Introducing Tree-Based-Regression Models for Prediction of Hard Rock TBM Performance with Consideration of Rock Type

Alireza Salimi1,2 · Jamal Rostami3 · Christian Moormann1 · Jafar Hassanpour4

Received: 17 October 2021 / Accepted: 27 March 2022 / Published online: 25 May 2022
© The Author(s) 2022

Abstract
Prediction of machine performance is a fundamental step for planning, cost estimation/control and selection of the machine type when using a tunnel boring machine (TBM). Penetration rate (PR) and machine utilization (U) are the two principal measures of TBM performance for evaluating the feasibility of using a machine in a given ground condition. However, despite the widespread use of TBMs and established track records, accurate estimation of machine performance could still be a challenge, particularly in complex geological conditions. Since different types of rocks have varied texture (cementation and grain size), and respond differently to cutting forces in the TBM tunnelling, incorporating the effects of rock type in performance prediction models can improve the accuracy of the estimates. The aim of this study was to develop models for predicting penetration rate of hard rock TBMs in different types of rock based on field penetration index (FPI), using multivariable regression analysis and machine learning algorithm, including classification and regression tree (CART). The proposed models offer estimated FPIs in different rock types, rock strength, and rock mass properties in the form of graphs (diagrams), which can be used to estimate TBM penetration rate. The proposed models have been developed based on the analysis of a comprehensive database of TBM performance in various rock types and offers more accurate estimates of machine performance by incorporating many of the key parameters available in typical geotechnical reports and contract documents. The models also exhibit sensitivity to rock mass parameters for predicting the penetration rate.

Highlights

- TBM performance was investigated and predicted in terms of rock texture (cementation & grain size).
- Geological and geotechnical factors, and machine specifications were considered.
- Development of new model via regression analysis.
- Application of Regression Tree to present a new chart/graph for estimation of TBM performance.

Keywords TBM performance · Penetration rate · Field penetration index (FPI) · Multivariable regression analysis · Classification and regression tree (CART)

1 Introduction

Tunnel Boring Machines (TBM) have been the predominant choice of tunnelling methods in various grounds, especially in hard rock applications with length over 1.5–2 km due to achievement of higher excavation speed, lower cost, improved safety, and environmental friendliness compared to traditional drill and blast method. Estimating TBM performance is a key parameter for tunnel design and selection of the appropriate machine type and specification. In last two decades, many performance predictions models have been
offered by various researchers to estimate penetration rate of hard rock tunnel boring machines (TBMs) in new tunneling projects which can be categorized in two main groups, namely theoretical and empirical methods (Khademi Hamidi et al. 2010). Theoretical models analyse cutting forces acting on disc cutter to estimate ROP based on force equilibrium equations. Laboratory cutting tests provide a basic understanding of rock fragmentation and the force-penetration behaviour of rocks are the basis for this class of performance prediction models. The main disadvantage of these models is that they do not completely represent the site parameters relative to rock mass conditions, in particular joints, as the TBM disc cutters would encounter in the field. Empirical models are primarily based on observation of field performance of the TBMs. In such cases where standard laboratory rock cutting facilities are not available, TBM performance may be predicted using formulas developed empirically.

Currently, three different 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 (FPI) (Nelson et al. 1983, Hassanpour et al. (2011, 2016) models are the most recognized TBM performance prediction and prognosis models in use around the world. The CSM model represents a semi-theoretical approach to TBM performance as it allows the calculation of the cutting forces that need to be applied on a disc cutter 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 the 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 have a major impact on the net speed of the TBM. To overcome this shortcoming, Yagiz (2002) and Ramezanzadeh (2005) modified the original CSM model by adding some rock mass properties as input parameters into the model, but with limited success.

Bruland (1998) updated and improved the NTNU model, which was originally proposed in 1978, based on field data originally collected from Norwegian tunnels, and subsequently expanded to other tunnelling projects around the world. The NTNU method uses some rock property indices such as Drilling Rate Index (DRI) estimated from rock brittleness “$S_{20}$” and hardness index “$S_{1}$” in addition to joint conditions to develop the estimated rate of penetration of TBM (Blindheim 1979). The NTNU model requires specialized tests which are not commonly performed in many projects. Filed 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. For instance, Hassanpour et al. (2011, 2016, 2021); Pourhashemi et al. (2021) and Goodarzi et al. (2021) has used FPI estimated as a function of UCS and RQD to develop new equations and charts for TBM performance prediction. These two models represent the available empirical approaches to estimate TBM penetration rate.

Apart from empirical and theoretical models, the use of machine learning (ML) techniques has received widespread attention in TBM performance prediction. Machine learning is a branch of artificial intelligence that consists of developing algorithms able to generalize behaviours from information provided in the form of examples. It is therefore, an inductive knowledge strategy (Salimi 2021; Coimbra et al. 2014). Capabilities and opportunities of using the machine learning algorithms and methods in underground tunnel construction have been discussed by Marcher et al. (2020) and Morgenroth et al. (2019). Several techniques, such as an artificial neural network (ANN), fuzzy logic, adaptive neuro-fuzzy inference system (ANFIS), particle swarm optimization (PSO) and support vector machine (SVM), random forest (RF), deep neural network (DNN) in approximating TBM performance parameters like penetration rate (PR) and advance rate (AR) have been highlighted by many scholars (Armaghani et al. 2017; Salimi et al. 2016; Koopialipoor et al. 2019; Mahdevari et al. 2014; Yagiz et al. 2009; Benardosand and Kaliampakos 2004; Ghasemi et al. 2014; Alvarez Grima et al. 2000). The flexible nature of the AI techniques makes them powerful tools in approximating and solving engineering problems more specifically when the problem is highly complex and nonlinear. However, the results of most of these studies show high correlation between their predicted rates and actual machine performance but cannot be used in estimating machine performance in other projects, since the related programs are not available to the end users. Furthermore, most of these machine learning methods (e.g., artificial neural networks or support vector machines) are difficult to apply as a large quantity of parameters must be provided or estimated to use these models. This means that they can be applied to predict the value of a target variable depending on data, but the rules or implicit patterns within the model cannot be interpreted. In the area of rock engineering/tunnelling, the suitability of data mining techniques is closely related to the applicability of the resulting model (Salimi 2021).

The main goal of the present work is to develop new models for estimation of TBM performance using FPI via statistical analysis (Regression analysis), as well as machine learning including tree-based-regression model such as, classification and regression tree (CART), which is known as graph/transparent solutions for the prediction of TBM performance in hard rock conditions. More attention is paid to introduce new models that incorporate the similarities in rock texture, cementation and grain size.

To reach this goal, compiled field data from eight tunnelling projects was compiled into a database and used in subsequent analysis. This includes Zagros water conveyance tunnel, Lot 2 in Iran (Hassanpour et al. 2009, 2016); Ghomrood water conveyance tunnel, Lots 3 and 4 in Iran (SCE Company 2004; Hassanpour et al. 2011); Karaj water conveyance tunnel, Lot 1 in...
Iran (Hassanpour et al. 2010); Golab conveyance water tunnel in Iran (Fatemi et al. 2016; ICE 2009); Maroshi-Ruparel water supply tunnel, Mumbai India (Jain et al. 2014; Jain 2014); Manapouri second tailrace tunnel, New Zealand (URS Company 2003; Delisio 2014; Deere et al. 2004); Deep Tunnel Sewerage System, in Singapore (Gong 2005) and Lötschberg Base Tunnel in Switzerland (Delisio and Zhao 2014; Delisio 2014).

## 2 Tunnelling Projects in TBM Field Performance Database

TBM performance data from various projects with the different rock mass conditions have been obtained and compiled in a TBM field performance database. Data from eight tunnelling projects including, Zagros water conveyance tunnel Lot 2, Manapouri second tailrace tunnel, Maroshi-Ruparel college tunnel, Ghomrood water conveyance tunnel Lots 3 and 4, Karaj water conveyance tunnel Lot 1, Golab conveyance water tunnel, Lötschberg Base Tunnel and Deep Tunnel Sewerage System for a total length of 92.93 km were selected for this investigation. The main characteristics of these TBM tunnelling projects are summarized in Table 1, while some more detailed information about the geological conditions along the tunnel alignment and the adopted construction method can be found in the literature. Also, Fig. 1 shows the geographical distribution of project sites.

### 3 Data Compilation

TBM field performance database contains different levels of information which defines the tunnel, rock mass conditions, and TBM performance parameters over the full length of a tunnel drive. The database contains over 666 data sets where ground conditions and machine performance were reliable input parameters were available and could be verified. The data sets comprised two main categories. The first category included machine performance parameters like net boring time, length of mined section and also the average of machine operational parameters (thrust, RPM, power and applied torque) throughout the section. The second part of database or category of information included some geological parameters such as intact rock properties (Compressive and tensile strength, quartz content, porosity), discontinuity characteristics such as spacing, weathering, surface condition and results of calculation of some rock mass parameters (like RQD, RMR) in selected tunnel sections. Also, the most important performance parameters containing average penetration rate (ROP), penetration per revolution (P), average

### Table 1 Main characteristics of tunnelling projects

| Projects | TBM-D (m) | Tunnel-L (km) | AD (km) | TBM type | Lithology |
|----------|-----------|---------------|---------|----------|-----------|
| Ghomrood water conveyance tunnel, Lots 3 and 4 (Iran) | 4.525 | 21.5 | 15 | Double shield (Wirth) | Limestone, Shale and Sandstone, Slate, Phyllite, Schist with quartzitic veins |
| Manapouri second tailrace tunnel (New Zealand) | 10.5 | 10 | 9.7 | Main beam open TBM (Robbins, Kvaerner-Markham) | Gneiss, Calc-silicate and quartzite and the intrusive rocks (Gabbro and Diorite) |
| Golab conveyance water tunnel (Iran) | 4.495 | 10 | 8 | Double shield (Wirth) | Periodic series of argillite shale and metamorphic sandstone, schist and amphibolite |
| Maroshi-Ruparel water supply tunnel Mumbai, (India) | 3.6 | 12.24 | 5.83 | Hard rock Gripper TBM (Wirth) | Fine compact basalt, Porphyritic basalt, Amygdaloidal basalt, Pyroclastic rocks (Tuff, Tuff breccia) and Intertrappeans (Shale)… |
| Zagros water conveyance tunnel, Lot 2 (Iran) | 6.73 | 26 | 15 | Double shield (Herrenknecht) | Limestone, Shale and Limy Shales |
| Karaj water conveyance tunnel, Lot 1 (Iran) | 4.65 | 15.9 | 8.7 | Double shield (Herrenknecht) | Tuffs, Shaly and Sandy Tuffs, Agglomerate |
| Lötschberg Base Tunnel (Switzerland) | 9.43 | 36.4 | 8.5 | Gripper TBM (Herrenknecht) | Crystalline Gneiss, Granodiorite and Granite, Granitic Gneiss, Amphibolite |
| Deep Tunnel Sewerage System (Singapore) | 4.82-4.45 | 38.5 | 22.2 | Hard rock shield TBM-EPB (Herrenknecht) | Bukit Timah granite (slightly to completely decomposed granite, silt, sandstone…)

1 Diameter; 2 Length; 3 Available Data; 4 Maroshi-Vakola; 5 Steg lateral adit, Main southern; 6 Only hard rock conditions (T05 and T06)
cutter load $F_n$, and field penetration index (FPI) have been estimated as follows:

$$\text{ROP} = \frac{L_b}{t_b}, \quad P = \frac{\text{ROP} \times 1000}{\text{RPM} \times 60},$$

$$\text{FPI} = \frac{F_n}{P}, \quad F_n = \left( T_h - F_f \right) / N_{\text{cutters}},$$

(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 (kN), $F_f$ is the estimated friction between the machine and the ground (kN) and $N_{\text{cutters}}$ is number of disc cutters installed on TBM cutterhead. To estimate the frictional force, machines were placed in two groups as reported in Table 1, gripper/open and double shield TBMs. In open type TBM the friction force which builds-up between machine and surrounding ground is much lower than shielded machines. In some cases, the front shoes of the machine are pressed against the walls and can impose a high pressure on the walls and thus high friction. However, for the most part, the friction of the machine can be included in the calculations by subtracting 20% machine weight from the total thrust force applied by the thrust cylinders (Delisio and Zhao 2014; Salimi et al. 2019a). For shielded TBMs the friction force builds-up between shield and surrounding ground, and hence is significantly higher than open machines, especially for double shield TBM. Previous studies have used 20% weight of the machine in non-squeezing grounds, or 20% of the rock load against the shield in low to medium level squeezing conditions. For highly squeezing conditions the value of friction forces could be higher than the machine thrust, leading to jamming. In such conditions, the use of an arbitrary percentage of weight of the machine is misleading. Further investigations are needed when shield TBMs are being utilized to assess the friction between shield and respected ground conditions (Salimi et al. 2019a; Salimi 2021). The general database structure is presented in Table 2.

One important issue to be noticed in this process is the missing data for different parameters in different records. Due to the difficulty of dealing with volumes of detailed data in several separate databases for different projects, it was necessary to reduce the number of data sets to a
manageable number. Heterogeneity of the data was also an issue which was caused using different protocols for recording TBM performance data for different tunnel job sites. Uniaxial Compressive Strength (UCS) is a commonly-used representative of rock strength in almost all of the TBM tunnel projects. Increasing in UCS causes a decrease in PR as noted by many researchers (Rostami 2013; Gong and Zhao 2009; Salimi et al. 2016). Rock mass behaviour is a function of rock material, frequency of joints, the existing joint conditions and surely influences the rock cutting by TBM. Joint spacing can be represented by RQD, joint frequency and volumetric joint count. In this study, based on availability of database, RQD and $J_v$ have been taken into account to represent joint frequency. As can be seen from Table 1, the database covers three main types of rock including, Igneous rocks (38%), Metamorphic rocks (31%), and Sedimentary rocks (31%) and the boring diameter varied between 3.6 & 10.5 m.

Various TBM performance indices have been proposed based on field penetration index (FPI), specific penetration (SP, inverse of FPI) and boreability index (BI, similar to FPI) by different researchers. Among these indices, FPI has got more attention compared the others. Considering the FPI as representative of TBM performance, it is commonly utilized to present the “Boreability” of the rock with changing geological/geotechnical circumstances which expresses the ease or difficulty of rock mass excavation by a TBM. The main advantage of FPI is that it allows penetration to be normalized for cutterload and thus it automatically takes care of machine thrust variations. FPI also has the capability to be used across different TBM diameters since it accounts for cutterhead RPM and number of cutters. Usually, stronger and less fractured rock masses are more difficult for cutting by disc cutters and boring by TBM require higher thrust levels to achieve a certain depth of penetration. So, higher values of FPI are low in such conditions (Hassanpour et al. 2011; Salimi et al. 2019a, b). Graphs presented in Fig. 2 show the histograms and distribution curves of different geological and TBM performance parameters recorded in the database based on different types of rock “G: Igneous rocks; M: Metamorphic rocks; S: Sedimentary rocks”. It should be noted that, $J_v$, is not available in all selected tunnel sections.

### 4 Developing New Models

In rock engineering practice, statistically based empirical equations have been extensively applied to predict target variables based on other operational or geological parameters. Empirical equations have great importance during the early stages of rock excavation and design works since they are more practical compared to extensive theoretical analyses. In the field of geomechanics, each rock type has its own texture, grain size, cementation and behaviour which affect the boreability and penetration rate of TBMs. Salimi et al. (2019a) considers Rock Type Code (RTC) as an input parameter in the proposed model to estimate TBM FPI. RTC was introduced by Laughton in 1990’s and have been used by Farrokh et al. (2012). Table 3 displays seven rock type categorizations. The first four classes are for “Sedimentary Rocks.” The fifth, sixth, and seventh classes are for “Metamorphic Rocks, Granite Rocks, and Volcanic Rocks”, respectively. It should be noted that, Gneiss (GN) is inherently metamorphic, but it is typically closer to granite in terms of its behaviour, especially where foliation is less pronounced. For this reason, it was categorized as GN in that analysis. To use rock type code as one of the selected input parameters to the model, code numbers including, 1 for G and GN, 2 for MV, 3 for SLK, 5 for C were employed (Salimi et al. 2019a, b).

According to the results of sensitivity analysis and parametric study of common models conducted by Fatemi et al. (2016) consideration of RTC has significant role to play in estimation of rock mass boreability. The same results have been found by Salimi (2021) and Salimi et al.
Generally speaking, rock texture, which consists of grains and matrix, directly correlated with the physical and mechanical properties of rock material and thus, rock drillability. For more illustration, Table 4 presents an example in which the basic RMR system is calculated for two rock masses in two different rock types. Overall, when comparing the most commonly used rock mass classification systems, the RMR classification is easiest to apply and shows better correlation with TBM performance, possibility due to the use of intact rock compressive strength as an input parameter. Moreover, RMR is frequently used in tunnel design process and reported from the logging of cores in the site investigation reports as well as back mapping of the tunnels. As such, input parameters of RMR system are often available for various projects. As can be seen from Table 4, despite the similar values found between two types of rocks (amphibolite which is an igneous crystalline rock and limestone that is a common sedimentary rock), the boreability of rock masses are different. Although, there are several factors which directly or indirectly can affect TBM performance, such as angle between tunnel axis and discontinuity planes $\alpha$
Introducing Tree-Based-Regression Models for Prediction of Hard Rock TBM Performance with…

(Alpha angle), crew experience, backup system and so on, but from geological points of view, it can be expected that, the differences are due to the rock texture and cementation. Therefore, as it anticipated, the boreability of rock masses are impacted by rock texture and cementation (Salimi et al. 2019b). Similar results have been also observed between other rock mass classifications such as Rock Structure Rating (RSR) by Wickham et al. (1972). In this study, given the available data from different rock types, performance prediction models are introduced based on similarities in rock textures categorized as G & GN; MV; SLK and C. Descriptive statistical distribution of variables in the database and input parameters for generated model for each rock type are summarized in Table 5. Figure 3 shows percentage distribution of different rock type categorization in this investigation.

As it was expected, higher value of FPI and its associated parameters can be found in hard massive rocks (“G & GN”), whereas lower values are attributed to soft/weak and more fractured rocks (“C”). This confirms the need for developing new models to include rock type categorization to reflect the similarities in their texture. It is worth to note that, in class G & GN, joint volumetric count (Jv) was available and showed better correlation with FPI compared to RQD. The reason for this phenomenon could be attributed to the limitation of RQD to be representative of rock mass fracturing degree in hard massive rocks since, it is an index with the maximum value of 100 which indicates the discontinuity spacing/frequency. Perhaps, this is why Gong and Zhao (2009) and Delisio and Zhao (2014) considered Jv being a better representative of joint frequency in their developed models (Salimi 2021).

To develop empirical models, the data in the databases was divided into two categories, i.e., training and testing/

| RMR input parameters           | Rock mass properties                                      | Rating |
|--------------------------------|-----------------------------------------------------------|--------|
| Rock mass 1 (Amphibolite)     | 170                                                       | 12     |
| Rock mass 2 (Limestone)       | 170                                                       | 12     |
| Uniaxial compressive strength (MPa) | 95                                                        | 20     |
| RQD (%)                        | 95                                                        | 20     |
| Spacing of discontinuities (m) | 0.8                                                       | 15     |
| Condition of discontinuities (Jc) | Slightly rough and moderately to highly weathered, wall rock surface separation < 1 mm | 20     |
| Groundwater general conditions | Damp                                                      | 10     |
| Basic RMR                      | 77                                                        | 77     |
| FPI (Field Penetration Index; kN/cutter/mm/rev) | 46.81                                                   | 38.48  |

Table 4 Comparison between different rock type categorizations with the same RMR value and TBM FPI

Table 5 Descriptive statistics of generated database based on different classes

| Class | N | Min | Max | Mean | Std. Deviation |
|-------|---|-----|-----|------|----------------|
| G & GN | 258 | 14.01 | 161.25 | 49.71 | 30.36 |
| MV    | 143 | 5.9  | 70.68 | 22.2  | 14.37 |
| SLK   | 139 | 9.52 | 45.44 | 19.42 | 6.44  |
| C     | 126 | 1.43 | 19.91 | 11.15 | 4.57  |

Fig. 3 Percentage distribution of different rock type codes
validation. This was done for developing and evaluating the proposed models. Many investigations recommended 20% of data for testing procedure. However, by partitioning the available data into 80% training and 20% testing, it can be expected that the number of samples which can be used for learning the model is drastically reduced and the results can depend on a particular random choice for the pair of (train, validation) sets. To address this issue, between 15 and 20% of the data was used for testing and validating the models. Further details of the dataset used in training of the models is presented in Table 6.

5 Machine Learning Methods for TBM Performance Models

Apart from empirical and theoretical models, the use of Machine Learning (ML) techniques has been used to examine relationship between FPI and geological parameters for each rock type. The background and pro/cons of using ML systems for TBM performance prediction has been discussed in previous publications (Zhang et al. 2020; Salimi et al. 2019a; Armaghani et al. 2017). In this context, decision trees are methods of a relatively easy application, being transparent and interpretable, since they allow obtaining patterns for a better explanation of a given phenomenon, showing the most important variables and their threshold values. This contribution reports the application of regression trees to assess the performance of TBM and offer graphs that can be used by others to reproduce the results and to predict TBM performance for future projects.

5.1 TBM Performance Prediction Models Using Regression Tree

One of the most popular techniques in data mining (analysis) is a DT (decision tree) in which a simple and comprehensible structure is used that can be utilized for classification, recognition, decision making as well as prediction of certain target parameters. There are several kinds of DT methods, among them, CART has been widely used with a high level of accuracy and performance for predicting problems in different engineering fields (Salimi 2021; Breiman et al. 1984; Tiryaki 2008). CART is a rule-based method introduced by Breiman
Fig. 4  Comparison between the calculated and predicted FPI based on rock type categorization via regression analysis for training and testing datasets
et al. (1984) and is based on whether the dependent variable is qualitative or quantitative; as such it can be categorized as a classification tree (CT) or regression tree (RT), respectively. This technique is recommended for use in situations where the form of the relationships between the dependent variable (response) and independent variables (predictors) is not exactly known before building a predictive model (Breiman et al. 1984). Furthermore, in CART analysis, there is no need to consider prior suppositions about the relationship between variables.

Given a set of samples, CART identifies one input variable and one break-point, before partitioning the samples into two child nodes. Starting from the entire set of available training samples (root node), recursive binary partition is performed for each node until no further split is possible or a certain terminating criterion is satisfied. At each node, best split is identified by exhaustive search, i.e., all potential splits on each input variable and each break-point are tested, and the one corresponding to the minimum deviations by, respectively predicting two child nodes of samples with their mean output variables is selected. After the tree growing procedure, typically an overly large tree is constructed, resulting in lack of model generalisation to unseen samples. A procedure of pruning is employed to remove sequentially the splits contributing insufficiently to training accuracy. The tree is pruned from the maximal sized tree, all the way back to the root node, resulting in a sequence of candidate trees. Each candidate tree is tested on an independent validation sample set and the one corresponding to the lowest prediction error is selected as the final tree (Breiman 2001; Wu et al. 2008; Yang et al. 2017; Salimi 2021).

To perform the RT, recursive partitioning and multiple regressions are carried out from the data-base. From the root node, the data splitting process in each internal node of a rule of the tree is repeated until a stop condition previously specified is reached. Stopping criteria can be defined for performing the regression tree algorithm—to keep the resultant tree from being too complicated for interpretation. The three main stopping criteria are: (1) a minimum number of observations in a node split; (2) the depth of the tree; and (3) the complexity parameter \( cp \). The complexity—measured by the coefficient \( cp \)—is the value analysed by the following rule. If the division in a specified node does not improve the fit of the model to data (on the value of set \( cp \)), then this division is ignored (Therneau and Atkinson 1997; Therneau et al. 2012; Tomczyk and Ewertowski 2013; Salimi 2021).

Alternatively, the optimal tree structure can be identified through ten-fold cross validation. In brief, in regression problems, where the output is a continuous number; CART can successfully predict the targeted outcome through using for example least absolute deviation (LAD) error. This study used CART RT models implemented in the R statistical computing software via “rpart library, party library and ggparty as well as mlbench library”. It is worth to note that, there is often a balance to be achieved in the depth and complexity of the tree to optimize predictive performance on some unseen data. Respecting to the tree depth, 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 5, 6, 7 and 8. To find this balance, one typically grows a very large tree as defined in the previous section and then prune it back to find an optimal subtree. The optimal subtree can be found using a cost complexity parameter that penalizes our objective function “least absolute deviation (LAD)” for the number of terminal nodes of the tree. For a given value of the smallest pruned tree that has the lowest penalized error, the optimum setting can be achieved. Smaller penalties tend to produce more complex models, which result in larger trees. Whereas larger penalties result in much smaller trees. Behind the scenes “CART RT via R computing software” is applied a range of cost complexity \( cp \) values to prune the tree considering the following parameters, 0.01, 0.001 and 0.0001 according to the literature review. To compare the error for each \( cp \) value, tenfold cross-validation performed so that the error associated with a given \( cp \) value is computed and the one with lowest RMSE or no significant differences selected. In brief, the determination of suitable combination of design parameters was taken in to account as a factor of paramount importance. This allowed the generation of operative robust tree-based-regression models with a high generalization capacity. Further information regarding the algorithm and its mathematical logic can be found in Breiman et al. (1984).

Similar data which has been employed in regression models are used for presenting a tree-based model for estimation of the TBM FPI in terms of rock type categorization in training and validation stages. Figures 5, 6, 7, 8 and Tables 9, 10, 11, 12 in Appendix A illustrate the preferable trees and present detailed information regarding the structure of the tree developed for TBM performance estimation based on rock type categorization, respectively.

Figure 9 displays the optimum tree size and the relationship between measured and predicted values obtained from the CART model for each rock type in training and testing stages is shown in Fig. 10. Furthermore, the relative variable importance for developed tree-based model for each rock type generated by CART is shown in Fig. 11. Relative variable importance standardizes the importance values for ease of interpretation which is defined as the percent improvement with respect to the most important predictor. An important variable is a variable that is used as a primary or surrogate splitter in the tree. The variable with the highest improvement score is set as the most important variable, and the other variables are ranked accordingly. As can be seen, the selected input parameters for model development, adequately reflect the effect of both intact and rock mass properties on TBM FPI. In general, it can be concluded that, the joint frequency on
Fig. 5 Regression tree developed for estimation of TBM FPI prediction “G & GN”
TBM performance has more significant role to play contrasting to intact rock properties such as UCS. This is in agreement with previous investigations (Hassanpour et al. 2011; Bruland 1998).

Also, the following formula can be used to calculate ROP (m/h) from the FPI predicted by developed models (Equations & Graphs):

$$\text{ROP (m/h)} = \frac{0.06 \times F_n \times \text{RPM}}{\text{FPI}}. \quad (6)$$

where $F_n$ is the average cutter load (kN/cutter), RPM is cutterhead speed (revolution per minute), and FPI is field penetration index (kN/cutter/rev/min).
6 Comparison of the Developed Models

The performance of the proposed models was evaluated according to statistical criteria including the root-mean-square-error (RMSE), mean-squared-error (MSE) and $R^2$ as follows:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y - y')^2}
\]  \hspace{1cm} (7)

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (y - y')^2}{\sum_{i=1}^{N} (y - \bar{y})^2}
\]  \hspace{1cm} (8)
where $y$, $y'$ and $\bar{y}$ are the measured, predicted and mean of the variable $y$, respectively; and $N$ is the total number of datasets. It is worth noting that the excellent model is considered where $R^2 = 1$, and RMSE as well as MSE equal to 0. The results of applying these models are summarized in Table 7. The results show that CART models offer higher accuracy in predicting FPI. As discussed before, the use of statistical analysis alone cannot offer satisfactory results and application of machine learning (ML) methods can improve the result of regression analysis, and in particular, tree-based modelling is an excellent alternative to regression analysis. In addition, it can handle data that are not normally distributed. This is a clear advantage in this field since most data do not follow normal distribution. Also, CART models are easier in visual representation, making a complex predictive model much easier to interpret. Additionally, decision trees are less likely to be influenced by outliers or missing values since it has no assumptions about space distributions and classifier structure (Salimi 2021).

6.1 Comparison of the Proposed and Existing Models

Among the different models which have been presented in the last decade, the model proposed by Hassanpour et al. (2011) was developed based on the FPI model and has similarities in input parameters and shows promising results compared to the common prediction models such as $Q_{TBM}$ (Hassanpour et al. 2016). The Formula and associated chart introduced by Hassanpour et al. (2011) is presented in Eq. 10, and is very applicable/constructive and reflect the practical approach in an early stage of tunnel design and construction. The model has been developed based on two commonly available inputs including, UCS and RQD which are most often available in any tunnelling
projects around the world. It is also worth to note that, in this study, the developed model for estimation of FPI in rock type G & GN, is based on UCS and Jv and a relevant formula that can convert Jv into RQD can be used to offer equivalency between the results of this study and that of Hassanpour (2011).

\[ FPI = \exp(0.008 \cdot UCS + 0.015 \cdot RQD + 1.384) \]  \hspace{5cm} (10)

Figure 12 shows the relationship between measured and predicted values obtained from the Hassanpour’s model for each rock type in testing stage. Since among the developed models in this investigation the CART model shows better results for each rock type categorizations, this model has been selected to be compared by the estimated FPI via Hassanpour’s model. For this purpose, variations of absolute error or \( E(\%) \) for each model and each rock type categorizations are calculated as follows:

\[ E(\%) = 100 \cdot \frac{|\text{Actual FPI} - \text{Estimated FPI}|}{\text{Actual FPI}} \]  \hspace{5cm} (11)

A summary of the statistical analysis performed on calculated rates and respective errors are presented in Table 8. As can be seen, the Hassanpour’s model provides better results in rock type categorization MV & SLK whereas shows higher error in other rock types including G & GN as well as C. The reason could be the consideration of RQD as joint frequency in hard massive rock masses where the RQD cannot represent the joint spacing adequately due to the limitation in maximum 100. Another cause can be related to the range of the complied data in the Hassanpour’s model. Additionally, the differences between geological characteristics of the sites used in development of the models have an impact on the accuracy of the models. However, it is worth to note that, the presented model by Hassanpour, shows promising/acceptable results when similar conditions such as range of UCS or disc cutter diameter (17") as well as geological characteristics are being applied in tunnel excavation with TBM. When choosing between empirical, theoretical, or ML models to predict TBM performance, one should pay attention to the application range of the model and geological conditions that the original model was based on.
Fig. 10 Comparison between the calculated and predicted FPI based on rock type categorization via regression tree (CART) for training and testing datasets
7 Model Limitations

The CART and empirical models developed in this investigation have some limitation in their application, similar to any other empirical models. The additional limitations are with respect to machine parameters, i.e., cutter diameter used on the machines, where the data used in this study are primarily from 432 mm (17”) disc cutters. Although the thrust force per cutter is a normalized value by cutter number in the developed model, the concentrated stress acting on the rock face at the contact point which initiates the fracture propagation is still greatly affected by cutter diameter and cutter tip width even if the force per cutter is the same as noted by Gong and Zhao (2009). Although, these machines have different diameters, they are similar in most of their specifications, particularly in cutterhead design and cutters arrangement i.e., the average spacing of disc cutters in all cutterheads was in the range of 60–90 mm. Consequently,

| Table 7 Performance indices for developed models | Train $R^2$ | RMSE | MSE | Test $R^2$ | RMSE | MSE |
|---------------------------------------------|------------|-----|-----|-------------|-----|-----|
| **Regression Model**                        |            |     |     |             |     |     |
| Class G & GN                                | 0.70       | 15.41 | 10.55 | 0.66        | 25.57 | 20.84 |
| Class MV                                    | 0.71       | 7.82 | 4.78 | 0.63        | 8.87  | 7.001|
| Class SLK                                   | 0.73       | 3.071 | 2.21 | 0.69        | 5.78  | 5.25 |
| Class C                                     | 0.72       | 2.43 | 1.93 | 0.66        | 3.36  | 2.92 |
| **CART Model**                              |            |     |     |             |     |     |
| Class G & GN                                | 0.88       | 10.89 | 7.95 | 0.84        | 12.91 | 10.12|
| Class MV                                    | 0.91       | 4.3 | 3.03 | 0.87        | 4.86  | 4.13 |
| Class SLK                                   | 0.92       | 2.11 | 1.65 | 0.86        | 2.86  | 2.3  |
| Class C                                     | 0.84       | 1.81 | 1.39 | 0.78        | 2.37  | 1.88 |

Fig. 11 Relative variable importance charts generated via “R” based on rock type categorization
when the machine parameters are changed (especially cutter diameter, cutter width and spacing), the model need to be used with consideration of the effects of these parameters. Perhaps existing models such as CSM formula which allows for variation of these parameters can be used for developing adjustment factors to extend the use of the proposed FPI numbers to the cases where disc diameter and tip width or spacing is outside the range of the available database. Furthermore, the estimated FPI and machine performance is not valid for mixed face or transitional working conditions; unstable blocky ground, and squeezing ground conditions.

**Fig. 12** Comparison between the calculated and predicted FPI based on rock type categorization via Hassanpour’s model in testing stage

**Table 8** Descriptive statistics of absolute errors (\(E\%\)) estimated for the prediction models “Hassanpour and CART” respecting to rock type categorizations

| Rock Type | Models    | N  | Range | Min  | Max  | Mean | Std. Deviation |
|-----------|-----------|----|-------|------|------|------|----------------|
| G & GN    | Hassanpour| 38 | 131.49| 0.55 | 122.47| 68.90| 32.37          |
|           | CART      | 38 | 92.26 | 0.31 | 69.03 | 20.05| 18.87          |
| MV        | Hassanpour| 23 | 136.42| 3.87 | 96.46 | 36.12| 39.41          |
|           | CART      | 23 | 40.13 | 0.58 | 40.68 | 12.96| 10.36          |
| SLK       | Hassanpour| 21 | 131.25| 2.36 | 77.53 | 31.612| 33.27         |
|           | CART      | 21 | 57.31 | 0.08 | 57.30 | 18.01| 15.32          |
| C         | Hassanpour| 16 | 115.83| 3.07 | 89.03 | 54.64| 29.65          |
|           | CART      | 16 | 49.86 | 2.24 | 42.35 | 23.14| 17.04          |
8 Discussion and Conclusion

In this study, a database of TBM field performance from eight tunnelling projects with total length of 92.93 km and boring diameter 3.6–10.5 m in different geological conditions was compiled and subjected to statistical analysis to derive empirical regression formulas for estimation of field penetration index (FPI). The data was subsequently analysed by machine learning (ML) methods and CART charts/graphs are offered for improving performance prediction for hard rock TBMs while incorporating rock type in the analysis.

Basically, given the ability of RT to perform recursive partitioning as an alternative method to the traditional multiple regressions, it was used for the analysis of a database of TBM field performance. The main advantage of CART (tree-based regression model) is that the end user does not need a computer code, nor have to be an expert in the field to use the model. In many applications, such as TBM performance prediction, CART offers better clarity of information, which makes the data understandable using graphic representations. It allows for selection of the most important variables, their threshold values, and finally implements proper rating and weights to each parameter based on the internal regression with the observed values. Moreover, the impact of each variable on the target can be obviously identified by the addressed tree structure in CART model.

The regression tree (RT) models could offer more accurate alternatives to the traditional multiple regression models. The proposed models in this study consisting of equations and graphs have been developed based on categorization of rock type, incorporating to its similarities in rock texture (cementation, grain size and shape). The results show that incorporating rock type is very useful when corresponding categories were used as input parameters for TBM performance complements the use of the most influential parameters including UCS (intact rock strength), Jv or RQD (degree of fracturing of the rock mass). The results also indicate that CART model outperforming the regression models with typical $R^2$ close to 90%, as compared to the multivariable regression equations that offer $R^2$ in the mid 70% range. The suggested formulas and graphs in this study that allows form incorporation of rock type, offer more accurate results compared to the previous generalized models based on CART.

### Appendix A

See Tables 9, 10, 11, 12.
Table 9 (continued)

| Node | Frequency | Split variable | Value | Predicted |
|------|-----------|----------------|-------|-----------|
| 50   | 4         | UCS            | \leq 156.05 | 59.8358   |
| 51   | 26        | UCS            | > 156.05  | 77.8175   |
| 52   | 22        | UCS            | \leq 169.65 | 33.6558   |
| 53   | 3         | UCS            | > 169.65  | 44.6687   |
| 54   | 3         | Jv             | \leq 3.765 | 95.25     |
| 55   | 2         | Jv             | > 3.765   | 73.998    |
| 56   | 11        | UCS            | \leq 67.75 | 26.033    |
| 57   | 3         | UCS            | > 67.75   | 32.0575   |
| 58   | 27        | Jv             | \leq 15.625 | 23.723    |
| 59   | 6         | Jv             | > 15.625  | 20.6235   |
| 60   | 39        | Jv             | \leq 9.845 | 42.5791   |
| 61   | 19        | Jv             | > 9.845   | 31.0365   |
| 62   | 14        | UCS            | \leq 185.805 | 47.4318  |
| 63   | 2         | UCS            | > 185.805 | 108.175   |
| 64   | 18        | UCS            | \leq 172.2 | 25.5007   |
| 65   | 5         | UCS            | > 172.2   | 35.5479   |
| 66   | 4         | UCS            | \leq 177.77 | 39.338    |
| 67   | 2         | UCS            | > 177.77  | 52.365    |

Table 10 Information related to each node in the CART model “MV”

| Node | Frequency | Split variable | Value | Predicted |
|------|-----------|----------------|-------|-----------|
| 1    | 143       | –              | –     | 22.2055   |
| 2    | 93        | UCS            | \leq 100.15 | 14.2415  |
| 3    | 50        | UCS            | > 100.15 | 37.0186   |
| 4    | 91        | RQD            | \leq 90.125 | 13.754   |
| 5    | 2         | RQD            | > 90.125 | 36.4218   |
| 6    | 38        | RQD            | \leq 95.7188 | 32.4683  |
| 7    | 12        | RQD            | > 95.7188 | 51.4277   |
| 8    | 34        | UCS            | \leq 56.5   | 10.1646   |
| 9    | 57        | UCS            | > 56.5    | 15.895    |
| 10   | 17        | RQD            | \leq 77.9375 | 27.4687  |
| 11   | 21        | RQD            | > 77.9375 | 36.5156   |
| 12   | 4         | UCS            | \leq 146.5  | 34.8888   |
| 13   | 8         | UCS            | > 146.5   | 59.6971   |
| 14   | 16        | UCS            | \leq 49    | 8.15823   |
| 15   | 18        | UCS            | > 49      | 11.948    |
| 16   | 21        | RQD            | \leq 65.25 | 13.5573   |
| 17   | 36        | RQD            | > 65.25   | 17.2587   |
| 18   | 7         | RQD            | \leq 67.5  | 25.2407   |
| 19   | 10        | RQD            | > 67.5    | 29.0283   |
| 20   | 2         | RQD            | \leq 79   | 35.1562   |
| 21   | 19        | RQD            | > 79      | 49.43     |
| 22   | 6         | UCS            | \leq 188   | 56.9795   |
| 23   | 2         | UCS            | > 188     | 67.85     |
| 24   | 10        | UCS            | \leq 38.5  | 7.63403   |
| 25   | 6         | UCS            | > 38.5    | 9.0319    |
| 26   | 16        | RQD            | \leq 82.5  | 11.2877   |
| 27   | 2         | RQD            | > 82.5    | 17.23     |
| 28   | 5         | RQD            | \leq 42.5  | 12.9798   |
| 29   | 16        | RQD            | > 42.5    | 15.4054   |
| 30   | 8         | UCS            | \leq 66.5  | 15.3481   |
| 31   | 28        | UCS            | > 66.5    | 17.8046   |
| 32   | 3         | UCS            | \leq 145.65 | 23.27    |
| 33   | 4         | UCS            | > 145.65  | 27.8683   |
| 34   | 3         | UCS            | \leq 130.55 | 23.5379  |
| 35   | 7         | UCS            | > 130.55  | 31.3813   |
| 36   | 6         | UCS            | \leq 137.05 | 28.8667  |
| 37   | 13        | UCS            | > 137.05  | 38.059    |
| 38   | 4         | UCS            | \leq 167.5 | 49.9116   |
| 39   | 2         | UCS            | > 167.5   | 60.5135   |
### Table 11: Information related to each node in the CART model “SLK”

| Node | Frequency | Split variable | Value | Predicted |
|------|-----------|----------------|--------|-----------|
| 1    | 139       | –              | –      | 19.4217   |
| 2    | 112       | RQD            | <= 83.75 | 17.0426   |
| 3    | 27        | RQD            | > 83.75 | 29.2904   |
| 4    | 54        | RQD            | <= 69.8125 | 15.3767 |
| 5    | 58        | RQD            | > 69.8125 | 18.5937   |
| 6    | 24        | UCS            | <= 167.5 | 27.5536   |
| 7    | 3         | UCS            | > 167.5 | 43.1847   |
| 8    | 23        | RQD            | <= 52.5  | 14.0504   |
| 9    | 31        | RQD            | > 52.5   | 16.3608   |
| 10   | 39        | UCS            | <= 85    | 17.653    |
| 11   | 19        | UCS            | > 85     | 20.5246   |
| 12   | 7         | UCS            | <= 104   | 23.0305   |
| 13   | 17        | UCS            | > 104    | 29.4161   |
| 14   | 21        | UCS            | <= 65    | 13.8659   |
| 15   | 2         | UCS            | > 65     | 15.9874   |
| 16   | 5         | RQD            | <= 58.625 | 16.0457  |
| 17   | 26        | RQD            | > 58.625 | 18.0025   |
| 18   | 36        | UCS            | <= 77.5  | 13.8759   |
| 19   | 3         | UCS            | > 77.5   | 17.9678   |
| 20   | 14        | UCS            | <= 119.5 | 19.9508   |
| 21   | 5         | UCS            | > 119.5  | 22.1312   |
| 22   | 3         | RQD            | <= 92.5  | 19.3573   |
| 23   | 4         | RQD            | > 92.5   | 25.7855   |
| 24   | 9         | UCS            | <= 131.75 | 27.7518  |
| 25   | 8         | UCS            | > 131.75 | 31.2884   |
| 26   | 15        | UCS            | <= 56.75 | 12.6534   |
| 27   | 6         | UCS            | > 56.75  | 14.3509   |
| 28   | 3         | UCS            | <= 49.95 | 15.6497   |
| 29   | 2         | UCS            | > 49.95  | 21.523    |
| 30   | 22        | UCS            | <= 68.75 | 15.6601   |
| 31   | 4         | UCS            | > 68.75  | 18.1667   |
| 32   | 17        | UCS            | <= 61.25 | 17.2502   |
| 33   | 19        | UCS            | > 61.25  | 18.7698   |
| 34   | 3         | UCS            | <= 116.495 | 19.7252 |
| 35   | 11        | UCS            | > 116.495 | 20.6777  |
| 36   | 2         | RQD            | <= 91.25 | 21.095    |
| 37   | 7         | RQD            | > 91.25  | 29.6538   |
| 38   | 2         | RQD            | <= 85.5  | 28.9541   |
| 39   | 6         | RQD            | > 85.5   | 32.0665   |

### Table 12: Information related to each node in the CART model “C”

| Node | Frequency | Split variable | Value | Predicted |
|------|-----------|----------------|--------|-----------|
| 1    | 126       | –              | –      | 11.1493   |
| 2    | 39        | RQD            | <= 31.25 | 5.64606  |
| 3    | 87        | RQD            | > 31.25 | 13.6162   |
| 4    | 26        | RQD            | <= 26.75 | 4.42851  |
| 5    | 13        | RQD            | > 26.75  | 8.08116   |
| 6    | 47        | RQD            | <= 63.5  | 12.4976   |
| 7    | 40        | RQD            | > 63.5   | 14.9306   |
| 8    | 2         | RQD            | <= 12.5  | 3.54738   |
| 9    | 24        | RQD            | > 12.5   | 4.50193   |
| 10   | 11        | UCS            | <= 38.5  | 7.69153   |
| 11   | 2         | UCS            | > 38.5   | 10.2241   |
| 12   | 6         | UCS            | <= 27.5  | 11.0229   |
| 13   | 41        | UCS            | > 27.5   | 12.7134   |
| 14   | 34        | UCS            | <= 89    | 14.6772   |
| 15   | 6         | UCS            | > 89     | 16.3668   |
| 16   | 2         | RQD            | <= 17.5  | 4.39395   |
| 17   | 22        | RQD            | > 17.5   | 5.68982   |
| 18   | 9         | UCS            | <= 22.25 | 6.26941   |
| 19   | 2         | UCS            | > 22.25  | 8.00756   |
| 20   | 4         | UCS            | <= 23.7  | 8.18883   |
| 21   | 2         | UCS            | > 23.7   | 12.4399   |
| 22   | 9         | UCS            | <= 57.5  | 12.4181   |
| 23   | 14        | UCS            | > 57.5   | 13.2831   |
| 24   | 12        | UCS            | <= 42.5  | 14.0875   |
| 25   | 22        | UCS            | > 42.5   | 15.7581   |
| 26   | 2         | UCS            | <= 94    | 15.93     |
| 27   | 4         | UCS            | > 94     | 16.5852   |
| 28   | 20        | UCS            | <= 35    | 4.324     |
| 29   | 2         | UCS            | > 35     | 5.90336   |
| 30   | 6         | RQD            | <= 36.25 | 12.2105   |
| 31   | 21        | RQD            | > 36.25  | 13.1447   |
| 32   | 10        | UCS            | <= 55.9  | 12.6175   |
| 33   | 4         | UCS            | > 55.9   | 13.5493   |
| 34   | 9         | RQD            | <= 73.5  | 15.1386   |
| 35   | 3         | RQD            | > 73.5   | 17.6167   |
| 36   | 9         | RQD            | <= 77.5  | 12.5765   |
| 37   | 13        | RQD            | > 77.5   | 15.1336   |
Funding  Open Access funding enabled and organized by Projekt DEAL.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alvarez Grima M, Bruins PA, Verhoef PWN (2000) Modelling tunnel boring machine performance by Neuro-Fuzzy method. Int J Tunnell Undergr Space Technol 15(3):259–269

Armaghani DJ, Mohamad ET, Narayanasamy MS, Narita N, Yagiz S (2017) Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. Tunnell Undergr Space Technol 63:29–43

Benardos A, Kaliampakos DC (2004) Modelling TBM performance with artificial neural networks. Tunnell Undergr Space Technol 19(6):597–605

Blindheim OT (1979) Boreability predictions for tunnelling. Ph.D. thesis. Department of geological engineering. The Norwegian Institute of Technology

Breiman L (2001) Statistical modeling: the two cultures. Stat Sci 16(3):199–231

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsorth, Belmont

Bruland A (1998) Hard rock tunnel boring: vol 1–10, Ph.D. thesis, Trondheim

Coimbra R, Rodriguez-Galiano V, Olóriz F, Chica-Olmo M (2014) Regression trees for modelling geochemical data—an application to Late Jurassic carbonates (Ammonitico Rosso). Comput Geosci 73:198–207

Company SCE (2004) Geological and Engineering Geological Report for Ghomrod Water Conveyance Tunnel Project (Lots 3 & 4) (Unpublished report)

Deere DW, Keis S, Watts C (2004) The Manapouri Tailrace Tunnel No. 2 construction—a very large TBM Tunnel in very strong rock. In: Ozdemir L (ed) North American Tunnelling, pp 421–432

Delisio A (2014) Field and numerical studies on the causes and effects of blocky rock conditions on TBM tunneling. Ph.D. Thesis. École Polytechnique Fédérale de Lausanne, Laboratoire de Mécanique des Roches (Unpublished)

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

Farrokh E, Rostami J, Laughton C (2012) Study of various models for estimation of penetration rate of hard rock TBMs. Tunnell Undergr Space Technol 30:110–123. https://doi.org/10.1016/j.tust.2012.02.012

Fatemi SA, Ahmadi M, Rostami J (2016) Evaluation of TBM performance prediction models and sensitivity analysis of input parameters. Int J Bull Eng Geol Environ. https://doi.org/10.1007/s10064-016-0967-2

Ghasemi E, Yagiz S, Ataei M (2014) Predicting penetration rate of hard rock tunnel boring machine using fuzzy logic. Bull Eng Geol Environ 73:23–35

Gong QM (2005) Development of rock mass characteristics model for TBM penetration rate prediction, Ph.D. thesis, School of civil and environmental engineering, Nanyang Technological University, Singapore (Unpublished)

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

Goodarzi S, Hassanpour J, Yagiz S, Rostami J (2021) Predicting TBM performance in soft sedimentary rocks, Case study of Zagros Mountains Water Tunnel Projects. Tunnell Undergr Space Technol. https://doi.org/10.1016/j.tust.2020.103705

Hassanpour J, Rostami J, Khamenechiyan M, Brunlad A (2009) Developing new equations for TBM performance prediction in carbon-ate-argillaceous rocks: a case history of Nowsoon water conveyance tunnel. Int J Geomech Geoenviron 4:287–297

Hassanpour J, Rostami J, Khamenechiyan M, Brunlad A, Tavakoli HR (2010) TBM performance analysis in pyroclastic rocks, a case history of Karaj Water Conveyance Tunnel (KWCFT). J Rock Mech Eng 4:427–445

Hassanpour J, Rostami J, Zhao J (2011) A new hard rock TBM performance prediction model for project planning. Tunnell Undergr Space Technol 26:595–603

Hassanpour J, Ghaedi Vanani AA, Rostami J, Cheshomi A (2016) Evaluation of common TBM performance prediction models based on field data from the second lot of Zagros water conveyance tunnel (ZWCT2). Tunnell Undergr Space Technol 52:147–1456

Hassanpour J, Firouzei Y, Hajipour G (2021) Actual performance analysis of a double shield TBM through sedimentary and low to medium grade metamorphic rocks of Ghomrod water conveyance tunnel project (lots 3 and 4). Bull Eng Geol Env. https://doi.org/10.1007/s10064-020-01947-z

Imensazan consulting engineers (ICE) (2009) Gobal water transfer tunnel engineering geology studies. (Unpublished)

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

Khadem Hamidi J, Shahriar K, Rezai B, Rostami J (2010) Performance prediction of hard rock TBM using Rock Mass Rating (RMR) system. Int J Tunnell Undergr Space Technol. https://doi.org/10.1016/j.tust.2010.01.008

Koojiapooroosh M, Tootoonchi H, Armaghani J, Mohamad ET, Hedayat A (2019) Application of deep neural networks in predicting the penetration rate of tunnel boring machines. Bull Eng Geol Environ 78:6347–6360. https://doi.org/10.1007/s10064-019-01538-7

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

Marcher T, Erharder G, Winkler M (2020) Machine Learning in tunneling—capabilities and challenges. Int J Geomech Tunnell. https://doi.org/10.1002/geot.202000001

Morgenroth J, Khan UT, Perras MA (2019) An overview of opportunities for machine learning methods in underground rock engineering design. MDPI Geosci 9(12):504. https://doi.org/10.3390/geosciences9120504

Nelson P, l’OuRourke 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, pp 227–237
Introducing Tree-Based-Regression Models for Prediction of Hard Rock TBM Performance with…

Pourhashemi SM, Ahangari K, Hassanpour J, Eftekhari SM (2021) TBM performance analysis in very strong and massive rocks; Case study: Kerman water conveyance tunnel project, Iran. Int J Geomech Geoeng. https://doi.org/10.1080/17486025.2021.1912410

Ramezanadeh A (2005) Performance analysis and development of new models for performance prediction of hard rock TBMs in rock mass. Ph.D. thesis. INSA, Lyon

Rostami J (1997) Development of a force estimation model for rock fragmentation with disc cutters through theoretical modelling and physical measurement of crushed zone pressure. Ph.D. thesis. Colorado School of Mines, Golden, Colorado

Rostami J (2013) Study of pressure distribution within the crushed zone in the contact area between rock and disc cutters. Int J Rock Mech Min Sci 57:172–186

Salimi A (2021) Investigation and Evaluation of Rock Mass Characteristics for Development of New TBM Performance Prediction Model in Hard Rock Conditions. Ph.D. thesis. Institute of Geotechnical Engineering (IGS), University of Stuttgart, Stuttgart (Published).

Salimi A, Rostami J, Moormann C, Delisio A (2016) Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs. J Tunnell under Space Technol 58:236–246

Salimi A, Rostami J, Moormann C (2019a) Application of rock mass classification systems for performance estimation of rock TBMs using regression tree and artificial intelligence algorithms. Int J Tunnell Undergr Space Technology. https://doi.org/10.1016/j.tust.2019.103046

Salimi A, Rostami J, Moormann C (2019b) Development of a New Models for Prediction of TBM Performance in Hard Rock Condition Based on Rock Type. In: TBM-DiGs:2019b-Tunnel Boring Machines in Difficult Grounds, 4th international conference. Colorado School of Mines, Golden

Therneau TM, Atkinson B, Ripley MB (2012) Package ‘rpart’

Tiryaki B (2008) Predicting intact rock strength for mechanical excavation using multivariate statistics, artificial neural networks, and regression trees. Eng Geol. https://doi.org/10.1016/j.enggeo.2008.02.003

Tomczyk MA, Ewertowski M (2013) Planning of recreational trails in protected areas: application of regression tree analysis and geographic information systems. Appl Geogr 40:129–139

U.R.S. Company (2003) Manapouri Power Station Second Tailrace Tunnel Engineering Geological Construction Report, Prepared for Meridian Energy Ltd. (Unpublished report)

Wickham GE, Tiedemann HR, Skinner EH (1972) Support determination based on geologic predictions. In: Proc. North American rapid excavation and Tunnelling conference: Chicago, New York: Society of Mining Engineering of the American Institute of Mining, Metallurgical and Petroleum Engineers, pp 43–64

Wu X, Kumar V, Quinlan RJ, Ghosh J, Yang Q, Motoda H et al (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14(1):1–37

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 tunnelling machines. Ph.D. thesis. Department of Mining and Earth Systems Engineering, Colorado School of Mines, Golden

Yagiz S, Gokceoglu C, Sezer E, Iplikci S (2009) Application of two non-linear prediction tools to the estimation of tunnel boring machine performance. Eng Appl Artif Intell 22:808–814

Yang L, Liu S, Tsoka S, Papageorgiou LG (2017) A regression tree approach using mathematical programming. Int J Expert Syst Appl 78:347–357

Zhang Q, Hu W, Liu Z, Tan J (2020) TBM performance prediction with Bayesian optimization and automated machine learning. Tunnell Undergr Space Technol. https://doi.org/10.1016/j.tust.2020.103493

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