

선형회귀분석과 머신러닝을 이용한 암석의 강도 및 암석학적 특징 기반 세르샤 마모지수 추정

홍주표¹ · 강윤성² · 고태영^{3*}

¹학생회원, 강원대학교 에너지·인프라 융합학과 석사과정

²비회원, 강원대학교 에너지·인프라 융합학과 석사과정

³정회원, 강원대학교 에너지자원·산업공학부 조교수

Estimation of Cerchar abrasivity index based on rock strength and petrological characteristics using linear regression and machine learning

Ju-Pyo Hong¹ · Yun Seong Kang² · Tae Young Ko^{3*}

¹Graduate Student, Dept. of Integrated Energy and Infra System, Kangwon National University

²Graduate Student, Dept. of Integrated Energy and Infra System, Kangwon National University

³Assistant Professor, Dept. of Energy and Resources Engineering, Kangwon National University

*Corresponding Author : Tae Young Ko, tyko@kangwon.ac.kr

OPEN ACCESS

Journal of Korean Tunnelling and
Underground Space Association
26(1)39-58(2024)
<https://doi.org/10.9711/KTAJ.2024.26.1.039>

eISSN: 2287-4747

pISSN: 2233-8292

Received December 27, 2023

Revised January 12, 2024

Accepted January 19, 2024



This is an Open Access article distributed under the terms of the Creative Commons Attribution

Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2024, Korean Tunnelling and Underground Space Association

Abstract

Tunnel Boring Machines (TBM) use multiple disc cutters to excavate tunnels through rock. These cutters wear out due to continuous contact and friction with the rock, leading to decreased cutting efficiency and reduced excavation performance. The rock's abrasivity significantly affects cutter wear, with highly abrasive rocks causing more wear and reducing the cutter's lifespan. The Cerchar Abrasivity Index (CAI) is a key indicator for assessing rock abrasivity, essential for predicting disc cutter life and performance. This study aims to develop a new method for effectively estimating CAI using rock strength, petrological characteristics, linear regression, and machine learning. A database including CAI, uniaxial compressive strength, Brazilian tensile strength, and equivalent quartz content was created, with additional derived variables. Variables for multiple linear regression were selected considering statistical significance and multicollinearity, while machine learning model inputs were chosen based on variable importance. Among the machine learning prediction models, the Gradient Boosting model showed the highest predictive performance. Finally, the predictive performance of the multiple linear regression analysis and the Gradient Boosting model

derived in this study were compared with the CAI prediction models of previous studies to validate the results of this research.

Keywords: Disc cutter wear, Cerchar abrasivity index, TBM, Regression analysis, Machine learning

초 록

TBM (Tunnel boring machine)은 터널 굴착 과정에서 여러 디스크 커터를 이용하여 암석을 절삭한다. 디스크 커터는 암석과의 지속적인 접촉과 마찰로 인해 마모된다. 디스크 커터의 표면이 마모되면 절삭 능력이 감소하고 굴착 효율이 떨어진다. 암석의 마모성은 디스크 커터 마모에 큰 영향을 미친다. 높은 마모도를 가진 암석은 커터에 더 큰 마모를 일으키며, 이는 디스크 커터의 수명을 단축시킨다. 세르샤마모지수(Cerchar abrasivity index, CAI)는 암석의 마모성을 평가하는데 널리 사용되는 지표로 CAI는 암석의 마모특성을 나타내며, 디스크 커터의 수명과 성능 예측에 필수적인 요소로 인식되고 있다. 본 연구의 목적은 암석의 강도, 암석학적 특성과 선형회귀, 머신러닝 기법을 이용하여 CAI를 효과적으로 추정하는 새로운 방법을 개발하는 것이다. 문헌 조사를 통해 CAI, 일축압축강도, 압연인장강도, 등가석영함량이 포함된 데이터베이스를 구축하고 파생변수를 추가하였다. 통계적 유의성과 다중공선성을 고려하여 다중선형회귀분석을 위한 입력변수를 선정하였고, 머신러닝 모델의 입력변수는 변수중요도 분석을 통해 선정하였다. 머신러닝 예측모델 중 Gradient Boosting 모델의 예측 성능이 가장 높게 나타나 최적의 CAI 예측 모델로 선정되었다. 마지막으로 본 연구에서 도출한 다중선형회귀분석과 Gradient Boosting 모델의 예측 성능을 선행연구들의 CAI 예측모델과 비교하여 연구 결과의 타당성을 확인하였다.

주요어: 디스크 커터 마모, 세르샤 마모지수, TBM, 회귀분석, 머신러닝

1. 서론

TBM (Tunnel boring machine)은 터널 굴착 과정에서 여러 디스크 커터를 이용하여 암석을 절삭한다. 이 디스크 커터들은 회전하면서 암석의 표면에 압력을 가하고, 암석을 깨뜨려 터널을 굴착한다. 디스크 커터는 암석과의 지속적인 접촉과 마찰로 인해 마모된다. 디스크 커터의 표면이 마모되면 절삭 능력이 감소하고 굴착 효율이 떨어진다. 암석의 마모성은 디스크 커터 마모에 큰 영향을 미친다. 높은 마모도를 가진 암석은 커터에 더 큰 마모를 일으키며, 이는 디스크 커터의 수명을 단축시킨다. 세르샤 마모지수(Cerchar abrasivity index, CAI)는 암석의 마모성을 평가하는데 널리 사용되는 지표로 CAI는 암석의 마모특성을 나타내며, 디스크 커터의 수명과 성능 예측에 필수적인 요소로 인식되고 있다.

CAI를 추정하기 위해 다양한 연구가 진행되어 왔다. 전통적인 방법은 직접적인 실험을 통하여 CAI값을 측정하는 것이다. 그러나 이 방법은 시간과 비용이 들고, 모든 암석 샘플에 적용하기 어려운 단점이 있다. 최근에는 머신러닝과 같은 데이터 과학 기술이 도입되어, 암석의 물성을 바탕으로 CAI를 추정하는 연구가 활발히 이루어지고 있다. 이러한 접근법은 실험적 방법의 한계를 극복하고, 더 효율적이고 간단하게 CAI 추정을 가능하게 한다.

본 연구의 목적은 암석의 강도와 암석학적 특성을 이용하여 CAI를 효과적으로 추정하는 새로운 방법을 개발하는 것이다. 이를 위해 암석의 일축압축강도(uniaxial compressive strength, UCS)와 압열인장강도(Brazilian tensile strength, BTS)를 사용하고 이 두 가지 강도를 통해 파생되는 여러 취성도(brittleness), 암석학적 특성으로 등가석영함량(equivalent quartz content, EQC)을 고려하였다. 이 변수들을 활용하여 선형회귀분석과 트리기반 머신러닝 기법을 통해 CAI를 추정한다. 선형회귀는 변수 간의 기본적인 관계를 탐색하는 데 유용하며, 트리기반의 기법은 복잡한 비선형 관계와 상호작용을 포착하는 데 강점을 가진다. 끝으로 선행연구를 통해 수집된 CAI의 예측식과 본 연구에서 개발된 모델을 동일한 데이터 셋에 적용하여 결과를 비교 분석하여 개발된 모델의 우수성을 평가하고자 한다.

2. 선행연구 검토

Capik and Yilmaz (2017)는 Cankurtaran와 Salmankas 터널 시공 데이터를 기반으로 하여 CAI와 일축압축강도, 등가 석영 함량, 점하중 강도, 압열인장강도, L형 슈미트 강도(Schmidt hardness L type) 및 N형 슈미트 강도(Schmidt hardness N type)와 선형회귀분석을 통해 상관관계를 분석하였고 양의 상관관계를 확인하였다. 다중회귀분석을 실시한 결과 일축압축강도와 등가 석영 함량이 입력변수일 때 결정계수 R^2 가 0.76으로 가장 높게 도출되었으며 식 (1)로 나타난다.

$$CAI = 0.243 + 0.16 UCS + 0.007 EQC \quad (1)$$

Ozdogan et al. (2018)은 CAI와 일축압축강도, 공극률(Porosity, PR), 쇼어 경도(Shore hardness index, SH)의 상관관계를 도출하기 위해서 30개의 데이터를 활용하여 단일 및 다중회귀분석을 수행하였다. CAI가 1을 초과할 때와 미만일 때를 구분하여 상관관계를 분석하였고 CAI가 1을 초과할 때 더 높은 상관관계를 보였다. 다중회귀분석에서 CAI가 1을 초과할 때 R^2 가 0.946으로 가장 높게 나오며 식 (2)로 표현된다.

$$CAI = -1.803 + 0.07303 SH - 0.790 PR - 0.00834 UCS \quad (2)$$

Altindag et al. (2009)은 CAI와 암석 물성 간의 상관관계를 구축하기 위해 일축압축강도, 압열인장강도 및 취성도를 활용하여 단일회귀분석을 수행하였다. 취성도(brittleness)는 암석의 파쇄성이나 연성의 정도를 측정하는 지표로서 일반적으로 인정되는 국제적인 표준이 없어 연구자마다 다양한 방식으로 정의되어 왔다(Meng et al., 2021). 취성도는 일축압축강도와 압열인장강도의 함수로 계산되며 Hucka and Das (1974), Özfirat et al. (2016)에 의해 식 (3)~식 (5)와 같이 제안되었다.

$$B_1 = \frac{\sigma_c}{\sigma_t} \tag{3}$$

$$B_2 = \frac{\sigma_c - \sigma_t}{\sigma_t + \sigma_t} \tag{4}$$

$$B_3 = \frac{\sigma_c + \sigma_t}{2} \tag{5}$$

CAI와 암석 물성 간의 회귀분석 결과는 다음과 같은 식 (6)~식 (11)로 도출되었고, CAI와 일축압축강도 및 B_3 모두 양의 상관관계를 가지는 것으로 분석되었다.

$$UCS = 31.43CAI + 25.84 \tag{6}$$

$$BTS = 5.887CAI^{0.241} \tag{7}$$

$$CAI = 1.786\ln(B_1) - 2.285 \tag{8}$$

$$CAI = 12.07B_2 - 7.954 \tag{9}$$

$$CAI = 0.104B_3 + 0.257 \tag{10}$$

Lee et al. (2012)은 CAI에 영향을 미치는 인자들을 분석하기 위하여 CAI와 국내 19종 암석의 일축압축강도, 압열인장강도, 탄성계수, 포아송 비(Poisson's ratio), 공극률, 쇼어 경도, 등가 석영 함량, 석영 함량과 회귀분석을 통해 상관관계를 분석하였다. 그 결과, 쇼어 경도, 등가 석영 함량과 CAI의 양의 상관관계를 확인하였고, 등가 석영 함량과 일축압축강도를 입력변수로 선형, 비선형 다중회귀분석을 실시한 결과 각각 R^2 가 0.75, 0.80으로 나타났다. 회귀식은 식 (11), 식 (12)로 나타난다.

$$CAI = 0.03257EQC + 0.00524UCS + 0.07186 \tag{11}$$

$$CAI = 0.25739 \times EQC^{0.49209} \times UCS^{0.0741} \tag{12}$$

Sun et al. (2019)은 27개의 중국의 도수터널 데이터를 바탕으로 소형 디스크 커터 마모 시험장치를 설계 및 제작하였고, CAI와 암석 마모 지수(Rock Abrasivity Index, RAI)의 회귀분석을 활용한 상관관계 분석으로 식 (13)을 도출하였다.

$$CAI = 0.996 \ln(RAI) - 0.656 \quad (13)$$

Jeong et al. (2019)은 국내 암석 유형을 대상으로 취성도와 CAI와의 상관관계를 조사하였다. 취성도($B_1 \sim B_3$)는 식 (3)~식 (5)로 계산되며, 등가 석영 함량과 취성도($B_1 \sim B_3$)는 암석 전체 데이터 및 암석 유형별로 다중회귀분석을 수행하였고 식 (14)~식 (16)을 도출하였다.

$$CAI = 0.41 \times B_1^{0.07} \times EQC^{0.42} \quad (R^2 = 0.78) \quad (14)$$

$$CAI = 0.51 \times B_2^{0.25} \times EQC^{0.42} \quad (R^2 = 0.77) \quad (15)$$

$$CAI = 0.37 \times B_3^{0.01} \times EQC^{0.42} \quad (R^2 = 0.77) \quad (16)$$

Tripathy et al. (2015)은 CAI를 예측하기 위하여 CAI, 점착력, 탄성계수(Elastic Modulus, E), 전단 계수(Shear modulus, G), 체적 탄성계수(Bulk modulus, K) P파 속도(P wave velocity, V_p), 이중 전단강도(Double shear strength), 단일 전단강도(Single shear strength), 경사 전단강도(Inclined shear strength), 압열인장강도, 일축압축강도 및 점하중 강도를 포함하는 데이터셋을 구축하였다. 변수 간 상관관계를 분석을 통해 다중회귀분석의 입력변수로 일축압축강도, P파 속도 및 탄성계수를 선정하였고, 다중회귀분석 결과, R^2 는 0.63으로 나타났고, 식 (17)을 도출하였다. 인공신경망 기법은 훈련 함수 경사 하강법, 학습 함수는 LEARNGDM을 적용하였고 입력변수는 일축압축강도, 점하중 강도, P파 속도 및 탄성계수로 105개의 데이터를 학습 데이터 85개, 검증 데이터 20개로 나누어 적용하였다. 모델의 R^2 는 0.97로 도출되었다.

$$CAI = -0.05 + 0.03 UCS - 8 \cdot 10^{-4} V_p + 0.08 E \quad (17)$$

Kwak and Ko (2022)는 CAI와 암석의 특성을 상관관계를 조사하기 위하여 암종, 일축압축강도, 압열인장강도, 석영 함량, 등가 석영 함량 등을 선형회귀분석, 다중 선형, 비선형회귀분석과 서포트 벡터 머신, 최근접 이웃 알고리즘, 의사 결정 트리, 랜덤 포레스트 및 인공신경망 기법을 이용하여 CAI를 예측하였고 모델 간 성능을 비교하였다. 그 결과, 랜덤 포레스트 모델의 성능이 MSE는 0.06, MAPE는 7.93, R^2 는 0.96으로 가장 예측 성능이

높은 모델로 분석되었다.

Kadkhodaei and Ghasemi (2019)는 Gene Expression Programming (GEP) 기법을 활용하여 RAI와 압열인장 강도로 CAI를 예측하는 연구를 수행하였다. 문헌 조사를 통해 수집된 106개의 데이터 중 90개 세트는 훈련 데이터로, 나머지 16개 세트는 테스트 데이터로 활용하였고 GEP 모델과 다중 선형회귀분석 및 다중 비선형 회귀분석의 R^2 는 각각 0.895, 0.567, 0.785로 분석되었다. 다중 선형회귀분석과 다중 비선형회귀분석의 회귀식은 식 (18), 식 (19)와 같다.

$$CAI = 1.030 + 0.027RAI + 0.034UCS \quad (18)$$

$$CAI = 0.623RAI^{0.333} \times UCS^{0.004} \quad (19)$$

Ko et al. (2016)은 CAI와 석영 함량, 일축압축강도, 압열인장강도 및 취성도를 단일 및 다중회귀분석을 통해 상관관계를 조사하였다. 문헌 조사를 통해서 데이터베이스를 구축하였고, 취성도($B_1 \sim B_3$)는 식 (3)~식 (5)로 계산하여 데이터베이스에 변수로 추가하였다. 화성암과 변성암을 구분하여 단일 및 다중회귀분석을 진행하였고, 다중 회귀 분석을 수행한 결과 화성암에서는 식 (20)을, 변성암에서는 식 (21)을 도출하였다.

$$CAI = 5.6738 - 0.0257QC + 0.0495UCS - 0.1299B_1 - 0.285B_3 \quad (R^2 = 0.70) \quad (20)$$

$$CAI = 1.636 - 0.0009QC + 0.0167UCS + 0.1958BTS - 0.089B_3 \quad (R^2 = 0.73) \quad (21)$$

3. 데이터 수집, 분석 및 입력변수선택

3.1 데이터 수집

본 연구에 사용된 데이터는 문헌 조사를 통해 수집되었으며(Alber, 2008; Rostami et al., 2014; Menschik, 2015; Abu Bakar et al., 2016; He et al., 2016; Ündül and Er, 2017; Hamzaban et al., 2018; Comakli, 2019; Jeong et al., 2019; Majeed and Abu Bakar, 2019; Torrijo et al., 2019; Majeed et al., 2020), CAI, UCS, BTS, EQC가 포함되어 있다. 또한, 기존 변수들의 함수로 표현되는 취성도 $B_1 \sim B_5$, 점착력(Cohesion, C_o), 암석마모지수(Rock Abrasivity Index, RAI)를 파생변수로 데이터베이스에 추가하였다.

취성도($B_1 \sim B_3$)는 식 (3)~식 (5), B_4 와 B_5 는 식 (22), 식 (23)과 같다(Altindag, 2003; 2010).

$$B_4 = \frac{\sigma_c \sigma_t}{2} \quad (22)$$

$$B_5 = \sqrt{\frac{\sigma_c \sigma_t}{2}} \quad (23)$$

여기서, σ_c 는 일축압축강도, σ_t 는 압열인장강도를 나타낸다.

BI는 본 연구에서 모델의 적합한 학습 성능을 위해 본 연구에서 정의한 취성도로 종속변수인 CAI와 독립변수인 일축압축강도와 압열인장강도에 활용하여 도출하였다.

$$y = a \cdot \sigma_c \cdot \sigma_t \quad (24)$$

여기서, a 는 상수항이다. 식 (24)를 가정하고 양변에 로그를 취한 후 선형회귀분석을 수행하여 식 (25), 식 (26)을 유도하였다.

$$\log(y) = A \log(\sigma_c) + B \log(\sigma_t) + a \quad (25)$$

$$y = \sigma_c^A \cdot \sigma_t^B \cdot e^a \quad (26)$$

여기서, A와 B는 각각의 회귀계수이다. 회귀분석 후 식 (26)에 지수 변환을 적용하여 기존의 스케일로 변환하였다. 이러한 과정을 통해 비선형관계를 표현할 수 있는 변수를 생성하였다. 식 (26)의 상수항은 제외하고 식 (27)과 같이 BI를 도출하였다.

$$BI = \sigma_c^{0.22} \sigma_t^{0.09} \quad (27)$$

점착력은 Piratheepan et al. (2012)이 제안하고, Moon and Yang (2020)이 정리한 일축압축강도와 압열인장강도로 추정하는 식 (28)을 통해 계산하여 파생변수로 추가하였다.

$$C_o = \frac{0.5 \sigma_c \sigma_t}{\sqrt{\sigma_t (\sigma_c - 3 \sigma_t)}} \quad (28)$$

마지막으로, RAI를 식 (29)로부터 계산하여 데이터베이스에 추가하였다(Plinninger, 2010).

$$RAI = EQC \times UCS \quad (29)$$

3.2 탐색적 데이터 분석

CAI 데이터베이스의 데이터는 302개이고 암종은 화성암, 변성암, 퇴적암 세 가지로 구분하여 각각 158개, 60개, 84개로 암종별 차이를 확인하기 위하여 암종별로 예측모델을 구현하였다. CAI 데이터베이스에 대한 히스토그램을 Fig. 1에 나타내었으며, CAI와 EQC를 제외한 변수들이 비정규분포의 양상을 보인다.

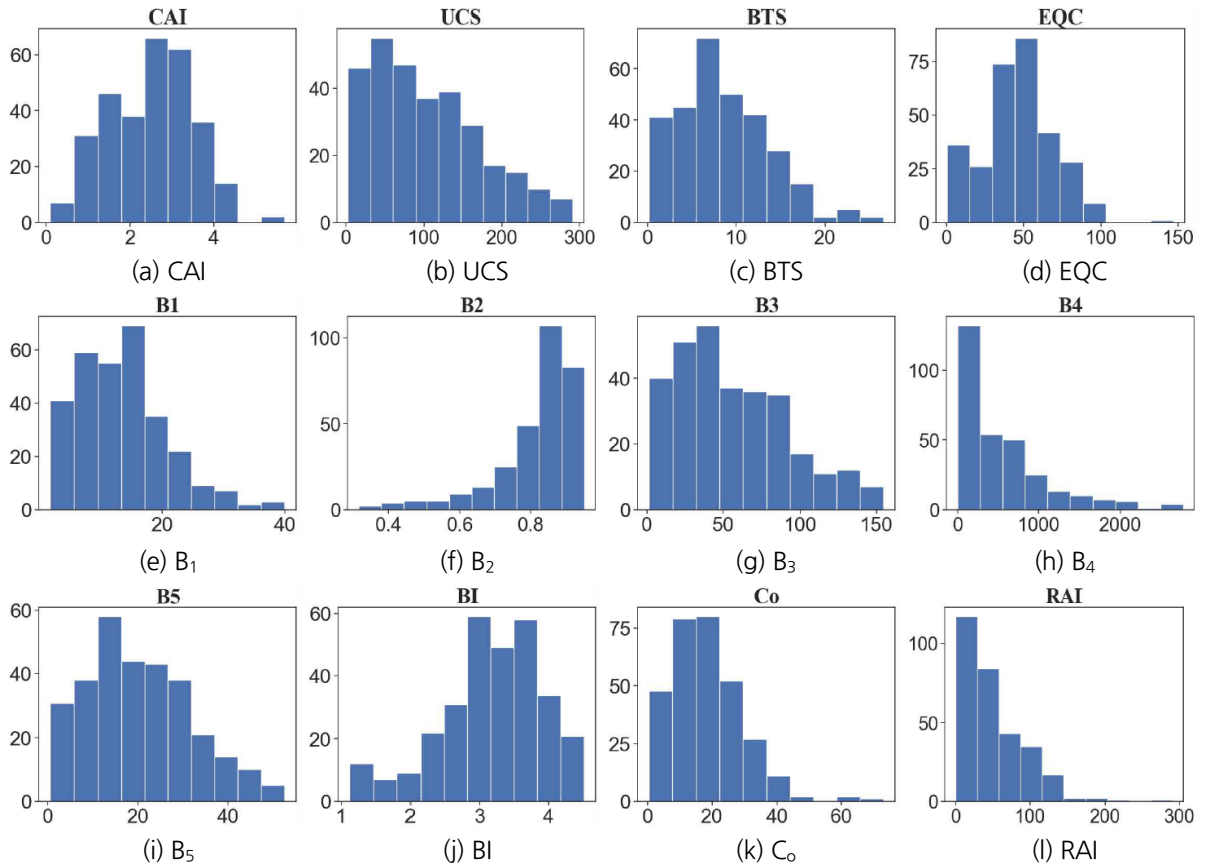


Fig. 1. Histogram of features

변수들 사이 상관관계를 분석하여 Fig. 2와 같이 히트맵으로 나타내었다. 상관계수가 1과 -1에 가까울수록 선형적 관계가 높다는 것을 의미하고 상관계수가 0에 가까운 경우 선형관계가 약하거나 없다는 것을 의미한다. 분석 결과 CAI는 RAI, EQC, BI, 그리고 B₅ 순으로 상관계수가 높게 나타났다. CAI와 변수 B₁, B₂를 제외하고는 양의 상관관계를 보이며 B₁과 B₂는 상관관계가 매우 약한 것으로 나타났다. 전체적으로 독립변수 간의 높은 상관관계가 나타나며, 이는 기존 변수의 함수로 표현되는 파생 변수이기 때문으로 보인다.

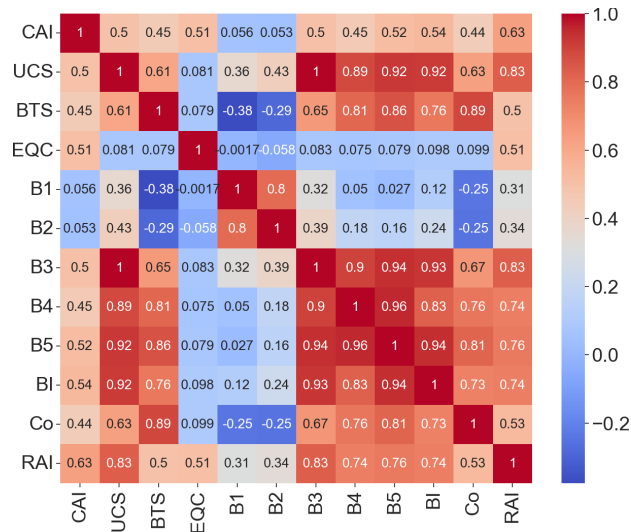


Fig. 2. Pair-wise correlation between features

Fig. 3은 전체 암종에 대한 CAI와 각 변수들 사이 산포도를 나타내며 CAI와 변수 간의 관계는 대부분 선형관계와 비선형 관계가 나타나며 B₁과 B₂는 뚜렷한 상관관계를 확인할 수 없다.

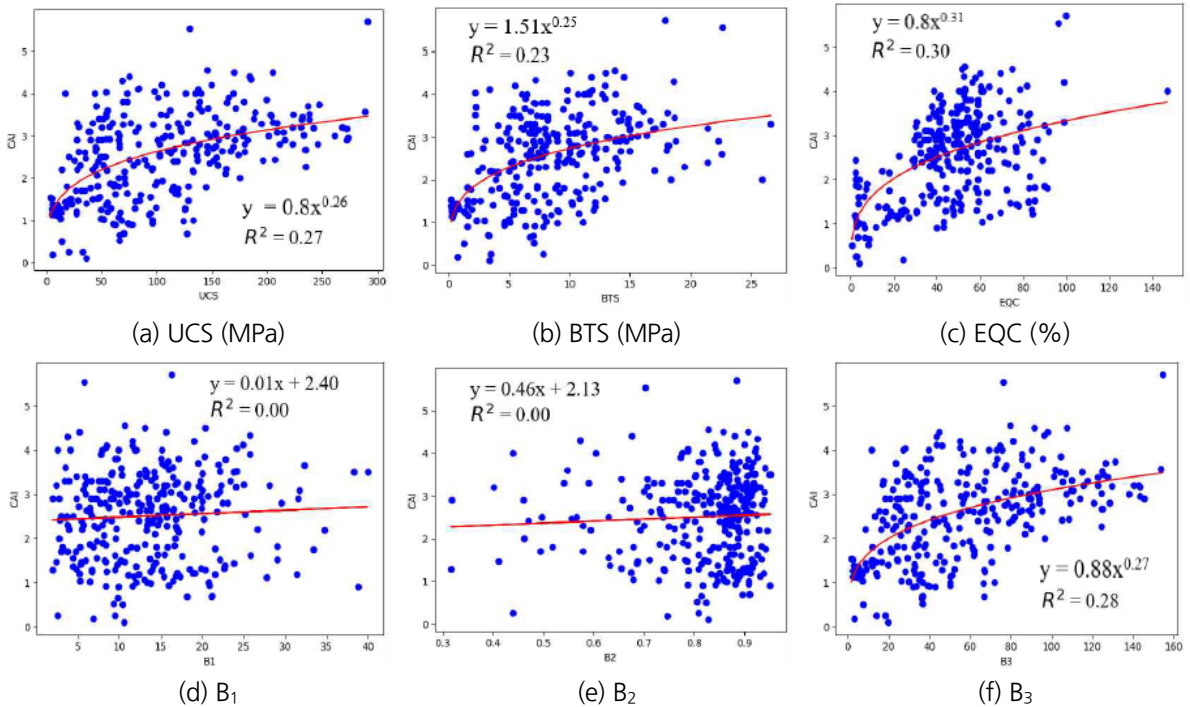


Fig. 3. Relationships between CAI and rock properties in all rock types

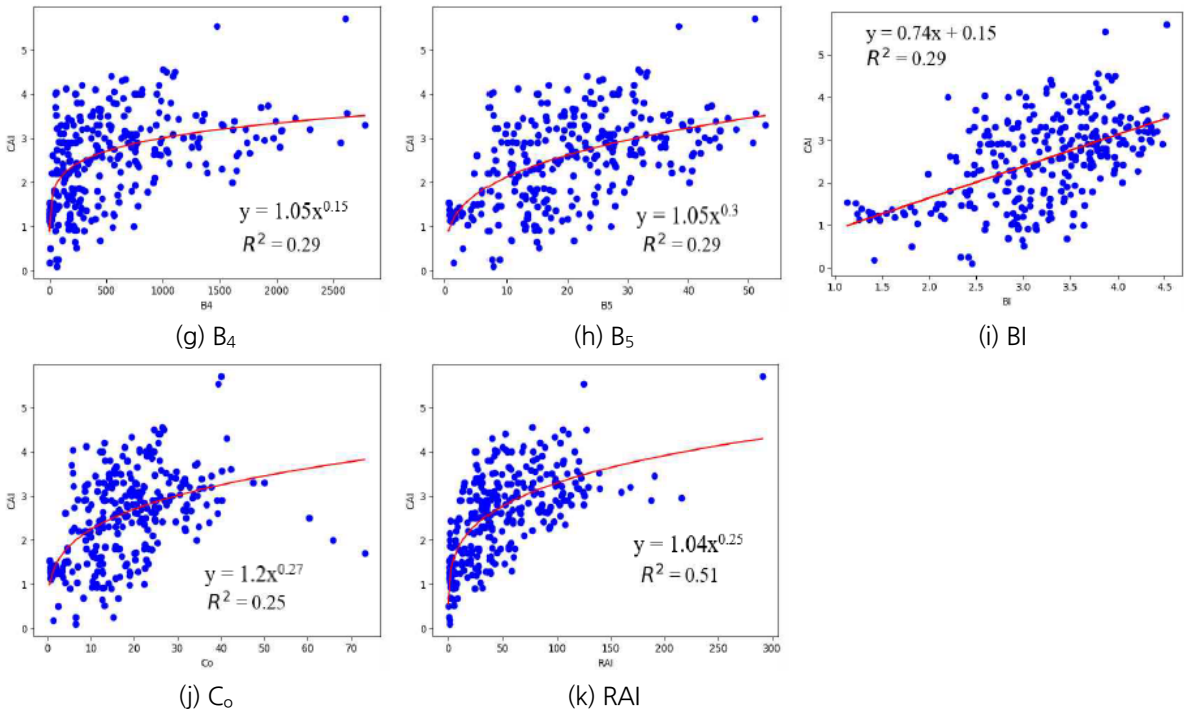


Fig. 3. Relationships between CAI and rock properties in all rock types (continued)

3.3 입력변수선택

3.3.1 다중선형회귀분석

다중선형회귀분석의 입력변수 선택은 모델의 성능과 해석력에 영향을 미치는 단계로 변수의 유의성과 다중공선성을 고려해야 한다. 변수 간 높은 선형상관관계로 발생하는 다중공선성은 모델의 안정성과 해석력에 악영향을 미친다. 최적의 입력변수 선정을 위하여 가능한 모든 변수 조합을 생성하여 각각에 대한 모델을 구현하고 평가하여 최적의 입력변수를 선택하는 완전 탐색법(Exhaustive search method)은 변수의 수가 증가함에 따라 조합의 수가 기하급수적으로 증가하므로 계산시간과 비용이 많이 소요되는 단점이 있다. 이에 대한 대안으로, 계산 효율성을 고려하여 전진 선택법(Forward selection method)과 후진 제거법(Backward elimination method) 및 단계적 선택법(Stepwise selection method)으로 변수 선택을 수행하였다. 전진 선택법은 출력 변수 상관관계 및 유의성이 가장 큰 입력변수부터 입력하여 하나씩 입력변수를 선택하는 방법이며, 후진 제거법은 전진 선택법과는 반대로 모든 변수를 선택한 후 설명력이 가장 낮은 입력변수부터 한 개씩 순차적으로 제거하여 선정한다. 단계적 선택법은 전진 선택법과 후진 제거법의 단점을 고려하여 전진 선택법을 수행 후 선택된 변수를 재검사하고 더 이상 중요하지 않은 변수를 제거하는 방법이다. 변수 선택 과정에서 변수들의 유의성은 p-value를 고려하였으며 p-value가 0.05보다 크면 해당 변수의 회귀계수가 통계적으로 유의미하지 않다고 판단하였다.

CAI 다중회귀분석 모델의 입력변수는 암종으로 구분하여 세 가지 변수 선택법으로 입력변수를 선정하였다. 다중공선성 문제를 해결하기 위하여 일반적으로 VIF가 10 이상인 변수는 다른 변수와 상관관계가 높다고 고려되기 때문에 VIF가 10 이상의 변수를 높은 순으로 제거하고 VIF를 다시 확인하여 전체 변수의 VIF가 10 이하가 될 때까지 반복하였다. Table 1는 선정된 입력변수 중 VIF가 10 이상인 변수를 제거한 입력변수들과 R²를 나타내었다. 변수의 R²를 비교하여 최종적으로 전체 암종에서 EQC와 BI, 화성암에서 UCS와 EQC, 변성암에서 BTS와 RAI, 퇴적암에서 BTS, EQC, B₂ 및 RAI를 선정하였다.

Table 1. Variables selected by the variable selection method along with VIF and R²

	Variable	Forward selection			Backward elimination				Stepwise selection	
		EQC	BI	RAI	UCS	BTS	EQC	EQC	BI	
All rock types	VIF	4.95	5.54	3.93	5.04	5.52	2.78		4.44	4.44
	R ²	0.47			0.50				0.50	
Igneous	Variable	UCS		EQC	UCS		EQC		UCS	EQC
	VIF	3.43		3.43	3.43		3.43		3.43	3.43
	R ²	0.44			0.44				0.44	
Metamorphic	Variable	BTS		RAI	BI		RAI		RAI	
	VIF	2.83		2.83	2.81		2.81		-	
	R ²	0.38			0.32					
Sedimentary	Variable	BTS	EQC	B ₂	RAI	UCS	BTS	EQC	B ₂	RAI
	VIF	4.26	4.74	5.02	4.05	7.56	5.85	2.37	5.26	-
	R ²	0.64				0.63				

3.3.2 머신러닝 모델

연구에서 머신러닝 모델들은 트리 기반의 모델들로 사이킷런(Scikit-learn) 라이브러리에서 제공하는 ‘feature_importances_’ 기능을 통해 각 변수의 모델 예측 중요도를 수치적으로 나타낼 수 있다. ‘feature_importances_’ 기능은 각 변수가 모델 예측에 기여한 정도를 평가하는 데 사용되는데 트리의 각 분할에서 해당 변수가 기여하는 불순도(Impurity) 감소량을 기반으로 계산된다. 불순도는 각 노드에서 데이터의 동질성을 나타내는 지표로, 노드의 분할이 불순도를 줄이는 방향으로 이루어짐으로써, 변수의 중요도를 결정한다.

트리 기반의 모델은 데이터의 복잡한 구조와 변수 간의 비선형적 상호작용을 고려할 수 있기 때문에, 선형 모델에서 문제가 되는 다중공선성의 영향을 상대적으로 덜 받는 것으로 여겨진다. 본 연구에서는 머신러닝 모델 각각의 변수 중요도를 확인하고 평균값을 계산하여 공통적인 중요도가 높은 변수를 확인하였다(Fig. 4). 변수가 많이 짐으로 발생하는 모델의 복잡성을 고려하여 상위 4개의 변수를 머신러닝 모델의 입력변수로 선정하였다. 최종적으로 CAI 예측모델은 전체 암종에서 BTS, EQC, B₃ 및 RAI를 화성암에서는 UCS, EQC, B₃ 및 RAI를 변성암에서는 BTS, EQC, C₀ 및 RAI를 퇴적암에서는 EQC, B₁, B₂ 및 RAI 입력변수로 선정하였다.

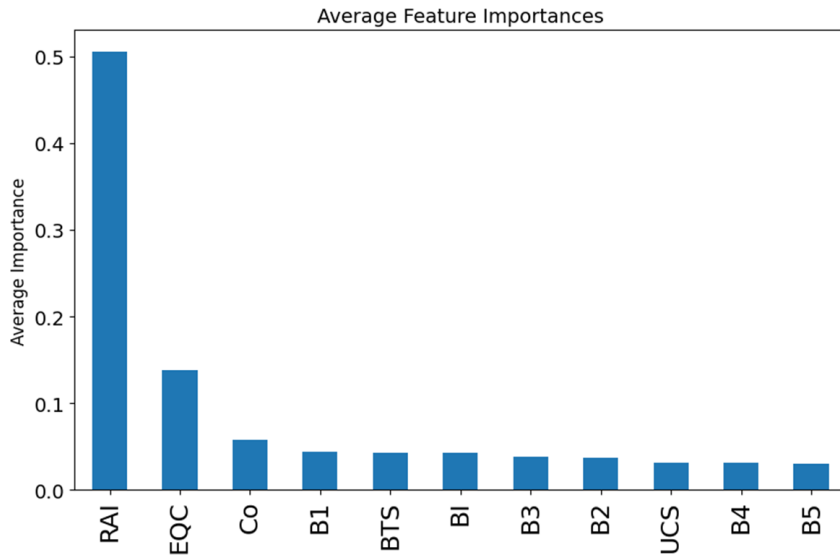


Fig. 4. Variable importance in the CAI prediction model in all rock types

4. 결과 및 성능 비교

4.1 다중선형회귀분석

예측 모델의 성능을 확인하기 위하여 전체 데이터를 학습용과 검증용 8:2의 비율로 나누었다. 모델의 성능 평가 지표는 결정계수(R^2)와 평균 제곱근 오차(Root-mean-square error, RMSE)를 사용하였다. R^2 는 모델의 설명력으로 해석되며, 예측모델이 실제 값을 얼마나 잘 설명할 수 있는지를 나타낸다. 일반적으로 R^2 는 0에서 1사이의 값을 가지며, 이 값이 1에 근접할수록 모델의 설명력이 높은 모델임을 의미한다. RMSE는 회귀모델의 예측값과 실제 값의 차이를 이용해 예측 성능을 판단하는 지표이며, 일반적으로 수치가 낮을수록 성능이 우수하다고 판단한다. 모델의 안정성과 데이터의 패턴이나 순서에 따른 편향을 최소화하기 위해 5겹 교차검증을 수행하여 5개 평가지표의 평균값을 최종 평가지표로 활용하였다.

다중선형회귀분석의 결과는 Table 2와 Fig. 5에 나타내었다.

Table 2. Summary of cross-validated results from multiple linear regression analysis for CAI

Rock type	Input variables	Train data		Test data	
		RMSE	R^2	RMSE	R^2
All	EQC, BI	0.72	0.50	0.73	0.48
Igneous	UCS, EQC	0.68	0.45	0.70	0.37
Metamorphic	BTS, RAI	0.63	0.38	0.65	0.24
Sedimentary	BTS, EQC, B ₂ , RAI	0.54	0.64	0.59	0.53

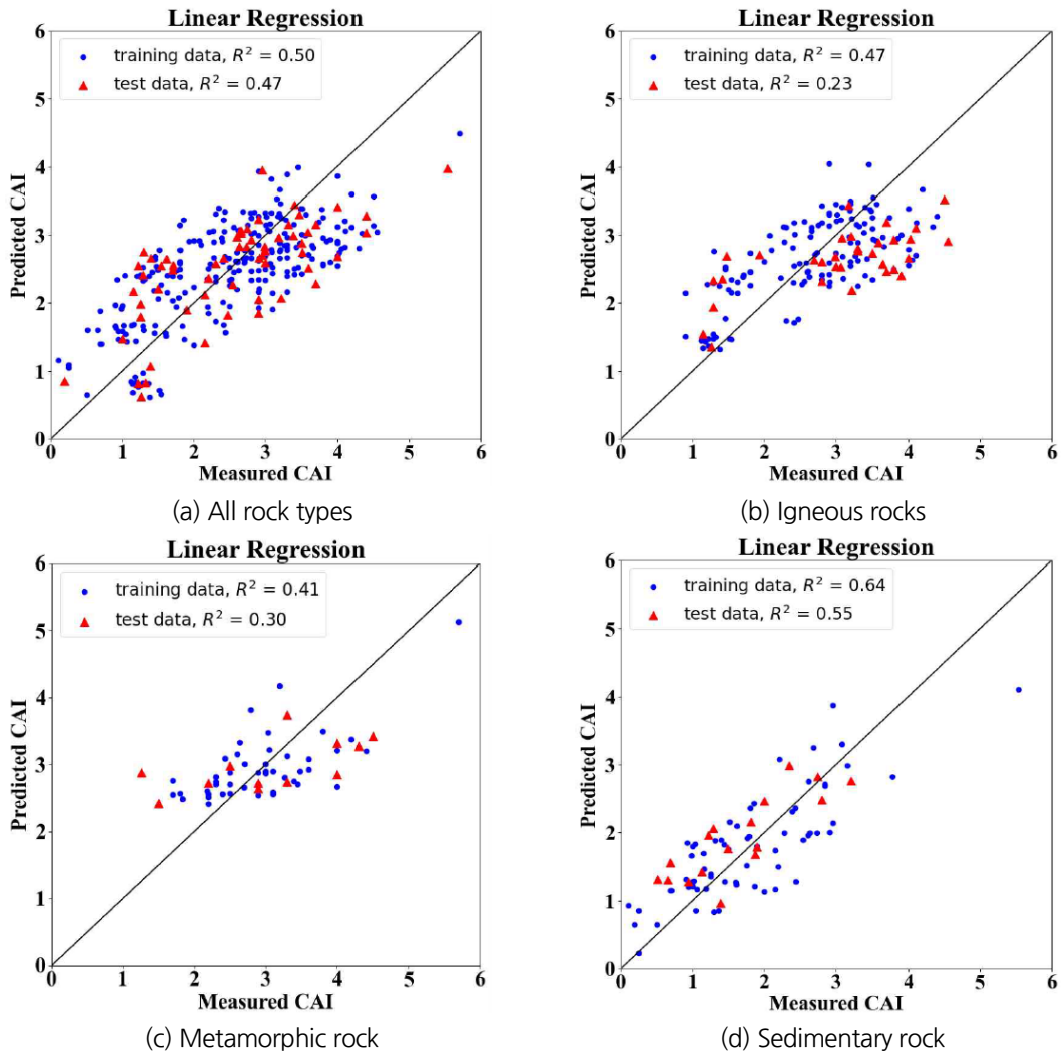


Fig. 5. CAI multiple linear regression analysis based on rock types

본 연구에서 도출된 CAI 예측식은 전체 암종, 화성암, 변성암 및 퇴적암에 대해 각각 식 (30)~식 (33)로 나타났다.

$$CAI = 0.02EQC + 0.68BI - 0.58 \quad (R^2 = 0.50) \text{ for all rock types} \quad (30)$$

$$CAI = 0.006UCS + 0.02EQC + 0.89 \quad (R^2 = 0.44) \text{ for igneous rocks} \quad (31)$$

$$CAI = 0.06BTS + 0.007RAI + 2.09 \quad (R^2 = 0.38) \text{ for metamorphic rocks} \quad (32)$$

$$CAI = 0.09BTS + 0.01EQC + 2.46B_2 + 0.005RAI - 1.49 \quad (R^2 = 0.64) \text{ for sedimentary rocks} \quad (33)$$

4.2 머신러닝 모델

본 연구에서 CAI 예측을 위해 사용된 머신러닝 알고리즘들은 Decision Tree, Random Forest, Extra Tree, XGBoost, Gradient Boosting 및 AdaBoost이다. 최적의 머신러닝 모델을 구현하기 위하여 하이퍼파라미터 튜닝 (Hyperparameter tuning)을 실시하였다. 하이퍼파라미터는 머신러닝 모델이 학습할 수 없어 사용자가 직접 조정 하는 값으로 적절한 훈련 모델의 학습을 위해 설정하는 변수이다. 지정한 파라미터 안에서 가장 우수한 성능의 하이퍼파라미터 조합을 찾는 그리드 서치(Grid search) 튜닝 방법을 사용하여 예측모델에 대해 최적의 파라미터를 결정하였다.

전체 암종에 대한 머신러닝 기반 CAI 예측 모델들의 학습 데이터와 검증 데이터의 예측 결과를 Fig. 6에 나타 내었으며, 암종별 CAI 예측 모델들의 성능 평가 결과는 Table 3에 나타내었다. 전체적으로 학습 데이터의 예측 성능이 검증 데이터의 예측 성능보다 높게 나타나며 이는 학습 데이터의 모델이 다소 과적합이 되었다고 볼 수 있 다. 이는 학습 데이터의 개수가 머신러닝 학습에 다소 부족했을 수도 있고 문헌 조사를 통해 수집된 데이터의 각 암석의 특이성으로 인한 데이터의 대표성이 부족해서 일 수도 있다.

전체 암종 데이터를 비교하였을 때 학습 데이터와 검증 데이터에서 R^2 가 각각 0.71, 0.51로 나타난 Gradient Boosting의 모델 가장 예측 성능이 우수하게 나타나 최적의 모델로 선정하였다.

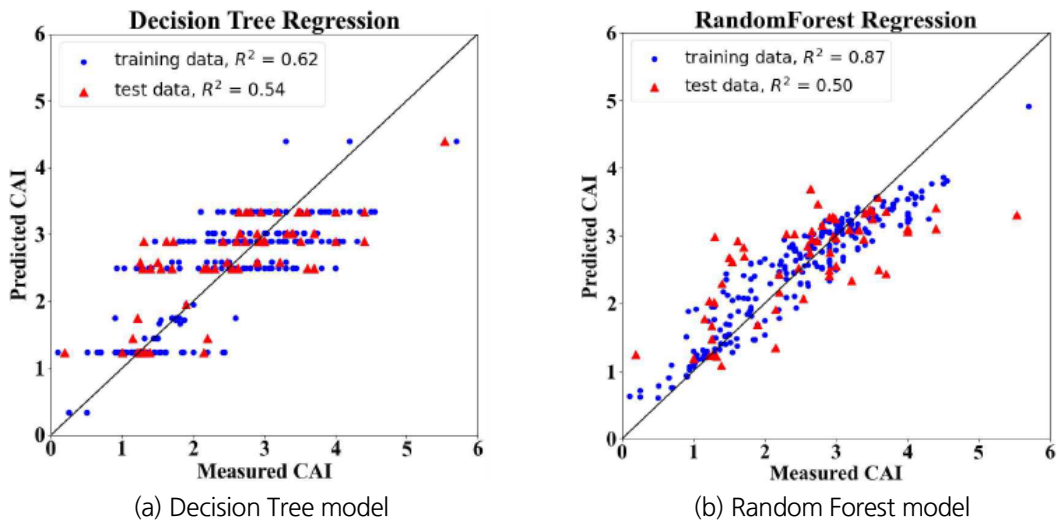
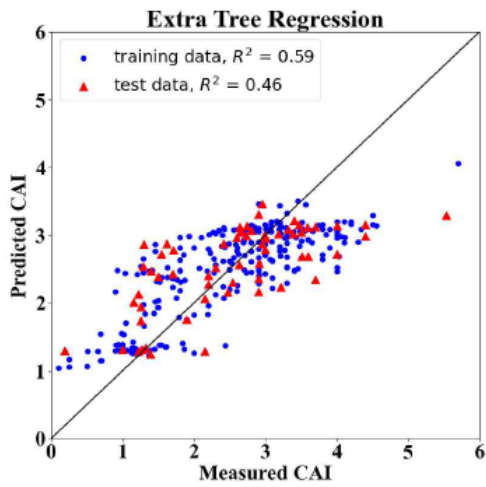
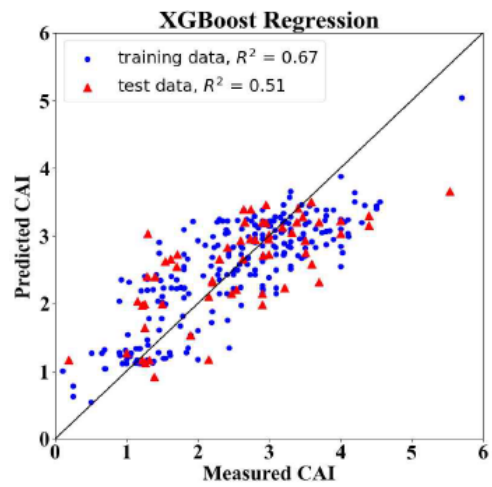


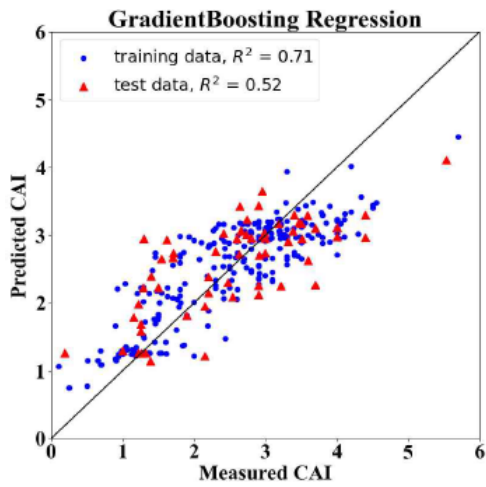
Fig. 6. Results of comparison between measured and predicted CAI for machine learning models in all rock types



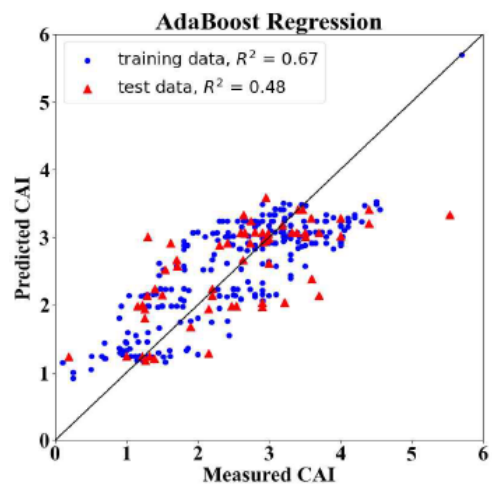
(c) Extra Tree model



(d) XGBoost model



(e) Gradient Boosting model



(f) AdaBoost model

Fig. 6. Results of comparison between measured and predicted CAI for machine learning models in all rock types (continued)

Table 3. Performance evaluation of all CAI machine learning models

Rock type	Model	Train data		Test data	
		RMSE	R ²	RMSE	R ²
All	Decision Tree	0.62	0.62	0.74	0.45
	Random Forest	0.37	0.87	0.71	0.50
	Extra Tree	0.65	0.59	0.72	0.48
	XGBoost	0.59	0.66	0.71	0.50
	Gradient Boosting	0.55	0.71	0.70	0.51
	AdaBoost	0.60	0.65	0.71	0.50
Igneous	Decision Tree	0.63	0.53	0.71	0.35
	Random Forest	0.26	0.92	0.67	0.43
	Extra Tree	0.59	0.59	0.68	0.41
	XGBoost	0.27	0.91	0.66	0.45
	Gradient Boosting	0.38	0.82	0.68	0.42
	AdaBoost	0.37	0.84	0.67	0.44
Metamorphic	Decision Tree	0.62	0.41	0.67	0.20
	Random Forest	0.48	0.65	0.66	0.25
	Extra Tree	0.62	0.41	0.68	0.20
	XGBoost	0.33	0.83	0.69	0.19
	Gradient Boosting	0.23	0.92	0.69	0.17
	AdaBoost	0.39	0.77	0.64	0.29
Sedimentary	Decision Tree	0.21	0.95	0.63	0.45
	Random Forest	0.23	0.93	0.59	0.53
	Extra Tree	0.50	0.70	0.63	0.47
	XGBoost	0.41	0.79	0.60	0.52
	Gradient Boosting	0.38	0.82	0.65	0.43
	AdaBoost	0.21	0.94	0.56	0.57

4.3 선행 연구와 성능 비교

본 연구에서 도출된 CAI 다중선형회귀분석 모델과 최적의 모델로 선정된 Gradient Boosting 모델과 선행연구에서 제시한 CAI 예측모델의 예측 결과를 도시하였으며(Fig. 7), 예측 성능을 RMSE로 평가한 결과는 Table 4에 나타내었다. 그 결과 본 연구에서 도출된 선형회귀분석 모델과 Gradient Boosting 모델의 RMSE가 선행연구 예측모델들의 RMSE보다 더 낮게 나타나 본 연구 결과의 타당성을 확인하였다.

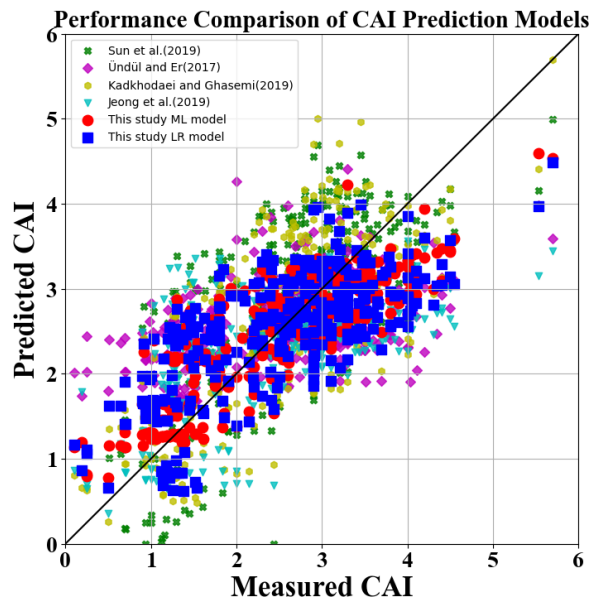


Fig. 7. Performance comparison of CAI prediction models

Table 4. Summary of CAI prediction models

No.	Reference	Model equation	RMSE
1	Sun et al. (2019)	$CAI = 0.996 \ln(RAI) - 0.656$	0.99
2	Ündül and Er (2017)	$CAI = 1.659 + 0.000546 UCS + 0.0991 BTS$	0.90
3	Kadkhodaei and Ghasemi (2019)	$CAI = 0.623 RAI^{0.333} \times UCS^{0.004}$	0.79
4	Jeong et al. (2019)	$CAI = 0.6 EQC^{0.36} \times B_5^{0.03}$	0.87
5	This study LR model	$CAI = 0.02 EQC + 0.68 BI - 0.58$	0.72
6	This study ML model	Gradient Boosting model	0.56

5. 결론

본 연구에서는 디스크 커터 마모지수 CAI를 예측하기 위한 데이터베이스를 구축하고, 이를 통해 마모지수와 암석 물성 간 상관관계를 조사하였다. 다중선형회귀분석을 수행하였고, 머신러닝 기법으로 예측모델을 구현 후 성능을 비교하여 최적의 모델을 선정하였다. 본 연구에서 도출한 다중선형회귀분석 모델과 최적의 머신러닝 기반 모델을 선행연구의 예측모델과 성능을 비교하였다. 이를 통해 도출된 연구 결과는 다음과 같다.

1. 본 연구에 사용된 데이터는 문헌 조사를 통해 수집되었고, CAI, UCS, BTS, EQC가 포함되며, 기존 변수들의 함수로 표현되는 취성도($B_1 \sim B_5$), 점착력(Cohesion, C_0), 암석마모지수(Rock Abrasivity Index, RAI)를 파생 변수로 데이터베이스에 추가하였다. CAI 데이터베이스의 데이터는 302개이고 암종은 화성암, 변성암, 퇴적

암 세 가지로 구분하여 각각 158개, 60개, 84개로 암종별 차이를 확인하기 위하여 암종별로 예측모델을 구현하였다.

2. 다중선형회귀분석을 위한 변수 선택은 P-value를 고려한 전진 선택법, 후진 제거법, 단계적 선택을 적용하였고, 그 후 다중공선성을 고려하여 분산팽창지수를 확인하였다. 전체 암석 유형에서 EQC와 BI, 화성암에서 UCS와 EQC, 변성암에서 BTS와 RAI, 퇴적암에서 BTS, EQC, B₂ 및 RAI를 입력변수로 선정하였다. 다중선형회귀분석을 암종별로 수행하여 예측 성능을 확인하였고 예측식을 도출하였다.
3. 머신러닝 기반에 모델의 변수 선택은 각 모델의 변수 중요도의 평균을 확인하여 입력변수를 선정하였으며, 전체 암종에서 BTS, EQC, B₃ 및 RAI를 화성암에서는 UCS, EQC, B₃ 및 RAI를 변성암에서는 BTS, EQC, C₀ 및 RAI를 퇴적암에서는 EQC, B₁, B₂ 및 RAI 입력변수로 선정하였다. 머신러닝 예측모델 중 Gradient Boosting 모델의 학습 데이터와 검증 데이터의 R²가 각각 0.71, 0.51로 가장 우수하여 최적의 모델로 선정하였다.
4. 본 연구에서 도출한 다중선형회귀분석 모델과 최적의 모델을 선행연구의 예측모델과 RMSE로 예측 성능을 비교하였고, 본 연구의 예측모델의 예측 성능이 더 높게 나타나 연구 결과의 타당성을 확인하였다.
5. 본 연구를 통해 개발된 예측모델은 선행연구의 모델들과 비교하여 우수한 성능을 보였으나, 이는 현재 사용된 데이터베이스에 한정된 결과이기 때문에 일반화에는 어려움이 있어 후속 연구에서는 추가적인 데이터를 통해 예측모델의 성능의 개선이 필요하다.

감사의 글

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2022R1F1A1063228).

저자 기여도

고태영은 연구 개념 및 설계, 원고 검토를 하였고, 강운성은 원고 작성을 하였고, 홍주표는 데이터 수집, 분석 및 해석과 원고 초안을 작성하였다.

References

1. Abu Bakar, M.Z., Majeed, Y., Rostami, J. (2016), "Effects of rock water content on CERCHAR Abrasivity Index", *Wear*, Vol. 368-369, pp. 132-145.
2. Alber, M. (2008), "Stress dependency of the Cerchar abrasivity index (CAI) and its effects on wear of selected rock cutting tools", *Tunnelling and Underground Space Technology*, Vol. 23, No. 4, pp. 351-359.
3. Altindag, R. (2003), "Correlation of specific energy with rock brittleness concepts on rock cutting", *Journal of the Southern African Institute of Mining and Metallurgy*, Vol. 103, No. 3, pp. 163-171.

4. Altindag, R. (2010), "Assessment of some brittleness indexes in rock-drilling efficiency", *Rock Mechanics and Rock Engineering*, Vol. 43, pp. 361-370.
5. Altindag, R., Sengun, N., Sarac, S., Mutluturk, M., Guney, A. (2009), "Evaluating the relations between brittleness and Cerchar abrasion index of rocks", *Proceedings of the EUROCK 2009 ISRM Regional Symposium*, Cavtat, Croatia, pp. 195-200.
6. Capik, M., Yilmaz, A.O. (2017), "Correlation between Cerchar abrasivity index, rock properties, and drill bit lifetime", *Arabian Journal of Geosciences*, Vol. 10, pp. 1-12.
7. Comakli, R. (2019), "Effects of the physico-mechanical properties of low-strength pyroclastic rocks on cutter wear of roadheaders", *Wear*, Vol. 428-429, pp. 205-216.
8. Hamzaban, M.T., Memarian, H., Rostami, J. (2018), "Determination of scratching energy index for Cerchar abrasion test", *Journal of Mining and Environment*, Vol. 9, No. 1, pp. 73-89.
9. He, J., Li, S., Li, X., Wang, X., Guo, J. (2016), "Study on the correlations between abrasiveness and mechanical properties of rocks combining with the microstructure characteristic", *Rock Mechanics and Rock Engineering*, Vol. 49, pp. 2945-2951.
10. Hucka, V., Das, B. (1974), "Brittleness determination of rocks by different methods", *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, Vol. 11, No. 10, pp. 389-392.
11. Jeong, H., Choi, S., Lee, S., Jeon, S. (2019), "Relationship between brittleness of rock, CAI and cutting performance of a TBM disc cutter", *Proceedings of the 2019 World Congress on Advances in Structural Engineering and Mechanics*, Jeju, pp. 17-21.
12. Kadkhodaei, M.H., Ghasemi, E. (2019), "Development of a GEP model to assess CERCHAR abrasivity index of rocks based on geomechanical properties", *Journal of Mining and Environment*, Vol. 10, No. 4, pp. 917-928.
13. Ko, T.Y., Kim, T.K., Son, Y., Jeon, S. (2016), "Effect of geomechanical properties on Cerchar Abrasivity Index (CAI) and its application to TBM tunnelling", *Tunnelling and Underground Space Technology*, Vol. 57, pp. 99-111.
14. Kwak, N.S., Ko, T.Y. (2022), "Machine learning based regression analysis for estimating Cerchar abrasivity index", *Geomechanics and Engineering*, Vol. 29, No. 3, pp. 219-228.
15. Lee, S.D., Jung, H.Y., Jeon, S.W. (2012), "Determination of rock abrasiveness using Cerchar abrasiveness test", *Tunnel and Underground Space*, Vol. 22, No. 4, pp. 284-295.
16. Majeed, Y., Abu Bakar, M.Z. (2019), "Effects of variation in the particle size of the rock abrasion powder and standard rotational speed on the NTNU/SINTEF abrasion value steel test", *Bulletin of Engineering Geology and the Environment*, Vol. 78, pp. 1537-1554.
17. Majeed, Y., Abu Bakar, M.Z., Butt, I.A. (2020), "Abrasive evaluation for wear prediction of button drill bits using geotechnical rock properties", *Bulletin of Engineering Geology and the Environment*, Vol. 79, pp. 767-787.
18. Meng, F., Wong, L.N.Y., Zhou, H. (2021), "Rock brittleness indices and their applications to different fields of rock engineering: A review", *Journal of rock mechanics and geotechnical Engineering*, Vol. 13, No. 1, pp. 221-247.

19. Menschik, F. (2015), Analysis of performance and wear of electrical rock hammer drills, Ph.D. Thesis, Technical University of Munich, pp. 44-45.
20. Moon, K., Yang, S.B. (2020), "Cohesion and internal friction angle estimated from Brazilian tensile strength and unconfined compressive strength of volcanic rocks in Jeju Island", Journal of the Korean Geotechnical Society, Vol. 36, No. 2, pp. 17-28.
21. Ozdogan, M.V., Deliormanli, A.H., Yenice, H. (2018), "The correlations between the Cerchar abrasivity index and the geomechanical properties of building stones", Arabian Journal of Geosciences Vol. 11, pp. 604.
22. Özfirat, M.K., Yenice, H., Şimşir, F., Yaralı, O. (2016), "A new approach to rock brittleness and its usability at prediction of drillability", Journal of African Earth Sciences, Vol. 119, pp. 94-101.
23. Piratheepan, J., Gnanendran, C.T., Arulrajah, A. (2012), "Determination of c and ϕ from IDT and unconfined compression testing and numerical analysis", Journal of Materials in Civil Engineering, Vol. 24, No. 9, pp. 1153-1164.
24. Plinninger, R.J. (2010), "Hardrock abrasivity investigation using the Rock Abrasivity Index (RAI)", Proceedings of the 11th IAEG Congress, Auckland, New Zealand, pp. 3445-3452.
25. Rostami, J., Ghasemi, A., Alavi Gharahbagh, E., Dogruoz, C., Dahl, F. (2014), "Study of dominant factors affecting Cerchar abrasivity index", Rock Mechanics and Rock Engineering, Vol. 47, pp. 1905-1919.
26. Sun, Z., Zhao, H., Hong, K., Chen, K., Zhou, J., Li, F., Zhang, B., Song, F., Yang, Y., He, R. (2019), "A practical TBM cutter wear prediction model for disc cutter life and rock wear ability", Tunnelling and Underground Space Technology, Vol. 85, pp. 92-99.
27. Torrijo, F.J., Garzón-Roca, J., Company, J., Cobos, G. (2019), "Estimation of Cerchar abrasivity index of andesitic rocks in Ecuador from chemical compounds and petrographical properties using regression analyses", Bulletin of Engineering Geology and the Environment, Vol. 78, pp. 2331-2344.
28. Tripathy, A., Singh, T.N., Kundu, J. (2015), "Prediction of abrasiveness index of some Indian rocks using soft computing methods", Measurement, Vol. 68, pp. 302-309.
29. Ündül, Ö., Er, S. (2017), "Investigating the effects of micro-texture and geo-mechanical properties on the abrasiveness of volcanic rocks", Engineering Geology, Vol. 229, pp. 85-94.