Preliminary Study of Utilizing Machine Learning to Predict Cutter Wear in Tunnel Boring Machines

V.Tan, K.F.Ho, J.Kumarasamy, C.Veeresh, Y.H.Zhang & H.Bernard
Land Transport Authority, Singapore

Y.M.Kim
University of Glasgow, Singapore

ABSTRACT: Cutters are essential tools for Tunnel Boring Machines (TBM) but are subject to wear and require timely replacement through Cutter Head Intervention (CHI). CHI is usually planned based on engineering judgement derived from TBM operations and interpreted ground conditions. However, conservative engineering judgment may result in early CHI, which can be unnecessary if the cutters have not worn out. Conversely, less conservative engineering judgement may result in late cutter replacement, potentially impacting the cutter head or other relevant parts of the TBM. With the availability of big data generated during tunnelling, machine learning, one of the artificial intelligences (AI) technologies, provides an opportunity for the engineer to obtain a second opinion. This preliminary study attempts to leverage several machine learning algorithms using regression and classification approaches to predict cutter wear in a slurry-type TBM in the Bukit Timah Granite formation by analyzing TBM operational parameters. The results demonstrated that the machine learning algorithms, trained by past tunnelling project data, could reasonably predict cutter wear, leading to a more efficient CHI regime.

1 INTRODUCTION

1.1 Background

Detecting cutter wear poses a challenge in the tunnelling industry, and it is essential to develop accurate and effective approaches to address it during tunnelling operations. The cutters on a TBM are essential cutting tools for tunnel boring and as they wear down or become damaged (Figure 1), the efficiency of the TBM decreases, leading to increased downtime and higher costs/risks (Bligin et al., 2012). There are two main wear mechanisms of cutters: normal abrasive wear and abnormal wear caused by highly variable loads which happens more frequently in mixed soil profiles (Liu et al., 2020). Replacements of cutters are required when they reach their normal wear limit or experience abnormal wearing. Such replacements are typically carried out during CHI when the TBM is stopped to conduct the replacements.

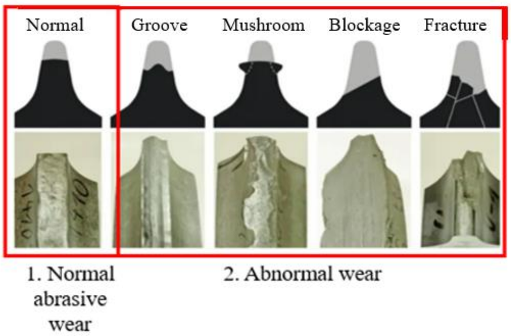


Figure 1. Common mechanisms of cutter wear (modified from Karami et. al., 2021).

Current methodologies of predicting cutter wear to ensure timely replacements include using empirical formulas based on interpreted ground conditions to estimate frequency of cutter replacements (Ko et. al., 2020) or a popular emerging method is the utilization of sensor monitoring to obtain real time information on cutter wear status (Lan et. al., 2019). However, both methods consist of a few limitations. Although the empirical method provides reasonable estimations, it still largely relies on the interpretation of ground profiles which has a degree of uncertainty due to potential variations in actual ground conditions. On the other hand, sensor monitoring requires the sensors to be installed on the TBM cutter head which is exposed to harsh working environments such as high temperature and vibration that may not be sustainable over time.

With the availability of big data generated during tunnelling operations, supervised machine learning has then emerged as a potential complementary approach for predicting cutter wear in TBM by leveraging its ability to learn from existing data and identify complex patterns. Several studies have indicated that common TBM operational parameters such as thrust, torque, penetration rate and cutter rotation speed can provide a reasonable indication of the extent of cutter wear (Fukui et. al., 2006), (Toth et. al., 2013). Therefore, it is worthwhile to investigate the use of supervised machine learning to derive the relationship between TBM operational parameters and cutter wear. This paper aims to provide a preliminary exploratory study on the use of a machine learning-based prediction model for cutter wear detection in Singapore's tunnelling projects.

1.2 Project Used for Study

For this study, data was obtained from Land Transport Authority's (LTA) Thomson Line Contract T211 project that was carried out in the Bukit Timah Granite Formation. The project consisted of four tunnels using four slurry-shield TBMs. In this preliminary stage, only one TBM was selected to obtain the training data used to develop the machine learning model. This TBM will be termed as TBM A.

TBM A is a 6.63m slurry shield TBM consisting of three types of cutters, namely the double cutters, face cutters and gauge cutters (Figure 2). Each cutter is labelled with a position index that indicates its position radius, with smaller numbers representing cutters closer to the centre and larger numbers representing those closer to the outer circumference. Based on TBM A's CHI records, the tunnel transitioned through a granite profile with weathering grades of GII to GVI. Figure 3 shows the presence of rock and soil profiles, as well as mixed profiles, occurring at Zone 2.

Type of Cutter	Position Index
Double	1-8
Face	9-31
Gauge	32-47

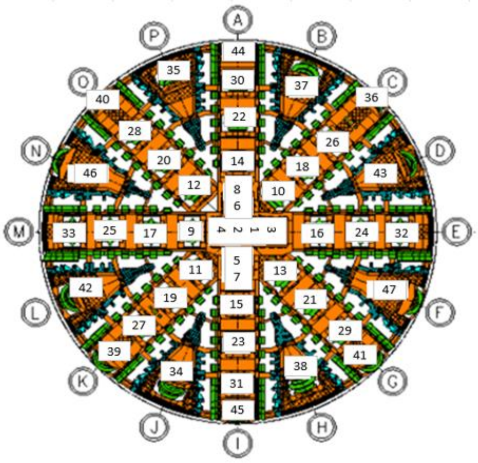


Figure 2. Details of TBM used for study.

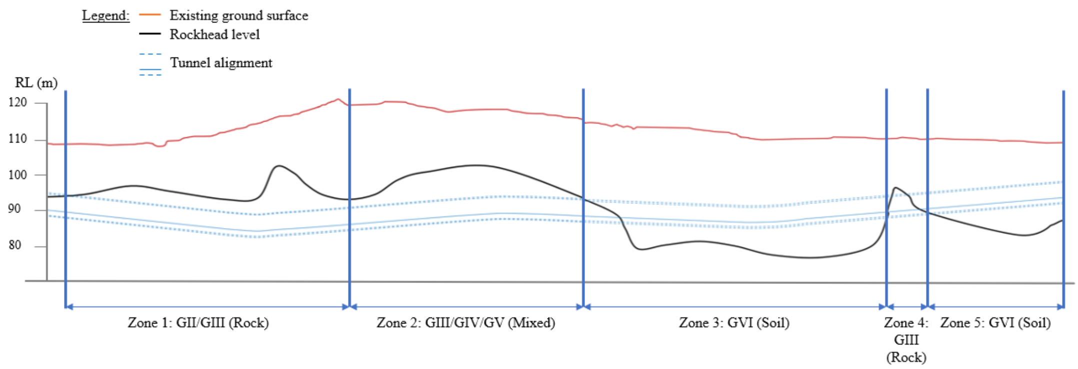


Figure 3. Details of soil profiles encountered based on TBM A's CHI records.

A study was also conducted to analyse TBM A's cutter replacement frequency. The objective was to derive any useful insights or patterns. As depicted in Figure 4, many abnormal wears were detected at

Zone 2, which corresponded to the heterogeneous soil layers. Moreover, the frequency of cutter replacements was higher for the gauge cutters, while the double cutters had the least replacement frequency.

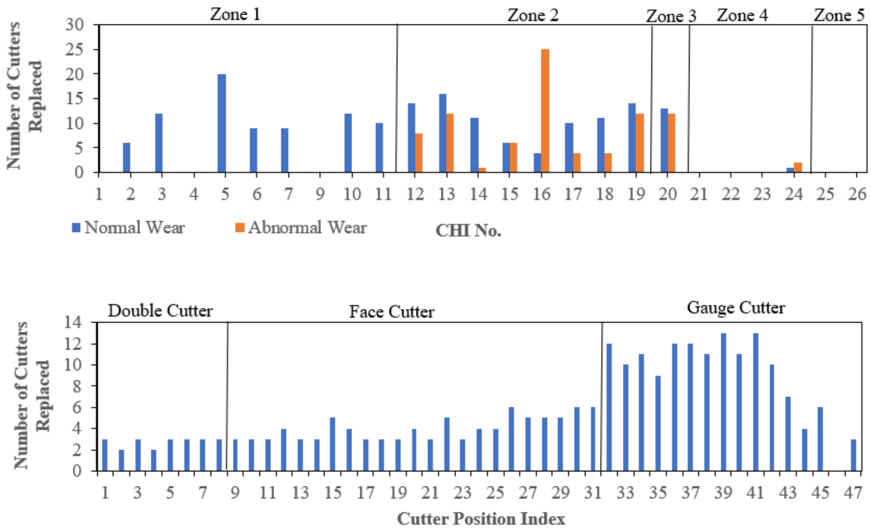


Figure 4. Details of TBM A’s cutters’ replacement frequencies.

2 DATA COLLATION

2.1 Input Data

This study uses TBM operational parameters for real-time inputs instead of interpreted data. The parameters were obtained from LTA's TEMS database, which captures TBM A's operational parameters throughout the entire operation process. To ensure selected input feature data hold sufficient weight to the output, irrelevant dataset was eliminated in three stages during preprocessing.

i. Data Filtering:
Raw data were extracted from the thrusting stage of the TBM since it is the stage when the cutters are in full operation. A recognition function was used to filter the data whereby all the values of thrust, advance rate, torque and rotation speed must not be 0.

ii. Data Sensitivity Analysis:
To ensure that the inputs accurately reflected the adverse ground conditions that cause cutter wear, a sensitivity analysis was conducted. Box-whisker plots were used to visually represent the different variations of operational parameters generated as the TBM transitioned through the five distinct soil zones as identified previously. This ensured that the parameters selected are relevant.

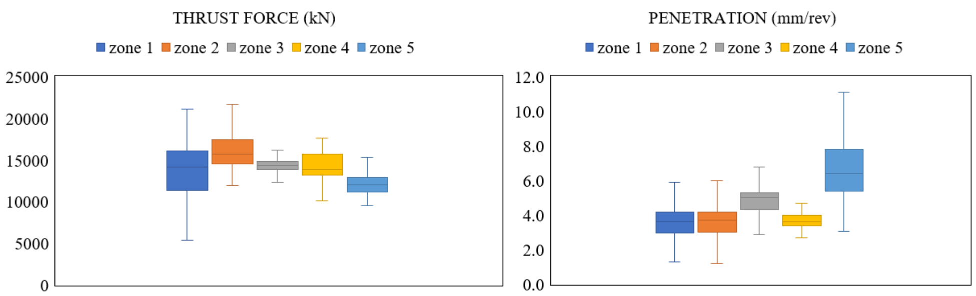


Figure 5. Sample box-whisker plots of chosen parameters to ensure data sensitivity.

iii. Data Correlation Analysis
Multicollinearity happens when two or more input variables are highly correlated with each other. This may lead to overfitting where the model learns the noise in the data. To determine the correlation, the Pearson correlation coefficient value was utilized as shown in Equation (1).

$$p(X,Y) = \frac{COV\ (X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

Where *p(X, Y)* = Pearson correlation coefficient between the two variables; *COV(X, Y)* = covariance of the two variables; $\sigma_X \sigma_Y$ = standard deviation values of the two variables.

As shown in Table 1, the correlation analysis was conducted using the correlation matrix of the selected input parameter features whereby features that are strongly correlated with one another can be identified. Generally, when a pair of feature has a correlation coefficient of more than 0.8, one of it can be removed to reduce the model complexity. For example, the advance rate and excavation rate are strongly correlated hence one of the input features can be removed.

Table 1. Correlation matrix of selected input features.

	Previous Wear	Position	Distance	Thrust	Torque	Penetration	Rot. No.	Exc. Rate	Rot. Speed	Adv. Speed	MC
Previous Wear	1										
Position	-0.0022	1									
Distance	-0.165	-0.012	1								
Thrust	0.25	-0.064	0.174	1							
Torque	0.251	-0.045	-0.232	0.725	1						
Penetration	-0.233	0.042	0.296	-0.717	-0.599	1					
Rot. No.	0.202	-0.053	-0.421	0.51	0.658	-0.541	1				
Exc. Rate	0.002	0.012	0.474	-0.036	-0.235	0.321	-0.545	1			
Rot. Speed	0.122	-0.049	0.053	0.613	0.347	-0.314	0.079	0.605	1		
Adv. Speed	-0.094	0.023	0.521	-0.305	-0.50	0.658	-0.627	0.942	0.428	1	
MC	-0.05	0.025	0.354	-0.239	0.078	0.338	-0.274	0.179	-0.283	0.238	1

2.2 Output Data

In order to use supervised machine learning to predict cutter wear, the extent of the wear must first be quantified. The cutter wear data were obtained from the CHI records of TBM A. The quantification of cutter wear has the potential to be a numerical or category output, hence this study explored the usage of both a regression and categorical prediction model to see which would fetch a better result.

To express the wear extent as a numerical value, the formula in Equation (2) was utilized.

$$Wear\ (\%) = \frac{Wear\ Value\ (mm)}{Limit\ Wear\ Value\ (mm)} \times 100 \quad (2)$$

Where wear value is the wear extent that was measured using a gauge tool in the CHI and limit wear value as the stipulated limit for each cutter discs type.

On the other hand, the primary concern is to determine if the cutters need to be replaced or not, hence it is also possible to simply quantify them as categorical labels ‘Replaced’ and ‘Unreplaced’.

2.3 Final Collated Training Dataset

After the extraction and analysis of training data from the CHI records and TEMS database, Tables 2 & 3 show the final chosen set of sample training datapoints that were fed into the regression and classification prediction models respectively. For the regression model, 676 datapoints were used for training while 921 datapoints were used to train the classification model. As tunnel boring operation is a time-series operation, the input parameters took the average value between each replacement to best represent the nature of soil encountered during the boring process.

Table 2. Sample input and output datapoints used for regression model.

Inputs										Output
Thrust (kN)	Torque (kNm)	Penetration (mm/rev.)	Distance (m)	Previous Wear (%)	Position Index	Rotation Number	Rotation Speed (rev/min)	Excavation Rate (m^3/min)	Moisture Content (%)	Wear (%)
11365	625	4.17	64.53	3.33	1	368	2.36	0.283	22	6.67
18800	2098	1.94	42.06	50	3	887	3.02	0.206	25	50
18331	2232	1.75	19.61	0	17	909	3.26	0.206	12	26.67
19209	2125	2.11	22.46	86.67	20	867	2.81	0.303	25	13.33
16771	1474	2.85	41.95	0	42	683	2.99	0.288	12	100

Table 3. Sample input and output datapoints used for classification model.

Inputs										Output
Thrust (kN)	Torque (kNm)	Penetration (mm/rev.)	Distance (m)	Previous Wear (%)	Position Index	Rotation Number	Rotation Speed (rev/min)	Excavation Rate (m^3/min)	Moisture Content (%)	Labels
11365	625	4.17	64.53	3.33	1	368	2.36	0.283	22	Unreplaced
18800	2098	1.94	42.06	50	3	887	3.02	0.206	25	Replaced
18331	2232	1.75	19.61	0	17	909	3.26	0.206	12	Unreplaced
19209	2125	2.11	22.46	86.67	20	867	2.81	0.303	25	Replaced
16771	1474	2.85	41.95	0	42	683	2.99	0.288	12	Replaced

The training dataset was randomly split into 90% for training and 10% isolated for testing, which adheres to the standard machine learning practice for acceptable splitting ratios. The training data was utilized to fit the model and fine-tune its hyperparameters, while the testing data was isolated for evaluating its performance by comparing the model's predictions to the actual outputs in the testing data.

3 PREDICTION MODELS

3.1 Machine Learning Algorithms

In supervised machine learning, various algorithms exist for training models. It's important to note that there is no single perfect algorithm for prediction models (Singh et. al., 2016). Experimenting with different algorithms is crucial to find the best fit for the specific use case. This paper will initially explore two common algorithms, Decision Trees (DT) and K-Nearest Neighbors (KNN), known for their ability to handle complex data and create regression/classification models.

The DT algorithm uses simple decision rules inferred from prior dataset features to predict the value of a target variable (Pedregosa et. al., 2011). It starts from a root node and branches to decision-making nodes, ultimately reaching a final leaf node (Figure 6). Each internal node tests a feature in the dataset, with each branch representing a possible test result and each leaf node indicating the final classification of the target variable (Che et. al., 2011). The hyperparameters that control the DT learning process are the depth of tree, minimum samples per leaf and split, and the criterion that determines the split.

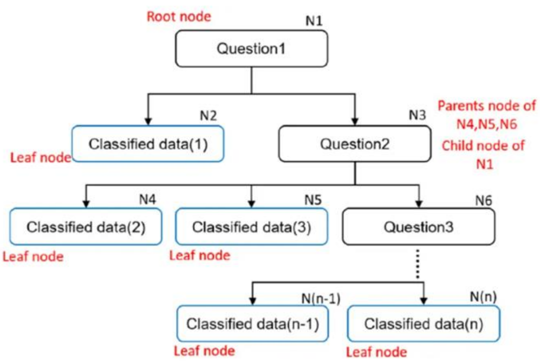


Figure 6. Conceptual diagram of the DT algorithm (Lee at. al., 2020).

The KNN is an algorithm that leverages proximity to perform classifications of predicted data. The prediction is made based on the datapoints that appear the closest around the unknown datapoint (Cunningham et. al., 2021). The quantity of the closest datapoints to determine a prediction is defined by the number of neighbors set in the algorithm. For example, as shown in Figure 7, data points are being depicted in a feature space whereby the K-Nearest Neighbors classifier has 7 nearest neighbors being set in the algorithm. For a classification problem, the unknown data shown will be set as Class B since 4 out of 7 nearest neighbors belong to Class B. Whereas for a regression problem, the algorithm will take the mean value of the 7 nearest neighbors. The hyperparameters that control the KNN learning process are the number of neighbors and the choice of distance measure to define the closest points.

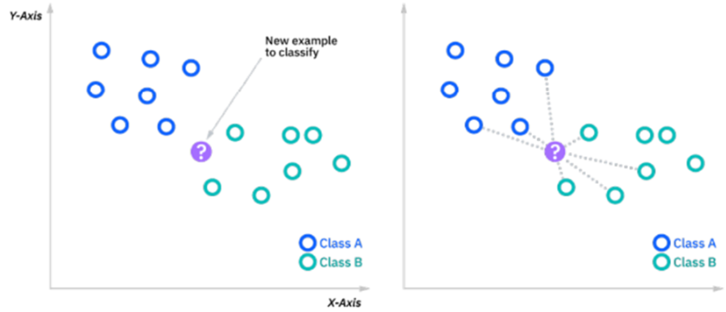


Figure 7. Conceptual diagram of the KNN algorithm (IBM, n.d.).

In this study, the coding of both algorithms was done Python Jupyter Notebook whereby the loading of the training dataset and the fitting of the algorithms took place. To ensure the best performance results, computational techniques such as the use of StandardScalar function to standardize the scale of the different input feature values for better analysis and GridSearchCV function to derive the optimal combination of hyperparameters that fetches the best performance scoring were utilized.

3.2 Performance Metrics

The performance of a machine learning prediction model is defined by how similar the predicted output is to the actual real-life output. Utilizing statistical measures is then a common methodology to define the performance of the developed machine learning algorithm.

For the regression model that produces a numerical value, common statistical indices will be used to evaluate the goodness of fit and the errors. The indices utilized are as shown in Equations (3) to (5).

$$R^2 = 1 - \frac{RSS}{TSS} \tag{3}$$

Where R^2 = coefficient of determination; *RSS* = sum of squares of residuals; *TSS* = total sum of squares.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}e_i^2} \tag{5}$$

Where MAE = mean absolute error; RMSE = root mean square error; n = number of data points; e = error value between the actual and predicted data point.

As for the classification model that produces categorical labels, the confusion matrix (Table 4) was used to evaluate the model’s performance. Leveraging the matrix, critical ratios Precision and Recall were used. Precision measures the percentage of correctly predicted positive outcomes out of all positive predictions, which affects TBM productivity. Recall measures the percentage of correctly predicted positive outcomes out of all actual positive outcomes, which affects TBM efficiency. Hence, in this context, these two ratios are more critical than evaluating the overall accuracy of the model.

Table 4. Confusion matrix for classification prediction model.

		Predicted	
		Replaced	Unreplaced
Actual	Replaced	True Positive (TP)	False Negative (FN)
	Unreplaced	False Positive (FP)	True Negative (TN)

$$Precision\ (\%) = \frac{TP}{TP + FP} \times 100 \tag{6}$$

$$Recall\ (\%) = \frac{TP}{TP + FN} \times 100 \tag{7}$$

4 RESULTS VALIDATION

4.1 Results of Testing Dataset

The performance of the prediction models was first validated on the 10% randomly sampled dataset that was isolated from the training dataset obtained from TBM A.

For the regression model, the statistical measures R^2, MAE and RMSE were used to evaluate the model's performance. Table 5 shows the data points used for testing, and Tables 6 & 7 display the performance results obtained using the DT and KNN algorithms, along with the hyperparameters set to achieve those results. Both algorithms produced similar R^2 values of 0.71 and 0.75, respectively, and exhibited an acceptable error margin, considering the cutter wear values from the testing dataset go up to 100%.

Table 5. Testing datapoints used for regression models.

No. of Datapoints	Wear % Range	Wear % Mean Value
68	3.33-100%	47.86%

Table 6. Results of regression model using DT algorithm.

Hyperparameters Set				**Results**		
Tree Depth	Min. Samples Per Leaf	Min. Samples Per Split	Criterion	R^2	MAE	RMSE
9	3	5	MSE	0.71	12.63	18.08

Table 7. Results of regression model using KNN algorithm.

Hyperparameters Set		**Results**		
No. of Nearest Neighbors	Choice of Dist. Measure	R^2	MAE	RMSE
4	Euclidean	0.75	11.65	17.12

For the classification model, the confusion matrix generated by both algorithms was evaluated to determine the models' performance (Figure 8). Similarly, both algorithms displayed similar performance as seen in Tables 8 & 9. It can be observed that the Precision and Recall rates of the unreplaced cutters are higher of up to 0.93. This is expected since the frequency of unreplaced cutters is usually much higher. However, the prediction model's ability to detect replacements is more critical. Nevertheless, when evaluating the rates of the replaced cutters, they were still within an acceptable range of at least 0.70.

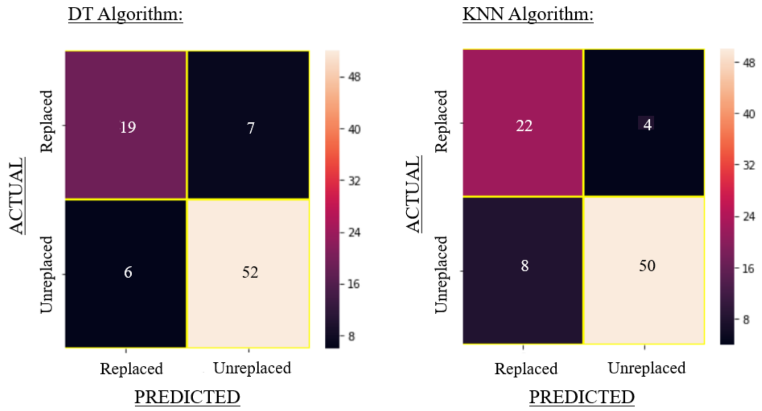


Figure 8. Extracted confusion matrix from classification models developed using DT and KNN algorithms.

Table 8. Results of classification model using DT algorithm.

Hyperparameters Set				**Results**			
Tree Depth	Min. Samples Per Leaf	Min. Samples Per Split	Criterion	Precision (Replaced)	Recall (Replaced)	Precision (Unreplaced)	Recall (Unreplaced)
8	5	5	Gini	0.70	0.73	0.88	0.86

Table 9. Results of classification model using KNN algorithm.

Hyperparameters Set		**Results**			
No. of Nearest Neighbors	Choice of Dist. Measure	Precision (Replaced)	Recall (Replaced)	Precision (Unreplaced)	Recall (Unreplaced)
5	Euclidean	0.73	0.85	0.93	0.86

4.2 Applying Trained Model on More Unseen Data

In addition to evaluating the randomly sampled testing dataset, it is also crucial to analyse the model's practical applications with more unseen data. In this regard, the trained model was further validated by applying it to the cutter wear data of another TBM in the same project. This TBM is termed as TBM B. Data from four specific CHIs in TBM B (namely CHI numbers 2, 4, 11, and 13) were used to evaluate the trained model's performance. These CHIs were selected because they consisted of variations in the types and number of cutters that need to be replaced, as shown in Figure 9. The trained model was evaluated based on its ability to identify the different variations of cutters that need to be replaced in each CHIs and to predict accurately that no replacements were necessary in CHI 13.

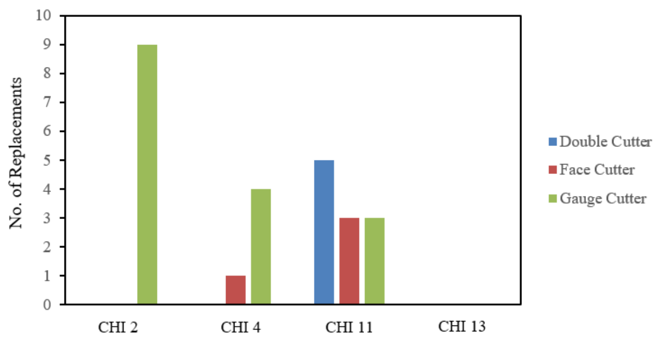
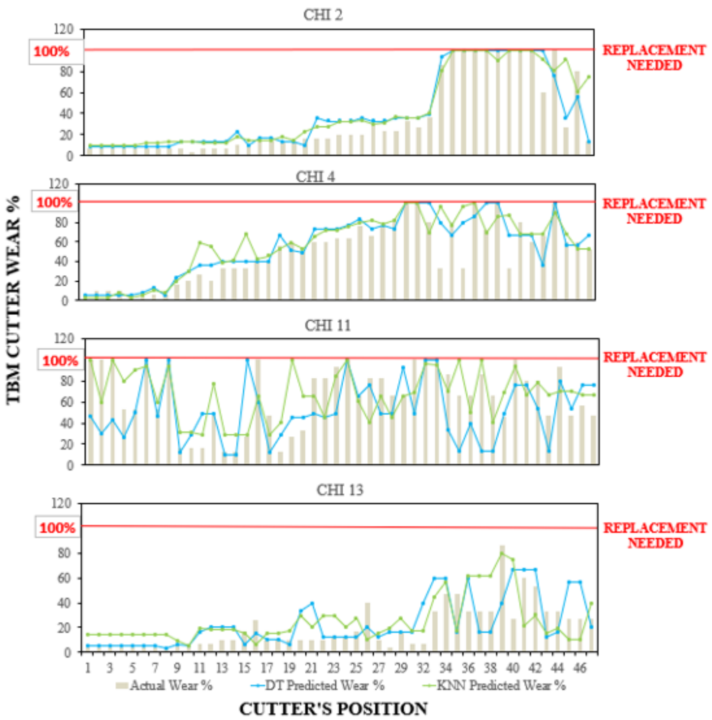


Figure 9. CHIs in TBM B chosen for validating trained model.

To validate the performance of the prediction models, the actual and predicted wear quantification data were extracted and compared. Comparison tables were used to analyse the actual versus predicted number of replacements required for each type of cutter disc in each CHI. The extracted wear quantification as numerical values from the regression prediction model are presented in Figure 10, while the extracted classification labels from the classification model are shown in Figure 11.

The predicted wear quantifications generally followed the trend of the actual data in both regression and classification approaches, despite some errors and misclassifications for individual cutters. However, in a more practical approach, the number of replacements as shown in the comparison tables was accurately identified. In the regression model, most predicted replacements for each cutter type were within a narrow margin of +/-1, indicating high precision. The classification model showed slightly lower precision, but most replacements were still within an acceptable margin of +/-2. Notably, the DT and KNN algorithms (for both classification and regression models) struggled to predict replacements for double cutters in CHI 11, while correctly predicting no replacements needed for CHI 13.



No. of Replacements for CHI 2

Type of Cutters	Actual	Predicted DT	Predicted KNN	Within +/- 1 DT	Within +/- 1 KNN
Double (Position Index 1-8)	0	0	0	✓	✓
Face (Position Index 9-31)	0	0	0	✓	✓
Gauge (Position Index 32-47)	9	9	8	✓	✓

No. of Replacements for CHI 4

Type of Cutters	Actual	Predicted DT	Predicted KNN	Within +/- 1 DT	Within +/- 1 KNN
Double (Position Index 1-8)	0	0	0	✓	✓
Face (Position Index 9-31)	1	1	1	✓	✓
Gauge (Position Index 32-47)	4	5	3	✓	✓

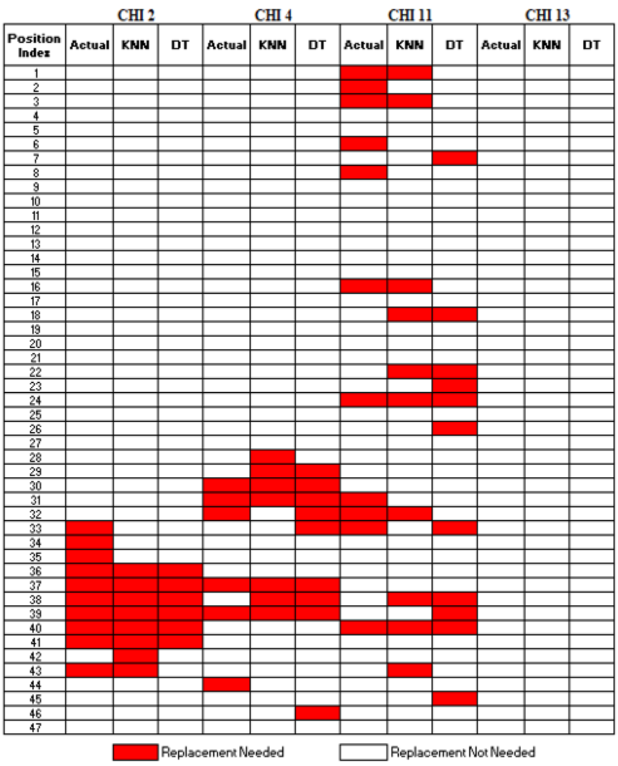
No. of Replacements for CHI 11

Type of Cutters	Actual	Predicted DT	Predicted KNN	Within +/- 1 DT	Within +/- 1 KNN
Double (Position Index 1-8)	5	2	2	✗	✗
Face (Position Index 9-31)	3	2	2	✓	✓
Gauge (Position Index 32-47)	3	2	2	✓	✓

No. of Replacements for CHI 13

Type of Cutters	Actual	Predicted DT	Predicted KNN	Within +/- 1 DT	Within +/- 1 KNN
Double (Position Index 1-8)	0	0	0	✓	✓
Face (Position Index 9-31)	0	0	0	✓	✓
Gauge (Position Index 32-47)	0	0	0	✓	✓

Figure 10. Extracted validation results and comparison tables from regression model.



No. of Replacements for CHI 2

Type of Cutters	Actual	Predicted DT	Predicted KNN	Within +/- 1 DT	Within +/- 1 KNN	Within +/- 2 DT	Within +/- 2 KNN
Double (Position Index 1-8)	0	0	0	✓	✓	✓	✓
Face (Position Index 9-31)	0	0	0	✓	✓	✓	✓
Gauge (Position Index 32-47)	9	5	7	✗	✗	✗	✓

No. of Replacements for CHI 4

Type of Cutters	Actual	Predicted DT	Predicted KNN	Within +/- 1 DT	Within +/- 1 KNN	Within +/- 2 DT	Within +/- 2 KNN
Double (Position Index 1-8)	0	0	0	✓	✓	✓	✓
Face (Position Index 9-31)	1	2	3	✗	✗	✓	✓
Gauge (Position Index 32-47)	4	3	4	✓	✓	✓	✓

No. of Replacements CHI 11

Type of Cutters	Actual	Predicted DT	Predicted KNN	Within +/- 1 DT	Within +/- 1 KNN	Within +/- 2 DT	Within +/- 2 KNN
Double (Position Index 1-8)	5	1	2	✗	✗	✗	✗
Face (Position Index 9-31)	3	5	4	✗	✓	✓	✓
Gauge (Position Index 32-47)	3	5	4	✗	✓	✓	✓

No. of Replacements for CHI 13

Type of Cutters	Actual	Predicted DT	Predicted KNN	Within +/- 1 DT	Within +/- 1 KNN
Double (Position Index 1-8)	0	0	0	✓	✓
Face (Position Index 9-31)	0	0	0	✓	✓
Gauge (Position Index 32-47)	0	0	0	✓	✓

Figure 11. Extracted validation results and comparison tables from classification model.

5 ERROR ANALYSIS

The machine learning models utilizing the DT and KNN algorithms for both classification and regression approaches showed promising results in predicting cutter wear for a slurry shield TBM in Bukit Timah Granite Formation. Nevertheless, to identify areas for model improvement, an analysis was conducted to determine the sources of errors. This section highlights the insights gained through error analysis using the regression dataset.

5.1 Higher Errors for Double Cutter Replacements

As shown in Figure 12, the error trends reflected in both DT and KNN algorithms were similar. Notably, the detection of replacement requirements for double cutters resulted in significantly more errors compared to the other types of cutters, such as gauge and face cutters. This could be attributed to the significantly lesser replacement data available for double cutters, in comparison to the other types. To minimise the errors occurring for double cutters replacement, sampling techniques can be explored. The proposed sampling techniques will be further elaborated in Section 6 of this paper.

Figure 12. Error analysis based on cutter type.

5.2 Poor Detection of Abnormal Wear

Another significant observation made was that the occurrence of errors in detecting abnormal wear type of cutters was significantly higher than that of normal wear type. This may be due to the limitations of using average values of TBM operational parameters taken between each replacement, which may not adequately capture the occurrence of mixed soil profiles where abnormal wear tends to frequently occur. To improve the model's accuracy in detecting abnormal wear, feature engineering techniques can be introduced to better represent the occurrence of mixed soil profiles. The proposed feature engineering techniques will be further elaborated in Section 6 of this paper.

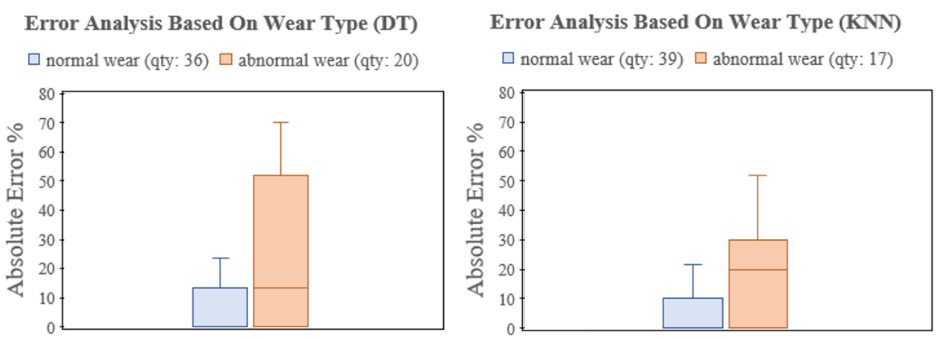


Figure 13. Error analysis based on cutter wear type.

6 FUTURE WORKS AND DISCUSSIONS

To propose model improvements and robustness for an efficient machine learning-based prediction model to improve tunnelling projects' CHI regime, the following future works are proposed:

i. Explore Sampling Techniques to Deal with Imbalanced Dataset:
Having an imbalanced dataset is one of the most common problems for machine learning whereby the minority class has the tendency to have more errors due to its limited dataset. This is the case for this project as the quantity of double cutters is significantly lesser than the face and gauge cutters.
To counter this problem, sampling techniques can be utilized to achieve a more balanced dataset whereby the accuracy of picking out the double cutters' replacements will not be compromised. This can be done through methods such as oversampling the minority class, undersampling the majority class or introducing synthetic samples such as utilizing the Synthetic Minority Oversampling Technique (SMOTE) (Junsomboon et. al., 2017). Additionally, it may be worthwhile to compartmentalize the machine learning models. For example, having one prediction model for double and face cutters and another separate model solely for gauge cutters can reduce the errors occurring for each type of cutter.

ii. Conduct Feature Engineering for Better Representation of Mixed-Face Soil Profiles:
To improve the model's performance in detecting abnormal wear, feature engineering can be conducted to ensure the instance of encountering a mixed soil profile is better represented.
An exploratory study can be done on the features to be introduced such as the soil parameters that are obtained during the CHI. Introducing additional statistical measures of the operational parameters such as the maximum or variance values may have a better indication of the complexity and variability of the soil profiles encountered by the TBM. This in turn will improve the accuracy of the prediction models by having better detection of abnormal wear occurring more frequently in mixed-face soil profiles.

iii. Expand the Dataset to Include Tunnels Operating in Other Soil Formations and With Different TBM Types:
In this preliminary study, the efficiency of the machine learning based prediction models was only validated against a slurry shield TBM operating in the Bukit Timah Granite ground formation. As the cutter wear pattern may be different across different TBM types operating in other ground formations, further studies should be conducted to determine the prediction models' ability to detect cutter wear in various tunnelling operation scenarios.

iv. Explore More Machine Learning Algorithms:
This paper has explored the use of DT and KNN algorithms to develop prediction models. With the expansion of the training dataset, more algorithms can be explored. For example, exploring the usage of Random Forest (an ensemble learning algorithm based on DT) or neural networks (a deep learning algorithm).

7 CONCLUSION

This preliminary study highlights the potential of using machine learning to detect cutter wear in TBM during tunnelling operations. The results obtained from the testing dataset showed that the regression prediction models using DT and KNN algorithms obtained similar R^2 values of 0.71 and 0.75 respectively, while the classification prediction models using both algorithms obtained Precision and Recall rates ranging between 0.70 to 0.85. Additionally, the performance of the trained prediction models was further validated by applying them to more unseen data using a practical approach, demonstrating that both regression and classification approaches accurately predicted the required replacements for each type of cutter in specific CHIs with a narrow margin of +/- 1 and 2, respectively. These findings suggested that both regression and classification approaches using KNN and DT algorithms are feasible, although it is essential to consider the trade-offs between these two approaches in terms of interpretability, computation time, and model complexity.

While the results are promising, the development of an accurate and reliable prediction model requires further exploration. As discussed in Section 6, sampling techniques and feature engineering can be investigated to improve the robustness of the prediction models. Furthermore, since the study was solely validated against slurry shield TBM operating in the Bukit Timah Granite Formation, the next step of the study should expand the training dataset to include more diverse ground conditions and TBM types. It is also crucial to ensure that the model's predictions align with the practical implications of replacing cutters, such as cost and downtime.

Overall, this paper demonstrates the usage of machine learning as a viable alternative to enhance the TBM CHI process by providing real-time predictions on the necessity for cutter replacements using operation parameter inputs generated during tunnelling. Such predictions can facilitate informed decision-making, ultimately improving the efficiency of tunnelling projects.

8 ACKNOWLEDGEMENT

This study serves as the final year Capstone Project of the first author at her university, Singapore Institute of Technology. The first author would like to thank her academic supervisor, Professor Kim Yong Min for his invaluable guidance and feedback, which have enriched this project. Additionally, she would also like to express gratitude to Land Transport Authority for generously providing the TBM data, which has played a pivotal role in the completion of this project.

9 RFERENCES

Bilgin, N., Copur, H., & Balci, C. (2012). Effect of replacing disc cutters with chisel tools on performance of a TBM in difficult ground conditions. *Tunnelling and underground space technology*, *27*(1), 41-51.

Che, D., Liu, Q., Rasheed, K., Tao, X. (2011). Decision Tree and Ensemble Learning Algorithms with Their Applications in Bioinformatics. In: Arabnia, H., Tran, QN. (eds) Software Tools and Algorithms for Biological Systems. Advances in Experimental Medicine and Biology, vol 696. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-7046-6_19

Cunningham, P., & Delany, S. J. (2021). K-nearest neighbour classifiers-a tutorial. *ACM Computing Surveys (CSUR)*, *54*(6), 1-25.

Fukui, K., & Okubo, S. (2006). Some attempts for estimating rock strength and rock mass classification from cutting force and investigation of optimum operation of tunnel boring machines. *Rock Mechanics and Rock Engineering*, *39*(1), 25.

Junsomboon, N., & Phienthrakul, T. (2017). Combining over-sampling and under-sampling techniques for imbalance dataset. In *Proceedings of the 9th International Conference on Machine Learning and Computing* (pp. 243-247).

Karami, M., Zare, S. & Rostami, J. Tracking of disc cutter wear in TBM tunneling: a case study of Kerman water conveyance tunnel. *Bull Eng Geol Environ* **80**, 201–219 (2021). https://doi.org/10.1007/s10064-020-01931-7

Ko, T. Y., & Lee, S. S. (2020). Effect of rock abrasiveness on wear of shield tunnelling in bukit timah granite. *Applied Sciences*, *10*(9), 3231.

IBM. (n.d.). KNN. IBM. Retrieved April 11, 2023, from https://www.ibm.com/topics/knn

Lan, H., Xia, Y., Ji, Z., Fu, J., & Miao, B. (2019). Online monitoring device of disc cutter wear–Design and field test. *Tunnelling and Underground Space Technology*, *89*, 284-294.

Lee, Gi-Jun & Ryu, Hee-Hwan & Kwon, Tae-Hyuk. (2020). A study on the optimum cutter spacing ratio according to penetration depth using decision tree-based and SVM regressions. Journal of Korean Tunnelling and Underground Space Association. 22. 501-513. 10.9711/KTAJ.2020.22.5.501.

Liu, B., Yang, H., & Karekal, S. (2020). Effect of water content on argillization of mudstone during the tunnelling process. *Rock Mechanics and Rock Engineering*, *53*, 799-813.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 12, 2825-2830.

Singh, A., N. Thakur and A. Sharma, "A review of supervised machine learning algorithms," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2016, pp. 1310-1315

Tóth, Á., Gong, Q., & Zhao, J. (2013). Case studies of TBM tunneling performance in rock–soil interface mixed ground. *Tunnelling and Underground Space Technology*, *38*, 140-150.