Geomechanics and Engineering, *Vol. 38, No. 6 (2024) 633-645*
https://doi.org/10.12989/gae.2024.38.6.633

Analysis of disc cutter replacement based on wear patterns using artificial intelligence classification models

Yunhee Kim[a], Jaewoo Shin[b] and Bumjoo Kim*

Department of Civil and Environmental Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-Gu, Seoul, 04620, Republic of Korea

(*Received November 28, 2023, Revised January 28, 2024, Accepted February 7, 2024*)

Abstract. Disc cutters, used as excavation tools for rocks in a Tunnel Boring Machine (TBM), naturally undergo wear during the tunneling process, involving crushing and cutting through the ground, leading to various wear types. When disc cutters reach their wear limits, they must be replaced at the appropriate time to ensure efficient excavation. General disc cutter life prediction models are typically used during the design phase to predict the total required quantity and replacement locations for construction. However, disc cutters are replaced more frequently during tunneling than initially planned. Unpredictable disc cutter replacements can easily diminish tunneling efficiency, and abnormal wear is a common cause during tunneling in complex ground conditions. This study aims to overcome the limitations of existing disc cutter life prediction models by utilizing machine data generated during tunneling to predict disc cutter wear patterns and determine the need for replacements in real-time. Artificial intelligence classification algorithms, including K-nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), and Stacking, are employed to assess the need for disc cutter replacement. Binary classification models are developed to predict which disc cutters require replacement, while multi-class classification models are fine-tuned to identify three categories: no replacement required, replacement due to normal wear, and replacement due to abnormal wear during tunneling. The performance of these models is thoroughly assessed, demonstrating that the proposed approach effectively manages disc cutter wear and replacements in shield TBM tunnel projects.

Keywords: artificial intelligence; disc cutter wear pattern; excavation data; multi-class classification model; shield TBM

1. Introduction

Tunnel Boring Machine (TBM), as a mechanized excavation method, is progressively expanding due to the acceleration of urban underground development. Consequently, various studies are underway to investigate the stability and excavation efficiency in complex or soft ground conditions(Kovari and Ramoni 2006, Jeong *et al*. 2018, Kim *et al*. 2018, Rezaei *et al*. 2019, Kim *et al*. 2020). In a shield TBM, the disc cutter is subject to natural wear via forces normal to its face and rolling forces generated by thrust and torque during the excavation process. When a disc cutter reaches its wear limit, it must be replaced. However, unpredictable, non-uniform disc cutter wear patterns reduce construction efficiency and increase project time and costs. In particular, when tunneling through complex ground conditions, disc cutters frequently experience abnormal wear such as chipping or and/or uneven wear patterns. The Colorado School of Mines (CSM) model (Roastami 1997), Norwegian University of Science and Technology (NTNU) model (Bruland 1998, Macias 2016), and Gehring model (Gehring 1995) are widely used to predict the life of disc cutters during the design stage for hard-rock TBMs. However, these models are primarily based on experimental data from a single rock mass and field experience and thus may not accurately predict the on-site location where replacement will be required; rather, they tend to only predict the volume or distance that a single disc cutter can excavate for a particular type of rock.

Various studies have investigated disc cutter wear patterns. Frenzel (2008) proposed a number of factors influencing disc cutter wear, while Connors (2017) described the challenges of slurry shield TBM tunneling in mixed ground conditions that included igneous rock. Disc cutter wear mechanisms and wear patterns under heterogeneous ground conditions have also been widely studied (Ren *et al.* 2018, Ellecosta *et al.* 2018, Farrok and Kim 2018, Butt *et al*. 2021, Yang *et al*. 2021, Fang *et al.* 2021, Karami *et al*. 2021). Recently, with the development of artificial intelligence (AI), predictions of disc cutter wear and replacement cycles have been made using various AI algorithms (Afradi *et al.* 2021, Mahmoodzadeh *et al*. 2021, Yu *et al*. 2021, Kim *et al*. 2022).

The present study analyzed the replacement of disc cutters and their wear patterns at two slurry shield TBM construction sites in accordance with the TBM excavation distance. Site A is a railway TBM site previously described and utilized in the model for evaluating the necessity of disc cutter replacement by Kim *et al*. (2022). The dataset from Site A in the previous study (Kim *et al*. 2022) assumed to be replaced of disc cutters with wear of over 80% of wear rate. However, in the study, the disc cutter replacement data

*Corresponding author, Professor
E-mail: bkim1@dongguk.edu
[a]Ph.D.
[b]Ph.D. Student

http://www.techno-press.org/?journal=gae&subpage=7
ISSN: 2005-307X (Print), 2092-6219 (Online)

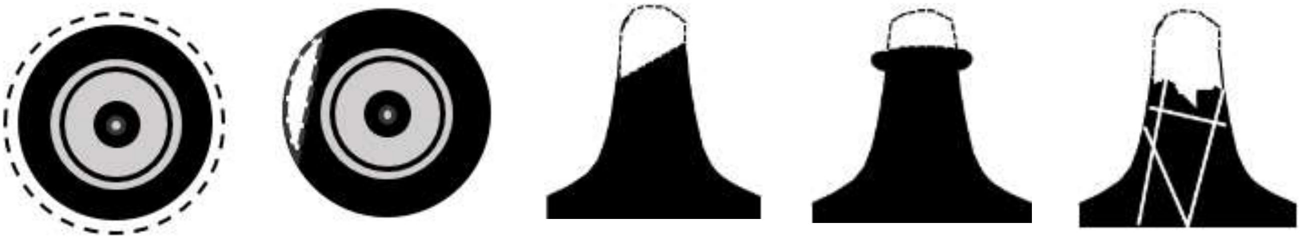


Fig. 1 General disc cutter wear patterns under heterogeneous ground conditions (after Ellecosta 2018)

was constructed using information form a single point in time when the disc cutter replacement took place during the excavation. Despite an increased imbalance between replacement and non-replacement data compared to the previous study, the utilization of real data contributed to achieving better learning outcomes in machine learning. In this study, excavation data, ground data, and CHI (Cutter head Intervention) data were additionally collected and analyzed from a domestic tunneling site (called as Site B) using a shield TBM with a similar diameter. Two datasets for binary and multi-class classification models were constructed from TBM data recorded during excavation and CHI data. The datasets from Sites A and B were used in four classification models: K-nearest neighbor (KNN), support vector machine (SVM), decision tree (DT), and stacking ensemble algorithms. To generalize the model, the datasets from the two construction sites were combined and used in these classification models. The binary classification model was used to predict which disc cutters required replacement, while the multi-classification models were developed to produce three classifications: no replacement required, replacement due to normal wear, and replacement due to abnormal wear during tunneling.

The rest of this study is organized as follows. Section 2 describes the general disc cutter wear patterns, especially focusing on abnormal wear and their underlying reasons. Section 3 presents an on-site analysis for two slurry TBM sites covering ground conditions, TBM specifications, excavation data, and analysis of disc cutter wear pattern and life during excavation. The data preprocessing and the machine learning (ML) algorithms used in the study are summarized in Section 4. Section 5 then presents and discusses the results for each classification model, and a conclusion is given in Section 6.

2. Wear patterns of disc cutter

A disc cutter is a TBM cutting tool that naturally wears during excavation. However, this wear exhibits different patterns depending on the rock strength, cutter type (Xiaokang *et al.* 2023), ground complexity, cutter head design (e.g., number of cutters, cutter spacing, and open ratio), clogged disc cutters, and jammed bearings (Farrokh 2013). The occurrence of abnormal wear is the result of multiple interacting factors, and accurately predicting it can be challenging. However, proactive prediction and management are essential for enhancing the efficiency of TBMs. The complex interplay of various causes underscores the need for a comprehensive approach. Real-time monitoring and regular inspections are crucial for minimizing wear and optimizing the performance of TBMs. Effectively addressing these factors through prediction and management strategies can significantly contribute to improving TBM efficiency. Fig. 1 presents general disc cutter wear patterns for heterogeneous ground conditions. Ellecosta (2018) identified five types of macroscopic wear for cutter rings and microscopic wear from scanning electron microscopic analysis of the cutter surface and optical light microscopic analysis of metallographic cross-sections. Lislerud (1997) separated primary wear mechanisms into two classes: macroscopic and microscopic. Macroscopic wear mechanisms include thermal fatigue under high compressive stress or impact shock in mixed face conditions, and microscopic wear mechanisms include surface fatigue, chemical reactions, adhesion, and abrasion.

In general, normal wear is an idealized disc cutter wear pattern where, while tunneling through a single rock layer, the cutter ring undergoes uniform wear along its linear rotation distance. Various patterns of abnormal wear, however, occurring in complex/difficult ground conditions, excavation scenarios, TBM mechanical faults, and more. Uneven wear is a representative pattern of abnormal wear, occurring when the housing installed on the cutter head is deformed or when the rotation of the disc cutter is stopped due to jamming. In addition, when the excavating in complex ground conditions, causing wear on only one side of the cutter ring. Fracture-types like breakage or brittle fracture or chipping occur when encountering extremely hard rock during the excavation process, causing some portions of the cutter ring to detach due to brittle failure at the rock boundary. Mushrooming occurs when excavating homogeneous rock with high strength, causing deformation of the disc cutter's appearance when the hardness of the disc cutter is relatively lower than that of the rock. In domestic sites, disc cutter replacements have occurred due to various forms of abnormal wear, leading to a decrease in the efficiency of adjacent cutters on the cutter head. This deterioration affects the efficiency of the excavation process, leading to increased thrust, torque, and irregularities, ultimately resulting in a decrease in excavation speed and necessitating equipment inspections. It is challenging to individually assess various patterns of disc cutter wear. In this study, we differentiated between models that determine replacement and those that

Table 1 Parameters determined from borehole loading and lab tests at Site A and B

SITE	Ground layer	Cohesion (kPa)	Friction angle (°)	Deformation modulus (MPa)	Poisson ratio (-)
Site A	Sedimentary layer (cohesion soil)	20	0	3	0.40
	Sedimentary layer (sandy soil)	0	28	15	0.35
	Sedimentary layer (gravel layer)	0	33	70	0.33
	Weathered soil	15	30	60	0.33
	Weathered rock	30	33	250	0.30
	Soft rock	500	33	3,000	0.29
	Hard rock	3,000	38	9,000	0.22
Site B	Soft rock	-	-	8,570-56,500	0.26 - 0.3
	Hard rock	10.7–20.6	41–51	31,000-60,400	0.21-0.27

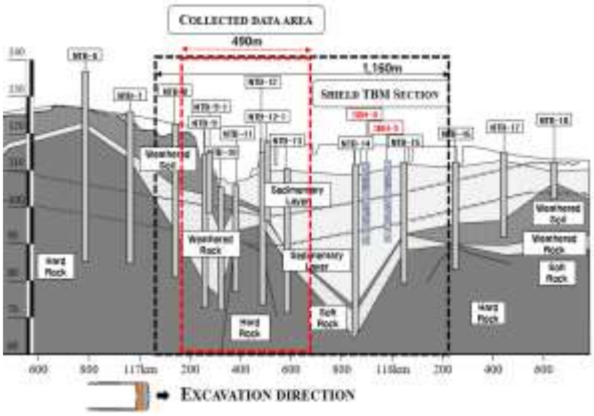


Fig. 2 Longitudinal geological profile at Site A

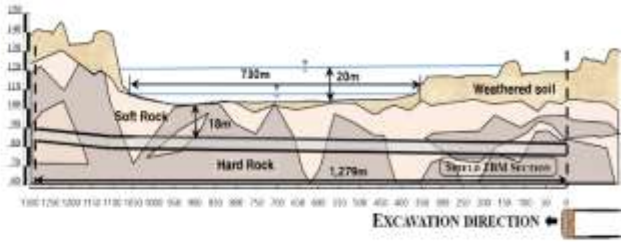


Fig. 3 Longitudinal geological profile at Site B

classify normal and abnormal wear when replacements are necessary.

3. Site analysis

In chapter 3, information on two Shield TBM sites referred to as A and B was presented, encompassing ground conditions, machine data, and the analysis results of disc cutter wear. The description of TBM Site A is derived from and has been modified based on Kim et al. (2022).

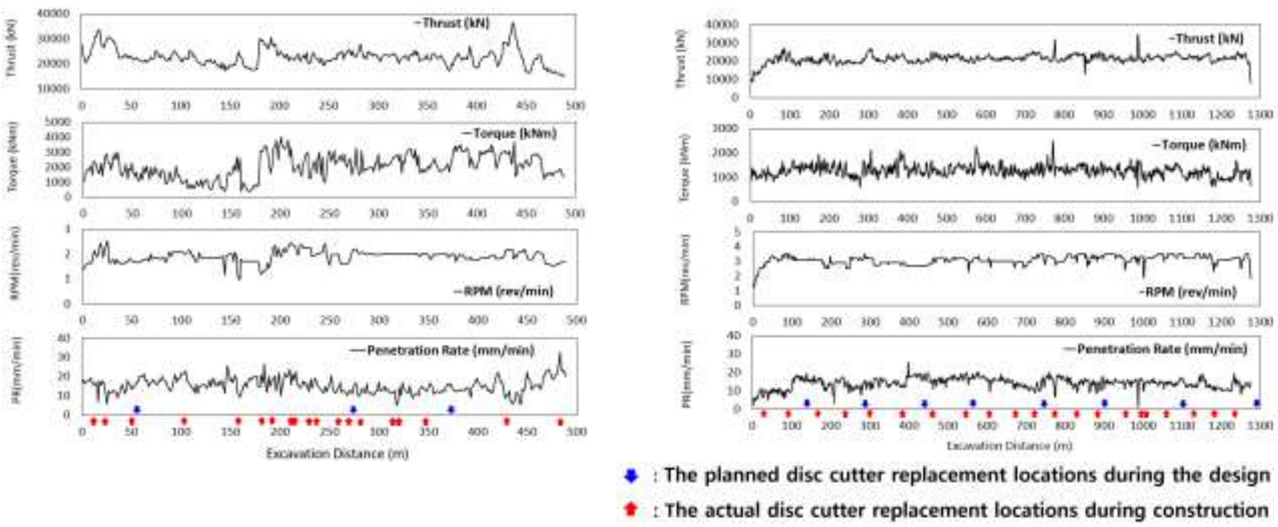
3.1 Ground conditions

Site A is a railway construction site in South Korea with a total tunnel length of 3,930 m, of which 1,160 m was excavated using a slurry shield TBM. The site was mainly characterized by biotite granite and contains vertical joints.

The main mineral components include orthoclase, microcline, plagioclase quartz, and biotite. Fig. 3 presents the complex strata at Site A, consisting of a sedimentary layer, weathered soil, weathered rock, and soft and hard rock. The data collection area for the study, which is indicated by the red dashed box in Fig. 3, was characterized by more complicated ground conditions than the other sections. In addition, 18–70 cm diameter boulders were also present at the depth of the excavated tunnel. Site B is a metro construction site in Seoul and about 730 m of the total 1,279 m of the shield TBM construction section was positioned directly under the Han River. As shown in Fig. 3 the water level was around 3 to 20 m and the minimum overburden depth was 18 m. Site B was characterized by gneiss, with soft rock and hard rock accounting for about 40% and 60% respectively. Table 1 summarizes the soil and rock parameters for each layer determined from borehole loading tests, triaxial tests and rock core test from site A and B. According to rock core testing at site B, the strength of the rock was 22–116 MPa, while the soil was classified as SM and GP-GM following the Unified Soil Classification System (USCS). The quartz content was 39–51%, representing very abrasive ground conditions under the Han River bedrock.

3.2 TBM equipment

The main specifications of the slurry shield TBM at site A and B are presented in Table 2. A slurry shield TBM manufactured by Herrenknecht (Germany) was used at Site A. It was equipped with a displacement cylinder, articulation cylinder, and drilling line to prevent face plate jamming, correct meandering, facilitate curvature construction, identify the geological conditions, and reinforce the ground. The TBM was designed for use in excavating mixed ground. It also maintained the chamber pressure with a bubble chamber, while jamming with excavated rock or gravel was prevented by installing a crusher. A total of 44 17-inch disc cutters were installed, while 110 scrappers were installed on the cutter head.



(a) Site A (b) Site B

Fig. 4 Shield TBM data during excavation from (a) Site A and (b) Site B

Table 2 Major specifications of the shield TBM at Site A and B

Specifications	Site A	Site B
TBM diameter (mm)	8,410	7,660
External diameter (mm)	8,370-8,390	7,580
Length (mm)	10,500	10,700
Total Thrust (kN)	59,464	56,000
Torque (kN·m)	4,680-6,786	5,325-13,320
RPM (rpm)	Max. 3.1	Max. 6.0
Penetration Rate (mm/min)	50	60
Maximum tunnel face pressure (bar)	4.5	6.5
Disc cutter diameter (inch)	17	17
Number of disc cutter(EA)	44	53

The disc cutters configuration in cutter head includes 40 single and 2 double disc cutters. The double disc cutters are used as center cutters, while among the single cutters, 38 are used as face cutters and 2 as gauge cutters. A slurry TBM manufactured by Kawasaki (Japan) was employed at Site B because this site passed under the Han River and was thus subject to high water pressure. The TBM equipment was designed to ensure a cutter head open ratio of around 20% to prevent blockages when crushing the hard and soft rock layers. The power and torque of the cutter head were set to account for the complicated ground conditions. The disc cutters at site B consists of a total of 53 disc cutters, with 45 single disc cutters comprising 2 gauge cutters and 43 face cutters, arranged from the center, and 4 double disc cutters.

3.3 Excavation data and CHI design

Excavation data are generated when the TBM is in operation and controlled according to the ground conditions. During this process, a shield TBM accumulates hundreds of data points that indicate the state of the ground conditions. The factors that have the most influence on disc cutter wear are the strength and abrasiveness of the ground but, because this is difficult to predict continuously, TBM data are used as input for prediction models. In the present study, thrust, torque, cutter head rotation per minute (RPM), and penetration rate (PR), which are strongly associated with disc cutter wear, were employed as input features for the prediction disc cutter wear pattern and replacement models. Thrust refers to the normal force on the tunnel face used to advance underground and is the standard for judging rock strength relatively. It also acts as a normal force on each disc cutter. Torque is the rotational force, while RPM is the rotational speed of the cutter head, which generates the rolling force on the disc cutters. Higher thrust can increase the normal force, and higher torque and RPM increase the rolling force on the disc cutters, leading to more wear and breakage. The PR is the principal measure of TBM performance and is calculated as the product of the disc cutter press-in depth (mm/rev) and cutter head RPM. Fig. 4(a) presents the machine data generated for a 490 m section at Site A, which is indicated by the red box in Fig. 2. The thrust, torque, RPM, and PR at Site B (1,279 m) are presented in Fig. 4(b).

The rate of advance and the distance or excavation volume for each disc cutter need to be calculated during the design stage to estimate the excavation period and cost. Using models such as the Gehring, CSM, NTNU and KICT-SNU (Ministry of Construction & Transportation, 2007) models to predict the cutter life can allow the distance or excavation volume to be calculated, thus the number of disc cutters required for a project can be predicted. At Site A, it was predicted that three replacements would be needed for the 490 m study section but in construction phase disc cutter inspections and replacements were carried out at 19 locations shown in Fig. 4(a). At Site B, around seven CHIs were planned for the 1,279 m section (Fig. 4(b)) but 21 CHIs were performed

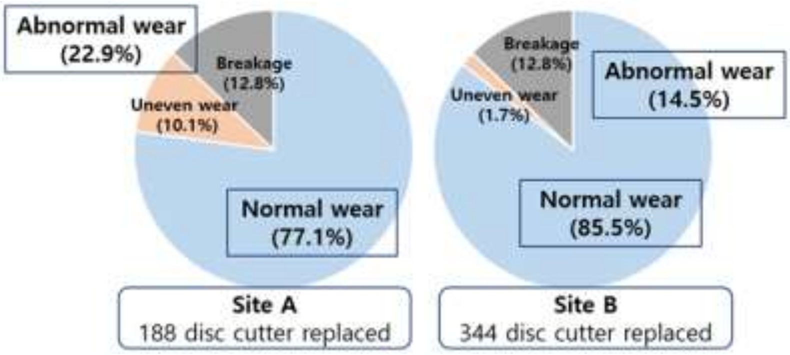


Fig. 5 The wear pattern ratio during disc cutter replacement at site A and B

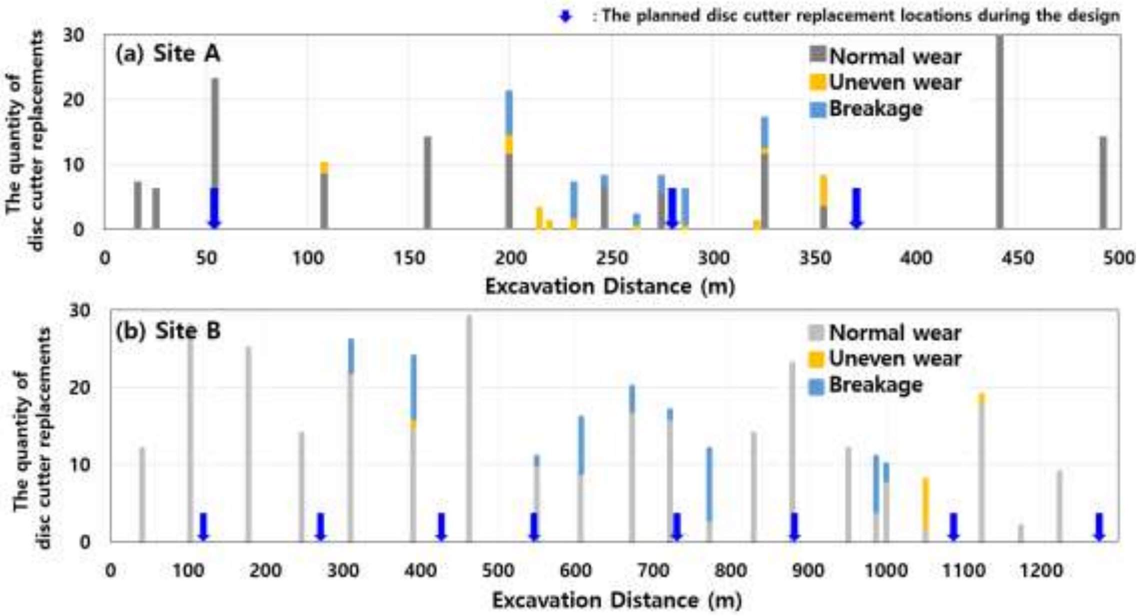


Fig. 6 The quantity of disc cutter replacements based on the wear pattern in excavation distance at (a) Site A and (b) Site B

during excavation. The distance that one disc cutter can excavate before reaching its wear limit was calculated to be 204 m in hard rock and 232 m in soft rock using the cutter life index (CLI) and the KICT-SNU method.

3.4 Disc cutter wear patterns at sites

The CHI data for Sites A and B were analyzed to determine the total number of disc cutters replaced during excavation and the frequency of the different wear patterns. Fig. 5 illustrates the ratio of disc cutter wear patterns observed at site A and B during excavation. Most disc cutters were replaced due to normal wear, but at Sites A and B, disc cutter replacements occurred due to abnormal wear at rates of 22.9% and 14.5%, respectively.

Fig. 6 shows the quantity of disc cutter replacements based on the wear pattern in excavation distance at Site A and B. Of the total 199 disc cutter replacements within the 490 m section at site A, 145 were due to normal wear, 19 were due to uneven wear, and 24 were due to breakage. The machine was stopped 19 times to exchange the disc cutter rather than the three times planned at the design stage. The total number of disc cutters required at site B was 344, with 294 of these due to normal wear, 6 due to uneven wear, and 44 due to breakage. Normal and abnormal wear thus accounted for 85.5% and 14.5% of total replacements, respectively. The replacement rate for normal wear was about 8% higher than that of Site A, and the replacement rate for uneven wear was lower by 2%.

In site A, frequent disc cutter replacements occurred particularly between excavation distances of 200 m to 350 m, with an increased occurrence of abnormal wear replacements compared to other sections during TBM excavation. As a result of analyzing the ground in this excavation section, it is constituted to a complex layer with a mixture of weathered soil, cohesive soil, and fractured rock, with a significant presence of large boulder leading to increased disc cutter breakage. From the ground profile presented in Fig. 3 at site B, it is observed that soft rock and hard rock appear alternately, and in sections where the geological/rock conditions change, there is a noticeable increase in both machine data variations and disc cutter breakages.

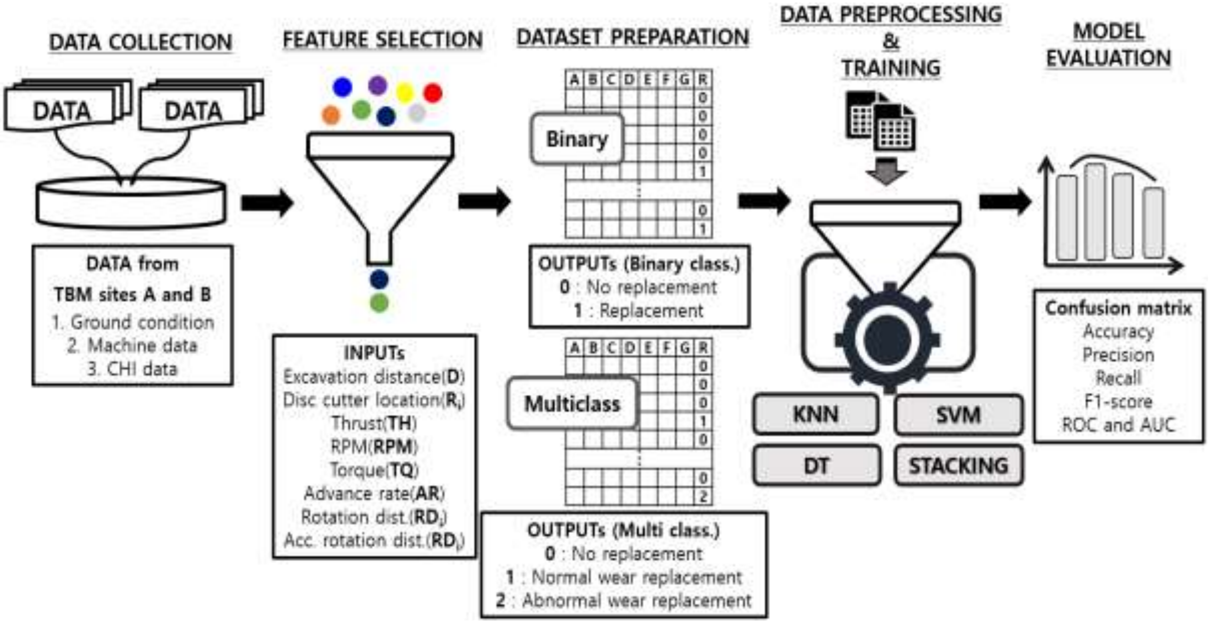


Fig. 7 Schematic diagram of the study process

4. Dataset and ML algorithms

This section presents the process used to construct the datasets for the binary and multi-class classification models. Fig. 7 presents the study process for the present study, including data collection, feature selection, model development, and evaluation

4.1 Selected features in dataset

It is essential to extract influential factors highly correlated with disc cutter wear as input data to develop a model capable of determining the replacement of disc cutters. The factors directly influencing disc cutter wear such as the uniaxial compressive strength, mineral content, or ground complexity are challenging to utilize directly as input data for an artificial intelligence model due to discontinuous data. Therefore, in this study, indirectly assess ground-related elements and used key machine data, such as thrust, torque, and rotation speed, which exert forces on the disc cutter. Additionally, advance rate, rotation distance and accumulated rotation distance of each disc cutter were incorporated. All data are continuously collected throughout the tunneling process and it is suitable for the convenient training and learning of AI models. The factors most strongly affecting the wear of the disc cutters were identified using Pearson correlation analysis, and these were used as input features into the models for the binary and multi-class models. The Pearson correlation result for selected input and output features in the models is shown in Fig. 8. In this process, ML algorithms selected only optimal features that ensured superior model performance.

4.2 Dataset composition

The datasets for each site were constructed using Shield TBM machine data, rotation advance distance in each disc

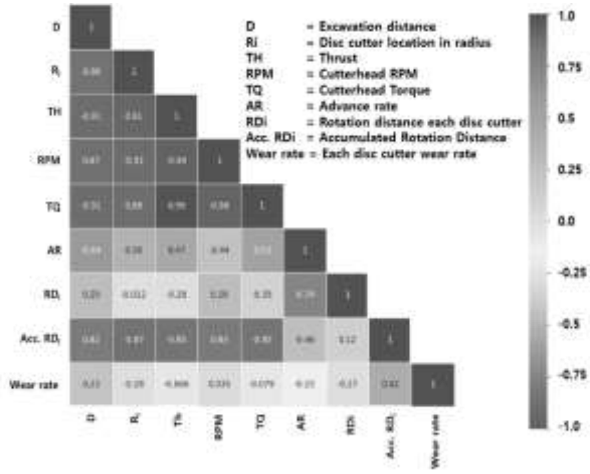
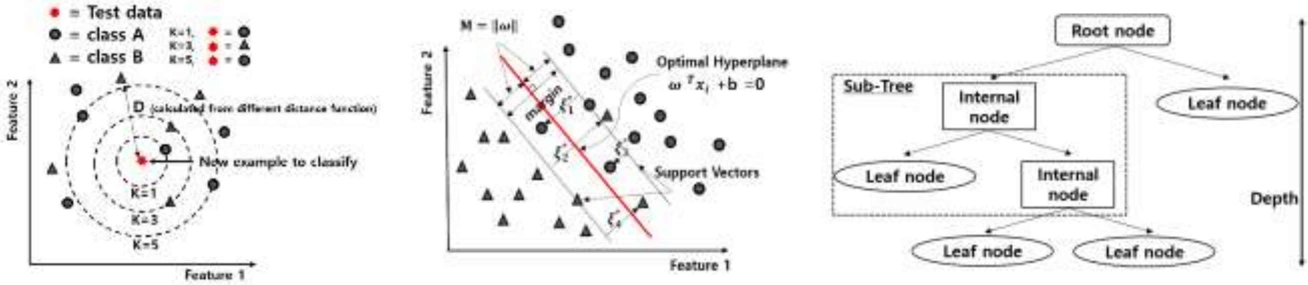


Fig. 8 Pearson correlation results for input and output features

cutter presented in Section 4.1 as input. These data were collected at excavation distances of 490 m at site A and 1279 m at site B. The output was established based on the analysis results of CHI data. For site A, only 24 out of the 44 disc cutters were utilized due to human error detection on data, while for site B, 45 disc cutters, excluding the center cutter, were used as the model's data. As illustrated in Fig. 7, the output was represented as 0 or 1 in the binary classification model and 0,1 or 2 in multi-class classification model. The number of data samples for the model is 7848 for site A and 36720 for site B. In the algorithm, the dataset was split into a 70/30 ratio for training and testing sets.

4.3 Data preprocessing

Data preprocessing was also used to improve model training and overall performance. Outliers in the data were first removed, after which min–max scaling and the



(a) KNN (b) SVM (c) DT

Fig. 9 Machine learning classification models (after Kim *et al.* (2022))

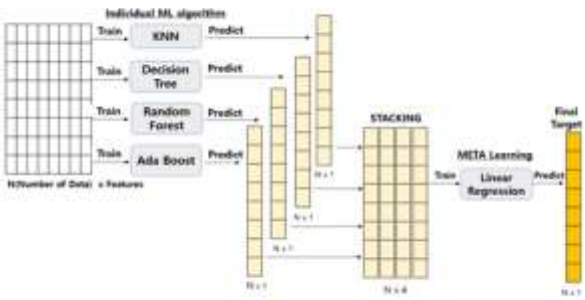


Fig. 10 Stacking ensemble model (after Kim *et al*. (2022))

synthetic minority oversampling technique (SMOTE) were used for preprocessing. Min–max scaling transforms the range of the features to a scale of [0,1] to remove the effects of differences in the scale of the features. SMOTE is an oversampling technique used in classification problems to handle unbalanced datasets in which one class is dominant. The two preprocessed datasets were then used as input for the KNN, SVM, DT, and stacking algorithms to predict disc cutter replacement.

4.4 ML classification algorithms

The classification ML algorithms KNN, SVM, DT, and the ensemble stacking technique were used in the present study. The description of the ML algorithms in this section is based on Kim *et al*. (2022).

KNN: The KNN algorithm classifies target data into the same class as the closest k neighboring data points in feature space (Fig 9(a)). To optimize this algorithm, the hyperparameters, such as the k-value and distance equations (e.g., Minkowski (Eq. (1)), Euclidean (Eq. (2)), Manhattan(Eq. (3))) need to be adjusted.

$$D_{Minkowski}(x,y) = \sqrt[p]{\sqrt{\sum_{i=1}^{n} |x_i - y_i|^p}} \tag{1}$$

$$D_{euclidean}(x,y) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \tag{2}$$

$$D_{Manhattan}(x,y) = \sum_{i=1}^{n} |x_i - y_i| \tag{3}$$

where x_i is the training features, y_i is the test features.

SVM: The optimal hyperplane in the SVM algorithm can be defined in Eq. (4) and shown in Fig. 9(b). The objective of SVM is to find the optimal hyperplane with weights and bias that maximize the margin, allowing for effective separation between different classes in the feature space.

$$f(x) = x^T \beta + \beta_0 \tag{4}$$

where x is the weight vector training features, β is the input feature vector, and β_0 is the bias.

In a non-linear problem, employing a kernel function (K) enables the introduction of additional dimensions to the original data, effectively transforming it into a linear problem within the resulting higher-dimensional space. In the SVM model predicting the disc cutter replacement, the RBF(Radial Basis Function) kernel is selected among linear, polynomial, and RBF kernels for model construction(Eq. (5)). In addition, the values for *C*, gamma, and max iterations are also adjusted to improve performance.

$$K_{RBF}(x,x') = \exp\left(-\gamma \|x - x'\|^2\right) \tag{5}$$

DT: The DT model employs tree-based classification that employs root nodes, internal nodes, and leaf nodes. Fig.9(c) illustrates a structure of DT. The DT algorithm classifies data by sequentially splitting it based on specific features, ultimately aiming for leaves that predominantly consist of a single target value. Throughout the classification process, the algorithm evaluates the probability of misclassification using impurity measures and seeks to minimize it. The hyper parameters max depth, maximum features, and minimum/maximum leaf nodes are adjusted to improve model performance.

Stacking: The stacking technique is not an individual machine learning algorithm but rather an ensemble model that predicts the output via selected individual ML algorithms and makes the final prediction based on Meta learner. The present study uses KNN, random forest, DT, and AdaBoost as the individual models. The outcomes from these individual models serve as input in a meta-learner step to derive the final target result. Liner regression is used as meta learner, and Fig. 10 illustrates the constructed staking model in the study.

Confusion matirix for binary classification		Predicted class	
		Replacement (1)	No replacement (0)
Actual class	Replacement (1)	TP (True Positive)	FN (False Negative)
	No replacement (0)	FP (False Positive)	TN (True Negative)

(a) Binary class classification

Confusion matirix for multiclass(3x3) classification		Predicted class		
		Replacement for abnormal wear (2)	Replacement for normal wear (1)	No replacement (0)
Actual class	Replacement for abnormal wear (2)	TP_2	E_{21}	E_{20}
	Replacement for normal wear (1)	E_{12}	TP_1	E_{10}
	No replacement (0)	E_{02}	E_{01}	TP_0

(b) Multi-class classification

Fig. 11 Confusion matrices for classification model (modified from Tharwat 2020)

4.5 Evaluation of classification model performance

The classification models were evaluated in the present study using a confusion matrix. Figs. 11(a) and 11(b) present the confusion matrices for the binary and multi-class classification models, respectively. In the ML training phase, the models are trained on the input data to learn important features, and the model uses these to predict the class label for test data containing the same features. The actual and predicted classes in the test output can then be compared. The confusion matrix for the binary classification matrix presented in Fig. 11(a) contains the following cases (Kohavi and Provost 1998):

- **TP** is a **True Positive**, in which a disc cutter that needs to be replaced is correctly identified as needing to be replaced.
- **FN** is a **False Negative,** in which a disc cutter that needs to be replaced is incorrectly identified as not needing to be replaced.
- **FP** is a **False Positive**, in which a disc cutter that does not need to be replaced is incorrectly identified as needing to be replaced.
- **TN** is a **True Negative**, in which a disc cutter that does not need to be replaced in reality is correctly identified as not needing to be replaced.

The accuracy, precision, recall, F1-score, and area under the curve (AUC) for the receiver operating characteristic (ROC) curve were calculated using the confusion matrices defined in Eqs. (6)-(9) and displayed in Fig. 12. Accuracy is calculated as the number of predictions that are correct from all predictions, while precision is the ratio of true positives to the total number of positives predicted. Recall is calculated as the ratio of true positives to the total number of actual positive cases in the dataset. The AUC for the ROC indicates how reliable a decision is. The more the ROC curve curves to the upper left, the higher the AUC value, which represents a better predictive model.

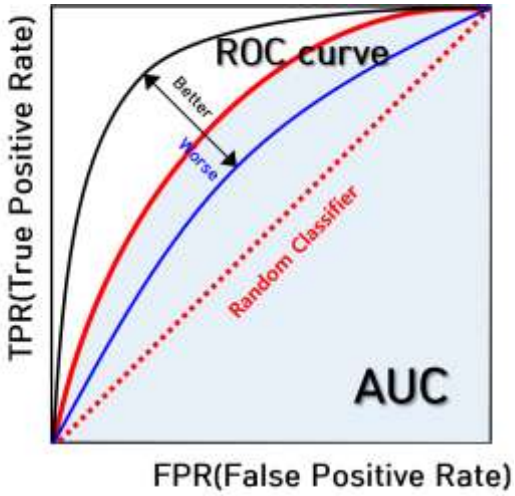


Fig. 12 ROC curve and AUC in classification algorithm performance evaluation indexes

$$\text{Accuracy (\%)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100 \tag{6}$$

$$\text{Precision (\%)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{7}$$

$$\text{Recall (\%)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{8}$$

$$\text{F1} - \text{score (\%)} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \tag{9}$$

5. Results and discussion

5.1 Performance of the binary classification models

The binary classification models predicted whether a disc cutter needed replacement or not, and three datasets (Site A, Site B, and Sites A and B together) were used to train the KNN, SVM, DT, and stacking ensemble models. Table 3 summarizes the performance for datasets A and B in terms of accuracy, precision, recall, F1 score, and AUC. The stacking ensemble model exhibited high accuracy and precision with the A and B datasets (Fig. 13), while the precision of the KNN, SVM, and DT models which are the individual machine learning algorithms were lower than their recall performance index. Precision calculated from the confusion matrix is the ratio of the number of disc cutters that were correctly predicted to need replacement (i.e., TP to the total number of replacements predicted (i.e., TP + FP). Low precision means that the model often falsely predicted replacement when the disc cutter was still usable. The reason for the lower precision performance in individual ML algorithms stems from issue of data imbalance in the test dataset. However, the stacking algorithm improved precision but had a little lower recall than the other algorithms. Recall refers to the ratio of the number of disc cutters that are correctly predicted to be

Table 3 Performance evaluation results for the binary classification models

Site	Algorithm	Confusion matrix				Accuracy	Precision	Recall	F1 score	AUC
		TP	FN	FP	TN					
A	KNN	37	2	94	2222	95.9	28.2	94.9	43.5	95.4
	SVM	32	7	365	1951	84.2	8.0	82.1	14.7	83.2
	DT	*35*	*4*	*42*	*2274*	98.1	45.5	89.7	60.3	94.0
	Stacking	*28*	*11*	*6*	*2310*	99.3	82.4	71.8	76.7	85.8
B	KNN	*107*	*10*	*140*	*10759*	98.6	43.3	91.5	58.8	95.1
	SVM	*105*	12	2878	8021	73.8	3.5	89.7	6.8	81.7
	DT	95	22	225	10674	97.8	29.7	81.2	43.5	89.6
	Stacking	97	20	9	10890	99.7	91.5	82.9	87.0	91.4

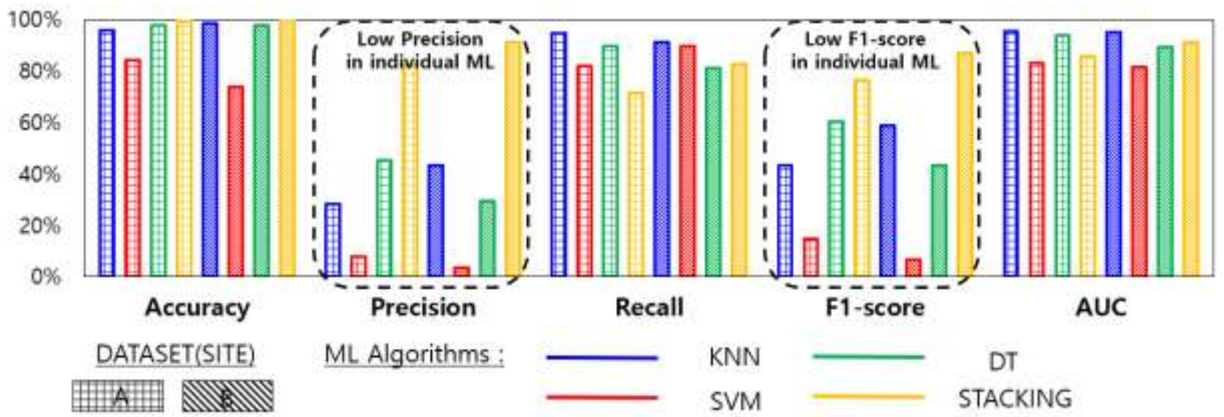


Fig. 13 Accuracy, precision, and recall results for the binary classification models (Dataset from site A and B)

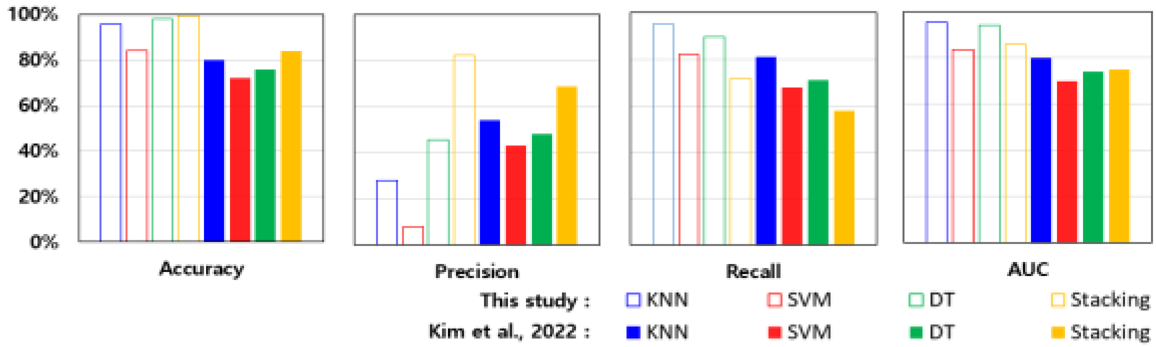


Fig. 14 Comparison performance evaluation indexes of the binary classification models from the study and Kim *et al.* (2022))

replaced (i.e., TP) to the number of disc cutters that need replacement in the data (i.e., TP + FN). In this study, it is essential to correctly predict disc cutter that need replacement, so recall is considered very important. Recall from the stacking ensemble model was 24% lower than from the KNN, which was the best-performing model for this metric. Despite this, when considering all performance indicators, the stacking algorithm was considered to outperform the other models in binary classification.

Kim *et al.* (2022) classified a disc cutter in the dataset as needing to be replaced when it exceeded 90% of the wear limit. However, the present study classified only those disc cutters that were actually replaced in the raw data as needing to be replaced. Except for precision, the performance matrices of the individual models were much improved with the classification standard used in the present study when compared to Kim *et al.* (2022) (Fig. 14). The reason for the low precision was that the number of data points for replaced disc cutters was significantly lower compared to the previous analysis, resulting in a data imbalance.

Table 4 Performance evaluation results for the multi-class classification models

Site	Models	Confusion matrix	Class	Each class			Overall		
				Accuracy	Precision	Recall	Accuracy	Precision	Recall
A	KNN	$\begin{bmatrix} 8 & 2 & 0 \\ 9 & 32 & 4 \\ 28 & 71 & 2986 \end{bmatrix}$	2	98.8	18.0	80.0	96.4	98.5	96.4
			1	97.3	30.0	71.0			
			0	96.7	100.0	97.0			
	SVM	$\begin{bmatrix} 5 & 2 & 1 \\ 9 & 5 & 17 \\ 405 & 65 & 1846 \end{bmatrix}$	2	82.3	1.2	63.0	78.8	97.5	78.8
			1	96.1	6.9	16.0			
			0	79.3	99.0	80.0			
	DT	$\begin{bmatrix} 6 & 2 & 0 \\ 3 & 24 & 4 \\ 47 & 54 & 2215 \end{bmatrix}$	2	97.8	11.0	75.0	96.7	77.0	75.0
			1	97.3	30.0	77.0			
			0	95.5	100.0	96.0			
	Stacking	$\begin{bmatrix} 8 & 0 & 0 \\ 2 & 21 & 8 \\ 0 & 9 & 2307 \end{bmatrix}$	2	99.9	80.0	100.0	99.2	99.2	99.2
			1	99.2	70.0	68.0			
			0	99.3	100.0	100.0			
B	KNN	$\begin{bmatrix} 7 & 5 & 1 \\ 2 & 88 & 14 \\ 6 & 44 & 10849 \end{bmatrix}$	2	99.9	47.0	54.0	99.4	99.5	99.4
			1	99.4	64.0	85.0			
			0	99.4	100.0	100.0			
	SVM	$\begin{bmatrix} 8 & 4 & 1 \\ 11 & 74 & 19 \\ 387 & 1410 & 9102 \end{bmatrix}$	2	96.3	2.0	62.0	83.4	98.8	83.4
			1	86.9	5.0	70			
			0	83.5	100.0	84.0			
	DT	$\begin{bmatrix} 5 & 5 & 3 \\ 3 & 85 & 16 \\ 99 & 523 & 10277 \end{bmatrix}$	2	99.0	4.7	38.0	94.1	98.9	94.1
			1	95.0	14.0	82.0			
			0	94.2	100.0	94.0			
	Stacking	$\begin{bmatrix} 7 & 5 & 1 \\ 2 & 83 & 19 \\ 0 & 11 & 10888 \end{bmatrix}$	2	99.9	78.0	54.0	99.7	99.6	99.7
			1	99.7	84.0	80.0			
			0	99.7	100.0	100.0			

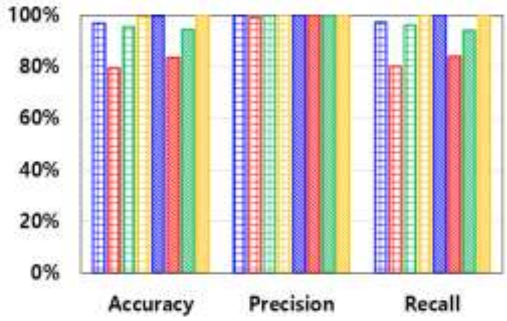
5.2 Results for multi-class classification models

The multi-class classification models predicted whether the disc cutters required replacement due to normal wear (Class 1), required replacement due to abnormal wear (Class 2), or did not need to be replaced (Class 0). The three datasets (Site A and Site B) were used to train the KNN, SVM, DT, and stacking ensemble models for multi-class classification. Table 4 summarizes the performance results for datasets A and B from confusion matrix. The accuracy, precision, and recall were evaluated for each class and for each class.

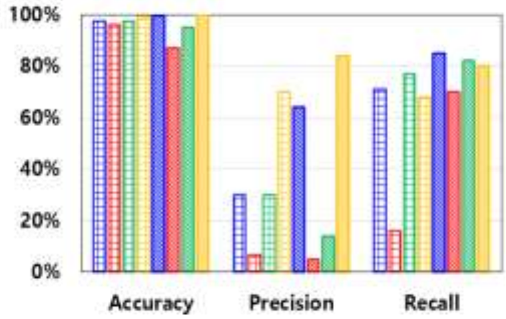
Fig 15 presents the accuracy, precision, and recall indexes for each class. Similar to the binary classification results, the stacking algorithm outperformed the other algorithms. In particular, the stacking algorithm exhibited greater precision for Classes 1 and 2, which was lower for the other models due to the data imbalance. When comparing the performance results of the model for the two sites, the accuracy for each class was higher for Site B. This is attributed to the increased variability in machine data and the more diverse disc cutter wear patterns due to the complex ground conditions at Site A, which had an impact on the learning process. The recall for class 2, which indicates the ratio of correctly predicting disc cutter replacements due to actual abnormal wear, showed higher values in the model trained on Site A, where replacements for abnormal wear were more frequent.

5.3 Evaluation of the Models with the site combined dataset

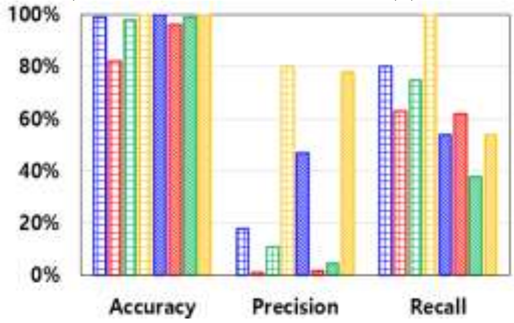
A combined dataset (A+B) was established using the data for the two sites to improve model generalization. The site combined dataset was used to analyze the binary and multi-class classification algorithms. Fig. 16 presents a comparison of the performance evaluation results for the site combined dataset and those from the optimal model (Stacking) for the Site A and B datasets individually in binary classification model. The performance evaluation indexes for the combined dataset were intermediate between the individual datasets and the same results were observed for the multi-class classification models. Fig. 17 illustrates comparison precision and recall index for each class 1 and 2 for dataset A, B, and site combined dataset in stacking algorithm. Precision and recall are inversely proportional, and the two indexes cannot increase simultaneously. In this multi class classification model, the classification of abnormal wear (class 2) needs to be well-performed for the model to be evaluated positively. Both in class 1 and 2, the precision demonstrated an increase in the combined dataset, and the recall were examined between each individual dataset. The integration of two sites dataset confirmed that the model performance could maintain a level similar to the individual dataset model performances.



(a) Class 0 (No replacement)



(b) Class 1 (Normal wear replacement)



(c) Class 2 (Abnormal wear replacement)

Fig. 15 Performance evaluation results for each class in the multi-class classification models

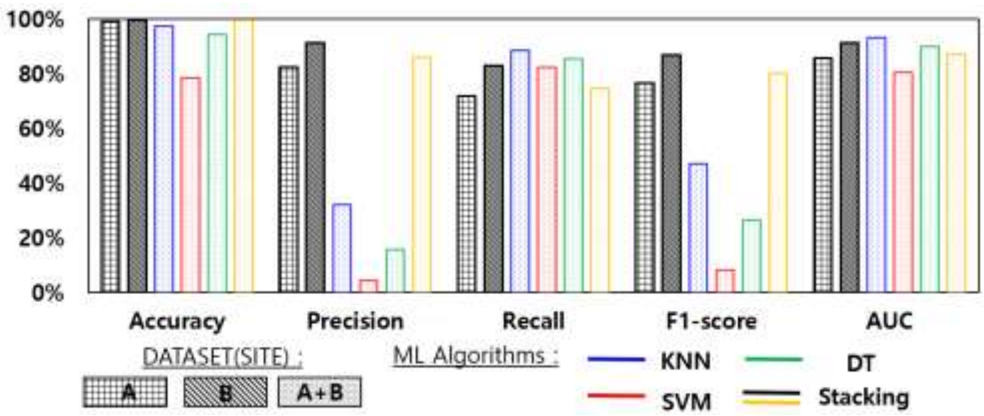


Fig. 16 Comparison of the performance evaluation index for the combined dataset(A+B) and the individual datasets for Sites A and B

6. Conclusions

In this study, a model was developed to assess the disc cutter wear condition and replacement status during tunneling. The determination of the disc cutter replacement timing during excavation is not easily discernible by operators, and replacement leads to equipment stoppage, causing downtime. Existing disc cutter lifespan prediction models can only estimate the consumption rate during the design phase and are not applicable for real-time predictions during construction. Therefore, in this research, an AI classification model was constructed using continuously generated machine data to predict disc cutter replacement, as well as classify normal and abnormal wear during tunneling, and its performance was evaluated.

The following conclusions were drawn from the analysis of disc cutter wear and the evaluation of the prediction models:

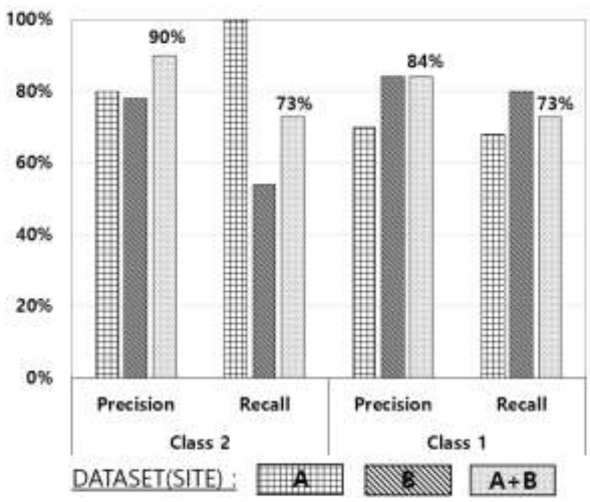


Fig. 17 Comparison of the precision and recall for each class and dataset from stacking algorithm

1. The replacement of disc cutters and wear patterns were analyzed at two slurry shield TBM sites. At Site A, 188 disc cutters were replaced; of these, 77.1% were replaced due to abnormal wear, while 22.9% were replaced due to abnormal wear. At Site B, 85.5% of 344 disc cutters were replaced due to normal wear, 24.5% due to abnormal wear, and 1.7% due to uneven wear. It was observed that more complex strata and larger gravel tended to produce more replacements due to abnormal wear.
2. It was predicted during the design stage that the disc cutter would need to be stopped for replacements three times at Site A and seven times at Site B. However, these predictions assumed that all disc cutters mounted on the cutter head would be replaced when the machine was stopped. During construction, the equipment had to be stopped at more locations, reducing the construction efficiency and increasing the number of replacement disc cutters. Due to this inaccuracy in the design stage predictions, the need for a prediction model for disc cutter replacement was evident.
3. The stacking ensemble algorithm was found to be the best-performing algorithm for both binary classification (determining whether a disc cutter should be replaced or not) and multi-class classification (identifying replacement due to normal wear, replacement due to abnormal wear, or no replacement). Although recall for the stacking ensemble model was lower than that for the other models, it exhibited higher accuracy overall. Therefore, the stacking algorithm was considered the optimal choice for the prediction of disc cutter replacement.
4. In the multi-class classification model, the precision for abnormal wear replacement (Class 2) and normal wear replacement (Class 1) were low using the individual algorithms. However, better results were observed for the other performance metrics using the stacking ensemble algorithm. The algorithm for identifying normal wear based on recall performed exceptionally well with the KNN algorithm, while the algorithm for detecting abnormal wear and deciding on replacements was deemed superior with the Stacking ensemble algorithm.
5. The data from the two sites was combined to generalize the model. The stacking algorithm exhibited better performance for the combined dataset, thus this model can be generalized by training it with combined datasets for different sites. In addition, the SMOTE technique was used for all datasets to resolve data imbalances. In order to increase both accuracy and precision, this imbalance must be resolved, which requires the use of the SMOTE technique and the collection of more data from various construction sites for training.

This study, as an initial step toward predicting of replacement time/period of disc cutters during construction, aims to develop a well-performing model through further investigation, involving data diversification and the use of high performance AI algorithms and various attempts from different perspectives. It is anticipated that these studies will contribute to reducing the construction costs of shield TBM tunnels and improving the economic efficiency of tunnel construction.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022R1F1A1065540).

References

Afradi, A., Ebrahimabadi, A. and Hallajian, T. (2021), "Prediction of the number of consumed disc cutters of tunnel boring machine using intelligent methods", Mining of Mineral Deposits.

Bruland, A. (1988), "Hard rock tunnel boring", Dissertation, Norwegian University of Science and Technology.

Butt, S. and Meschke, G. (2021). "Interaction of cutting disc with heterogeneous ground", *PAMM*, **20**(1). https://doi.org/10.1002/pamm.202000060.

Connors, R. (2017), "The challenges of tunnelling with slurry shield machines in mixed ground", The David Sugden Young Engineers Writing Award.

Ellecosta, P., Käsling, H. and Thuro, K. (2018), "Tool wear in TBM hard rock drilling–backgrounds and special phenomena", *Geomech. Tunn.*, **11**(2), 142-148. https://doi.org/10.1002/geot.201800006.

Fang, Y., Yao, Z., Xu, W., Tian, Q., He, C. and Liu, S. (2021), "The performance of TBM disc cutter in soft strata: A numerical simulation using the three-dimensional RBD-DEM coupled method", *Eng. Fail. Anal.*, **119**, 104996. https://doi.org/10.1016/j.engfailanal.2020.104996.

Farrokh, E. (2013), "Study of utilization factor and advance rate of hard rock TBMs", PhD Thesis. PSU.

Frenzel, C., Käsling, H. and Thuro, K. (2008), "Factors influencing disc cutter wear", *Geomech. Tunn.*, **1**(1), 55-60. https://doi.org/10.1002/geot.200800006.

Gehring, K. (1995), "Leistungs- und Verschleißprognose im maschinellen Tunnelbau", *Felsbau.*, **13**(6), 439-448.

Jeong, H., Zhang, N. and Jeon, S. (2018), "Review of technical issues for shield TBM tunneling in difficult grounds", *J. Korean Tunn. Undergr. Sp. Assoc.*, **28**(1), 1-24.

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

Kovari, K. and Ramoni, M. (2006), "Urban tunnelling in soft ground using TBMs", *Proceedings of the International Conference and Exhibition on Tunnelling and Trenchless Technology: Tunnelling and Trenchless technology in the 21st century*, Subang Jaya-Selangor Darul Ehsan, January.

Kohavi, R. and Provost, F. (1998). "Glossary of terms", Editorial for the Special Issue On Applications of Machine Learning and the Knowledge Discovery Process.

Kim, D., Khanh, P., Park, S., Oh, J. and Choi, H. (2020), "Determination of effective parameters on surface settlement during shield TBM", *Geomech. Eng.*, **21**(2), 153-164. https://doi.org/10.12989/gae.2020.21.2.153.

Kim, K., Oh, J., Lee, H., Kim, D. and Choi, H. (2018), "Critical face pressure and backfill pressure in shield TBM tunneling on soft ground", *Geomech. Eng*., **15**(3), 823-831. https://doi.org/10.12989/gae.2018.15.3.823.

Kim, Y., Hong, J., Shin, J. and Kim, B. (2022), "Shield TBM disc cutter replacement and wear rate prediction using machine learning techniques", *Geomech. Eng*., **29**(3), 249-258. https://doi.org/10.12989/gae.2022.29.3.249.

Lislerud, A. (1997), "Principles of mechanical excavation", Tamrock Corp. POSI VA 97-12.

Macias, F.J. (2016), "Hard rock tunnel boring: Performance predictions and cutter life assessments", Dissertation, Norwegian University of Science and Technology.

Mahmoodzadeh, A., Mohammadi, M., Ibrahim, H.H., Abdulhamid, S.N., Ali, H.F.H., Hasan, A.M. and Mahmud, H. (2021), "Machine learning forecasting models of disc cutters life of tunnel boring machine", *Automat. Constr*., **128**, 103779.

Ministry of Construction & Transportation. (2007), "Development of an optimal excavation design model for rapid tunnel mechanization construction", Construction & Transportation R&D report.

Ren, D.J., Shen, S.L., Arulrajah, A. and Cheng, W.C. (2018), "Prediction model of TBM disc cutter wear during tunnelling in heterogeneous ground", *Rock Mech. Rock Eng.*, **51**, 3599-3611. https://doi.org/10.1007/s00603-018-1549-3.

Rezaei, A.H., Shirzehhagh, M. and Golasand, M.R.B. (2019), "EPB tunneling in cohesionless soils: A study on Tabriz Metro settlements", *Geomech. Eng*., **19**(2), 153-165. https://doi.org/10.12989/gae.2019.19.2.153.

Rostami, J. (1997), "Development of a force estimation model for rock fragmentation with disc cutters through theoretical modeling and physical measurement of crushed zone pressure", Dissertation, Colorado School of Mines.

Tharwat, A. (2020), "Classification assessment methods", *Appl. Comput. Inform*., **17**(1), 168-192.

Xiaokang, S., Yusheng, J., Zongyuan, Z., Zhiyong, Y., Zhenyong, W., Jinguo, C. and Quanwei, L. (2023), "TBM disc cutter ring type adaptability and rock-breaking efficiency: Numerical modeling and case study", *Geomech. Eng*., **34**(1), 103-113. https://doi.org/10.12989/gae.2023.34.1.103.

Yang, Z., Sun, Z., Fang, K., Jiang, Y., Gao, H. and Bai, Z. (2021), "Cutting tool wear model for tunnel boring machine tunneling in heterogeneous grounds", *Bull. Eng. Geol. Environ*., **80**(7), 5709-5723. https://doi.org/10.1007/s10064-021-02298-z.

Yu, H., Tao, J., Huang, S., Qin, C., Xiao, D. and Liu, C. (2021), "A field parameters-based method for real-time wear estimation of disc cutter on TBM cutter head", *Automat. Constr*., **124**, 103603. https://doi.org/10.1016/j.autcon.2021.103603.