepsilon + \xi_i^*$$

$$(w \cdot \phi(x) + b) - y_i \leq \varepsilon + \xi_i$$

$$\xi_i, \xi_i^* \geq 0, \dots, l$$

where  $\xi_i$  and  $\xi_i^*$  are slack variables that specify upper and lower training errors that are subject to error tolerance, and  $C$  is a positive constant that determines the degree of penalized loss when an error occurs.

The dual form of the nonlinear SVR can be described as [43]:

Minimize  $(\underline{\alpha}_i, \bar{\alpha}_i)$ :

$$\begin{aligned} & \frac{1}{2} \sum_{i,j=1}^l (\underline{\alpha}_i - \bar{\alpha}_i)(\underline{\alpha}_j - \bar{\alpha}_j) \phi(x_i) \cdot \phi(x_j) + \varepsilon \sum_{i=1}^l (\underline{\alpha}_i + \bar{\alpha}_i) \\ & - \sum_{i=1}^l y_i (\underline{\alpha}_i - \bar{\alpha}_i) \end{aligned} \quad (8)$$

Subject to:

$$\sum_{i=1}^l y_i (\underline{\alpha}_i - \bar{\alpha}_i) = 0$$

$$0 \leq \underline{\alpha}_i \leq C, \quad i = 1, 2, \dots, l$$

$$0 \leq \bar{\alpha}_i \leq C, \quad i = 1, 2, \dots, l$$

Selection of a suitable nonlinear function  $\phi(x_i)$  and the computation of  $\phi(x_i) \cdot \phi(x_j)$  in the feature space could be a difficult task. The input space computation can be performed using a kernel function  $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$  to yield the inner products in feature space, circumventing the problems intrinsic in determining the feature space. Functions that meet the condition of Mercer in feature space can be proven to match dot products. Any functions that fulfill the theorem of Mercer can, therefore, be used as a kernel. The following are some commonly used kernels in SVM [43]:

$$K(x_i, x_j) = x_i \cdot x_j \quad \text{Linear kernel} \quad (9)$$

$$K(x_i, x_j) = [\gamma(x_i, x_j) + c]^d \quad \text{Polynomial kernel} \quad (10)$$

$$K(x_i, x_j) = \tanh[\gamma(x_i, x_j) + c] \quad \text{Sigmoid kernel} \quad (11)$$

$$K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2) \quad \text{Radial basis function kernel} \quad (12)$$

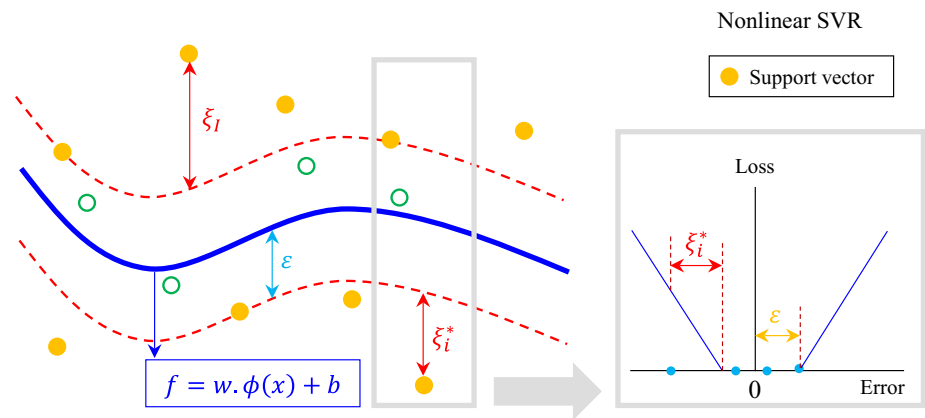
Finally, the feature of the kernel allows the nonlinear SVR decision function to be expressed as Eq. 13:

$$f(x_i) = \sum_{k=1}^l (-\underline{\alpha}_k + \bar{\alpha}_k) K(x_i, x_k) + b \quad (13)$$

### 2.2.3 DT

Decision Tree (DT) as a nonparametric data mining technique is an efficient statistical tool to develop prediction algorithms for a response variable. DT comes with several advantages; for example, no assumption is needed about the distribution of explanatory variables, it is not affected by high correlations among independent variables, it can be applied to many types of dependent variables including categorical, numerical, and survival data, the most important variables explaining the dependent variable are included in decision trees, and the insignificant variables are excluded. Although DT has initially been proposed for

**Fig. 1** Nonlinear SVR with Vapnik’s  $\epsilon$ -insensitive loss function [43]



large datasets, it also provides accurate estimates for small datasets.

Classification and Regression Tree (CART) is a specific type of DT algorithms. It covers both classification and regression trees. In the regression tree, the predicted outcome is considered a real number. In this case, variance reduction is used for splitting the nodes.

The steps for the algorithm are as follows:

1. First, the variance of the target is calculated
2. The datasets are divided into different attributes, and the variance of each branch is subtracted from the variance before the split. This is known as variance reduction.

The variance reduction of a node  $N$  is defined as:

$$I_V(N) = \frac{1}{|S|^2} \sum_{i \in S} \sum_{j \in S} \frac{1}{2} (x_i - x_j)^2 - \left( \frac{1}{|S_t|^2} \sum_{i \in S_t} \sum_{j \in S_t} \frac{1}{2} (x_i - x_j)^2 + \frac{1}{|S_f|^2} \sum_{i \in S_f} \sum_{j \in S_f} \frac{1}{2} (x_i - x_j)^2 \right) \tag{14}$$

where  $S$ ,  $S_t$ , and  $S_f$  are the set of presplit sample indices, set of sample indices for which the split test is true, and set of sample indices for which the split test is false, respectively. Each of the above summands is indeed variance estimates, though, written in a form without directly referring to the mean.

The attribute with the highest VR will be the decision node.

3. The datasets are divided based on the values of selected attributes. If the variance of a branch is more than zero, it divided again.
4. Continue the process until all the data are processed.

### 2.3 Automatic resource forecasting model

Initial observations are crucial for training the machine learning tools, and depending on the type of projects and

problems can be different. More observations for the training result in more precise results. Some of the most important observations in the road tunnels before construction are the inlet and outlet portals and location of boreholes, and perhaps these observations are often the only observations available for training at this stage.

In this work, a geological or geotechnical parameter (for example, parameter  $X$ ) in the tunnel path is considered first, so that there are preliminary observations about it in the tunnel path. These observations are then used to train the GPR, SVR, and DT tools. After the training process, parameter  $X$  state can be forecasted along the entire path of the tunnel. In the next stage, in order to predict the construction time and costs of the tunnel, the previously constructed road tunnels are applied to train the GPR, SVR, and DT tools. In this way, by studying the previously constructed road tunnels, the different time and costs per each meter of construction based on the different states of the parameter  $X$  are applied to train the prediction tools. After training, according to the forecasted states of the parameter  $X$  along the tunnel path using the GPR, SVR, and DT tools, the construction time and costs of the tunnel can be forecasted. It can be noted that the selected previously constructed tunnels should be similar to the tunnel under investigation in terms of the maintenance system, drilling method, tunnel cross section, which affects the tunnel construction time and costs.

Since the initial data are either low or not accurate before the construction stage of the tunnel, these can affect the predictions made on the ground conditions and the time and costs of the tunnel construction. To cope this problem, it is possible to update the previous prediction results during tunnel construction by using the actual data on parameter  $X$ , construction time, and construction costs obtained in the constructed parts as the new train datasets. Therefore, in order to update the previous prediction results during tunnel construction, in addition to the initial datasets, the field-observed datasets also can be utilized to train

the GPR, SVR, and DT tools. In Fig. 2, all the prediction stages of this article are presented in a schematically way.

In the next section, an engineering application of the GPR, SVR, and DT tools is presented on Garan road tunnel to predict the RMR parameter state along the tunnel route and the construction time and costs of the tunnel before and during construction.

### 3 Engineering application

#### 3.1 Engineering background

The new road among Sanandaj and Marivan cities is an under-development venture in the northwest of Iran. In light of the section through the Zagros Mountains, a few passages extend on this course, including the Garan, Hamru, Baghan, and Gezerdareh, which have been actualized or are under development. The Sanandaj–Marivan old road is a standout among the most hazardous roads in Iran, where numerous mishaps happen each year, prompting money-related misfortunes and passing of individuals. The length of the old Sanandaj–Marivan road is 126 km.

On the new road, the length will be decreased to 105 km, which can lessen the traveling time and mishaps.

The present examination is directed on Garan tunnel with a length of 1900 m and a cross-segment region of 97 m<sup>2</sup> as a major aspect of the under-development Sanandaj–Marivan road. The channel of Garan burrow is viewed as the eastern mouth (to the Sanandaj-S), and its outlet is viewed as the western (to the Marivan-M). The area of Garan tunnel is shown in Fig. 3.

The inlet and outlet portals of Garan tunnel are situated on the SM39 + 490 and SM41 + 390 (km + m), individually. The topographical profile of the Garan tunnel with the situation of the four boreholes is shown in Fig. 4a. As per the accessible information, a sum of four kinds of designing geography in the passage course is recognizable and isolated from one another: limestone (Li), shale (Sh), sand shales with shale–limestone arrangement (ShL), and limestone–shale succession (LSH).

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

- IPE 180: spacing 0.75–1.5 m
- Rock bolts: fully grouted,  $\phi 25$  mm,  $L$ : 4–6 m

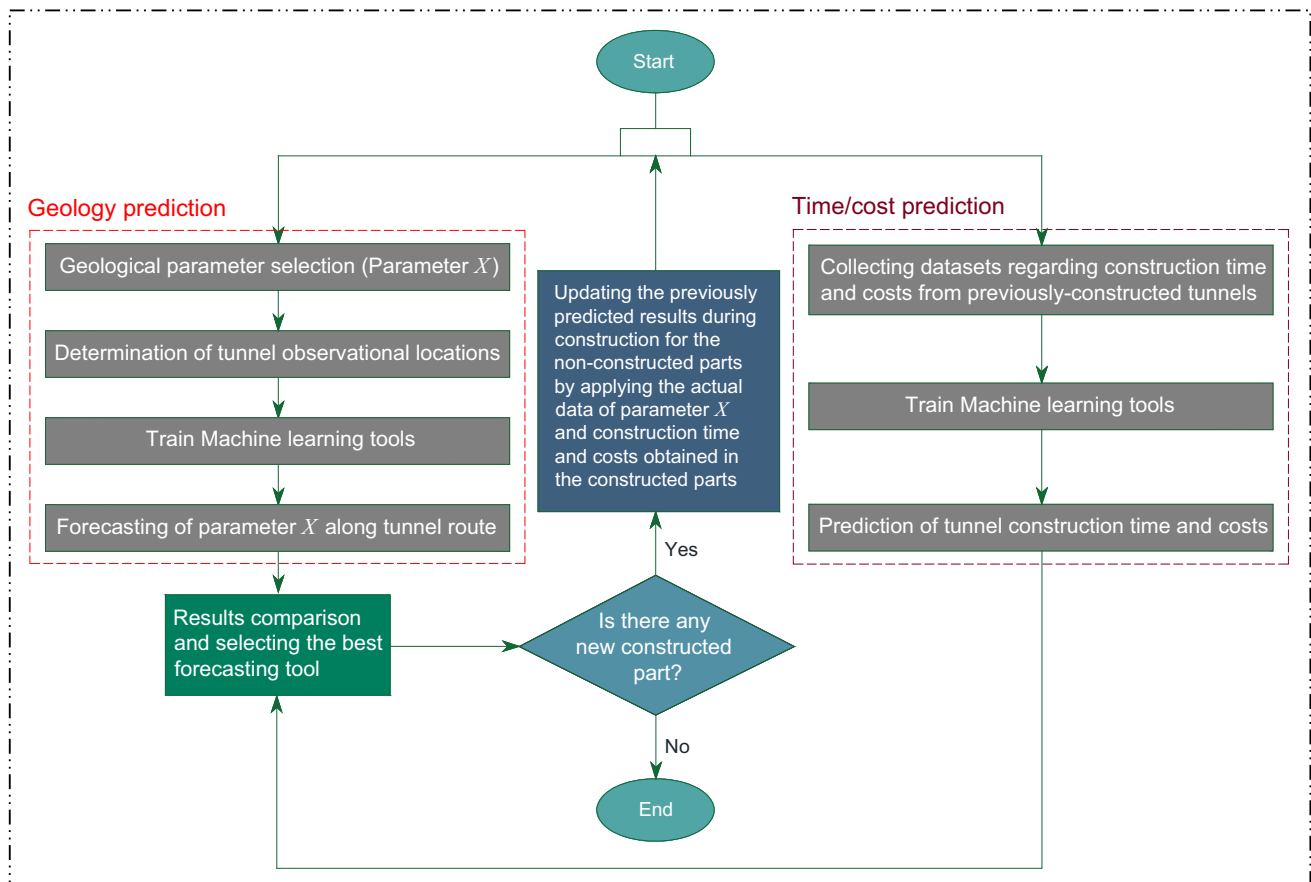


Fig. 2 The prediction stages of parameter  $X$  and the time and costs of tunnel construction by using the GPR, SVR, and DT tools

- Shotcrete: 22 cm reinforced by two-layer mesh  $\phi 6@100 \times 100$  mm.

In Fig. 4b, Garan tunnel cross section associated with projected/applied support system is presented.

### 3.2 Sample selection and model establishment

In this article, fivefold cross-validation is adopted at each stage of the predictions using the MATLAB tool, 2018, through the regression learner app. In this application, there are several different model types of GPR, SVR, and DT tools. At the forecasting stage, all the model types are used individually, and the results of the model with the highest accuracy were considered. In the MATLAB application, the GPR tool consists of four model types named rational quadratic, squared exponential, Matern 5/2, and exponential. The SVR tool consists of six model types called linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian. Also, the DT tool consists of three models of a fine tree, medium tree, and coarse tree.

Fitting a GPR model involves estimating the following model parameters from the data:

- Covariance function  $k(x_i, x_j|\theta)$  parameterized in terms of kernel parameters in vector  $\theta$ .
- Noise variance ( $\sigma^2$ ).
- Coefficient vector of fixed basis functions ( $\beta$ ).

The value of the ‘KernelParameters’ name–value pair argument is a vector that consists of initial values for the signal standard deviation ( $\sigma_f$ ) and the characteristic length scales ( $\sigma_l$ ). The ‘fitrgp’ function uses these values to determine the kernel parameters. Similarly, the ‘Sigma’ name–value pair argument contains the initial value for the noise standard deviation  $\sigma$ . During optimization, ‘fitrgp’ creates a vector of unconstrained initial parameter values  $\eta_0$  by using the noise standard deviation and the kernel parameters. ‘fitrgp’ analytically determines the explicit basis coefficient  $\beta$ , specified by the ‘Beta’ name–value pair argument, from estimated values of  $\theta$  and  $\sigma^2$ . Therefore,  $\beta$  does not appear in the  $\eta_0$  vector when ‘fitrgp’ initializes numerical optimization.

For SVR, the regularization and portion width parameters were tuned utilizing programmed hyper-parameter improvement technique given in MATLAB with the ‘fitsvm’ work. ‘fitsvm’ trains a Support Vector Machine (SVM) regression model on a low- through moderate-dimensional predictor dataset. It supports mapping the predictor data using kernel functions and supports sequential minimal optimization (SMO), iterative single data algorithm (ISDA), or L1 soft-margin minimization via quadratic programming (L1QP) for objective function minimization.

DT has a plethora of hyperparameters that require fine-tuning in order to derive the best possible model that reduces the generalization error as much as possible. Usually, the tree complexity is measured by one of the following metrics: the total number of nodes, the total number of leaves, tree depth, and a number of attributes used. `max_depth`, `min_samples_split`, and `min_samples_leaf` are all stopping criteria, whereas `min_weight_fraction_leaf` and `min_impurity_decrease` are pruning methods.

In each prediction step, to assess the performances of the results, seven statistical evaluation criteria are used. These criteria are the coefficient of determination ( $R^2$ ), mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), relative RMSE (RRMSE), mean absolute percentage error (MAPE), and mean relative error (MRE), respectively, given in Eqs. 15–21. All of these criteria were applied to validate the prediction results obtained by the intelligence approached in the previous researches [44–49]:

$$R^2 = \left( \frac{\sum_{i=1}^n (f(x_i) - \bar{f}(x))(f^*(x_i) - \bar{f}^*(x))}{\sqrt{\sum_{i=1}^n (f(x_i) - \bar{f}(x))^2 \sum_{i=1}^n (f^*(x_i) - \bar{f}^*(x))^2}} \right)^2 \tag{15}$$

$$MAE = \left( \frac{1}{n} \right) \sum_{i=1}^n |f(x_i) - f^*(x_i)| \tag{16}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (f(x_i) - f^*(x_i))^2 \tag{17}$$

$$RMSE = \sqrt{\left( \frac{1}{n} \right) \sum_{i=1}^n (f(x_i) - f^*(x_i))^2} \tag{18}$$

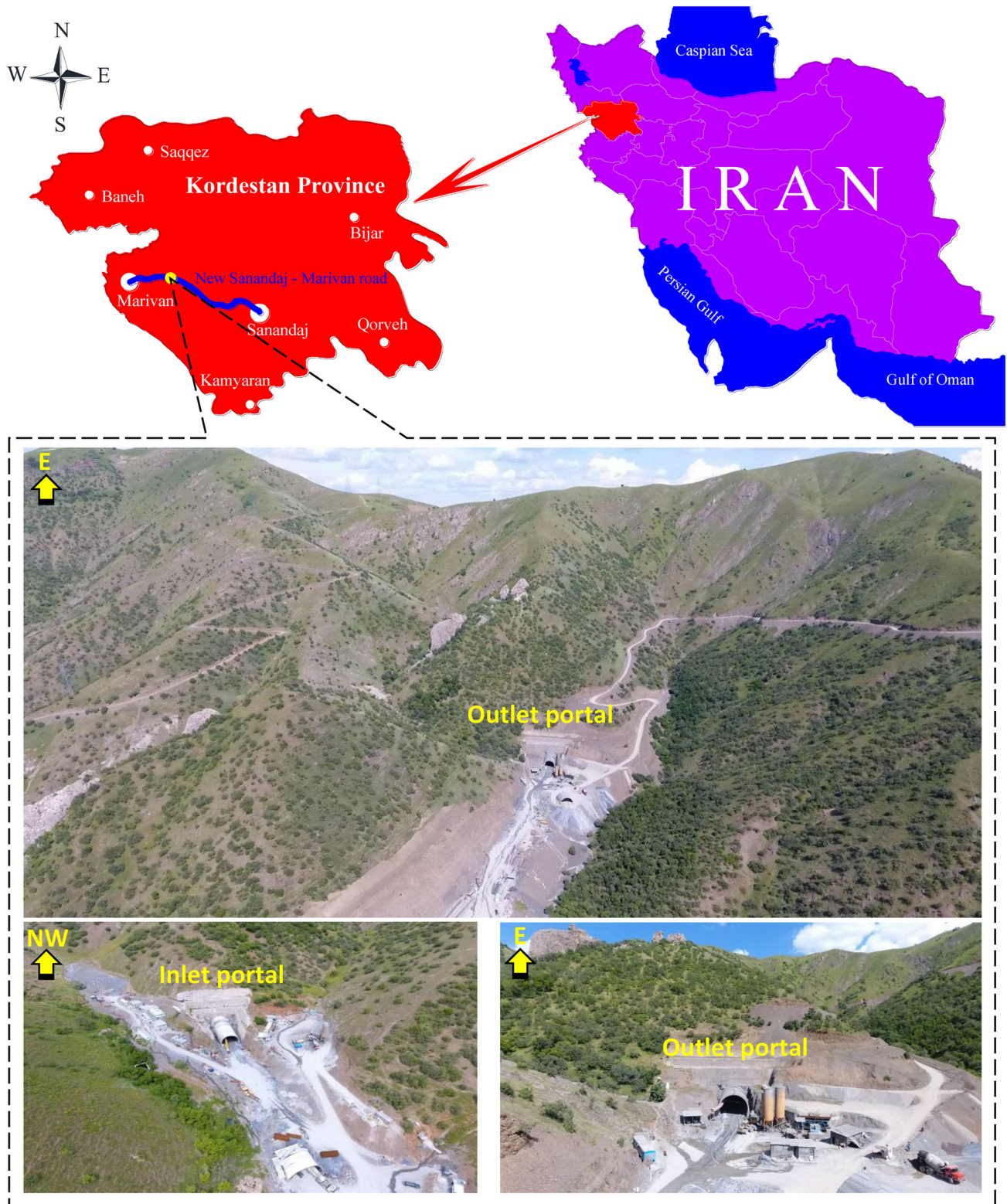
$$RRMSE = \sqrt{\left( \frac{1}{n} \right) \sum_{i=1}^n \left( \frac{f(x_i) - f^*(x_i)}{f(x_i)} \right)^2} \tag{19}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{f(x_i) - f^*(x_i)}{f(x_i)} \right| \tag{20}$$

$$MRE = \left( \frac{1}{n} \right) \sum_{i=1}^n \frac{|f(x_i) - f^*(x_i)|}{|f(x_i)|} \tag{21}$$

where  $f(x)$  is the actual value and  $f^*(x)$  is the predicted value,  $\bar{f}(x)$  are the means of actual and predicted values, and  $n$  is the number of datasets.

Because the discontinuities, water conditions, and uniaxial compressive strength of intact rock are very effective in determining the ground conditions of the tunnel route, and since the geomechanical parameter of RMR considers all of the aforementioned properties, the RMR parameter is used to define the geology conditions of the tunnel route. It

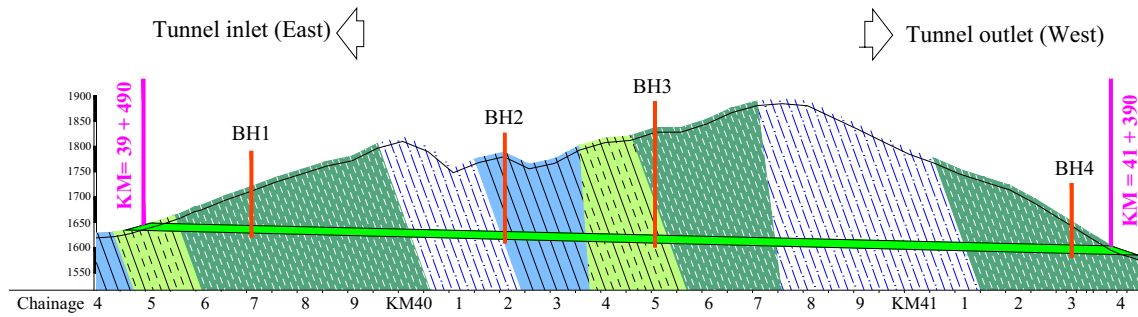


**Fig. 3** Project location of Garan tunnel

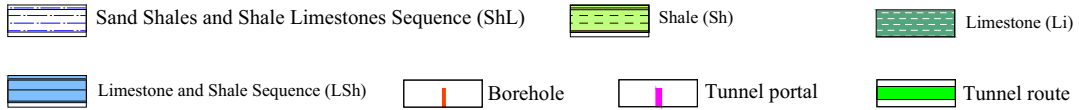
should be noted that in this study, the RMR values are expressed as an interval, but in the GPR, SVR, and DT tools, the mean of these two numbers is used. For example,

if the RMR parameter is 5–10, the value in the tools is considered equal to 7.5.

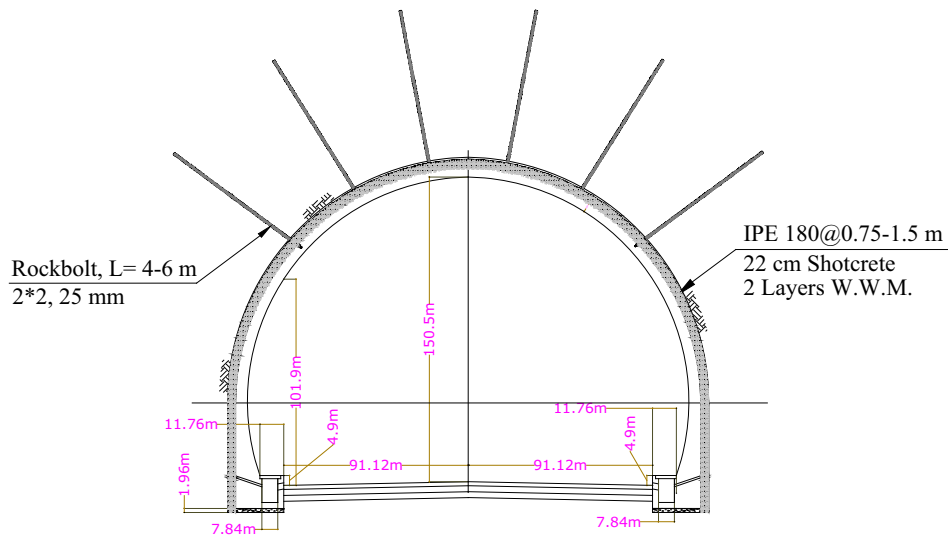




LEGEND:



(a)



(b)

Fig. 4 a Geological map of Garan tunnel, b Garan tunnel cross section associated with its projected support system

Table 2 The pre-construction training datasets of the RMR along Garan tunnel rout

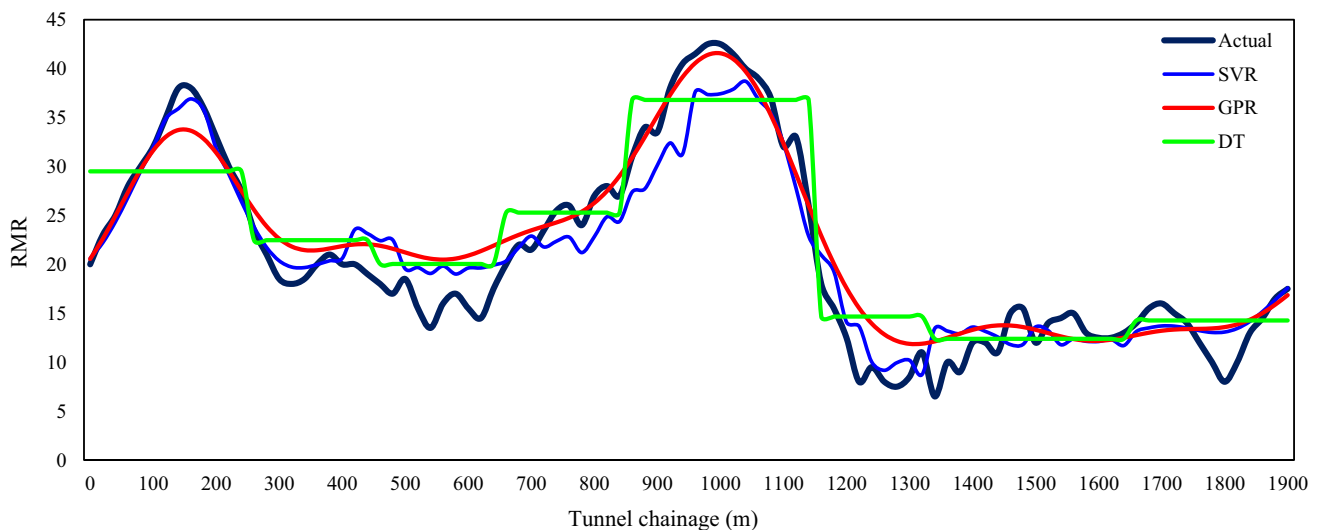
Tunnel chainage (m)	RMR	Tunnel chainage (m)	RMR	Tunnel chainage (m)	RMR	Tunnel chainage (m)	RMR
0 (inlet portal)	15–25	500	35–40	990 (BH3)	40–45	1500	10–15
50	25–30	550	30–35	1000	35–40	1550	5–10
100	25–30	600	30–35	1050	35–40	1600	5–10
150	30–35	650	30–35	1100	30–35	1650	5–10
200	30–35	700	25–30	1150	25–30	1690 (BH4)	10–15
210 (BH1)	25–35	710 (BH2)	20–25	1200	25–30	1700	10–15
250	20–25	750	25–30	1250	25–30	1750	15–20
300	25–30	800	20–25	1300	20–25	1800	15–20
350	25–30	850	25–30	1350	15–20	1850	15–20
400	30–35	900	35–40	1400	15–20	1900 (outlet portal)	15–20
450	25–30	950	40–45	1450	15–20		

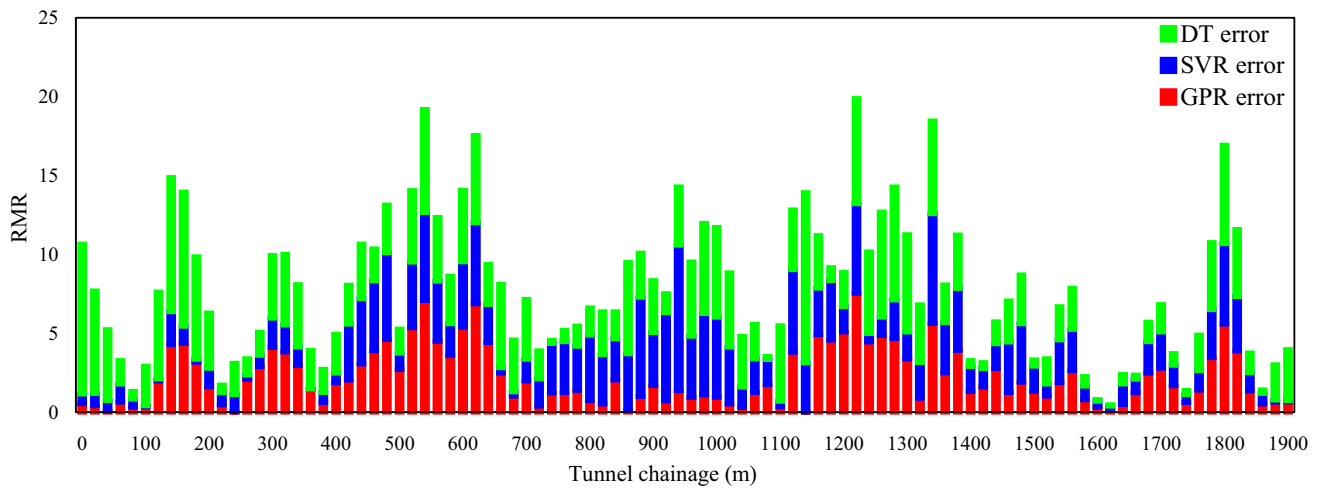
**Table 3** Pre-updating initial values of the GPR, SVR, and DT parameters for the RMR parameter prediction

Tool	Parameter	Value or type
GPR	KernelFunction (form of the covariance function)	'SquaredExponential'
	BasisFunction (explicit basis in the GPR model)	'Constant'
	Beta (initial value of the coefficients for the explicit basis)	21.7096
	Sigma (initial value for the noise standard deviation of the Gaussian process model)	4.2506
	FitMethod (method to estimate parameters of the GPR model)	Exact Gaussian Process Regression
SVR	Kernel function	'Gaussian'
	Epsilon (half the width of epsilon-insensitive band)	1.1490
	Solver (optimization routine)	'SMO'
	Bias	21.0725
	Mu	948.5612
	Sigma	558.1550
DT	'PredictorSelection'—algorithm used to select the best split predictor	'allsplits'
	SplitCriterion—split criterion	'MSE'
	'Prune'—flag to estimate optimal sequence of pruned subtrees	on'
	'MaxNumSplits'—maximal number of decision splits	138
	'MinLeafSize'—minimum number of leaf node observations	12
	'MinParentSize'—minimum number of branch node observations	24

**Table 4** The pre-updating results for RMR parameter

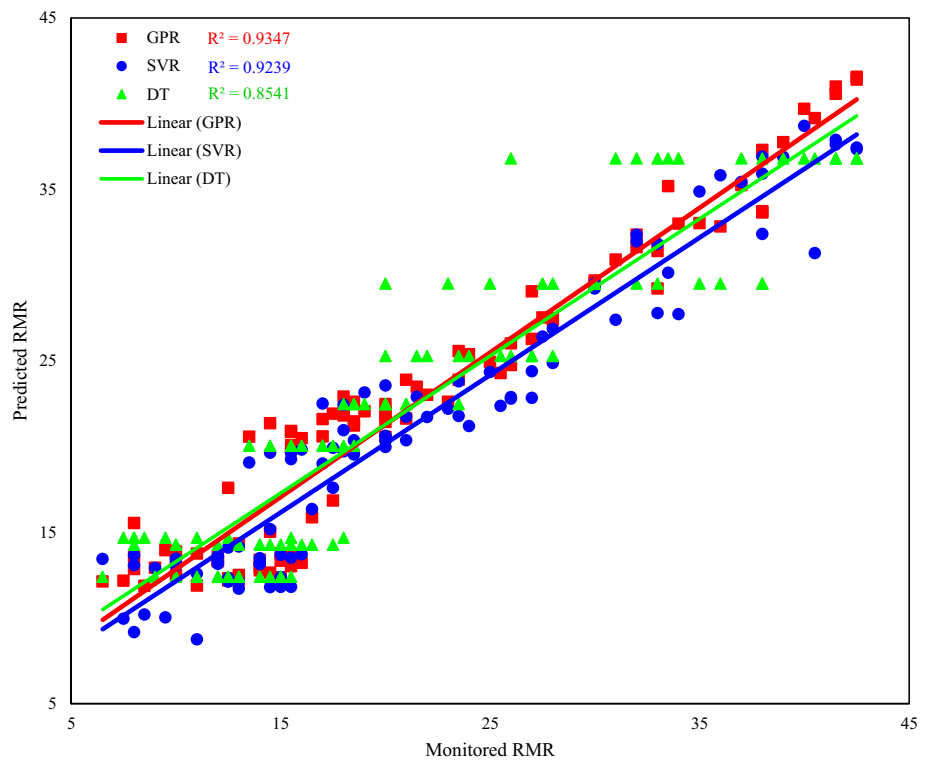
Parameter	Tool	$R^2$	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
RMR	GPR	0.83	2.9415	16.1023	4.0127	0.2534	17.2342	0.172342
	SVR	0.77	3.2662	22.3540	4.7280	0.2479	18.0983	0.180983
	DT	0.76	3.7534	22.5331	4.7469	0.2821	20.3073	0.203073

**Fig. 5** The pre-updating predicted results of GPR, SVR, and DT tools for RMR parameter along Garan tunnel route



**Fig. 6** The pre-updating results of prediction errors of GPR, SVR, and DT tools for RMR parameter along Garan tunnel route

**Fig. 7** The pre-updating results of prediction performance of GPR, SVR, and DT tools on the RMR parameter



**Table 5** The pre-updating prediction results for the RMR parameter along the Garan tunnel route

Parameter	Tool	$R^2$	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
RMR	GPR	0.9347	2.2518	8.3195	2.8843	0.2451	15.7904	0.157904
	SVR	0.9239	2.9303	12.080	3.4756	0.2605	16.8646	0.168646
	DT	0.8541	3.2037	15.472	3.9334	0.2805	19.2777	0.192777

## 4 Results analysis and comparison

In this paper, predictions are made in two steps of pre-updating and post-updating. In the following subsections, the results of each step will be discussed in more detail.

### 4.1 Pre-updating results

Before starting a tunneling project, a geological map is prepared by the engineers (Fig. 4a). This map is obtained



**Table 7** Pre-updating values of the GPR, SVR, and DT parameters for predicting the construction time and costs

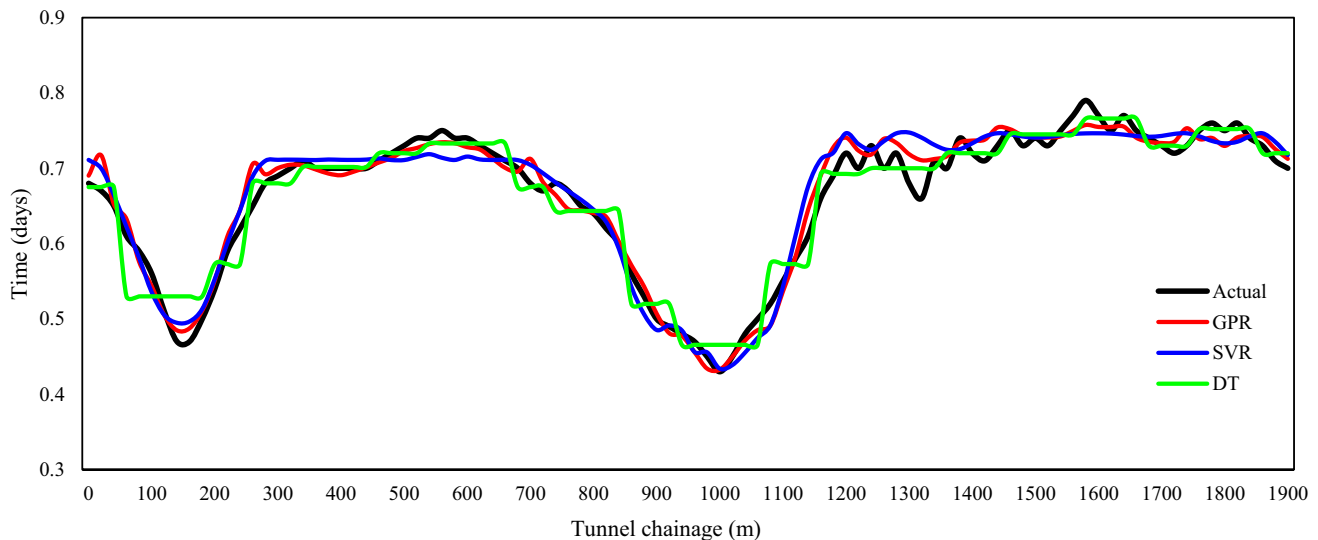
	Tool	Parameter	Value or type	
Time	GPR	KernelFunction (form of the covariance function)	'Exponential'	
		BasisFunction (explicit basis in the GPR model)	'Constant'	
		Beta (initial value of the coefficients for the explicit basis)	0.5613	
		Sigma (initial value for the standard deviation of the Gaussian process model)	0.0296	
		FitMethod (method to estimate parameters of the GPR model)	'Exact Gaussian Process Regression'	
	SVR	Kernel function	'Gaussian'	
		Epsilon (half the width of epsilon-insensitive band)	0.0254	
		Solver (optimization routine)	'SMO'	
		Bias	0.4877	
		Mu	46.5461	
	DT		Sigma	26.0222
		'PredictorSelection'—algorithm used to select the best split predictor	'allsplits'	
		SplitCriterion—split criterion	'MSE'	
		'Prune'—flag to estimate optimal sequence of pruned subtrees	'on'	
		'MaxNumSplits'—maximal number of decision splits	140	
Costs	GPR	'MinLeafSize'—minimum number of leaf node observations	4	
		'MinParentSize'—minimum number of branch node observations	10	
		KernelFunction (form of the covariance function)	'Exponential'	
		BasisFunction (explicit basis in the GPR model)	'Constant'	
		Beta (initial value of the coefficients for the explicit basis)	5.0107e+03	
	SVR	Sigma (initial value for the standard deviation of the Gaussian process model)	40.2888	
		FitMethod (method to estimate parameters of the GPR model)	'Exact Gaussian Process Regression'	
		Kernel function	'Gaussian'	
		Epsilon (half the width of epsilon-insensitive band)	50	
		Solver (optimization routine)	'SMO'	
	DT	Bias	5.2248e+03	
		Mu	46.5461	
		Sigma	26.0222	
		'PredictorSelection'—algorithm used to select the best split predictor	'allsplits'	
		SplitCriterion—split criterion	'MSE'	
		'Prune'—flag to estimate the optimal sequence of pruned subtrees	'on'	
		'MaxNumSplits'—maximal number of decision splits	140	
		'MinLeafSize'—minimum number of leaf node observations	4	
		'MinParentSize'—minimum number of branch node observations	10	

**Table 8** The pre-updating prediction results of construction time and construction costs parameters

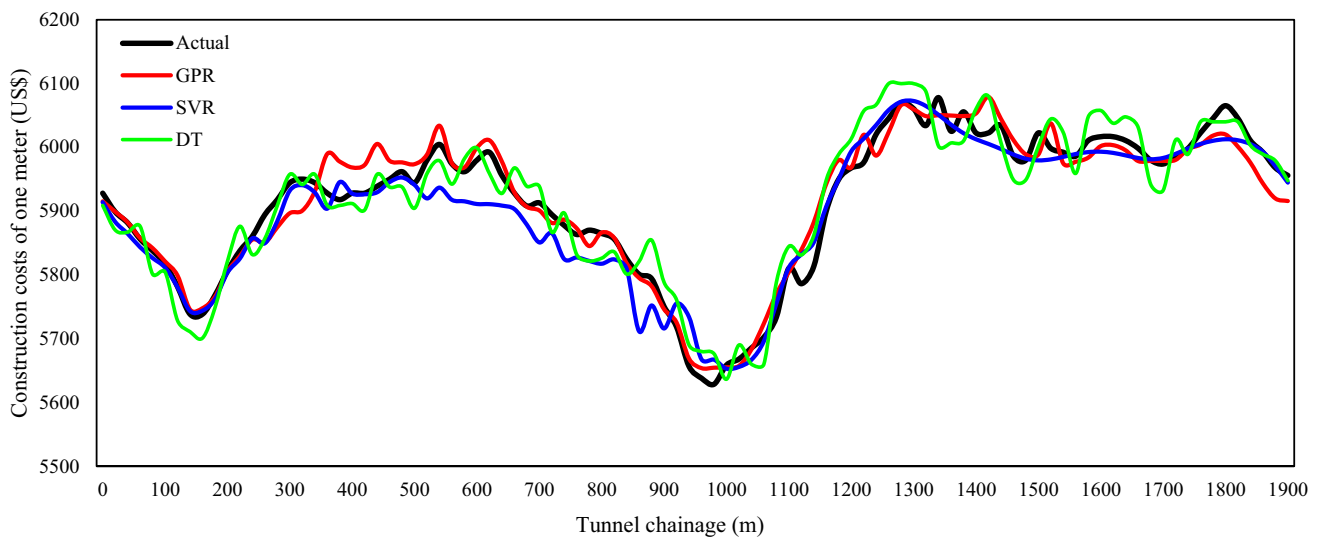
Parameter	Tool	$R^2$	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
Time (days)	GPR	0.97	0.026009	0.00097798	0.031273	0.05261	3.98974	0.03989
	SVR	0.97	0.026206	0.00104670	0.032352	0.05555	4.11389	0.04113
	DT	0.97	0.027596	0.00115760	0.034024	0.05742	4.34550	0.04345
Costs (US\$)	GPR	1.00	31.348	1948.9	44.146	0.00616	0.40179	0.00401
	SVR	0.99	39.510	2280.4	47.753	0.00713	0.42149	0.00421
	DT	0.99	37.884	2691.0	51.875	0.00746	0.46303	0.00463

from preliminary investigations of the study area through different surveys of geotechnical, geophysical, etc.

In this map, the status of important geological parameters that influence the tunnel construction is specified in the



**Fig. 8** The pre-updating prediction results of construction time obtained by GPR, SVR and DT along Garan tunnel route



**Fig. 9** The pre-updating prediction results of construction costs obtained by GPR, SVR, and DT along the Garan tunnel route

tunnel path. One of these parameters is the RMR parameter. In this paper, these data are used as the preliminary training data of the forecasting models. Note that these data are not very accurate due to the unknown subsurface conditions before the tunnel construction, but at the first stage of forecasts, there is no alternative. The initial training data of the RMR parameter available for the Garan tunnel route are presented in Table 2.

For this step of predictions (pre-updating prediction step for RMR parameter), the initial values for the hyperparameters of the GPR, SVR, and DT tools are presented in Table 3. The validation criteria results of fivefold cross-validation are presented in Table 4. According to the results, all the machine learning tools of GPR, SVR, and

DT have presented a good accuracy in the predictions. All the predictions are close together. However, the GPR provides more accurate results compared to the SVR and DT models.

To obtain the RMR status along the Garan tunnel route, after the fivefold cross-validation, the obtained models of the GPR, SVR, and DT tools were exported, and the datasets available along Garan tunnel route were applied as the test datasets in the exported models. Finally, the RMR parameter status was predicted by the GPR, SVR, and DT tools, as shown in Fig. 5. Also, to validate the prediction results accuracy of the GPR, SVR, and DT tools along the Garan tunnel route, the actual RMR status of the tunnel route is shown in Fig. 5. Looking at Fig. 5, it can be seen

that all the methods have presented close results together. Also, a comparison of the prediction results with the actual model of the Garan tunnel route shows the high accuracy of the GPR, SVR, and DT tools in predicting the RMR parameter.

The prediction errors of the GPR, SVR, and DT tools in predicting the RMR parameter along the Garan tunnel route are shown in Fig. 6. The mean error values for the GPR, SVR, and DT tools are less than 2.2518, 2.9303, and 3.2037, respectively. In Fig. 7, the  $R^2$  values of the regression results by the GPR, SVR, and DT tools versus the field-observed RMR parameter are shown. The  $R^2$  values for the GPR, SVR, and DT tools are evaluated by 0.9347, 0.239, and 0.8541, respectively.

Through comparisons made between the prediction results of the RMR parameter along the Garan tunnel route with the actual mode, Table 5 was obtained. In Table 5, the results of the different validation criteria for this step of prediction are presented. From all of these results obtained for the Garan tunnel route, it can be concluded that the GPR, SVR, and DT results are close together and to the actual mode, but still, GPR performs better prediction results than SVR, and SVR results are more accurate than DT.

In the next step, to predict the time and cost of constructing the tunnel, several input data presented in Table 6 were used as the train datasets of the GPR, SVR, and DT. These data were collected from the previously constructed road tunnels, which were similar to the Garan tunnel in terms of drilling, support system, area, and cross-sectional shape of the tunnel. The data of Table 6 explain the relationship between the RMR parameter and the time and costs of tunnel construction. For this step of prediction, the

initial values used for the parameters of the GPR, SVR, and DT are presented in Table 7.

After applying the datasets presented in Table 6 as the training datasets of the GPR, SVR, and DT tools, the different evaluation statistical criteria resulting from the fivefold cross-validation were obtained as shown in Table 8. Looking at Table 8, still, like the previous prediction step (for RMR parameter), the GPR results have more accuracy than the SVR and DT results. Also, SVR provides higher accuracy compared to the DT.

Then, the fivefold cross-validation models obtained by the regression learner app of the MATLAB 2018 tool were exported and the construction time and construction costs of Garan tunnel route were applied in the exported models as the test datasets. Finally, according to Figs. 8 and 9, respectively, the construction time and construction costs of the Garan tunnel were predicted in the whole tunnel route. Also, to validate the prediction results along the Garan tunnel route, the actual construction time and costs are shown in Figs. 8 and 9, respectively. Considering Figs. 8 and 9, it can be concluded that all the GPR, SVR, and DT tools have presented good results, and they all are close together.

The prediction errors of the GPR, SVR, and DT for the two parameters of construction time and construction costs along the Garan road tunnel are shown in Figs. 10 and 11, respectively. For the construction time, the mean error values of the GPR, SVR, and DT are less than 0.0143 days, 0.0170 days, and 0.0175 days, respectively. These values for the construction costs are less than US\$20.6073, US\$23.3594, and US\$30.4157, respectively. In Figs. 12 and 13, the  $R^2$  values of the regression results by the GPR, SVR, and DT tools versus the field-observed construction time and costs along the tunnel route are shown. For the

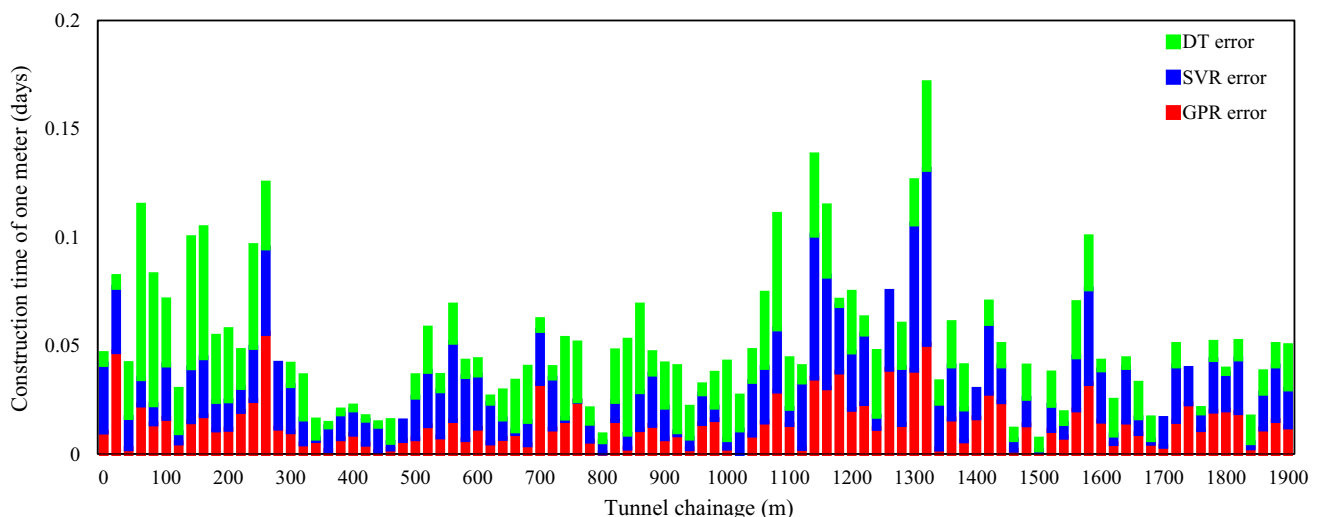
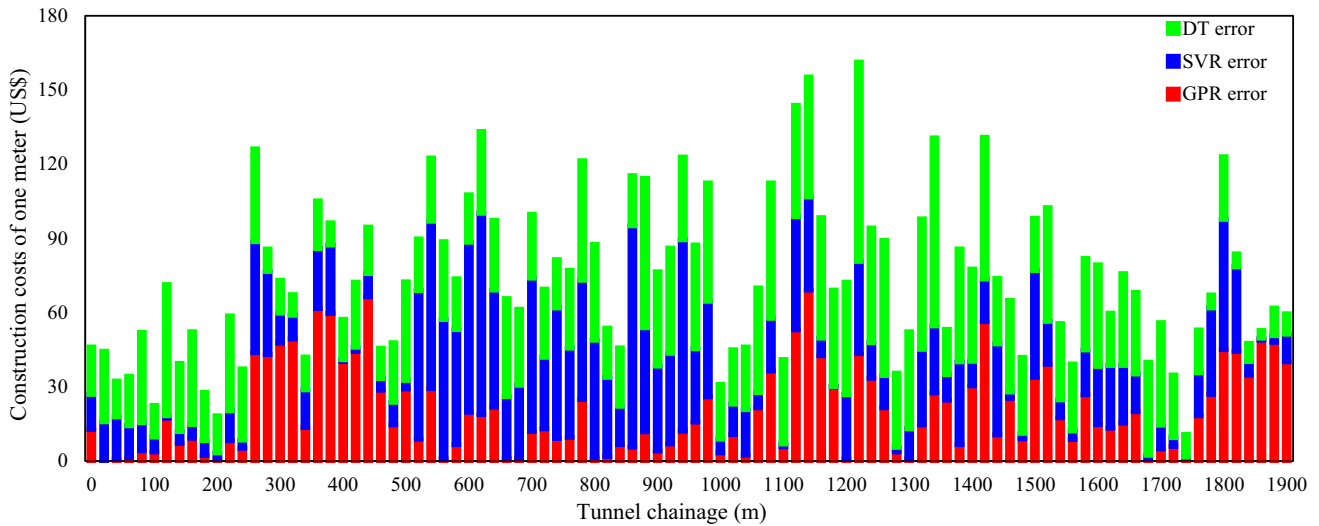
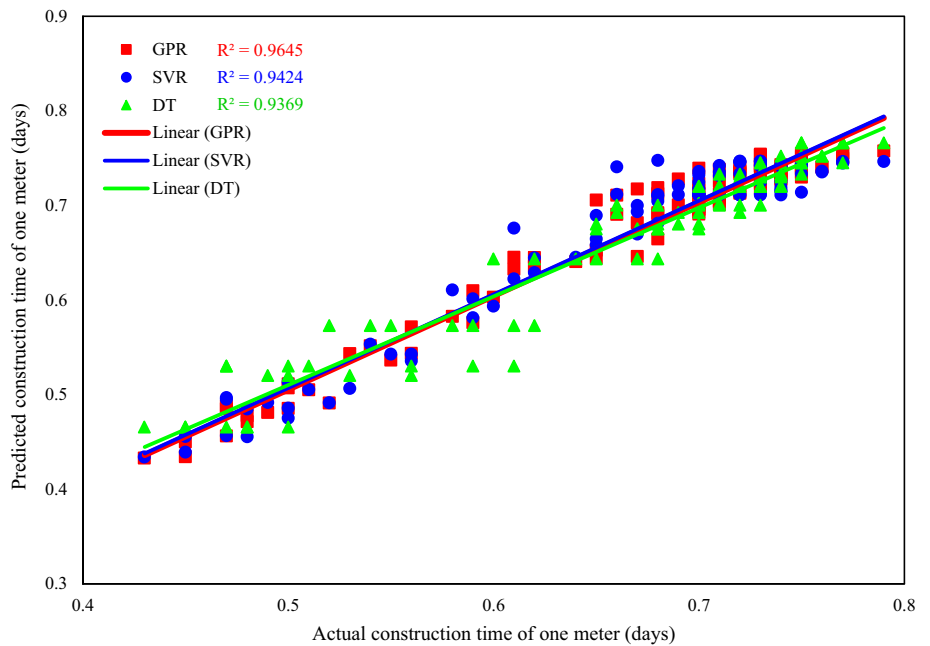


Fig. 10 The pre-updating prediction errors of construction time obtained by GPR, SVR and DT along Garan tunnel route



**Fig. 11** The pre-updating prediction errors of construction costs obtained by GPR, SVR and DT along garan tunnel route

**Fig. 12** The pre-updating prediction performance of GPR, SVR, and DT tools on construction time



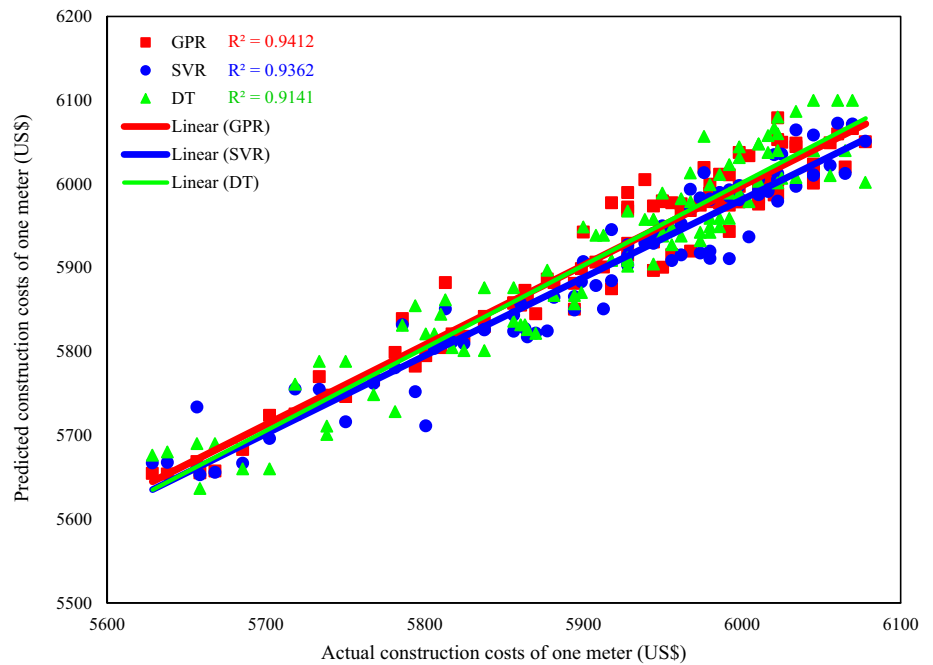
construction time (Fig. 12), the  $R^2$  values of the GPR, SVR, and DT are equal to 0.9645, 0.9424, and 0.9369, respectively. Also, for the construction costs (Fig. 13), the  $R^2$  values of the GPR, SVR, and DT are equal to 0.9412, 0.9362, and 0.9141, respectively.

Through comparison of the construction time and cost results predicted by the GPR, SVR, and DT tools along the Garan tunnel route with the actual mode, the statistical evaluation criteria applied in this article were estimated (Table 9). From all of these results, it can be seen that the GPR, SVR, and DT tools have presented good results. Still, the GPR results are better than the SVR and DT. Also, SVR has presented more accurate results than DT.

Until now, the status of the RMR, construction time, and construction costs were predicted by the GPR, SVR, and DT tools before Garan tunnel construction (pre-updating step). All the results predicted by the GPR, SVR, and DT tools were compared together through the different statistical evaluation criteria in the two modes of fivefold cross-validation and along Garan tunnel route. Comparing the predicted results with the field-observed mode, it was concluded that the GPR tool provides more accurate results than the SVR and DT. Also, the SVR performs better results than DT. All of these predictions were related to the construction of the Garan tunnel, in which datasets were gathered from the pre-existing tunnel and the previously



**Fig. 13** The pre-updating prediction performance of GPR, SVR, and DT tools on construction costs



**Table 9** The pre-updating prediction results of construction time and costs parameters along the Garan tunnel route

Parameter	Tool	$R^2$	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
Time (days)	GPR	0.9645	0.0143	0.0003	0.0181	0.0277	2.1762	0.0217
	SVR	0.9424	0.0170	0.0005	0.0235	0.0360	2.8591	0.0285
	DT	0.9369	0.0175	0.0005	0.0238	0.0409	2.8780	0.0287
Costs (US\$)	GPR	0.9412	20.6073	739.9957	27.2028	0.0045	0.3473	0.0034
	SVR	0.9362	23.3594	962.1546	31.0186	0.0052	0.3954	0.0039
	DT	0.9141	30.4157	1151.9117	33.9398	0.0057	0.5149	0.0051

constructed road tunnels. In the next step, in addition to the initial datasets, new input datasets obtained during the tunnel construction are used to train the GPR, SVR, and DT tools, and the previously predicted results will be updated. Also, the impact of these updates on the predicted results will be investigated.

**4.2 Post-updating results**

To update the previously forecasted results, about 200 m from both inlet and outlet portals of the tunnel was assumed as the constructed parts, and the data obtained in these locations were used as the new train datasets. Also, for the post-updating stage, the initial values used for the parameters of the GPR, SVR, and DT tools for each of the three parameters of RMR, construction time, and construction costs are presented in Table 10.

To update the previously predicted results of the RMR parameter, in addition to the pre-updating training datasets (Table 2), the more training datasets were obtained from the constructed parts of the tunnel and the number of the

dataset was raised. Therefore, machine learning tools are applicable to the broader range [50]. The post-updating results of the RMR parameter are presented in Table 11 through the statistical evaluation criteria. For this prediction step, the GPR results still have more accuracy than the SVR and DT. Also, SVR still provides better results than DT.

After applying the actual RMR parameter of the Garan tunnel route in the fivefold cross-validation exported models, the RMR parameter was predicted along the whole tunnel route by all of the GPR, SVR, and DT tools (Fig. 14). According to Fig. 14, all the results predicted by the GPR, SVR, and DT tools are very good and close together.

In Fig. 15, the error values of the GPR, SVR, and DT tools in predicting the RMR parameter are shown along the Garan tunnel route for the post-updating step. The mean error values for the GPR, SVR, and DT tools are less than 0.6049, 1.1269, and 2.0827, respectively. In comparison with the pre-updating results, the average values of errors in predicting the status of the RMR parameter are reduced

**Table 10** The post-updating values of the GPR, SVR, and DT parameters for the prediction of RMR, construction time, and construction costs

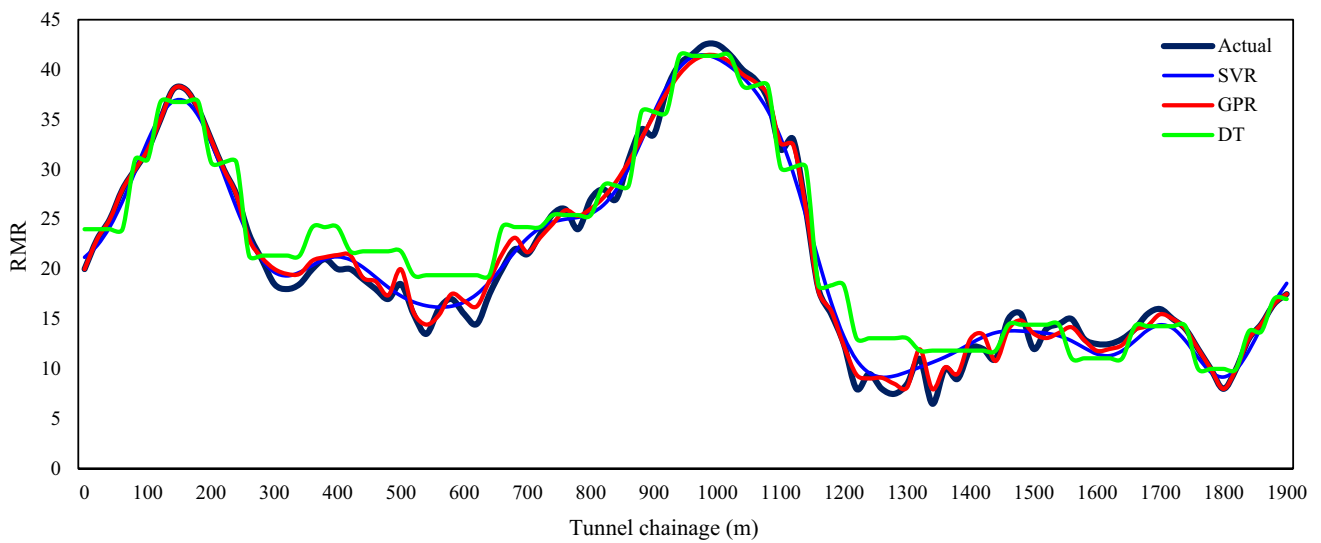
	Tool	Parameter	Value or type
RMR	GPR	KernelFunction (form of the covariance function)	'Matern5/2'
		BasisFunction (explicit basis in the GPR model)	'Constant'
		Beta (initial value of the coefficients for the explicit basis)	21.6675
		Sigma (initial value for the standard deviation of the Gaussian process model)	3.9565
		FitMethod (method to estimate parameters of the GPR model)	'Exact Gaussian Process Regression'
	SVR	Kernel function	'Gaussian'
		Epsilon (half the width of epsilon-insensitive band)	1.1861
		Solver (optimization routine)	'SMO'
		Bias	21.1476
		Mu	948.7417
	DT	'PredictorSelection'—algorithm used to select the best split predictor	'allsplits'
		SplitCriterion—split criterion	'MSE'
		'Prune'—flag to estimate the optimal sequence of pruned subtrees	'on'
		'MaxNumSplits'—maximal number of decision splits	150
'MinLeafSize'—minimum number of leaf node observations		4	
Time	GPR	KernelFunction (form of the covariance function)	'Exponential'
		BasisFunction (explicit basis in the GPR model)	'Constant'
		Beta (initial value of the coefficients for the explicit basis)	0.5621
		Sigma (initial value for the standard deviation of the Gaussian process model)	0.0267
		FitMethod (method to estimate parameters of the GPR model)	'Exact Gaussian Process Regression'
	SVR	Kernel function	'Gaussian'
		Epsilon (half the width of epsilon-insensitive band)	0.0252
		Solver (optimization routine)	'SMO'
		Bias	0.4839
		Mu	35.7018
	DT	'PredictorSelection'—algorithm used to select the best split predictor	'allsplits'
		SplitCriterion—split criterion	'MSE'
		'Prune'—flag to estimate the optimal sequence of pruned subtrees	'on'
		'MaxNumSplits'—maximal number of decision splits	222
'MinParentSize'—minimum number of branch node observations		10	
Costs	GPR	KernelFunction (form of the covariance function)	'Exponential'
		BasisFunction (explicit basis in the GPR model)	'Constant'
		Beta (initial value of the coefficients for the explicit basis)	5.0031 + 03
		Sigma (initial value for the standard deviation of the Gaussian process model)	34.8192
		FitMethod (method to estimate parameters of the GPR model)	'Exact Gaussian Process Regression'
	SVR	Kernel function	'Gaussian'
		Epsilon (half the width of epsilon-insensitive band)	43.5693
		Solver (optimization routine)	'SMO'
		Bias	2.2174e+03
		Mu	35.7018
	DT	Sigma	25.4124
		'PredictorSelection'—algorithm used to select the best split predictor	'allsplits'
		SplitCriterion—split criterion	'MSE'
		'Prune'—flag to estimate the optimal sequence of pruned subtrees	'on'

**Table 10** continued

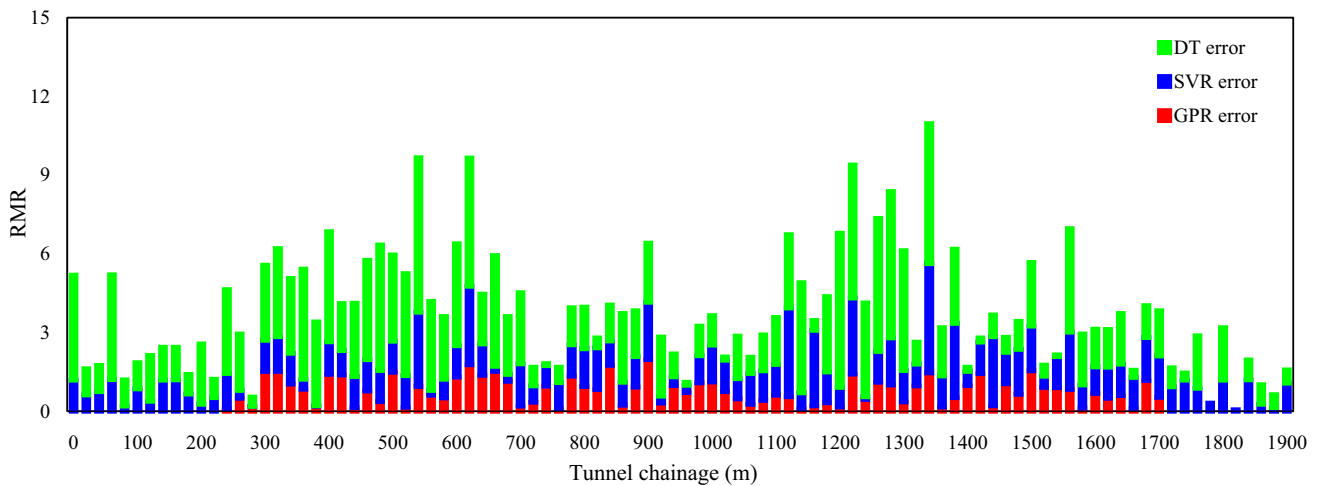
Tool	Parameter	Value or type
	‘MaxNumSplits’—maximal number of decision splits	222
	‘MinLeafSize’—minimum number of leaf node observations	1
	‘MinParentSize’—minimum number of branch node observations	10

**Table 11** The post-updating results of the RMR parameter

Parameter	Tool	$R^2$	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
RMR	GPR	0.9065	1.6391	9.5817	3.0954	0.1504	8.3182	0.083182
	SVR	0.9167	2.1726	11.606	3.4068	0.1859	11.2034	0.112034
	DT	0.8691	2.5492	12.916	3.5939	0.2115	14.2883	0.142883

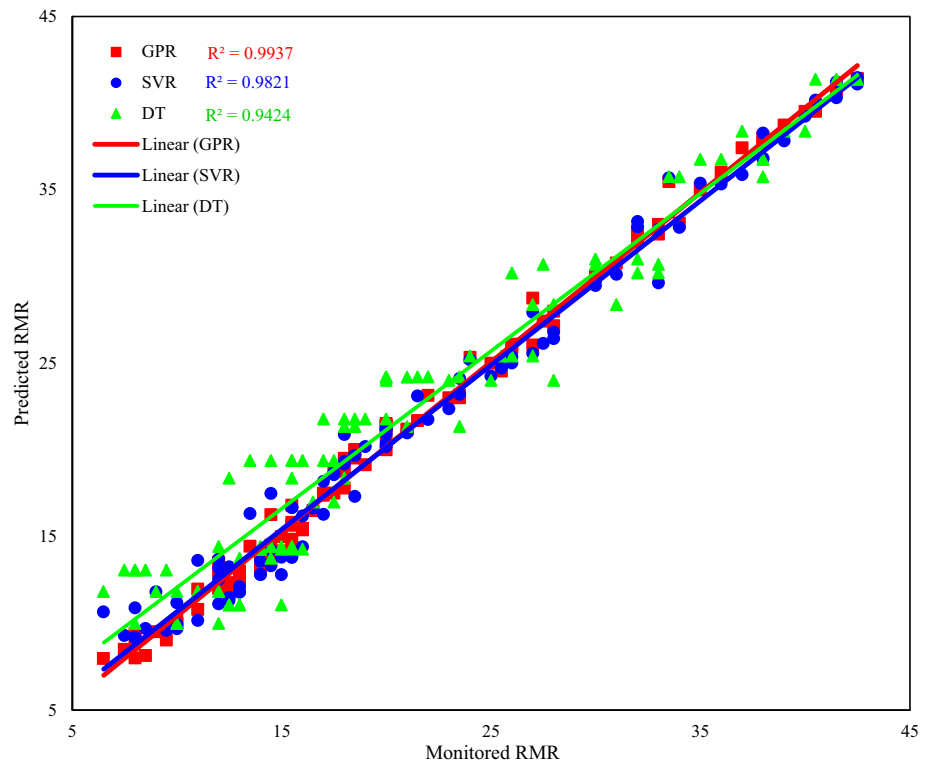


**Fig. 14** The post-updating prediction results of GPR, SVR, and DT for the RMR parameter along the Garan tunnel route



**Fig. 15** The post-updating prediction error results of GPR, SVR, and DT tools for RMR parameter along Garan tunnel route

**Fig. 16** The updated prediction performance of GPR, SVR, and DT on the RMR parameter



**Table 12** The post-updating prediction results of the RMR parameter along the Garan tunnel route

Parameter	Tool	R <sup>2</sup>	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
RMR	GPR	0.9937	0.6049	0.6633	0.8144	0.0564	3.7215	0.037215
	SVR	0.9821	1.1269	1.8436	1.3577	0.1133	7.3370	0.073370
	DT	0.9424	2.0827	6.7651	2.6009	0.2114	13.7110	0.137110

by 1.647, 1.8034, and 1.121, for GPR, SVR, and DT, respectively. Also, according to Fig. 16, in predicting the status of the RMR parameter, the R<sup>2</sup> values for the GPR, SVR, and DT increased by 0.059, 0.0582, and 0.0883, respectively.

Comparing the updated predictions of the RMR parameter along the Garan tunnel route with the actual RMR, the different statistical validation criteria are evaluated as shown in Table 12. From all of these results, it can be concluded that the GPR, SVR, and DT results are close together and to the actual mode, but GPR performs better than SVR, and SVR performs better than DT. Also, there is a significant upgrade potential while constructing the tunnel in predicting the RMR parameter.

The actual datasets of the RMR, construction time, and construction costs parameters that were obtained during Garan tunnel construction (200 m from both inlet and outlet portals) are given in Table 13. In order to update the previous prediction results of the construction time and costs parameters along the Garan tunnel route, in addition to the preliminary datasets obtained from the previously

constructed road tunnels (Table 6), the new obtained datasets presented in Table 13 were also used to train the forecasting tools.

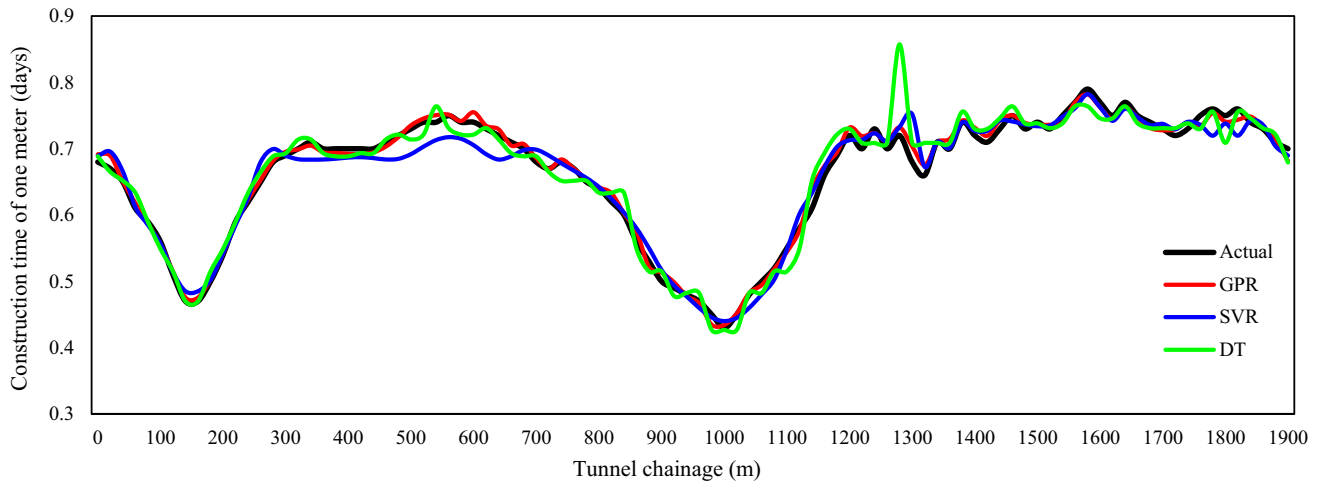
The post-updating results for the construction time and costs obtained by the GPR, SVR, and DT tools are compared together in Table 14 through the different statistical evaluation criteria. As the pre-updating step, GPR presented more accurate results than the SVR and DT tools. Also, the SVR tool still has shown a better potential ability in the predictions than the DT.

Using the construction time and costs datasets of the Garan tunnel route as the input datasets in the exported models of the fivefold cross-validation, the GPR, SVR, and DT tools have predicted the construction time and costs along Garan tunnel route as shown in Figs. 17 and 18, respectively. In Figs. 17 and 18, the updated forecasts of the GPR, SVR, and DT tools on the construction time and costs parameters are compared with the actual parameter one in each location along the tunnel route. According to Figs. 17 and 18, the more proximity between the predicted

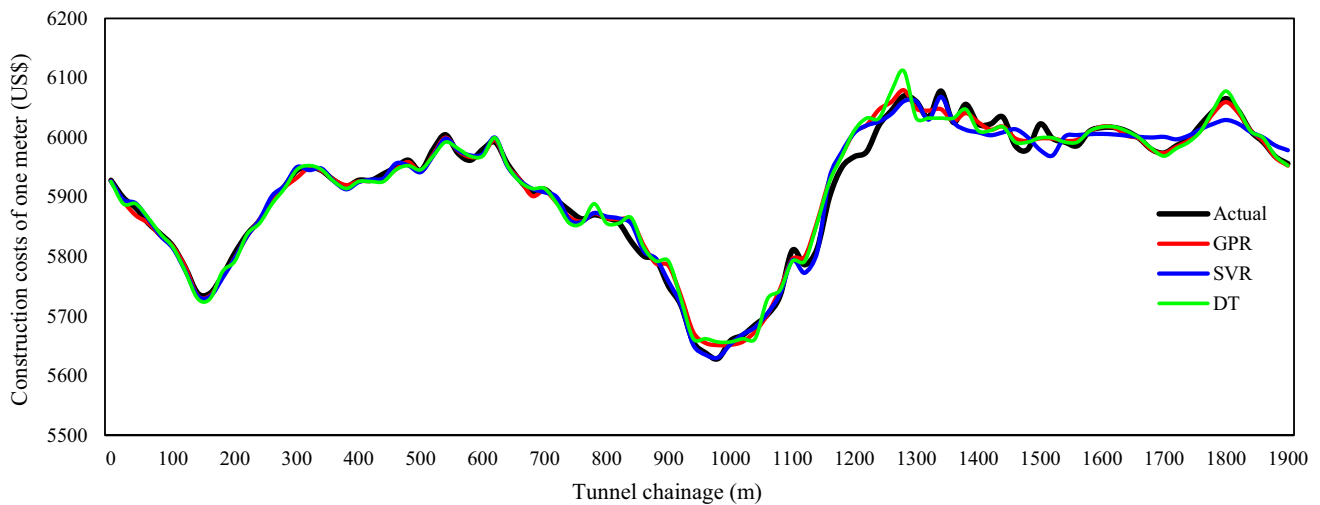


**Table 14** The post-updating prediction results of the construction time and construction costs parameters

Parameter	Tool	$R^2$	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
Time (days)	GPR	0.97	0.022581	0.00084215	0.02902	0.04767	3.57142	0.03571
	SVR	0.97	0.024631	0.0010258	0.032028	0.04820	3.55539	0.03555
	DT	0.97	0.025457	0.0011001	0.033167	0.04931	3.62189	0.03621
Costs (US\$)	GPR	1.00	25.223	1444.9	38.012	0.00569	0.34509	0.00345
	SVR	1.00	34.293	1757.6	41.923	0.00637	0.36274	0.00362
	DT	0.99	30.768	2187.8	46.774	0.00662	0.38458	0.00384



**Fig. 17** The post-updating prediction results of construction time obtained by GPR, SVR, and DT along Garan tunnel route

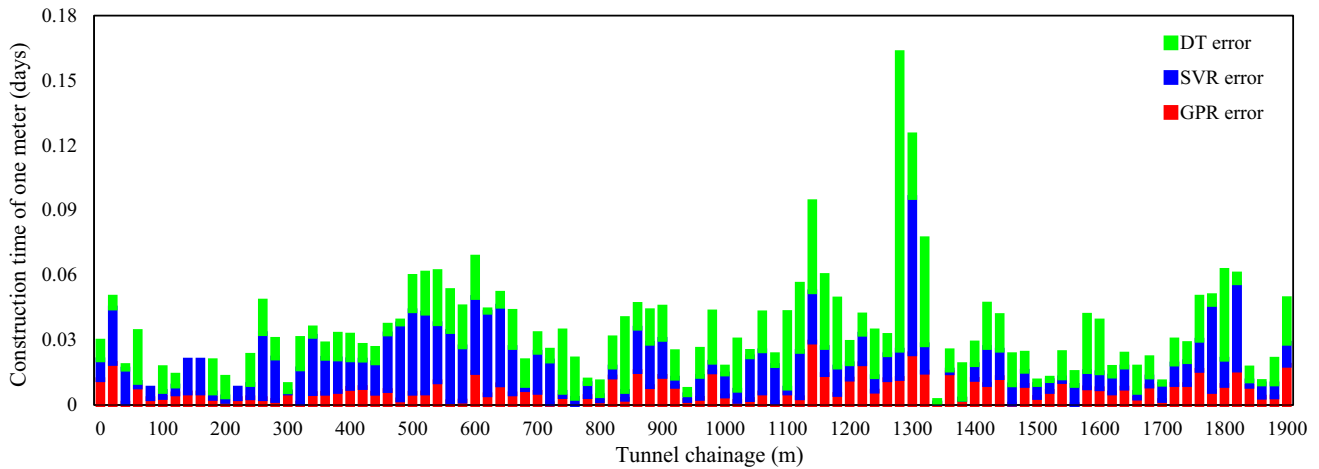


**Fig. 18** The post-updating prediction results of construction costs obtained by GPR, SVR, and DT along the Garan tunnel route

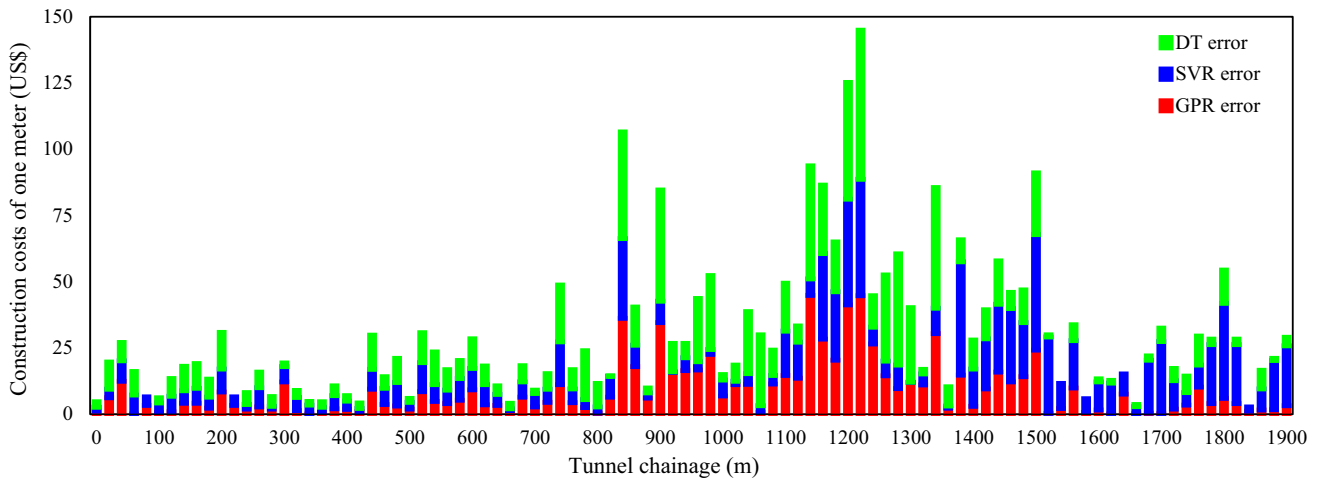
### 5 Discussion

Undoubtedly, reducing the uncertainties of the underground conditions in tunneling projects is one of the most serious issues. The most important of these uncertainties is the geological and geotechnical conditions of the tunnel route. The uncertainty of these conditions, in turn, increases the risk of the time and costs of the tunnel construction.

Therefore, it is necessary to reduce the uncertainty about the geological and geotechnical conditions of the tunnel route in order to reduce risk about the time and costs of constructing the tunnel. Usually, in road tunnels, limited data are available regarding the geological and geotechnical conditions of the tunnel route. Under such circumstances, it is not possible to estimate the time and costs of constructing the tunnel at an acceptable level. To reduce



**Fig. 19** The post-updating prediction error results of GPR, SVR, and DT tools for the construction time along Garan tunnel route



**Fig. 20** The post-updating prediction error results of GPR, SVR, and DT tools for the construction costs along Garan tunnel route

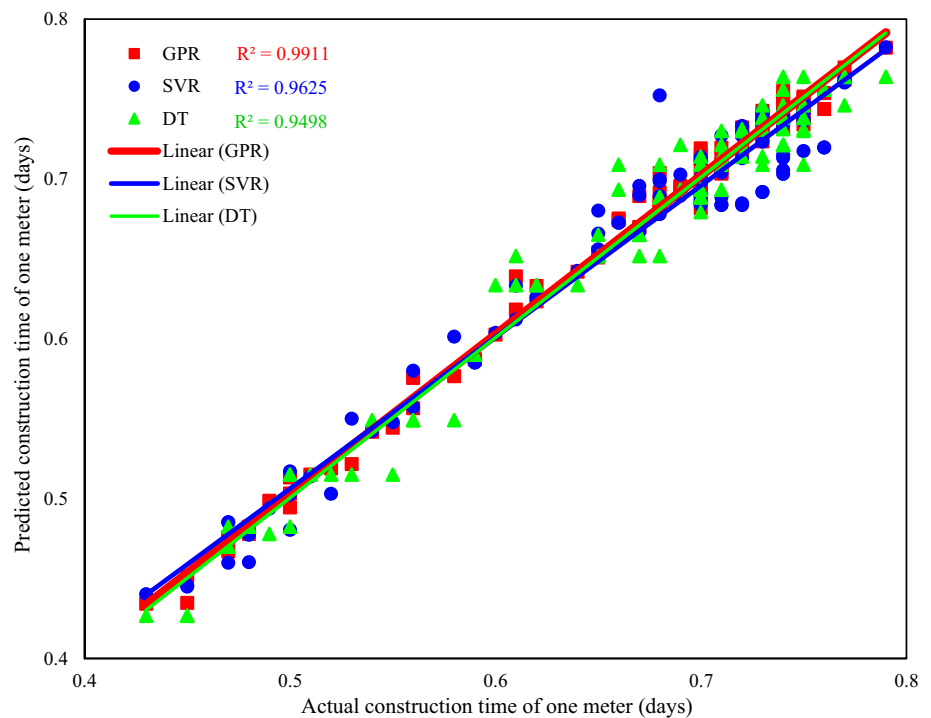
these uncertainties, the GPR, SVR, and DT tools were used to predict tunnel geology and time and costs of its construction. To illustrate the correct functioning of these tools, they were used in a case study of Garan road tunnel. To predict the tunnel route geology, the RMR parameter was considered. Initially, before the construction of the tunnel, the data about the RMR parameter were available only in the inlet and outlet portals and borehole locations. In this stage, only these data were applied to train the prediction tools. Then, the status of the RMR parameter was predicted in each location in the tunnel path by each of the tools. Since no data were available on the construction time and costs before construction of the tunnel, so to train the GPR, SVR, and DT tools, the previously constructed road tunnels were used. With the help of these old tunnels, for different values of the RMR parameter, different construction time and costs were obtained and used as the train datasets of the prediction tools. Then, according to the predetermined RMR values, the time and costs estimations

for each meter of construction in any position in the tunnel path were predicted by the GPR, SVR, and DT tools.

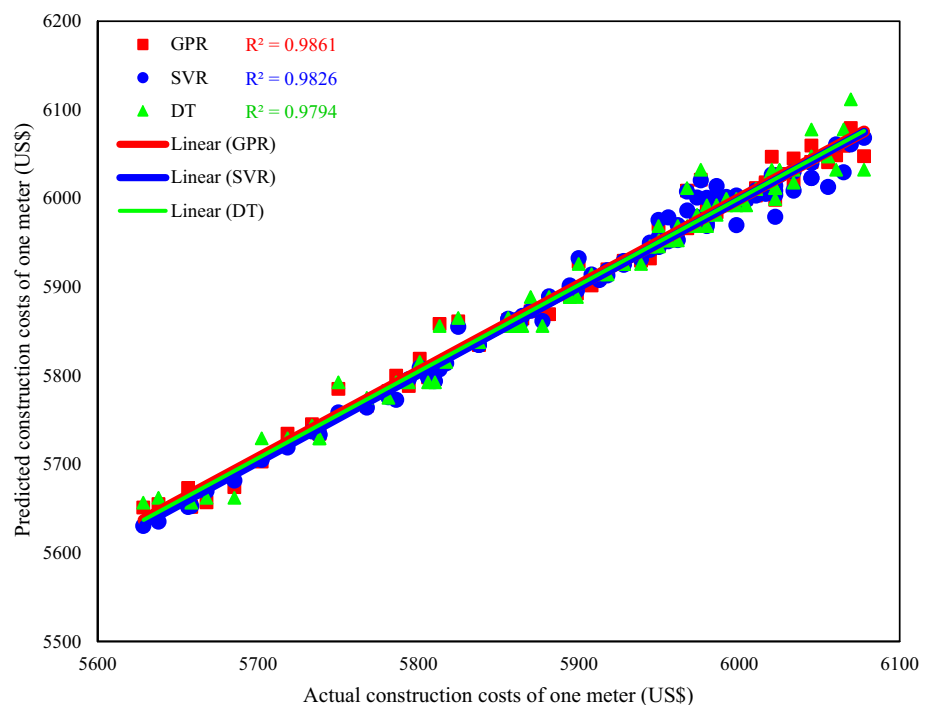
In the next step, to show the impact of upgrades during the tunnel construction on the previous predictions, it was assumed that only 200 m was constructed on the inlet and outlet of the tunnel route. Using actual data on the RMR parameter and construction time and costs in the constructed parts as the new train datasets, the previous forecasted results were updated.

To validate the prediction results of the GPR, SVR, and DT tools at each step, the fivefold cross-validation was adopted using the MATLAB tool 2018 through the regression learner app. In this application, there were several different model types of GPR, SVR, and DT tools. At the forecasting stage, all the model types were used individually, and the results of that model were taken into account that was more accurate. To assess the performances of selected models, seven statistical evaluation criteria of  $R^2$ , MAE, MSE, RMSE, RRMSE, MAPE, and

**Fig. 21** The updated prediction performance of GPR, SVR, and DT tools on construction time



**Fig. 22** The updated prediction performance of GPR, SVR, and DT tools on construction costs



MRE were used. All of these fivefold cross-validation results are summarized in Table 16. Also, in each step of the predictions, the fivefold cross-validation models were exported through MATLAB 2018 tool, and the datasets available along Garan tunnel route were applied in the models to predict the RMR, construction time, and construction costs along Garan tunnel path. For the Garan

tunnel route predictions, the predicted results were compared with the actual mode, and like the fivefold cross-validation results, also, the different statistical evaluation criteria were evaluated for these results. All of these results are summarized in Table 17.



**Table 15** The post-updating prediction results of the construction time and costs parameters along the Garan tunnel route

Parameter	Tool	$R^2$	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
Time (days)	GPR	0.9911	0.0073	0.00008	0.0090	0.0139	1.1011	0.0110
	SVR	0.9625	0.0138	0.00033	0.0182	0.0270	2.0689	0.0206
	DT	0.9498	0.0142	0.00046	0.0215	0.0321	2.1552	0.0215
Costs (US\$)	GPR	0.9861	8.7907	178.2121	13.3496	0.0022	0.1491	0.0014
	SVR	0.9826	10.3969	217.8698	14.7604	0.0024	0.1746	0.0017
	DT	0.9794	10.7844	261.7609	16.1790	0.0027	0.1827	0.0018

**Table 16** A summary of the all fivefold cross-validation results predicted by the GPR, SVR, and DT tools

Prediction step	Tool	Parameter	$R^2$	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
Pre-updating step	GPR	RMR	0.83	2.9415	16.1023	4.0127	0.2534	17.2342	0.172342
		Construction time (days)	0.97	0.026009	0.00097798	0.031273	0.05261	3.98974	0.03989
		Construction costs (US\$)	1.00	31.348	1948.9	44.146	0.00616	0.40179	0.00401
	SVR	RMR	0.77	3.2662	22.3540	4.7280	0.2479	18.0983	0.180983
		Construction time (days)	0.97	0.026206	0.00104670	0.032352	0.05555	4.11389	0.04113
		Construction costs (US\$)	0.99	39.510	2280.4	47.753	0.00713	0.42149	0.00421
	DT	RMR	0.76	3.7534	22.5331	4.7469	0.2821	20.3073	0.203073
		Construction time (days)	0.97	0.027596	0.00115760	0.034024	0.05742	4.34550	0.04345
		Construction costs (US\$)	0.99	37.884	2691.0	51.875	0.00746	0.46303	0.00463
Post-updating step	GPR	RMR	0.90	1.6391	9.5817	3.0954	0.1504	8.3182	0.083182
		Construction time (days)	0.97	0.022581	0.00084215	0.02902	0.04767	3.57142	0.03571
		Construction costs (US\$)	1.00	25.223	1444.9	38.012	0.00569	0.34509	0.00345
	SVR	RMR	0.91	2.1726	11.606	3.4068	0.1859	11.2034	0.112034
		Construction time (days)	0.97	0.024631	0.0010258	0.032028	0.04820	3.55539	0.03555
		Construction costs (US\$)	1.00	34.293	1757.6	41.923	0.00637	0.36274	0.00362
	DT	RMR	0.86	2.5492	12.916	3.5939	0.2115	14.2883	0.142883
		Construction time (days)	0.97	0.025457	0.0011001	0.033167	0.04931	3.62189	0.03621
		Construction costs (US\$)	0.99	30.768	2187.8	46.774	0.00662	0.38458	0.00384

## 6 Conclusions

In this article, the effects of geological uncertainties in tunnel construction time and costs were tried to be reduced through the machine learning tools of GPR, SVR, and DT. To train the prediction tools, data were gathered from the under-studying tunnel observations and the previously constructed tunnels. To assess the effect of updated outcomes during the tunnel construction, the tools used were updated for one time after constructing 200 m from both inlet and outlet portals of the tunnel. Also, in each stage of the predictions, the fivefold cross-validation was adopted. Finally, all the predicted results were validated through seven validation criteria of MAE, MSE, RMSE, RRMSE, MAPE, MRE, and  $R^2$ . The validations showed that the GPR results were more accurate than the SVR results, and

SVR has more accuracy than the DT. Also, during the tunnel construction, updating the previously predicted results by applying the actual data obtained in the constructed parts can significantly increase the accuracy of the predictions comparing to the pre-updating mode. In conclusion, according to the results of this study, machine learning methods are useful tools for solving problems with complex mechanisms such as geology, construction time, and construction costs of a road tunnel project. Furthermore, the existing models are suitable to use as predictors for tunnel future levels.

In this study, only one feature (RMR) was considered to predict the tunnel path geology, and multiple features can be considered to have a better description of tunnel geology.

**Table 17** A summary of all results along the Garan tunnel route obtained through the exported models of the fivefold cross-validation results

Prediction step	Tool	Parameter	R <sup>2</sup>	MAE	MSE	RMSE	RRMSE	MAPE (%)	MRE
Pre-updating step	GPR	RMR	0.9347	2.2518	8.3195	2.8843	0.2451	15.7904	0.157904
		Construction time (days)	0.9645	0.01430	0.0003000	0.01810	0.0277	2.1762	0.0217
		Construction costs (US\$)	0.9412	20.6073	739.99570	27.2028	0.0045	0.3473	0.0034
	SVR	RMR	0.9239	2.9303	12.080	3.4756	0.2605	16.8646	0.168646
		Construction time (days)	0.9424	0.01700	0.0005000	0.02350	0.0360	2.8591	0.0285
		Construction costs (US\$)	0.9362	23.3594	962.15460	31.0186	0.0052	0.3954	0.0039
	DT	RMR	0.8541	3.2037	15.472	3.9334	0.2805	19.2777	0.192777
		Construction time (days)	0.9369	0.01750	0.0005000	0.02380	0.0409	2.8780	0.0287
		Construction costs (US\$)	0.9141	30.4157	1151.9117	33.9398	0.0057	0.5149	0.0051
Post-updating step	GPR	RMR	0.9937	0.6049	0.6633	0.8144	0.0564	3.7215	0.037215
		Construction time (days)	0.9911	0.00730	0.0000800	0.00900	0.0139	1.1011	0.0110
		Construction costs (US\$)	0.9861	8.79070	178.21210	13.3496	0.0022	0.1491	0.0014
	SVR	RMR	0.9821	1.1269	1.8436	1.3577	0.1133	7.3370	0.073370
		Construction time (days)	0.9625	0.01380	0.0003300	0.01820	0.0270	2.0689	0.0206
		Construction costs (US\$)	0.9826	10.3969	217.86980	14.7604	0.0024	0.1746	0.0017
	DT	RMR	0.9424	2.0827	6.7651	2.6009	0.2114	13.7110	0.137110
		Construction time (days)	0.9498	0.01420	0.0004600	0.02150	0.0321	2.1552	0.0215
		Construction costs (US\$)	0.9794	10.7844	261.76090	16.1790	0.0027	0.1827	0.0018

## Compliance with ethical standards

**Conflict of interest** Authors declare no conflict of interest.

## References

- Flyvbjerg B, Holm MS, Buhl S (2002) Underestimating costs in public works projects: error or lie? *J Am Plan Assoc* 68:279–295. <https://doi.org/10.1080/01944360208976273>
- Wang S, Li L, Shi S, Cheng S, Hu H, Wen T (2020) Dynamic risk assessment method of collapse in mountain tunnels and application. *Geotech Geol Eng*. <https://doi.org/10.1007/s10706-020-01196-7>
- Zhou H, Zhao Y, Shen Q, Yang L, Cai H (2020) Risk assessment and management via multi-source information fusion for undersea tunnel construction. *Autom Constr* 111:103050. <https://doi.org/10.1016/j.autcon.2019.103050>
- Wang X, Shi K, Shi Q, Dong H, Chen M (2020) A normal cloud model-based method for risk assessment of water inrush and its application in a super-long tunnel constructed by a tunnel boring machine in the arid area of Northwest China. *Water* 12:644. <https://doi.org/10.3390/w12030644>
- Shahrour I, Bian H, Xie X, Zhang Z (2020) Use of smart technology to improve management of utility tunnels. *Appl Sci* 10:711. <https://doi.org/10.3390/app10020711>
- Mahmoodzadeh A, Zare S (2016) Probabilistic prediction of the expected ground conditions and construction time and costs in road tunnels. *J Rock Mech Geotech Eng* 8:734–745. <https://doi.org/10.1016/j.jrmge.2016.07.001>
- Mahmoodzadeh A, Mohammadi M, Daraei A, Rashid TA, Sherwani AFH, Faraj RH, Darwesh AM (2019) Updating ground conditions and time-cost scatter-gram in tunnels during excavation. *Autom Constr* 105:102822. <https://doi.org/10.1016/j.autcon.2019.04.017>
- Flyvbjerg B (2006) From Nobel Prize to project management: getting risks right. *Proj Manag J* 37:5–15. <https://doi.org/10.1177/875697280603700302>
- Kermanshachi S, Safapour E (2020) Gap analysis in cost estimation, risk analysis, and contingency computation of transportation infrastructure projects: a guide to resource and policy-based strategy establishment. *Practi Period Struct Des Constr* 25:06019004. [https://doi.org/10.1061/\(ASCE\)SC.1943-5576.0000460](https://doi.org/10.1061/(ASCE)SC.1943-5576.0000460)
- Alsultan M, Jun J, Lambert JH (2020) Program evaluation of highway access with innovative risk-cost-benefit analysis. *Reliab Eng Syst Saf* 193:106649. <https://doi.org/10.1016/j.res.2019.106649>
- Cerezo-Narváez A, Pastor-Fernández A, Otero-Mateo M, Ballesteros-Pérez P (2020) Integration of cost and work breakdown structures in the management of construction projects. *Appl Sci* 10:1386. <https://doi.org/10.3390/app10041386>
- Ahn SJ, Han SU, Al-Hussein M (2020) Improvement of transportation cost estimation for prefabricated construction using geo-fence-based large-scale GPS data feature extraction and support vector regression. *Adv Eng Inform* 43:101012. <https://doi.org/10.1016/j.aei.2019.101012>
- Min SY, Kim TK, Lee JS, Einstein HH (2008) Design and construction of a road tunnel in Korea including application of the decision aids for tunneling—a case study. *Tunn Undergr Space Technol* 23:91–102. <https://doi.org/10.1016/j.tust.2007.01.003>
- Moret Y, Einstein HH (2016) Construction cost and duration uncertainty model: application to high-speed rail line project. *J Constr Eng Manag* 142:05016010. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001161](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001161)

15. Sousa RL, Einstein HH (2012) Risk analysis during tunnel construction using Bayesian networks: Porto Metro case study. *Tunn Undergr Space Technol* 27:86–100. <https://doi.org/10.1016/j.tust.2011.07.003>
16. Chung TH, Mohamed Y, AbouRizk S (2006) Bayesian updating application into simulation in the North Edmonton Sanitary Trunk tunnel project. *J Constr Eng Manag* 132:882–894. [https://doi.org/10.1061/\(ASCE\)0733-9364\(2006\)132:8\(882\)](https://doi.org/10.1061/(ASCE)0733-9364(2006)132:8(882))
17. Benardos AG, Kaliampakos DC (2004) Modelling TBM performance with artificial neural networks. *Tunn Undergr Space Technol* 19:597–605. <https://doi.org/10.1016/j.tust.2004.02.128>
18. Isaksson T, Stille H (2005) Model for estimation of time and cost for tunnel projects based on risk evaluation. *Rock Mech Rock Eng* 38:373–398. <https://doi.org/10.1007/s00603-005-0048-5>
19. Moayedi H, Mosallanezhad M, Rashid ASA, Jusoh WAW, Muazu MA (2020) A systematic review and meta-analysis of artificial neural network application in geotechnical engineering: theory and applications. *Neural Comput Appl* 32:495–518. <https://doi.org/10.1007/s00521-019-04109-9>
20. Galende-Hernández M, Menéndez M, Fuente MJ, Palmero GIS (2018) Monitor-While-Drilling-based estimation of rock mass rating with computational intelligence: the case of tunnel excavation front. *Autom Constr* 93:325–338. <https://doi.org/10.1016/j.autcon.2018.05.019>
21. Wauters M, Vanhoucke M (2014) Support vector machine regression for project control forecasting. *Autom Constr* 47:92–106. <https://doi.org/10.1016/j.autcon.2014.07.014>
22. Cheng MY, Wu YW (2009) Evolutionary support vector machine inference system for construction management. *Autom Constr* 18:597–604. <https://doi.org/10.1016/j.autcon.2008.12.002>
23. Tixier AJP, Hallowell MR, Rajagopalan B, Bowman D (2016) Application of machine learning to construction injury prediction. *Autom Constr* 69:102–114. <https://doi.org/10.1016/j.autcon.2016.05.016>
24. Cheng MY, Wu YW, Chen KL (2012) Risk preference based support vector machine inference model for slope collapse prediction. *Autom Constr* 22:175–181. <https://doi.org/10.1016/j.autcon.2011.06.015>
25. Gao X, Shi M, Song X, Zhang C, Zhang H (2019) Recurrent neural networks for real-time prediction of TBM operating parameters. *Autom Constr* 98:225–235. <https://doi.org/10.1016/j.autcon.2018.11.013>
26. Torabi-Kaveh M, Sarshari B (2019) Predicting convergence rate of Namaklan twin tunnels using machine learning methods. *Arab J Sci Eng*. <https://doi.org/10.1007/s13369-019-04239-1>
27. Rohmer J, Foerster E (2011) Global sensitivity analysis of large-scale numerical land-slide models based on Gaussian-process metamodeling. *Comput Geosci* 37:91–927
28. Liu R, Ye Y, Hu N, Chen H, Wang X (2019) Classified prediction model of rock burst using rough sets-normal cloud. *Neural Comput Appl* 31:8185–8193. <https://doi.org/10.1007/s00521-018-3859-5>
29. Ning F, Shi Y, Cai M, Xu W, Zhang X (2020) Manufacturing cost estimation based on a deep-learning method. *J Manuf Syst* 54:186–195. <https://doi.org/10.1016/j.jmsy.2019.12.005>
30. Barzegar R, Sattarpour M, Deo R, Fijani E, Adamowski J (2019) An ensemble tree-based machine learning model for predicting the uniaxial compressive strength of travertine rocks. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-019-04418-z>
31. Noorian-Bidgoli M, Jahed Armaghani D, Khamesi H (2016) Feasibility of PSO-ANN model for predicting surface settlement caused by tunneling. *Eng Comput* 4:705–715. <https://doi.org/10.1007/s00366-016-0447-0>
32. Goh ATC, Zhang W, Zhang Z, Xiao Y, Xiang Y (2018) Determination of earth pressure balance tunnel-related maximum surface settlement: a multivariate adaptive regression splines tool. *Bull Eng Geol Environ* 77:489–500. <https://doi.org/10.1007/s10064-016-0937-8>
33. Tijić K, Car-Puši D, Šperac M (2019) Costs estimation in road construction using artificial neural network. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-019-04443-y>
34. Hashemi ST, Ebadati OM, Kaur HA (2019) Hybrid conceptual costs estimating model using ANN and GA for power plant projects. *Neural Comput Appl* 31:2143–2154. <https://doi.org/10.1007/s00521-017-3175-5>
35. Gao W, Karbasi M, Hasanipanah M, Zhang X, Guo J (2018) Developing GPR model for forecasting the rock fragmentation in surface mines. *Eng Comput* 34:339–345. <https://doi.org/10.1007/s00366-017-0544-8>
36. Mohammadi M, Hossaini MF (2017) Modification of rock mass rating system: interbedding of strong and weak rock layers. *J Rock Mech Geotech Eng* 9:1165–1170. <https://doi.org/10.1016/j.jrmge.2017.06.002>
37. Bieniawski ZT (1973) Engineering classification of jointed rock masses. *S Afr Inst Civ Eng* 15:335–344
38. Bieniawski ZT (1989) Engineering rock mass classifications: a complete manual for engineers and geologists in mining, civil, and petroleum engineering. Wiley-Interscience, New York, pp 40–47. ISBN 0-471-60172-1
39. Williams CKI (1998) Prediction with Gaussian processes: from linear regression to linear prediction and beyond. In: Jordan MI (ed) *Learning in graphical models*. NATO ASI series (Series D: behavioural and social sciences), vol 89. Springer, Dordrecht, pp 599–621. [https://doi.org/10.1007/978-94-011-5014-9\\_23](https://doi.org/10.1007/978-94-011-5014-9_23)
40. Rasmussen CE (2004) Gaussian processes in machine learning. In: Bousquet O, von Luxburg U, Rätsch G (eds) *Advanced lectures on machine learning*. ML 2003. Lecture notes in computer science, vol 3176. Springer, Berlin, pp 63–71. [https://doi.org/10.1007/978-3-540-28650-9\\_4](https://doi.org/10.1007/978-3-540-28650-9_4)
41. Rasmussen CE, Williams CKI (2006) *Gaussian processes for machine learning*. MIT Press, Cambridge. ISBN 026218253X
42. Maity R, Bhagwat PP, Bhatnagar A (2010) Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrol Process* 24:917–923. <https://doi.org/10.1002/hyp.7535>
43. Yu PS, Chen ST, Chang IF (2006) Support vector regression for real-time flood stage forecasting. *J Hydrol* 328:704–716. <https://doi.org/10.1016/j.jhydrol.2006.01.021>
44. Behnia D, Ahangari K, Noorzad A, Moeinossadat SR (2013) Predicting crest settlement in concrete face rockfill dams using adaptive neuro-fuzzy inference system and gene expression programming intelligent methods. *J Zhejiang Univ Sci A* 14:589–602. <https://doi.org/10.1631/jzus.A1200301>
45. Wang C, Wang X, Xia Z, Zhang C (2019) Ternary radial harmonic Fourier moments based robust stereo image zero-watermarking algorithm. *Inf Sci* 470:109–120. <https://doi.org/10.1016/j.ins.2018.08.028>
46. Garg A, Tai K, Vijayaraghavan V, Singru PM (2014) Mathematical modelling of burr height of the drilling process using a statistical-based multi-gene genetic programming approach. *Int J Adv Manuf Technol* 73:113–126. <https://doi.org/10.1007/s00170-014-5817-4>
47. Wang C, Wang X, Li Y, Xia Z, Zhang C (2018) Quaternion polar harmonic Fourier moments for color images. *Inf Sci* 450:141–156. <https://doi.org/10.1016/j.ins.2018.03.040>
48. Shahin MA (2014) Load-settlement modeling of axially loaded

- steel driven piles using CPT-based recurrent neural networks. *Soils Found* 54:515–522. <https://doi.org/10.1016/j.sandf.2014.04.015>
49. Wang C, Wang X, Xia X, Ma B, Shi YQ (2019) Image description with polar harmonic Fourier moments. *IEEE Trans Circuits Syst Video Technol.* <https://doi.org/10.1109/TCSVT.2019.2960507>
50. Ahangari K, Moeinossadat SM, Behnia D (2015) Estimation of tunnelling-induced settlement by modern intelligent methods. *Soils Found* 55:737–748. <https://doi.org/10.1016/j.sandf.2015.06.006>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.