

ORIGINAL ARTICLE

# 이수식 TBM 데이터와 랜덤포레스트를 이용한 일축압축강도 분류 예측에 관한 연구

강태호<sup>1</sup>, 최순욱<sup>1\*</sup>, 이철호<sup>1</sup>, 장수호<sup>2</sup>

<sup>1</sup>한국건설기술연구원 지반연구본부 수석연구원, <sup>2</sup>한국건설기술연구원 지반연구본부 선임연구위원

## A Study on the Prediction of Uniaxial Compressive Strength Classification Using Slurry TBM Data and Random Forest

Tae-Ho Kang<sup>1</sup>, Soon-Wook Choi<sup>1\*</sup>, Chulho Lee<sup>1</sup>, and Soo-Ho Chang<sup>2</sup>

<sup>1</sup>Senior Researcher, Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology

<sup>2</sup>Senior Research Fellow, Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology

\*Corresponding author: [soonugi@kict.re.kr](mailto:soonugi@kict.re.kr)

Received: December 5, 2023

Revised: December 11, 2023

Accepted: December 11, 2023

### ABSTRACT

Recently, research on predicting ground classification using machine learning techniques, TBM excavation data, and ground data is increasing. In this study, a multi-classification prediction study for uniaxial compressive strength (UCS) was conducted by applying random forest model based on a decision tree among machine learning techniques widely used in various fields to machine data and ground data acquired at three slurry shield TBM sites. For the classification prediction, the training and test data were divided into 7:3, and a grid search including 5-fold cross-validation was used to select the optimal parameter. As a result of classification learning for UCS using a random forest, the accuracy of the multi-classification prediction model was found to be high at both 0.983 and 0.982 in the training set and the test set, respectively. However, due to the imbalance in data distribution between classes, the recall was evaluated low in class 4. It is judged that additional research is needed to increase the amount of measured data of UCS acquired in various sites.

**Keywords:** Slurry TBM, Machine learning, Uniaxial compressive strength, Classification model, Random forest

### 초록

최근 국내외에서 기계학습 기법으로 TBM 굴진 데이터와 지반데이터를 분석하는 지반 분류예측 연구가 증가하고 있다. 본 연구에서는 다양한 분야에서 널리 사용되고 있는 머신러닝 기법들 중 의사결정트리 기반 랜덤포레스트 모델을 3곳의 이수식 TBM 현장에서 획득한 기계 데이터와 지반 데이터에 적용하여 일축압축강도에 대한 다중 분류예측 연구를 하였다. 일축압축강도의 다중 분류 예측을 위해서 학습과 테스트 데이터를 7:3으로 분할하였으며, 최적의 파라미터를 선정을 위해서 분할 교차검증을 포함하는 그리드 서치를 활용하였다. 의사 결정 트리를 기반으로 한 랜덤 포레스트를 사용하여 일축압축강도 분류 학습을 수행한 결과, 다중 분류 예측 모델의 정확도는 학습 세트와 테스트 세트에서 각각 0.983 및 0.982로 모두 높게 나타났다. 다만, 클래스 간 데이터 분포의 불균형으로 인하여 클래스 4에서는 재현율



이 낮게 평가되었다. 다양한 현장에서 획득한 일축압축강도의 측정 데이터양을 늘리는 연구가 필요한 것으로 판단된다.

**핵심어:** 이수식 TBM, 기계학습, 일축압축강도, 분류모델, 랜덤포레스트

## 1. 서론

국내외 터널공사에서 TBM (Tunnel Boring Machine) 공법은 안정성과 소음, 진동 등의 환경적인 부분에서 타 공법 대비 효과적이어서 다양한 지반 조건에서 널리 사용되고 있다. 터널이 굴착되는 지반의 지질 조건은 전반적인 터널 공사비와 굴진 속도를 결정하는 데 중요한 역할을 한다. 터널 설계 시 수행되는 지반조사는 제한적인 조건에서 이루어진 결과로 실제 굴착 중에는 암반 상태가 다르거나 예측하지 못한 암반 이상지대가 출현하여 TBM의 굴진 성능이 저하되고 굴착 현장의 안정성 문제가 발생할 수 있다. TBM 굴진 작업은 굴착면을 이루는 암반의 역학적 특성과 지질학적 조건, TBM의 운전조건 등에 복합적으로 영향을 받기 때문에 굴착 중인 암반조건뿐만 아니라 향후 암반상태의 변화를 예측하는 것이 필요하다.

Mokhtari and Mooney(2020)는 지도학습 기반의 기계학습 기법과 토압식 TBM 굴진 데이터를 활용하여 굴진속도 예측 모델을 개발하였다. 토압식 TBM의 굴진 속도에 영향을 미치는 주요 특성이 추력(thrust), 커터 헤드 토크(torque), 폼 유량(foam flow) 및 스크류 컨베이어 토크 등이라고 하였으며, 굴진 대상 지반 조건들이 이러한 특성들에 영향을 많이 주고 있고, 그에 따라 TBM 운영도 달라진다고 언급하였다.

Kang et al.(2020, 2021, 2022)은 기계학습기법들을 활용하여 이수식 TBM 현장에서 획득한 굴진데이터와 RQD, 일축압축강도, 탄성파속도에 대하여 암반 특성 분류 예측을 수행하였고 토압식 TBM 현장에서 획득한 토사지반의 200체 통과율과 표준관입시험의 N값 기준으로 지반 분류 예측, 토압식 TBM의 기계데이터를 활용한 굴진성능 예측을 하였다. 그 결과로부터, 단일현장 기반의 분류 예측 모델에서 발생하는 자료 부족에 기인한 과적합 문제를 언급하였고 예측성능 일반화를 해결하기 위해서는 좀 더 다양한 데이터가 필요함을 제시하였다.

Ayawah et al.(2022)는 암반특성 분류를 위한 다양한 AI 및 기계학습 모델을 활용한 연구에 대하여 검토 및 사례 연구를 수행하였으며, 암반 특성 분류예측에서 중요한 영향인자로 총추력, 커터토크, 커터헤드 회전속도, 굴진률, 개별 에너지, 디스크 수직력 등을 제시하였다. Yang et al.(2022)는 토사 지반의 표준관입시험 N값과 암석의 RMR (Rock Mass Rating)을 활용한 4가지 분류항목으로 구분하고 분류학습 모델인 랜덤포레스트, 그래디언트 부스팅, 서포트벡터머신을 활용하여 굴진데이터를 학습하여 지반 타입 분류 예측하는 연구를 하였다. 그 결과, 지반 타입을 예측하는데 총 추력, 관입속도, 추진력과 토크의 비율 등이 가장 큰 영향을 미친다고 하였다.

본 연구에서는 이수식 TBM 현장 3곳에서 획득한 기계데이터와 지반정보를 대상으로 암석의 공학적 분류기법과 다양한 분야에서 사용되고 있는 의사결정트리 기반의 랜덤포레스트 기법을 활용하여 일축압축강도에 대해 다중분류 예측하는 연구를 수행하였다.

## 2. 일축압축강도 분류와 분류학습 알고리즘

### 2.1 일축압축강도 분류

흙을 분류하는 목적은 비슷한 성질을 보이는 흙을 물리적, 역학적으로 유사한 거동을 보이는 그룹으로 구분하여 공학적인 성질

을 파악하기 위함이다. 마찬가지로 암석의 분류는 암석, 암반을 유사한 거동을 보이는 그룹으로 구분하여 특성을 파악하는 기준을 마련하고자 함이다. 암석 분류는 목적과 특성에 따라 지질학적 분류, 공학적 분류 등으로 구분되며, 현재 주요 암반분류방법은 암반 하중(Terzaghi, 1946), RQD (Rock Quality Designation, Deere et al., 1967), 암석강도(Deere and Miller, 1966), RSR (Rock Structure Rating, Wickham et al., 1972), RMR (Rock Mass Rating, Bieniawski, 1989), Q-system (Barton et al., 1974), 탄성파 속도(Anon, 1979), 지반공학적 분류(ISRM, 1981) 등이 사용되고 있다. 암반분류는 암반의 중요한 특성을 쉽고 명확한 용어를 사용하여 간편하고 신속하게 분류가 가능하여야 한다. 암반 분류의 단계는 기반암을 풍화암, 연암, 보통암, 경암으로 구분하여 암종에 대한 분류와 암반 상태에 대해 암질, 색조, 풍화상태, 강도, 분연속면 간격, 절리면 거칠기 등을 구분하는 암질에 대한 분류로 나뉘어져 있다. 이러한 암반을 분류 하는 기준은 지질학적 특성에 따른 분류, 공학적 특성에 의한 분류 등에 따라 다소 다를 수 있다.

본 연구에서는 암석의 특성에 대한 분류에서 대표적 공학적 특성 기준인 일축압축강도(Table 1)를 기계학습의 분류 기준 항목으로 선정하였다.

**Table 1.** Classification of rocks based on UCS (ISRM, 1979)

Classification	Uniaxial compressive strength (UCS), MPa
Extra weak	0.25 – 1
Very weak	1 – 5
Weak	5 – 25
Medium Strong	25 – 50
Strong	50 – 100
Very strong	100 – 250
Extra strong	> 250

## 2.2 의사결정트리 기반의 랜덤포레스트 분류 모델

본 연구에서는 목표변수(target)를 레이블(label)하고 훈련 데이터(training dataset)을 이용해 학습을 진행하는 지도학습(supervised learning)기반 알고리즘을 사용하였다. 또한 이수식 쉘드 TBM 현장의 암석 특성에 대한 분류 예측을 목표로 하고 있으므로 분류 기법을 사용하였다. 분류 기법에는 많은 알고리즘이 있지만, 다중 분류가 가능하고 기본적인 알고리즘인 의사결정트리 방법과 알고리즘 결합을 통해 분류 결과를 향상시키고 과적합을 방지할 수 있는 앙상블 모델 중 배깅(Bagging, Bootstrap Aggregating) 모델인 랜덤포레스트를 사용하여 학습을 수행하였다.

의사결정트리는 주어진 데이터를 분류하고 규칙을 찾는 방법으로 가지 분리하여 변수를 구분하고 영역을 만들어 클래스 값을 구한다(Fig. 1). 의사결정트리 기법은 연속형, 범주형 변수에 모두 사용 가능하고 차원축소나 변수 선택 등 의사결정 생성과정을 통해 많은 변수 중에서 상대적으로 종속변수에 주는 영향을 파악 할 수 있다(Breiman et al., 1984). 트리기법은 그룹별로 특징을 발견하거나 어느 집단에 속하는지 여부를 파악하여 세분화 하는데 효율적인 기법이다. 변수의 정규화나 표준화 같은 전처리가 필요하지 않으며, 특정 변수의 값이 누락되어도 사용 가능하나, 학습데이터에 과대 적합되는 특징 때문에 새로운 데이터에 적용하면 예측 성능이 좋지 않아 일반화가 어렵다는 단점이 있다.

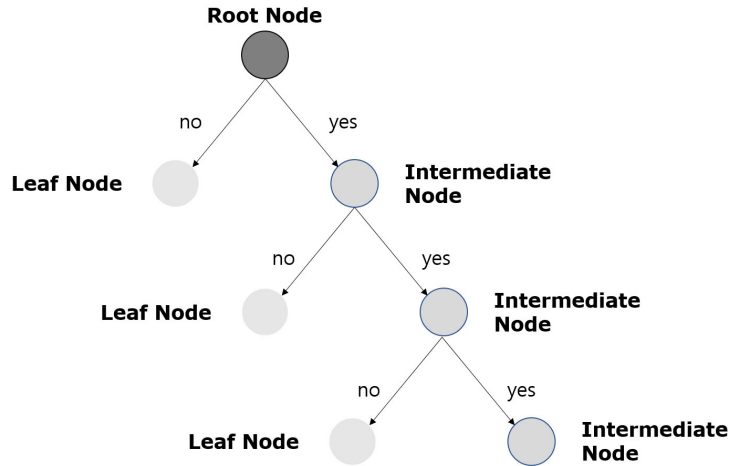


Fig. 1. Diagram of decision tree algorithms

이러한 단점들을 해결하고 더 좋은 예측 성능을 위해서 개발된 알고리즘이 앙상블 모델(Ensemble Learning Method)이다. 앙상블 모델은 의사결정나무 기반으로 서로 다른 분류 알고리즘 모델을 생성하고 결합하여 새로운 규칙을 만들어 최적의 결과로 분류하는 방향으로 진행되는 방법으로 대표적인 앙상블 기법은 부스팅(Boosting)과 배깅 모델이 있다. Kearns and Valiant(1994)가 처음으로 제안한 부스팅 기법은 단일 모델을 활용하여 순차적으로 학습하고 발생하는 오차를 보완하기 위해 다음 분류기에 가중치를 부여하고 다시 학습을 진행하는 방식이다. 이와 다르게, Breiman(1996)에 의해 제안된 배깅은 무작위로 중복 샘플(bootstrap samples)을 생성하여 병렬적으로 사용하면서 다양하게 학습하는 모델이고, 대표적인 방법으로는 랜덤 포레스트(Randomforest)가 있다(Fig. 2). 모델을 병렬처리 하면서 피처를 일부만 랜덤하게 사용하기 때문에 가지치기의 자원이 감소하고 다양성을 확보함에 따라 과적합을 방지하여 정확도가 향상되는 장점이 있다.

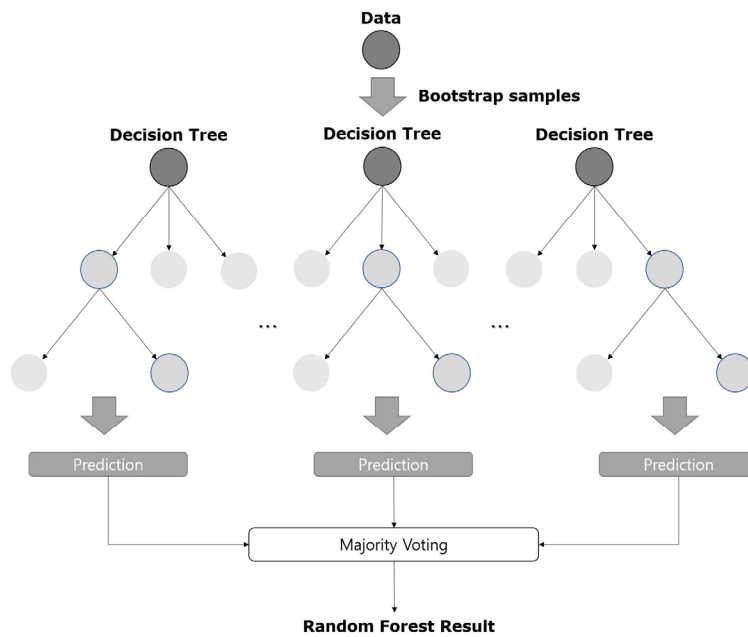


Fig. 2. Diagram of Random forest algorithms

### 3. 일축압축강도 분류기준 및 데이터세트 구성

#### 3.1 현장 개요 및 지반 정보

학습 대상인 이수식 TBM 터널 3개 현장의 장비(S-01 ~ 03)로부터 굴진자료를 수집하고 터널 설계 시 작성된 지반조사보고서와 터널 종단면도를 수집하여 분석하였다.

설계 시 시행된 지반조사에서 TBM 터널 구간에는 총 72개의 시추가 이루어졌으며, 시추 시료에 대해 수행된 암석 물성 시험 결과를 활용하여 분석하였다. TBM S-01 장비가 통과하는 터널 구간의 암반 구성은 연암파쇄대가 12%, 연암이 23%, 경암이 65%이고, RQD는 25 이하가 24%, 25~50가 30%, 50 이상이 46%로 나타났으며, 암석시험 결과 일축압축강도는 25.95~116 MPa로 대체로 양호한 암질이 분포하고 있었다. S-02 장비가 통과하는 지반의 상부에는 실트질 모래 내지는 실트질 점토의 붕적층이 발달되어 있으며, 터널 통과 구간의 기반암 구성이 연암이 95%, 경암 5%이고, RQD가 25%이하의 구간이 57.9%, 25% 이상인 구간이 42.1%로 나타났으며, 일축압축강도는 7.9 ~ 101.9 MPa로 다소 연약한 암질이 포함되었다. 또한 터널 통과 구간에 풍화가 진행된 연약대도 존재한다. S-03 장비의 굴진 지반에는 상부에 얇은 점토와 사질토로 구성된 퇴적층 및 풍화가 진행된 풍화토가 존재하며, 터널 통과구간의 경우에는 연암 약 64.3%, 경암 약 35.7%로 구성되며, 현장에서 채취한 암석의 일축압축강도가 9.08 ~ 67.13 MPa 범위로 풍화가 진행되어 불량한 암반구간도 다소 확인되었다(Fig. 3).

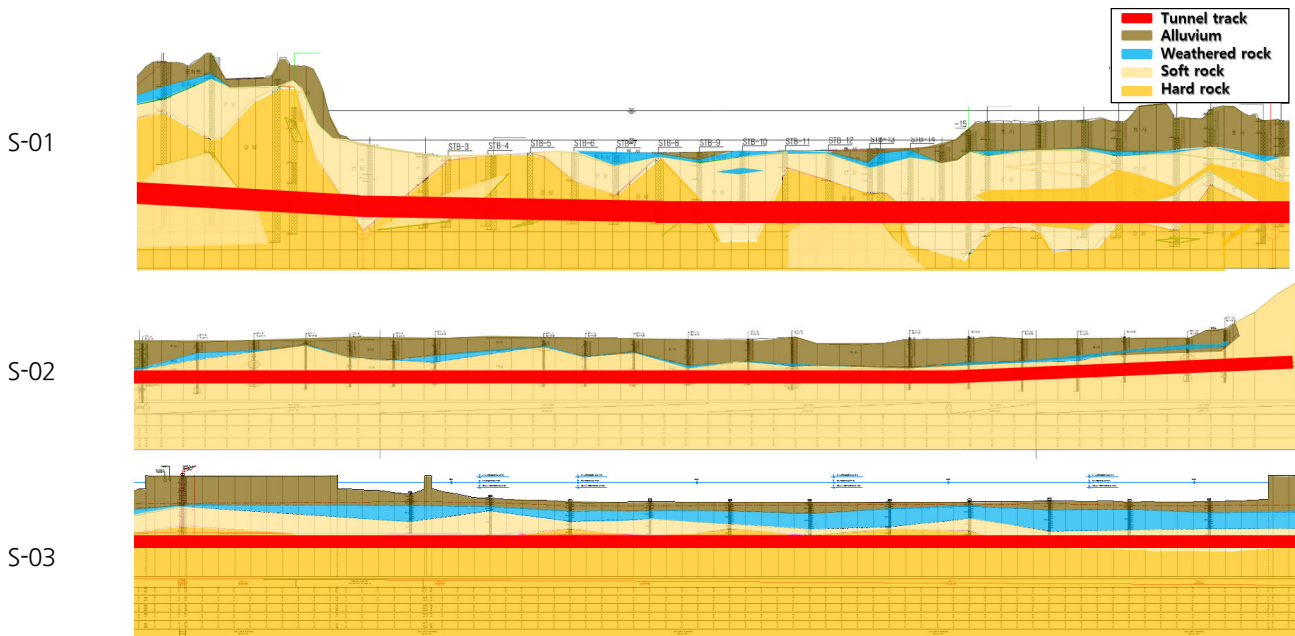


Fig. 3. Geological profile along tunnel track

#### 3.2 일축압축강도 분류기준

총 3대의 이수식 쉴드 TBM 장비에서 획득한 굴진 데이터를 분석하여 암반 특성 분류 학습을 위한 일축압축강도 분류 기준을 구성하였다. 앞서 2.1 절에서 언급한 바와 같이 기준항목인 일축압축강도에 대한 분류기준 Table 1과 굴진 데이터 분석결과를 분류 학

습에 필요한 클래스 선정에 활용하였다. 일축압축강도가 약함(Weak, 25 MPa이하)의 경우를 암반 상태가 안 좋은 클래스 1로 하고, 보통(Medium Strong, 25-50 MPa), 강함(Strong, 50-100 MPa), 매우 강함(Very Strong, 100 MPa이상)의 순서로 클래스 2, 3, 4로 하여 일축압축강도에 대한 분류 클래스를 설정하였다(Table 2).

**Table 2.** Criteria for classification of Uniaxial compressive strength (UCS)

Class range	Uniaxial compressive strength (MPa)	Number of data	Percentage of data (%)
1	UCS < 25	5936	31.7
2	25 < UCS ≤ 50	4060	21.7
3	50 < UCS ≤ 100	7686	41.0
4	UCS ≥ 100	1056	5.6

### 3.3 데이터세트 구성

암석의 일축압축강도 분류 예측 연구의 데이터 구성에 활용된 정보는 터널 대상 구간의 지반정보, 노선정보, TBM 기계 데이터 등이 있다. 본 연구에서 사용한 기계 데이터를 얻은 이수식 쉴드 TBM 장비 3대의 주요 제원은 Table 3과 같다. 장비의 직경은 3.5~7.3 m이고, 최대추력은 9.6~56 MN, 최대토크는 5.4~53.3 kN·m, 최대 RPM은 3.6~4.6이다.

**Table 3.** Summary of Slurry TBM specification

Slurry TBM shield machine	S-01	S-02	S-03
Outside diameter (mm)	7,300	3,590	3,500
Inside diameter (mm)	6,600	3,530	3,436
Max. shield jack thrust force (MN)	56 (2 × 28 shield jack)	12 (1.0 × 12 shield jack)	12 (1.0 × 12 shield jack)
Max. cutterhead torque (MN·m)	5.33	1.41	0.54
Max. RPM (r/min)	3.6	4.6	4.6
Cutter (EA)	45	38	22
Segment ring length (m)	1.5	1.2	1.2

굴진 데이터 분석 결과, 사용된 장비 3대(S-01~03)의 평균 추력이 21.67, 4.74, 5.02 MN으로 장비 최대 성능의 약 38.7, 39.5, 46.4%이고, 커터헤드 평균 토크는 1.25, 0.26, 0.1 MN·m로 장비제원 최대 토크의 약 17.8, 18.7, 23.45 % 범위에서 주로 운용되었으며, 평균 토압은 188.29 ~303.75 kPa로 안정적으로 유지되었다. 그리고 이수식 TBM 장비 3대의 평균 굴진 속도는 약 13.02~17.33 mm/min이고, 커터헤드의 평균 회전속도는 2.68~3.17 r/min이다. 암석 분류 예측 데이터의 분석에 사용된 주요 데이터의 통계학적 특성과 데이터 항목 별 분포 특성은 Fig. 4와 Table 4에서 확인할 수 있다.

TBM 굴진 현장에서는 압력 관련 변수, 속도 관련 변수, 온도 관련 변수, 위치/방향 관련 변수, 굴진시간이력 관련 변수 등 광범위한 기계 데이터 수천 개가 생성될 수 있으며, 학습 대상 현장에서는 약 1000여개가 기록되었다. 따라서 TBM의 굴진 데이터를 기반으로 정확하고 효과적인 분류 예측을 위해서는 목표에 맞는 특성변수 항목의 선택과 선택된 데이터의 처리가 무엇보다 중요하다.

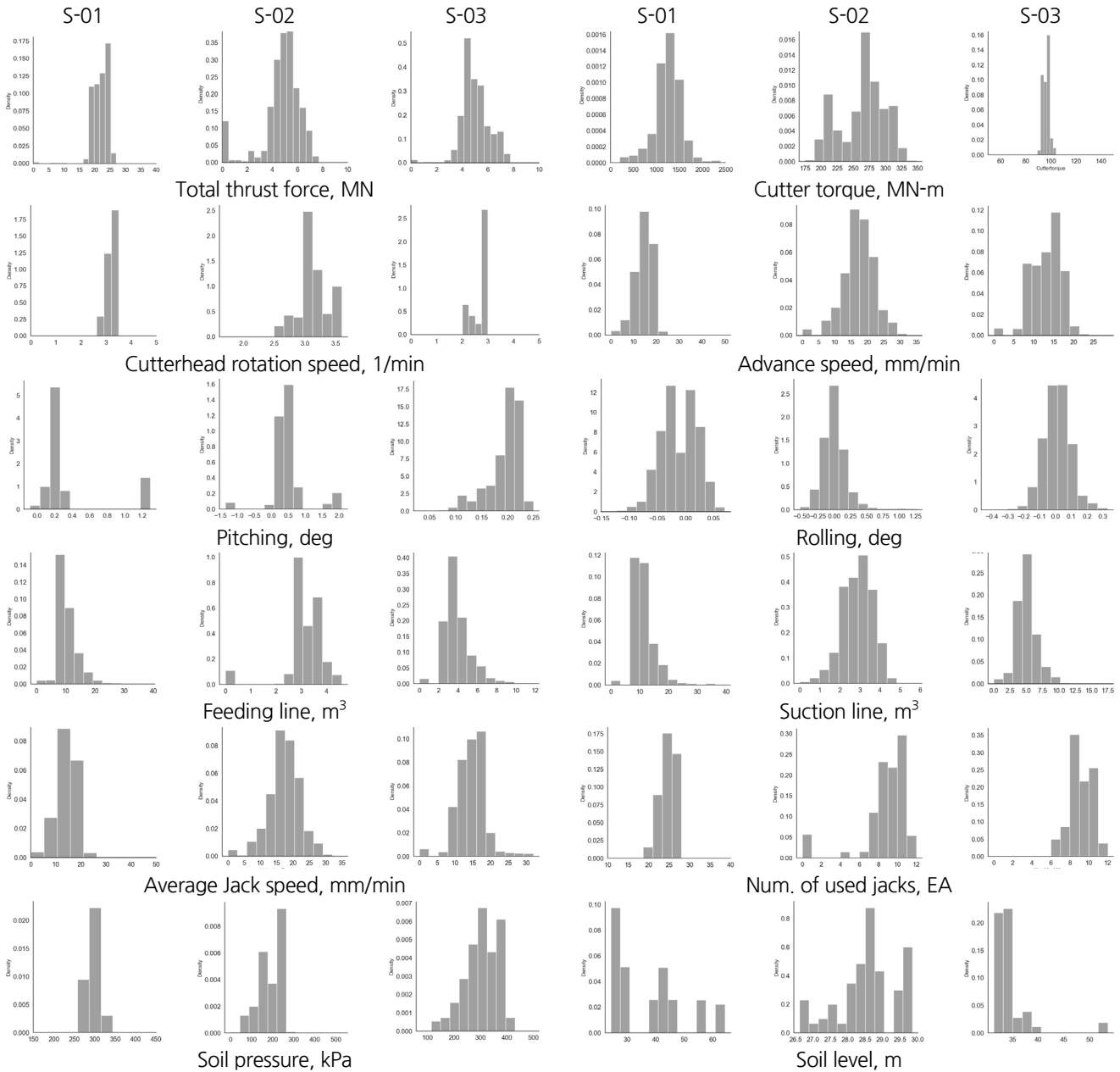


Fig. 4. Distribution of data according to ground conditions for the TBM

Table 4. Statistical description of the database

Factors	Min	Q1 (25%)	Q2 (50%)	Q3 (75%)	Max	Average	Standard deviation	
Total thrust force (MN)	S-01	0.00	19.95	22.05	23.75	26.96	21.67	2.91
	S-02	0.00	4.26	5.01	5.65	7.70	4.74	1.60
	S-03	0.00	4.31	4.81	5.64	7.76	5.02	1.09
Cutter torque, kN-m	S-01	0.00	1100.00	1262.00	1421.00	2388.00	1245.01	294.62
	S-02	0.00	236.70	270.00	288.70	348.10	264.22	34.90
	S-03	0.00	94.00	96.00	98.00	124.00	96.20	3.34

**Table 4.** Statistical description of the database (continued)

Factors		Min	Q1 (25%)	Q2 (50%)	Q3 (75%)	Max	Average	Standard deviation
Cutterhead rotation speed, 1/min	S-01	0.00	3.02	3.21	3.28	3.50	3.17	0.23
	S-02	0.00	3.10	3.10	3.20	3.60	3.14	0.23
	S-03	0.00	2.30	2.80	3.00	3.00	2.68	0.36
Advance speed, mm/min	S-01	0.00	11.76	14.29	16.67	50.00	14.06	3.89
	S-02	0.00	15.00	17.14	20.73	35.00	17.33	4.84
	S-03	0.00	10.53	13.33	15.38	28.57	13.02	3.59
Pitching, deg	S-01	-0.08	0.18	0.22	0.27	1.29	0.37	0.39
	S-02	-1.36	0.32	0.40	0.51	2.08	0.48	0.50
	S-03	0.03	0.18	0.21	0.22	0.25	0.20	0.03
Rolling, deg	S-01	-0.14	-0.03	-0.01	0.01	0.07	-0.01	0.03
	S-02	-0.53	-0.11	-0.01	0.09	1.26	0.01	0.20
	S-03	-0.43	-0.04	0.01	0.06	0.33	0.01	0.09
Feeding line, m3	S-01	0.00	8.12	9.50	11.81	38.50	10.48	4.02
	S-02	0.00	2.90	3.10	3.60	4.60	3.12	0.80
	S-03	0.00	3.02	3.57	4.48	11.81	3.89	1.40
Suction line, m3	S-01	0.00	9.12	10.47	12.78	39.52	11.50	4.13
	S-02	0.00	2.40	2.90	3.40	5.80	2.86	0.75
	S-03	0.00	4.16	4.87	5.81	17.48	5.06	1.66
Average Jack speed, mm/min	S-01	0.00	12.00	14.00	17.00	63.00	14.04	3.87
	S-02	0.00	15.00	19.00	25.00	90.00	20.33	7.40
	S-03	0.00	12.00	15.00	16.00	32.00	14.36	4.10
Num. of used jacks, EA	S-01	0.00	24.00	25.00	26.00	28.00	24.91	2.20
	S-02	0.00	8.00	9.00	10.00	12.00	8.33	2.43
	S-03	0.00	8.00	9.00	10.00	12.00	8.69	1.30
Soil pressure, kPa	S-01	0.00	285.00	295.00	302.00	344.00	295.76	14.10
	S-02	0.00	156.91	205.94	235.36	529.56	188.29	53.80
	S-03	0.00	264.78	304.01	353.04	500.14	303.75	64.55
Soil level, m	S-01	24.25	26.79	37.71	41.32	64.22	36.97	12.36
	S-02	26.61	28.09	28.56	29.01	29.85	28.58	0.86
	S-03	31.36	32.42	33.52	34.57	53.54	34.49	4.04

본 연구에서는 굴진 데이터에서 암석의 분류 예측에 기여하지 않는 다양한 전압 측정값, 계측 온도, 위치 및 자세 정보와 같은 특성변수를 분석에서 제외하였다. 또한, 통계적인 분석을 통하여 각 특성 변수의 전체 데이터를 25% 단위로 나눌 때에 Q1(1사분위수), Q2(2사분위수), Q3(3사분위수)로 나타낼 수 있으며, 그 결과로 데이터 분포 특성과 이상치를 판단하여 최종적으로 암반 분류 예측에 사용될 특성변수를 선정하였다. 선정된 특성변수는 주요 기계데이터인 총추력(total thrust force), 커터토크(cutter torque), 커터헤드 회전속도(cutterhead rotation speed), 굴진속도(advance speed), 피칭(pitching), 롤링(rolling), 송니유량(feeding line discharge), 배니유량(suction line discharge), 평균 잭속도(average Jack speed), 사용 잭수(number of used jacks), 토압(soil pressure), 토피고(cover depth)이다.



## 4. 분석 결과

### 4.1 모델 성능평가와 하이퍼파라미터 결정

본 연구에서는 일축압축강도 분류 예측을 위한 지도학습 모델을 구축하고자 앞서 2.2절에서 설명한 의사결정트리 기반으로 다중 분류가 가능하고 과적합 방지에 유리한 랜덤포레스트 기법을 선정하였다. 자료 분석과 기계학습에는 파이썬 3.8(python 3.8) 프로그램과 오픈 소스 라이브러리를 사용하였으며, 사용한 구체적인 분석과정과 학습 절차는 Fig. 5와 같다. 기계학습 알고리즘의 적합성을 평가하기 위해서 학습 데이터(training data)와 테스트 데이터(test data)를 7:3 비율로 분할하였다. 학습된 기계학습 분류 모델의 예측성능을 평가하는 방법으로는 Table 5와 같이 분류 모델의 정답과 오답을 보여주는 혼돈행렬(confusion matrix)을 이용한 분류 모델 평가 지표인 정확도(accuracy), 정밀도(precision), 재현율(recall), F1-스코어(score)를 사용하였다(Table 6).

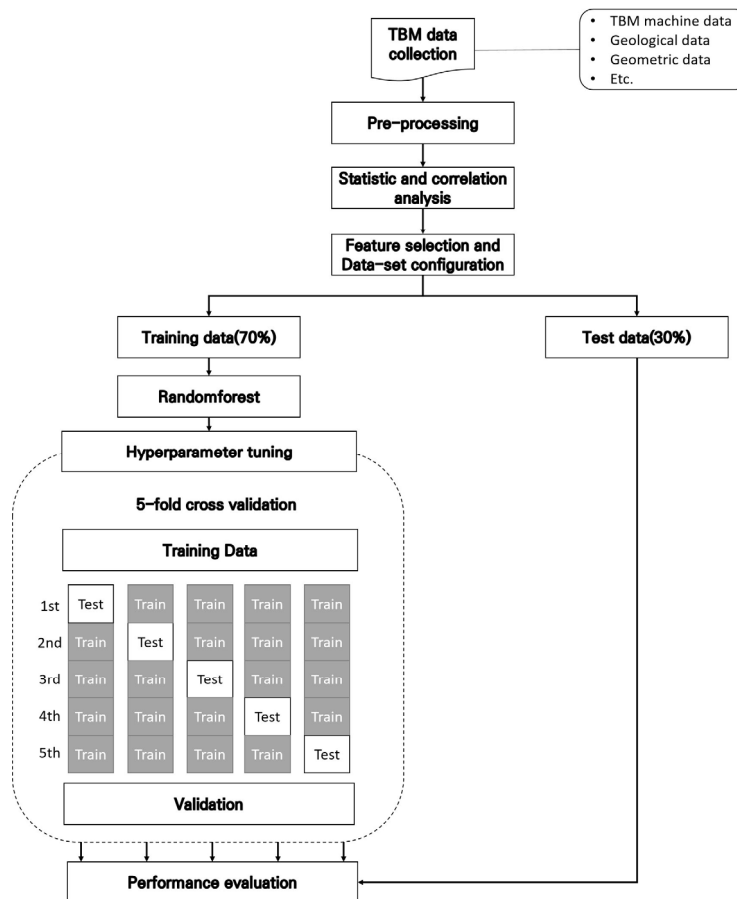


Fig. 5. Flowchart of Machine Learning model procedure for classification prediction

Table 5. Confusion matrix of classification

Class	Positive(1)	Negative(0)
Positive(1)	True Positive	False Positive
Negative(0)	False Negative	True Negative

**Table 6.** Classification Performance Measures

	Formulas for measuring performance	Definition of the Terms
Accuracy	$\frac{TP+ TN}{TP+ TN+ FP+ FN}$	Accuracy is the proportion of the total number of predictions that are correct
Precision	$\frac{TP}{TP+ FP}$	Precision is ratio of total number of correctly classified positive examples and the total number of predicted positive examples
Recall	$\frac{TP}{TP+ FN}$	Recall is also referred as True Positive Rate or Sensitivity
F1-score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	F1 score is a weighted average of the recall (sensitivity) and precision

데이터 전처리(Preprocessing)과정으로 3대의 장비에서 획득한 데이터 각 값에서 최솟값을 뺀 것을 최댓값에서 최솟값으로 뺀 값으로 나누어 주는 데이터 스케일링(Min-Max Scaling)을 적용하고, 데이터세트에서 TBM 장비 제원과 굴진 데이터를 확인하여 결측치와 이상치를 제거하였다. 3대의 장비에서 획득된 각각의 데이터를 전처리한 후에 통합하여 일축압축강도를 분류 예측을 위한 전체 학습 데이터를 구성하였다. 하이퍼파라미터(hyperparameter)는 머신러닝 모델을 만들 때 직접 조정해주는 값으로 최적의 훈련 모델을 구현하기 위해 결정하는 변수이다. 데이터 훈련 과정에서 알고리즘 모델에 적합한 파라미터 조합을 찾기 위한 하이퍼 파라미터 튜닝(hyperparameter tuning)을 할 수 있으며, 이는 머신러닝 모델링에서 성능에 영향을 줄 수 있는 하이퍼파라미터에 대해서 최적에 가까운 값을 찾기 위한 조정 작업이라 할 수 있다.

본 연구에서는 일축압축강도 분류 예측의 학습 결과가 일부 데이터에 과적합(overfitting)되는 것을 방지하기 위해 분할 교차 검증(k-fold cross validation)을 적용하였다. 분할 교차 검증은 앞서 7:3로 분할된 학습 데이터(training data)를 다시 k 묶음으로 나누어 학습 데이터와 테스트 데이터를 k개 구성하여 k번 학습을 반복하고 k번의 학습 결과에 대해 평균을 내어 편향되지 않는 최종 성능을 구하는 방법이다(Fig. 5). 본 연구에서는 5겹 교차 검증(5-fold cross validation)의 성능비교를 통해 가장 좋은 성능을 내는 하이퍼파라미터 조합을 찾는 그리드서치(GridSearchCV) 튜닝 방법을 사용하여 일축압축강도 분류 예측 학습모델에 대해 적절한 파라미터를 결정하였다(Table 7).

**Table 7.** Summary of Hyperparameter for Classification models

Model	Hyperparameter
Ensemble Learning Method, Bagging, Bootstrap Aggregating	Randomforest Max_depth: 3, max features: 5, n_estimators: 100

### 4.2 학습 결과

4개 클래스의 일축압축강도 분류 기준과 3대의 이수식 쉴드 TBM 굴진 데이터, 그리고 지반정보를 기반으로 랜덤포레스트 기법을 이용하여 학습을 수행하였다. 혼돈행렬은 실제 값과 예측 값의 일치 여부를 시각적으로 보여주며, 이는 일축압축강도 분류 예측 모델의 성능을 평가하는 기본적인 판단 기준이 된다. 랜덤포레스트를 활용한 예측 모델의 혼돈행렬이 학습 데이터 세트와 테스트 데이터세트 모두에서 실제 값과 예측 값이 전반적으로 잘 일치하는 것을 확인하였다(Fig. 6). 일반적으로 모델의 적합성을 나타내는 정확도는 1.0에 가까울수록 분류 성능이 우수하고 결과의 정확도가 높은 것을 나타내지만, 클래스에 대한 예측결과의 불균형이

발생할 수 있으므로 정밀도(precision), 재현율(recall), f-스코어를 종합적으로 보고 모델성능을 판단한다(Table 6). 일축압축강도에 대한 분류 예측 모델의 학습 결과는 Table 8과 같다.

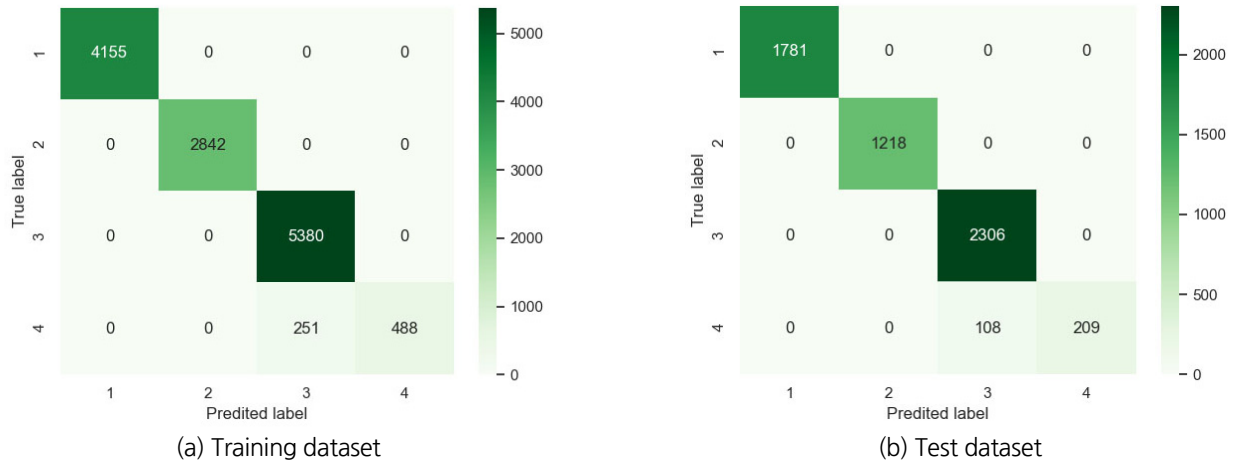


Fig. 6. Confusion matrix of classification prediction for Shield TBM Datasets

Table 8. Evaluating the performance of the classification for Shield TBM Datasets

Randomforest Model	Training dataset				Test dataset			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Class 1	0.983	1.00	1.00	1.00	0.982	1.00	1.00	1.00
Class 2		1.00	1.00	1.00		1.00	1.00	1.00
Class 3		0.96	1.00	0.98		0.96	1.00	0.98
Class 4		1.00	0.68	0.81		1.00	0.69	0.82

일축압축강도(UCS) 항목의 분류 기준에 대한 랜덤포레스트 기법의 정확도는 학습 세트와 테스트 세트에서 0.983, 0.982로 모두 분류 성능이 우수하였고, 정밀도가 0.96~1.0, 재현율이 0.68~1.0, f-스코어가 0.81~1.0로 학습세트와 테스트 세트 모두에서 높게 나타났다. 랜덤포레스트를 적용한 일축압축강도에 대한 다중 분류 예측에서 전반적인 성능 평가 지표가 우수한 성능을 나타냈으나, 클래스 4의 경우 재현율이 학습 세트에서는 0.68, 테스트 세트에서는 0.69로 다른 분류 클래스(class 1 ~ 3) 보다 상대적으로 낮게 나타났다. 그 이유는 타겟 변수인 일축압축강도가 범주별로 불균형한 분포로 나타났기 때문이다.

### 4.3 일축압축강도 분류 예측

앞서 4.2절에서는 이수식 TBM 장비 3대에 대한 지반조사결과와 굴진 데이터를 활용하여 일축압축강도 분류 예측을 수행하였고, 학습 결과로 랜덤포레스트 기반의 일축압축강도 분류 예측 모델의 성능을 확보하였다. Fig. 7는 학습 과정에서 별도로 분리해 놓은 평가 세트를 대상으로 실제 일축압축강도의 분류등급과 기계학습 모델의 학습 결과를 활용한 예측 값을 비교한 결과이다. 다중 분류 항목에서 클래스 1 ~ 3은 전반적으로 우수한 예측 성능을 보여주고 있다. 다만, 클래스 4에 대한 예측 확률이 0.4 ~ 0.6이고, 예측 성능도 차이가 있는 것으로 보인다. 이는 4.2절에서 클래스 4에 대한 재현율이 상대적으로 낮게 평가된 결과로 상대적으로 예측 성능의 불확실성이 높은 것으로 나타났다.

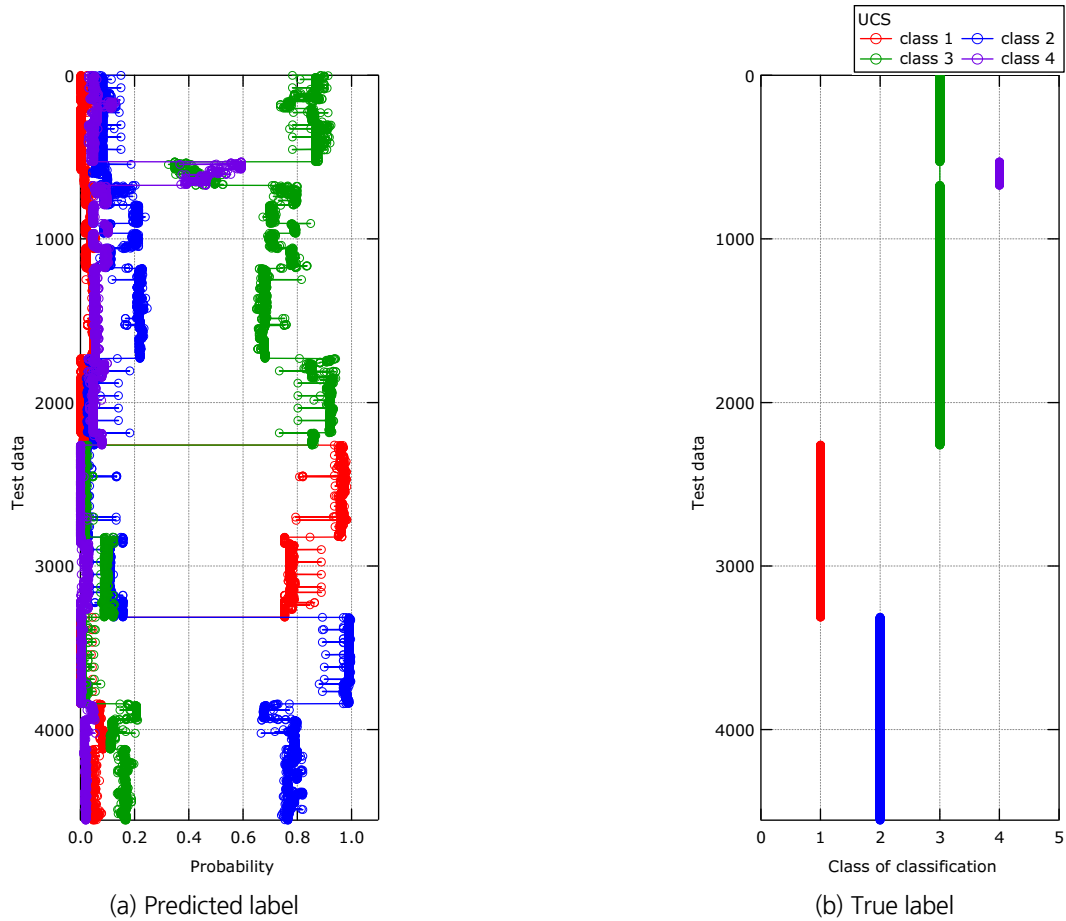


Fig. 7. Results of Predicted UCS classification using test data.

## 5. 결론

본 연구에서는 3대의 이수식 TBM의 지반정보와 굴진정보를 기반으로 기계학습을 사용하여 일축압축강도에 대한 다중 분류 예측을 하였다. 3개 현장에서 확보한 일축압축강도 분류 데이터를 대상으로 의사결정나무 기반의 랜덤포레스트 알고리즘을 적용한 다중 분류 학습을 통해 모델에 적합한 하이퍼파라미터를 선정하였으며, 학습 결과를 토대로 최적 분류 모델을 도출하는 연구를 수행하였다.

특성변수의 분포특성과 변수 간 상관도 분석을 통해 굴진정보인 총추력, 커터토크, 커터헤드 회전속도, 굴진속도, 피칭, 롤링, 송니유량, 배니유량, 평균 잭속도, 사용 잭수, 토압, 토피고를 최종 변수로 선정하고 하저구간을 통가하는 3대의 쉴드 TBM 현장의 지반조사에서 측정된 일축압축강도 데이터 각각 12개, 11개, 13개와 매칭되는 굴진데이터 약 18,000개로 데이터 세트를 구성하였다. 그리고, 학습 모델의 검증을 위해 학습 데이터와 테스트 데이터를 7:3 비율로 나눠서 분류 모델학습 성능평가를 하였다. 분할 교차 검증(5-fold cross validation) 기반의 그리드서치(GridSearchCV) 튜닝 방법으로 성능비교를 통해 분류 학습 모델이 가장 좋은 성능을 내는 하이퍼파라미터 조합을 결정하였으며, 랜덤포레스트 분류 모델의 정확도가 0.96 이상의 좋은 성능을 보여주는 것을 확인하였다.

랜덤 포레스트를 이용한 일축압축강도의 다중 분류 항목 중에서 클래스 4의 재현률이 상대적으로 낮고 그 결과 예측 결과에서도 성능이 낮아지는 것을 확인하였다. 이는 타겟 변수인 일축압축강도가 범주별로 불균형한 분포로 나타났기 때문이다. 실제 터널 시공 현장은 설계단계에서부터 굴착하기 좋은 암반상태를 우선적으로 검토하여 노선을 선정하기 때문에 악한 지반이나 아주 강한 암반의 발생빈도가 낮아진다. 따라서 관련 데이터를 충분히 확보하기 어렵기 때문에 일축압축강도 분류 클래스별 분포가 불균형을 이루게 되는 원인이 된다.

결론적으로 추후에 다양한 조건의 학습데이터를 충분히 확보하고, 데이터 샘플링 기법(언더, 오버 샘플링 등)에 대한 검토를 하여 타겟 변수 분포의 불균형 문제를 조절할 필요가 있다. 본 연구의 결과는 3개 현장의 이수식 TBM장비에서 획득한 데이터를 기반으로 도출된 것으로 일반화 성능을 검증하기에는 부족함이 있으므로, 추후 다양한 장비와 현장의 데이터를 추가한 연구 개발이 필요하다.

## 감사의 글

본 연구는 국토교통부 국토교통과학기술진흥원이 시행하고 한국도로공사가 총괄하는 “스마트건설기술개발 국가R&D사업(과제번호: 23SMIP-A158708-04)”의 지원으로 수행되었습니다.

## REFERENCES

- Anon, O.H., 1979, Classification of rocks and soils for engineering geological mapping, Part I: Rock and soil materials, Bulletin of the International Association of Engineering Geology, 19, 364-371.
- Ayawah, P.E., Sebbeh-Newton, S., Azure, J.W., Kaba, A.G., Anani, A., Bansah, S., and Zabidi, H., 2022, A review and case study of Artificial intelligence and Machine learning methods used for ground condition prediction ahead of tunnel boring Machines, Tunnelling and Underground Space Technology, 125, 104497.
- Barton, N., Lien, R., and Lunde, J., 1974, Engineering classification of rock masses for the design of tunnel support, Rock Mech., 6, 189-236.
- Bieniawski, Z.T., 1989, Engineering rock mass classifications. John Wiley & Sons.
- Breiman, L., 1996, Bagging predictors, Machine Learning, 24, 123-140.
- Breiman, L., Friedman, J., Stone, C.J., and Olshen, R.A., 1984, Classification and Regression Trees. CRC press.
- Deere, D.U. and Miller, R.P., 1966, Engineering classification and index properties for intact rock, Tech. Rep. No. AFWL-TR-65-115, AirForce Weapons Lab., Kirtland Air Base, New Mexico.
- Deere, D.U., Hendron, A.J., Patton, F.D., and Cording, E.J., 1967, Design of surface and near surface construction in rock, 8th U.S. Symposium on Rock Mechanics: Failure and breakage of rock: New York, Society of Mining Engineers, American Institute of Mining, Metallurgical, and Petroleum Engineers.
- ISRM, 1979, Suggested methods for determining the uniaxial compressive strength and deformability of rock materials, International Journal of Rock Mechanics and Mining. Sciences and Geomechanics Abstracts, 16, 135-140.
- ISRM, 1981, Suggested method for laboratory determination of direct shear strength, Rock characterisation testing and monitoring(Brown ET ed.), International Society for Rock Mechanics, Pergamon: Oxford, UK.

- Kang, T.H., Choi, S.W., Lee, C., and Chang, S.H., 2020, A Study on Prediction of EPB shield TBM Advance Rate using Machine Learning Technique and TBM Construction Information, *Tunnel and Underground Space*, 30(6), 540-550.
- Kang, T.H., Choi, S.W., Lee, C., and Chang, S.H., 2021, A Study on the Prediction of Rock Classification Using Shield TBM Data and Machine Learning Classification Algorithms, *Tunnel and Underground Space*, 31(6), 494-507.
- Kang, T.H., Choi, S.W., Lee, C., and Chang, S.H., 2022, Soil Classification by Machine Learning Using a Tunnel Boring Machine's Operating Parameters, *Applied Sciences*, 12(22), 11480.
- Kearns, M. and Valiant, L.G., 1994, Cryptographic limitations on learning Boolean formulae and finite automata, *Journal of the Association for Computing Machinery*, 41(1), 67-95.
- Mokhtari, S. and Mooney, M.A., 2020, Predicting EPBM advance rate performance using support vector regression modeling, *Tunn. Undergr. Space Technol.*, 104, 103520.
- Terzaghi, K., 1946, Rock defects and load on tunnel supports, *Introduction to rock tunnelling with steel supports*, (R.V. Proctor & T.L. White eds.) Commercial Sheering & Stamping Co: Youngstown, USA.
- Wickham, G.E., Tiedemann, H.R., and Skinner, E.H., 1972, Support determination based on geologic predictions, In Lane, K.S., Garfield, L.A. (eds.). *Proc. 1st North American Rapid Excavation & Tunnelling Conference (RETC)*, Chicago. 1. American Institute of Mining, Metallurgical and Petroleum Engineers (AIME), New York, 43-64.
- Yang, H., Song, K., and Zhou, J., 2022, Automated recognition model of geomechanical information based on operational data of tunneling boring machines, *Rock Mechanics and Rock Engineering*, 1-18.