The 2025 World Congress on
Advances in Structural Engineering and Mechanics (ASEM25)
BEXCO, Busan, Korea, August 11-14, 2025

Data-Driven Modelling of EPBM Advancement Rates Using Various ML Algorithms in Weathered Rock Condition

*Aman Kumar, Deepanshu Shirole

Department of Civil Engineering, Indian Institute of Technology, Delhi, India

ABSTRACT

This paper aims to establish data-driven models for the prediction of advance rate (AR) of earth pressure balance-tunnel boring machine (EPB-TBM) in weathered rocks. Specifically, for this purpose, linear regression (LR), support vector regression (SVR), and random forest (RF) machine learning algorithms have been employed, and their efficacy in predicting EPB-TBM AR has been evaluated. For this, a representative dataset of a metro project tunnel passing through weathered rock was analyzed. In particular, for this, initially the geological data in the nearfield of the tunnel was statistically interpolated using the Kriging technique, which enabled objective sub-surface profiling. Following this, rock mass rating (RMR), compressive strength of intact rock (UCS), face pressure, main drive speed, main drive torque, advance thrust force, cutterhead contact force, average chamber pressure, screw conveyor torque, polymer flow rate and foam flow rates were used as input features in the linear regression (LR), support vector regression (SVR) and random forest (RF) machine learning (ML) models. Specifically, a 90:10 ratio of train-test split was used for training and testing of the models for the cumulative 1201 datasets. The analysis showed that LR, SVR, and RF are able to predict the AR of the EPB-TBM with a coefficient of determination (R^2) of 80%, 89%, and 88%, respectively. Polymer flow rate, main drive torque, and foam flow rate have been found to be significant parameters of first degree of importance, followed by face pressure, average chamber pressure, advance thrust force, and cutterhead contact force. From the geology of the intact rock, the RMR of the rock mass is found to be a better parameter than the UCS of the intact rock. A partial dependence plot of the individual parameter has been plotted; the advance thrust force has an unchangeable influence on advancement rates, while main drive speed and main drive torque are found to be inversely and directly related to the advancement rate, respectively. Bound optimization has been performed to obtain the optimum set of input values that will maximize the output, i.e., advance rate (AR). RMR, UCS of the intact rock and face pressure have been kept as static bound, while average earth pressure, main drive torque, screw conveyor torque and cutterhead contact force have been kept as relative bound, and main drive speed, advance thrust force, polymer

flow rate, and foam flow rate have been kept as absolute bound. A representative value of static bound has been kept to perform the constrained inverse modelling.

1. INTRODUCTION

EPBM has been widely used in the urban environment, specifically for metro tunneling. EPBM balances the earth pressure by employing face pressure, utilizing excavated material and additives at the soil face, and creates a balanced environment for the mining operation. The earth pressure balancing often presents several challenges, making it difficult for the study purpose. The understanding of the advancement rate of EPBM depends upon various geological, mechanical, and external factors. TBM performance prediction modelling in the beginning is prominently done by theoretical models, which are useful in explaining the nature of physical forces acting in excavation, but do not serve the purpose of performance prediction well. Later, scholars used empirical models to explain the performance prediction, which also have significant insufficiencies. Recently, with the advancement of computing tools and the availability of enormous TBM recorded sensor data, it has become fascinating to use ML algorithms on actual cutting behavior of the TBM to model its performance prediction. Although several scholars have used different ML algorithms to model the EPB-TBM advancement behavior, however, there is no established model for the prediction of the EPB-TBM advancement rates as of now. Nonetheless, sufficient work has been done on the hard rock TBM, and characterization of its advancement rates has been quite well understood.

Comparison of various hard rock TBM models as summarized in Table 1, presents a comparative summary of five research studies that focus on predicting the rate of penetration (ROP) of Tunnel Boring Machines (TBMs) in hard rock conditions using various machine learning and optimization techniques. These studies, conducted between 2004 and 2014, highlight the evolution of modeling strategies, tools used, datasets, and performance metrics for improving TBM efficiency.

The key focus of each study is to model and predict how quickly a TBM can progress through rock based on various input parameters. To achieve this, each study employs different artificial intelligence or machine learning techniques. One of the earliest studies by (Benardos, 2004) uses a simple Neural Network, while subsequent research like (Yagiz, June 2009), (Mokhtari S, 2020)and (Armaghani, 2017) employ Artificial Neural Networks (ANN) and advanced hybrid methods involving Imperialist Competitive Algorithm (ICA) and Particle Swarm Optimization (PSO). (Mahdevari, 2014) takes a different approach by using Support Vector Regression (SVR), and (Yagiz S, 2011) uses PSO as a standalone optimization tool. The studies (Yagiz, June 2009) have been conducted on different tunnel sites across the world. A critical component of these models is the selection of input parameters—the geological and operational variables used to train and test the models. These inputs range from rock and geological properties such as Rock Quality Designation (RQD), Uniaxial Compressive Strength (UCS), Brazilian Tensile Strength (BTS), Brittleness Index (BI), and Quartz Content, to structural factors like the angle between weak planes and tunneling direction. Operational variables such as thrust force, cutterhead torque, revolution per minute (RPM), and specific energy consumption are also included in more advanced models. This diverse set of parameters allows the models to capture the complex interactions between geology and machine performance.

Model accuracy is expressed using either percentage error or the coefficient of determination (R^2). Higher R^2 values indicate better model performance. (Benardos, 2004) reported a modest 8% error, while later models showed significant improvements. For example, (Yagiz, June 2009) achieved $R^2 > 80$, (Armaghani, 2017) achieved $R^2 > 95$, and (Mahdevari, 2014) reported $R^2 = 94.9\%$, indicating a strong predictive performance of their SVR model.

Table 1. Comparison of various hard rock TBM models.

Hard rock TBM rate of penetration performance model					
Models	(Benardos, 2004)	(Yagiz, June 2009)	(Armaghani, 2017)	(Mahdevari, 2014)	(Yagiz S, 2011)
Type of machine	Hard Rock TBM	Hard Rock TBM	Hard rock TBM	Hard Rock TBM	Hard Rock TBM
Tools used	Neural Network	Artificial Neural Network	ANN, ICA & PSO, and a Hybrid of them	Support Vector Regression	Particle Swarm Optimization
Project Site	Athens	Queen's Water Tunnel, New York	Water tunnel, Pahang, Malayasia	Queen's Water Tunnel, New York	Queens Water Tunnel No. 3, New York
Input Parameters	Degree of weathering, Overload Factor, RMR, RQD, UCS, Tunnel Overburden, Ground Water Table, Rock Permeability, TBM Operations	UCS, BTS, BI, Distance, DPW, Angle alpha between weak plane and tunneling direction	UCS, BTS, RQD, RMR, Weathered Zone, Quartz content, Thrust force per Cutter, RPM	UCS, BTS, BI, Distance, DPW, Angle alpha between weak plane and tunneling direction, Thrust force, Cutterhead power and torque, Specific Energy	UCS, BTS, BI, DPW, Alpha, Measured ROP
% COD	$R^2 > 92$	$R^2 > 80$	$R^2 > 95$	$R^2 > 94.9$	$R^2 > 73$

The "EPBM Rate of Penetration Performance Model," as summarized in Table 2, provides a comparative overview of four studies focusing on modeling the Rate of Penetration (ROP) for earth pressure balance tunnel boring machine (EPB-TBM). The objective across all the studies is to predict the ROP based on various mechanical and operational parameters using machine learning and statistical techniques. Each study differs in terms of modeling tools, project locations, and selected input variables.

In the first study in Table 2 (Mokhtari S, 2020), employed Support Vector Regression (SVR) to model the ROP using data from the Northgate Link Tunnel in Seattle. The model achieved a high predictive performance with an R^2 value greater than 88. Input parameters included a comprehensive list of machine dynamics and environmental conditions, such as screw conveyor torque (1 and 2), cutterhead torque, net thrust, foam flow rates, tail shield annular pressure, and clearances around the tail shield (left, right, top, and bottom). The inclusion of parameters like front body rolling and stroke difference left-to-right indicates an attempt to capture nuanced mechanical behavior during excavation.

The subsequent model by (Mokhtari S., 2020) used the same tunnel project site but shifted to a more interpretable modeling technique, elastic net regression, which blends both Lasso and Ridge regression properties. While the R^2 score dropped slightly to greater than 75, the choice of model emphasizes interpretability and feature selection efficiency. Input parameters largely overlap with the previous study but focus more on internal machine metrics like cutterhead RPM, depth below ground surface, and depth below groundwater table, showing a shift towards variables that reflect more direct control over boring conditions.

Table 2. Comparison of various EPBM models

EPB-TBM rate of penetration performance model				
Models	(Mokhtari S, 2020)	(Mokhtari S., 2020)	(Elbaz K., 2019)	(Gao X., 2019)
Types	EPBM	EPBM	EPBM	EPBM
Tools used	Support Vector Regression	Elastic Net	ANFIS	RNN
Project Site	Northgate Link Tunnel, Seattle	Northgate Link Tunnel, Seattle	China Metro Tunnel Project	Subway Tunnel
Input feature	Screw-1,2 Torque, Cutterhead Torque, Net Thrust, Foam Flow Rates, Stroke Difference, Front body rolling, Average stroke, TS Annular Pressure, Left, right, top, and bottom tail clearance	Cutterhead RPM, Net thrust, Foam flow rates, cutterhead torque, screw conveyor torque1 72, Depth below ground surface and water	Cutterhead RPM, Cutterhead Torque, Screw Conveyor rotation speed	Cutterhead torque, thrust force, Chamber pressure

% COD	R^2> 88	R^2> 75	R^2> 85	R^2> 73

In contrast, (Elbaz K., 2019) adopted a soft computing approach using ANFIS (Adaptive Neuro-Fuzzy Inference System) to predict ROP during the China Metro Tunnel Project. ANFIS is particularly effective for modeling systems with complex, nonlinear relationships and uncertain inputs. The model utilized fewer but highly relevant parameters such as cutterhead RPM, torque, and screw conveyor rotation speed. Despite the smaller input set, the model performed well with an R^2 value exceeding 85, suggesting that ANFIS was able to extract meaningful relationships from the data.

(Gao X., 2019) explored a variety of combinations of algorithms in Recurrent Neural Network for modeling EPBM performance during a Subway Tunnel project with success in R^2 score of about 73%. The input parameters include cutterhead torque, thrust force, and chamber pressure—all critical to EPBM operations in pressurized ground conditions. These parameters reflect direct mechanical stressors influencing penetration rates.

For this study, EPB-TBM data of an underground metro tunnel project of a city in the southern part of India have been used. Machine learning algorithms, from simpler to complex, have been utilized, such as linear regression, support vector regression, and random forest. The flow chart for the data-driven modelling, similar to (Morshedlou A., 2024) is shown in Figure 1 below.

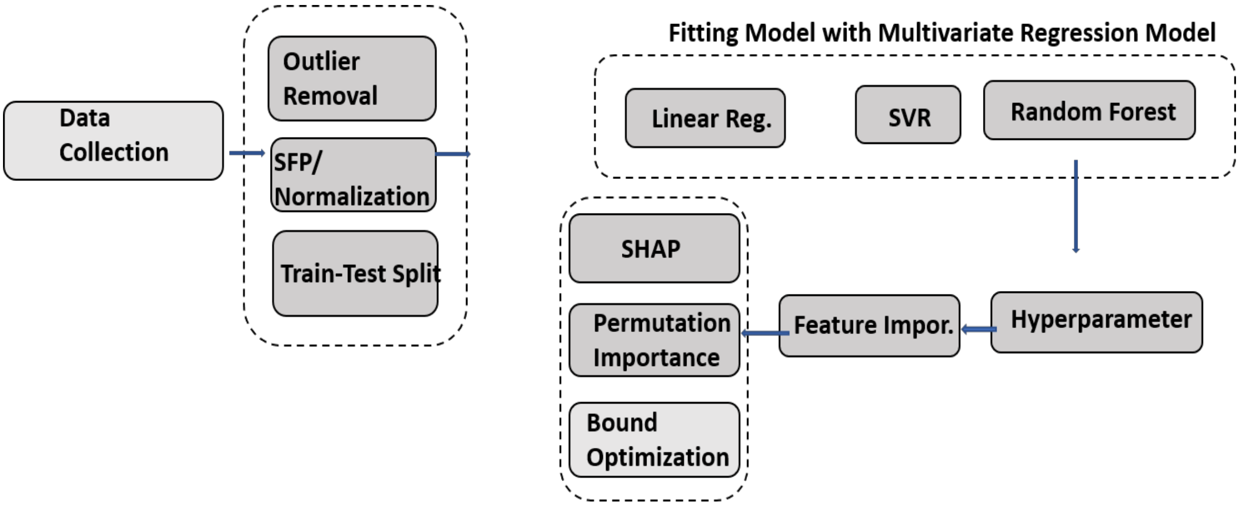


Figure 1. Flow chart showing data-driven modelling. (Morshedlou A., 2024)

The features used for this study are rock mass rating (RMR), compressive strength of intact rock (UCS), face pressure, main drive speed, main drive torque, advance thrust force, cutterhead contact force, average chamber pressure, screw conveyor torque, polymer flow rate and foam flow rates. Feature matrices are detected for outliers and subsequently removed, followed by normalization and train-test splitting of the datasets. LR, SVR, and RF models are trained and tested. Hyperparameter tuning has been performed, followed by the selection of significant features. SHAP and permutation importance analysis are performed. Bound optimization has been carried out on a sample. The paper is organized as follows. Project details and sub-surface profiling have been discussed in section 2, followed by EPBM configuration details and data collection and cleaning is discussed. Data-driven modelling has been discussed in section 3. Parameter selection, training, testing of the model, and regularization have been discussed in the same section. The discussion part, as section 4, further provides findings of the model.

2. PROJECT BACKGROUND AND SUB-SURFACE PROFILING

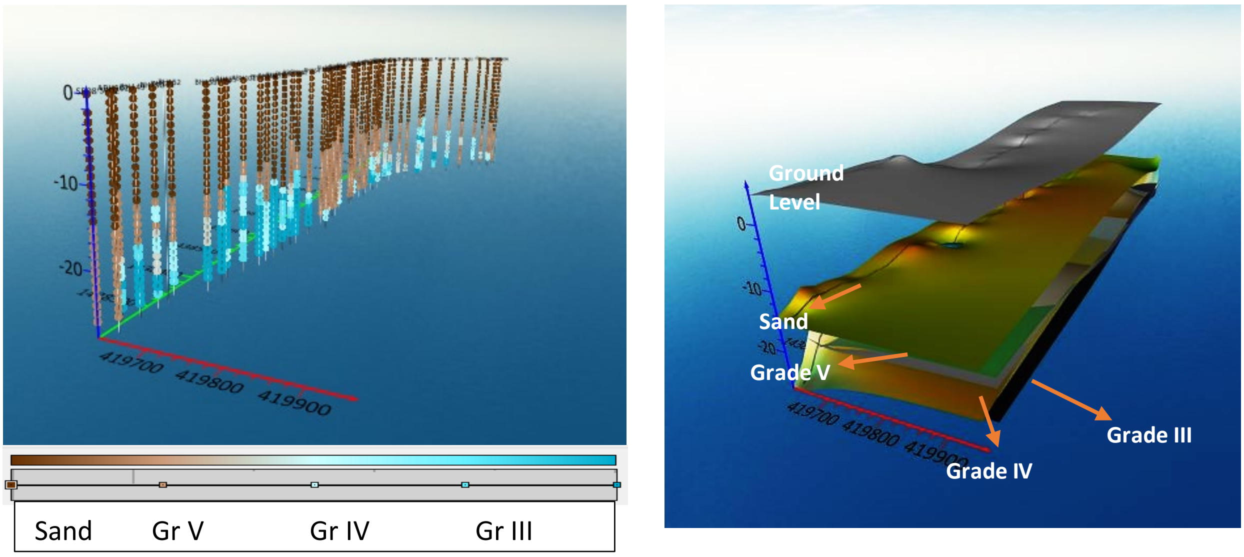
For modelling the EPB-TBM advance rates, geological and machine data from the metro tunnel project from a city in the southern part of India have been taken. The tunnel alignment stretch for this study is a twin tunnel, of approximately 1.2 km length each. The highly weathered section of the tunnel is especially selected for the study. The diameter of the tunnel is a standard metro tunnel of 6.6 m and was constructed using two identical EPB-TBMs and both tunnels have been completed recently.

Figure 2. Project tunnel alignment.

The study area has distinct geological features. The upper layer is made of transported sandy soil, varying in thickness up to 20 m, underlaid by metamorphic rock of various weathering grades. A small shallow river passes through the middle of the alignment and makes a significant contribution to local geology. A total 47 numbers of geotechnical boreholes were drilled in the nearfield of the tunnel alignment for the soil investigation. Various field tests and laboratory tests (on 11 drilled boreholes) have been performed on the samples to characterize the geological parameters along the tunnel alignment. The field bore log record and the lab report shared by the project team have been digitized.

Borehole profile, showing type and extent of various soils such as Sand, Grade V, IV, III, II, AND I have been drawn using Surfer software as shown in Figure 3 (a). Grades are defined on the basis of Rock Quality Designation (RQD) of the core obtained during drilling of the boreholes. RQD value of 0-10 is categorized as grade V, 11-30 as grade IV, 31-60 as grade III, 61-80 as grade II, and more than 80 as grade I rocks. The gridding method using the Kriging Interpolation technique has been used for the development of various sub-surfaces (Lisa, 2016). To explain the methods in simple terms, suppose the depths of the sea are known at various points; if someone wants to generate the sea bed, it can be generated using Kriging Interpolation. Kriging is a geostatistical interpolation method widely used in geology to predict spatially distributed variables based on known sample data. It estimates unknown values by weighting surrounding measured points, considering both the distance and the degree of variation between them. In geology, Kriging is essential for creating accurate maps of mineral deposits, soil properties, groundwater levels, and contamination spread. It can be utilised to estimate the geological parameters for this study.

For Kriging Interpolation, a linear type of semivariogram has been utilized, the slope has been kept at, default setup of 0.020, search neighborhood consists of all the points. The specific parameters of the Ordinary Kriging can be varied based on the judgment of the researcher. The individual subsurfaces thus obtained, as shown in Figure 3(b), have been stacked over one another, since sub-surfaces contain spatial values. Later, stacked sub-surfaces have been intersected with a vertical plane passing through the tunnel axis, and the cross-section thus obtained is the sub-surface profile along the tunnel alignment.



(a) (b)

Figure 3. (a) Geotechnical borewell Record along tunnel alignment & 3(b) 3-D Subsurface layers along tunnel alignment.

Two different subsurface profiles can be obtained by intersecting with the vertical plane passing through either of the tunnel axes. Figure 4, shown below, depicts the sub-surface profile thus generated for the downline tunnel.

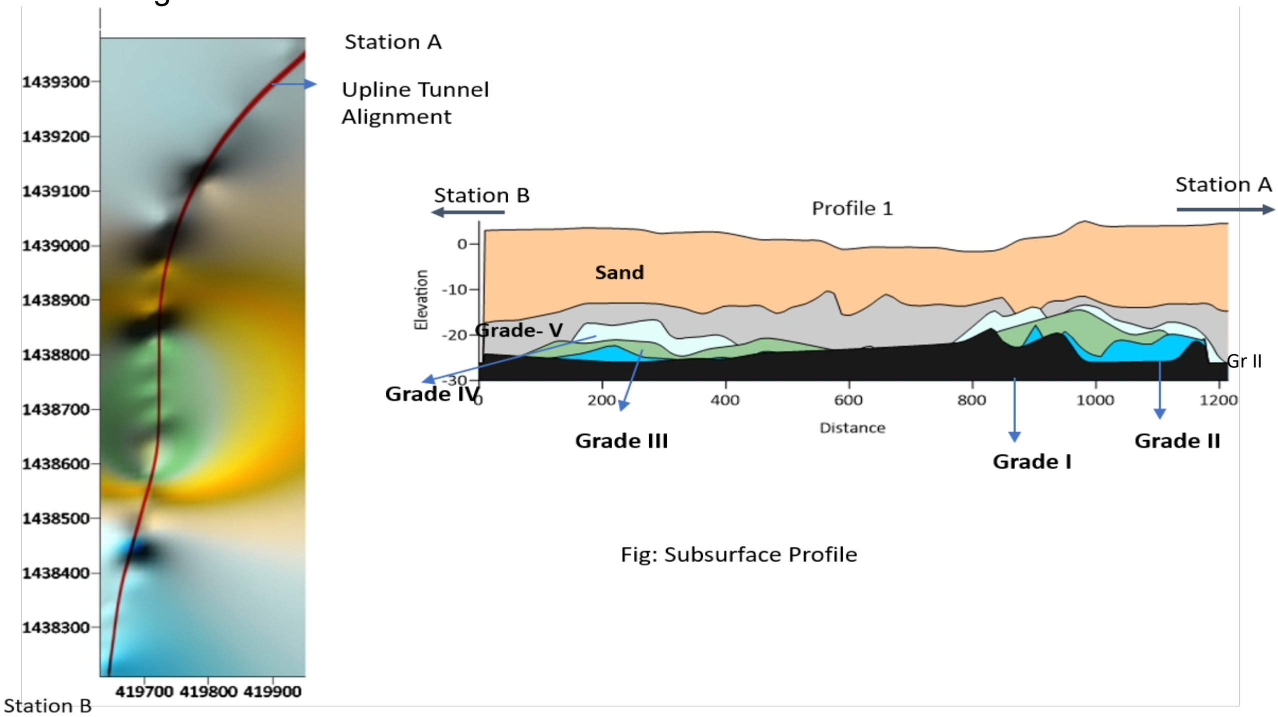


Figure 4. Subsurface profile along the downline tunnel alignment.

The vertical axes of the sub-surface profile have been enlarged 5 times, to make the diagram presentable, since the length of the tunnel is 1.2 km and the depth lies between 15-25 m. Subsequently, the tunnel longitudinal drawing has been masked over the generated sub-surface profile using a tool called ARCGIS, as shown in Figure 5. One can clearly see part of the tunnel passing through sandy strata, mixed strata, and within the weathered rocky geology. The segment of the tunnel, specifically passing through the weathered rocky strata, has been chosen for the modelling, further. Similarly, subsurface profile generation and ring association have been performed for the other tunnel as well.
Rings are the structural unit of a tunnel, it facilitate in indexing of the tunnel construction and also help for study purposes. Rings associated with weathered rock are categorized as shown in Table 2. A total of 1201 rings are selected from weathered rock, 624 from the upline tunnel, and 577 from the downline tunnel.

Table 3. The rings association of weathered rock

Tunnel	Rings		Soil-Type
Upline TBM 2	From	To	Weathered Rock
	170	794	624

Downline TBM 1	190	620	430
	650	797	147
		Σ = 1201 Rings	

170-794

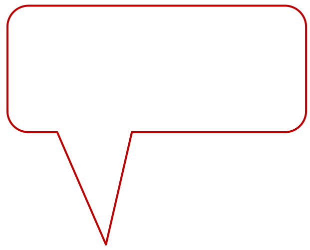


Figure 5. Upline tunnel alignment vis-à-vis sub-surface profile and its magnified section

2.1 EPBM CONFIGURATION AND OPERATION

Two single-shield EPB, identical TBMs have been used for tunneling between Station A and Station B. TBM 1 and TBM 2 are used for the downline and upline tunnels, respectively. The articulation system used in these TBMs is active, allowing precise steering and better control during tunneling operations. The TBMs have a shield length of 10 meters, providing structural support to the excavated tunnel.

The cutting face diameter of the TBM is 6.6 meters, suitable for creating tunnels for metro or utility lines. The cutterhead has an opening ratio of 32%, indicating the proportion of the cutterhead face that is open for excavated material to fall into the excavation chamber. Cutterhead equipped with disc cutters and scrapers, which are efficient for cutting through a variety of ground materials. Each machine uses 16 thrust jacks arranged in 4 groups, responsible for pushing the machine forward. Jacks can cumulatively exert a force of 42.5 MN, ensuring steady and powerful advancement. EPBM is powered by eight electric main drives, providing a reliable and efficient driving mechanism. The TBMs have a substantial torque capacity of 4239 kN-m, enabling them to rotate the cutterhead through resistant ground. The cutterhead can operate at a maximum rotational speed of 4.555 rpm, offering a balance between penetration rate and torque. It uses a single screw conveyor for muck removal. The conveyor screw is 0.8

meters in diameter, facilitating effective soil transportation. The machine can discharge excavated material at a rate of 6.48 m^3/min. The TBM can operate at a maximum pressure of 4 bar, suitable for maintaining face stability in varying ground conditions. The machine can advance at a rate of 80 mm/min, which reflects its cutting performance under optimal conditions.

2.2 DATA COLLECTION, PREPARATION, AND PARAMETER SELECTION

As discussed in the above section, the part of the tunnel alignment that lies completely within the weathered rock is selected for the analysis. For the upline tunnel, weathered rock strata lie between rings 170 to 794, and for the downline tunnel, weathered rock condition exists in rings 190 to 620 and 650 to 797. The rest of the segment of the tunnel alignment is traversing through mixed geology (a mix of rock and sand) and thus renders the analysis very complex, which is not in the scope of this study.

Out of 47 geotechnical boreholes drilled along the twin tunnel alignment of approximately 1.2 km, 11 boreholes were tested in the laboratory. RMR and UCS (of intact rock) are established among other things.

Rock mass rating provides a comprehensive evaluation of the weathered rock condition. It is the composite score of UCS of the intact rock, Rock Quality Designation, Spacing of discontinuities, ground water condition, and orientation of discontinuities. RMR rating of 81-100 is designated as very good, 61-80 as good, 41-60 as fair, 21-40 as poor, and 0-20 as very poor. The RMR value and the UCS of the rock taken from the report are further interpolated using Kriging interpolation techniques of ordinary types, having spherical semi-variogram, with search radius 100 m, and the number of nearest neighbors is kept at 10. The representative value of both parameters for each ring has been extracted using the ArcGIS tool. RMR Values range from 17 to 76, indicating high variability in weathered rock condition, inevitable from the fact that the tunnel lies in the vicinity of the seashore. UCS value ranges from 27 to 68 MPa. Face pressure values are given by a geologist based on overburden, water head, and surcharge.

TBM is a complex and advanced machine with numerous inbuilt sensors and recorders. It records various operational datapoints every 5 seconds. For the selected ring, mean value of advancement rates, main drive rpm, main drive torque, advance thrust force, cutterhead contact force, average chamber pressures, foam flow rates, polymer flow rates and screw conveyor torque were obtained from the data recorder of the machine. The feature matrix has been prepared with advancement (or ring no) as the index and comprises of advancement rate as the dependent variable and other input features as independent variables. There was a total of 3 features from geology and 7 machine features.

The average chamber pressure wasn’t directly reported. EPB-TBM bulkhead has 3 pairs of sensors, in this case, which measure the excavation chamber pressure at three locations. A linear straight line was passed along the three pressure sensor locations, pressure force acting on the circular face as per the vertical linear pressure line was integrated and subsequently converted into average chamber pressure.

Thus, our dataset comprises a total of 1201 pieces of data. Data is cleaned using an outlier removal technique, based on quartile and interquartile range. The upper bound and lower bound is defined as (Q1-1.5*IQR) and (Q3-1.5*IQR). Dataset is cleaned against 4 parameters, namely advancement speed (mm/min), advance thrust force (kN), average chamber pressure (bar) and main drive speed (rpm). Advancement rates may be exceptionally high or low based on local variability of some external and internal factors; thus, it needs to be curated. Operator, while driving the EPB-TBM, often tries to keep the variation in main drive speed as minimal as possible and fixes it based on their previous knowledge and experience. But, often in cautious situations and under other circumstances, such as after intervention, the speed may drop, quite less or much more, thus it needs to be removed for a fair study. Operators are often used to abuse the machine based on the exceptionally high thrust force in cases of low advancement rates, and may opt for a lesser than desirable thrust force in case of local damages. Thus, it renders data to be filtered based on thrust force as well. Sometimes, due to malfunctioning or damage of the earth pressure sensors in the bulkhead, the values of earth pressures, thus obtained, are extremely low, which need to be removed from the analysis. After outlier removal, we are left with 1045 datapoints. ML learning algorithm for testing and training is performed on these cleaned datasets.

Before splitting the dataset between train and test. The entire dataset has been partitioned into 4 parts on the basis of the magnitude of advancement rates, namely low, medium-low, medium-high, and high. This practice is called stratified functional partitioning (SFP), which ensures the uniform splitting of the dataset from the entire range of advancement rate values, thus reducing the chances of biases and variances. After SFP, the category of dataset has been split into train and test data in a ratio of 90:10 and later coalesced together to form composite train and test samples comprising almost uniform values of advancement rates from each category of low, medium-low, medium-high, and high. Further, before training independent variables are standardized using a scaling function as scaled_x_train and scaled_x_test to bring uniformity among the various different dimensional parameters. However dependent variable y has not been scaled.

3. DATA-DRIVEN MODELLING

There are numerous machine learning regression algorithms that can be used for modelling the advanced rates of EBP-TBM. Among all, three distinct regression algorithms have been employed, namely linear regression (LR), support vector regression (SVR), and the random forest (RF) algorithm, which were found to perform satisfactorily on the error parameters. All coding and modelling functions were performed in a Jupyter notebook in Python. The model and codes can be accessed at the GitHub link https://github.com/Amanchs-ui/TBM_TUNNELING.

Linear regression is the simplest and often provides adequate accuracy, and its simpler structure provides an interpretable description of the relationship between EPBM parameter and AR.
The hypothesis function of the linear regression model is depicted as Equation 1.

$$h(x) = Ø_0x_0 + Ø_1x_1 + Ø_2x_2 + Ø_3x_3 + Ø_4x_4 + Ø_5x_5 + Ø_6x_6 + Ø_7x_7 \ldots\ldots \quad (1)$$

where x1, x2, x3 …. are feature vectors and $Ø_1, Ø_2, Ø_3$ ……… are known as the coefficient of the features, $Ø_0$ is specifically known as intercepts. Algorithm tries to optimize the hypothesis function based on the learning from the training data. The bigger the value of coefficient of the features, the more important it will be. Test-train split ratio for this algorithm has been kept at 10:90.

Table 4. Values of various coefficients of Linear Regression Model

Feature	**Coefficient**	**Values**
Intercept	Ø0	10.92
Advance thrust force [kN]	Ø1	-0.56
Main drive total contact force [kN]	Ø2	-0.51
Main drive speed [rpm]	Ø3	-0.16
Main drive torque [MNm]	Ø4	1.50
Face Pressure[bar]	Ø5	0.53
EEP Middle Sensor [Bar]	Ø6	-1.11
EEP Bottom Sensors [Bar]	Ø7	-0.06
Screw conveyor torque [kNm]	Ø8	-0.09
Foam Polymer Flow Rate [l/min]	Ø9	1.67
Foam Flow Rate [l/min]	Ø10	1.35
RMR [No]	Ø11	-0.19
UCS [Mpa]	Ø12	0.39

The coefficient of determination (R^2) for the above model was found to be 0.80, while root mean squared error (RMSE) were found to be 1.95.

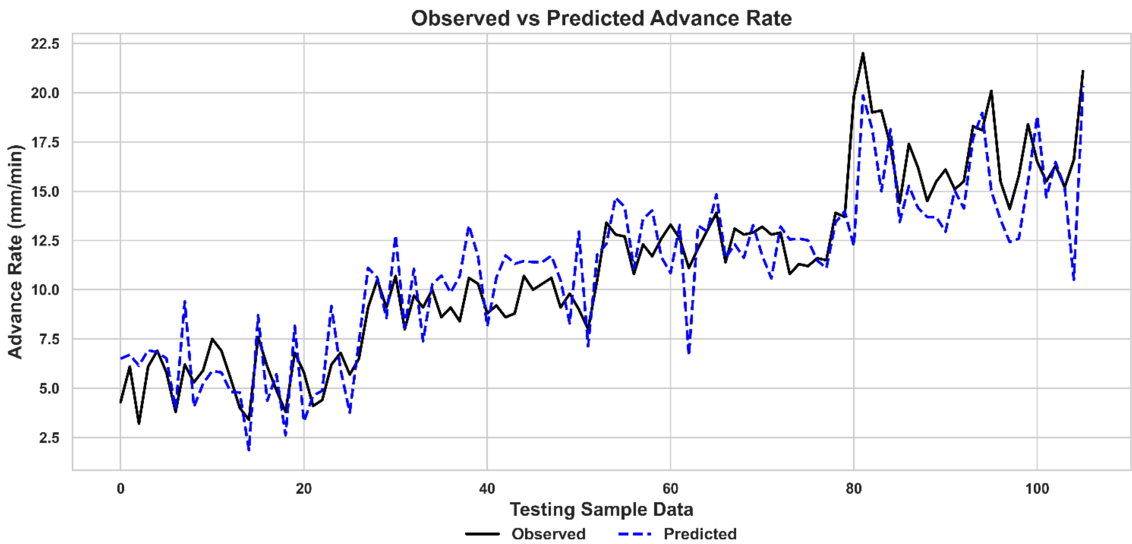


Figure 6. Comparison between observed vs predicted advanced rate in linear regression model on test data

One of the disadvantages of the linear regression model is that it only depicts linear relationships between independent and dependent variables.

As one can observe visually in the above plot, the predicted line is very sensitive, and it's a classic case of overfitting. For fine-tuning, the overfitting LASSO and Ridge regularization have been utilized.

LASSO (Least Absolute Shrinkage and Selection Operator) is a regularization technique used in linear regression that adds a penalty equal to the absolute value of the magnitude of coefficients (L1 norm). This method helps to prevent overfitting by shrinking less important feature coefficients to zero, effectively performing variable selection. As a result, LASSO is useful when dealing with datasets that have many features, allowing the model to become simpler and more interpretable. It is especially effective when only a subset of input features is expected to have a strong influence on the target variable.
LASSO with cross validation (10 subsets) has been performed to find the best alpha, which comes out to be 0.0033, and test MSE and COD come as 3.36 and 0.83, not much improvement from the original model.

Support Vector Regression (SVR) is a powerful machine learning technique used for time series forecasting. Based on the principles of Support Vector Machines (SVM), SVR aims to find a function that approximates future values within an acceptable error margin. It is especially useful for capturing complex, non-linear relationships in time series data.
In SVR, data is often first scaled, and time series features are engineered (like lagged values) to convert the sequential data into a supervised learning problem. The RBF (Radial Basis Function) kernel is commonly used to model nonlinear trends and patterns. SVR is robust to noise and overfitting, making it effective for small to medium-sized datasets.

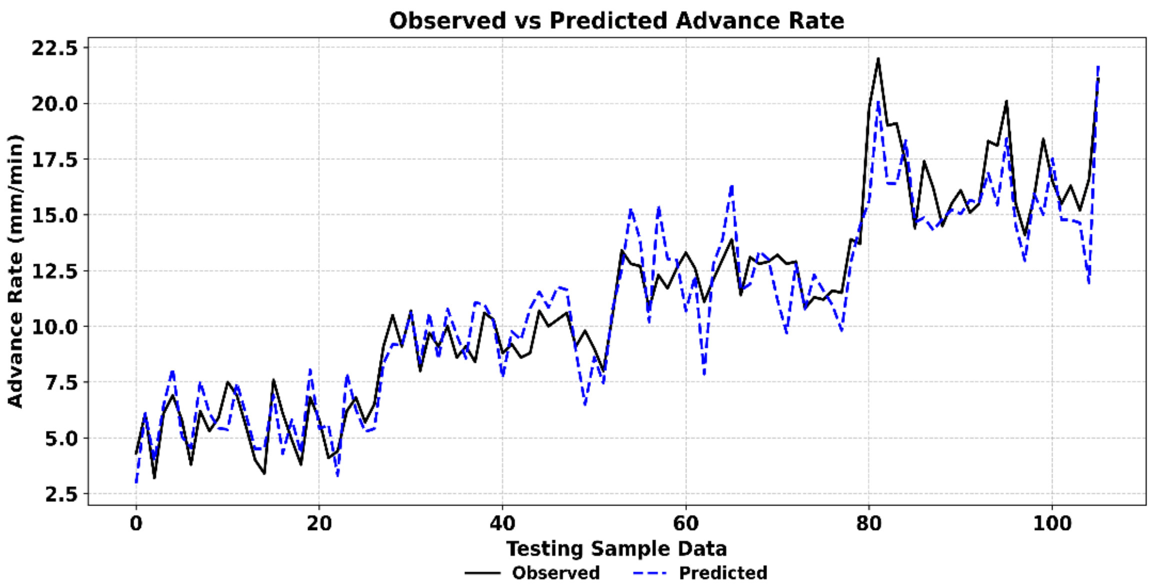


Figure 7. Comparison between observed vs predicted advanced rate in the SVR model on test data

The train test split ratio used for the SVR algorithm is 10:90. As mentioned above, radial basis function kernel was employed. The trend obtained from predicted vs observed value of advancement rates on the test sample data is as follows.

The coefficient of determination (R^2) for the SVR model is found to be 91, and the root mean squared error (RMSE) is found to be 1.35. The SVR model also seemed to be overfitting. To generalize, the overfitting hyperparameter tuning has been done. GridsearchCV with coefficient of determination as a scoring metric has been performed to find the best set of 'C', gamma, and epsilon, which comes as 50, 0.01, and 0.1, respectively. However, hyperparameter tuning didn't lead to any improvement in the coefficient of determination.

Random Forest Regression is an ensemble learning method that can be effectively applied to time series forecasting. It builds multiple decision trees and averages their predictions to improve accuracy and control overfitting. In time series tasks, Random Forest is typically used by creating lagged features from past observations to predict future values.

This method is capable of capturing complex non-linear relationships and handling missing values or noisy data, making it suitable for real-world time series like sales forecasting, stock market analysis, and energy consumption prediction. Random Forest is robust and less sensitive to hyperparameter tuning compared to other models.

However, it does not inherently account for time dependencies or seasonality, so proper feature engineering (e.g., creating lag, rolling statistics, or time-based features) is crucial. Despite this, its ability to model non-linear patterns and interactions makes it a powerful and flexible tool for time series forecasting.

The test train split used for the random forest algorithm is 90:10. The trend between predicted vs observed values of the advancement rates is as follows.

The coefficient of determination (COD) for the RF model is found to be 0.90 and root mean squared error (RMSE) was found to be 1.44 respectively.

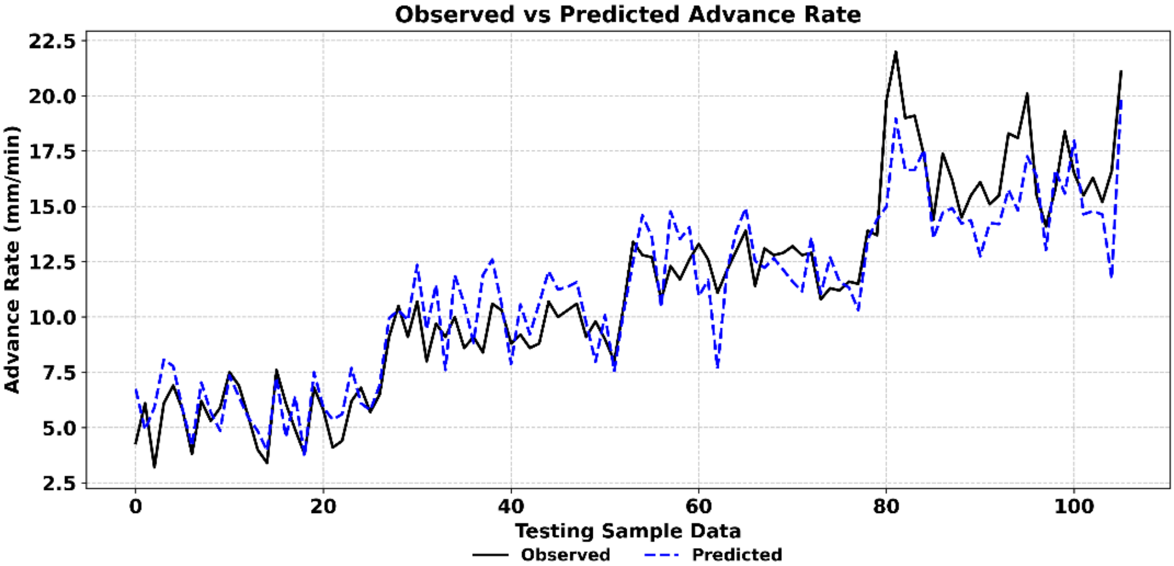


Figure 8. Comparison between observed vs predicted advanced rate in RF model on test data.

The RF model was further tried to be improved with the help of GridsearchCV to find the best hyperparameter for the model. The best parameterization gives maximum depth of the model, maximum sample leaf, and minimum sample split and n-estimator as 'none', 1, 2, and 200, respectively. However, the hyperparameter fixing didn't improve the coefficient of determination significantly.

A summary table of the stratified functional partitioning, train-test-split, coefficient of determination score, and root mean square values of linear regression, support vector regression, and random forest has been depicted below.

Table 5. Summary of error parameters of various models

Model Name	k-fold CV	T/T/S	R2 Score	RMSE
Linear Regression	4	0.1	0.80	1.95
Random Forest	4	0.1	0.88	1.56
Support Vector Regression	4	0.1	0.89	1.47

4. DISCUSSION

However, the general performance prediction behavior of linear regression, support vector regression, and random forest has been discussed in the above chapter. The nuances and specifics of every algorithm have been discussed further in this chapter.

- As shown in Figure 9 below, predicted vs observed values of the advance rates across three algorithms, linear regression, support vector regression, and random forest regression. Visually, it can be observed that there is an underestimation of prediction at the higher values of advance rates (mm/min) across all the algorithms, or it is biased at higher values of advance rates. Although these biases are less in the case of support vector regression and random forest than in linear regression. However, hyperparameter tuning have been done for all the algorithms, but fraction of biases still persists.

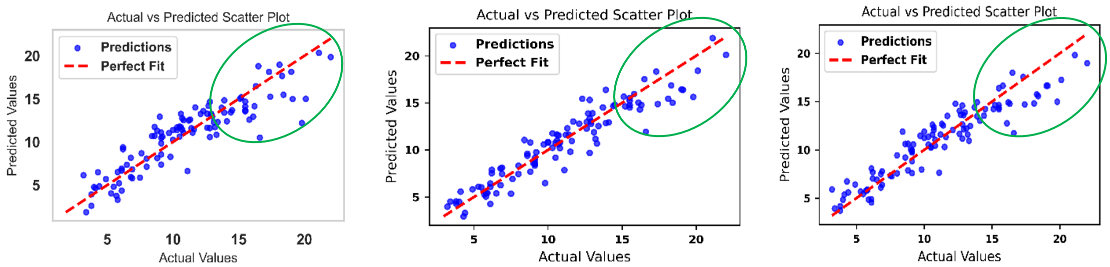


Figure 9. Scatter plot (a) linear regression (b) support vector regression (c) random forest.

- There were two plots used to figure out the importance of the parameter in modelling advanced rates. Shapely additive explanations, i.e., SHAP analysis, give the marginal contributions of the individual parameter in determining advance rates (mm/min). Permutation feature importance is another parameter used in analyzing significant parameters. It reshuffles the individual parameter and checks its influence on the dependent parameters. SHAP and LASSO feature coefficients have been plotted for linear regression as shown in Figure 10. Subsequently, permutation feature importance has been plotted for the support vector regression model and the random forest model.

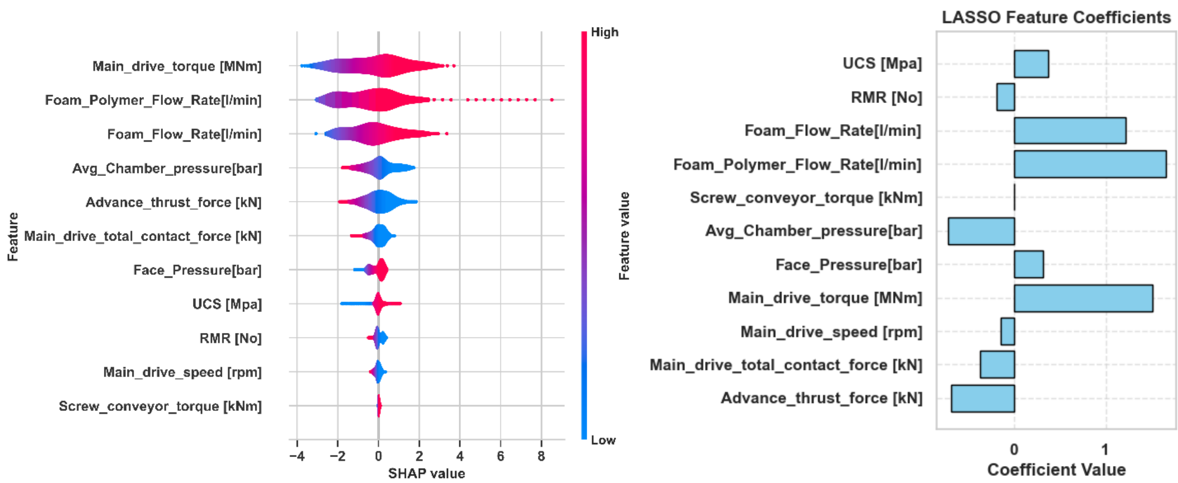


Figure 10. Showing SHAP and LASSO feature coefficients of linear regression

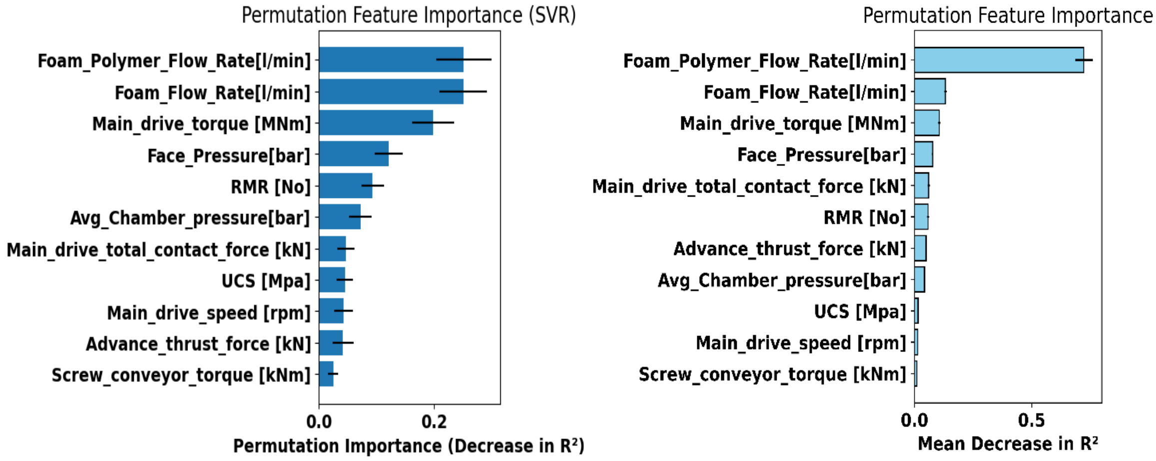


Figure 11. Permutation feature importance (a) Support vector regression (b) random forest

From the above feature importance plots, it can be inferred that polymer flow rate (l/min), main drive torque (MNm), and foam flow rate (l/min) are the features of first-degree importance. While face pressure, average chamber pressure, advance thrust force, and cutterhead contact force are the features of second-degree importance.

- As discussed above, the limitation of all the models to underestimate the value of the advance rate (mm/min) at higher values. A residual plot, as shown in Figure 12, for the worst 20 performing cases has been investigated, and the dataset has been manually observed to see if there were any recognizable patterns associated with underestimation. However, no such pattern could be found; biases at higher values of advance rates (mm/min) were due to obscure slump or surge in some particular value of parameters.

Table 6. Showing True, Predicted, and Residual Values

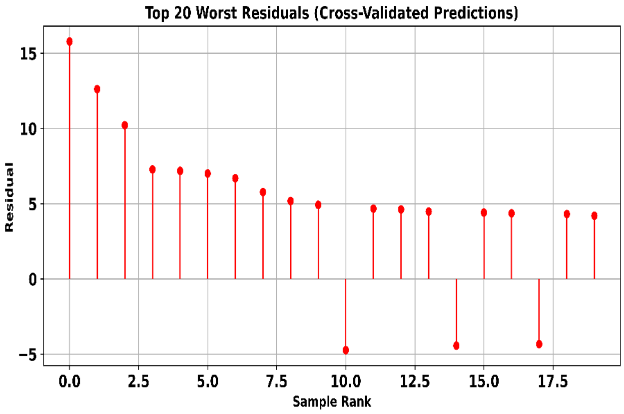


Figure 12. Residuals plot Predicted vs Actual in SVR (Top20)

Index	True	Predicted	Residual
759	20.6	4.79	15.81
801	16.4	3.89	12.51
953	16.2	6.91	9.29
425	9.8	17.85	-8.05
838	21.3	13.48	7.82
793	18.6	11.36	7.24
849	15	8.69	6.31
772	19.3	13.09	6.21
256	8.4	14.48	-6.08
643	12	6.09	5.91
590	11.4	5.74	5.66
904	20.5	14.85	5.65
973	19.8	14.16	5.64
796	19.5	13.88	5.62
824	23.2	17.71	5.49
912	14.7	20.04	-5.34
752	14.9	9.60	5.30
925	16.6	11.31	5.29
573	13.5	18.71	-5.21
898	18.9	13.78	5.12

- The independent variables can be further categorized as purely independent variable, controllable variable and indirectly controllable variable. Purely independent variable such as RMR value, UCS value of the intact rock and face pressure (bar), are the naturally existing features, it cannot be manipulated in tunnel construction in general. Controllable variable such as main drive speed (rpm), advance thrust force (kN), foam polymer flow rate (l/min) and foam flow rates (l/min) are controllable variable, since operator have significant control and command over that. Indirectly controllable variable such as main drive torque (MNm) and screw conveyor torque (kNm) are automatic turbo generated, machine tries to adjust automatically the torque in a way to match the cutterhead speed with thrust. Average chamber pressure is a result of various parameter, controlled indirectly.

Table 7. Feature variable categorization and respective bounds

Purely Independent Variable	Controllable Variable	Indirectly controllable Variable
RMR Value	Main Drive speed	Main Drive torque
UCS Value	Advance Thrust force	Average Chamber Pressure
Face Pressure	Polymer Flow Rates	Screw Conveyor Torque
	Foam Flow rates	Cutterhead contact force
Fixed/ Static Bounds	Absolute Bounds	Relative Bound

Now, bound optimization has been performed keeping the kind of variable in mind, it's a kind of constrained inverse modelling, which is to find the optimal set of input values (features) that will maximize the target output i.e. advance rate (rpm), while ensuring that all values stay within acceptable or realistic bound.

Table 8. Features and their respective bounds

Features	Correlation/ Bounded by	Nominal Vlaues
Purely Independent Variable / Fixed Bounds		
RMR (No)	Fixed	45
UCS (Mpa)	Fixed	51
Face Pressure (Bar)	Fixed	1.84
Controllable Variable/ Absolute Bounds		
Main_drive_speed [rpm]	4.555	3.1885
Advance_thrust_force [kN]	42000	29400
Foam_Flow_Rate[l/min]	As per manuf. Reco.	477.4
Polymer_Flow_Rate[l/min]	As per manuf. Reco.	1.96
Indirectly Controllable Variable/ Relative Bounds		
Main_drive_torque [MNm]	Average Chamber pressure (ACP)	1.81 + 0.69 * 0.49/0.43 (ACP-2.03)
Main_drive_total_contact_force [kN]	Advance Thrust Force (ATF)	3766+0.44*855/1713*(ATF- 13092)
Average_Chamber_Pressure[Bar]	Face Pressure (FP)	2.03 + 0.40 * 0.478/0.135* (FP-1.854)
Screw_conveyor_torque [kNm]	Foam Flow rate (FFR)	13.443 - 0.28*10.363/128*(FFR-306.68)

Purely independent variables such as RMR, UCS of the intact rock, and face pressure values are fixed or static bonds. Main drive speed (rpm), advance thrust force (kN), polymer flow rates (l/min), foam flow rates (l/min) are absolute bounds, thus, main drive speed (rpm) and advance thrust force (kN) are limited by machine capacity (70% of the TBM capacity) and polymer flow rates(l/min) and foam flow rates (l/min) are as per manufacturer recommendation. Indirectly controllable variables such as main drive torque (MNm), average chamber pressure (bar), screw conveyor torque (kNm), and cutterhead contact force (kN) are indirectly controlled variable, and can be limited with functional constrain, derived from Pearson correlation coefficient established from the dataset, and roughly correlated with the strongly correlated features correspondingly. which has been shown here in Table 8.

Now, optimization of the support vector regression model and random forest model has been carried out, with the values of RMR (No), UCS of intact rock (MPa), and face pressure values kept at 45, 51, and 1.84, respectively, for a sample example. The following results have been found.

Table 9. Features and their optimum values as per inverse modelling

Feature Name	SVR Model	RF Model Value	Type
Face_Pressure[bar]	1.84	1.84	Fixed Input
RMR [No]	45	45.00	Fixed Input
UCS [Mpa]	51	51.00	Fixed Input
Advance_thrust_force [kN]	12387	12638	Optimized Input
Main_drive_speed [rpm]	1.984	2.02	Optimized Input
Polymer_Flow_Rate[l/min]	1.96	1.85	Optimized Input
Foam_Flow_Rate[l/min]	477.4	446.43	Optimized Input
Avg_Chamber_pressure[bar]	2.01	2.01	Computed Feature
Main_drive_torque [MNm]	1.7944	1.79	Computed Feature
Main_drive_contact_force [kN]	3624.9	3680.17	Computed Feature
Screw_conveyor_torque [kNm]	9.687	10.37	Computed Feature
Max Predicted Advancement Rate [mm/min]	**19.25**	**17.40**	**Model Output**

The average predicted advance rate has been found as 18.32 mm/min.

- One of the applications of this kind of modeling is to gain actionable insights into EPB-TBM operation. A partial dependence plot has been drawn for the features, which means that keeping all the input variables fixed at their mean values, what will be the behavior of advance rates (mm/min) if we vary a particular feature. Main drive speed (mm/min), main drive torque (MNm) and advance thrust (kN) are the big three of the TBM operation. Partial dependence plot of the big three are plotted in the Figure 13.

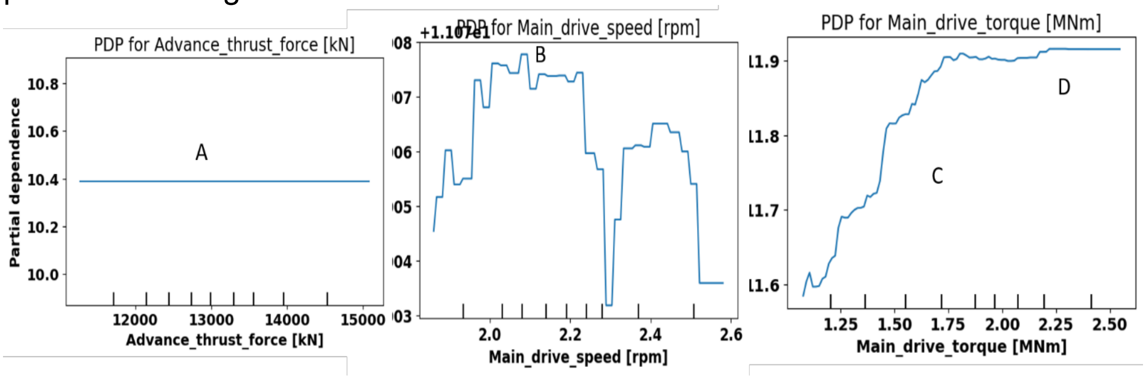


Figure 13. Partial dependence plot for the (a) advance thrust force (kN) (b) main drive speed (rpm) and (c) main drive torque (MNm)

The above figure can be explained with points- A, B, C and D. There seemed to be negligible influence on the advance rates (mm/min) on increasing the advance thrust force as shown with A mark. Optimal range of main drive speed seemed to be 1.8 to 2 mm/min as highlighted by B point. Advance rates (mm/min) seems to increase modestly linearly with the increase in main drive torque (MNm) as depicted by C point and seems to flatten with further increment in main drive torque (MNm), indicative of wastages of resources and energy as depicted by D point.

5. CONCLUSION

Recent advancement in the development of ML algorithms gives ample opportunity to analyze multidimensional variables such as EPB-TBM advancement by utilizing data-driven modelling. This paper is work in that direction. Data obtained from a metro project in the southern part of India is utilized for modelling EPBM advancement rates in weathered rock conditions. Geological as well as machine parameters have been utilized explicitly. Efforts have been made to see the prediction behavior of various algorithms, since every algorithm learns and predicts differently. Linear regression, support vector machine, and random forest have been reported to give a reliable measure of prediction. Investigation on the worst 20 errors found no pattern. Furthermore, SHAP analysis and permutation feature importance suggests that polymer flow rate (l/min), main drive torque (MNm) and foam flow rate (l/min) as primary features, followed by face pressure (bar), average chamber pressure (bar), advance thrust force (kN) and main drive contact force (kN) as secondary parameters. Bound optimization has been performed, categorizing the input features into three categories as fixed bounds, absolute bounds, and relative bounds. Univariate Partial dependence plot for the advance thrust force (kN), main drive speed (rpm), and main drive torque (MNm) depicted in the study would give the operator an actionable insight. Among RMR (No) and UCS (Mpa) of intact rock, RMR (no) is found to be better predictive parameter in support vector regression and random forest regression. However, further cohesion and internal angle of friction parameter may potentially serve as important parameter, which needs to be investigated.

REFERENCES

Armaghani, M. E. (2017). DDevelopment of Hyrid Intelligent Model for Predicting TBM Penetration Rate in Hard Rock Condition. *Tunneling and Underground Space Technology*, Vol. 63, 29-43.

Benardos, K. (2004). Modelling TBM Performance with Artificial Neural Networks. *Tunneling and Underground Space Technology*, 19(6): 597-605.

Elbaz K., S. S. (2019). Optimization of EPB Shield Performance with Adaptive Neuro-Fuzzy Intference System and Genetic Algorithm. *Applied Sciences*, 780-797.

Gao X., S. M. (2019). Recurrent Neural Networks for Real Time Prediction of TBM Operating Parameters. *Automation in Construction*, Vol-98, 225-235.

Lisa, M. (2016). *Advancing Understanding of the Relationship Between Soil Conditioning and Earth Pressure Balance Tunnel Boring Machine Chamber and Shield Annulas Behavior.* Golden, Colorado.

Mahdevari, S. K. (2014). A Support Vetor Regression Model for Predicting Tunnel Boring Machine Penetration Rates. *International Journal of Rock Mechanics and Mining Sciences*, 214-230.

Mokhtari S, M. M. (2020). Predicting EPBM Advance Rate Performance Using Support Vector Regression. *Tunneling and Underground Space Technology*, Vol-104, .

Mokhtari S., N. W. (2020). White-Box Regression (Elastic Net) of Earth Pressure Balance Shield Machine Advance Rate. *Automation in Construction*, Vol-115.

Morshedlou A., R. J. (2024). Introducing a New Model for Prediction of Mean Cutting Forces Acting on Conical Pick Cutters. *Rock Mechanics and Rock Engineering*, Vol. 57, 1695-1716.

Yagiz S, K. H. (2011). Prediction of Hard Rock TBM Penetration Rate Using Particle Swarm Optimization. *International Jounal of Rock Mechanics and Mining Sciences*, Vol- 48, 427-433.

Yagiz, S. E. (June 2009). Application of Two Non-Linear Prediction Tools to the Estimation of Tunnel Boring Machine Performance. *Engineering Application of Artificial Intelligence*, 809-817.