Geotech Geol Eng
DOI 10.1007/s10706-017-0380-z

ORIGINAL PAPER

Examining Feasibility of Developing a Rock Mass Classification for Hard Rock TBM Application Using Non-linear Regression, Regression Tree and Generic Programming

A. Salimi · J. Rostami · C. Moormann · J. Hassanpour

Received: 10 July 2017 / Accepted: 3 October 2017

Abstract Geotechnical and geological parameters have the greatest impact on the performance of hard rock tunnel boring machines (TBMs). This includes the rock and rock mass properties that affect the rate of penetration (ROP) as well as the machine utilization that is heavily dependent on ground support type and related machine downtime and delays. However, despite the widespread use of TBMs and established track records, accurate estimation of machine performance is still a challenge, especially in complex geological conditions. The past studies have tried to use rock mass classification systems for improving the accuracy of the machine performance prediction. Rock mass classifications has been primarily developed for design of ground support, and as such, have not offered a good fit for estimation of TBM performance. This paper will review performance of a hard rock TBM in a 12.24 km long tunnel and offers analysis of field performance data to evaluate the relationship between various lithological units and TBM operation. The results of statistical analysis of the initial 5.83 km long tunnel indicate strong relationships between geomechanical parameters and TBM performance parameters. Site specific models, including Non-linear regression analysis (NLRA), Classification and regression tree (CART), and Genetic Programming (GP) have been used for analysis of a TBM performance relative to the ground condition data. The current study has looked at the possibility of developing a new rock mass classification system for TBM application by using the above noted analysis. Preliminary results indicate that CART can be used for offering a proper rating scheme for a rock mass classification system that can be used for TBM applications.

Keywords TBM performance · Penetration rate · Non-linear regression analysis · Classification and regression tree · Genetic programming

A. Salimi (✉) · C. Moormann
Institute of Geotechnical Engineering, University of Stuttgart, Pfaffenwaldring 35, 70569 Stuttgart, Germany
e-mail: alireza.salimi@igs.uni-stuttgart.de

J. Rostami
Earth Mechanics Institute, Department of Mining Engineering, Colorado School of Mines, 1600 Illinois St., Golden, CO 80401, USA

J. Hassanpour
School of Geology, College of Science, University of Tehran, Tehran, Iran

1 Introduction

Hard rock tunnel boring has become more or less the standard method of choice for tunnels of various sizes up to 14.5 m with length over 1.5–2 km in rock. Estimating the performance of TBM is a vital part in tunnel design, and for the selection of the most appropriate excavation machine. During the past two

Published online: 09 October 2017

decades, many TBM performance prediction models for evaluation of TBM have been proposed by various researchers. In brief, TBM performance prediction models can be categorized into two distinguished approaches, namely theoretical and empirical methods (Rostami et al. 1996). Currently, three models including Colorado School of Mines or CSM (Rostami 1997) and Norwegian University of Science and Technology or NTNU (Bruland 1998) as well as field penetration index or FPI (Nelson 1983; Hassanpour et al. 2011) models are the most recognized TBM performance prediction and prognosis models in use around the world.

The CSM model allows calculation of the cutting forces that need to be applied on a disc cutter in order to reach a certain penetration into the rock. This method offers the advantages of being able to consider the geometry of the problem (the diameter and tip geometry of disc and the spacing or distance between the grooves) in detail. However, the original CSM model does not consider the natural discontinuities of the rock mass, which has an important influence on the net speed of the TBM. As a follow up work, Yagiz (2002) modified the original CSM model by adding some rock mass properties as input parameters into the model. Also, Ramezanzadeh (2005) has continued on this work and developed a database of TBM field performance for over 60 km of tunnels. He offered adjustment factors for CSM model to account for joints and discontinuities.

The NTNU method uses some rock drilling indices, such as Drilling Rate Index (DRI) estimated from rock brittleness "S_{20}" and hardness index "S_J" in addition to joint conditions to develop the estimated rate of penetration of TBM (Blindheim and Bruland 1998). The model has evolved and updated throughout the years and Bruland (1998) updated and improved the NTNU model based on field data mainly collected from Norwegian tunnels. NTNU model requires specialized tests, so called rock drillability / boreability tests, which are not commonly performed in many projects.

Field penetration index (FPI) has been introduced by Nelson et al. (1983) and has been subsequently used as a means for predicting the performance of TBMs. Hassanpour et al. (2011) has recently used FPI estimated as a function of RQD and UCS to develop a new chart for prediction of TBM performance. Furthermore, a new model has been proposed by Delisio and Zhao (2014) for prediction of TBM performance in blocky rock conditions (FPI_{blocky}) based on uniaxial compressive strength (UCS) and volumetric joint count (Jv). Also, Salimi et al. (2016a, b) proposed a new empirical equation for prediction of TBM performance (FPI) in terms of UCS and joint spacing.

As a growing trend in recent years, machine learning (ML) techniques such as artificial neural networks (ANNs), support vector machines (SVMs) and an adaptive neuro-fuzzy inference system (ANFIS) have been successfully conducted in both civil and mining engineering (Fattahi and Babanouri 2017; Singh et al. 2017; Tonnizam Mohamad et al. 2016; Sharma and Singh 2017). Mahdevari et al. (2014) used a support vector regression analysis (SVR) to predict penetration rate based on data from the Queens Water Tunnel, in New York City. Also, particle swarm optimization (PSO) technique has been utilized by Yagiz and Karahan (2011) with the same data from Queens Water Tunnel for prediction of TBM penetration rate. Furthermore, Salimi et al. (2016a, b) applied two common ML algorithms, namely SVR and ANFIS for prediction of TBM performance (FPI).

In this study, a database of actual machine performance from tunneling project namely, Maroshi-Ruparel water supply tunneling project in Mumbai, India with 5.83 km available data (Jain et al. 2014; Jain 2014) has been compiled. To examine the effect of various parameters on penetration rate, principle components analysis (PCA), non-linear regression analyses (NLRA), classification and regression tree (CART), and genetic programming (GP) method were applied to develop a new model for TBM performance estimation.

2 Description of the Project Used for This Study

The detail information related to geology of the area, TBM specifications, physio-mechanical properties of rocks for Maroshi-Ruparel tunnel project can be found in Jain et al. (2014). The main characteristics of the TBM tunneling project are summarized in Table 1.

Table 1 Main characteristics of tunneling project

Project	Tunnel length (km)	Available data (km)	TBM type and manufacture	TBM diameter (m)
Maroshi-Ruparel water supply tunnel, Mumbai, (India)	12.24	5.83 (Maroshi-Vakola)	Hard rock Gripper TBM (Wirth)	3.6

Geological zone	Formation	Lithology	UCS (MPa)	Max. depth (m)
Deccan traps (Lava flows Basaltic rocks	Upper traps (Upper Cretaceous to Lower Eocene)	Fine compact basalt, Porphyritic basalt, Amygdaloidal basalt, Pyroclastic rocks (Tuff, Tuff breccia) and Intertrappeans (Shale)	16–143	82

3 TBM Field Performance Database

In this study, data on geological setting and ground conditions, TBM operational parameters and machine performance represented by rate of penetration were collected during pre-construction and construction phases. The data is organized in a special database including 72 tunnel sections where the records of ground conditions and machine performance were reliable and could be verified. The data sets included two main categories. The first category contained machine performance parameters such as, net boring time, length of mined section as well as the average of machine operational parameters like thrust, RPM, applied torque and power throughout the section. These parameters were gathered from the daily operating records and TBM data logger. Besides, the most important performance parameters containing average penetration rate (ROP), penetration per revolution (P) and field penetration index (FPI) have been estimated using the formula as listed below:

$$ROP = \frac{L_b}{t_b}, \quad P = \frac{ROP \times 1000}{RPM \times 60}, \quad FPI = \frac{F_n}{P}, \quad F_n = \left(T_h - F_f\right)/N_{cutters} \tag{1}$$

where ROP is rate of penetration (m/h), L_b is boring length (m), t_b is boring time (h), P is cutter penetration per revolution (mm/rev), RPM is cutterhead rotational speed (rev/mm), FPI is Field Penetration Index expressed in (kN/cutter/mm/rev), F_n is cutter load or normal force, T_h is the applied thrust of the machine, and F_f is the estimated friction between the machine and the ground. The second part of database included some geological parameters such as intact rock properties (Compressive strength, porosity and…), discontinuity characteristics such as spacing, surface condition, weathering/alteration, ground water and also results of calculation of some rock mass parameters (like RQD, RMR, Q and GSI) in selected tunnel sections. Descriptive statistical distribution of variables in the database and input parameters for generated model is summarized in Table 2. Since the parameters including joint condition (J_C) and ground water condition (G_W) in RMR systems are qualitative (descriptive), the partial rating of these parameters are used in this analysis. Also, it is important to note that, where multiple joint sets were identified, different strategies could be adopted to incorporate their impact. One approach is to focus on the critical joint set which can have the highest impact (most assist or hinder) on TBM penetration rate. Another approach is to use a combination of the joints as prescribed by NTNU system and using Ks_{tot}. The last approach was to compute an average α angle for all existing joint sets. This approach has some down sides since it assumes an arithmetic averaging of the joint orientation to represent the cumulative impact of the joints. The approach used in this study was the first one i.e. using the critical joint set, which was selected to be the set with highest frequency and minimum joint spacing.

4 Rock Mass Classification Systems

Over the years, various rock mass classification systems have been presented in mining and civil engineering. According to Bieniawski (1989), a rock

Table 2 Descriptive statistics of generated database for this study (Jain et al. 2014; Jain 2014)

Variable	N	Min	Max	Mean	SD	Variance
UCS (MPa)	72	16	143	72.77	29.05	843.90
J_S (m)	72	0.05	1.25	0.41	0.241	0.059
RQD (%)	72	20	95	62.45	21.44	459.77
J_C rating in RMR	72	8	26	15.54	4.86	23.66
G_W rating in RMR	72	4	15	12.80	2.79	7.82
Alpha (°)	72	1	79	42.36	20.04	401.84
RMR (basic)	72	25	90	53.63	15.80	249.89
Q	72	0.12	165.85	11.46	24.49	600.14
GSI	72	20	85	47.88	15.49	240.15
ROP (m/h)	72	1.38	4.11	2.40	0.69	0.477
FPI (kN/cutter/mm/rev)	72	5.90	23.76	13.65	4.35	18.95

mass classification scheme is intended to classify the rock masses, provide a basis for estimating deformation and strength properties, supply quantitative data for support estimation and present a platform for communication between exploration, design and construction groups. Also, these models are commonly employed in many empirical design practices and planning in rock engineering contrasting with their original intent and applications. A good example is usage of available rock mass classification systems in estimation of TBM performance in different tunneling projects because of the simplicity and worldwide acceptance of the classification systems in general engineering practices, and in particular in underground mining and construction. To address this conflict, extensive investigations have been carried out for determining the TBM performance based on rock mass classification systems.

Some of the new versions of the models and rock mass classification systems have been modified for application in TBM projects. Barton (2000) proposed a new model based on Q system, namely Q_{TBM} which includes many input parameters. Also, Alber (1996, 2000) presented RMR_{TBM} which is modified RMR system for use in TBM tunneling projects. Sapigni et al. (2002) correlated TBM performance parameters, containing penetration rate (PR) and field penetration index (FPI) to RMR classification system. Innaurato et al. (1991) adapted Cassinelli et al. (1982) model by adding the uniaxial compressive strength (UCS) in the model. The model takes into account the effect of intact and rock mass, but the latter is defined by infrequent geomechnical quality index which is rarely available in the geotechnical characterization of a tunnel. In addition, Rock Mass Excavibility (RME) which has been proposed by von Preinl et al. (2006), can be correlated directly with the performance of TBM, based on the assessment of intact and rock mass characteristics. The fact is that, the RME index is quite similar to RMR and quite easy to estimate, however the cutterload acting on single disc is not considered. Cutterload, as shown by Rostami (1997) can have major effect on the penetration rate. Some of the recent works in this field has been done by Hassanpour et al. (2009, 2011), where a correlation was established between FPI as representative of TBM performance and UCS and RQD as representative of intact rock and rock mass properties, respectively. Using FPI addresses the issue of machine thrust or cutterload which was noted before.

Among the most commonly used rock mass classification systems, the RMR classification is easiest to apply and based on the results of Sapigni et al. (2002), RMR shows a better correlation with TBM penetration rate, possibility due to the use of intact rock property as an input parameter. This matter also has been confirmed by Hassanpour et al. (2011) which found better correlation between PR and RMR, compared to Q-system and GSI. Hence, the RMR classification has been selected for analyzing the TBM performance in this paper.

5 TBM Performance Estimation Using RMR Classification System

Various TBM performance indices have been introduced based on field penetration index (FPI), specific penetration (SP, inverse of FPI) and boreability index (BI, similar to FPI). Among the indices, the FPI developed by Nelson et al. (1983) and has been successfully applied to the determination of the correlation between the RMR and TBM performance. The advantage of the FPI is that, it accounts for the effect of two major TBM operational parameters including the TBM thrust and RPM, and by extension diameter of the TBM. This means that, it allows penetration to be normalized for cutterload and thus it automatically takes care of machine thrust variations. Usually, stronger and less fractured rock masses are more difficult to cut by disc cutters and boring by TBM and requires use of higher thrusts to achieve a certain level of penetration. So, higher values of FPI are usually seen in strong and massive rocks. In contrast, lower thrust values are needed for excavation of poor quality rock masses (Weaker and more fractured) due to crack initiation and propagation is enhanced by pre-existing fractures (Gong et al. 2006). It means that, the values of FPI are low (Hassanpour et al. 2011). Hence the FPI has been selected to be a representative of rock mass boreability. According to the result of Hassanpour et al. (2011) basic RMR shows better correlation with machine performance than RMR_{89}. Therefore, basic RMR ratings has been adopted for current study. The basic RMR classification system is determined from the sum of ratings (weighting) of inputs parameters as described in other publications (Bieniawski 1973). The assigned weightings for various parameters are fine tuned for tunnel design process for given rock mass condition. The ratings, however, are not necessarily appropriate for different purposes such as excavation applications. The relationship between FPI and basic RMR is shown in Fig. 1, which shows weak correlation between basic RMR and FPI ($R^2 = 0.52$) for the data in the tunneling project being studied.

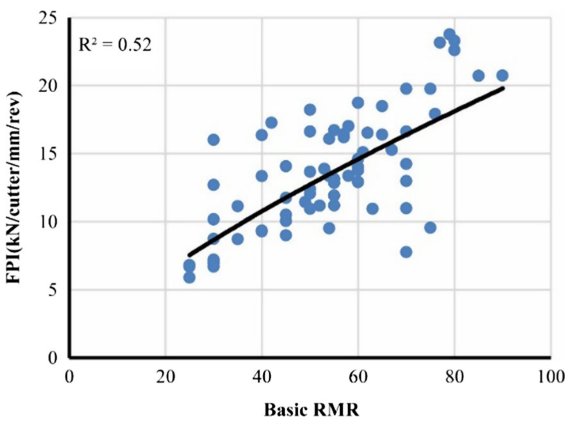


Fig. 1 Correlation between basic RMR and measured FPI

Subsequent examination of the five input parameters of RMR as separate parameters was conducted and Fig. 2 illustrates the correlations between the individual independent variables in basic RMR and measured FPI. The results indicate that the UCS shows the highest R^2 value, 0.65 and R^2 decreases in the order of the J_S (0.61), RQD (0.52), J_C partial rating in RMR (0.31) and G_W partial rating in RMR (0.003).

The last parameter in RMR classification, namely the water conditions, although important for the evaluation of stability conditions, does not appear to be a good predictor for penetration rate and therefore it can be concluded that the ground water condition has negligible effects on the TBM FPI.

Khademi Hamidi et al. (2010) considered α (Alpha) as an alternative adjustment factor for discontinuity orientation in RMR_{89}. The influence of the angle (α) on TBM penetration rate has been widely reported by many researchers (Aeberli and Wanner 1978; Lislerud 1988; Bruland 1998; Thuro and Plinninger 2003; Gong and Zhao 2009). The relationship between FPI and α angle in Fig. 3, is almost consistent with that of the results of field studies by Bruland (1998) and Yagiz (2008) and numerical simulation by Gong et al. (2005). Some studies suggest that the best TBM performance occurs at an Alpha angle between 45° and 90°. The impact of joint angle is controlled by the nature of the joints and joint spacing. For single joints this angle was found to be between 35° and 45° in this study, and 60° in previous studies, i.e. NTNU model. The difference could also be derived from the index by which the TBM performance is analyzed. TBM performance analyzed based on FPI in this study, while the performance indicator used other studies were penetration rate.

6 Developing a New Empirical Model

In rock engineering, formulae based statistical analysis of empirical data have been widely accepted for prediction of a variable based on other operational or geological parameters. Thus, empirical equations have played a great role in tunnel engineering due to their practical base as compared to more theoretical

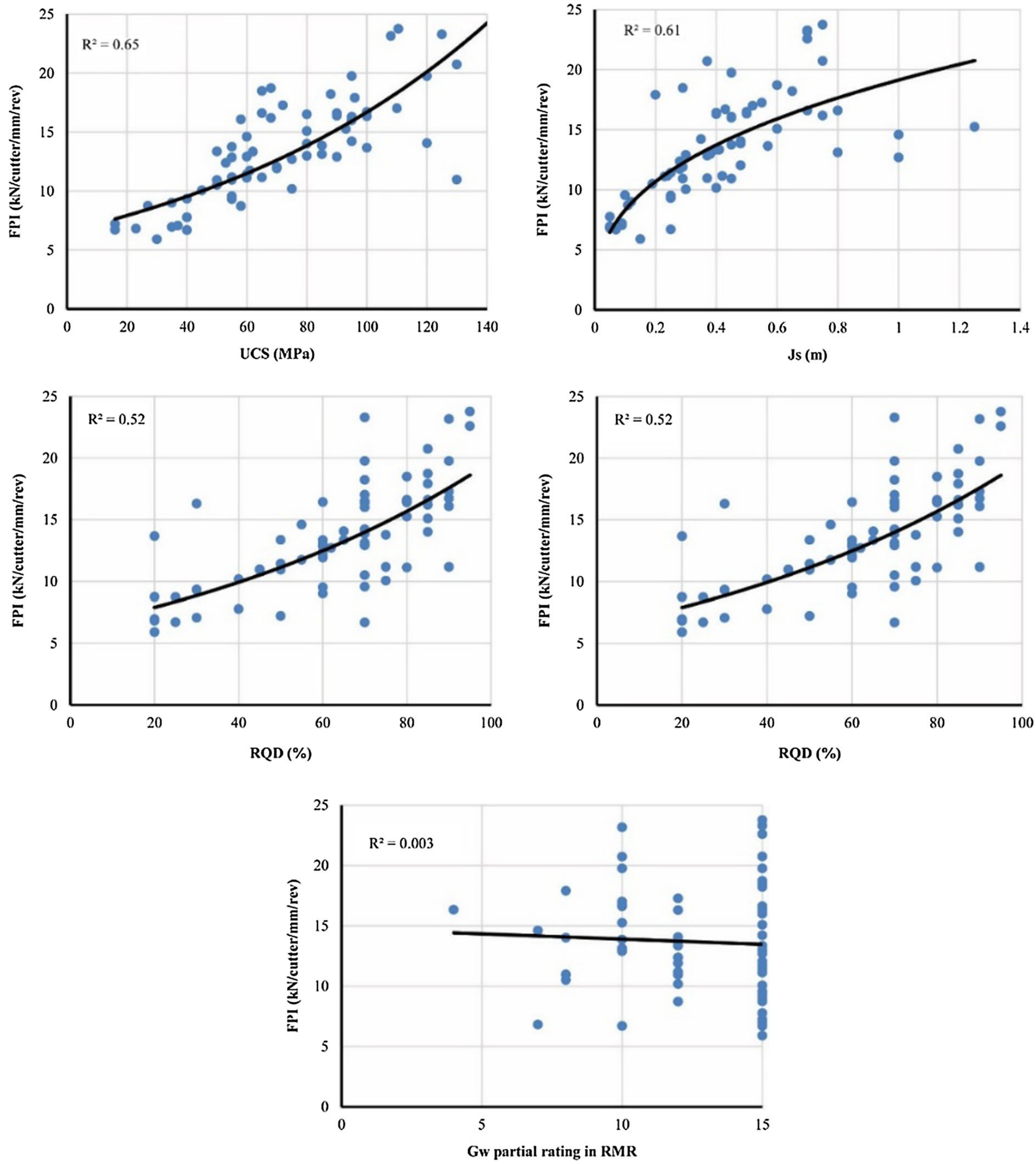


Fig. 2 The relationship between measured FPI and RMR five input parameters

analysis. In the current investigation, first goal was to identify parameters with highest impact on TBM performance; followed by seeking correlation between these parameters and TBM performance. For this purpose, the principle component analysis (PCA) was performed to identify the critical input parameters, followed by developing new empirical equations through containing classification and regression tree (CART) and genetic programming (GP).

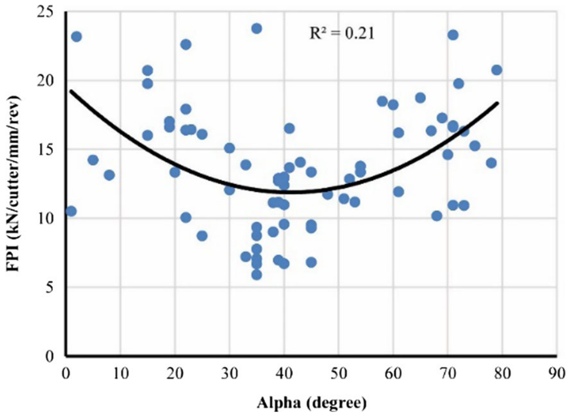


Fig. 3 Correlation of joint orientation (Alpha) with measured FPI

6.1 Principle Component Analysis

PCA is a classical method that provides a sequence of the best linear approximations to a given high-dimensional observation, and it has received much attention in the literature in recent years. With minimal effort, PCA provides a roadmap on how to reduce a complex data set to a lower dimension. More information about the methodology and procedure of PCA can be found in Salimi et al. (2016a, b). In this paper, PCA was performed on a set of target parameters (output) and factors (input parameters), and the ratio of variance of first component to total variance (variance ratio) were calculated. Accordingly, this ratio can be determined by the similarity between parameters affecting TBM performance (Fig. 4). Although the factor containing four inputs (UCS, RQD, J_C and Alpha) was shown to have the highest influence on FPI, the difference in comparison with three factors setting including (UCS, RQD, J_C) is minimal, hence FPI has been considered as a function of these three inputs i.e. (UCS, RQD, J_C); and therefore, these parameters were selected as input parameters for developing a new empirical equation and respected artificial intelligence methods (CART and GP). The impact of Alpha on TBM performance is well documented and will be addressed in the follow up studies with larger database of combined TBM performance in various projects.

6.2 Non-linear Regression Analysis

Non-linear multiple regression analyses have been utilized in rock engineering and engineering geology to solve complex problems. As stated by Davis (2002), most of the problems in geology involve complex and interacting forces, impossible to isolate and study individually. Hence, the Non-linear regression analysis method was conducted to find an empirical equation with the best fit with existing data. The R program environment was used as a modeling tool to apply multi-variable non-linear regression analysis. After a series of modeling, the best combination of rock mass parameters to predict the field penetration index (FPI), a new equation was obtained as follow:

$$FPI = e^{1.644} \times UCS^{0.007} \times RQD^{0.006} \times Jc^{0.001}. \quad (2)$$

Based on the results of statistical analysis, the coefficient of regression (R^2) was found to be 0.76. A comparison between the measured and predicted results from Eq. 2 is shown in Fig. 5. The dotted line is 1:1 line.

The correlation between input parameters and output were found in 3-dimensioanl surface plots. Figure 6 shows the correlation between UCS, RQD, J_C and FPI. The plots show that an increase in UCS and RQD value and UCS and J_C value results in higher value of FPI, as does increase in RQD and J_C. This seem intuitive and is in an agreement with field observation and previous literature. In brief, strength of rock material (UCS) and two parameters defining joint condition (RQD and J_C) can be used for estimation of ROP. The topology of the graphs in Fig. 6 a and b seem a bit odd for UCS ranging from 50-100 and this could be due to the distribution of input parameters, however when larger data set is used this issue is expected to be addressed.

7 Artificial Intelligence Algorithms

In recent years, machine learning (ML) techniques, such as artificial neural networks (ANNs), support vector machines (SVMs) and neuro-fuzzy methods have been successfully conducted for estimation of TBM performance. Alvarez Grima et al. (2000) used neuro-fuzzy method for prediction of TBM

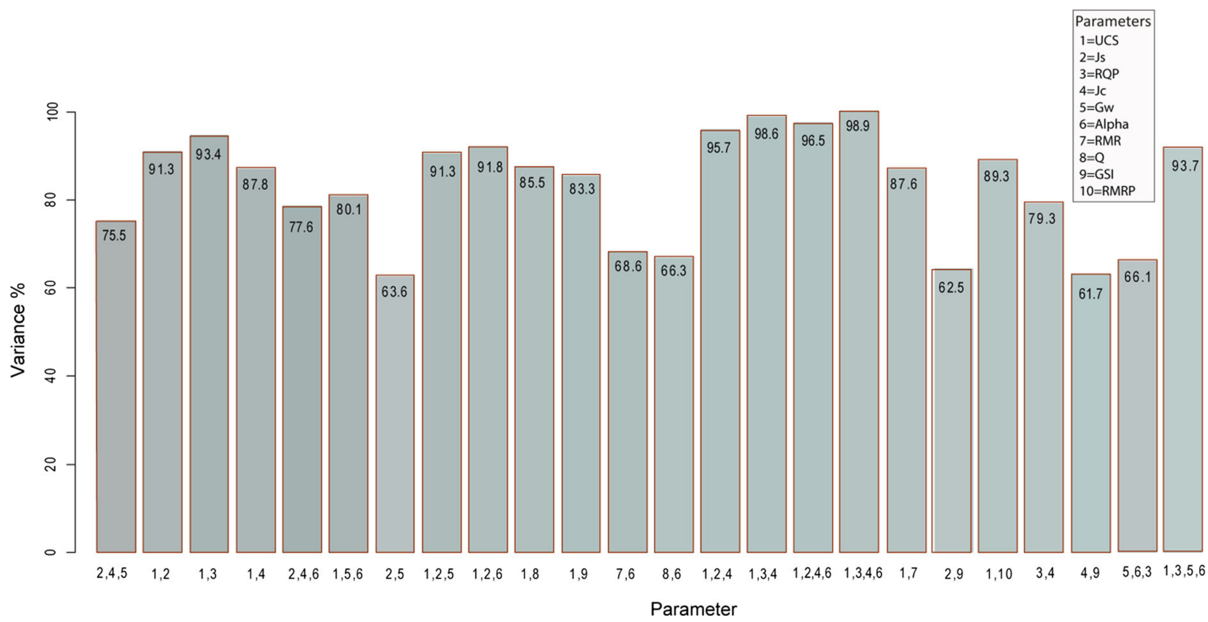


Fig. 4 Principle components analysis for some features in this study

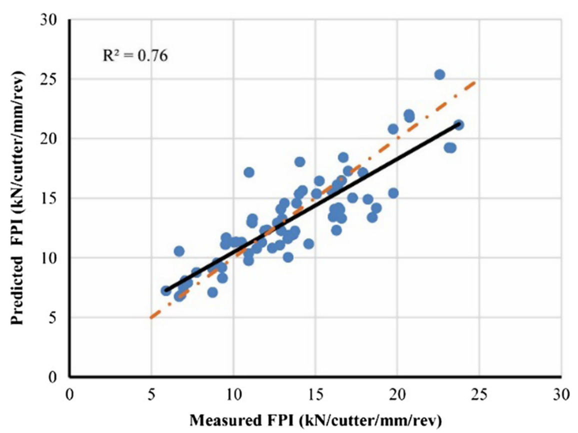


Fig. 5 Comparison between the measured and predicted FPI from Eq. 2

performance. Gholamnejad and Tayarani (2010) applied artificial neural network to estimate TBM penetration rate. In addition, Armaghani et al. (2017) applied two different hybrid intelligent models, containing particle swarm optimization (PSO)-artificial neural network (ANN) and imperialism competitive algorithm (ICA)-ANN and also simple ANN for predicting the TBM penetration rate in the Pahang-Selangor Raw Water Transfer (PSRWT) tunnel in Malaysia. The results of most of these studies has been a "black box" programs that show high correlation between their predicted rates and actual machine performance, but cannot be used by others in estimating machine performance in other projects. In this respect, two machine learning algorithms, containing classification and regression tree (CART) and genetic programing (GP) were utilized to assess the TBM performance which are not acting as "black box" and can be easily applied in other projects.

7.1 Classification and Regression Tree (CART)

One of the most promising AI methods for this application is Decision tree, which is one of the most popular techniques in data mining. It uses simple and comprehensible structure for data analysis and can be used for classification, reorganization, decision making, as well as developing prediction models. In comparison to other machine learning algorithms such as, artificial neural network (ANN) which have complex structure and are known as black-box, application of decision tree for prediction is usually preferred due to their simplicity, and low computational cost. The decision tree offers a tree structure that can be presented, easily interpreted, and defined as a transparent and reproducible. One of the approaches used for implementation of decision tree is

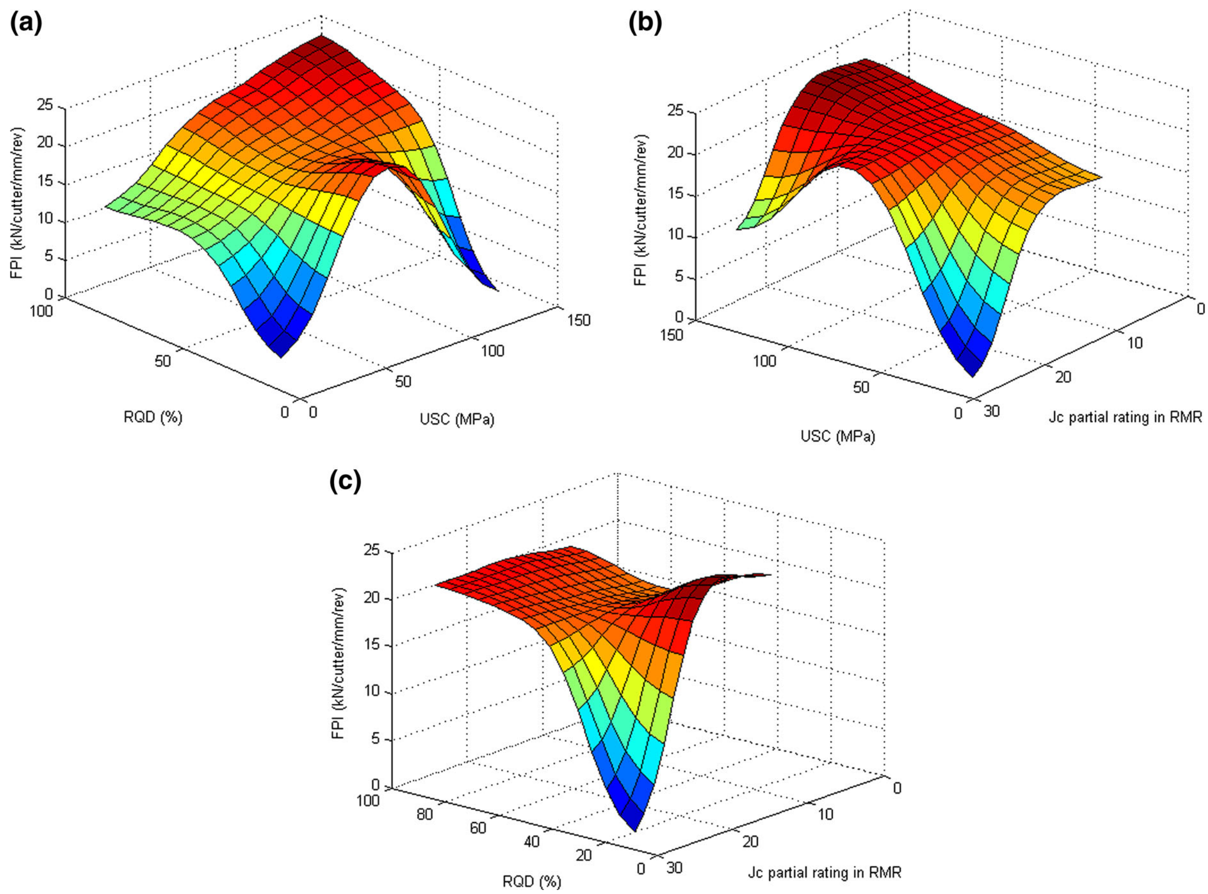


Fig. 6 Surface graph showing relationship of FPI with **a** UCS and RQD **b** UCS and joint condition **c** RQD and joint condition

classification and regression tree (CART), introduced by Brieman et al. (1984).

This method has been widely used for predicting problems in different engineering fields with high level of accuracy. CART can be used based on whether the dependent variable is being either qualitative or quantitative. It can be categorized as classification tree (CT) or regression tree (RT), respectively. RT can perform recursive portioning as an alternative method to the traditional multiple regressions. Decision tree offers a graph the same as a tree consisted of root, branches, leaves and nodes which is illustrated according to a series of question ("Yes" or "No") and each question is characterized by a variable. In the procedure of developing a decision tree, a variable is selected as root or the first node and it is partitioned into several internal nodes in a series of features. The decision tree is usually drawn from top to down; so that the root is located at the top and a chain which includes root, branch, and leaf extend to the bottom. Each node can be partitioned into two branches which are related to a certain input parameter and characterized as a specific range of input parameters. Further information regarding the algorithm and its mathematical logic can be found in Salimi et al. (2016a, b) and Breiman et al. (1984).

In order to build the CART model, the MATLAB programing language was used. It is good to note that, the suggested interval software for maximum "tree depth" and the "number of the intervals" were [2–10] and [1–10] respectively. The higher in depth, the model becomes more complicated and harder for production of the tree; and if the depth of the tree is low, the efficiency of the model will be reduced and some parameters may be omitted. Hence, the related tree depth was reduced to [3–8] and five CART models with different controlling parameters were created by trial and error. Results of these exercises are shown in

Table 3 Statistical indices for evaluation of different CART models

No	Model	Maximum tree depth	Number of interval	RMSE
1	CART	5	8	1.83
2	CART	6	6	1.91
3	CART	4	7	2.26
4	CART	7	10	1.79
5	CART	8	9	1.57

Table 3. In order to analyze the performance of the models root-mean-square-error (RMSE) has been calculated and used to quantify the errors generated by each model. As shown in Table 3, model 5 with maximum tree depth of 8 having 9 intervals is the best model. Figure 7 illustrates the preferable tree. The developed tree model has 25 nodes which are specified by squares and their related numbers and the name of variable related to the node and interval changes are shown. The detailed information regarding the structure of the tree is presented in Table 4. The relationship between measured and predicted values obtained from the CART model is shown in Fig. 8. Correlation coefficient (R^2) between measured and predicted FPI by CART model is 0.87. In Fig. 8, the dotted line is 1:1 line.

The interpretation of the model for estimation of FPI value in CART algorithm is very simple. For instance, suppose a dataset with UCS, J_C and RQD values of 70, 15 and 60, respectively. Considering the above assumptions, a root can be passed using different range of $63.5 \leq \text{UCS} < 73.5$, $J_C \geq 13$ and $\text{RQD} < 70$ and finally it will reach to node number 20 which predicts the value 12.05 for FPI. This process is repeated for other datasets and relevant FPI values can be predicted. Also, as can be seen from the graph there is no need to have all input parameters i.e. UCS, RQD and J_C simultaneously and in some parts having the single parameter (UCS values) is sufficient for estimation of FPI in this case study. When using a UCS alone, it seems like the effect of rock mass fracturing is not considered. Generally, geological units which include rocks with low values of UCS are more fractured. In fact, values of UCS also reflect the effect of fracturing on TBM performance. Therefore, it can be concluded that in some parts of tunnel sections the effect of rock mass joining and fracturing is hidden in UCS values. Another reason could be the rang of input parameters notably UCS.

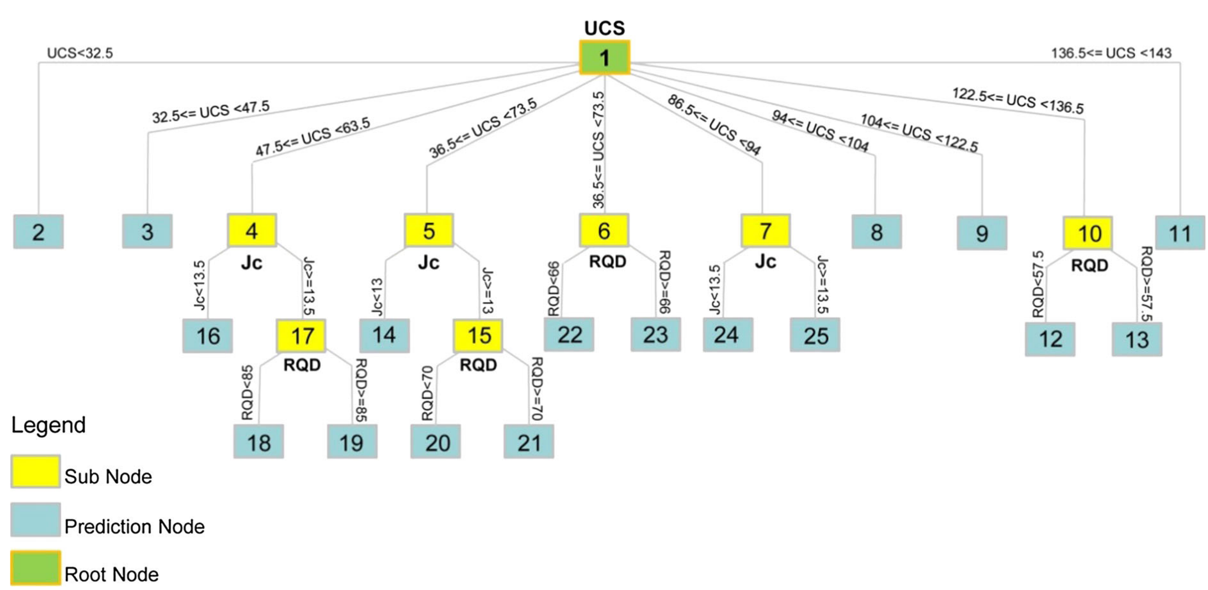


Fig. 7 Regression tree developed for FPI

Table 4 Information related to each node in the CART model

Node	Frequency	Split variable	Value	Predicted
1	72	–	–	13.656
2	5	UCS	[16, 32.5]*	7.072
3	7	UCS	[32.5, 47.5]	8.121
4	20	UCS	[47.5, 63.5]	11.875
5	8	UCS	[63.5, 73.5]	15.300
6	8	UCS	[73.5, 86.5]	13.553
7	6	UCS	[86.5, 94]	15.963
8	8	UCS	[94, 104]	16.363
9	5	UCS	[104, 122.5]	19.548
10	4	UCS	[122.5, 136.5]	18.930
11	1	UCS	[136.5, 143]	22.590
12	1	RQD	[45, 57.5]	10.970
13	3	RQD	[57.5, 85]	21.583
14	2	J_C	[10, 13]	11.535
15	6	J_C	[13, 25]	16.555
16	8	J_C	[8, 13.5]	10.809
17	12	J_C	[13.5, 2]	12.586
18	11	RQD	[30, 85]	12.268
19	1	RQD	[85, 90]	16.080
20	1	RQD	[60, 70]	12.050
21	5	RQD	[70, 90]	17.456
22	3	RQD	[40, 66]	11.947
23	5	RQD	[66, 85]	14.516
24	1	J_C	[10, 13.5]	12.890
25	5	J_C	[13.5, 25]	16.578

* The software keeps the left bracket open

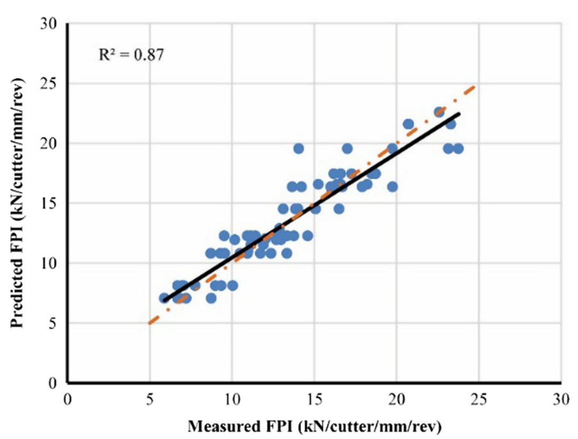


Fig. 8 Comparison between the measured and predicted FPI from CART model

7.2 Genetic Programming (GP)

Genetic programing (GP) is an extension of genetic algorithm (GA) which is considered as a branch of evolutionary computing (EC), first presented by Cramer (1985) and then developed by Koza (1992). GP is a modified version of the GA which is based on Darwin's theory of "survival of the fitness" and is a biological model of evaluation and natural selection. The main difference between GP and other EAs like GA is related to the structure of the problem solutions (individuals). Individuals in GA are linear coded binary strings of fixed length that known as chromosomes, whereas in GP individuals are the computer programs which follow LISP language. LISP is an acronym of the processing list and can be applied for executing the data strings. The programs are called symbolic expression (s-Expression) and can be presented in the form of tree structure with different sizes and shapes. The population in GP is initialized with randomly generated the programs which are composed of terminal set (T) and function set (F). The fact is that, if the terminal and function set are not selected appropriately for the problem, the desired results cannot be achieved.

Terminals are input (independent) variables of the model and a set of constant values that assigned as GP designer according to nature of the problem. The function set is arithmetical, logical, Booleans or user-defined functions (e.g. +, −, *, /, ^2, ^3, Q, sin, cos, tan, ln, and, or, not, nor and etc.). A typical program, presenting the expression is demonstrated in Fig. 9. In this example, the function set (F) is consisted of multiplication, division, addition, subtraction and the

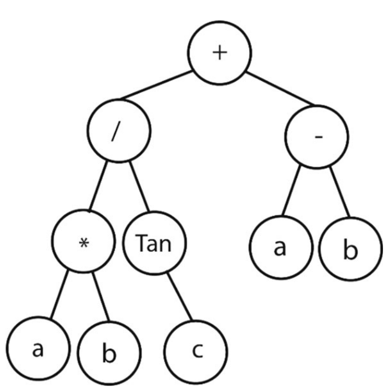


Fig. 9 A typical parse tree structure of the function ab/(tan(c)) + (a − c)

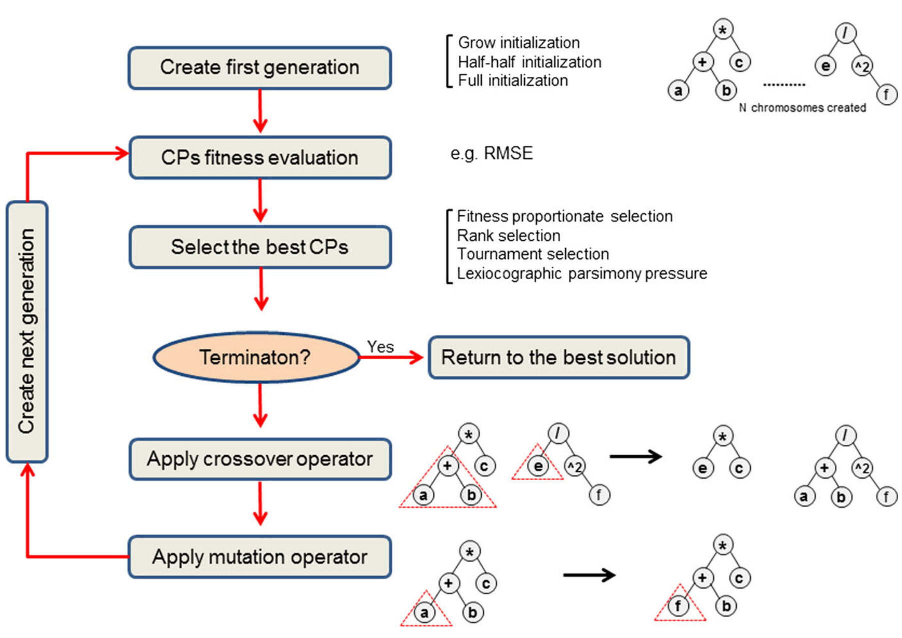
tangent function, $F = \{\times, /, +, -, \tan\}$and the terminal set (T) is consisted of N = 3 variable as $T = \{a, b, c\}$. The flowchart of GP system is shown in Fig. 10. As can be seen, the process of GP algorithm starts with a random generation of the first population of CPs. For this purpose, three common initialization methods, including grow, half, and full initialization are available. After generating the initial population; the fitness of each individual data will be evaluated by a fitness function and root-mean-square-error (RMSE) is used for this investigation and then if the stopping conditions (best solution or number of generation) are not attained, the process will continue. It is good to note that, three important genetic operators, containing reproduction, crossover, and mutation respectively, are applied in GP algorithm.

Reproduction associates selecting a computer program from the current population of programs relative to the fitness criteria, and allowing it reproduction by coping it into the new population as noted by Koza (1992). Several different types of reproduction operations, including fitness-proportionate reproduction or the roulette wheel selection, tournament selection, rank, and lexicographic parsimony pressure selection are commonly applied in GP algorithm. Further information about the mentioned methods can be found in Koza (1992). Crossover is the predominant search operator in GP which is highly stochastic. This operator is a binary operator, where it selects two parse trees (parents) probabilistically and swaps two parts between them that finally make two new computer programs (children or offspring). Another genetic operator is mutation which is a unary operator, introduces random changes in the individual. Mutation selects randomly individuals and can occur on any operational and final nodes of parse trees. If the selected node is an operational node, it changes with another node with sub-trees that contain a new function otherwise if the selected node is a final node; it substitutes just with another node. After applying these genetic operators, new programs with modified structures will be created as the second generation. This process continues to achieve the maximum generation (Koza 1992; Silva and Almeida 2003, Shirani Faradonbeh et al. 2017).

In the modeling procedure of GP, GPlab which is genetic programming toolbox for MATLAB developed by Silva and Almeida (2003) was utilized to propose a new mathematical equation for prediction of TBM performance.

As a first step, terminal sets need to be assigned. In this study, UCS, RQD and Jc are introduced as inputs of GP model. Besides, to improve the ability of GP in function finding, 6 random constants ranging from − 10 to + 10 were defined. Therefore, the used terminal sets include as follow:

Fig. 10 A general flowchart for GP



$$F = \{UCS, RQD, Jc, 5.9168, 3.6931, 9.7011, 4.2831, 7.2241, 1.5137\} \quad (3)$$

To minimize the error and maximize correlation coefficient (R^2), the most common functions suggested by other researchers were taken into consideration and used as follow:

$$F = \{+, -, \times, /, \hat{}2, \hat{}3, 1/x, Sqrt, Sin, Cos, \tan, Arc\tan, Exp, Ln\} \quad (4)$$

The fitness of each individual parameter was evaluated by root-mean-square-error (RMSE). It is worth to note that in addition to genetic operators which need to be arranged, several important parameters composed of number of population, the initialization method, selection method, number of generation and etc. are necessary to be assigned. In order to determine the proper values of these parameters, several models with different conditions were evaluated based on trial-and-error procedure. The optimum combination of GP parameters is obtained as listed in Table 5. The program is evolved through 500 generations and the best solution so-far, is considered as the result of genetic programing. Equation 5 is the mathematical phrase of the best generated computer program by GP.

$$FPI = \left(\frac{UCS}{7.2241}\right) + a\tan\left[\left(\frac{RQD - UCS}{RQD}\right) \times 7.2241\right] + (1.5137 \times \ln(Jc)) \quad (5)$$

Tree structure of the best GP model for TBM performance prediction is depicted in Fig. 11. The

Table 5 GP parameters for constructed model

GP parameters	Values
Terminal set	UCS, RQD, Jc
Fitness function	RMSE
Number of population	500
Number of generation	500
Initialization method	Half
Section method	Roulette wheel
Genetic operators	
Crossover	0.9
Mutation	0.05
Reproduction	0.2

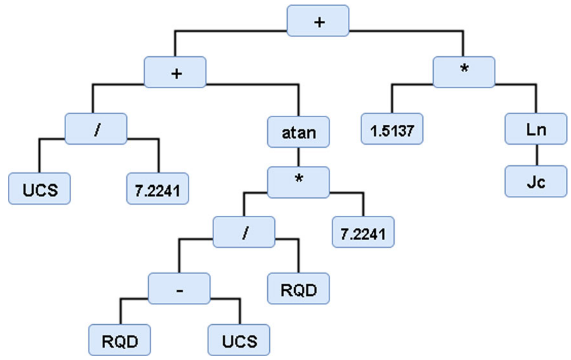


Fig. 11 Tree structure of the best GP model

relationship between measured and predicted values obtained from the GP model is shown in Fig. 12. Correlation coefficient (R^2) between measured and predicted FPI by GP model is 0.81. In Fig. 12, the dotted line is 1:1 line. It should be iterated that these results are preliminary and the study was done to evaluate the capability of the procedure for use in this application. The initial success will be tried with larger database of TBM field performance in near future.

8 Conclusions

Overall the results of the current study and analysis of existing models for prediction of TBM field performance has revealed that the RMR system, or in general all the existing rock mass classification systems with the present rating scheme have limited accuracy in prediction of TBM performance. To mitigate this limitation and develop new models, the principle component analysis (PAC) was applied to find the critically important parameters to asses boreability of rock mass and develop new models for estimating TBM performance. The hypothesis was tested by using a set of data from a recent project in India. Based on the PCA results, the most effective parameters on boreability of rock mass contain UCS, RQD and J_C. Because the influence of joint orientation is not monotonic and peaks around 45-60 degrees, it has proven to be difficult to incorporate join orientation in the prediction models and this parameter has not appeared as an influential factor in PCA analysis, an experience that has been observed by other researchers in the past. This investigation shows that the effect of joint orientation on TBM penetration rate is a function

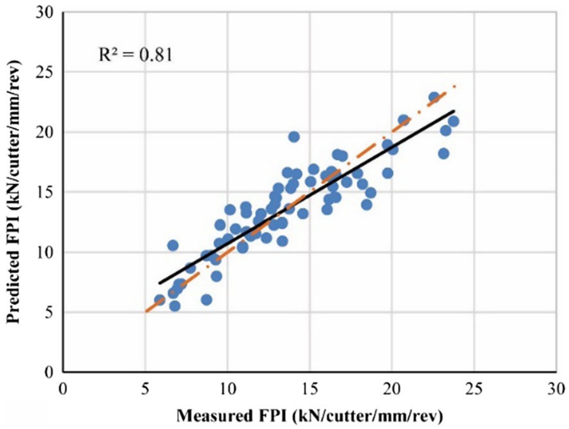


Fig. 12 Comparison between the measured and predicted FPI from GP model

of joint fracture spacing (i.e. RQD or J_S), as was also indicated by NTNU model (Bruland 1998).

As noted earlier, most of the studies of TBM performance predictions have been investigated using soft computing methods, such as ANN and SVR that are predominantly based on black box procedure. Although these models can predict TBM performance with a high level of accuracy, they cannot propose a mathematical equations or graphic representations using inputs to estimate performance of TBM. GP and CART are two applications that can solve this problem and present new equations as well as graphs (diagrams) for TBM performance prediction, respectively.

In this study, non-linear regression analysis (NLRA); classification and regression tree (CART), and genetic programing (GP) models has been investigated to develop new models for estimation of FPI. Among them, the CART model shows better accuracy for TBM performance estimation, however, it is prudent to apply all three methods for TBM performance prediction and compare the results, perhaps use the average value for FPI for prediction. Obviously, the range of input data, rock types, machine types, etc. in this study has been very limited and as such, the results cannot be considered to be universal. The current study was performed to examine the capability of the listed procedures for development of the TBM performance prediction models and as such, a limited database is used for the initial studies. Follow up works will be performed on larger data sets to offer a more comprehensive model.

Acknowledgements The authors wish to extend their sincere thanks to Dr. Prasnna Jain and Prof. T. N. Singh for sharing their database of TBM field performance for this study. Also, the authors express their appreciation to Mr. Ehsan Sharafian for his assistance and comments on this study.

References

Aeberli U, Wanner WJ (1978) On the influence of discontinuities at the application of tunneling machines. In: Proceedings of the 3rd international congress IAEG, Madrid, p 7–14

Alber M (1996) Prediction of penetration, utilization for hard rock TBMs. In: Proceedings of the international conference of Eurock '96, Balkema, Rotterdam, p 721–725

Alber M (2000) Advance rates of hard rock TBMs and their effects on project economics. Tunn Undergr Space Technol 15(1):55–64

Alvarez Grima M, Bruines PA, Verhoef PNW (2000) Modeling tunnel boring machine performance by neuro-fuzzy method. Int J Tunn Undergr Space Technol 15:259–269

Armaghani DJ, Tonnizam Mohamad E, Sundaram Narayanasamy M, Narita N, Yagiz S (2017) Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. Int J Tunn Undergr Space Technol 63:29–43

Barton N (2000) TBM tunneling in jointed and faulted rock. Balkema, Brookfield

Bieniawski ZT (1973) Engineering classification of jointed rock masses. Trans S Afr Inst Civ Eng 15:335–344

Bieniawski ZT (1989) Engineering rock mass classifications. Wiley, New York, p 251

Blindheim OT, Bruland A (1998) Boreability testing. Norwegian TBM tunneling. Norwegian Tunnelling Society, Oslo

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and Regression Trees. Chapman and Hall/CRC

Bruland A (1998) Hard rock tunnel boring. Ph.D. thesis, Norwegian University of Science and Technology, Trondheim

Cassinelli F, Cina S, Innaurato N, Mancini R, Sampaolo A (1982) Power consumption and metal wear in tunnel-boring machines: analysis of tunnel boring operation in hard rock. In: Tunnelling '82, London, Inst Min Metall, p 73–81

Cramer NL (1985) A representation for the adaptive generation of simple sequential programs. In: Grefenstette JJ (ed) Proceedings of the first international conference on genetic algorithms and their applications. Erlbaum

Davis JC (2002) Statistics and data analysis in geology, 3rd edn. Wiley, New York p, p 638

Delisio A, Zhao J (2014) A new model for TBM performance in blocky rock conditions. Int J Tunn Undergr Space Technol 43:440–452

Faradonbeh RS, Salimi A, Monjezi M, Ebrahimabadi A, Moormann C (2017) Roadheader performance prediction using genetic programming (GP) and gene expression programming (GEP) techniques. Int J Environ Earth Sci. doi:10.1007/s12665-017-6920-2

Fattahi H, Babanouri N (2017) Applying optimized support vector regression models for prediction of tunnel boring

machine performance. Int J Geotech Geol Eng. doi:10.1007/s10706-017-0238-4
Gholamnejad J, Tayarani N (2010) Application of artificial neural networks to the prediction of tunnel boring machine penetration rate. Min Sci Technol (China) 20(5):727–733
Gong QM, Zhao J (2009) Development of a rock mass characteristics model for TBM penetration rate prediction. Int J Rock Mech Min Sci 46(1):8–18
Gong QM, Zhao J, Jiao YY (2005) Numerical modeling of the effects of joint orientation on rock fragmentation by TBM cutters. Tunn Undergr Space Technol 20:183–191
Gong QM, Jiao YY, Zhao J (2006) Numerical simulation of influence of joint spacing on rock fragmentation by TBM cutters. Int J Tunnell Undergr Space Technol 21(1):46–55
Hassanpour J, Rostami J, Khamehchiyan M, Bruland A (2009) Developing new equations for TBM performance prediction in carbonate-argillaceous rocks: a case history of Nowsood water conveyance tunnel. Geomech Geoeng Int J 4:287–297
Hassanpour J, Rostami J, Zhao J (2011) A new hard rock TBM performance prediction model for project planning. Tunn Undergr Space Technol 26:595–603
Innaurato N, Rondena E, Zaninetti A (1991) Forecasting and effective TBM performance in a rapid excavation of tunnel in Italy. In: Proceedings of the 7th international congress on rock mechanics ISRM. Balkema, Rotterdam. Aachen, Germany, p 1009–1014
Jain P (2014) Evaluation of engineering geological and geotechnical properties for the performance of a tunnel boring machine in Deccan traps rocks—a case study From Mumbai, India. Ph.D. thesis, Indian Institute of Technology Bombay, India (Unpublished)
Jain P, Naithani AK, Singh TN (2014) Performance characteristics of tunnel boring machine in basalt and pyroclastic rocks of Deccan traps—a case study. Int J Rock Mech Geotech Eng 6:36–47
Khademi Hamidi J, Shahriar K, Rezai B, Rostami J (2010) Performance prediction of hard rock TBM using Rock Mass Rating (RMR) system. Tunn Undergr Space Technol 25(4):333–345
Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge
Lislerud A (1988) Hard rock tunnel boring: prognosis and cost. Tunn Undergr Space Technol 3(1):9–17
Mahdevari S, Shahriar K, Yagiz S, Akbarpour Shirazi M (2014) A support vector regression model for predicting tunnel boring machine penetration rates. Int J Rock Mech Min Sci 72:214–229
Nelson PP (1983) Tunnel boring machine performance in sedimentary rock. Ph.D. thesis, Cornell University, Ithaca, NY
Nelson P, O'Rourke TD, Kulhawy FH (1983) Factors affecting TBM penetration rates in sedimentary rocks. In: Proceedings, 24th US symposium on rock mechanics, Texas A&M, College Station, TX, p 227–237
Ramezanzadeh A (2005) Performance analysis and development of new models for performance prediction of hard rock TBMs in rock mass. Ph.D. thesis, INSA, Lyon, France
Rostami J (1997) Development of a force estimation model for rock fragmentation with disc cutters through theoretical modeling and physical measurement of crushed zone pressure. Ph.D. thesis, Colorado School of Mines, Golden, Colorado, USA
Rostami J, Ozdemir L, Nilsen B (1996) Comparison between CSM and NTH hard rock TBM performance prediction models. In: Proceedings, the annual conference of the institution of shaft drilling technology (ISDT), Las Vegas
Salimi A, Faradonbeh RS, Monjezi M, Moormann C (2016a) TBM performance estimation using a classification and regression tree (CART) technique. Int J Bull Eng Geol Environ. doi:10.1007/s10064-016-0969-0
Salimi A, Rostami J, Moormann C, Delisio A (2016b) Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs. Int J Tunn Undergr Space Technol 58:236–246
Sapigni M, Berti M, Behtaz E, Busillo A, Cardone G (2002) TBM performance estimation using rock mass classification. Int J Rock Mech Min Sci 39:771–788
Sharma LK, Singh TN (2017) Regression based models for the prediction of unconfined compressive strength of artificially structured soil. Int J Eng Comput. doi:10.1007/s00366-017-0528-8
Silva S, Almeida J (2003) GPLAB, a genetic programming tool box for MATLAB. In: Proceeding from the Nordic MATLAB conference, Copenhagen, Denmark, pp 273–278
Singh R, Kumar Umrao R, Ahmad M, Ansari MK, Sharma LK, Singh TN (2017) Prediction of geomechanical parameters using soft computing and multiple regression approach. Int J Meas 99:108–119
Thuro K, Plinninger RJ (2003) Hard rock tunnel boring, cutting, drilling and blasting: rock parameters for excavatability. ISRM 2003-technology roadmap for rock mechanics, South African Institute of Mining and Metallurgy, p 1–7
Tonnizam Mohamad E, Faradonbeh RS, Armaghani DJ, Monjezi M, Abd Majid MZ (2016) An optimized ANN model based on genetic algorithm for predicting ripping production. Int J Neural Comput Appl. doi:10.1007/s00521-016-2359-8
von Preinl ZTB, Celada Tamames B, Galera Fernandez JM, Alvarez Hernandez M (2006) Rock masse excavability indicator: new way to selecting the optimum tunnel construction method. Tunn Undergr Space Technol 21(3–4):237
Yagiz S (2002) 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. thesis, Department of Mining and Earth Systems Engineering, Colorado School of Mines, Golden, Colorado, USA
Yagiz S (2008) Utilizing rock mass properties for predicting TBM performance in hard rock condition. Tunn Undergr Space Technol 23(3):326–339
Yagiz S, Karahan H (2011) Prediction of hard rock TBM penetration rate using particle swarm optimization. Int J Rock Mech Min Sci 48:427–433