Geomechanics and Engineering, *Vol. 30, No. 1 (2022) 75-91*
https://doi.org/10.12989/gae.2022.30.1.075

Several models for tunnel boring machine performance prediction based on machine learning

Arsalan Mahmoodzadeh[1a], Hamid Reza Nejati*[1], Hawkar Hashim Ibrahim[2b], Hunar Farid Hama Ali[3c], Adil Hussein Mohammed[4d], Shima Rashidi[5e] and Mohammed Kamal Majeed[6f]

[1]*Rock Mechanics Division, School of Engineering, Tarbiat Modares University, Tehran, Iran*
[2]*Department of Civil Engineering, College of Engineering, Salahaddin University-Erbil, 44002 Erbil, Kurdistan Region, Iraq*
[3]*Department of Civil Engineering, University of Halabja, Halabja, Kurdistan Region, Iraq*
[4]*Department of Communication and Computer Engineering, Faculty of Engineering, Cihan University-Erbil, Kurdistan Region, Iraq*
[5]*Department of Computer Science, College of Science and Technology, University of Human Development, Sulaymaniyah, Kurdistan Region, Iraq*
[6]*Information Technology Department, Faculty of Science, Tishk International University (TIU), Erbil, Kurdistan Region, Iraq*

(Received August 20, 2021, Revised May 2, 2022, Accepted May 4, 2022)

Abstract. This paper aims to show how to use several Machine Learning (ML) methods to estimate the TBM penetration rate systematically (TBM-PR). To this end, 1125 datasets including uniaxial compressive strength (UCS), Brazilian tensile strength (BTS), punch slope index (PSI), distance between the planes of weakness (DPW), orientation of discontinuities (alpha angle-α), rock fracture class (RFC), and actual/measured TBM-PRs were established. To evaluate the ML methods' ability to perform, the 5-fold cross-validation was taken into consideration. Eventually, comparing the ML outcomes and the TBM monitoring data indicated that the ML methods have a very good potential ability in the prediction of TBM-PR. However, the long short-term memory model with a correlation coefficient of 0.9932 and a route mean square error of 2.68E-6 outperformed the remaining six ML algorithms. The backward selection method showed that PSI and RFC were more and less significant parameters on the TBM-PR compared to the others.

Keywords: feature selection; machine learning; penetration rate; tunnel boring machine; tunneling

1. Introduction

Aiming the classification of rocks, access to geological data, and analysis of tunnels construction design, many digitalization ways have been provided by researchers. Digitalization of tunnel construction is one of the most crucial engineering problems that can provide a suitable solution in less time and at a lower cost. With the growing and developing cities and humanity's need for more intercity communication, the demand for tunnels is growing. For this reason, different tunnel construction methods are being developed and paid much attention today. One of these methods is the mechanized tunneling method, in which tunnel boring machines (TBMs) are used for tunnel construction. TBM performance prediction is a very important issue that has been seriously considered by researchers (Gholamnejad and Tayarani 2010, Li *et al.* 2020). TBM-penetration rate (TBM-PR) and TBM-advanced rate (TBM-AR) are two of the machine performance factors that need to be considered during tunnel construction. Prediction of these factors can be very important in the design analysis of tunnel construction and reduction of uncertainties, including time delays and cost overruns.

Theoretical models are the most common methods for predicting TBM achievement, empirical models (Hassanpour *et al.* 2010, Goh *et al.* 2018), and machine learning (ML) models (Alvarez Grima *et al.* 2000, Benardos and Kaliampakos 2004, Salimi *et al.* 2016, Jahed Armaghani *et al.* 2017, Minh *et al.* 2017, Koopialipoor *et al.* 2018, Fattahi 2019, Zhou *et al.* 2019, Xu *et al.* 2019, Jahed Armaghani *et al.* 2019). The mechanical equilibrium equation for a single cutter was used to create theoretical models. In the classical theoretical model of the Colorado School of Mines (CSM); many parameters are considered, like rock uniaxial compressive strength (UCS) and geometrical cutter properties. Within the CSM model, load on the single cutter was measured based on the mechanical equilibrium equation. The rock mass discontinuity influence

*Corresponding author, Ph.D.
E-mail: h.nejati@modares.ac.ir
[a]M.Sc,
E-mail: m.arsalan@modares.ac.ir
[b]M.Sc,
E-mail: hawkar.ibrahim@su.edu.krd
[c]M.Sc,
E-mail: hunar.hamaali@uoh.edu.iq
[d]M.Sc,
E-mail: adil.mohammed@cihanuniversity.edu.iq
[e]M.Sc,
E-mail: shima.rashid@uhd.edu.iq
[f]M.Sc,
E-mail: mohammed.kamal@tiu.edu.iq

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

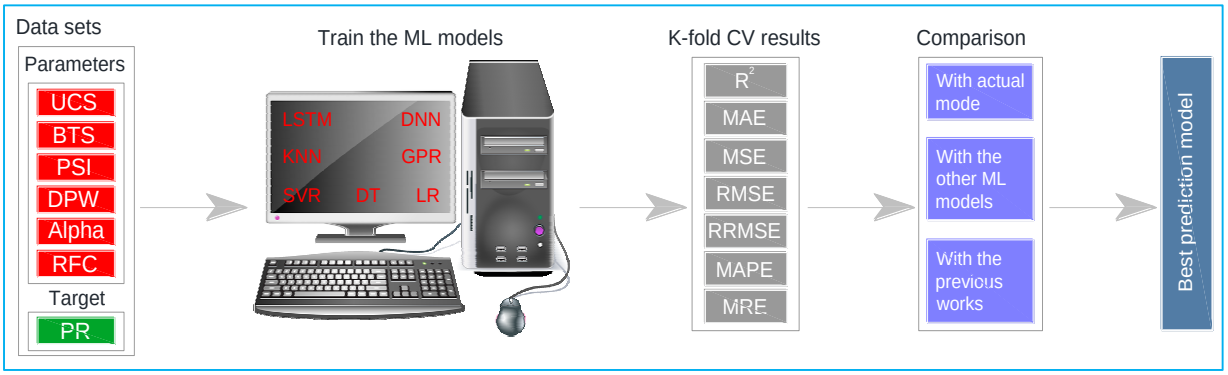

Fig. 1 Article structure in a schematically way

on the performance of TBM was neglected in the CSM model. The theoretical model concentrates on the single cutter's force and its analytic. The force acting on the TBM is reduced to a static force that does not represent the dynamic interaction of the TBM with the rock mass. Moreover, the experimental models are designed to provide a statistical relationship between the output parameters of TBMs and geological conditions. The experimental models' reliability was contingent on the quantity and quality of data. The disadvantage of experimental models is that the relationship must be established for each geological event, which limits their applicability in a wide variety of conditions.

Recently, Machine learning (ML) digital technics have shown their potential ability to solute different challenging problems (Liu *et al.* 2021a, Xiang *et al.* 2021, Liu *et al.* 2021b, Bai *et al.* 2021). With the increase in total data and the ease with which more computational resources can be acquired, a substantial increase in the application of ML techniques to different tasks has been observed. ML technics have been applied and developed in many fields and technologies, and although they have not yet developed as much as other disciplines, they have eventually reached the three engineering fields of tunneling, geotechnics, and geology. ML models rely upon geological conditions calculated data and TBM system parameters. This model is capable of considering multiple complex variables in an indirect way including TBM output and the geological situation in the surrounding region. ML methods for estimating TBM tunneling parameters have been developed by various researchers. Even so, since the number of data sets used in the previous studies was mainly small, their results cannot be trusted. Also, in most previous studies, only one of the ML methods has been used, while applying several ML methods and comparing their prediction results can be very effective. In this way, the ability of the other ML methods can be examined, and the best model can be specified.

This study aims to forecast the TBM-PR, which is an important TBM performance indicator, by utilizing 1125 data sets collected from the Alborz service tunnel on Tehran–Shomal motorway in Iran. Seven ML approaches include long-short term memory (LSTM), K-nearest neighbors (KNN), deep neural networks (DNN), Gaussian process regression (GPR), support vector regression (SVR), decision tree (DT), and linear regression (LR) have been used in the prediction of TBM-PR. Each of the 1125 data sets includes six parameters of uniaxial compressive strength (UCS), Brazilian tensile strength (BTS), punch slope index (PSI), the distance between the planes of weakness (DPW), the orientation of discontinuities (alpha angle-α), rock fracture class (RFC), and a target of TBM-PR. In order to judge the performance of machine learning methods, it is possible to provide more confidence in the results through verification of the 5-fold cross-validation (CV). Lastly, the predictive outcomes of the seven ML methods applied in this study have been compared to the TBM monitoring data, to each other, and the other ML methods in the previous studies in the field done by the other researchers. Lastly, the best prediction model has been specified and suggested for the prediction of the TMB-PR of tunnel future levels. Also, to determine the most influential parameters considered on the life of disc cutters, a backward selection method has been used.

The remaining sections of this report are organized as follows: section 2 discussed the database used in this article, while section 3 discusses various methods for testing the outcomes of such intelligent models. Furthermore, section 4 explains in detail the ML methods used in this research and analyses the predicted results. Moreover, in section 5, the prediction results and prior study in this area are compared for each intelligent model and other models. Besides, sections 6 and 7 present the discussion and conclusions, respectively. The schematic description of the structure of this article is shown in Fig. 1.

2. Database

In order to access the database, the Alborz service tunnel on Tehran– Shomal motorway project in Iran is considered in this study. In the Iranian capital of Tehran, there is a new motorway known as the Tehran- Shomal motorway, which connects the cities of Tehran and Chalus that is located at the Caspian Sea. It is equal to 121 km in length. Currently, traffic is limited to narrow roads that travel through the Alborz mountains, and the trip takes between five and six hours. Once the project is completed, travel time would be significantly shortened to less than two hours, and capacity will be increased overall. The alignment of the highway has over 30 twin tunnels with double lanes. With a length of

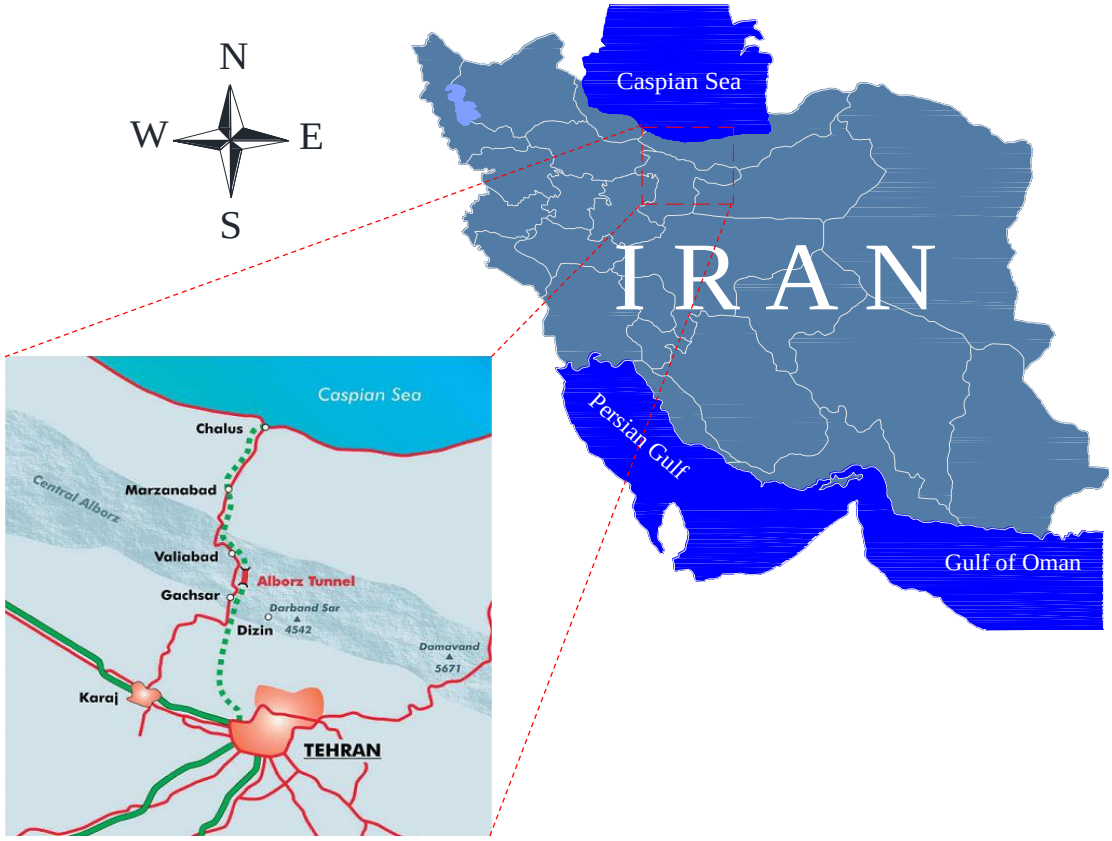

Fig. 2 Project location of the Alborz tunnel

Table 1 A brief review of the database used

	UCS	BTS	PSI	DPW	α	RFC	TBM-PR
count	1125.000	1125.000	1125.000	1125.000	1125.000	1125.000	1125.000
mean	147.5267	9.142311	32.82284	0.804133	38.30808	2.952889	1.983396
std	12.757017	0.813436	6.666596	0.263426	9.784725	1.443633	0.163017
min	118.30000	6.300000	25.00000	0.000000	0.000000	1.000000	1.710000
25%	139.08000	8.600000	28.30000	0.800000	36.10000	2.000000	1.880000
50%	143.25000	9.400000	29.80000	0.800000	38.50000	3.000000	1.930000
75%	151.53000	9.800000	35.00000	0.800000	39.70000	4.000000	2.030000
max	199.70000	11.40000	58.00000	2.000000	89.00000	7.000000	2.950000

6,400 m and an altitude of 2,400 m, the Alborz tunnel is the longest. The location of the Alborz service tunnel is shown in Fig. 2.

Between the two main tunnels is the service tunnel, and it is often used as an entranceway to the other tubes for maintenance and site exploration. Currently, the construction of the Alborz tunnel is completed by TBM. The entrance of the tunnel is considered the northern mouth (to the Shomal-S), whereas the tunnel's exit is considered the southern mouth (to the Tehran-T). The Lithology of the tunnel route mainly consists of Tuffs, Andesite, Anidrite, Limestone, and Sandstone. The compressive strength of the rocks of the tunnel route varies from 20 to 120 Mpa. The longest fault is located in the ST5339-5361, where the water flows into the tunnel is high from this fault and provides conditions for squeezing the rocks of the tunnel pathway.

Fig. 3 illustrates the longitudinal geological profile of the Alborz service tunnel.

According to the previously studied TBM-PR prediction and the availability of the data along the Alborz service tunnel route, six parameters of UCS, BTS, PSI, DPW, α, and RFC effective on the TBM-PR are considered in this study. A database including 1125 data sets was collected along the tunnel route that was accessed after the tunnel construction. Table 1 summarizes the database.

3. Results analysis methods

The predictive accuracy of intelligent methods is evaluated using seven statistical evaluation criteria. These criteria are defined by Eqs. (1)-(7) as the coefficient of

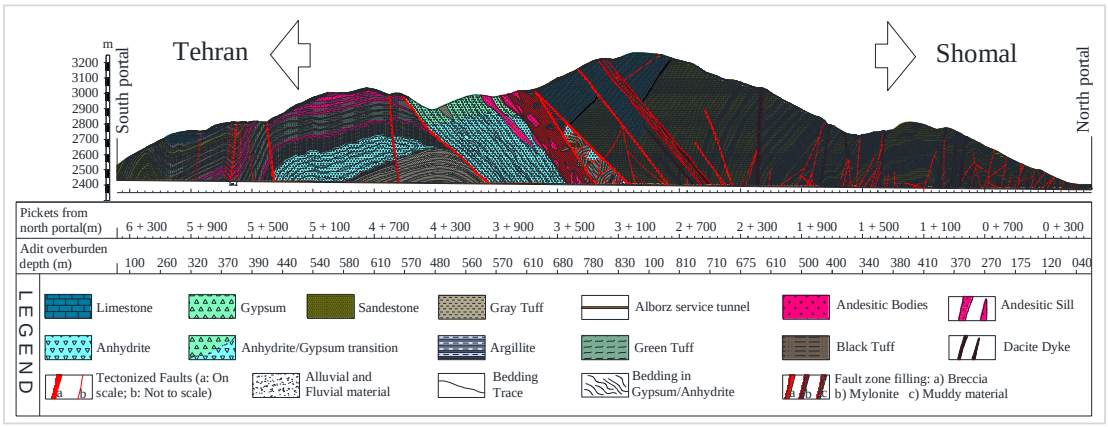

Fig.3 Geological map of Alborz service tunnel

determination (R^2), mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), relative root mean square error (RRMSE), mean absolute percentage error (MAPE), and mean relative error (MRE)

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

$$\mathrm{MAE} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} |f(x_i) - f^*(x_i)| \quad (2)$$

$$\mathrm{MSE} = \frac{1}{n} \sum_{1}^{n} (f(x_i) - f^*(x_i))^2 \quad (3)$$

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^{n} (f(x_i) - f^*(x_i))^2} \quad (4)$$

$$\mathrm{RRMSE} = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^{n} \left(\frac{f(x_i) - f^*(x_i)}{f(x_i)}\right)^2} \quad (5)$$

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{f(x_i) - f^*(x_i)}{f(x_i)} \right| \quad (6)$$

$$\mathrm{MRE} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \frac{|f(x_i) - f^*(x_i)|}{|f(x_i)|} \quad (7)$$

where $f(x)$ represents the actual value and $f^*(x)$ denotes the expected value, $\bar{f}(x)$ and $\overline{f^*}(x)$ are the means of actual and predicted values, respectively, and n represents the total number of data sets.

4. TBM-PR prediction using intelligent methods

In this analysis, GPR, SVR, DT, and LR modeling were performed using the MATLAB program 2018. Anaconda version 3.6 is applied to perform the LSTM, DNN, and KNN algorithms. Anaconda is a metapackage that contains all Python packages.

There are different cross-validation (CV) methods for model selection, such as the hold-out method, K-fold CV, leave one out CV, and bootstrap method. In this analysis, the K-fold CV (K=5) is considered to validate the prediction performance. In the K-fold CV method, the original training set is split in K folds of the same size using random sampling without repetition. The models are trained K times, each one of them using as training set K-1 folds and 1 (the not used one) as a validation set. The prediction error is the average of the K individual errors. Error variance can be used as a measure of the stability of the model. The advantage of this method is that it matters less how the data is divided. The model is less prone to having selection bias. Also, since each observation is used for the confirmation process only once, and every observation is used for confirmation and training, this technique has been better than a repetitive random sub-sampling procedure (Zhang *et al.* 2020).

4.1 LSTM

Deep learning is a standard form of neural networks with many layers. The knowledge included within these deep learning networks is more easily remembered than that contained within ordinary neural networks. A RNN is a type of neural network that combines a loop network. Knowledge persistence exists in the networks across these loops. This network collects inputs and data from previous networks, executes specified procedures, and outputs the next network output. Some applications may only need current knowledge, while others may also need additional information from the past. A learning lag exists in redundant neural networks in which the gap between the desired input and the requirement is improved. However, LSTMs networks are a special kind of RNN, capable of studying these situations. Such networks are intentionally designed so that, on the permanent networks, the long-term dependency problem can be mitigated.

Recalling knowledge over a long period of time is an important function of LSTM. It is a common choice because the model's precision is always dependent on the sum of prior knowledge. The basic LSTM module, also known as the repeating module, interacts with four layers of

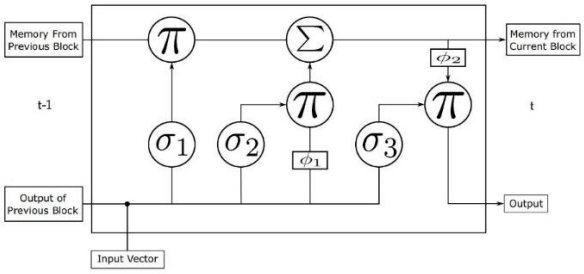

Fig. 4 LSTM repeating module (Kumar *et al.* 2018)

neural networks, as seen in Fig. 4. In the module, there are three gates with activation functions of σ1, σ2, and σ3, and it contains two output activation functions such as Φ1 and Φ2, as seen in Fig. 4. To depict addition and multiplication element-by-element, the Σ and π symbols are used. A bullet (•) symbol is used to describe concatenating mathematical procedures. The cell state is the main component of LSTM, in which the current block memory (S_t-1) is produced from the prior block memory. The information flow is then promptly approved further down the line. The first layer (S_t) in the network determines the sum of the prior knowledge to flow, and Eq. (8) shows the activity performed by this layer (Kumar *et al.* 2018).

In the network, two layers and their associations influence the mechanism of storing new data in the cell state. A sigmoid layer (σ_2) that find out the (I_t) value to be updated see Eq. (9) and Φ_1 tanh layer, which makes a new candidate value ($\tilde{S}_t$) as seen in Eq. (10). To make a case, both are to be added together with the state. The cell state is lastly replaced with Eq. (11).

$$c_{ft} = \sigma_1\left(W_{cf}.[O_{t-1}, x_t] + b_{cf}\right) \quad (8)$$

$$I_t = \sigma_2(WI.[O_{t-1}, x_t] + bI) \quad (9)$$

$$S_t'' = tanh(WS.[O_{t-1}, x_t] + bS) \quad (10)$$

$$S_t = c_{ft} \times S_{t1} + I_t \times S_{t-1}'' \quad (11)$$

The key features of the LSTM models, which motivate us to use it, is as follows:

The data of our problem has a time dependency feature. i.e., the variables are dependent each other in terms of time. Therefore, we decide to use RNN-based models to consider this time dependency. We can say that when we move from RNN to LSTM (Long Short-Term Memory), we are introducing more & more controlling knobs, which control the flow and mixing of Inputs as per trained Weights. So, LSTM gives us the most Control-ability and, thus, better Results. But also comes with more Complexity and Operating Cost. Consequently, for the sake of having the best accuracy, we choose LSTM instead of conventional RNN.

The process of understanding the data plays a key role in the process of choosing the right algorithm for the right problem. Some algorithms can work with smaller sample sets while others require tons and tons of samples. Certain algorithms work with categorical data while others like to work with numerical input. Because our datasets are large We propose the use of LSTM Algorithm.

Anaconda version 3.6 is used to implement the LSTM algorithm. The optimization of the LSTM's hyper-parameters is separated into three stages:

- First, a generic model for the regression of the LSTM model process is constructed, and meta-heuristic algorithm of particle swarm optimization (PSO) is used to optimize the hype-parameters of the LSTM model.
- In the second stage, the previous stage's LSTM model and the starting value of the parameters to train the model using input data were utilized, and then the Bayesian theory was utilized to optimize the LSTM hyper-parameters.
- In the third stage, in order to get the final prediction results, new data was fed into the LSTM model after the first PSO algorithm stage and the initial value of the parameter has been optimized so that the optimization problem can be implemented immediately and the parameter value does not easily fall into an optimal local value.

Fig. 5 illustrates the LSTM model used in this article. Looking at Fig. 4, the input and output layers are designed based on the six input parameters of UCS, BTS, PSI, DPW, α, and RFC, and one target of TBM-PR. Four hidden layers are considered with 40, 20, 10, and 5 neurons from the first layer to the fourth layer, respectively. Table 2 contains information about the parameters used in the LSTM model.

Fig. 6 explains the 5-fold CV results of the LSTM model in comparison to the real PR along the Alborz service tunnel. The error values for the LSTM model are very little and negligible, as shown in Fig. 6. For more analysis of the LSTM performance on the TBM-PR, the other evaluation criteria results are also presented in Table 3. All of these findings indicate the LSTM model's possible ability to forecast TBM-PR. Thus, using the database applied in this paper, the LSTM model presented in this paper can be applied for the TBM-PR prediction of future tunneling projects with high confidence.

Table 2 Details of the effective Parameters that have been investigated for the LSTM model

Parameter	Value or type`
Number of neurons	Input layer: 6, First hidden layer: 40, Second hidden layer: 20, Third hidden layer: 10, Fourth hidden layer: 5, Output layer: 1
Number of LSTM hidden layers	4
Optimizer	'Nadam'
Activation function	'ReLu'
Kernel_initializer	'lecun_uniform'
Batch_size	10
Loss	'mse'
Epochs	100
Validation_split	0
verbose	1

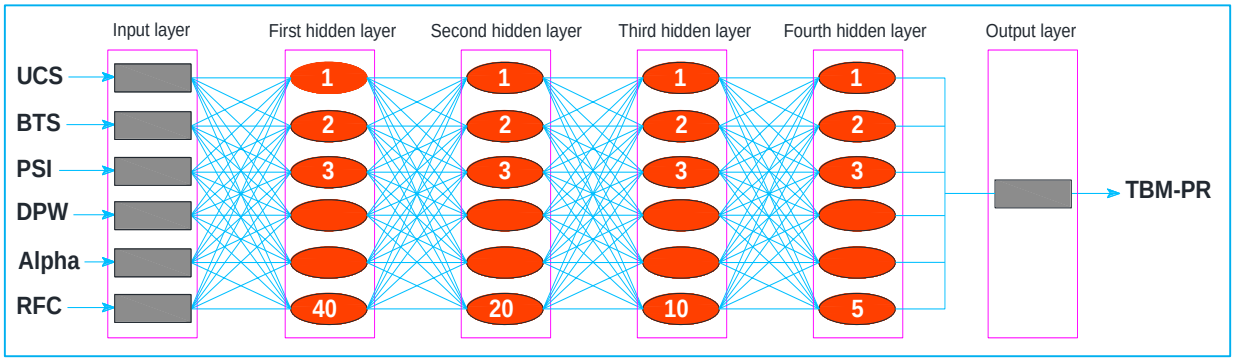

Fig. 5 Structure of the LSTM model

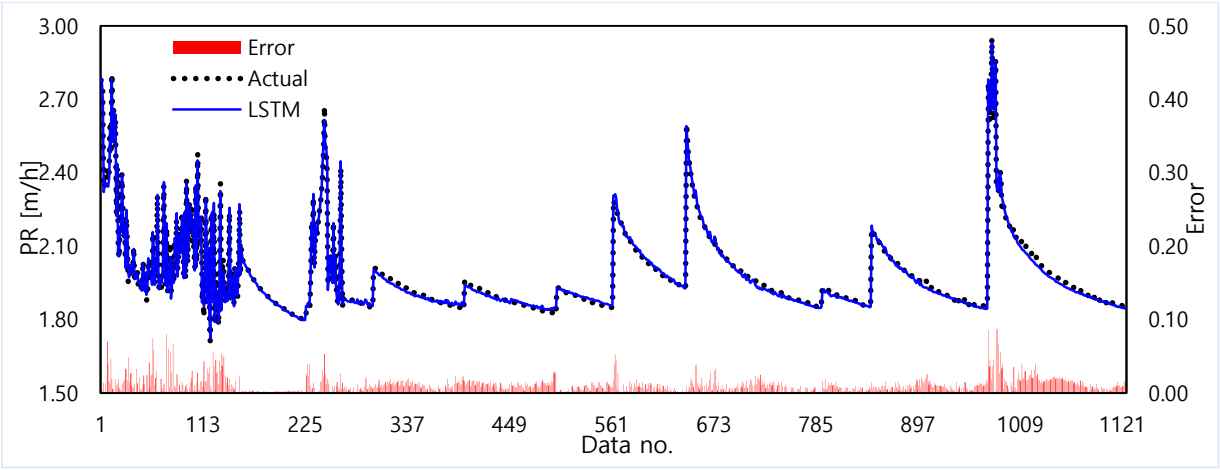

Fig. 6 TBM-PR prediction results of the LSTM model

Table 3 Assessment criteria outcomes of the LSTM model

R^2	MAE	MSE	RMSE	RRMSE	MAPE [%]	MRE
0.9932	2.68E-06	8.06252E-09	8.97915E-05	3.23096E-05	9.63285E-05	9.63285E-07

4.2 DNN

In the layered neural network of DNN, a deep feed-forward neural network is available. The hidden units between its outputs and inputs are more than one layer. The layers contain input layers, which are followed by mid-layers, hidden layers, and finally, the output layer. Consequently, the input, hidden, and output layers are all connected to the network's neighboring layers. Furthermore, DNN is especially well-suited to analyzing raw input data because it can recognize patterns and learn useful features from it without the need for rigorous feature engineering, hand-crafted guidelines, or data pre-processing. Moreover, with the rise in training data, its efficiency further improves (Hannun *et al.* 2019). DNNs have a much wide range of applications, from simple text creation to machine vision tasks, and the early uses of DNN are in automatic translation (Seifert *et al.* 2017).

The key features of the DNN models, which motivate us to it them, is as follows:

Compared to its predecessors, the main advantage of DNN is that it automatically detects the important features without any human supervision. In this case, we have no idea about the dependency of feature; therefore, we use DNN as one of the comparators to decrease the dimensionality of datasets and see the effects of the automatic dimensionality reduction in the results.

Anaconda 3.6 is also used to apply the DNN algorithm. The DNN method's most critical parameters are the number of hidden layers, the number of neurons within each layer, the activation function, the batch size, the kernel initializer, the epochs, and the verbose. To optimize the hyper-parameters of the DNN model, the same three steps as stated for the LSTM model have been used. As a result, the optimal implementation of the DNN model is presented in Fig. 7. As shown in Fig. 7, the DNN model has one input layer, four hidden layers and one output layer. Each of the hidden layers contains 100 neurons. Table 4 contains further details on the parameters considered for the DNN process.

Fig. 8 illustrates the DNN model's 5-fold CV outcomes in comparison to the real PR along the Alborz service tunnel, and the DNN model's error values are very small. For more analysis of the DNN model performance on the TBM-PR, the other assessment criteria results are also shown in Table 5. All of these results demonstrate the DNN model's possible capacity to forecast TBM-PR. Therefore, taking into account the database utilized in this paper, the DNN model presented in this paper can be applied for the TBM-PR prediction of future tunneling projects with high confidence.

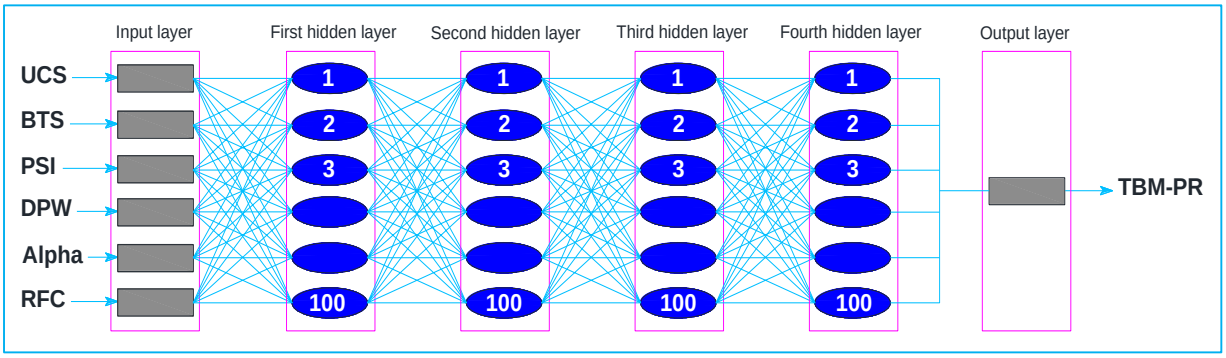

Fig. 7 Structure of the DNN model

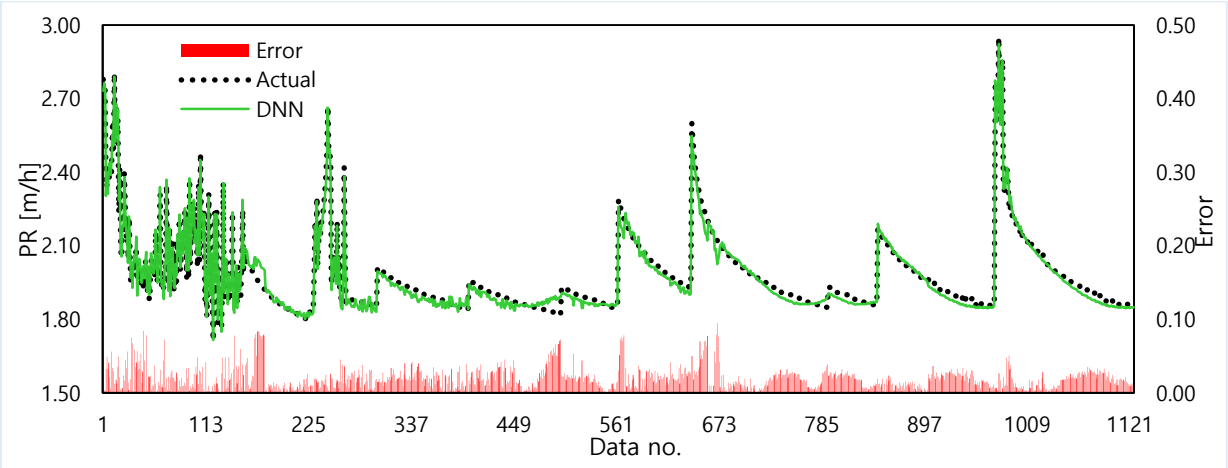

Fig. 8 TBM-PR prediction results of the DNN model

Table 4 Details of the effective Parameters that have been investigated for the DNN model

Parameter	Value or type
Number of neurons	Input layer: 6, First hidden layer: 100, Second hidden layer: 100, Fourth hidden layer: 100, Fifth hidden layer: 100, Output layer: 1
Number of hidden layers	4
Kernel_initializer	'he_normal'
Activation function	'reLu'
Optimizer	'Nadam'
verbose	1
Batch_size	10
Epochs	100

Table 5 Assessment criteria outcomes of the DNN model

R^2	MAE	MSE	RMSE	RRMSE	MAPE [%]	MRE
0.9784	4.25E-05	2.03097E-06	0.001425121	0.000512799	0.001528872	1.52887E-05

4.3 KNN

KNN is a simple and well-managed machine learning algorithm that can be applied to classification and regression problems. It is important to apply and comprehend, but it is very difficult to do as the amount of data in use expands considerably. KNN functions through the determination of the ranges between a query and all the examples in the results, choosing the specified number of examples for the query (K), then voting for the most frequent label (in the situation of classifying) or averaging the labels (In the situation of a regression).

The key features of the KNN model, which motivate us to use it, is as follows:

KNN is a non-parametric method we used for prediction in this paper. It is one of the easiest ML approaches have been recently used. It is a lazy learning model with local approximation. We use this model considering the following terms:

The key Advantages:

- Easy and simple machine learning model.
- Few hyperparameters to tune.

Hyperparameters:

- KNN mainly involves two hyperparameters, K value & distance function.
- K value: how many neighbors to participate in the KNN algorithm. k should be tuned based on the validation error.
- Distance function: Euclidean distance is the most used similarity function. Manhattan distance, Hamming Distance, Minkowski distance are different alternatives.

Assumptions:

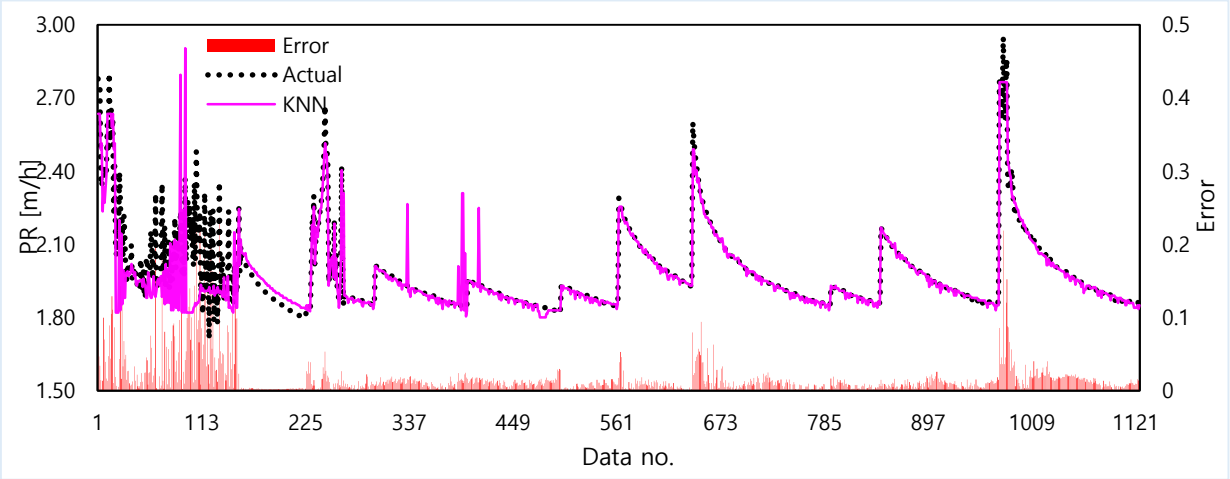

Fig. 9 TBM-PR prediction results of the KNN model

Table 6 Assessment criteria outcomes of the SVR model

R^2	MAE	MSE	RMSE	RRMSE	MAPE [%]	MRE
0.7635	5.60E-05	3.5293E-06	0.0018786	460.0006759910	0.0020154152	.01541E-05

- There should be a clear understanding of the input domain.
- Feasibly moderate sample size (due to space and time constraints).
- Collinearity and outliers should be treated prior to training.

Comparison with other models:

- KNN and other models' general difference is the large real time computation needed by KNN.

As a matter of fact, choosing the optimal KNN's parameters and hyperparameters is beyond the scope of our work. However, considering the time complexity of GridSearch, we try to use the knowledge of the following references to select the potential values for K and then we try to examine these potential values to choose the best one. In this regard, the values 2, 3, 4, 7, and 9 were examined to choose the optimal value. After careful consideration, the best value of K was 2 to give the best prediction of the disc cutter life.

Fig. 9 illustrates the 5-fold CV outcomes of the KNN model and compares them to the real PR along the Alborz service tunnel. It can be seen in Fig. 9, KNN model produces small error values. For more analysis of the KNN model performance on the TBM-PR, the other evaluation criteria results are also presented in Table 6. All the results clarify the medium capability of the KNN model for the TBM-PR prediction. Therefore, by taking into account the database utilized in this paper, the KNN model presented in this paper can be applied for the TBM-PR prediction of future tunneling projects with some risk.

4.4 GPR

GP could be described as a nonparametric machine learning algorithm based on supervised learning, divided into two major groups: regression and classification. Whereas the regression target values are continuous quantity forecasts, the classifying results are discrete labels. However, the GP is defined as a collection of arbitrary variables in which any finite subset of the variables has a joint multivariate Gaussian distribution. The covariance functions and mean functions specify a GP. The process of $f\ (x)$ with the covariance function $k(x,x')$ and the mean function $m(x)$ could be defined as

$$\mu(x) = E\ [f(x)] \tag{12}$$

$$k(x,x') = E[\big(f(x) - \mu(x)\big)\big(f(x') - \mu(x')\big)] \tag{13}$$

Hence, the GP is defined as

$$f(X) \sim GP\big(\mu(X), k(X,X')\big) \tag{14}$$

For the notational simplicity, the $\mu(x)$ function (i.e., mean function) is assumed to be zero. In addition, the $k(x,x')$ function (i.e. covariance function) will be calculated by a function known as kernel function.

Training data for real process f is denoted as

$$D = \{\big(X^{(i)}, y^{(y)}\big)\ \big|\ i = 1,2,\dots,n\} \tag{15}$$

where X is an input vector of dimension d, y is the corresponding real-valued noisy observation such that $y = f(X) + \in$. Let D be the collection of d x n input matrix, X, and target vectors, y.

$$D = \{X, y\} \tag{16}$$

The challenge now would be to calculate the required values f^* for unknown input X^*. As per the Gaussian process prior, the f and f^* joint distribution is

$$\begin{bmatrix} f \\ f^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K + \sigma_n^2 I & K^{*T} \\ K^* & K^{**} \end{bmatrix}\right) \tag{17}$$

where σ_n^2 is the noise variance, I is the Identity matrix, $K^*$is a matrix denoting the kernel function assessed for each member of X^* x X and similarly K is defined over X x X and K^{**}over X^* x X^*. Hence, the conditional distribution of f^* given X^*, X and f can be found

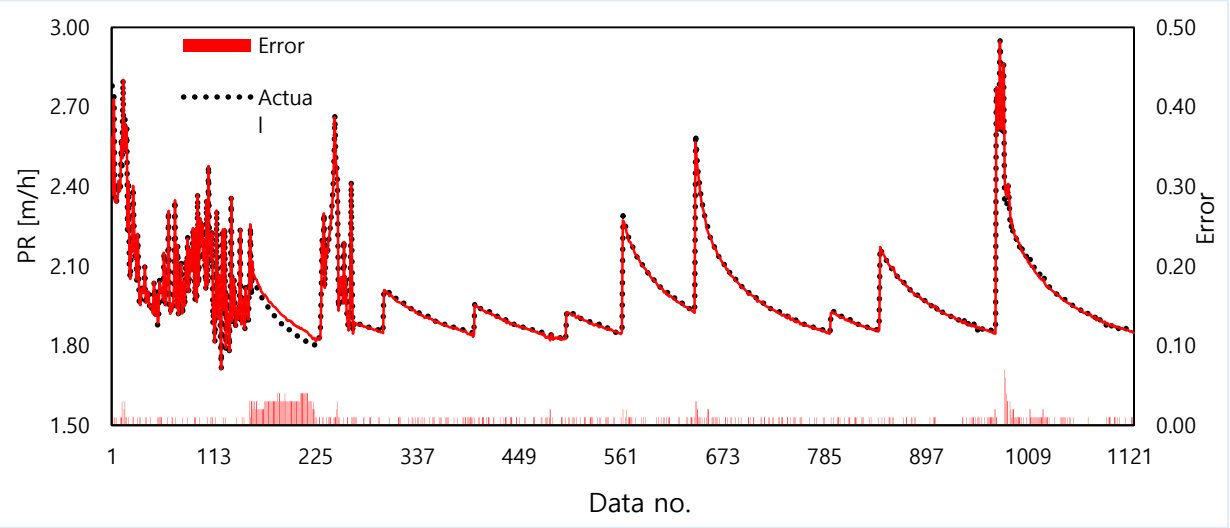

Fig. 10 TBM-PR prediction results of the GPR model

$$f^*|X^*, X, f \sim N\left(K^*(K+\sigma_n^2 I)^{-1} f, K^{**} - K^*(K+\sigma_n^2 I)^{-1} K^{*T}\right) \quad (18)$$

To reduce the mean squared error (MSE), the best estimator of f^* will be the mean of the above normal distribution

$$\hat{f}^* = K^*(K+\sigma_n^2 I)^{-1} f \quad (19)$$

The GP technique could be represented as the linear combination of n kernel functions

$$\hat{f}^* = \sum_{i=1}^{n} \alpha^{(i)} k\left(x^*, x^{(i)}\right) \quad (20)$$

where α is the coefficient vector, and the optimal MSE coefficient vector will be $\alpha = (K+\sigma_n^2 I)^{-1} f$. Thus, the most crucial aspect of the GPR is the covariance (Kernel) function.

The key features of the GPR model, which motivate us to use it, is as follows:

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

In the regression learner app embedded in MATLAB software 2018, four different models for GPR include squared exponential, rational quadratic, exponential, and Matern 5/2. After modeling by this program, the model type with the most accurate results is taken into account. Also, the optimization mode is considered in the app, so that the app itself optimizes the amount and type of hyper-parameters of the GPR model.

Table 7 For TBM-PR prediction, the optimal model form and hyper-parameter values of GPR were determined

Parameter	Value or type
KernelFunction	Rational Quadratic
FitMethod	Exact Gaussian process regression
BasisFunction	'Constant'
Sigma	0.0112
Beta	2.3223

Three primary hyperparameter tuning approaches have been proposed in literatures, including random search, grid search, metaheuristic-based search (smart search) although there are also other approaches which are less popular. Contrary to the "dumb" alternatives of grid search and random search, metaheuristic-based hyperparameter tuning, including PSO, is much less parallelizable. Instead of producing all the candidate points up front and investigating the batch in parallel, metaheuristic-based tuning approaches pick a few hyperparameter settings, investigate their performance, then determine where to sample next. These methods are intrinsically iterative and sequential process, which is not parallelizable. On one hand, making fewer evaluations and reducing the total time complexity are the primary goal of any computation algorithm and on the other hand, metaheuristic-based search algorithms require computation time to find out where to place the next set of samples. Besides, metaheuristic-based search algorithms also contain parameters of their own that need to be tuned.

These shortcomings motivate us to choose random search, which has the least time and space complexity.

Considering the Bergstra theorem that stated that "if the close-to-optimal region of hyperparameters occupies at least 5% of the grid surface, then random search with 60 trials will find that region with high probability", The hyperparameters of the GPR and SVR methods, which are shown in Tables 2 and 3, were optimized with a random

Table 8 Assessment outcome measurements of the GPR model

R^2	MAE	MSE	RMSE	RRMSE	MAPE [%]	MRE
0.9897	8.89E-06	8.88889E-08	0.000298142	0.00010728	0.000319848	3.19848E-06

search approach as follows: the values of hyperparameter are modeled with an exponential probability density function, which produces values for each hyperparameter to be investigated according to the models' performance in the validation set. After 60 iterations with differently produced values from the probability functions, the hyperparameters that generated the highest accuracy are selected and stored. The optimal hyperparameters are chosen based on the median of the hyperparameters' values calculated for each fold. Finally, the 5-fold cross-validation is repeated to guarantee that all test instances are investigated with the optimal hyperparameters, which leads to more stable practical and results.

The optimized type and value of the GPR hyper-parameters produced by the regression learner app are presented in Table 7.

Fig. 10 depicts the 5-fold CV results produced by the GPR model. Referring to Fig. 10, it can be observed that there is a very high accuracy between the GPR prediction outcomes and the real mode of the TBM-PR with negligible error values. Table 8 provides the results for all the other evaluation criteria. All of these findings point to the high capability of the GPR model considered in this study for the TBM-PR prediction in tunneling projects. Therefore, the GPR model applied in this work can be used as a good prediction model in the tunnel future levels for the TBM-PR prediction.

4.5 SVR

Vapnik (2000) modified his first version model (ε-support vector regression, SVR) by changing the ε-insensitive loss function. This modification permits the SVR model to use the margin idea in the regression process. Margin in the modified model can be described by the summation of the distances of hyperplane from the closest points of two classes. Minimizing errors between the actual training data and the hyperplane are the main target of the SVR. Kernel function idea has introduced by Vapnik (2000) for non-linear SVR. Readers are directed to Vapnik (2000) to understand more about SVR.

The regularization and portion width parameters for SVR were fine-tuned using a MATLAB-based programmed hyper-parameter improvement technique called 'fitrsvm'. On a low-through moderate-dimensional predictor data set, 'fitrsvm' trains a support vector machine (SVM) regression model. 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.

The key features of the SVR model, which motivate us to use it, is as follows:

One of the main advantages of SVR is that its computational complexity does not depend on the dimensionality of the input space. It performs lower computation compared to other regression techniques. Additionally, it has excellent generalization capability, with high prediction accuracy and is robust to outliers.

Table 9 For TBM-PR prediction, the optimal model form and hyper-parameter values of SVR were determined

Parameter	Value or type
Solver	'SMO'
Epsilon	0.0111
Kernel Function	'Quadratic'
Bias	1.9627

The regression learner app of MATLAB 2018 software was applied in this article to make SVR forecasts on the TBM-PR. Linear, quadratic, cubic, fine gaussian, medium gaussian, and coarse gaussian SVR models are available in this app. All of these models made TBM-PR predictions, and the best one was chosen. When selecting an optimal mode for the SVR hyper-parameters in MATLAB 2018, the optimization mode was also taken into account. The optimal model form and hyper-parameter values for the SVR system are shown in Table 9.

According to the SVR prediction results obtained from the 5-fold CV as in Fig. 11, the accurate results are produced by the SVR model comparing with the actual TBM-PR along the Alborz service tunnel path. Also, as shown in Fig. 11, the non-considerable errors are obtained by the SVR model. According to the R^2 value and the results for all other evaluation criteria presented in Table 10, and it could be introduced the SVR model considered in this study as a good ML method for the TBM-PR prediction in the tunneling projects.

4.6 DT

The DT is a classification and regression approach that is based on the technique of non-parametric survived learning. It also includes a set of if-then-else decision rules. The model's finest perdition occurs when the DT digs deeper and deeper to find the best fit with the actual data. The DT has a number of advantages. First, there is no need to make assumptions about the distribution of explanatory variables. Second, substantial relationships between independent factors have no effect on DT outcomes. Third, DT can encompass a wide range of dependent variables, including survived data, category data, and numerical data. Fourth, this strategy includes the powerful factors while excluding the least powerful variables that characterize the dependent variable. The DT can predict both small and large datasets well, despite the fact that it was originally designed to predict only large datasets well.

The key features of the DT model, which motivate us to use it, is as follows:

DTs are a type of supervision learning algorithm that repeatedly splits the sample based on certain questions about the sample. These are very useful for prediction problems. They are relatively easy to understand and very effective. DTs represent several decisions followed by

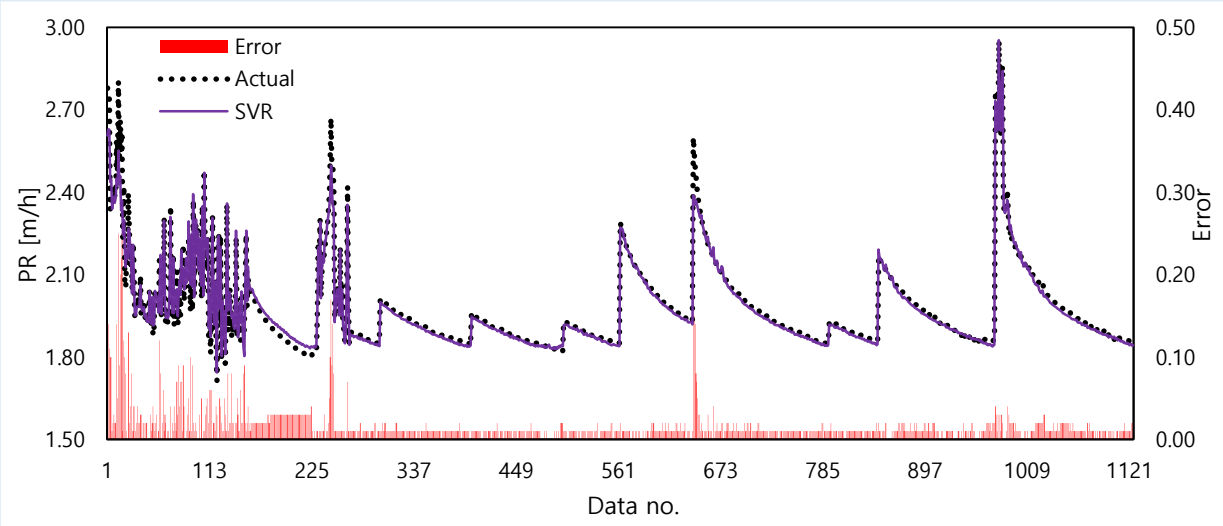

Fig. 11 TBM-PR prediction results of the SVR model

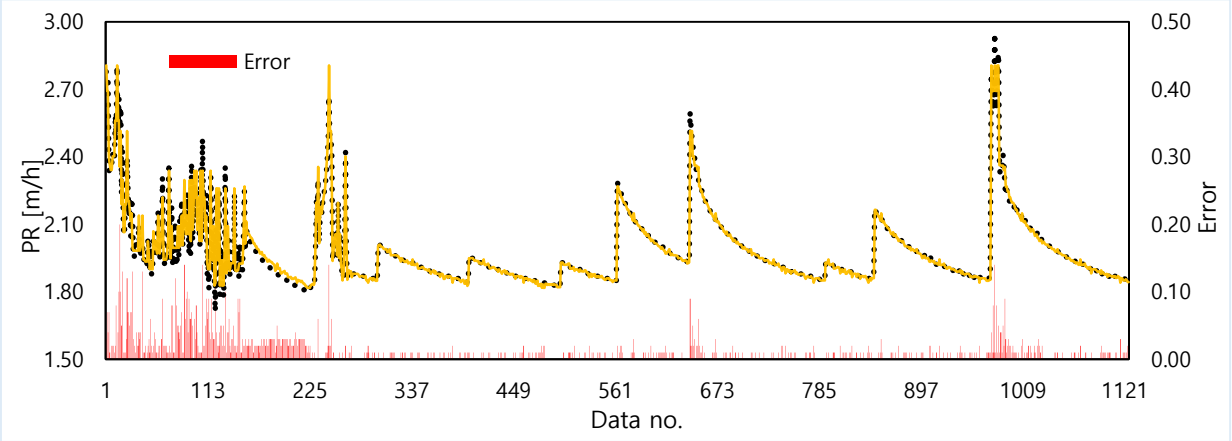

Fig. 12 TBM-PR prediction results of the DT model

different chances of occurrence. This technique helps us to define the most significant variables and the relation between two or more variables. In our problem, we have various variables, which are related to each other, so we select DT as one of comparing models. In other words, a significant advantage of a decision tree is that it forces the consideration of all possible outcomes of a decision and traces each path to a conclusion. It creates a comprehensive analysis of the consequences along each branch and identifies decision nodes that need.

Key advantages:

- No preprocessing needed on data.
- No assumptions on the distribution of data.
- Handles collinearity efficiently.
- DT can provide an understandable explanation over the prediction.

The DT algorithm is explained as follows:

1. First, the calculated, targeted variance is computed.

2. The database is divided into multiple portions based on the various features, and the variance of each sectioned component is subtracted from the variance previous to the division. This is known as variance reduction. Node N can be defined by the variance reduction as

$$I_V(N) = \frac{1}{|S|^2}\sum_{i\in S}\sum_{j\in S}\frac{1}{2}(x_i - x_j)^2 - \left(\frac{1}{|S_t|^2}\sum_{i\in S_t}\sum_{j\in S_t}\frac{1}{2}(x_i - x_j)^2 + \frac{1}{|S_f|^2}\sum_{i\in S_f}\sum_{j\in S_f}\frac{1}{2}(x_i - x_j)^2\right) \quad (21)$$

S is a group of samples that is not separated yet, S_t is a group of separated samples with true result and S_f is a group of separated samples with a false result. Without referring to the mean, each of the summands presented above is indeed variance estimates though written in a form. In each summation term in Eq. 21, variance estimation is required in such a way the mean is not referred to directly.

The attribute's node of choice is determined by the greatest VR.

3. The datasets are divided based on the values of selected attributes. If a part's variance is greater than zero, it is separated a second time.

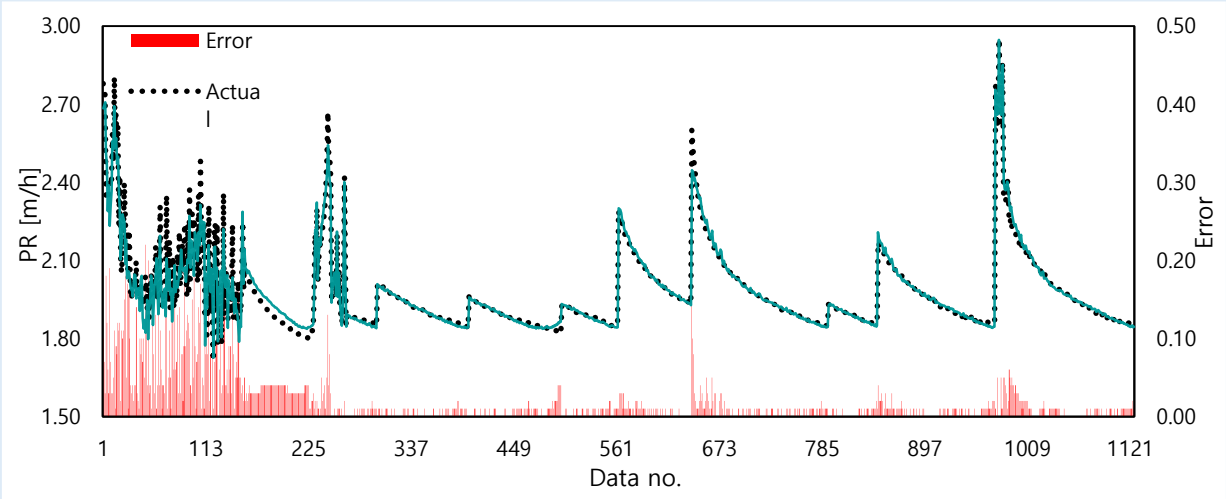

Fig. 13 TBM-PR prediction results of the LR model

Table 10 Assessment criteria outcomes of the SVR model

R^2	MAE	MSE	RMSE	RRMSE	MAPE [%]	MRE
0.9729	1.51E-04	2.56889E-05	0.005068421	0.001823763	0.00543741	5.43741E-05

Table 11 For TBM-PR prediction, the optimal model form and hyper-parameter values of DT were determined

Parameter	Value or type
"PredictorSelection"	"allsplits"
"MaxNumSplits"	1124
"Prune"	"on"
"MinLeafSize"	4
"SplitCriterion"	"mse"
"MinParentSize"	10

4. Continue the trial until all of the data has been analyzed. In the DT approach, three model, such as medium, coarse, and fine, are embedded in the MATLAB 2018. The FOS predictions were performed through these three models and eventually considered the model that provided more precise results. The information about the optimized DT's hyper-parameter parameters considered in this analysis is provided in Table 11.

Fig. 12 shows the 5-fold CV prediction outcomes obtained by the DT model, and these results are also compared with the monitored TBM-PR of the Alborz service tunnel path. To more analyzing the DT model results, the other evaluation criteria results are evaluated as in Table 12. All the results show the capability of the DT model on the TBM-PR forecast in tunneling projects. In this regard, taking into account the database utilized in this article, the presented DT model can be applied for the other future tunneling projects to obtain accurate predictions on the TBM-PR.

4.7 LR

The linear regression (LR) method is used to model the relationship between a scalar response (dependent variable) and one or more explanatory (independent variables). When there is only one explanatory component, simple linear regression is used. The method is referred to as multiple LR when more than one explanatory variable is used. This is not to be confused with multivariate LR, which predicts several correlated dependent variables rather than a single scalar variable.

Table 12 Assessment criteria outcomes of the DT model

R^2	MAE	MSE	RMSE	RRMSE	MAPE [%]	MRE
0.9782	2.67E-05	0.00000080	0.000894427	0.000321841	0.000959543	9.59543E-06

In LR, relationships are modeled by means of linear predictor functions, with unknown variables determined from the data, and the term "linear models" refers to models of this kind. The conditional mean of the result given the values of the explanatory variables (or predictors) is most frequently assumed to be an affine function of those values; the conditional median or some other quantile is assumed to be a less frequently assumed affine function of those values. As in other methods of regression analysis, LR emphasizes on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of both of these parameters that is the domain of multivariate analysis

To allow TBM-PR predictions through the LR process, the regression learner app from MATLAB Tool 2018 was used. For the LR method, the MATLAB tool includes four models: linear, interactions, robust, and stepwise. All of these models were given forecasts, and the interaction model eventually revealed the system's precise performance. Furthermore, the TBM-PR predicted by the LR model through the 5-fold CV method is shown in Fig. 13. In comparison with the actual TBM-PR, good accuracy is presented by the LR model with few errors. The other evaluation criteria results obtained as in Table 13 mention the good ability of the LR model on the TBM-PR prediction in tunneling projects and considering the database utilized in this article; it can be applied for the TBM-PR prediction of the other tunneling future levels.

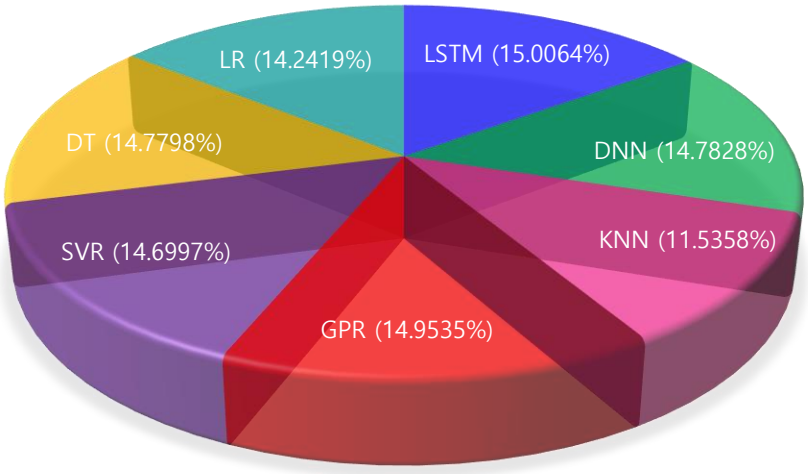

Fig. 14 The cumulative percentages of the ML methods accuracy based on the R^2 values

Table 14 A summary of the ML methods' predictive assessment criterion outcomes

	LSTM	DNN	KNN	GPR	SVR	DT	LR
R^2	0.9932	0.9784	0.7635	0.9897	0.9729	0.9782	0.9426
MAE	2.68E-06	4.25E-05	5.60E-05	8.89E-06	1.51E-04	2.67E-05	8.89E-05
MSE	8.06E-09	2.03E-06	3.53E-06	8.89E-08	2.57E-05	0.0000008	8.89E-06
RMSE	8.98E-05	0.001425121	0.001878646	0.000298142	0.005068421	0.000894427	0.002981424
RRMSE	3.23E-05	0.000512799	0.000675991	0.00010728	0.001823763	0.000321841	0.001072802
MAPE	9.63E-05	0.001528872	0.002015415	0.000319848	0.00543741	0.000959543	0.003198477
MRE	9.63E-07	1.53E-05	2.02E-05	3.20E-06	5.44E-05	9.60E-06	3.20E-05

5. Results comparison

In Fig. 14, the cumulative percentage of the relative prediction accuracies presented by all of the ML methods is shown. As in Fig. 14, the cumulative percentages of the ML methods accuracy based on the R^2 values are very close to each other except the KNN method. In Table 14, all the evaluation criteria resulted from the seven ML technics are presented. According to Table 14, all the results presented by the LSTM model are more accurate than the other models. After the LSTM model, the GPR model has presented more accurate results. Except for the KNN model, all the other models have predicted the TBM-PR close to each other.

Providing a comparison between the TBM performance (PR) prediction results of this study and the other forgoing researches, in which the various parameters were deemed to the prediction of the TBM performance using the different ML technics, is not bad and can be very important for the other future studies in the field. To sum up this, Table 15 listed most of the previous researches down on the TBM performance prediction by the various ML models from 2000 to 2019. In the previous researches presented in Table 15, the prediction of TBM-PR and TBM advance rate (TBM-AR) are considered as the two important TBM performances. Table 15 also listed the outcomes obtained by the seven ML models utilized in this study.

The R^2 and RMSE indicate the comparisons of all models' results. In Table 15, by Torabi *et al.* (2013) achieved the most reliable prediction results using the ANN method by applying four parameters of UCS, cohesion (C), friction angle (φ), and Poisson's ratio (υ). However, only 31 and 8 data sets were applied in their model for the training and testing, respectively. Many of the other researches were also utilized a small data set. But it should be noted that the performance ability of the ML tool can hardly be determined using a low data set. Also, any of the researches mentioned in Table 15 not used the K-fold CV method. However, if they used the K-fold CV method, acceptable results may not be obtained since the results might not be appropriate if the testing data sets are substituted with some of the training data sets. For this purpose, in this article, 1125 data sets were used for modeling, and the K-fold (K=5) CV method was applied to analyze the ML methods' performance ability. As described in Section 4, the advantage of the K-fold CV method is that it matters less how the data is divided, the model is less prone to having selection bias, and since each observation is used for the confirmation process only once, and every observation is used for confirmation and training, this technique has been better than a repetitive random sub-sampling procedure (Zhang *et al.* 2020). Therefore, the results of this study are more confident than the other previous studies. At all, looking at Table 15, in comparison with the previous studies, a very good accuracy has been produced by the ML methods applied in this study (LSTM, DNN, GPR, SVR,

Table 15 The proposed models are compared to other soft computing models available in the literature

Reference	Inputs	Method	Output	Training	Testing	R^2	RMSE
Alvarez Grima *et al.* (2000)	RPM*, TF, D_c	ANN	PR, AR	512	128	---	1.02
	RPM, TF, D_c	ANFIS	PR, AR	512	128	---	4.96
Benardos and Kaliampakos (2004)	RQD, RMR, k, N, WTS, WZ	ANN	AR	8	8	---	1.0175
Yagiz *et al.* (2009)	DPW, α	ANN	PR	---	---	---	0.0800
	DPW, α	NLMR	PR	---	---	---	0.1150
Mikaeil *et al.* (2009)	DPW, UCS, BTS, α, PSI	FIS	PR	151	151	---	---
Gholamnejad and Tayarani (2010)	UCS, DPW, RQD	ANN	PR	120	65	0.9390	---
Yagiz and Karahan (2011)	UCS, CFF, RPM, D_c, TF	PSO	PR, AR	80%	20%	0.7370	---
Gholami *et al.* (2012)	RQD, UCS, Js, Jc	ANN	PR	---	---	0.85	2.65
	RQD, UCS, Js, Jc	LMR	PR	---	---	0.66	3.93
Oraee *et al.* (2012)	RQD, UCS, DPW	ANFIS	PR	120	57	0.69	---
Salimi and Esmaeili (2013)	UCS, BTS, PSI, SD, β	LR	PR	33	13	0.5400	0.5400
	UCS, BTS, PSI, SD, β	NLR	PR	33	13	0.7800	0.4000
	UCS, BTS, PSI, SD, β	ANN	PR	33	13	0.8300	0.3800
Shao *et al*. (2013)	DPW, α, UCS, PSI	ELM	PR	133	20	---	0.1426
	DPW, α, UCS, PSI	LS-SVM	PR	133	20	---	0.1889
	DPW, α, UCS, PSI	PLS	PR	133	20	---	0.1919
	DPW, α, UCS, PSI	GP	PR	133	20	---	0.2398
Torabi *et al*. (2013)	UCS, C, φ, υ	ANN	PR, UI	31	8	0.9990	---
	UCS, C, φ, υ	SPSS	PR, UI	31	8	0.7900	---
Mahdevari *et al.* (2014)	DPW, α, BTS, BI, UCS, TF, CT, CP, SE	SVR	PR	130	20	0.9493	---
	DPW, α, BTS, BI, UCS, TF, CT, CP, SE	MVR	PR	130	20	0.7715	---
Ghasemi *et al.* (2014)	UCS, BI, DPW, RB, α	FL	PR	121	30	0.8930	0.1300
	UCS, BI, DPW, RB, α	LMR	PR	121	30	0.7039	0.2200
	UCS, BI, DPW, RB, α	NLMR	PR	121	30	0.6952	0.3400
	UCS, BI, DPW, RB, α	PSO	PR	121	30	0.6968	0.2100
Yagiz and Karahan (2015)	DPW, α, UCS, BI	DE	PR	120	30	0.8070	---
	DPW, α, UCS, BI	HS-BFGS	PR	120	30	0.8060	---
	DPW, α, UCS, BI	GWO	PR	120	30	0.8060	---
	DPW, α, UCS, BI	PSO	PR	120	30	0.7370	---
Salimi *et al.* (2016)	RQD, RMR, Q, GSI, UCS, BTS, α, J_s	ANFIS	PR	---	---	0.8800	2.3700
	RQD, RMR, Q, GSI, UCS, BTS, α, J_s	SVR	PR	---	---	0.9200	1.3600
Khandelwal and Jahed Armaghani (2016)	Qu, rock type, UCS, D_c	ANN	PR	38	9	0.8210	---
	Qu, rock type, UCS, D_c	MR	PR	38	9	0.3440	---
	Qu, rock type, UCS, D_c	GA-ANN	PR	38	9	0.9400	---
Jahed Armaghani *et al.* (2017)	RQD, RMR, UCS, BTS, RPM, TF, WZ	PSO-ANN	PR	1028	257	0.9050	0.0340
	RQD, RMR, UCS, BTS, RPM, TF, WZ	ICA-ANN	PR	1028	257	0.9120	0.0350
	RQD, RMR, UCS, BTS, RPM, TF, WZ	ANN	PR	1028	257	0.6660	0.0710
Minh *et al.* (2017)	DPW, α, UCS, BI	FL	PR	153	153	0.7140	---
Koopialipoor *et al.* (2018)	RQD, UCS, RMR, BTS, WZ, TF, RPM	GMDH	PR	167	42	0.9240	0.1690
	RQD, UCS, RMR, BTS, WZ, TF, RPM	MR	PR	167	42	0.9220	0.1830
Fattahi (2019)	UCS, DPW, RQD	RVR	PR	148	37	0.9760	---
Zhou *et al.* (2019)	RQD, UCS, RMR, BTS, TF, RPM	ANN	AR	1028	257	0.8745	0.0472
	RQD, UCS, RMR, BTS, TF, RPM	GP	AR	1028	257	0.9162	0.0388
Xu *et al.* (2019)	UCS, BTS, RQD, WZ, TF, RPM	NN	PR	25	16	0.9240	0.1800
	UCS, BTS, RQD, WZ, TF, RPM	SVM	PR	25	16	0.9140	0.1830
	UCS, BTS, RQD, WZ, TF, RPM	KNN	PR	25	16	0.9070	0.2040
	UCS, BTS, RQD, WZ, TF, RPM	CART	PR	25	16	0.8970	0.2160
	UCS, BTS, RQD, WZ, TF, RPM	CHAID	PR	25	16	0.8500	0.2520
Jahed Armaghani *et al.* (2019)	UCS, BTS, RMR, RQD, q, WZ, TF, RPM	PSO-ANN	AR	1029	257	0.9610	0.0220
	UCS, BTS, RMR, RQD, q, WZ, TF, RPM	ICA-ANN	AR	1029	257	0.9510	0.0260
This study	UCS, BTS, PSI, DPW, α, RFC	LSTM	PR	1125	225	0.9932	8.97915E-05
	UCS, BTS, PSI, DPW, α, RFC	DNN	PR	1125	225	0.9784	1.425121E-3
	UCS, BTS, PSI, DPW, α, RFC	KNN	PR	1125	225	0.7635	1.878646E-3
	UCS, BTS, PSI, DPW, α, RFC	GPR	PR	1125	225	0.9897	2.981420E-4
	UCS, BTS, PSI, DPW, α, RFC	SVR	PR	1125	225	0.9729	5.068421E-3
	UCS, BTS, PSI, DPW, α, RFC	DT	PR	1125	225	0.9782	8.944270E-4
	UCS, BTS, PSI, DPW, α, RFC	LR	PR	1125	225	0.9426	2.981424E-3

*RPM: Revolution per minute; TF: Thrust force; D_c: Cutter diameter; ANN: Artificial neural network; ANFIS: Adaptive network-based fuzzy inference system; AR: Advance rate; PR: Penetration rate; RQD: Rock quality designation; RMR: Rock mass rating; k: Permeability; N: Overload factor; WTS: Water table surface; WZ: Weathering zone; NLMR: Non-linear multivariate regression; CFF: Core fracture frequency; J_s: Joint spacing; J_c: Joint condition; LMR: linear multivariate regression; PSO: Particle swarm optimization; DPW: Distance between the planes of weakness; α: Angle between the plane of weakness and TBM-driven direction; UCS: Uniaxial compressive strength; RB: Rock brittleness; FL: Fuzzy logic; CT: Cutterhead torque; CP: Cutterhead power; SE: Specific energy; HS-BFGS: hybrid harmony search; GWO: Gray wolf optimizer; MVR: Multi-variable regression; GSI: Geological strength index; MR: Multiple regression; GA-ANN: Hybrid genetic algorithm; Q: Q-value; SVR: Support vector regression; PSI: Peak slope index; BTS: Brazilian tensile strength; SD: Spacing of discontinuities; β: Orientation of discontinuities with respect to the tunnel axis; ELM: Extreme learning machine; LS-SVM: Least square support vector machine; PLS: Partial least square; GP: Genetic programming; C: Cohesion; φ: friction angle; υ: Poisson's ratio; UI: Utilization index; LR: Linear regression; NLR: Non-linear regression; GMDH: Group method of data handling; PSO-ANN: Particle swarm optimization- Artificial neural network; ICA-ANN: imperialist competitive algorithm-Artificial neural network; RVR: Relevance vector regression; q: Quartz content; RFC: rock fracture class. LSTM: Long-short term memory; DNN: Deep neural networks; KNN: K-nearest neighbors; GPR: Gaussian process regression; SVR: Support vector regression; Decision tree; LR: Linear regression

Table 16 First step of feature selection

	Estimate	SE	tStat	p-value	Significance code
(Intercept)	0.47385	0.04869	9.7321	1.53E-21	***
UCS	0.0030681	0.00031554	9.7236	1.65E-21	***
BTS	0.057991	0.0031682	18.304	1.15E-65	***
PSI	0.014097	0.00067378	20.922	2.76E-82	***
DPW	0.010827	0.0089835	1.2052	0.22837	
α	0.0012983	0.00024077	5.3924	8.48E-08	***
RFC	0.0018946	0.001573	1.2045	0.22867	

Table 17 The second step of feature selection

	Estimate	SE	tStat	p-value	Significance code
(Intercept)	0.48868	0.047639	10.258	1.17E-23	***
UCS	0.0030631	0.0003154	9.7117	1.84E-21	***
BTS	0.058011	0.0031669	18.318	9.15E-66	***
PSI	0.014039	0.00067209	20.889	4.32E-82	***
α	0.0013488	0.00023717	5.687	1.65E-08	***

DT, and LR). The KNN model may not be discussed as a popular method in this article due to the less accurate presentation than the other models used in this study. In comparison with the earlier studies down in the field using different ML technics, the KNN model of this article can be considered one of the popular methods for the TBM performance prediction

6. Discussion

In this work, the proposed Gaussian process regression methodology is addressed in terms of its generalization. Generalization is a principle utilized to describe the capability of a model to interact and adapt to new information. Hence, after working with data not utilized during training, a model can assimilate new input and properly forecast. The ability to generalize is the foundation for a model's success and practical performance. No generalization can be obtained for a model which was well trained in training data. Although the model can successfully predict the training data, it is considered worthless when new data are provided. A model begins to 'memorize' rather than 'learn' the training data; this is considered as overfitting.

Overfitting can be avoided by using feature selection, which minimizes the amount of features and hence the computational cost of the model. In this case, we utilize the stepwise method to select the most important set of characteristics from the total available features in the supplied dataset (Zhang *et al*. 2020).

Stepwise regression has three strategies:

1. **Forward selection** begins with no predictors in the model and iteratively adds the most useful predictors until the improvement is no longer statistically significant.
2. **Backward selection** which begins with all predictors in the model, iteratively removes the least significant predictors, and finishes when all predictors are statistically significant.
3. **Stepwise selection** is a combination of forward and reverse selections. You begin with no forecasters and gradually add the most useful predictors (like forward selection). Remove any variables that no longer improve model fit after adding each new variable (like backward selection).

In this work, we used stepAIC() [MASS package], which uses AIC to select the best model. It has a direction option that can take the following values: I “both” (for stepwise regression, both forward and backward selection); “backward” (for backward selection); and “forward” (for forward selection) (for forward selection). It returns the best possible final model. StepAIC is one of the most extensively used feature selection search methods in R. To arrive at the final collection of features, we strive to minimize the stepAIC value as much as possible. Three stars (or asterisks) represent a very significant p-value in the results shown in the tables below. As a result, a small p-value for the intercept and slope suggests that we may reject the null hypothesis, allowing us to conclude that the predictor and the target variables have a strong association. A p-value of 5% (.05) or less is considered a good cut-off point.

During the first stage, they fitted the model to include all the predictors and the target. The lowest values of p (which must be less than 0.05) are of the share of PSI, BTS, UCS, and α, respectively, as shown in Table 16. So, these four parameters are chosen, and then step 2 is taken.

The model now has only four predictors, which are PSI, BTS, UCS, and α. It can be noticed that, in Table 17, the smallest range of features are {PSI, BTS, UCS, α}, as shown in the same table that the highest impact parameter for TBM-PR is the PSI parameter.

In this paper, the effect of several different parameters on the TBM-PR was investigated. It should be noted that there are many parameters that affect the TBM-PR, which can vary depending on the geological conditions, the skill of the operator team, the type of TBM and so on. For example, under the same conditions in a tunnel, if the skill of the TBM operator team is high, they can achieve a higher PR in the tunnel excavation. Or if the type of TBM, the number of cutting discs, the arrangement of the discs and many other parameters of the TBM are different, the PR will also be different.

In this paper, the data were obtained from only one tunnel (Alborz service tunnel). In fact, the models were trained with the help of data obtained from several parts of the Alborz tunnel, and then predictions were made for other parts of the tunnel. Since the tunnels excavated using the TBMs are usually very long, the actual data obtained from the excavated sections of a tunnel can be used and the designs of the non-excavated sections of the tunnel can be updated. This means that during the tunnel excavation, the forecasts for the non-excavated parts of the tunnel can be made more accurate step by step. Now, if a person wants to use the trained model by this in another tunnel as a user, he/she must pay attention to the data considered in the model. The trained model can be used for other tunnels when the geological parameters are almost identical. Or at least in that tunnel we have access to all the parameters used in the trained model. Also, different TBMs with different capabilities are used in tunnels

If the type of TBM is different, one cannot expect a proper performance from the trained model for that tunnel. Because when the type of TBM changes, the parameters related to the TBM should also be considered in the model. For example, machine diameter, arrangement of cutting discs, type of discs, and so on. One can expect proper performance from a trained model in different types of tunnels when the data used to train the model are obtained from different types of tunnels in different geological conditions. Therefore, the more comprehensive and more complete data with a wider range is used in the trained model, the more efficient the model will be for different conditions and different tunnels. In this article, we trained the ML models to predict the TBM-PR in the Alborz service tunnel. We showed that with the help of actual data obtained during the excavation of the Alborz tunnel, the LSTM model can be trained well to predict the TBM-PR of the non-excavated parts of the tunnel.

References

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

Benardos, A.G. and Kaliampakos, D.C. (2004), "Modelling TBM performance with artificial neural networks", *Tunn. Undergr. Sp. Tech*., **19**(6), 597-605. https://doi.org/10.1016/j.tust.2004.02.128.

Fattahi, H. (2019), "Tunnel boring machine penetration rate prediction based on relevance vector regression", *Int. J. Optim. Civil Eng*., **9**(2), 343-353. https://www.sid.ir/EN/JOURNAL/ViewPaper.aspx?ID=699797.

Goh, A.T.C., Zhang, W.G., Zhang, Y.M., Xiao, Y. and Xiang, Y.Z. (2018), "Determination of earth pressure balance tunnel-related maximum surface settlement: a multivariate adaptive regression splines approach", *Bull. Eng. Geol. Environ*., 77, 489-500. https://doi.org/10.1007/s10064-016-0937-8.

Ghasemi, E., Yagiz, S. and Ataei, M. (2014), "Predicting penetration rate of hard rock tunnel boring machine using fuzzy logic", *Bull. Eng. Geol. Environ*., **73**, 23-35. https://doi.org/10.1007/s10064-013-0497-0

Gholamnejad, J. and Tayarani, N. (2010), "Application of artificial neural networks to the prediction of tunnel boring machine penetration rate", *Min. Sci. Tech*., **20**(5), 0727-0733. https://doi.org/10.1016/S1674-5264(09)60271-4.

Gholami, M., Shahriar, K., Sharifzadeh, M. and Hamidi, J.K. (2012), "A comparison of artificial neural network and multiple regression analysis in TBM performance prediction", In: ISRM Regional Symposium-7th Asian Rock Mechanics Symposium, International Society for Rock Mechanics. https://onepetro.org/ISRMARMS/proceedingsabstract/ARMS712/All-ARMS712/ISRM-ARMS7-2012-115/39877.

Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P. and Ng, A.Y. (2019), "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network", *Nature Medicine*, **25**(1), 65. https://www.nature.com/articles/s41591-018-0268-3.

Hassanpour, J., Rostami, J., Khamehchiyan, M., Bruland, A. and Tavakoli, H.R. (2010), "TBM performance analysis in pyroclastic rocks: a case history of Karaj water conveyance tunnel", *Rock Mech. Rock Eng*., **43**(4), 427-445. https://doi.org/10.1007/s00603-009-0060-2.

Jahed Armaghani, D., Mohamad, E.T., Narayanasamy, M.S., Narita, N. and Yagiz, S. (2017), "Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition", *Tunn. Undergr. Sp. Tech*., **63**, 29-43. http://dx.doi.org/10.1016/j.tust.2016.12.009.

Jahed Armaghani, D., Koopialipoor, M., Marto, A. and Yagiz, S. (2019), "Application of several optimization techniques for estimating TBM advance rate in granitic rocks", *J. Rock Mech. Geotech. Eng*., **11**, 779-789. https://doi.org/10.1016/j.jrmge.2019.01.0 02.

Kumar, J., Goomer, R. and Singh, A.K. (2018), "Long short term memory recurrent neural network (lstm-rnn) based workload forecasting model for cloud datacenters", *Procedia Comput. Sci*., **125**, 676-682. https://doi.org/10.1016/j.procs.2017.12.087.

Khandelwal, M. and Jahed Armaghani, D. (2016). "Prediction of drillability of rocks with strength properties using a hybrid GA-ANN technique", *Geotech. Geol. Eng*., **34**, 605-620. https://doi.org/10.1007/s10706-015-9970-9

Koopialipoor, M., Nikouei, S.S., Marto, A., Fahimifar, A., Jahed Armaghani, D. and Mohamad, E.T. (2018). "Predicting tunnel boring machine performance through a new model based on the group method of data handling", *Bul. Eng. Geol. Environ*., **78**, 3799-3813. https://doi.org/10.1007/s10064-018-1349-8.

Li, J., Li, P., Guo, D., Li, X. and Chen, Z. (2020), "Advanced prediction of tunnel boring machine performance based on big data", *Geosci. Front*., **12**(1), 331-338, https://doi.org/10.1016/j.gsf.2020.02.011.

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

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

http://dx.doi.org/10.12989/gae.2021.25.1.001.

Mikaeil, R., Naghadehi, M.Z. and Sereshki, F. (2009). "Multifactorial fuzzy approach to the penetrability classification of TBM in hard rock conditions", *Tunn. Undergr. Sp. Technol*., **24**(5), 500-505. https://doi.org/10.1016/j.tust.2008.12.007.

Mahdevari, S., Shahriar, K., Yagiz, S. and Shirazi, M.A. (2014), "A support vector regression model for predicting tunnel boring machine penetration rates", *Int. J. Rock Mech. Min. Sci*., **72**, 214-229. http://dx.doi.org/10.1016/j.ijrmms.2014.09.012.

Minh, V.T., Katushin, D., Antonov, M. and Veinthal, R. (2017), "Regression models and fuzzy logic prediction of TBM penetration rate", *Open Eng*., **7**, 60-68. https://doi.org/10.1515/eng-2017-0012.

Oraee, K., Khorami, M.T. and Hosseini, N. (2012), "Prediction of the penetration rate of TBM using adaptive neuro fuzzy inference system (ANFIS)", *Proceeding of SME Annual Meeting & Exhibit, From the Mine to the Market*, Now It's Global, Seattle, WA, USA. 297-302. https://doi.org/10.13140/2.1.3707.7445.

Salimi, A. and Esmaeili, M. (2013), "Utilizing of linear and non-linear prediction tools for evaluation of penetration rate of tunnel boring machine in hard rock condition", *Int. J. Min. Mineral Eng*., **4**(3), 249-264. https://doi.org/10.1504/IJMME.2013.053172.

Salimi, A., Rostami, J., Moormann, C. and Delisio, A. (2016), "Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs", *Tunn. Undergr. Sp. Tech*., **58**, 236-246 http://dx.doi.org/10.1016/j.tust.2016.05.009.

Shao, C., Li, X. and Su, H. (2013), "Performance Prediction of Hard Rock TBM Based on Extreme Learning Machine", (Eds., Lee, J., Lee, M.C., Liu, H., Ryu, J.H.) Intelligent Robotics and Applications. ICIRA 2013. Lecture Notes in Computer Science, **8103**. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40849-6_40

Seifert, C., Aamir, A., Balagopalan, A., Jain, D., Sharma, A., Grottel, S. and Gumhold, S. (2017), "Visualizations of deep neural networks in computer vision: A survey", In Transparent Data Mining for Big and Small Data, 123-144.

Vapnik, V.N. (2000), "The nature of statistical learning theory. New York", https://doi.org/10.1007/978-1-4757-3264-1.

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

Xu, H., Zhou, J., Asteris, P.G., Jahed Armaghani, D. and Tahir, M.M. (2019). "Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate", *Appl. Sci*., **9**, 3715. https://doi.org/10.3390/app9183715.

Yagiz, S., Gokceoglu, C., Sezer, E. and Iplikci, S. (2009). "Application of two non-linear prediction tools to the estimation of tunnel boring machine performance", *Eng. Appl. Artifi. Intel*., **22**(4-5), 808-814. https://doi.org/10.1016/j.engappai.2009.03.007.

Yagiz, S. and Karahan, H. (2011), "Prediction of hard rock TBM penetration rate using particle swarm optimization", *Int. J. Rock Mech. Min. Sci*., **48**(3), 427-433. https://doi.org/10.1016/j.ijrmms.2011.02.013.

Yagiz, S. and Karahan, H. (2015). "Application of various optimization techniques and comparison of their performances for predicting TBM penetration rate in rock mass", *Int. J. Rock Mech. Min. Sci*., **80**, 308-315. http://dx.doi.org/10.1016/j.ijrmms.2015.09.019.

Torabi, S.R., Shirazi, H., Hajali, H. and Monjezi, M. (2013). "Study of the influence of geotechnical parameters on the TBM performance in Tehran–Shomal highway project using ANN and SPSS", *Arabian J. Geosci*., **6**, 1215-1227. https://doi.org/10.1007/s12517-011-0415-3.

Zhou, J., Yazdani Bejarbaneh, B., Jahed Armaghani, D. and Tahir, M.M. (2020). "Forecasting of TBM advance rate in hard rock condition based on artificial neural network and genetic programming techniques", *Bull. Eng. Geol. Environ*., **79**, 2069-2084. https://doi.org/10.1007/s10064-019-01626-8.

Zhang, P., Wu, H.N., Chen, R.P. and Chan, T.H.T. (2020), "Hybrid meta-heuristic and machine learning algorithms for tunneling-induced settlement prediction: A comparative study", *Tunn. Undergr. Sp. Tech*., **99**, 103383. https://doi.org/10.1016/j.tust.2020.103383.

CC