



# Predicting Convergence Rate of Namaklan Twin Tunnels Using Machine Learning Methods

Mehdi Torabi-Kaveh<sup>1</sup> · Bahram Sarshari<sup>2</sup>

Received: 9 July 2019 / Accepted: 6 November 2019 / Published online: 20 November 2019  
© King Fahd University of Petroleum & Minerals 2019

## Abstract

Convergence prediction of tunnels has always been one of the important issues of geotechnical projects. Developing prediction models is a good approach to predict convergence, when there is no knowledge of the possibility of convergence occurrence in the future. In this study, convergence rates of two tunnels from the Namaklan twin tunnel were predicted by different types of machine learning methods. Artificial neural networks (ANNs), multivariate linear regression (MLR), multivariate nonlinear regression (MNR), support vector regression (SVR), Gaussian process regression (GPR), regression trees and ensembles of trees (ET) were applied to predict the convergence rate (CR). The six parameters of cohesion ( $c$ ), internal friction angle ( $\Phi$ ), uniaxial compressive strength of rock mass ( $\sigma_c$ ), rock mass rating, overburden height ( $H$ ) and the number of installed rock bolts (NB) were selected as predictor parameters. The dataset was collected via field investigations and laboratory experiments. The results showed that the MLP-ANN model can successfully predict the CR with the determination coefficient ( $R^2$ ) of 0.93. The RBF-ANN model is also successful in predicting the CR with  $R^2$  of 0.81. The SVR, MNR and MLR models were also constructed to obtain an empirical formula for predicting the CR. Comparing among the three models showed that the SVR model is more successful ( $R^2 = 0.66$ ) than the MNR and MLR models with  $R^2$  of 0.65 and 0.61, respectively. However, the SVR model is placed in the next rank of the ANN models. Among the rest models, except the ET model ( $R^2 = 0.66$ ), the RT and GPR models have no good capability for the prediction of the CR. In total, assessing the statistical indices indicated that the ANNs are superior to the other models in predicting the CR. However, the SVR model could be considered to be a reliable predictive model for convergence rate estimation.

**Keywords** Tunnel convergence rate · Namaklan twin tunnels · Machine learning methods · Convergence rate prediction

## 1 Introduction

Evaluation of tunnel convergence is one of the most important issues during the excavation and even after utilization. This subject is more important concerning the new Austrian tunneling method (NATM). The occurrence of convergence at a high rate is associated with great problems such as reducing the advance rate and safety which increase operating costs. Two major factors are considered as a cause of convergence after the excavation. First is strain induced by stress distortion due to tunnel face advance, and second is

the behavior of the material after excavation (e.g., the time-dependent creep, swelling and squeezing) [1]. Predicting convergence magnitude is essential to select excavation and support methods. Hence, tunneling studies have focused on the prediction of convergence using geomechanical parameters and excavation parameters. Researchers have used different approaches to predict tunnel convergence magnitude. In this regard, many researchers applied numerical methods to predict tunnel convergence [2–6]. However, in most cases, such models are not usable to other projects. On the other hand, there is a limitation about coverage of all effective parameters on convergence in the form of an empirical model [7]. Therefore, several researchers have recently used artificial neural networks (ANNs) and new statistical methods in predicting convergence in relation to tunnels and powerhouses. Mahdevari and Torabi [8] used artificial neural networks (ANNs) to predict tunnel convergence. They used two different ANN models including multilayer perceptron

✉ Mehdi Torabi-Kaveh  
m.torabikaveh@yazd.ac.ir; mehditorabikaveh@yahoo.com

<sup>1</sup> Department of Geology, Faculty of Science, Yazd University, Yazd, Iran

<sup>2</sup> Engineering Geology Expert in Pars Consulting Engineers Company, Tehran, Iran



(MLP) and radial basis function (RBF). They reported that the proposed MLP model has more efficiency in predicting convergence compared to the other methods. Also, the study revealed that cohesion, friction angle, elastic modulus and uniaxial compressive strength are the most effective factors in estimating convergence, whereas tensile strength has the least effect. In research conducted by Adoko et al. [9], multivariate adaptive regression spline (MARS) and ANN were applied to predict the diameter convergence of a high-speed railway tunnel in weak rock. They suggested that MARS can be considered as a reliable alternative to ANN when a geomechanical problem such as the tunnel convergence requires nonlinear modeling. In another study, Zarei et al. [10] used three different methods including statistical, ANN and numerical modeling to develop a convergence criterion for the Chehel Chay water conveyance tunnel. They also concluded that ANN is more success in estimating convergence. Rajabi et al. [11] used some geomechanical and design parameters to predict the maximum horizontal displacement of sidewalls of a powerhouse. They predicted the maximum horizontal displacement by an ANN model, and after doing sensitivity analysis, they reported that overburden depth is the most effective parameter versus tensile strength which is the least effective parameter on displacement. Also, in a recent study, Hajihassani et al. [12] used ANN to predict the tunneling-induced ground movements as a flexible nonlinear approximation function. They reported that the proposed ANN model could be applied to predict three-dimensional ground movements induced by tunneling with high accuracy. Chen et al. [13] applied three different ANNs including back-propagation (BP), RBF and the general regression neural network (GRNN) in developing an appropriate model to predict the maximum surface settlement caused by EPB shield tunneling and compared their reliability. The GRNN model was found to be more accurate than the BP and RBF models and was proposed as a capable model to predict the behavior of ground settlement.

Although different methods have been used to identify the relationships between geotechnical properties and tunnel convergence as mentioned above, there still exists uncertainty in convergence prediction for tunnels excavated in different lithologies. Also, in spite of extensive use of new machine learning methods in prediction aims concerning geotechnical engineering problems [14–17], approximately in most of the studies on the subject of tunnel convergence prediction, prediction models have been limited to common linear and nonlinear regressions and ANNs. Accordingly, this research tries to extend the knowledge about convergence prediction of the tunnel by investigating the applicability of different methods in modeling convergence rate. In the present study, new and common approaches of machine learning systems including support vector regression (SVR), Gaussian process regression (GPR), regression trees (TR)

and ensembles of trees will be applied to predict convergence rate of Namaklan twin tunnel (west of Iran). Also, some common prediction models such as multivariate linear regression (MLR), multivariate nonlinear regression (MNR) and ANNs will be used to predict the convergence with the aim of comparing among the prediction approaches. This study uses some commonly employed and perhaps the most representative geotechnical parameters including uniaxial compressive strength (UCS), rock mass rating (RMR) and rock quality designation (RQD), rock quality system ( $Q$ ), cohesion ( $C$ ), internal friction angle ( $\Phi$ ), uniaxial compressive strength of rock mass ( $\sigma_c$ ) and uniaxial tensile strength of rock mass ( $\sigma_t$ ), along with some construction parameters such as overburden height and number of rock bolts as inputs for the models. Then, correlations will be done between predicted and observed convergence rates along the tunnel route.

## 2 Project Descriptions and Geology

Namaklan tunnel consists of twin D-shape bored tunnels of 6.35 m outer diameter. The tunnels were bored in the fourth section of Arak–Khorramabad Highway, between the cities Borujerd and Khorramabad as shown in Fig. 1. The lengths of the tunnels are 656 m and 652 m for west and east tunnels, respectively, with the horizontal distance of the twin tunnels being 45 m (more than 2.5 times the tunnel diameter) (Fig. 1). For both the tunnels, the diameter of excavation is 15.6 m and the diameter of the final lining is 13.6 m.

The route of the twin tunnel is composed of an alternation of marly limestone, limestone, shale and marl (Fig. 2) [18]. Due to the performance of tectonic forces, systematic discontinuities were mostly developed in hard layers. Based on the ratio of resilient and non-uniform layers in different zones, the route of the tunnel is characterized by gapping and different zones with blocks in different dimensions. The twin tunnels are parallel to each other (approximately 45 meters apart), and the characteristics of the defined zones are the same. In both the tunnels, the slope of the layering is from the outlet side to the inlet of the tunnel. Rock mass zoning along the tunnels has been done based on lithology type, characteristics of discontinuities (direction, density, spacing, etc.), weathering depth and overburden height of the tunnel. Different zones were characterized along the tunnel route that their characteristics were recorded in measuring stations and used as prediction parameters in the analysis part.

The shale and marl units have very low-to-low permeability, and they are very weak to weak from the hydrodynamic potential point of view. The limestone and the marly limestone units show a wide range of surface solution features, which confirms that they have low-to-moderate permeability. Also, the emergence of a spring at the outlet

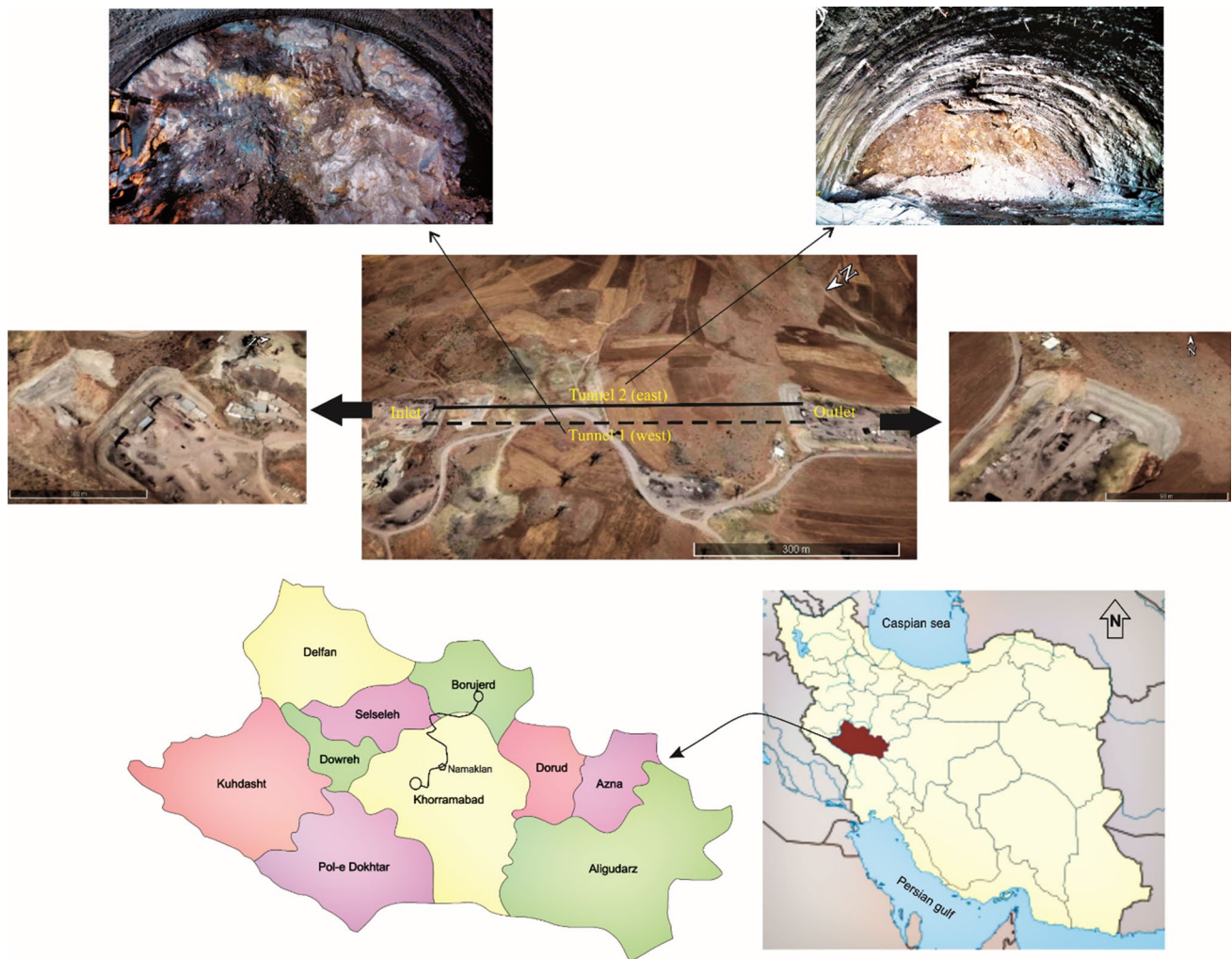


Fig. 1 Location of Namaklan twin tunnel on Iran map

of the tunnel and monitoring drilled boreholes indicate the high level of groundwater in the studied area [18].

Morphology of a region is a function of its lithology units and tectonics condition. The rock unit outcrops in the region include Chaghavandi overthrust units and quaternary alluviums. Figure 3 illustrates the morphology found in the study area. The limestone and marly limestone of Chaghavandi units are relatively hard and form sharp landscape features in the region. Quaternary deposits of old and new alluvium are very loose, with numerous small and large streams that form flat sections. The inlet trenches of both tunnels were excavated within marl limestone, which are located underneath thicker alluvial deposits, but the tunnel outlet trenches were created in a weathered zone of the marl limestone unit and a part of marl and shale units. In this regard, a collapse was occurred at the outlet of the tunnel due to the existence of weak rocks in this part (Fig. 4).

### 3 Data Collection and Input Parameters

The data were collected via field investigations along the tunnel route. During field measurements, rock mass properties of forty-five zones were recorded to evaluate rock mass quality. During the field study, the datasets were collected per each tunnel segment where the rock mass of each zone showed similar characteristics. Due to the significant effects of structural parameters and overburden pressure on the development of convergence, geological parameters should be considered in the stage of convergence rate assessment. Therefore, geological parameters such as rock quality designation (RQD), rock mass rating (RMR), rock quality system ( $Q$ ) and overburden height ( $H$ ) were measured for all the tunnel zones and were used in the construction of prediction models. After excavating a tunnel opening, the strength of rock mass, which is a



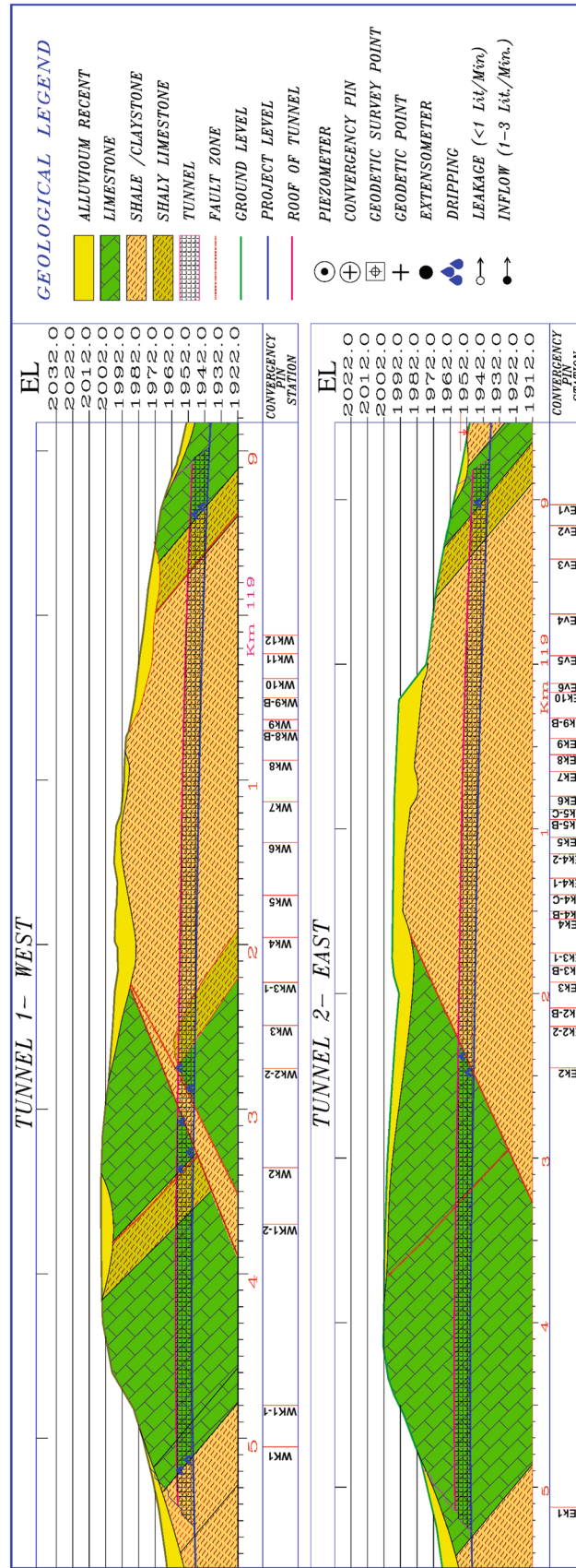
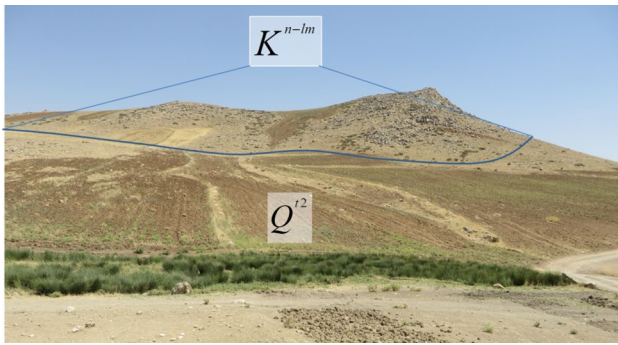


Fig. 2 Geological longitudinal section of the tunnel route [18]



**Fig. 3** Sharp morphology of Chaghavandi units ( $K^{n-lm}$ ) in the region versus quaternary deposits ( $Q^{r2}$ ) with flat morphology

function of rock mass properties, is an important factor in probable convergence occurrence around a tunnel. Hence, considering rock mass properties is essential in predicting the convergence rate of a tunnel. In this regard, major rock mass properties including cohesion ( $C$ ), internal friction angle ( $\Phi$ ), uniaxial compressive strength of rock mass ( $\sigma_c$ ) and uniaxial tensile strength of rock mass ( $\sigma_t$ ) as outputs of RocLab program were calculated according to the Hoek–Brown and Mohr–Coulomb criteria and were applied in developing prediction models. Finally, uniaxial compressive strength (UCS) of intact rock, the rock mass parameters ( $C$ ,  $\Phi$ ,  $\sigma_c$  and  $\sigma_t$ ), the geological data (RQD, RMR,  $Q$  and  $H$ ) along with number of installed rock bolts (NB) were used to predict the convergence rate (CR) of each zone. After each cycle of excavation, convergence

control points were established and convergence meters were applied to measure the CR values. The measurements were frequently repeated according to a daily program until stabilization of the points. Table 1 presents some details of the recorded parameters and the measured CR values. Also, Fig. 5 illustrates occurred convergences in some parts of the tunnel route.

#### 4 Predicting of the Tunnel Convergence Rate Using Statistical Models

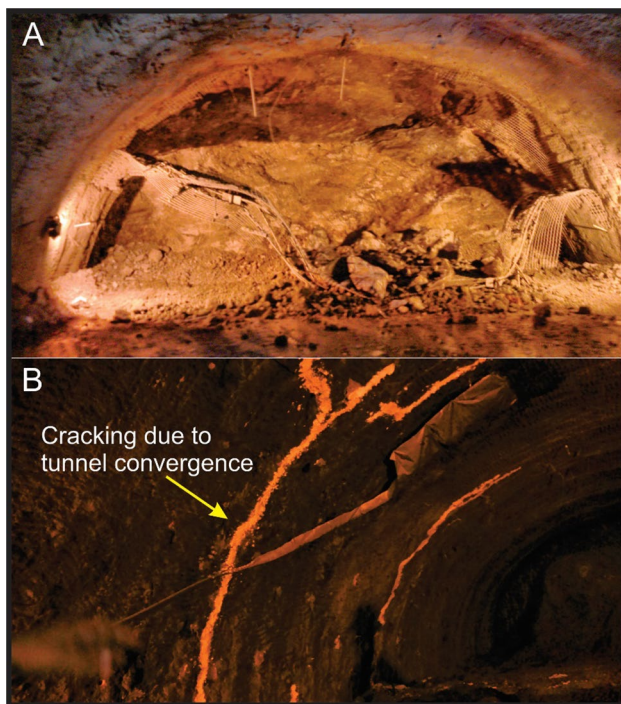
Regression-based equations extracted from statistical models are considered as easy-use prediction tools among available predicting models [19]. Considering that convergence rate of a tunnel is related to different geomechanical parameters, so the prediction of convergence rate based on only one parameter is not reliable. Hence, in this study, multivariate regression analyses (linear and nonlinear forms) were applied to create prediction models of the convergence rate of the tunnel using the geomechanical parameters as inputs. Multiple linear regression (MLR) analysis was performed by SPSS software [20]. Backward method was selected to perform multivariate linear regressions. In this method, all independent variables are first entered into the equation and the effect of all variables is evaluated on dependent variable. Less efficient variables are excluded one after another, and eventually, these steps continue until the test error reaches a significant level of 10%. Backward elimination starts with all of the predictors in the model. The variable that is least



**Fig. 4** Collapse at the outlet of the west tunnel (tunnel 1) due to the existence of weak rocks

**Table 1** Details of input and output parameters

Type and number of data	Symbol	Unit	Minimum	Maximum	Mean	SD
Inputs/45 for each parameter	UCS	MPa	3.00	52.00	8.11	12.22185
	$H$	m	9.00	46.00	32.2	9.15065
	RQD	%	10.0	95.00	23.0667	18.75851
	RMR	–	13.0	49.00	23.2000	6.75412
	$Q$	–	0.001	26.00	0.6520	3.86945
	NB	N/m	0.00	12.00	6.4889	3.22365
	$\sigma_t$	MPa	0.00	0.02	0.0032	0.00507
	$\sigma_c$	MPa	0.01	0.84	0.1036	0.20267
	$c$	MPa	0.05	1.80	0.2526	0.45445
	$\Phi$	Degree (°)	14.4	27.60	18.6624	3.35269
Output/45	CR	mm/day	0.03	2.10	0.7320	0.62321

**Fig. 5** **a** Breaking the support system due to the occurrence of convergence in tunnel 1; **b** creating cracks in the wall of tunnel 2 in result of convergence

significant that is the one with the largest  $p$  value is removed, and the model is refitted. Each subsequent step removes the least significant variable in the model until all remaining

variables have individual  $P$  values smaller than some value, such as 0.05 or 0.10. In fact, in this method, important variables are known and remain in the equation. Every independent variable with a low correlation coefficient with the dependent variable was omitted. The summary of the obtained models is presented in Table 2. As can be seen from the table, with excluding the less important variables,  $R^2$  shows negligible changes for the models and so model 4 which has the lowest and the most effective variables ( $H$ , RMR,  $\Phi$ ,  $c$ ,  $\sigma_c$  and NB) among the models is selected as a prediction model.

The following multiple linear regression equation was obtained for prediction of convergence rate (CR) as output considering six inputs of the model 4:

$$CR = 4.35 + 6.23H - 0.08RMR - 3.9NB + 6.948\sigma_c - 2.26c - 0.01\Phi \quad (1)$$

where CR is the predicted convergence rate. A scatter plot of the predicted CR versus the measured CR is shown in Fig. 6. Given the unreliable determination of coefficient ( $R^2 = 0.61$ ) obtained in linear analysis, it was tried to find a multivariate nonlinear regression (MNR) model that incorporates all the parameters used in model 4 in the linear regression. Several different forms of nonlinear regression equations were used, and it was found one with the smallest number of parametric constants, still providing the most accurate result (with highest  $R^2$ ). The selected model is of the following form:

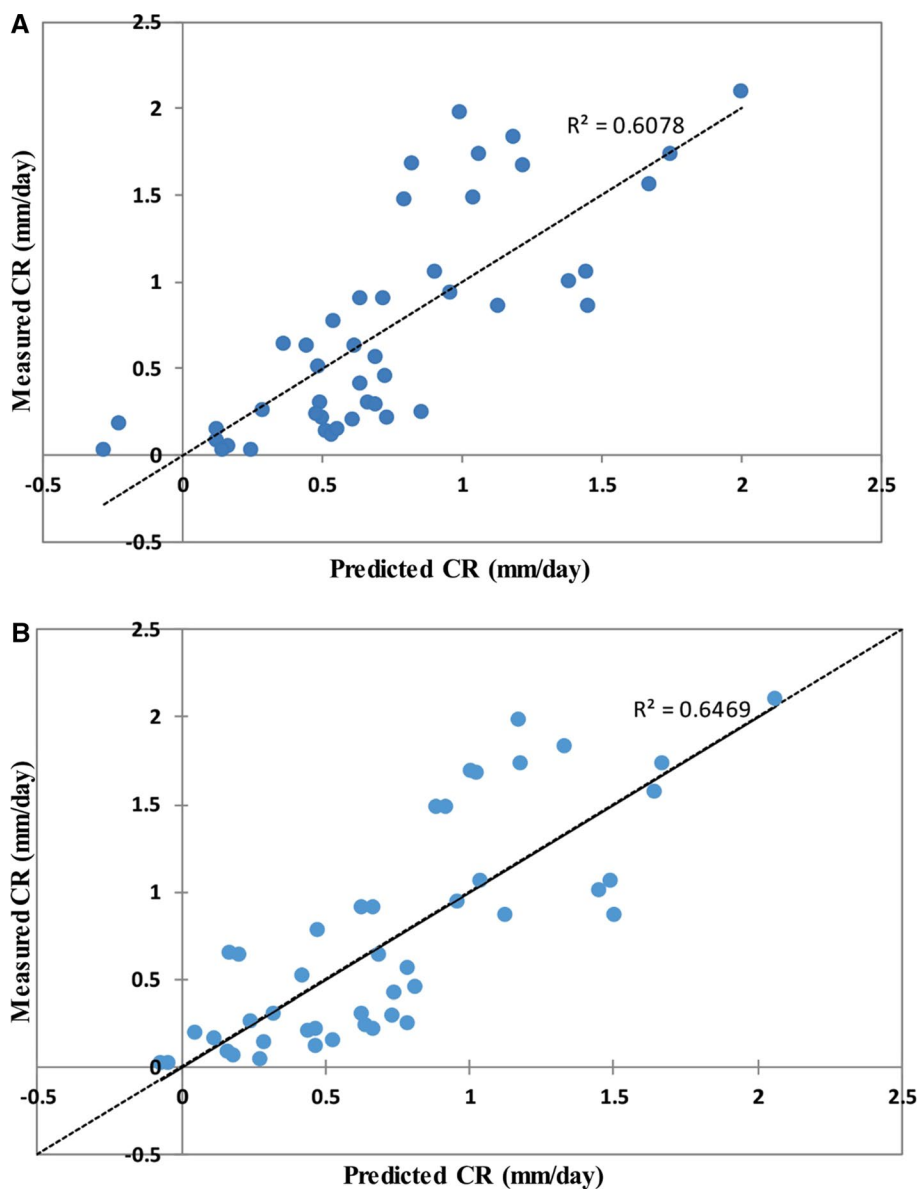
**Table 2** Model summary for the multivariate linear regressions

Model	Predictors	$R$	$R^2$	SE of the estimate	Excluded variables
1	$\Phi$ , $H$ , $Q$ , NB, $\sigma_t$ , RMR, RQD, $c$ , $\sigma_c$	0.781	0.610	0.43661	UCS
2	$\Phi$ , $H$ , NB, $\sigma_t$ , RMR, RQD, $c$ , $\sigma_c$	0.781	0.610	0.43052	UCS, $Q$
3	$\Phi$ , $H$ , NB, RMR, RQD, $c$ , $\sigma_c$	0.780	0.609	0.42491	UCS, $Q$ , $\sigma_t$
4	$\Phi$ , $H$ , NB, RMR, $c$ , $\sigma_c$	0.780	0.608	0.41972	UCS, $Q$ , $\sigma_t$ , RQD





**Fig. 6** Scatter plots of the predicted CR versus the measured CR for the linear (a) and nonlinear (b) multivariate regressions



$$CR = 11.641 - 0.048H + 0.142RMR - 0.013NB + 8.258\sigma_c - 0.96c - 0.987\Phi + 0.001H^2 - 0.005RMR^2 - 0.001NB^2 + 4.271\sigma_c^2 - 1.41c^2 + 0.02\Phi^2 \tag{2}$$

In spite of what was expected, the nonlinear model could not significantly enhance the accuracy of the model for predicting the CR in comparison with the linear regression model, as the nonlinear regression increased the value of  $R^2$  from 0.61 in MLR to 0.65 in MNR (Fig. 6).

In this study, some statistical indices including mean of the mean absolute error (MAE), the squares of the errors (MSE), root mean square error (RMSE) and determination coefficient ( $R^2$ ) were used to evaluate the performance of

prediction models by comparing the predicted CR values with the true values. The following equations given expressions of MAE, MSE and RMSE, respectively:

$$MAE = \frac{1}{n} \sum_{i=1}^n |(Y_i - Y'_i)| \tag{3}$$

where  $Y'_i$  is the predicted value,  $Y_i$  is the measured value and  $n$  is the number of all variables.

**Table 3** Comparative table of the developed prediction models

Model	Statistical indices for the models	Characteristics of the selected models	Total rank
Multivariate linear regression (MLR)	$R^2=0.61$ MSE=0.149 RMSE=0.386 MAE=0.311	Analysis method: backward	6
Multivariate nonlinear regression (MNR)	$R^2=0.648$ MSE=0.134 RMSE=0.366 MAE=0.299	Function: quadratic form	4
<i>ANNs</i>			
MLP	$R^2=0.925$ MSE=0.029 RMSE=0.169 MAE=0.122	Training algorithm: Levenberg–Marquardt Transformation functions: hidden layer (tansig) and output layer (purelin) Performance function: MSEREG Gradient: $1.04e-13$ Mu: 0.01 Validation check: 56 epochs	1
RBF	$R^2=0.81$ MSE=0.073 RMSE=0.269 MAE=0.218	Training algorithm: Levenberg–Marquardt Transformation functions: hidden layer (softmax) and output layer (purelin) Performance function: MSEREG	2
<i>Support vector regression (SVR)</i>			
eps-regression SVM <sup>a</sup>	$R^2=0.659$ MSE=0.148 RMSE=0.385 MAE=0.268	SVM-Type: eps-regression SVM-Kernel: radial Cost:1 Gamma: 0.1666667 Epsilon: 0.1 Number of support vectors: 39	3
Linear SVM	$R^2=0.32$ MSE=0.26 RMSE=0.51 MAE=0.42		
Quadratic SVM	$R^2=0.32$ MSE=0.26 RMSE=0.51 MAE=0.42		
Cubic SVM	$R^2=0.18$ MSE=0.45 RMSE=0.67 MAE=0.51		
Fine Gaussian SVM	$R^2=0.18$ MSE=0.45 RMSE=0.67 MAE=0.51		
Medium Gaussian SVM	$R^2=0.45$ MSE=0.21 RMSE=0.46 MAE=0.36		
Coarse Gaussian SVM	$R^2=0.29$ MSE=0.27 RMSE=0.52 MAE=0.41		
<i>Regression tree (RT)</i>			





**Table 3** (continued)

Model	Statistical indices for the models	Characteristics of the selected models	Total rank
Fine tree <sup>a</sup>	$R^2=0.53$ MSE=0.17923 RMSE=0.42335 MAE=0.33832	Preset: Fine tree Minimum leaf size: 4 Surrogate decision splits: Off	7
Medium tree	$R^2=0.52$ MSE=0.18 RMSE=0.42 MAE=0.32		
Coarse tree	$R^2=0.00$ MSE=0.38 RMSE=0.62 MAE=0.53		
<i>Gaussian process regression (GPR)</i>			
Exponential GPR <sup>a</sup>	$R^2=0.53$ MSE=0.17959 RMSE=0.42378 MAE=0.34787	Preset: Exponential GPR Basis function: Constant Kernel function: Exponential Use isotropic kernel: true	8
Rational quadratic GPR	$R^2=0.50$ MSE=0.19 RMSE=0.44 MAE=0.37	Kernel scale: Automatic Kernel sigma: Automatic Sigma: Automatic Standardize: true	
Squared exponential GPR	$R^2=0.50$ MSE=0.19 RMSE=0.43 MAE=0.37	Optimize numeric parameters: true	
Matern GPR	$R^2=0.51$ MSE=0.19 RMSE=0.43 MAE=0.37		
Ensemble trees (ET)	$R^2=0.63$ MSE=0.1401 RMSE=0.3743 MAE=0.32546	Preset: Bagged tree Minimum leaf size: 8 Number of learners: 30	5

<sup>a</sup>The best model with the highest performance

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2 \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2} \tag{5}$$

RMSE can define the performance of different models in predicting target values. In this regard, if RMSE to be < 10%, 10–20%, 20–30% and > 30%, the performance of a model is excellent, good, fair and poor, respectively. Comparing among the statistical indices (MSE, RMSE and MAE) calculated for both models which are presented in Table 3 shows that the MNR model has lower errors in the prediction of the CR values and so has more efficiency in predicting than the MLR model. Although based on the RMSE values, the performance of both models is poor.

## 5 Predicting of the Tunnel Convergence Rate Using ANNs

ANN is a new alternative to other conventional statistical techniques which are often limited by strict assumptions of normality, linearity, variable independence, one-pass approximation, the curse of dimensionality, etc. Some of the advantages of the neural network include requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables and the availability of multiple training algorithms [21–23]. But some following limitations decrease the ability of ANN in comparison with statistical models. Requirement of a large diversity of training for operation is a common limitation for neural networks. The knowledge acquired during the training of the model is stored implicitly, and hence, it is hard to come up with

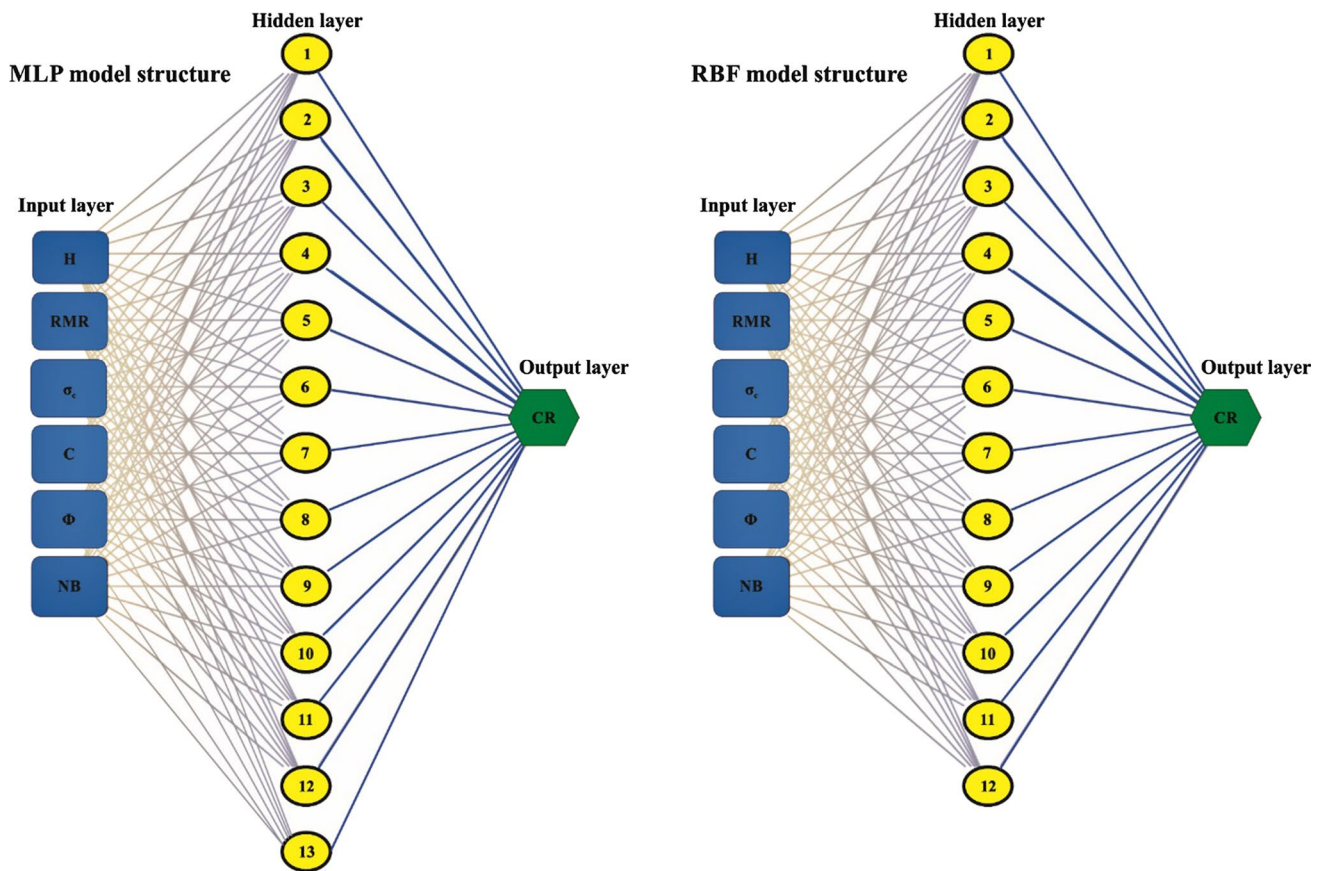


Fig. 7 Structures of the constructed ANNs

a reasonable interpretation of the overall structure of the network. Besides, slow convergence speed, less generalizing performance, ‘black box’ nature, greater computational burden, arriving at a local minimum and overfitting problems are intrinsic disadvantages of ANN.

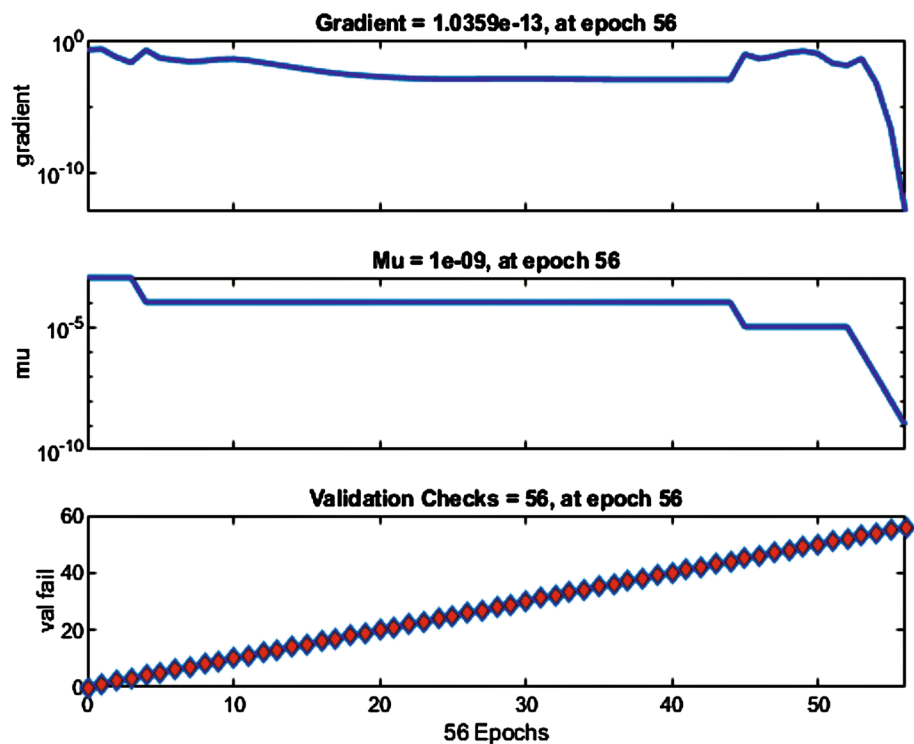
The multilayer perceptron (MLP) and the radial basis function (RBF) are two famous neural networks which have been widely used in many areas of geomechanics. RBF is considered as an alternative to MLP, because MLP is nonlinear in its parameters and it should perform several attempts of trial and error to choose the best from a set of locally optimum parameters. While RBF network has linear parameters and is capable of universal approximations and learning without local minima, it guarantees convergence to optimum parameters. Also, it is reported that the RBF-type networks learn faster than MLP networks and train with the least initial set of data [24].

To examine the capability of MLP and RBF in predicting convergence rate of the tunnel, two models were developed using  $H$ , RMR,  $\Phi$ ,  $c$ ,  $\sigma_c$  and NB parameters as inputs and CR as the target of the models. In the MLP model for all neurons in the hidden layer, hyperbolic tangent transfer function (tansig) was selected to transform real-valued

arguments ranging from  $-1$  to  $+1$ . In the case of the output layer (target), linear transfer function (purelin), which takes real-valued arguments and returns them unchanged, was applied. In the RBF model, the activation functions for the hidden layer and the output layer were the radial basis function and purelin, respectively.

An automatic computation was used to find the best number of units of the hidden layer within the minimum and maximum values of a range, where the best number of hidden units was the one that yields the smallest error in the testing data. Optimization and determination of the best MLP and RBF models were carried out by evaluating several different networks with different hidden neurons and spread parameters. The best created MLP and RBF models have one hidden layer with 13 hidden neurons and one hidden layer with 12 hidden neurons, respectively (Fig. 7). The Levenberg–Marquardt algorithm and clustering were, respectively, used for training the MLP and the RBF models via neural network toolbox of MATLAB [25] until reaching the best-fitted results (Table 3). Seventy percent of the data were used to train the neural network. The rest 30% of the input data were applied for validating and testing the neural network results.

**Fig. 8** Performance error plot of the optimum MLP model



In the case of the MLP model, at the epoch of 56 of learning the best network was obtained (Fig. 8). As can be seen from the figure, after reaching the best validation check, it means the least mean squared error (MSE), the training was stopped and the network weights were stored as the best network weights. Figure 9 shows that there is a meaningful correlation between the measured and predicted CR, which reveals that the selected network has high accuracy in predicting the tunnel convergence rate. The success of the network in learning and generalizing is shown in Fig. 9. As can be seen from the figure, the test data are estimated with a high regression of 0.957. Also, the linear regression between the input and output data is 0.962 for all network (Fig. 9). Good results were found for the RBF model. Figure 10 shows a scatter plot between the predicted and measured CR values using the RBF model with the 12 hidden neurons. As can be seen from the figure, the determination coefficient ( $R^2$ ) is 0.81 which illustrates a good correlation between the measured and predicted CR. Also, the model has reliable values of MSE, RMSE and MAE (Table 3).

## 6 Predicting of the Tunnel Convergence Rate Using SVR

Application of the support vector regression (SVR) models in solving nonlinear problems is growing among researchers who deal with geotechnical issues. This approach focuses on forecasting analysis [26, 27]. There are some advantages

which give much better performance to support vector machine (SVM) than ANN. These advantages include a global and unique solution, a simple geometric interpretation, a structural risk minimization and low prone to overfitting. Hence, some types of SVR models with different kernel types were developed to predict CR of the tunnel and their prediction accuracies were compared. The models were constructed by Regression Learner Application in MATLAB [25] and the e1071 package which is the first and most intuitive package in R software. RMR, H, NB,  $\sigma_c$ ,  $c$  and  $\Phi$  were selected as inputs of the models. Six models with kernel types of linear, quadratic, cubic, fine Gaussian, medium Gaussian and coarse Gaussian were created in MATLAB (Table 3). It was found that except medium Gaussian model which has a relatively fair determination coefficient ( $R^2$  of 0.45), other models have very weak performance in predicting the CR values compared to the multivariate regressions and the ANNs. Therefore, it was tried to develop another SVM model based on the kernel function type of radial basis function (RBF). A reason for using the kernel function is its ability in transforming our data from nonlinear form to linear form. It allows the SVM to find a fit, and then, data are mapped to the original space. The model shows relatively good ability in predicting CR with  $R^2=0.66$  (Fig. 11) but contrary to our expectation, the SVM model is not more efficient than the ANN models. Equation 6 represents the equation developed by the constructed SVM model:

Fig. 9 Regression plots corresponding to the MLP model

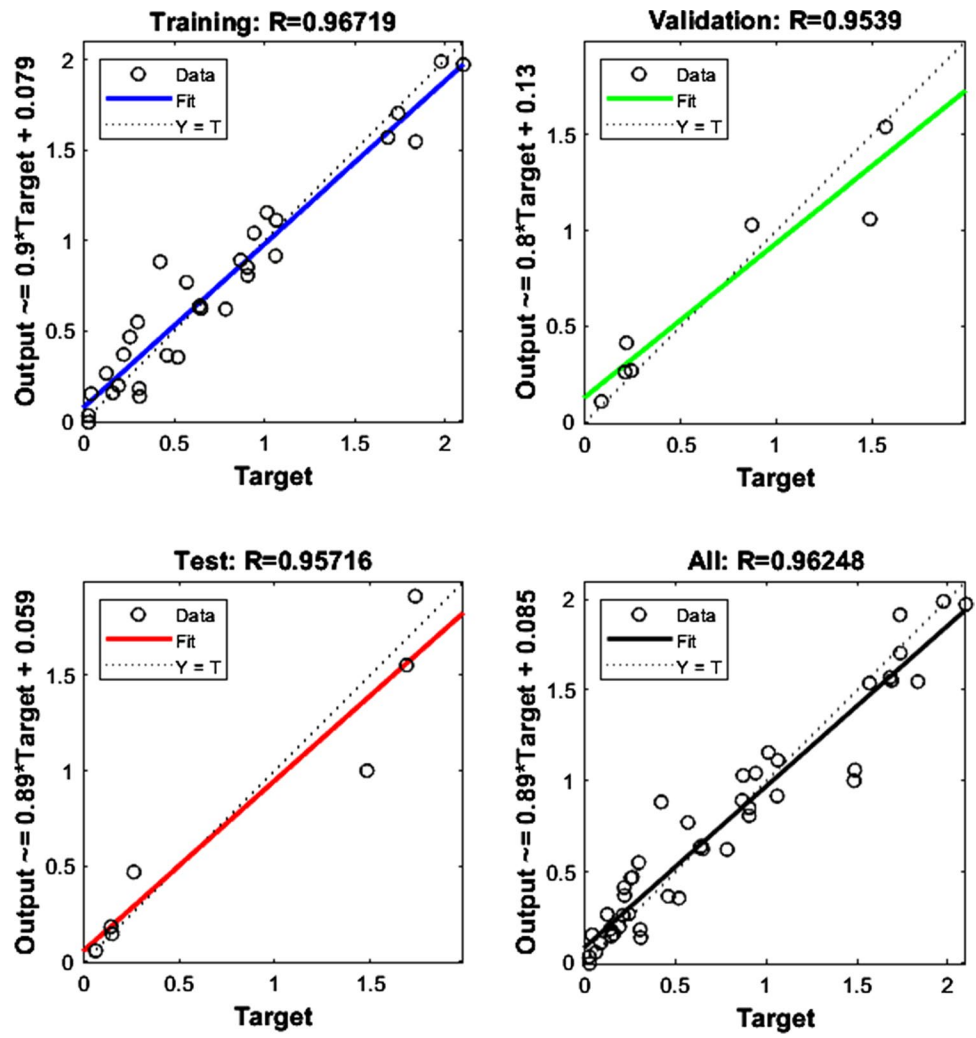
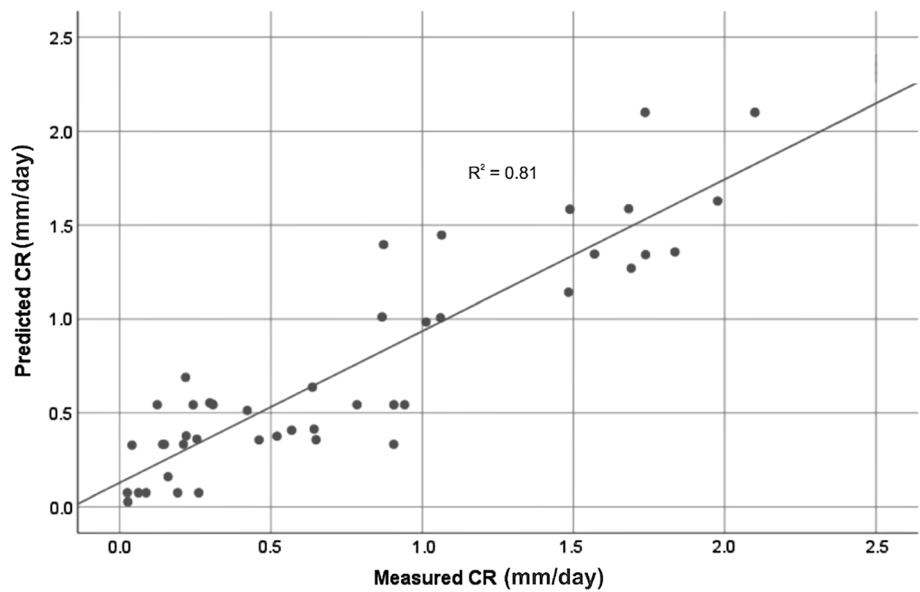
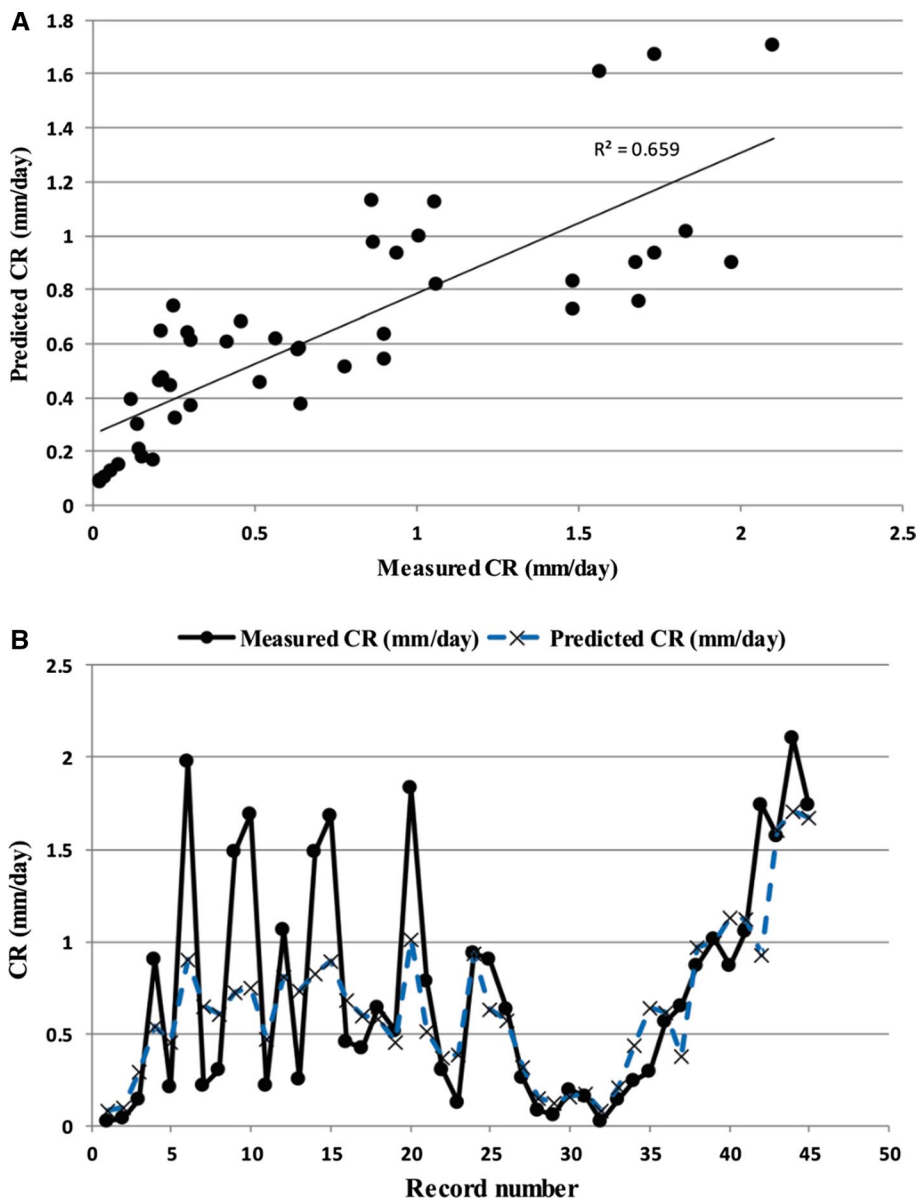


Fig. 10 Scatter plot of the predicted CR versus the measured CR for the RBF-ANN model





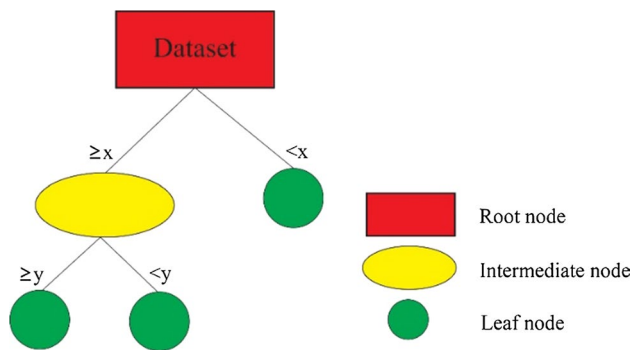
**Fig. 11 a** Scatter plot of the predicted CR versus the measured CR for the SVR model; **b** comparison of the predicted versus measured values



$$CR = 0.63H - 6.52RMR - 1.64NB - 3.5\sigma_c - 3.57c - 6.18\Phi + 0.446 \quad (6)$$

### 7 Predicting of the Tunnel Convergence Rate Using RT

The regression tree (RT) is considered a powerful and fast technique for fitting and prediction aims. RT is defined via a decision tree that assesses the effect of predictors (inputs) on output variable(s) (response). The RT method converts input variables into a mixture of continuous and categorical



**Fig. 12** Structure of a regression tree

variables in predicting a single output [28]. The task of learning in regression trees involves the prediction of real numbers instead of discrete batch values which are common in classification trees. The most advantage of RT is the easy interpretation of the summarized results which make it a reliable method in prediction aims [29].

RT is a graph similar to a tree that contains a sequence of questions that are defined by variables and a set of fitted response values (Fig. 12). In response to that whether a predictor satisfies a given condition, answer to each question is ‘yes’ or ‘no.’ The answer to a question decides that the process shifts to another question or stops at a fitted response value. When the answer is ‘yes,’ the process continues in the left branch. Root, branches, leaves and nodes are the main components of a tree. In a decision tree, nodes form circles and branches connect nodes. During the process, root is considered as the first node which is a variable and then several internal nodes are created as a result of dividing that node based on a series of features. Decision tree structure in such a way that the root is placed at the top of a chain and leaf (including root, branch and node) forms the end of it. A decision tree is called a classification tree when the output of a tree is a discrete set of values, whereas a real set of values for the output of a tree converts it to a regression tree.

To explain the mathematics behind classification in the regression tree, a brief mathematical formula is presented the following. Decision regression tree algorithms work through recursive partitioning of the training set to obtain subsets that are as pure as possible to a given target class. Each node of the tree is associated with a particular set of records  $T$  that is split by a specific test on a feature. For example, a split on a continuous attribute  $A$  can be induced by test  $A \leq x$ . The set of records  $T$  is then partitioned into two subsets that lead to the left branch of the tree and the right one:

$$T_l = \{t \in T : t(A) \leq x\}$$

and

$$T_r = \{t \in T : t(A) > x\}$$
(7)

Similarly, a categorical feature  $B$  can be used to induce splits according to its values. For example, if  $B = \{b_1, \dots, b_k\}$ , each branch  $i$  can be induced by the test  $B = b_i$ .

The divide step of the recursive algorithm to induce a decision tree takes into account all possible splits for each feature and tries to find the best one according to a chosen quality measure: the splitting criterion. If your dataset is induced on the following scheme:

$$A_1, \dots, A_m, C$$
(8)

where  $A_j$  are attributes and  $C$  is the target class, all candidate’s splits are generated and evaluated by the splitting criterion. Splits on continuous attributes and categorical ones are generated as described above. The selection of the

best split is usually carried out by impurity measures. The impurity of the parent node has to be decreased by the split. Let  $(E_1, E_2, \dots, E_k)$  be a split induced on the set of records  $E$ , a splitting criterion that makes use of the impurity measure  $I(\cdot)$  is:

$$\Delta = I(E) - \sum_{i=1}^k \frac{|E_i|}{|E|} I(E_i)$$
(9)

Standard impurity measures are the Shannon entropy or the Gini index. More specifically, CART uses the Gini index that is defined for the set  $E$  as follows. Let  $p_j$  be the fraction of records in  $E$  of class  $c_j$ :

$$p_j = \frac{|\{t \in E : t[C] = c_j\}|}{|E|}$$

then

(10)

$$\text{Gini}(E) = 1 - \sum_{j=1}^Q p_j^2$$

where  $Q$  is the number of classes. It leads to a 0 impurity when all records belong to the same class.

Since the data in this study are real values, a regression tree was applied to predict the convergence rate. The same datasets used for developing the above models were applied to create RT models via Regression Learner Application in MATLAB [25]. All the regression tree model types (fine, medium and coarse) in the application were constructed to predict the CR values, and the best model with dataset was selected. A fine tree was determined as the best model that its characteristics are presented in Table 3. Figure 13a illustrates the scatter plot of the CR values predicted by the regression tree model versus the measured values. As can be seen from the figure, the RT model with  $R^2$  of 0.44 is not successful in predicting CR values, as almost a significant part of the data points are too scattered relative to the respective regression line of CR values. Also, it was revealed that there is a huge difference between predicted and observed CR values which confirms the inefficiency of the model in predicting CR (Fig. 13b).

## 8 Predicting of the Tunnel Convergence Rate Using ET

Considering the low capability of the RT model in the prediction of CR, the ensemble method was applied to enhance the applicability of the regression model in predicting CR. The ensemble method combines multiple weak regression trees with the aim of forming an accurate and strong regression tree model. In this method, multiple diverse regression models are created based on different samples of the original



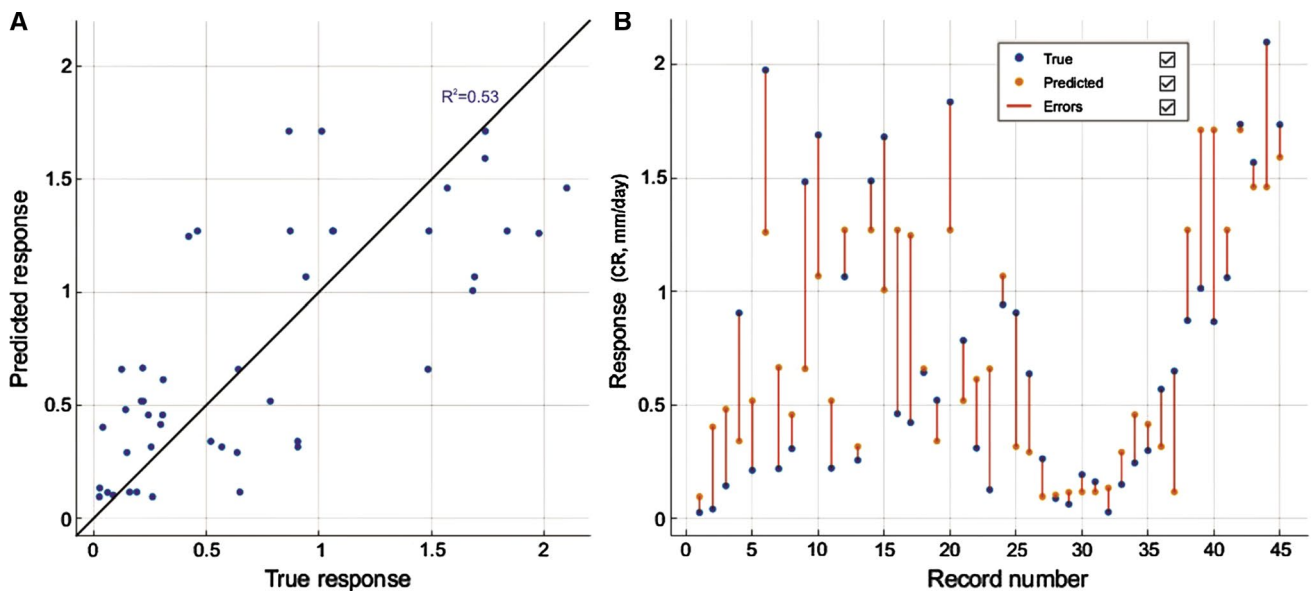


Fig. 13 a Scatter plot of the predicted CR versus the measured CR for the RT model; b comparison of the predicted versus measured values

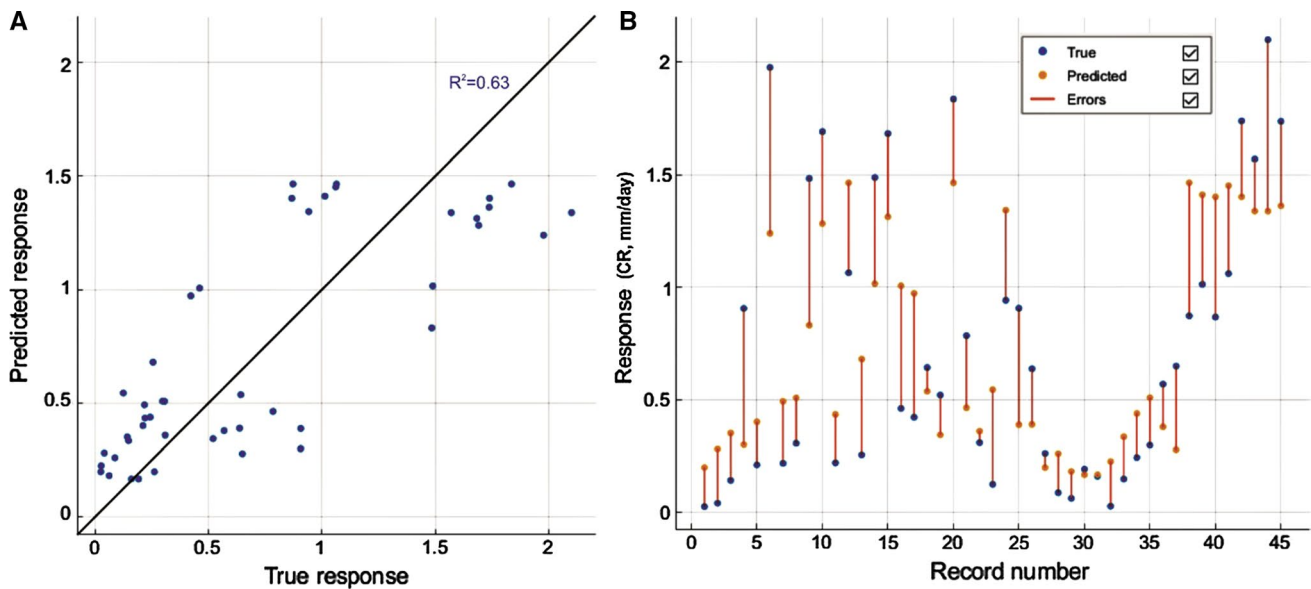


Fig. 14 a Scatter plot of the predicted CR versus the measured CR for the ET model; b comparison of the predicted versus measured values

dataset, and finally, their outputs are combined. The result of the combination is a reduction of the variance in the model. Three types of ensemble methods (bagging, boosting and random trees) are commonly used as prediction models [30]. Among them, bagging is a simple and effective ensemble algorithm which utilizes a series of training sets according to random sampling and applies the regression tree algorithm to each dataset. Eventually, it calculates predictions of the new data by taking average among the models. This advantage makes it a better selection concerning large datasets. In this study, the bagging ensemble method was applied to

predict CR values using Regression Learner Application in MATLAB [25]. All the ensemble model types (boosted trees and bagged trees) in the application were constructed to predict CR values, and the best model was selected based on the obtained values of  $R^2$ , RMSE and MSE for the models. Table 3 presents characteristics of the best-developed ensemble model which is a bagged model. The reliability of the developed model was evaluated by comparing the obtained  $R^2$ , RMSE and MSE values for the ensemble model and the RT model. The comparison showed that applying ensemble method could significantly enhance the accuracy

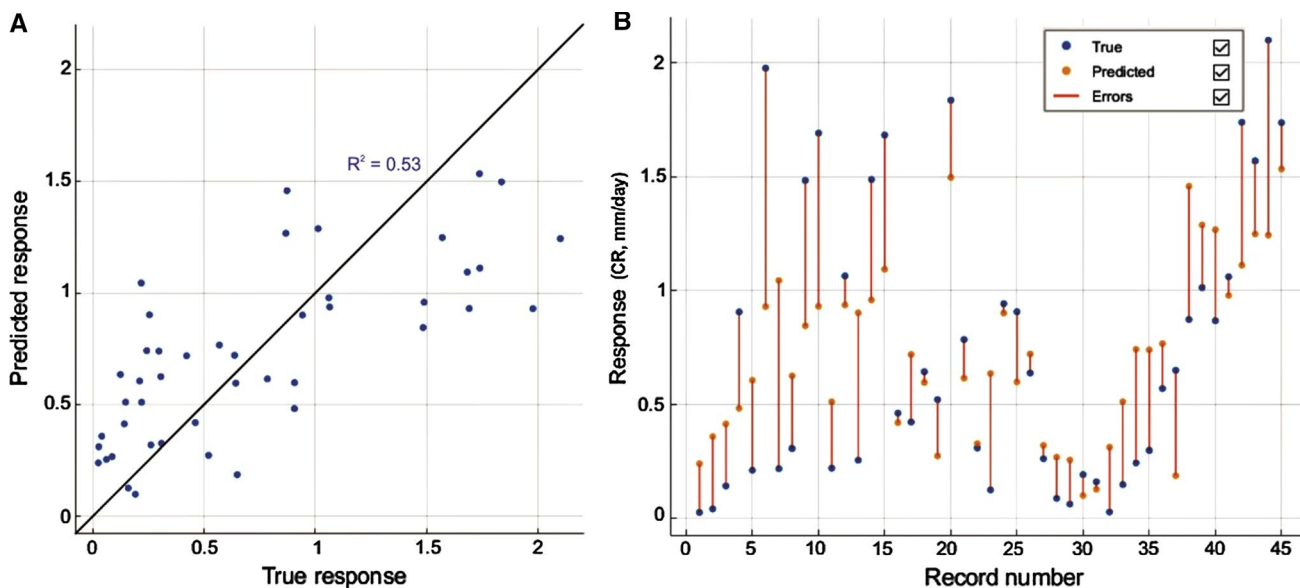


Fig. 15 a Scatter plot of the predicted CR versus the measured CR for the GPR model; b comparison of the predicted versus measured values

of the prediction model (Fig. 14a). So that the ensemble model has higher determination coefficient ( $R^2=0.63$ ) and lower MSE, RMSE and MAE values (0.14, 0.37 and 0.34, respectively) than that of the RT model (0.18, 0.43 and 0.34, respectively) (Table 3). However, the developed model has still a relatively low coefficient of determination that it reduces its applicability for new datasets. Also, Fig. 14b shows that differences between the predicted and observed CR values were more reduced compared to that of the RT model.

### 9 Predicting of the Tunnel Convergence Rate Using GPR

GPR is a nonparametric probabilistic modeling approach to generalize nonlinear and complex functions in datasets [31]. Hence, the approach has been used by many researchers in various studies for solving different problems in engineering [32]. Considering that GPR uses kernel functions, so it can handle nonlinear data efficiently. The parameters of GPR are a set of random variables, as any finite number of them has a joint Gaussian distribution. GPR has some advantage which makes it differ from other common prediction models. For example, GPR has a practically simpler understanding and implementing compared to back-propagation neural networks [33]. The use of kernel functions in the GPR model has made it close to SVM [34]. Furthermore, compared to the other regression kernel-based models, GPR provides reliability of responses to the given input data which is known to the probabilistic model [35].

The following expression is considered for predicting of  $y$  (output) by GPR:

$$y_i = f(X_i) + \epsilon_i \tag{11}$$

where  $f(x_i)$  and  $\epsilon_i$  are latent functions and Gaussian noise, respectively. GPR treats  $f(x_i)$  as a random variable. The following equation gives the joint distribution of  $y$ :

$$P(y) = N(0, K(x, x) + \sigma^2 I) \tag{12}$$

where  $K(x, x)$  is the kernel function and  $I$  is an identity matrix. The predictive distribution of  $y_{D+1}$  corresponding to a new given input  $x_{D+1}$  is given by the following expression:

$$\begin{pmatrix} y \\ y_{D+1} \end{pmatrix} \sim N(0, K_{D+1}) \tag{13}$$

where  $K_{D+1}$  is the covariance matrix, and its expression is given by:

$$K_{D+1} = \begin{bmatrix} K(x, x) + \sigma^2 I & K(x, x_{D+1}) \\ K(x_{D+1}, x)^T & K(x_{D+1}, x_{D+1}) \end{bmatrix} \tag{14}$$

The distribution of  $y_{D+1}$  is Gaussian with mean and variance [36]:

$$\mu = K(x_{D+1})^T [K(x, x) + \sigma^2 I]^{-1} y \tag{15}$$

$$\Sigma = K(x_{D+1}, x_{D+1}) - K(x_{D+1}, x) [K(x, x) + \sigma^2 I]^{-1} K(x, x_{D+1}) \tag{16}$$

The details of GPR are given by Williams and Rasmussen [37].



**Table 4** Descriptive analysis of the new data

Type and number of data	Symbol	Unit	Minimum	Maximum	Mean	SD
Inputs/10 for each parameter	$H$	m	38.9	58.81	46.69	14.59
	RMR	–	19.75	23.00	21.00	18
	NB	N/m	0.375	0.50	0.40	0
	$\sigma_c$	MPa	0.124	0.275	0.18	0.11
	$c$	MPa	0.047	0.10	0.07	0.04
	$\Phi$	Degree (°)	16.05	17.1	16.45	15.46
Output/10	CR	mm/day	0.687	2.80	1.09	0.22

Four types of GPR kernel functions including squared exponential GPR, matern GPR, exponential GPR and rational quadratic GPR were considered to develop GPR models. The models have been developed using Regression Learner Application in MATLAB [25] (Table 3). According to the reported  $R^2$  values in Table 3, all the GPR models have similar prediction capabilities, where the exponential model, with a slight difference in the coefficient of determination ( $R^2 = 0.53$ ), located at a higher rank than the other GPR models. As shown in Fig. 15a, it is clear that the GPR model does not give reliable precision, so that the predicted CR values are relatively far away from the measured CR values (Fig. 15b). To evaluate the performance of the GPR model, the predicted CR values were compared with the measured values using the statistical indices ( $R^2$ , MSE, RMSE and MAE) (Table 3). It is observed that the GPR model has a relatively low capability in predicting the CR values.

## 10 Comparing the Prediction Models

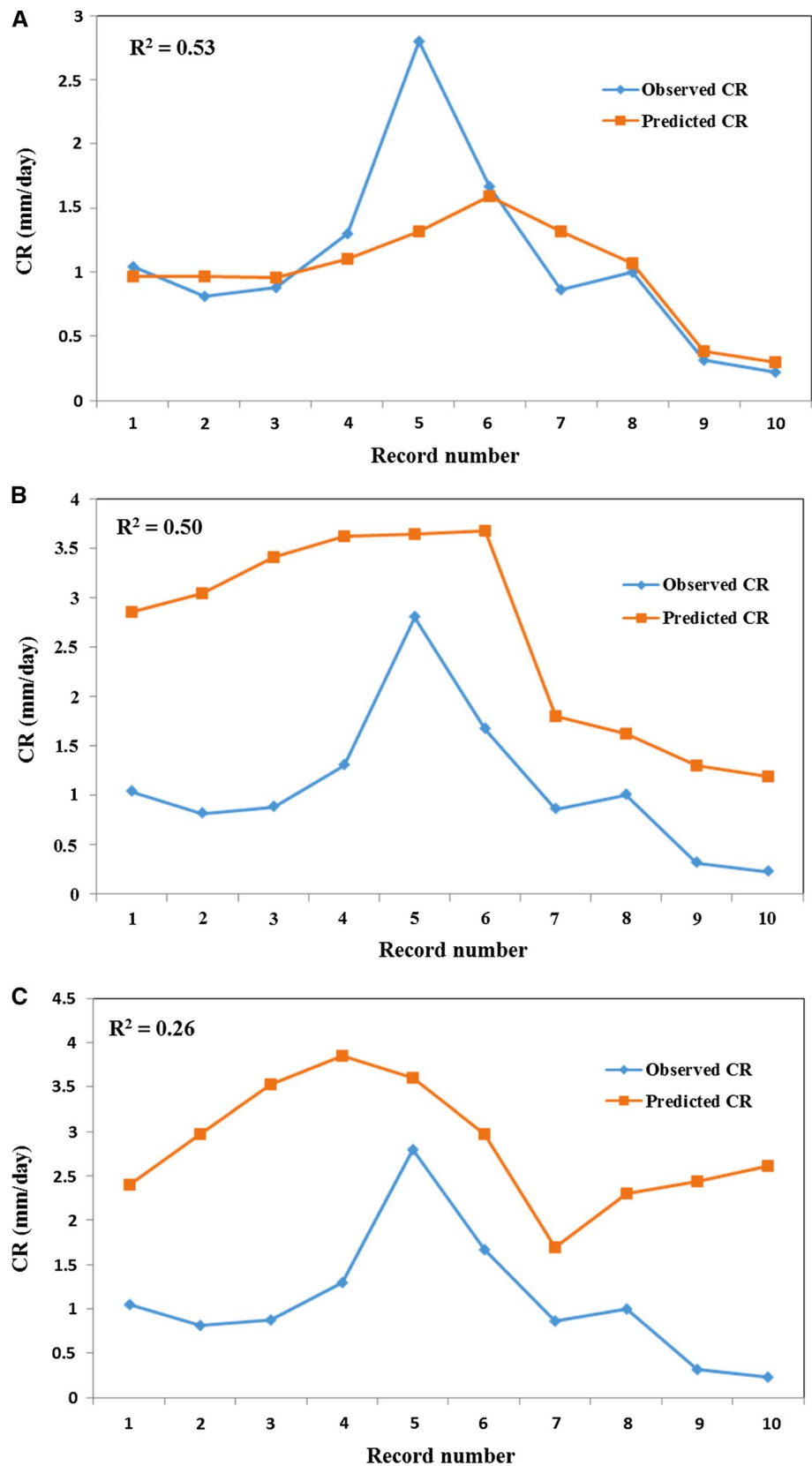
In this part, the capability of the developed models to predict CR values is evaluated. For investigating the capacity performance of the models, the four statistical indices ( $R^2$ , MSE, RMSE and MAE) were used as judgment criterion. It is approved that, when  $R^2$  is equal to 1 and MSE, RMSE and MAE are zero, a model will be excellent. Therefore, the models were ranked based on their statistical indices to predict the CR values (Table 3). The results show that the MLP and RBF models have good results than the other models with a remarkable difference. The best model for predicting CR is MLP, and RBF, SVR, MNR, ET, MLR, RT and GPR fall into the next ranks, respectively. A high complication degree of the dataset, with considering a high number of the variables, can be considered as a reason to justify the superiority of the ANN models. Hence, ANNs are commonly used to treat super complicated problems, in which too many variables to be simplified in a model. Among the other models, the SVR model has more accuracy in predicting the CR values ( $R^2 = 0.66$ ). A comparison between the SVR model and the ANN models reveals that

although the ANNs have higher capability than the SVR model, the SVR model can be applied as a simpler method in predicting CR values due to its ability in transforming the dataset from nonlinear form to linear form. Although the MNR model has  $R^2$  of 0.65 which is close to that for the SVR model and it also has lower MSE and RMSE values than the SVR model, the SVR is more capable in prediction because it could predict the CR values meaningfully. The MNR model has meaningless values (such as negative values) in predicted values. This kind of meaningless prediction mostly happens while trying to predict something which is out of range of the regression and data do not follow a normal distribution. The ET and the MLR models show results close together and have relatively reliable determination coefficients ( $R^2$  of 0.63 and 0.61, respectively). The RT model has relatively low success in predicting the CR values in spite of its ability in the use of kernel functions in modeling. Also, similar results were found for the GPR model, of course with slightly lower the statistic metric indices than the RT model (Table 3). This model has the least accuracy in predicting the CR values among the models with the rank of 8/8.

## 11 Validation of the Prediction Models

New data were collected from another project in Iran to check the validity of the prediction models (Table 4). The data belong to tunnel 2 of the Kermanshah–Khosravi Railway (west of Iran). Lithology of the tunnel route is shale. The shale rocks are mostly weak, but in some parts, their strength is high due to increasing silica content. Some of the equations extracted from the models were used to calculate convergence rate values for the tunnel according to input values presented in Table 4. Figure 16 shows a comparison between CR values obtained from the models with observed CR values for the tunnel. The results reveal that there is an acceptable correlation between the calculated CR and the observed CR for the SVR model, while the correlations for the MLR and the MNR prediction models are fair and weak, respectively (Fig. 16). The SVR model could estimate the CR values relatively close to the observed CR

**Fig. 16** Correlation between the observed CR values of the Kermanshah–Khosravi Railway tunnel and the predicted CR values from **a** the SVR model, **b** the MLR model and **c** the MNR model



values, which is evident of its high capability in prediction (Fig. 16a), whereas the regression models overestimated the CR values compared to the observed CR values (Fig. 16b, c), and therefore, they are not reliable predictors. Although it should be noted that due to the use of data related to similar rocks to examine validity of the models, they may not be extensible for other projects with different lithologies.

## 12 Conclusions

This study tried to establish different prediction models to develop meaningful relationships between some strength and design parameters (UCS,  $C$ ,  $\Phi$ ,  $\sigma_c$ ,  $\sigma_v$ , RQD, RMR,  $Q$ ,  $H$  and NB) with the recorded convergence rate values of the Namaklan twin tunnel in the west of Iran. MLR, MNR, MLP-ANN, RBF-ANN, SVR, RT, ET and GPR models are eight models that were developed for CR prediction. Multivariate linear regression with the backward analysis method showed that six parameters including  $H$ , NB, RMR,  $c$ ,  $\Phi$  and  $\sigma_c$  are the most important parameters in predicting the CR values. Hence, the parameters were used as inputs of all the models. The results showed that the MLP-ANN model can accurately predict the CR values ( $R^2=0.925$ ). The RBF-ANN model has been characterized as the second possible alternative for the prediction of the CR with  $R^2$  of 0.81. Although ANNs are capable in predicting the CR values, the SVR model demonstrated to be a promising alternative. The SVR model transformed the dataset from nonlinear form to linear form and developed an equation as a quick tool to estimate the CR values. Linear and nonlinear multiple regressions were also performed to generate a formula to predict the CR values. The MLR, ET and MNR models were relatively successful in predicting the measured CR with a determination coefficient of 0.61, 0.63 and 0.65, respectively. Also, it was observed that the RT and GPR models are the weakest in predicting the CR values among the prediction models. Finally, it can be concluded that the ANNs approaches could combine statistical techniques with machine learning techniques in a black box and make themselves the best possible models for the prediction of CR. As another result, the SVR model could be considered as a feasible prediction tool for predicting the convergence rate of the tunnel along with the ANNs. Also, the validity of the SVR, MLR and MNR models was checked by new data from another project. It was found that the capability of the SVR model is still reliable for predicting the reported CR values. However, more research works are needed to make the most accurate relationship to predict CR of tunnels excavated in different geological conditions with various types of lithology.

## References

- Kontogianni, V.; Psimoulis, P.; Stiros, S.: What is the contribution of time-dependent deformation in tunnel convergence? *Eng. Geol.* **82**(4), 264–267 (2006)
- Kontogianni, V.A.; Stiros, S.C.: Predictions and observations of convergence in shallow tunnels: case histories in Greece. *Eng. Geol.* **63**(3–4), 333–345 (2002)
- Eberhardt, E.: Numerical modelling of three-dimension stress rotation ahead of an advancing tunnel face. *Int. J. Rock Mech. Min. Sci.* **38**(4), 499–518 (2001)
- Paternesi, A.; Schweiger, H.F.; Scarpelli, G.: Finite element simulations of twin shallow tunnels. In: *ISRM Regional Symposium-EUROCK 2015 2015*. International Society for Rock Mechanics and Rock Engineering
- Yazdani, M.; Sharifzadeh, M.; Kamrani, K.; Ghorbani, M.: Displacement-based numerical back analysis for estimation of rock mass parameters in Siah Bisheh powerhouse cavern using continuum and discontinuum approach. *Tunn. Undergr. Sp. Technol.* **28**, 41–48 (2012)
- Chen, R.-P.; Lin, X.-T.; Kang, X.; Zhong, Z.-Q.; Liu, Y.; Zhang, P.; Wu, H.-N.: Deformation and stress characteristics of existing twin tunnels induced by close-distance EPBS under-crossing. *Tunn. Undergr. Sp. Technol.* **82**, 468–481 (2018)
- Barton, N.: Some new Q-value correlations to assist in site characterisation and tunnel design. *Int. J. Rock Mech. Min. Sci.* **39**(2), 185–216 (2002)
- Mahdevari, S.; Torabi, S.R.: Prediction of tunnel convergence using artificial neural networks. *Tunn. Undergr. Sp. Technol.* **28**, 218–228 (2012)
- Adoko, A.-C.; Jiao, Y.-Y.; Wu, L.; Wang, H.; Wang, Z.-H.: Predicting tunnel convergence using multivariate adaptive regression spline and artificial neural network. *Tunn. Undergr. Sp. Technol.* **38**, 368–376 (2013)
- Zarei, H.; Ahangari, K.; Ghaemi, M.; Khalili, A.: A convergence criterion for water conveyance tunnels. *Innov. Infrastruct. Solut.* **2**(1), 48 (2017)
- Rajabi, M.; Rahmannedjad, R.; Rezaei, M.; Ganjalipour, K.: Evaluation of the maximum horizontal displacement around the power station caverns using artificial neural network. *Tunn. Undergr. Sp. Technol.* **64**, 51–60 (2017)
- Hajihassani, M.; Kalatehjari, R.; Marto, A.; Mohamad, H.; Khosrotash, M.: 3D prediction of tunneling-induced ground movements based on a hybrid ANN and empirical methods. *Eng. Comput.* (2019). <https://doi.org/10.1007/s00366-018-00699-5>
- Chen, R.-P.; Zhang, P.; Kang, X.; Zhong, Z.-Q.; Liu, Y.; Wu, H.-N.: Prediction of maximum surface settlement caused by earth pressure balance (EPB) shield tunneling with ANN methods. *Soils Found.* **59**(2), 284–295 (2019)
- Ren, Q.; Wang, G.; Li, M.; Han, S.: Prediction of rock compressive strength using machine learning algorithms based on spectrum analysis of geological hammer. *Geotech. Geol. Eng.* **37**(1), 475–489 (2019)
- Seker, S.E.; Ocak, I.: Performance prediction of roadheaders using ensemble machine learning techniques. *Neural Comput. Appl.* **31**(4), 1103–1116 (2019)
- Ghasemi, E.; Kalhori, H.; Bagherpour, R.; Yagiz, S.: Model tree approach for predicting uniaxial compressive strength and Young's modulus of carbonate rocks. *Bull. Eng. Geol. Environ.* **77**(1), 331–343 (2018)
- Pham, B.T.; Prakash, I.; Bui, D.T.: Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **303**, 256–270 (2018)



18. Taha Consulting Engineers Company: Report of geological and engineering geology study of III and IV parcels of Namaklan tunnel. In: Tehran, Iran. (2014) (in Persian)
19. Yilmazkaya, E.; Dagdelenler, G.; Ozcelik, Y.; Sonmez, H.: Prediction of mono-wire cutting machine performance parameters using artificial neural network and regression models. *Eng. Geol.* **239**, 96–108 (2018)
20. Spss I. IBM SPSS statistics version 21. International Business Machines Corp., Boston, MA (2012)
21. Tu, J.V.: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epidemiol.* **49**(11), 1225–1231 (1996)
22. Koopialipour, M.; Tootoonchi, H.; Armaghani, D.J.; Mohamad, E.T.; Hedayat, A.: Application of deep neural networks in predicting the penetration rate of tunnel boring machines. *Bull. Eng. Geol. Environ.* **78**(8), 6347–6360 (2019)
23. Moosazadeh, S.; Namazi, E.; Aghababaei, H.; Marto, A.; Mohamad, H.; Hajihassani, M.: Prediction of building damage induced by tunnelling through an optimized artificial neural network. *Eng. Comput.* **35**(2), 579–591 (2019)
24. Moody, J.; Darken, C.J.: Fast learning in networks of locally-tuned processing units. *Neural Comput.* **1**(2), 281–294 (1989)
25. MATLAB and Statistics Toolbox Release 2018a The MathWorks, I., Natick, Massachusetts, United States (2018)
26. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I.: Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **13**(2), 361–378 (2016)
27. Hong, H.; Pradhan, B.; Jebur, M.N.; Bui, D.T.; Xu, C.; Akgun, A.: Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environ. Earth Sci.* **75**(1), 40 (2016)
28. Breiman, L.; Friedman, J.; Olshen, R.A.; Stone, C.J.: Classification and regression trees. Chapman & Hall, New York (1984)
29. Tomczyk, A.M.; Ewertowski, M.: Planning of recreational trails in protected areas: application of regression tree analysis and geographic information systems. *Appl. Geogr.* **40**, 129–139 (2013)
30. Dietterich, T.G.: An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach. Learn.* **40**(2), 139–157 (2000)
31. Rasmussen, C.E.: Gaussian processes in machine learning. In: Summer School on Machine Learning, pp. 63–71. Springer, Berlin, Heidelberg (2003)
32. Pal, M.; Deswal, S.: Modelling pile capacity using Gaussian process regression. *Comput. Geotech.* **37**(7–8), 942–947 (2010)
33. Seeger, M.; Williams, C.; Lawrence, N.: Fast forward selection to speed up sparse Gaussian process regression. EPFL (2003)
34. Quiñero-Candela, J.; Rasmussen, C.E.: A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* **6**(Dec), 1939–1959 (2005)
35. Yuan, J.; Wang, K.; Yu, T.; Fang, M.: Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression. *Int. J. Mach. Tool Manufact.* **48**(1), 47–60 (2008)
36. Williams, C.K.: Prediction with Gaussian processes: From linear regression to linear prediction and beyond. In: *Learning in Graphical Models*, pp. 599–621. Springer, Dordrecht (1998)
37. Williams, C.K.; Rasmussen, C.E.: Gaussian processes for regression. In: *Advances in Neural Information Processing Systems*, pp. 514–520. MIT Press, Denver, CO, USA (1996)

