TBM penetration rate prediction using machine learning models and models' generalization

Shengfeng Huang
Stevens Institute of Technology, Hoboken, NJ, USA

Pooya Dastpak
Stevens Institute of Technology, Hoboken, NJ, USA

Saadeldin Mostafa
Stevens Institute of Technology, Hoboken, NJ, USA

Misagh Esmaeilpour
Stevens Institute of Technology, Hoboken, NJ, USA

Rita Sousa
Stevens Institute of Technology, Hoboken, NJ, USA

ABSTRACT: Most existing models about penetration rate (PR) prediction have been developed and validated against data from one single project. This poses the question whether these models can perform well when faced with new data. We use two datasets of two tunnels built with the same construction method and in similar geological conditions. Different machine learning (ML) models are trained, validated, and tested with dataset from one tunnel and then generated to the other dataset. Additionally, the effect of several data processing techniques for splitting and scaling on the performance and generalization of the different models is tested. The results demonstrate that random forest (RF) and extreme gradient boosting (XGBoost) exhibit better performance than other models. Regarding generalization, CART and XGBooost model exhibit the best performance. The impact of splitting and scaling techniques on the generalization of the models becomes noticeable than on the performance of models.

Keywords: Penetration rate, EPBM, Machine learning, Generalization, split and scale technique.

1 INTRODUCTION

The choice between New Austrian Tunneling Method (NATM) and TBM is a decision often related to ground and length of the tunnel. NATM tends to be more economical for shorter distances and is more flexible when it comes to support systems. However, modern TBM can be used in a range of geologies, and it offers great advantages over NATM when it comes to the safety for the workers in case of difficult ground conditions. An essential task in tunneling is the reliable estimation of performance, often measured through penetration or penetration rate, which is needed for planning, cost estimation, and feasibility assessment. The ability of the TBM to collect large amounts of data during construction allows for better assessing and updating predictions during construction and supports decision making on site in a timely manner.

Most work on tunneling performance, has been done on predicting performance of TBM in rock using traditional statistical methods (Sapigni et al., 2002; Yagiz, 2008). With the proliferation of machine learning (ML) and the capability of TBM to collect large datasets during construction, a new generation of models has been developed. During the past two decades, several ML based models have been used to estimate AR or PR, such as ANN , Fuzzy Logic, support vector machine (SVM) (Mokhtari & Mooney, 2020; Xu et al., 2019). Among them, the most commonly used techniques are random forest (RF) (Huang, Dastpak, et al., 2023; Sun et al., 2018; Tao et al., 2015) and support vector regression (SVR) (Huang, Esmaeilpour, et al., 2023; Mokhtari & Mooney, 2020; Yang et al., 2020) due to robust ability in dealing with complex systems with large inputs. Also, many temporal models have been developed to predict TBM performance with time, including recurrent neural networks, long-short term memory network (B. Gao et al., 2021; X. Gao et al., 2019; Lin et al., 2022). Hybrid models that combine ML and optimization algorithms have better performance through tunning hyperparameters (Yang et al., 2020; Yu et al., 2022).

Despite the many models developed in the past decades, there is no consensus to what ML models best perform in predicting advance rate (AR) or Penetration (PR), particularly as new ML algorithms continue to be developed. Moreover, the ability of a model to perform well when data from a new tunnel becomes available is seldom tested mainly due to the scarcity of data. This is an important point as for a prediction model to be of use, its ability to generalize well to new data is essential. In this paper we test the ability of several ML models to predict TBM performance based on sensor data, and more importantly we test the ability of those models to generalize to new datasets. For that purpose, we will use the data from two tunnels in the same city with under similar ground conditions. Different ML models are tested and compared. The effect of several data processing techniques for scaling and splitting on the performance of the different models is also investigated. More importantly the generalization of the different models is tested. This is an important point as much of the existing research only uses data from one single tunnel to train, validate and test their performance prediction models.

2 DATA DESCRIPTION

2.1 Project description

The data used in this study was collected during the construction of two tunnels S and C of the light metro for the city of Porto in Portugal. Tunnel S is 3.7 km long and tunnel C is 2.3 km long. The construction method used in the excavation of the tunnels was an Earth Pressure Balance Machine (EPBM), capable of operating in mixed ground conditions. Through an extensive geological survey during the planning phase of the project, seven geomechanical groups have been defined, ranging from sound granite to saprolite and alluvial deposits. Both tunnels through similar conditions. Figure 1 shows alignments of both tunnels.

2.2 Data and data processing

To build the models, the input features were selected based on engineering judgment – nine (9) machine parameters were selected, as shown in Table 1. It is important to note that we did not have access to the as-built geologic properties for tunnel S. Therefore, we relied solely on machine parameters in building our models. The raw data from the sensors was processed and only data from the excavation phase was used (i.e., halt phase data, when TBM is stopped were excluded).

Different scaling and splitting techniques were also compared. Namely, we used three different scaling techniques: StandardScaler, MinMaxScaler and RobustScaler to test their influence in the performance of the models. StandardScaler scales the data to have a mean of zero and a variance of one, but it may not ensure balanced feature scales, especially in the presence of outliers. MinMaxScaler scales the data to the range [0, 1], ensuring that all features have the same scale. RobusScaler is designed to be less affected by outliers by centering the data using the median and scaling it based on the interquartile range, making it a good choice when working with data that has outliers.

Figure 1. Map of Metro do Porto. Tunnel C runs from Campanhã to Trindade and tunnel S runs from Salgueiros to São Bento (Adapted from Babendererde et al., 2004).

Table 1. List of selected input features.

Number	Feature
1	Torque cutting wheel
2	Pressure force cutting wheel
3	Thrust force
4	Torque screw
5	Cutting wheel speed of rotation
6	Thrust pressure
7	Earth pressure
8	Pressure foam lance
9	Excavated material flow

As for the splitting techniques, we compared Random split with Stratified split. In the Random split, data is split into training and testing randomly. In the Stratified split, the data is split in a way that ensure both training and testing data can be representative of all ranges of data, which may help reduce bias in the model.

2.3 Methods

Several models were tested, which include models such as k-nearest neighbor (KNN), support vector regression (SVR), artificial neural networks (ANN), random forest (RF), classification and regression trees (CART), and extreme gradient boosting (XGBoost). KNN is a non-parametric algorithm that makes predictions based on the closest labeled examples in the training data. SVR is a supervised learning algorithm that analyzes data and learns to predict the output values based on input data. ANN is a machine learning model inspired by biological neural networks, consisting of layers of interconnected nodes that process input data. RF and CART are decision tree-based models that recursively split the data into subsets based on certain criteria. XGBoost is a boosting algorithm that uses an ensemble of weak decision trees to make predictions.

For each model, we performed a grid search to optimize the hyperparameters, and evaluated the performance using the root mean squared error (RMSE) and coefficient of determination (R^2) metrics. These models were trained, validated, and tested on the data of tunnel line C and generalized to line S.

3 RESULTS

3.1 PR prediction model performance

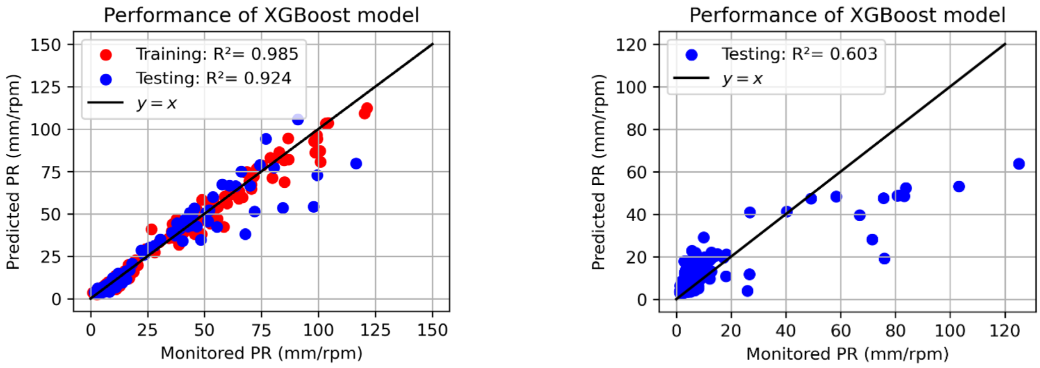
The results are displayed below on Table 2. Four datapoints whose value of PR are greater 150 mm/rpm were removed as they were considered outliers. The results show that models perform well when predicting PR, with R^2 ranges from 0.90 to 0.99 for training data and 0.88 to 0.94 for testing data. Among these models, RF, and XGBoost outperform other models with R^2 larger than 0.98 for training data and R^2 larger than 0.94 for testing data. On the other hand, model error, expressed in RMSE, is low in all cases.

Table 2. Performance of models under different split and scale techniques.

Model	Split techniques	Scale techniques	Tunnel line C				Generalization to tunnel line S	
			Training		Testing			
			R^2	RMSE	R^2	RMSE	R^2	RMSE
KNN	Random split	StandardScaler	0.940	4.311	0.940	4.211	-0.125	11.845
SVR			0.967	3.184	0.884	5.869	0.019	11.062
ANN			0.944	4.169	0.925	4.715	-0.034	11.360
RF			0.986	2.075	0.940	4.202	0.667	6.450
CART			0.980	2.460	0.907	5.248	0.646	6.643
XGBoost			0.990	1.727	0.938	4.297	0.748	5.611
KNN	Stratified split	StandarScaler	0.948	4.034	0.908	5.160	-0.302	12.744
SVR			0.904	5.483	0.887	5.725	0.289	9.419
ANN			0.958	3.729	0.928	4.557	0.033	10.982
RF			0.990	1.766	0.929	4.532	-0.216	12.316
CART			0.980	2.533	0.886	5.732	0.687	6.249
XGBoost			0.985	2.169	0.924	4.698	0.603	7.034
KNN	Stratified split	MinMaxScaler	0.947	4.074	0.906	5.220	-0.455	13.474
SVR			0.957	3.657	0.915	4.956	-0.084	11.631
ANN			0.956	3.715	0.932	4.446	0.333	9.123
RF			0.990	1.766	0.929	4.532	-0.216	12.316
CART			0.978	2.642	0.889	5.654	0.677	6.352
XGBoost			0.985	2.169	0.924	4.698	0.603	7.034
KNN	Random split	RobustScaler	0.945	4.114	0.932	4.492	0.361	8.932
SVR			0.964	3.346	0.914	5.046	0.405	8.615
ANN			0.932	4.570	0.921	4.830	0.015	11.086
RF			0.986	2.095	0.942	4.148	0.686	6.256
CART			0.980	2.460	0.907	5.248	0.646	6.643
XGBoost			0.990	1.727	0.938	4.297	0.748	5.611

Slight changes are observed in the performance of KNN and SVR models when using different scaling and split techniques, while the impact was relatively small for ANN, RF, CART, and XGBoost models. Due to the relatively balanced nature of the date with enough datapoints, the impact of split is less apparent. This means that by using either Random split or Stratified split, one can split the data into two parts which are representative of all ranges of data. Regarding the impact of scale techniques, the MinMaxScaler and RobustScaler outperforms StandardScaler technique since data is not normally distributed and thus StandardScaler is not the best fit. It is worth noting that scaling techniques do not have an impact on the performance of decision tree-based models such as random forests, CART, and XGBoost, as these models are not sensitive to the variance in the data.

The prediction performance showing the XGBoost model with Stratified split and MinMaxScaler technique predicted PR versus monitored PR using is presented in Figure 2(a). Overall, the results show good agreement between the predicted and monitored PR values. However, it is important to note that most of the PR data is concentrated in the range of 0-25 mm/rpm, while less data lies in the range of 25-150 mm/rpm. As a result, the model's fitting in this range is not as good as in the lower range. This may suggest that the model is not as accurate in predicting higher values of PR, and further improvements may be necessary to enhance the model's performance in this range.



(a) Performance on tunnel line C

(b) Generalization to tunnel line S

Figure 2. Performance and generalization of XGBoost using Stratified split and MinMaxScaler technique.

3.2 Generalization

Once the models described above were trained, validated, and tested, they were applied to the data of from tunnel line S to assess their generalization capability. The results are presented in Table 2 and one typical generalization performance of XGBoost model on tunnel S is shown in Figure 2(b).

The results show that models' generalization ability varies greatly depending on the techniques used. In most cases, the models' generalization performance is poor. However, in some cases, the R^2 value for generalization can reach up to 0.74. For example, the XGBoost model exhibits higher generalization ability with R^2 larger than 0.60, regardless of the split and scale techniques used. The RF and CART model also demonstrate a good generalization capability.

In contrast to the results obtained for tunnel C, we observed a noticeable impact of the split and scale techniques on the generalization performance of the models for tunnel S. The variance of R^2 can reach up to 0.90 for RF, followed by 0.81 for KNN. RF can exhibit a good generalization performance with R^2 of 0.69 using Random split, but it can also have a poor performance with R^2 of -0.22 using Stratified split. Less impact is observed for CART and XGBoost model, with variance of R^2 of 0.04 for CART and 0.14 for XGBoost. The models that best generalize are CART and XGBoost.

In addition, large errors of PR prediction can be observed in the range of 25-150 mm/rpm. This is most likely due to the insufficient training data within this range.

4 CONCLUSIONS

Models based on ML have proliferated in the last decade in tunneling. Despite their incredible performance when applied to a dataset corresponding to one project, there are few studies that look into the ML model's ability to react to new data and make accurate predictions, i.e., the capability of a model to generalize. This lack of studies is mainly due to the lack of available tunnel construction data. However, a model's ability to generalize is central to its success. A model that does not generalize well is practically useless.

This paper shows the preliminary results of a study that is being conducted to develop TBM performance prediction models and test their generalization to new data. The results show that new generation models that use boosting techniques such as XGBoost show a promising outcome but that more studies need to be done in order to develop models that can not only generalize well but can be deployed in real life situations. Emphasis of our current work is on developing online learning models that use past data, but continuously update in real-time as the excavation progresses.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Porto Metro Authorities for generously allowing us to use data from the construction of the Porto Metro, without which this work would not have been possible. This work was partially supported by the United States Department of Transportation (USDOT) Region 2 University Transportation Center (UTC) led by the Center for Advanced Infrastructure and Transportation (CAIT) at Rutgers University through research project #CAIT-UTC-REG 56.

REFERENCES

Babendererde, D. S., Hoek, D. E., Marinos, P. P., & Cardoso, P. A. S. (2004). *Geological risk in the use of TBMs in heterogeneous rock masses—The case of "Metro do Porto" and the measures adopted*. 15.

Gao, B., Wang, R., Lin, C., Guo, X., Liu, B., & Zhang, W. (2021). TBM penetration rate prediction based on the long short-term memory neural network. *Underground Space*, *6*(6), 718–731. https://doi.org/10.1016/j.undsp.2020.01.003

Gao, X., Shi, M., Song, X., Zhang, C., & Zhang, H. (2019). Recurrent neural networks for real-time prediction of TBM operating parameters. *Automation in Construction*, *98*, 225–235. https://doi.org/10.1016/j.autcon.2018.11.013

Huang, S., Dastpak, P., Esmaeilpour, M., Kaijian, L., & Sousa, R. (2023). Comparison between machine learning algorithms for TBM advance rate prediction. In *Expanding Underground—Knowledge and Passion to Make a Positive Impact on the World*. CRC Press.

Huang, S., Esmaeilpour, M., Dastpak, P., & Sousa, R. (2023). *EPBM Advance Rate Prediction Using Hybrid Feature Selection and Support Vector Regression Modeling*. 253–264. https://doi.org/10.2991/978-94-6463-104-3_22

Lin, S.-S., Zhang, N., Zhou, A., & Shen, S.-L. (2022). Time-series prediction of shield movement performance during tunneling based on hybrid model. *Tunnelling and Underground Space Technology*, *119*, 104245. https://doi.org/10.1016/j.tust.2021.104245

Mokhtari, S., & Mooney, M. A. (2020). Predicting EPBM advance rate performance using support vector regression modeling. *Tunnelling and Underground Space Technology*, *104*, 103520. https://doi.org/10.1016/j.tust.2020.103520

Sapigni, M., Berti, M., Bethaz, E., Busillo, A., & Cardone, G. (2002). TBM performance estimation using rock mass classifications. *International Journal of Rock Mechanics and Mining Sciences*, *39*(6), 771–788. https://doi.org/10.1016/S1365-1609(02)00069-2

Sun, W., Shi, M., Zhang, C., Zhao, J., & Song, X. (2018). Dynamic load prediction of tunnel boring machine (TBM) based on heterogeneous in-situ data. *Automation in Construction*, *92*, 23–34. https://doi.org/10.1016/j.autcon.2018.03.030

Tao, H., Jingcheng, W., & Langwen, Z. (2015). Prediction of hard rock TBM penetration rate using random forests. *The 27th Chinese Control and Decision Conference (2015 CCDC)*, 3716–3720. https://doi.org/10.1109/CCDC.2015.7162572

Xu, H., Zhou, J., G. Asteris, P., Jahed Armaghani, D., & Tahir, M. M. (2019). Supervised Machine Learning Techniques to the Prediction of Tunnel Boring Machine Penetration Rate. *Applied Sciences*, *9*(18), Article 18. https://doi.org/10.3390/app9183715

Yagiz, S. (2008). Utilizing rock mass properties for predicting TBM performance in hard rock condition. *Tunnelling and Underground Space Technology*, *23*(3), 326–339. https://doi.org/10.1016/j.tust.2007.04.011

Yang, H., Wang, Z., & Song, K. (2020). A new hybrid grey wolf optimizer-feature weighted-multiple kernel-support vector regression technique to predict TBM performance. *Engineering with Computers*, *38*(3), 2469–2485. https://doi.org/10.1007/s00366-020-01217-2

Yu, H., Zhou, X., Zhang, X., & Mooney, M. (2022). Enhancing earth pressure balance tunnel boring machine performance with support vector regression and particle swarm optimization. *Automation in Construction*, *142*, 104457. https://doi.org/10.1016/j.autcon.2022.104457