JOURNAL OF ROCK MECHANICS

VOL. 9, NO. 4, PP. 19-32.

https://doi.org/10.22034/IRSRM.2026.532101.0655

Machine Learning-Driven Prediction of TBM Penetration Rate in Pyroclastic Formations of the Alborz Mountains, Iran: Integrating LMR, ANN, and C&R Tree Models

Ali Asghar Ghaedi Vanani[1]*, Jafar Hassanpour[2]*, Mehdi Alihkani[3], Mohamad Eslami[4].
1. Department of Geology, Faculty of Basic Sciences, Tarbiat Modares University, Tehran, Iran.
2. School of Geology, College of Science, University of Tehran, Tehran, Iran.
3. Department of Civil Engineering, Islamic Azad University, Shahrekord Branch, Iran.
4. Department of Geology, Faculty of Basic Sciences, Shiraz University, Shiraz, Iran.

Received: 02 January 2026 Accepted: 23 February 2026
(* Corresponding Author: ali.ghaedi2012@gmail.com & hassanpour@ut.ac.ir)

Keywords

Tunnel Boring Machine (TBM)
Penetration Rate
Karaj Water Conveyance Tunnel
Artificial Neural Network (ANN)
Decision Tree (C&R Tree)

Abstract

This study develops and compares three machine learning models -Multiple Linear Regression (LMR), Artificial Neural Network (ANN), and Classification and Regression Tree (C&R Tree) - for predicting the Penetration Rate (PR) of a Tunnel Boring Machine (TBM). A comprehensive dataset of 161 field samples was compiled from 22.7 km of the Karaj Water Conveyance Tunnel, which is excavated through pyroclastic rock formations of the Alborz Mountains (Iran). This dataset incorporates key intact rock properties (UCS), rock mass characteristics (RQD, GSI, joint orientation), and critical TBM operational parameters (RPM, Fn, Fr). Model performance was evaluated using the coefficient of determination (R^2) and mean absolute error (MAE). The ANN model was found to achieve the highest predictive accuracy ($R^2 = 0.93$, MAE = 0.02), followed by the C&R Tree model ($R^2 = 0.8$ ، MAE = 0.05), while the LMR model demonstrated lower performance ($R^2 = 0.76$, MAE = 0.07). Notably, the C&R Tree model offered superior interpretability, with the Geological Strength Index (GSI) identified as the most influential parameter. Furthermore, this model generated explicit decision rules, thereby elucidating the specific geological conditions associated with the minimum (2.38 m/h) and maximum (4.47 m/h) penetration rates observed in the terminal nodes. In summary, while the ANN model provides superior numerical precision, the C&R Tree model delivers interpretable insights that effectively bridge quantitative prediction and qualitative engineering analysis. The synergistic application of both approaches is thus proposed as a robust framework for optimizing TBM operations and enhancing tunneling productivity.

I. INTRODUCTION

The analysis of Tunnel Boring Machine (TBM) performance is crucial for developing accurate evaluation models, given the unique characteristics of the machines, the diverse lithology encountered, and the geological structures present in each project [1]. Moreover, utilizing and comparing existing models provides insight into their respective strengths and limitations, aiding researchers in developing new models based on novel methodologies [2], [1]. Machine performance prediction models have been categorized as theoretical, empirical, statistical, and computational [3]. Theoretical models were devised to compute cutter forces based on intact rock properties, cutter specifications, and cutting geometry, all regarding a given penetration depth [4]–[16].

Several studies used different statistical techniques, such as simple bivariate regression, to predict FPI (kN/cutter/mm/rev) [17], [2], [3], [18], PR (m/h), and advanced rate (AR, m/day) [19]. By using linear multiple regression, they developed models to predict PR [20]–[22]. Also, a non-linear multiple regression has been used to predict FPI [17], [2], [1], [25], [18], P (mm/rev) [23], PR [24], PR, and AR [19]. However, the coefficient of determination for these models is reported to be less than 0.7. The accuracy of statistical models depends on various types of parameters that should be collected by field observations (e.g., the QTBM model), special laboratory tests in some models (e.g., the NTNU models), and different geological and lithological conditions based on which the database is developed (e.g., Hassanpour et al. models). Therefore, data outside the database's limits, inaccuracy in field observations, and lack of specific laboratory tests will lead to a decrease in the model's prediction accuracy [25], [19].

In recent years, new computational models, such as Artificial Neural Networks (ANNs), have acquired popularity for TBM performance prediction. For the prediction of PR, various studies used the ANN models [26]–[36], the ANN and support vector machine

(SVM) [22], [37], [38], tree-based regression models [39]–[41], and XGBoost models [42], [43]. ANNs are machine learning algorithms that can capture complex relationships and patterns in large datasets. By training on TBM performance data, ANNs can learn the nonlinear interactions between input parameters and performance indicators, allowing for more accurate predictions ($R^2 \geq 0.9$) (Table 1).

While ANNs often achieve high predictive accuracy, their "black-box" nature can limit practical interpretability for engineering decision-making. To address this need for transparent models, recent studies have successfully employed interpretable, tree-based machine learning algorithms for TBM performance prediction. For instance, Random Forest (RF) and Gradient Boosting models have been used to predict penetration rate (PR) and field penetration index (FPI) with high accuracy ($R^2 > 0.89$), while simultaneously providing clear rankings of variable importance [51], [52], [53]. More directly, Classification and Regression Tree (C&R Tree) models have been applied to identify key geological controls on TBM utilization and to generate actionable "if-then" decision rules that link specific rock mass conditions to expected performance outcomes [54], [55] (Table 2). These studies underscore the value of tree-based models in bridging the gap between complex data-driven prediction and interpretable engineering insight.

A review of recent domestic research reveals an evolutionary trajectory in the development of TBM penetration rate prediction models within the Iranian context. This evolution has progressed from employing basic time-series models [56] to conducting systematic comparisons of diverse methodologies [57], culminating in the adoption of sophisticated hybrid algorithms. Contemporary studies exhibit a pronounced focus on optimizing neural networks through metaheuristic algorithms [58], utilizing adaptive neuro-fuzzy inference systems (ANFIS) [59], and developing comprehensive multi-stage feature selection frameworks [60]. This trajectory underscores the emphasis of Iranian researchers on enhancing model complexity to achieve greater predictive accuracy, while concurrently addressing a wide spectrum of ground conditions—from hard rock to soft soil—and different TBM systems, including open-mode and Earth Pressure Balance (EPB) machines. Collectively, these studies establish a robust foundation for the advancement of indigenous predictive models that are specifically tailored to the unique geological and operational challenges encountered in Iranian tunneling projects.

This study employs the C&R Tree algorithm alongside LMR and ANN with a dual objective: to compare their predictive performance and to extract explicit, interpretable rules that elucidate the interaction between geological conditions and TBM operational efficiency in pyroclastic formations. This analysis is based on a comprehensive database compiled from 22.7 km of the Karaj Water Conveyance Tunnel (Lots 1 & 2), incorporating actual TBM operational data (e.g., RPM, Fn, Fr) as well as intact rock and rock mass characteristics (e.g., UCS, RQD, GSI, joint orientation and spacing).

Table 1. A summary of TBM performance prediction models using ANN techniques.

Source	Model	Input parameters	Output	Comment
Benardos and Kaliampakos (2004) [20]	ANN	N, RQD, UCS, RMR, overburden, permeability	AR	
Yagiz and Ozdemir (2009) [21]	ANN	UCS, σ_t, BTS, CAI	PR	Using 151 datasets; $R^2 = 0.9$
Gholami and Ghaffari (2012) [22]	ANN	UCS, RQD, Js, Jc	PR	Data of 121 tunnel sections; $R^2 = 0.72$
Salimi and Esmaeili (2013) [23]	ANN	PSI, UCS, BTS, DPW, α	PR	Data of 46 sections of the Karaj–Tehran tunnel; $R^2 = 0.83$
Salimi et al. (2016) [18]	ANFIS	UCS, BTS, RQD	FPI	Number of 75 data; $R^2 = 0.91$
Armaghani et al. (2017) [26]	PSO-ANN	UCS, BTS, RMR	PR	Using 1286 data samples in hard rock condition; $R^2 = 0.95$
Koopialipoor et al. (2019) [27]	ANN, DNN	UCS, RQD, BTS, RMR, WZ	PR	Using 1286 datasets; $R^2 = 0.93$
Zhou et al. (2021) [31]	ANN, GP	UCS, RQD, BTS, RMR, TFC, RPM	AR	Using 1286 datasets; $R^2 = 0.92$

Table 2. A summary of TBM performance prediction models using tree-based machine learning algorithms.

Source	Model	Input parameters	Output	Comment
Zhang et al. (2022), [51]	Random Forest (RF)	UCS, BTS, RQD, RMR, joint condition, depth, RPM, cutterhead torque	PR	RF outperformed ANN and SVM; $R^2 \approx 0.89$; highlighted feature importance interpretation.
Zhou et al. (2022), [52]	Gradient Boosting (XGBoost)	UCS, RQD, GSI, joint spacing, orientation, RPM, thrust	PR & AR	Superior accuracy ($R^2 > 0.94$) and provided gain-based feature importance.
Chen et al. (2021), [53]	C&R Tree	UCS, RQD, GSI, joint condition, groundwater, overburden	TBM Utilization (%)	Generated actionable decision rules to identify bottlenecks and geological risks.
Wang et al. (2020), [54]	Random Forest (RF)	Rock type index, quartz content, CAI, joint spacing, thrust per cutter	FPI	Emphasized model interpretability via variable importance and partial dependence plots for hard rock TBM.
Mikaeil et al. (2019), [55]	Decision Tree (C5.0) & AdaBoost	Rock brittleness indices, tensile strength, drillability indices	ROP (Rate of Penetration)	Compared single tree with ensemble; showed how boosting improves accuracy while maintaining rule traceability.

II. TBM MODELING METHODOLOGIES

This study uses LMR, ANN, and C&R Tree techniques to develop models for prediction PR (m/h) based on actual data of TBM operation and ground characters based on the 14 km of Lot 2 and 8.7 km of Lot 1 from the KWCT project (Fig. 1).

The LINEAR model: The LINEAR approach in SPSS automates linear regression analysis, including data preparation, outlier detection using Cook's Distance, and best subset selection [44]. Missing values are replaced, categorical predictors specified, and influential outliers identified. The best subset method evaluates all possible combinations of predictors and selects the model with the highest R^2. This ensures that the LMR models are robust and optimized for predicting TBM performance indicators such as Penetration Rate (PR). Multiple iterations of outlier removal and subset selection are performed to refine the final model [18].

Artificial Neural Networks (ANN, MLP): ANNs are multilayer feed-forward networks designed to capture nonlinear patterns in operational and geological data [45], [26], [28]. Input nodes receive TBM operational parameters and rock mass properties, which are transformed through hidden layers via weighted connections. Output nodes predict TBM performance indicators. The Multi-Layer Perceptron (MLP) architecture allows iterative training to optimize weights and biases, effectively modeling complex relationships without requiring a predefined mathematical correlation between inputs and outputs.

Classification and Regression Tree (C&R Tree): The C&R Tree algorithm is a non-parametric, rule-based machine learning technique used for both classification and regression tasks. It operates by recursively partitioning the dataset into homogeneous subgroups based on the most significant input parameters, with the goal of minimizing variance within each resulting node [50]. The splitting process employs criteria such as the Gini index or mean squared error reduction to select optimal thresholds for continuous predictors. This results in a transparent, tree-like structure of decision rules that explicitly illustrate the interaction between geological and operational factors influencing TBM penetration. The model was developed using IBM SPSS Modeler (v26), with pruning applied to prevent overfitting and enhance generalization to unseen data.

Evaluation of Models' Performance: Each model's performance is evaluated using several metrics, including coefficient of determination (R^2), mean square error (MSE), root-mean-squared error (RMSE), normalized root-mean-square error (NRMSE), mean absolute error (MAE), and absolute errors (E%).

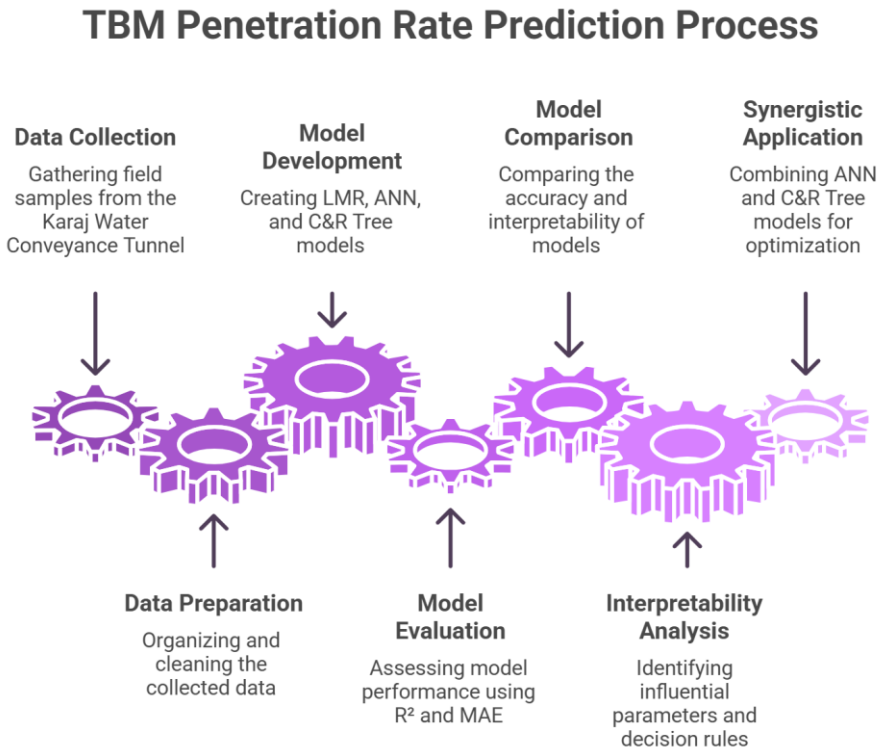

Fig. 1). Flow chart for the development of LMR, ANN, and C&R Tree models to predict PR (m/h).

III. GEOLOGY AND PROJECT DESCRIPTION

The Middle Alborz Mountains are located in the northern part of Iran, running in an east-west direction. Folded and thrust structures are present in this mountain range, which consists of a sequence of sedimentary rocks, such as conglomerate, sandstone, and siltstone, along with pyroclastic (volcanic) rocks [46], [47]. The central Alborz basins have an ongoing tectonic deformation rate of 2–5 mm per year, according to studies that involved GPS geodetic measurements and morphotectonics [48]. This active tectonic setting has resulted in the development of various fractures, faults, and joints within the pyroclastic-sedimentary rock formations.

The Karaj Water Conveyance Tunnel (KWCT) was built to transfer 16 m³/s of water from the Karaj Dam to the Tehran metropolitan area. The tunnel is about 30 kilometers long and has a diameter of 4.65 meters. Lot 1 (ET-K0) is situated at the southeast end and has a length of 15.7 kilometers, while Lot 2, also known as K0-P, is 14 kilometers long (Fig. 2) [2], [1]. The tunnel was excavated using a tunnel boring machine (TBM) with a double shield, manufactured by Herrenknecht. The cutter head of the TBM was equipped with disc cutters that had a diameter of 17 inches (432 mm) and a width of 31 inches (432 mm). They could withstand a load capacity of 250 kN [49].

IV. PREPARING OF DATABASE

Laboratory tests were conducted on various features of the intact rock, including petrographic characteristics, physical properties, UCS (MPa), DRI, and CLI. In addition to pre-construction testing, samples were collected and analyzed during the tunnel construction phase. For each tunnel section (geological unit), the mean values of intact rock properties were calculated and reported (Table 3). Table 3 indicates that the highest UCS values, ranging from 100 to 130 MPa, occur in gabbro and Monzogabbro units, while the lowest UCS values, between 30 and 50 MPa, are found in siltstone and shale units. Similarly, the lowest CLI values (10–15) are observed in gabbro and Monzogabbro units, whereas the highest CLI values (40–45) are present in GT units.

We incorporated the database related to the first 8.7 kilometers of the KWCT1 project by Hassanpour et al. (2010), along with data from the KWCT2 project. These 189 data from 61 tunnel units are similar to engineering, geological, and lithological characteristics (Table 4). This combined dataset was used to develop new models using both LMR and ANN techniques. Table 4 shows the descriptive statistics related to TBM performance parameters, intact rock (UCS, MPa), and rock mass parameters recorded, e.g., RQD%, GSI, joint orientation to tunnel axis (α), spacing of joints (Sj, m), and the normalized Q-value ($Qc = Q \times \sigma_c/100$), in the database of the KWCT project. The evidence shows that the minimum and maximum PR ranges from 1.38 to 4.95 (m/h), P is equal to 4.4 to 12.02 (mm/rev), and RPM has values from the lower limit of 3.04 to the upper limit of 8.22 (Table 4).

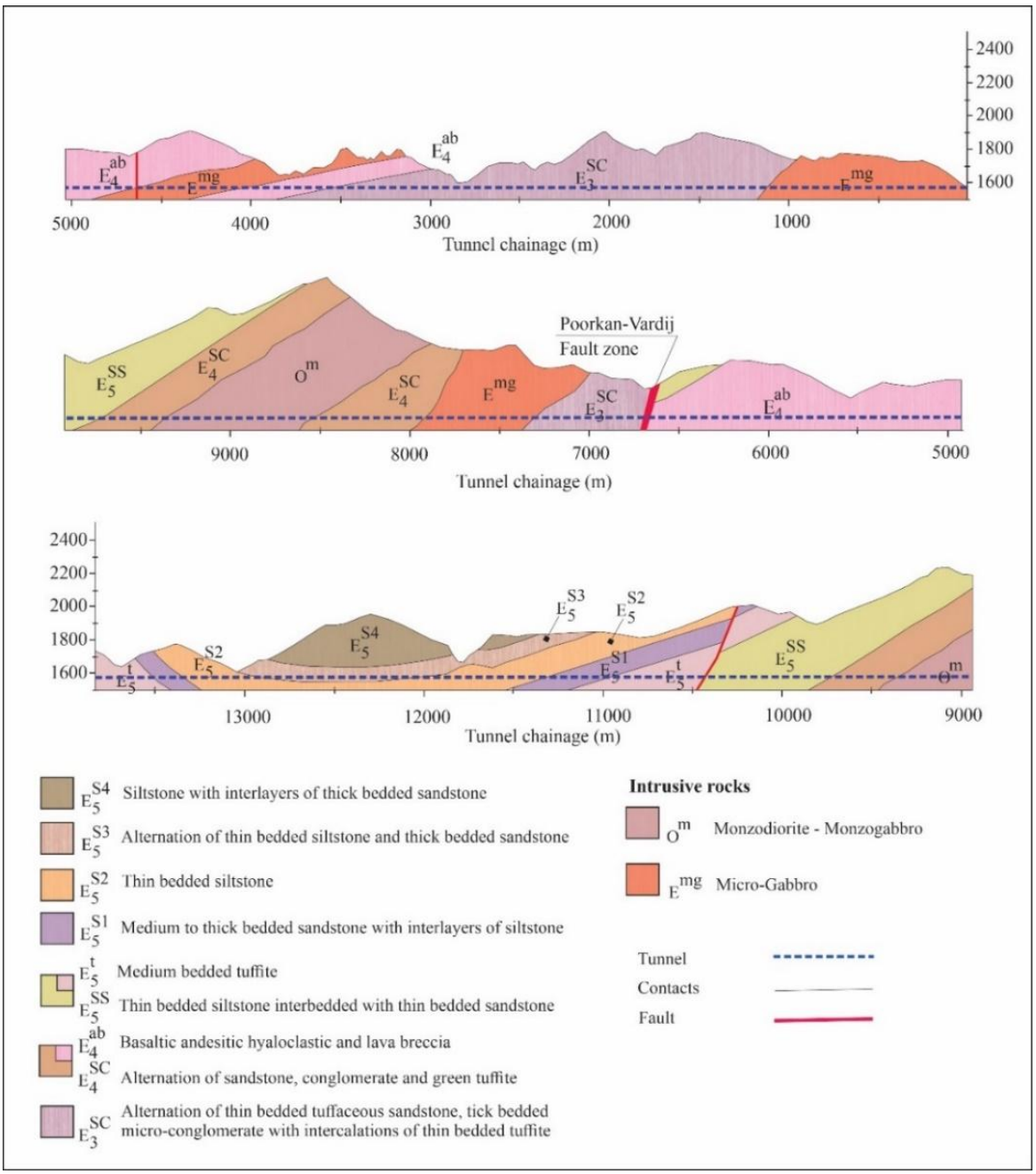

Fig. 2: Engineering geological cross section along the tunnel (KWCT2)

.Table 3. Main characteristics of lithological in the database of KWCT project.

Lithotype	Quartz content (%)	Porosity (%)	UCS (Mpa)	Drilling Rate Index (DRI)	Cutter Life Index (CLI)
Vitric tuff- green crystal tuff (GT)	5	1-10	30-100	60-65	40-45
Lithic Tuff (tf)	15	3-6	50-80	55-60	25-30
Sandstone, micro conglomerate (sc)	20	1-5	40-80	60-65	25-30
Siltstone, shale (ss)	10	1-5	30-50	40-45	35-40
Gabbro, Monzo- gabbro (Gb)	0-5	>5	100-130	55-60	10-15

Table 4. Descriptive statistics of the TBM performance, intact rock, rock mass parameters in the KWCT

Parameter	Range	Minimum	Maximum	Mean	Std. Deviation
PR (m/h)	3.37	1.38	4.۹5	3.09	0.84
P (mm/rev)	11.62	4.4	12.02	7.93	2.02
FPI (kN/Cutter/mm/rev)	22.46	6.45	28.9	19.52	5.41
RPM	5.18	3.04	8.22	6.55	0.79
Fn (kN/Cutter)	123.38	76.31	199.68	148.31	31.82
Fr (kN)	13.8	11.1	24.9	15.81	2.63
UCS (MPa)	110	40	150	84.08	25.04
***α* (degree)**	38.22	31.78	70	49.93	9.82
Sj (m)	1.5	0.01	1.51	0.54	0.36
RQD %	80	15	95	69.36	17.85
GSI	66.25	18.75	85	50.77	17.57
Qc	63.82	0.18	64	7.54	11.60

V. DEVELOPMENT MODELS

We developed the LMR and ANN models to predict the PR (m/h) using 161 data from 61 geologic units (22.7 km) of the KWCT1 and KWCT2 projects. In order to identify the most significant independent parameters and use them as useful inputs to LMR and ANN models, a simple regression has been established between the predictors (geological parameters and TBM operation data) and the dependent variable (PR) of the models (Fig. 3).

Figure 3 depicts the direct relationship between PR (m/h) and TBM performance data (RPM, Fr, and Fn) as well as the discontinuity angle, GSI, and Qc parameters. Also, the PR shows a linear inverse correlation with the parameters of UCS and RQD%.

Figures 4a depict the structure of the LMR model used to predict PR (m/h), as well as predictor importance (PI) values. In the PR model, the RPM (PI = 0.32), GSI (PI = 0.26), Fr (PI = 0.19), and Fn (PI = 0.04) have the highest direct correlation. The PR values decrease as RQD and UCS increase, with importance values of 0.11 and 0.08, respectively (Fig. 4a).

Figure 4b compares the values predicted by the developed LMR model to the actual values of PR, as well as the percentage of absolute error (E %). The maximum E% in the LMR model of PR is equal to 30%. The comparison of the observed and predicted values of PR confirms the validity of the LMR models, with R^2 values of 0.76 (Figs. 4a-b).

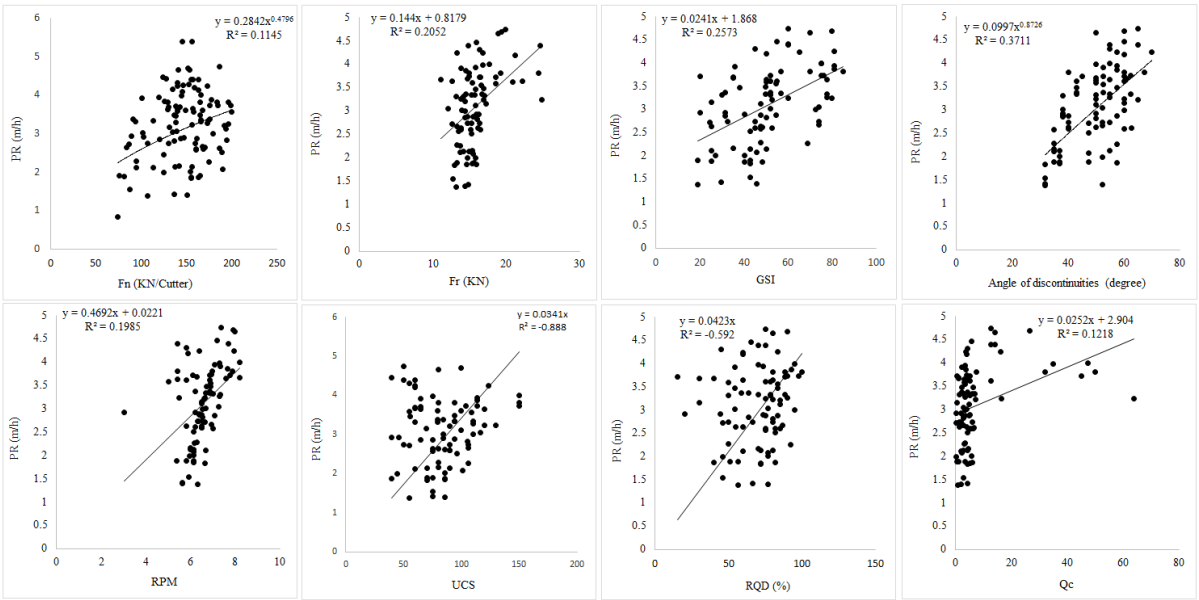

Fig. 3: the correlation between PR (m/h) with TBM operating parameters (RPM, Fr, and Fn) and ground parameters (RQD, GSI, UCS, Qc, and angle of discontinuities)

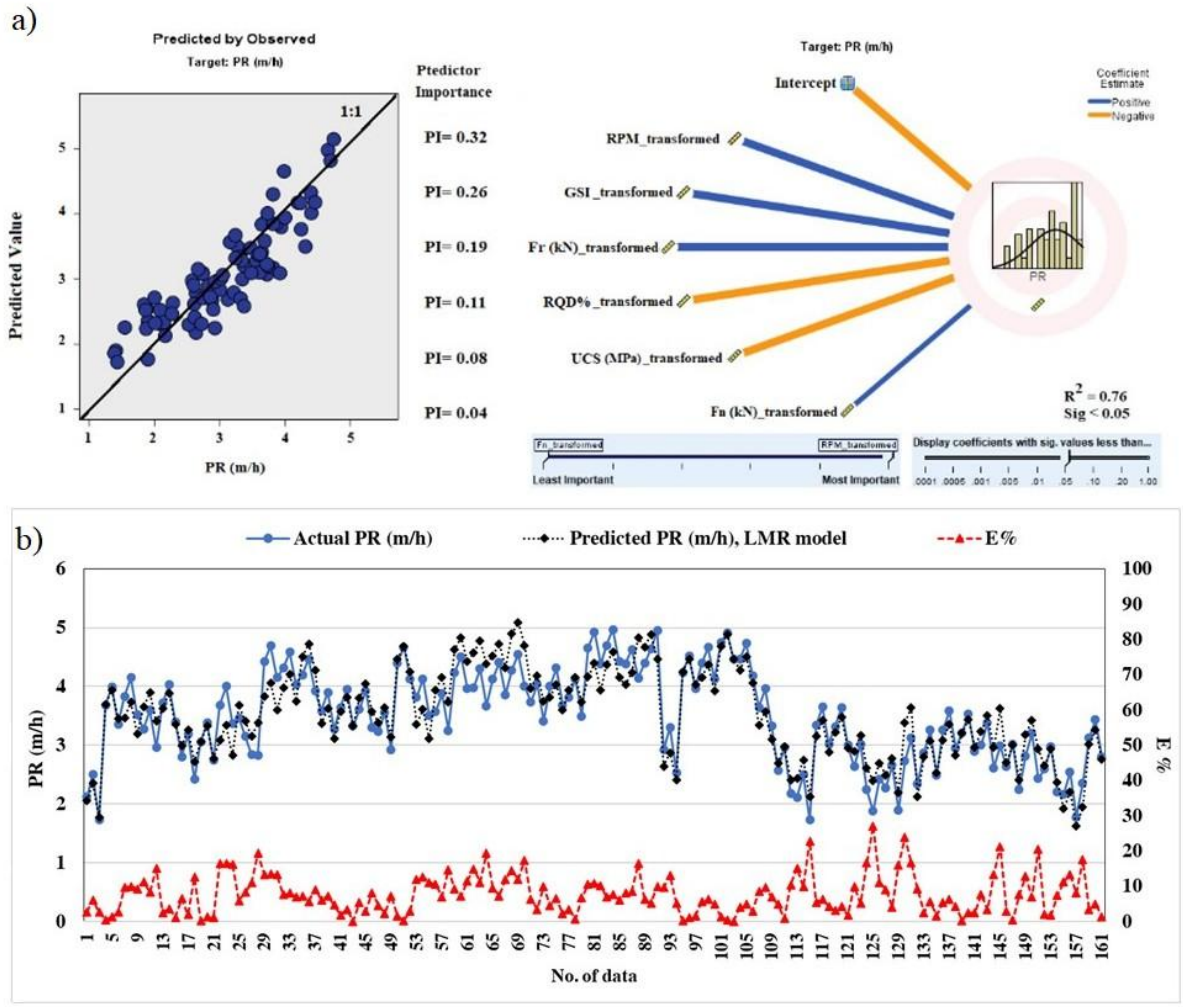

Fig. 4: a) the structure of LMR model of PR (m/h), and values of the Predictor Importance (PI). b) The comparison between the predicted values and the actual values of PR, along with the percentage of absolute error (E %).

Table 5 shows ANOVA analysis, standardized coefficients, t statistic, and p-value (or Sig.) corresponding to the developed equation of the PR (m/h) in table 5. Considering the significance level (Sig < 0.05), the equations for all three models are statistically valid with 95% confidence (Table 6).

Table 7 presents a summary of LMR models, showcasing input data, R^2 values, and errors derived from loss functions. The PR model is in prediction accuracy, with $R^2 = 0.80$, RMSE = 0.3281, and MAE = 0.0305. Also, the developed PR model shows lower error values (RMSE = 0.3281 and MAE = 0.0787) than the common models (Table 7).

Table 5. Summary of the LMR models for PR (m/h).

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
PR (m/h)	(Constant)	-2.687	0.574		-4.679	0.000
	GSI	0.037	0.005	0.77	7.889	0.000
	Fr (kN)	0.129	0.022	0.405	5.837	0.000
	RPM	0.575	0.072	0.546	7.961	0.000
	RQD%	-0.03	0.005	-0.634	-5.528	0.000
	UCS (kPa)	-0.013	0.003	-0.375	-4.2	0.000
	Fn (kN/Cutter)	0.008	0.002	0.316	3.495	0.001

Table 6. The equations developed for predicting PR (m/h).

Model	Eq.	Adjusted R Square	Sig.
PR (m/h)	$PR = 0.575RPM + 0.129Fr + 0.037GSI + 0.008Fn - 0.03RQD - 0.013UCS - 2.687$	0.76	<0.001

We employed 161 data points for the ANN models, with 70% for training (113 data points), 15% for validation (24 data points), and 15% for testing (24 data points). The model had one hidden layer, and the number of neurons was approximately half of the input parameters.

ANN architecture: A single hidden layer MLP was selected based on the dataset size and complexity. The number of neurons was set to approximately half the input parameters (a common heuristic to balance capacity and generalization), and the hyperbolic tangent activation function was used for its smooth gradient properties. This architecture is sufficiently flexible to capture nonlinear interactions without excessive parameters that could lead to overfitting.

Overfitting prevention measures: We employed: (1) Early stopping based on validation error, (2) L2 regularization (weight decay) to penalize large weights, and (3) cross validation during training to monitor generalization.

The final architecture (one hidden layer with neurons ≈ half the input parameters) was chosen to balance model capacity and generalization, as evidenced by consistent performance across training, validation, and test subsets. The close agreement between training, validation, and test performance ($R^2 \approx 0.93$) indicates that overfitting was effectively controlled

Table 7 summarizes the developed ANN model of PR, including the number of data points used for training, validation, and testing; input parameters; R^2 values; and errors obtained from the loss functions for each model. In table 7, the R^2 values for the training data are higher in all three models compared to the test and validation data, and the error values for the training data are lower. The PR model (ANN) provides $R^2 = 0.9199$ to 0.9373 for test and validation data, with MSE values between 0.2281 and 0.2088 and MAE values between 0.0477 and 0.0243. The highest R^2 values for the PR models are 0.9358 with RMSE equal to 0.3268, which shows slightly more accuracy in the PR model (Table 7).

Table 7. A summary of LMR and ANNs models, showcasing input data, R^2 values, and errors derived from loss functions

model	Number of data		Input data	R^2	MSE	RMSE	NRMSE	MAE
LMR-PR (m/h)	161		RPM, GSI, Fr (kN), RQD%, UCS (MPa), Fn (kN/cutter)	0.76	0.1085	0.3294	0.0931	0.0787
C&R Tree	Training	107	RPM, Fr (kN), GSI, RQD%, UCS (Mpa), α (degree), Qc, Fn(kN/cutter)	0.932	0.2081	0.4560	0.0943	0.0243
	Testing	54		0.801	0.1023	0.3476	0.0947	0.0511
ANN - PR (m/h)	Training	113		0.935	0.0952	0.3085	0.0191	0.0236
	Testing	24		0.919	0.2281	0.4775	0.0945	0.0477
	Validation	24		0.937	0.2088	0.4569	0.0943	0.0243
	All	161		0.935	0.1068	0.3268	0.0443	0.0241

The parameters RPM (PI = 0.24), Fr (PI = 0.22), GSI (PI = 0.13), RQD% (PI = 0.12), UCS (PI = 0.09), (PI = 0.08), Qc (PI = 0.06), and Fn (PI = 0.06) have the most significant impact on PR (m/h) (Fig. 5). Figure 5 compares the observed values with the predicted values in the ANN model for PR, with R^2 values of 0.92, based on the 1:1 line. Figure 6 shows the comparison between the values predicted by the ANN models and the actual values of PR for all the data, the validation data, and the test data. The results show that the maximum E% in the ANN models for PR is equal to 20%.

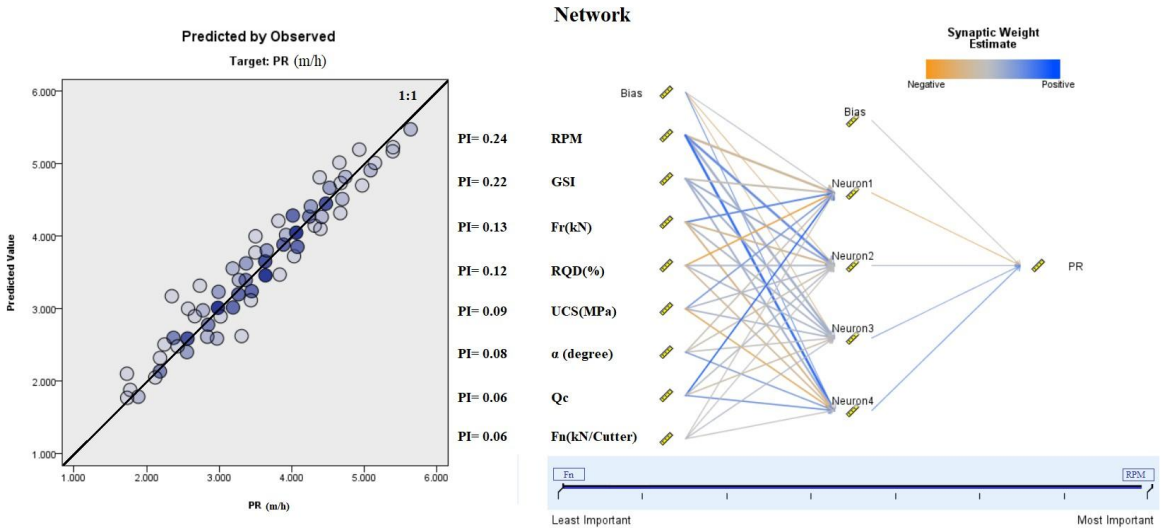

Fig. 5: the structure of ANN model of PR (m/h), and values of the Predictor Importance (PI).

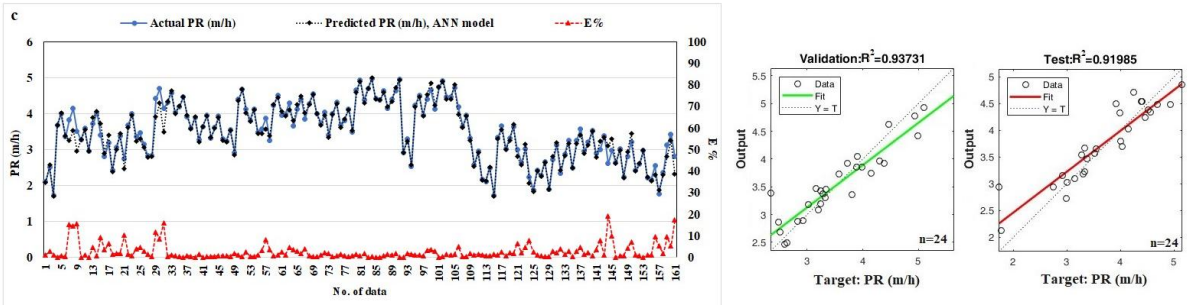

Fig. 6: Comparison between the values predicted by the ANN model and the actual values of PR for all the data, the validation data, and the test data.

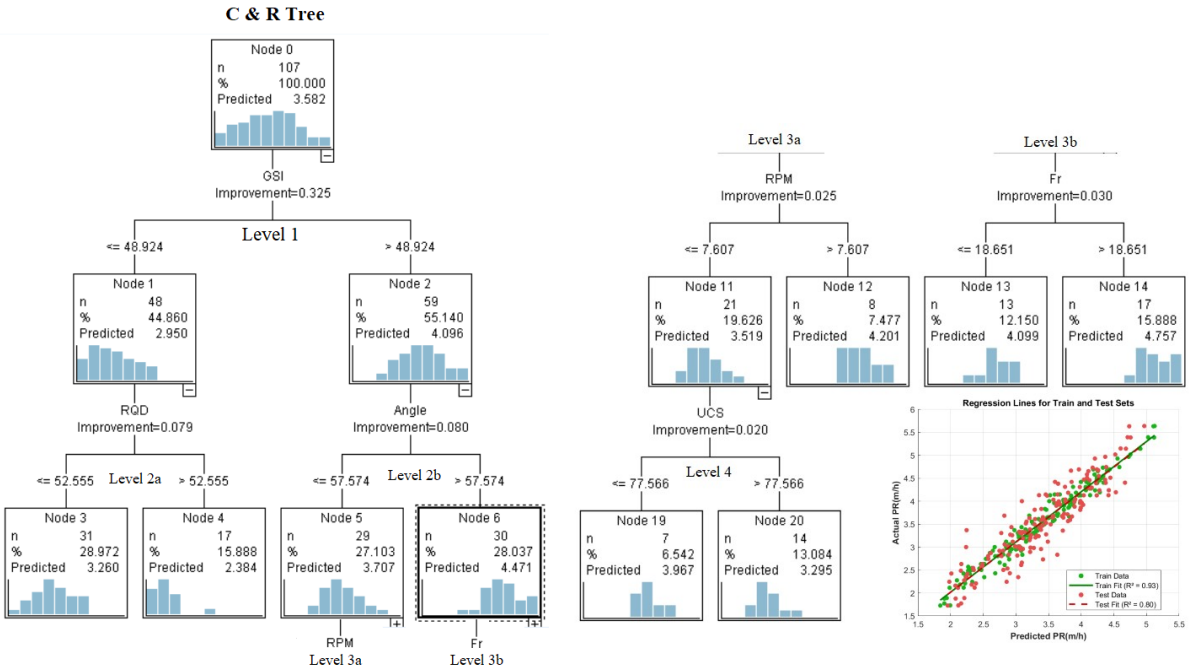

Fig. 7. The structure of Classification and Regression (C&R) Tree model of PR (m/h).

The structure and outcomes of the Classification and Regression (C&R) Tree analysis are as follows (Fig. 7):

Overall Tree Structure: The generated tree consists of a single root node (Node 0), four hierarchical levels, and a total of 20 nodes. The model was constructed using a total of 107 samples

(n=107). The predicted mean Penetration Rate (PR) for the entire dataset at the root node is 3.582 meters per hour. Primary Splitting Variable: The Geological Strength Index (GSI) created the first and most significant split in the tree, indicating its paramount role in explaining the variance in penetration rate for this project. The improvement score for this split was 0.325 (Fig. 7).

Rule 1: If `GSI ≤ 48.92`, data is directed to Node 1 (44.86% of the data). The mean PR in this node decreases to 2.950 m/h. This indicates that rock masses with lower quality (lower GSI) generally correspond to a lower average penetration rate.

Rule 2: If `GSI > 48.92`, data proceeds to Node 2 (55.14% of the data). The mean PR in this node increases to 4.096 m/h. This demonstrates the positive and substantial influence of better-quality rock (higher GSI) on drilling performance.

Secondary Splitting Variables: Within the branch corresponding to lower-quality rock (Node 1), the Rock Quality Designation (RQD) was selected as the subsequent splitting parameter (Improvement = 0.079). Conversely, within the branch for higher-quality rock (Node 2), the joint orientation (Angle) was chosen as the next splitting parameter (Improvement = 0.080). This reveals that for competent rock masses, joint orientation plays a more deterministic role in the penetration rate compared to other parameters.

Interpretation of Terminal Nodes (Leaves) and Inference Rules: The terminal nodes (leaves) of the tree, specifically Nodes 3, 4, 5, and 6, represent distinct geotechnical-operational scenarios with their corresponding average penetration rates. For instance:

Node 4 (n=17, 15.9% of data): This node corresponds to conditions where `GSI ≤ 48.92` and `RQD falls within a specific (presumably low) range`. The lowest mean penetration rate (2.384 m/h) is assigned to this node, which likely represents the most unfavorable geological conditions for tunneling in terms of overall rock mass quality. Node 6 (n=30, 28.0% of data): This node is defined by `GSI > 48.92` and `joint orientation within a specific range`. The analysis indicates that in pyroclastic rocks, a larger angle between joint sets and the tunnel axis facilitates rock fracturing, consequently yielding the highest observed mean penetration rate (4.471 m/h) in this node (Fig. 7).

Role of Operational Parameters: The provided charts (e.g., Regression Lines) also include an analysis of the rotational speed (RPM) parameter. These charts typically plot the model-predicted RPM against the actual RPM for both training and test datasets, demonstrating the model's capability to predict this key operational parameter. The Uniaxial Compressive Strength (UCS) parameter appears as a splitting variable in one of the Level 4 nodes (Nodes 19 and 20), generating specific predicted values for the ensuing branches (Fig. 7).

VI. Discussion

A number of studies [50] used the artificial neural network (ANN) method to analyze the most important parameters for predicting the TBM AR, including RQD, UCS, RMR, BTS, weathering zone, thrust force per cutter, and RPM. Other studies used it to forecast the TBM PR by various parameters, including UCS, rock type, Qc, BTS, RQD, RMR, thrust, and torque [51], [28]. Still others estimated FPI using the UCS, BTS, RMR, etc. [29], [18], [30]. Similar to the previous studies, our findings showed that the use of the ANN technique is more accurate than linear multiple regression models due to its high flexibility and lack of parameters as model inputs [31].

It should be noted that the ANN models, like general regression techniques, require sufficient and high-quality data. However, simple field observations and geological parameters (such as UCS and RQD) can be obtained without complex experiments and used in the models. The ANN and LMR models both used all of the input parameters. Because of the constraints of linear relationships, LMR models select the most suitable set of input parameters and disregard others. ANN models leverage non-linear relationships and maximize input parameters to achieve higher R^2 and minimize error in similar conditions. While ANN models allow for unrestricted input parameter selection, they cannot establish direct or inverse relationships between each independent parameter and the target parameter. To better understand how different parameters affect TBM performance, it is advisable to investigate linear models (LMR) alongside ANN models in different projects.

The C&R Tree analysis provided transparent, hierarchical rules, identifying GSI as the paramount factor influencing penetration rate (PR), followed by RQD and joint orientation. This variable importance aligns with findings from other tree-based models in hard rock TBM projects, which consistently rank composite rock mass indices as most critical [51, 52].

A key engineering insight is the model's quantification of performance extremes. It confirms that enhancing rock mass quality (GSI) offers the most significant leverage for improving PR in pyroclastic rocks. More importantly, it precisely pinpoints the highest-risk geological scenario—a combination of low GSI and low RQD—associated with the lowest PR (2.38 m/h). This capability to identify "risk nodes" for proactive mitigation is a recognized strength of decision tree models [54].

Ultimately, the C&R Tree's value lies in its diagnostic and explanatory power, complementing the high predictive accuracy of the ANN model. While the ANN serves as a superior forecasting tool, the C&R Tree acts as a decision-support framework. It elucidates the underlying logic of performance variations and highlights critical conditions, thereby bridging the gap between complex data-driven predictions and actionable engineering insight. This synergistic approach—using ANN for planning and C&R Tree for operational diagnosis—constitutes a

robust methodology for optimizing TBM projects in geologically complex environments [53, 55].

VII. Conclusion

This study investigated the prediction of the Tunnel Boring Machine (TBM) Penetration Rate (PR) using three distinct machine learning models: Linear Multiple Regression (LMR), Artificial Neural Network (ANN), and Classification and Regression Tree (C&R Tree). A comprehensive database of 161 data points from 61 geological units along the Karaj Water Conveyance Tunnel (KWCT) was compiled, incorporating TBM operational parameters (RPM, cutter thrust, rolling force) and rock mass characteristics (UCS, RQD%, GSI, joint spacing). The ANN model demonstrated superior predictive accuracy ($R^2 = 0.93$, MAE = 0.02), followed by the C&R Tree model ($R^2 = 0.850$, MAE = 0.05), while the LMR model offered a baseline performance ($R^2 = 0.76$, MAE = 0.07). The ANN's performance is attributed to its ability to capture complex nonlinear interactions, which linear models cannot fully represent.

Beyond accuracy, the study highlights a critical functional distinction between the models. While LMR provides interpretable linear relationships, the C&R Tree uniquely delivers a transparent, rule-based framework. It identified the Geological Strength Index (GSI) as the most influential parameter and generated explicit decision rules (e.g., conditions leading to the lowest and highest PR). This output transforms the model from a pure forecasting tool into a diagnostic system for identifying critical geological scenarios and understanding the hierarchical logic behind performance variations.

In practical terms, this suggests a synergistic implementation strategy: ANN models are optimal for high-fidelity performance forecasting and planning, whereas C&R Tree models are invaluable for real-time operational diagnosis, risk assessment, and providing explainable insights to engineers on-site. The convergence of high accuracy (ANN) and high interpretability (C&R Tree) provides a robust, dual-framework methodology. Implementing such an integrated approach can significantly enhance TBM project planning, optimize operational parameters, mitigate risks, and improve overall construction efficiency in similar geological settings.

VIII. Limitations and Future Research

While promising, the developed models possess inherent limitations. The ANN model, despite its high accuracy, operates as a "black box," offering limited insight into the causal mechanisms behind its predictions, which can hinder trust and actionable decision-making in critical scenarios. Conversely, the C&R Tree model, while highly interpretable, showed lower predictive accuracy compared to the ANN and may be more sensitive to variations in the training dataset, potentially affecting its generalizability. Furthermore, all models were trained on data from a specific geological setting (pyroclastic and sedimentary rocks of the Alborz zone), which may limit their direct applicability to projects in significantly different rock masses, such as massive igneous or weak, water-bearing soils. Future research should focus on: (1) developing hybrid or explainable AI (XAI) techniques that combine the predictive power of ANNs with the transparency of rule-based models; (2) validating and calibrating these models using extensive datasets from diverse geological environments to enhance their robustness and transferability; and (3) integrating real-time machine data streams to enable dynamic, updating prediction models for adaptive TBM operation management.

References

[1] J. Hassanpour, J. Rostami, and J. Zhao, “A new hard rock TBM performance prediction model for project planning,” Tunnelling and Underground Space Technology, vol. 26, no. 5, pp. 595–603, 2011.
[2] J. Hassanpour, J. Rostami, M. Khamehchiyan, A. Bruland, and H. R. Tavakoli, “TBM performance analysis in pyroclastic rocks: A case history of Karaj Water Conveyance Tunnel,” Rock Mechanics and Rock Engineering, vol. 43, no. 4, pp. 427–445, 2010.
[3] A. Delisio and J. Zhao, “Analysis and prediction of TBM performance in blocky rock conditions,” Tunnelling and Underground Space Technology, vol. 33, pp. 131–142, 2014.
[4] F. F. Roxborough and H. R. Phillips, “Rock excavation by disc cutter,” Int. J. Rock Mech. Min. Sci., vol. 12, no. 12, pp. 361–366, 1975.
[5] P. Nelson, T. D. O’Rourke, and F. H. Kulhawy, “Factors affecting TBM penetration rates in sedimentary rocks,” in Proc. 24th US Symp. Rock Mechanics, pp. 227–237, College Station, TX: Texas A&M, 1983.
[6] H. P. Sanio, “Prediction of the performance of disc cutters in anisotropic rock,” Int. J. Rock Mech. Min. Sci., vol. 22, no. 3, pp. 153–161, 1985.
[7] K. Sato, F. Gong, and K. Takura, “Prediction of disc cutter performance using a circular rock cutting ring,” Colorado School of Mines, Golden, USA, 1991.
[8] J. Rostami, “A new model for performance prediction of hard rock TBMs,” in Proc. Rapid Excavation and Tunneling Conf. (RETC), 1993.
[9] J. Rostami, “Development of a force estimation model for rock fragmentation with disc cutters through theoretical modeling and physical measurement of crushed zone pressure,” Ph.D. dissertation, Colorado School of Mines, 1997.
[10] J. Rostami and L. Ozdemir, “A new model for performance prediction of hard rock TBM,” in Proc. RETC, pp. 793–809, Boston, MA, 1993.
[11] J. E. O’Rourke, J. E. Spring, and S. V. Coudray, “Geotechnical parameters and tunnel boring machine performance at Goodwin Tunnel, California,” in North America Rock Mechanics Symp., Balkema, Rotterdam, pp. 467–473, 1994.

[12] L. Ozdemir, "Development of theoretical equations for predicting tunnel boreability," Ph.D. dissertation, Colorado School of Mines, 1997.
[13] S. Yagiz, "Development of rock fracture and brittleness indices to quantify the effects of rock mass features and toughness in the CSM Model basic penetration for hard rock tunneling machines," Ph.D. dissertation, Colorado School of Mines, 2002.
[14] S. Yagiz, "Utilizing rock mass properties for predicting TBM performance in hard rock condition," Tunnelling and Underground Space Technology, vol. 23, no. 3, pp. 326–339, 2008.
[15] S. Mahdevari, K. Shahriar, S. Yagiz, and M. Akbarpour Shirazi, "A support vector regression model for predicting tunnel boring machine penetration rates," Int. J. Rock Mech. Min. Sci., vol. 72, pp. 214–229, 2014.
[16] S. Pourhashemi, K. Ahangari, J. Hassanpour, and S. Eftekhari, "TBM performance analysis in very strong and massive rocks; case study: Kerman water conveyance tunnel project, Iran," Geomechanics and Geoengineering, vol. 17, no. 4, pp. 1110–1122, 2022.
[17] E. Farrokh, J. Rostami, and C. Laughton, "Study of various models for estimation of penetration rate of hard rock TBMs," Tunnelling and Underground Space Technology, vol. 30, pp. 110–123, 2012.
[18] A. Salimi, J. Rostami, C. Moormann, and A. Delisio, "Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs," Tunnelling and Underground Space Technology, vol. 58, pp. 236–246, 2016.
[19] D. J. Armaghani and M. Azizi, "Hybrid intelligent models for predicting TBM performance," Underground Space, 2021.
[20] A. G. Benardos and D. C. Kaliampakos, "Modelling TBM performance with artificial neural networks," Tunnelling and Underground Space Technology, vol. 19, no. 6, pp. 597–605, 2004.
[21] S. Yagiz, C. Gokceoglu, E. Sezer, and S. Iplikci, "Application of two non-linear prediction tools to the estimation of tunnel boring machine performance," Eng. Appl. Artif. Intell., vol. 22, no. 4, pp. 808–814, 2009.
[22] M. Gholami, K. Shahriar, M. Sharifzadeh, and J. K. Hamidi, "A comparison of artificial neural network and multiple regression analysis in TBM performance prediction," OnePetro, 2012.
[23] A. Salimi and M. Esmaeili, "Utilising of linear and non-linear prediction tools for evaluation of penetration rate of Tunnel Boring Machine in hard rock condition," Int. J. Mining and Mineral Engineering, vol. 4, no. 3, p. 249, 2013.
[24] C. Shao, X. Li, and H. Su, "Performance prediction of hard rock TBM based on extreme learning machine," in Intelligent Robotics and Applications, Springer, Berlin, Heidelberg, pp. 409–416, 2013.
[25] S. R. Torabi, H. Shirazi, H. Hajali, and M. Monjezi, "Study of the influence of geotechnical parameters on the TBM performance in Tehran–Shomal highway project using ANN and SPSS," Arabian J. Geosci., vol. 6, no. 4, pp. 1215–1227, 2013.
[26] D. J. Armaghani, E. T. Mohamad, M. S. Narayanasamy, N. Narita, and S. Yagiz, "Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition," Tunnelling and Underground Space Technology, vol. 63, pp. 29–43, 2017.
[27] M. Koopialipoor, S. S. Nikouei, A. Marto, A. Fahimifar, D. J. Armaghani, and E. T. Mohamad, "Predicting tunnel boring machine performance through a new model based on the group method of data handling," Bull. Eng. Geol. Environ., vol. 78, no. 5, pp. 3799–3813, 2019.
[28] Q. Xu, X. Huang, B. Zhang, Z. Zhang, J. Wang, and S. Wang, "TBM performance prediction using LSTM-based hybrid neural network model: Case study of Baimang River tunnel project in Shenzhen, China," Underground Space, vol. 11, pp. 130–152, 2023.
[29] A. Afradi and A. Ebrahimabadi, "Comparison of artificial neural networks (ANN), support vector machine (SVM) and gene expression programming (GEP) approaches for predicting TBM penetration rate," SN Appl. Sci., vol. 2, no. 12, p. 2004, 2020.
[30] J. Zhou, Y. Qiu, D. J. Armaghani, W. Zhang, C. Li, S. Zhu, and R. Tarinejad, "Predicting TBM penetration rate in hard rock condition: A comparative study among six XGB-based metaheuristic techniques," Geoscience Frontiers, vol. 12, no. 3, p. 101091, 2021.
[31] J. Zhou, Y. Qiu, D. J. Armaghani, W. Zhang, C. Li, S. Zhu, and R. Tarinejad, "Predicting TBM penetration rate in hard rock condition: A comparative study among six XGB-based metaheuristic techniques," Geoscience Frontiers, vol. 12, no. 3, p. 101091, 2021.
[32] J. Zhou, Y. Qiu, S. Zhu, D. J. Armaghani, M. Khandelwal, and E. T. Mohamad, "Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization," Underground Space, vol. 6, no. 5, pp. 506–515, 2021.
[33] J. Zhou, Y. Qiu, S. Zhu, D. J. Armaghani, C. Li, H. Nguyen, and S. Yagiz, "Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate," Eng. Appl. Artif. Intell., vol. 97, p. 104015, 2021.
[34] R. D. Cook, "Detection of influential observation in linear regression," Technometrics, vol. 42, no. 1, pp. 65–68, 2000.
[35] A. Delisio, J. Zhao, and H. H. Einstein, "Analysis and prediction of TBM performance in blocky rock conditions at the Lötschberg Base Tunnel," Tunnelling and Underground Space Technology, vol. 33, pp. 131–142, 2013.
[36] K.-L. Du and M. N. S. Swamy, "Multilayer perceptrons: Architecture and error backpropagation," in Neural Networks and

Statistical Learning, Springer, London, pp. 97–141, 2019.
[37] M. Alavi, “Tectonostratigraphic synthesis and structural style of the Alborz mountain system in Northern Iran,” J. Geodynamics, vol. 21, no. 1, pp. 1–33, 1996.
[38] M. Berberian, “The Southern Caspian: A compressional depression floored by a trapped, modified oceanic crust,” Can. J. Earth Sci., vol. 20, pp. 163–183, 1983.
[39] M. R. Ghassemi, M. Fattahi, A. Landgraf, M. Ahmadi, P. Ballato, and S. H. Tabatabaei, “Kinematic links between the Eastern Mosha Fault and the North Tehran Fault, Alborz range, northern Iran,” Tectonophysics, vol. 622, pp. 81–95, 2014.
[40] H. M. Hughes, “The relative cuttability of coal-measures stone,” Mining Science and Technology, vol. 3, no. 2, pp. 95–109, 1986.
[41] J. Wang, A. S. Mohammed, E. Macioszek, M. Ali, D. V. Ulrikh, and Q. Fang, “A novel combination of PCA and machine learning techniques to select the most important factors for predicting tunnel construction performance,” Buildings, vol. 12, no. 7, p. 919, 2022.
[42] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” Nature, vol. 323, no. 6088, pp. 533–536, 1986.
[43] S. Yagiz, J. Rostami, and L. Ozdemir, “Colorado School of Mines approaches for predicting TBM performance,” in Proc. ISRM EUROCK Conf., 2012.
[44] A. Palmstrom, “RMi – a rock mass characterization system for rock engineering purposes,” Ph.D. dissertation, Univ. of Norway, 1995.
[45] N. Barton, R. Lien, and J. Lunde, “Engineering classification of rock masses for the design of tunnel support,” Rock Mechanics, vol. 6, no. 4, pp. 189–236, 1974.
[46] Z. T. Bieniawski, Engineering rock mass classifications: A complete manual for engineers and geologists in mining, civil, and petroleum engineering. New York: John Wiley & Sons, 1989.
[47] A. Bruland, “Hard rock tunnel boring,” Ph.D. dissertation, Norwegian Univ. of Science and Technology, 2000.
[48] N. R. Barton, TBM tunnelling in jointed and faulted rock. Boca Raton, FL: CRC Press, 2000.
[49] H. M. Hughes, “The relative cuttability of coal-measures stone,” Mining Science and Technology, vol. 3, no. 2, pp. 95–109, 1986.
[50] J. Wang, A. S. Mohammed, E. Macioszek, M. Ali, D. V. Ulrikh, and Q. Fang, “A novel combination of PCA and machine learning techniques to select the most important factors for predicting tunnel construction performance,” Buildings, vol. 12, no. 7, p. 919, 2022.
[51] W. Zhang, Q. Zhang, X. Li, and Z. Liu, “Machine learning-based prediction of hard rock TBM penetration rate using in-situ data and rock properties,” Tunnelling and Underground Space Technology, vol. 120, p. 104287, 2022.
[52] J. Zhou, Y. Qiu, S. Zhu, D. J. Armaghani, M. Khandelwal, and E. T. Mohamad, “Optimizing TBM performance prediction: A comparative study of tree-based ensemble models,” Geoscience Frontiers, vol. 13, no. 1, p. 101264, 2022.
[53] J. Chen, R. Liu, Y. Li, and F. Wang, “A CART-based approach for evaluating geological risks on TBM tunneling efficiency,” Rock Mechanics and Rock Engineering, vol. 54, pp. 153–169, 2021.
[54] L. Wang, Y. Zhao, H. Li, and X. Chen, “Application of Random Forest for predicting hard rock TBM performance,” Journal of Rock Mechanics and Geotechnical Engineering, vol. 12, no. 5, pp. 1042–1053, 2020.
[55] R. Mikaeil, S. Shaffiee Haghshenas, J. Omar, and M. Ataei, “Evaluating the penetration rate of TBMs using decision tree and boosted decision tree algorithms,” Neural Computing and Applications, vol. 31, pp. 4493–4507, 2019.
[56] M. Mazhabi and S. M. Eftekari, "Application of time series model in predicting the penetration rate of full-face tunnel boring machines," Journal of Rock Mechanics, vol. 1, no. 4, pp. 63–73, 2017.
[57] S. A. Fatemi, M. Ahmadi, M. J. Azinfar, and A. Chamani, "Comparison of capabilities and efficiency of different methods in predicting TBM performance and proposing the best method for model presentation," Journal of Rock Mechanics, vol. 6, no. 2, pp. 31–42, 2022.
[58] H. Chakeri, S. S. H. Bazargan, R. Mirzaei, M. Darbar, and H. Karimi Ansari, "Estimation of tunnel boring machine penetration rate using neural network optimized with particle swarm and firefly algorithms, Case study: Tabriz Metro Line 2," Journal of Rock Mechanics, vol. 9, no. 1, pp. 47–60, 2025.
[59] A. Ameli Ghomsar, E. Farrokh, and K. Shahriar, "Prediction of EPB TBM penetration rate in soft grounds using adaptive neuro-fuzzy inference system (ANFIS)," Journal of Rock Mechanics, vol. 7, no. 2, pp. 15–27, 2023.
[60] A. Fallahi, S. M. Eftekari, and E. Taheri, "Evaluation of the effect of TBM operational parameters on penetration rate using machine learning and multi-stage feature selection approach," Journal of Rock Mechanics, vol. 9, no. 1, pp. 61–79, 2025.

نشریه‌ی مکانیک سنگ
JOURNAL OF ROCK MECHANICS
دوره نهم، شماره۴، زمستان ۱۴۰۴، صفحه ۱۹تا ۳۲
https://doi.org/10.22034/IRSRM.2026.532101.0655

پیش‌بینی نرخ نفوذ TBM با استفاده از یادگیری ماشین در سازندهای آذرآواری رشته‌کوه البرز، ایران: تلفیق مدل‌های رگرسیون خطی چندگانه، شبکه عصبی مصنوعی و درخت تصمیم

علی‌اصغر قائدی وانانی [۱*]، جعفر حسن‌پور [۲*]، مهدی علیخانی [۳]، محمد اسلامی [۴]

۱- دکترای زمین شناسی مهندسی، دانشگاه تربیت مدرس، تهران، ایران.
۲- دانشکده زمین شناسی، پردیس علوم، دانشگاه تهران، تهران، ایران.
۳- گروه مهندسی عمران، دانشگاه آزاد اسلامی، واحد شهرکرد، ایران.
۴- گروه زمین شناسی، دانشکده علوم پایه، دانشگاه شیراز،شیراز، ایران.

دریافت : ۱۴۰۴/۱۰/۱۲؛ پذیرش: ۱۴۰۴/۱۲/۰۴

(*نویسنده مسئول: ali.ghaedi2012@gmail.com و hassanpour@ut.ac.ir)

چکیده

این مطالعه به توسعه و مقایسه سه مدل یادگیری ماشین -رگرسیون خطی چندگانه(*LMR*) ، شبکه عصبی مصنوعی(*ANN*) ، و درخت طبقه‌بندی و رگرسیون(*C&R Tree*) - برای پیش‌بینی نرخ نفوذ (*PR*) ماشین حفار تونل (*TBM*) می‌پردازد. مجموعه‌داده جامعی شامل ۱۶۱ نمونه صحرایی از ۲۲/۷ کیلومتر تونل انتقال آب کرج، که در سازندهای آذرآواری رشته کوه البرز (ایران) حفر شده است، گردآوری شد. این مجموعه‌داده شامل ویژگی‌های سنگ دست‌نخورده (*UCS*)، مشخصات توده‌سنگ (*RQD*، *GSI*، جهت‌یابی درز)، و پارامترهای عملیاتی ماشین (*RPM*، *Fn*، *Fr*) است. عملکرد مدل‌ها با استفاده از ضریب تعیین (R^2) و میانگین خطای مطلق (*MAE*) ارزیابی شد. مدل *ANN* با بیشترین دقت پیش‌بینی) R^2 = ۰/۹۳ و ۰/۰۲ (*MAE* = شناخته شد و پس از آن مدل *C&R Tree* (R^2 = ۰/۸ و ۰/۰۵ (*MAE* = قرار گرفت، در حالی که مدل *LMR* عملکرد کمتری نشان داد (R^2 =0.76 و ۰.۰۷(*MAE* = . قابل توجه است که مدل *C&R Tree* با قابلیت تفسیرپذیری بالا، شاخص مقاومت زمین‌شناسی (*GSI*) را به عنوان مؤثرترین پارامتر شناسایی کرد. علاوه بر این، این مدل قوانین تصمیم‌گیری صریحی را ایجاد کرد که شرایط زمین‌شناسی خاص مربوط به کمترین (۲/۳۸ متر در ساعت) و بیشترین (۴/۴۷ متر در ساعت) نرخ نفوذ مشاهده‌شده در گره‌های پایانی را مشخص می‌کند. مجموعاً، در حالی که مدل *ANN* دقت عددی بالاتری ارائه می‌دهد، مدل *C&R Tree* بینش‌های تفسیرپذیری ارائه می‌دهد که به طور مؤثری پیش‌بینی کمی و تحلیل کیفی مهندسی را پیوند می‌زند. بنابراین کاربرد هم‌افزای هر دو رویکرد به عنوان چارچوبی مستحکم برای بهینه‌سازی عملیات *TBM* و افزایش بهره‌وری حفاری تونل پیشنهاد می‌شود.

واژگان کلیدی

ماشین حفار تونل (*TBM*)؛ نرخ نفوذ؛ یادگیری ماشین؛ تونل انتقال آب کرج؛ شبکه عصبی مصنوعی؛ درخت تصمیم‌گیری