

Machine Learning-Based Rock Class Prediction Using Eco-Geological Forest Growth Indicators for Forest Road Cut-Slope Design

Gwang-Jin Park*, Moon Su Park**

*Ph.D. Candidate, Department of Forest Resources, Sunchon National University, Suncheon, Korea

**Professor, Department of Forest Resources, Sunchon National University, Suncheon, Korea

[Abstract]

This study developed an eco-geological approach for predicting rock classes in forest road cut slopes using forest spatial information and machine learning. A total of 330 cut slopes in eastern Jeollanam-do, South Korea, were classified as Hard Rock, Soft Rock, Weathered Rock, or Soil. Topographic and geological variables were combined with forest-derived growth indicators, including HIAge, HIAC, HIDBH, and DBHAge. Seven machine learning algorithms were compared, and Random Forest achieved the highest performance, with an accuracy of 93.88% and an AUC of 0.9895. HIAge and HIAC showed high effect sizes and variable importance, indicating that vegetation growth characteristics may reflect bedrock weathering conditions. The results suggest that forest-derived indicators can serve as eco-geological proxies for high-precision rock class prediction and provide a cost-effective, non-invasive method for assessing ground conditions in forested mountainous areas.

▶ **Key words:** Rock Class Prediction, Machine Learning, Forest Spatial Information, Eco-Geological Proxy, Forest-Derived Growth Indicators

[요 약]

본 연구는 산림 공간정보와 머신러닝을 활용하여 임도 절토사면의 암반등급을 예측하는 생태지질학(Eco-geological) 방법을 제안하였다. 전라남도 동부지역의 임도 절토사면 330개소를 경암, 연암, 풍화암 및 토사로 분류하고, 지형·지질 변수와 수고, 흉고직경 및 임령으로부터 산출한 HIAge, HIAC, HIDBH, DBHAge를 예측변수로 활용하였다. 7개 기계학습 알고리즘을 비교한 결과, Random Forest가 정확도 93.88%와 AUC 0.9895로 가장 우수한 성능을 보였다. 특히 HIAge와 HIAC는 높은 효과크기와 변수 중요도를 나타내어 식생 성장 특성이 기반암의 풍화 상태를 일정 부분 반영할 가능성을 보여주었다. 이러한 결과는 임상파생변수가 암반등급 예측을 위한 생태지질학적 대리지표로 활용될 수 있으며, 광역 산림지역의 지반 상태를 비파괴적이고 경제적으로 평가하는데 기여할 수 있음을 시사한다.

▶ **주제어:** 암반등급 예측, 머신러닝, 산림공간정보, 생태지질학적 대리지표, 임상파생변수

-
- First Author: Gwang-Jin Park, Corresponding Author: Gwang-Jin Park
 - *Gwang-Jin Park (pk1007@nate.com), Department of Forest Resources, Sunchon National University
 - **Moon Su Park (parkms@scnu.ac.kr), Department of Forest Resources, Sunchon National University
 - Received: 2026. 05. 21, Revised: 2026. 06. 29, Accepted: 2026. 07. 15.

I. Introduction

산림지역의 기반시설인 임도(Forest Road)는 산림자원의 효율적 관리와 산림재해 대응을 위한 핵심 인프라로서, 노선 계획 및 시공 단계에서 지반 조건에 대한 정확한 평가가 요구된다[1][2]. 특히 암반등급(Rock Class, RC)은 절토 공법 선정, 굴착 장비 결정, 공사비 산정 및 사면 안정성 평가의 핵심 기준으로 활용되며, 암반등급 판정의 오류는 설계 경제성과 시공 안전성에 직접적인 영향을 미친다[3].

임도 절토사면 공사에서 실제 경암(Hard Rock) 구간이 연암 또는 풍화암으로 과소평가될 경우 브레이커 및 발파 장비의 추가 투입에 따른 공사비 증가가 발생할 수 있다. 반대로 토사(Soil) 또는 풍화암(Weathered Rock)을 경암으로 과대평가할 경우 불필요한 절토와 부적절한 배수계획으로 인해 사면 안정성이 저하되고 산사태 발생 위험이 증가할 수 있다. 따라서 암반등급의 정확한 예측은 경제적 시공계획 수립과 재해 예방을 위한 핵심 선행조건이라 할 수 있다.

전통적인 지반조사 방법인 시추조사(Boring Survey), 탄성파탐사(Seismic Exploration), 현장 관입시험 등은 높은 신뢰성을 제공하지만 조사 비용이 크고 적용 범위가 제한적이라는 한계를 가진다. 특히 접근성이 낮은 산악지역에서는 조사 장비의 진입이 어렵고 조사 밀도가 제한되기 때문에 광범위한 산림지역에 대한 연속적이고 객관적인 지반정보 확보가 현실적으로 어렵다. 또한 조사 결과의 해석 과정에서 전문가의 경험과 주관적 판단이 개입될 가능성이 존재하여 설계 단계의 불확실성을 증가시키는 요인으로 작용한다.

최근에는 이러한 문제를 해결하기 위해 기계학습(Machine Learning) 기반의 지반 예측기법이 다양한 공학 분야에서 활용되고 있다[4]. 기계학습은 입력변수와 목표변수 간의 복잡한 비선형 관계를 학습하여 예측모형을 구축하는 데이터 기반 접근법으로, 전통적인 경험식이나 통계적 회귀모형보다 복잡한 자연환경에서 높은 예측 성능을 제공하는 것으로 보고되고 있다. 특히 Random Forest, Support Vector Machine, Artificial Neural Network 및 Gradient Boosting 계열 알고리즘은 다수의 환경변수 간 상호작용을 효과적으로 처리할 수 있어 지반 공학 및 암반공학 분야에서 활용 사례가 증가하고 있다.

본 연구에서는 전라남도 동부지역 임도 절토사면 330개소의 현장조사 데이터를 활용하여 암반등급 예측모형을 구축하고자 한다. 기존에 활용되어 온 지형·지질 변수뿐만 아니라 수고(Tree Height), 흉고직경(DBH: Diameter at Breast Height), 임령(Stand Age)을 기반으로 한 임상파

생변수(HIAge, HIAC, HIDBH, DBHAge)를 추가적으로 도입하였다. 이를 통해 경암(HR), 연암(SR), 풍화암(WR), 토사(S)의 4개 암반등급을 예측하고, 산림 식생의 생장 특성이 지하 기반암의 풍화 및 물리적 특성과 연관성을 가진다는 생태지질학적(Eco-Geological) 가설의 타당성을 검증하고자 한다.

본 연구의 구체적인 목적은 다음과 같다.

첫째, 7개의 기계학습 알고리즘을 적용하여 암반등급 예측 성능을 비교하고 최적의 예측모형을 선정한다.

둘째, 수고, 흉고직경 및 임령 기반의 임상파생변수가 암반등급 예측에 미치는 영향을 분석하고 그 활용 가능성을 검증한다.

II. Preliminaries

1. Related works

1.1 Introduction of Ground Classification System and Machine Learning

암반등급(Rock Class)은 토목공학 및 산림공학 분야에서 설계, 시공성 평가, 공사비 산정 및 사면 안정성 검토를 위한 핵심 기초정보로 활용된다[5]. 국내 건설 및 산림공사 현장에서는 일반적으로 국토교통부 「건설공사 표준품셈」[6]과 한국도로공사 「도로설계요령」[7]에 따라 풍화도(weathering degree), 절리 간격(joint spacing), 일축압축강도(uniaxial compressive strength), 굴착 시공성(excavatability) 등을 종합적으로 고려하여 토사(Soil), 풍화암(Weathered Rock), 연암(Soft Rock), 경암(Hard Rock)으로 구분한다.

이러한 공학적 분류체계는 단순한 암석학적 분류가 아니라 굴착 장비 선정, 공사비 산정 및 시공계획 수립과 직접 연계되므로 실무적 중요성이 매우 크다[8]. 특히 임도 절토사면 공사에서는 암반등급 판정의 오류가 공사비 증가와 사면 안정성 저하를 동시에 초래할 수 있기 때문에 사전 예측 정확도가 매우 중요하다[9].

그러나 기존의 지반조사 방법은 시추조사(Boring Survey), 탄성파 탐사(Seismic Exploration), 현장 관입 시험(Field Penetration Test) 등 물리적 조사기법에 의존하고 있다. 이러한 방법은 높은 정확도를 제공하지만 조사 비용과 시간이 많이 소요되며, 접근성이 낮은 산악지역에서는 공간적 적용성이 현저히 제한된다. 특히 광역 산림지역에서는 제한된 조사점만을 기반으로 지반 상태를 추정해야 하므로 공간적 불확실성이 크게 증가한다.

최근 이러한 한계를 극복하기 위하여 머신러닝(Machine Learning) 기반의 지반정보 예측기법이 활발히 연구되고 있다[10]. 머신러닝은 입력 변수와 목표 변수 사이의 복잡한 비선형 관계를 학습하는 데이터 중심(Data-driven) 접근법으로, 기존 경험식이나 회귀모형보다 높은 예측 성능을 제공하는 것으로 알려져 있다.

대표적인 알고리즘인 Random Forest(RF)는 다수의 의사결정트리를 배깅(Bagging) 방식으로 결합함으로써 과적합을 억제하고 다중공선성에 강건한 특성을 가진다[11]. Support Vector Machine(SVM)은 고차원 공간에서 최대 마진 초평면을 탐색하여 우수한 일반화 성능을 제공하며[12], XGBoost는 Gradient Boosting과 정규화 기법을 결합하여 복잡한 비선형 관계를 효과적으로 학습하는 것으로 평가받고 있다[13].

실제 지질공학 분야에서는 RF, SVM, XGBoost 기반의 암반분류 연구가 80~95% 수준의 높은 분류 정확도를 보고하고 있으며[10], 산사태 취약성 평가에서도 AUC 0.90 이상의 우수한 성능이 제시되고 있다[14]. 그러나 기존 연구들은 다음과 같은 공통적인 한계를 가진다.

첫째, 표면 기반 변수 의존성으로 기존 연구 대부분은 DEM 기반 지형 변수(경사, 표고, 곡률), 지질도, 원격탐사 영상에 의존하여 지반을 예측하였다. 이러한 변수들은 지반 형성에 영향을 미치지만, 지하 기반암의 풍화 상태를 간접적으로만 반영할 뿐 직접적인 생태-지반 상호작용 정보를 포함하지 못한다.

둘째, 지반-생태 연계 메커니즘 미반영 문제로, 대부분의 ML 기반 암반분류 연구는 지반을 순수 지형·지질학적 문제로 접근하여, 기반암과 식생 간 상호작용을 반영하지 못하였다.

셋째, 공간 일반화 문제로 무작위 분할(Random Split)에 기반한 검증은 공간자기상관(spatial autocorrelation)으로 인해 성능을 과대평가할 수 있으며, 실제 미지 지역 적용 시 성능 저하 위험이 존재한다[15]. 따라서 차세대 지반 예측 연구는 기존 지형·지질 중심 접근을 넘어, 기반암-토양-식생의 통합 시스템 관점에서 지반을 이해할 필요가 있음을 시사한다.

1.2 Eco-geology and bedrock - vegetative interaction

생태지질학(Eco-Geology)은 지질학적 기반(Substrate)과 생태계 구조 및 기능 간의 상호작용을 규명하는 융합 학문 분야로 정의된다[16]. 최근 연구들은 기반암의 물리적·화학적 특성이 토양 생성, 수분 저장 능력, 양분 순환 및 식생 생산성을 결정하는 핵심 요인임을 보고하고 있다.

기반암은 풍화(weathering) 과정을 거쳐 레골리스(regolith)와 토양층을 형성하며, 이 과정에서 생성된 풍화층의 두께와 광물 조성은 토양의 수분 보유력과 양분 공급 능력을 결정한다. 특히 풍화대(weathered zone)는 식생이 이용 가능한 지하수 저장소(subsurface water reservoir) 역할을 수행하며, 가뭄 시기 식생의 생존성과 생산성을 좌우하는 중요한 환경 요인으로 작용한다.

Hahm et al.[16]은 기반암 조성이 산악 생태계의 생산성과 수관 구조를 설명하는 핵심 인자임을 보고하였으며, 이는 기반암 조건이 식생 성장을 지배하는 주요 요인임을 시사한다. 즉, 기반암의 풍화 상태는 토양 수분 환경을 변화시키고, 이는 다시 수목의 성장 특성에 영향을 미치는 연쇄적 구조를 형성한다.

이러한 관점에서 수고(Tree Height), 흉고직경(DBH), 임령(Stand Age) 등의 성장 변수는 단순한 임분 특성이 아니라 지하 기반암 상태를 간접적으로 반영하는 생태지질학적 대리지표(Eco-Geological Proxy)로 해석될 수 있다. 특히 동일한 수종과 유사한 임령 조건에서도 기반암 풍화 정도에 따라 수목의 상대 성장 효율이 달라질 수 있으며, 이는 지상부 식생 정보를 이용하여 지하 기반암의 상태를 추정할 수 있는 가능성을 제공한다.

따라서 본 연구는 기존의 지형·지질 중심 접근을 확장하여 식생 성장 특성과 기반암 풍화 상태 간의 연계 구조를 머신러닝 모델에 통합하고자 한다. 이는 지반공학과 생태학을 연결하는 새로운 융합적 접근으로서, 산림지역 암반 등급 예측 분야에 새로운 이론적 프레임워크를 제시한다.

1.3 Research Gaps and an Integrated Rock Class Prediction Framework Using Forest Spatial Information

국내에서는 임도 노선계획 단계에서 암반구간을 추정하기 위한 지형·암반 관련 연구가 상대적으로 부족한 실정이다. 전현준(2025)은 서부지방산림청 관할 12개 임도 예정 노선을 대상으로 현장조사를 실시하여 암반구간의 GPS 위치 정보를 취득하였다[20]. 다만, 현장조사는 암반구간의 위치 확인을 중심으로 수행되었으며, 분석에 사용된 설명 변수는 기존 주제도와 수치지형도로부터 추출하였다. 구체적으로 20 m 간격으로 설정한 총 1,233개 측점(암반구간 133개, 비암반구간 1,100개)을 대상으로 대축척 임상도, 산림입지토양도 및 수치지형도를 중첩하여 모암형, 토양형, 토성, 임상, 수종, 경사도 및 지형지수 등의 공간정보를 구축하였다. 이후 카이제곱 검정과 독립표본 t-검정을 적용하여 암반 유무에 따른 변수별 차이를 분석한 결과, 모암형, 토성 및 경사도 등이 암반구간 추정과 유의한 관련성을 가지는 것으로 나타났다.

그러나 이러한 선행연구는 다음과 같은 한계를 가진다. 첫째, 기존 연구는 주로 절개사면 단위의 국지적 암반 비율 산정이나 암반 유무에 따른 변수의 통계적 차이 검증에 집중되어 있어, 광역 산림공간정보를 기반으로 암반구간 또는 암반등급을 분류하는 예측모델로 확장되지 못하였다. 특히 전현준(2025)의 연구는 암반구간과 비암반구간 간의 특성 차이를 규명하는 데 중점을 두었기 때문에, 실제 임도 노선계획에 적용할 수 있는 분류·예측모델이나 알고리즘별 성능평가 결과를 제시하지 못하였다. 또한, 암반구간과 비암반구간의 표본 수가 각각 133개와 1,100개로 크게 불균형하고 조사대상도 12개 노선으로 제한되어 있어, 도출된 영향인자만으로 미조사 구간의 암반 유무를 안정적으로 분류하는 데 한계가 있다.

둘째, 기존 연구에서 활용된 지형지수는 일정 면적 단위의 개괄적인 지형 특성을 평가하기 위해 구축된 지표로서, 선형(線形) 구조를 가지는 임도 노선의 세부적인 지반 변화를 판별하는 데 제약이 있다. 특히 동일한 지형 범주 내에서도 기반암의 종류와 풍화 정도가 국지적으로 달라질 수 있으므로, 실제 지질 조건이 서로 다른 암반구간과 비암반구간이 동일하거나 유사한 지형 특성으로 표현될 가능성이 있다.

셋째, 기존 연구는 기반암의 풍화 특성과 산림식생 생장 간의 생태학적 상호작용을 충분히 고려하지 못하였다. 선행연구에서는 임상도에서 제공되는 임상과 수종 등의 범주형 속성을 주로 활용하였으며, 임령, 수고 및 흉고직경과 같은 정량적 생장정보 간의 관계를 파생변수로 가공하여 분석하지 않았다. 이에 따라 기반암의 분포와 풍화 상태가 토양의 생성과 수분·양분 조건을 매개로 산림식생의 생장에 영향을 미칠 수 있다는 생태지질학적 관계를 암반 예측에 활용하려는 접근은 제한적으로 이루어졌다.

넷째, 기존 연구에서는 암반구간 또는 암반등급의 분류 가능성을 정량적으로 평가할 수 있는 예측모델의 구축과 다양한 알고리즘 간 성능 비교가 충분히 이루어지지 않았다. 또한, 특정 조사대상지에서 구축된 자료와 분석 결과가 다른 지역이나 신규 임도 노선에도 적용될 수 있는지에 대한 공간적 일반화 가능성 역시 검토되지 못하였다.

따라서 본 연구에서는 이러한 연구 공백을 보완하기 위하여 임령, 수고 및 흉고직경 등 산림공간정보의 생장속성을 조합한 HIAge, HIAC, HIDBH 및 DBHAge 등의 임상파생변수를 구축하였다. 또한, 기존의 지형·지질·토양 변수와 임상파생변수를 통합하여 다양한 머신러닝 알고리즘의 암반등급 분류성능을 비교·평가하였다. 아울러 식생의 생장 특성이 기반암의 분포와 풍화 상태에 따른 지하 환경의

차이를 간접적으로 반영할 수 있다는 생태지질학적 가정에 기초하여, 임상파생변수의 통계적 변별력과 모델 내 예측 기여도를 검토하였다. 이를 통해 기존의 지형·지질 중심 접근을 산림식생의 생장정보와 결합한 융합적 암반등급 예측 프레임워크로 확장하고자 하였다.

다만, 본 연구의 모델 성능은 전체 자료를 학습집합과 시험집합으로 무작위 분할한 후, 모델 구축에 사용되지 않은 독립 시험집합을 대상으로 평가하였다. 따라서 본 연구에서 확인된 예측성능은 구축된 표본자료의 범위 내에서 해석할 필요가 있으며, 다른 지역이나 신규 임도 노선에 대한 공간적 일반화 가능성을 직접적으로 입증한 것은 아니다. 향후 연구에서는 지역 또는 임도 노선 단위의 공간 블록 교차검증과 독립적인 외부자료를 이용한 외부검증을 수행하여 모델의 공간적 전이성과 실무 적용 가능성을 추가로 평가할 필요가 있다.

III. The Proposed Scheme

1. Study Area and Data Configuration

본 연구는 전라남도 동부권에 위치한 고흥군, 보성군, 순천시 및 여수시 일대의 임도 절토사면을 대상으로 수행하였다. 연구 대상지는 남해안 산지지형과 내륙 산악지형이 혼재되어 있으며, 다양한 지질대와 산림 입지 조건을 포함하고 있어 암반등급 분포의 공간적 다양성을 확보할 수 있는 지역으로 판단하였다. 따라서 본 연구지역은 산림 지역 암반등급 예측모델의 적용 가능성을 검증하기 위한 대표적인 시험지(Test Bed)로 적합하다.

데이터 신뢰성 확보를 위하여 설계고 기준 약 1~3 m 범위 내에서 노출된 임도 절토사면 총 330개 지점을 조사 대상으로 선정하였다. 각 조사 지점에서는 현장 계측자료와 산림 공간정보를 함께 수집하여 분석에 활용하였다. 조사 대상은 지질 조건, 지형 조건 및 임분 특성이 서로 상이한 지역에 분포하도록 구성함으로써 특정 지질 유형 또는 산림 환경에 편중되지 않도록 하였다.

암반등급(Rock Class)은 현장 시공 경험이 풍부한 산림 기술자가 절토면에 노출된 암반 상태를 직접 조사하여 판정하였다. 판정 기준은 절리 간격(Joint Spacing), 굴삭기(0.7 m³급)의 굴착 시공성(Excavatability), 해머 타격에 의한 반발 경도(Rebound Hardness), 풍화 정도(Weathering Degree)를 종합적으로 고려하였으며, Figure 1은 국내 산림사업 및 건설공사 실무에서 활용되는 공학적 암반분류 기준에 따라 경암(Hard Rock: HR),

연암(Soft Rock: SR), 풍화암(Weathered Rock: WR), 토사(Soil: S)의 4개 범주로 구분하였다.

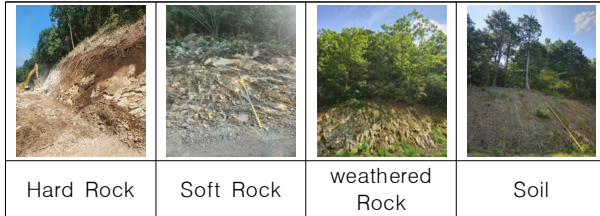


Fig. 1. Field survey view photo.

본 연구는 암반등급을 다중범주 분류(Multi-Class Classification) 문제로 정의하였으며, 각 범주 간의 균형 있는 학습을 위하여 데이터 분할 단계에서 클래스 불균형(Class Imbalance)의 영향을 최소화하고자 하였다. 이를 위해 전체 330개 관측치를 훈련 집합(Training Set)과 시험 집합(Test Set)으로 70:30 비율로 분할하였다. 70:30 분할 방식은 기계학습 분야에서 일반적으로 활용되는 검증 전략으로, 제한된 표본 규모에서 충분한 학습 데이터 확보와 독립적인 성능 검증 간의 균형을 제공하는 것으로 알려져 있다[18].

또한 데이터 분할 과정에서는 층화 무작위 추출(Stratified Random Sampling)을 적용하여 4개 암반등급의 상대적 비율이 훈련 집합과 시험 집합 모두에서 원본 데이터와 동일하게 유지되도록 하였다. 최종적으로 훈련 집합은 232개 표본, 독립 시험 집합은 98개 표본으로 구성되었다. 이러한 분할 전략은 특정 암반등급에 대한 학습 편향을 최소화하고 모델 평가의 신뢰성을 향상시키기 위한 목적에서 수행되었다.

2. Variable Configuration and Forest-derived Variables

Table 1. Data Configuration

Item	Value	Description
dependent variable	RC	Rock Class (HR, SR, WR, S)
topographic variables	SLOmean, ELEmean, CUGeneral	Average slope, elevation, general curvature
random variable	DBH, HI, Age	the diameter of the chest
Derivative ratio variable	HIAC, HIAC, HIDBH, DBHAge	ratio of labor and chest diameter to wage age/age age
Categorical variables	TN, REFROCK, PRRCKMDN	Water species, geological college, and parent rock classification

Table 1은 본 연구에 투입된 변수의 구성과 학술적 특성을 정리한 것이다. 종속변수는 암반등급(Rock Class: RC)이며, 독립변수는 지형변수(평균경사 SLOmean, 평균표고 ELEmean, 일반곡률 CUGeneral), 임분 원시변수(흉고직경 DBH, 수고 HI, 임령 Age), 임상파생변수(HIAC, HIAC, HIDBH, DBHAge), 그리고 범주형 변수인 수종(TN), 지질대(REFROCK), 모암분류(PRRCKMDN)로 구성된다.

기존 산림생장학(Forest Growth Science)에서 수고(Tree Height), 흉고직경(DBH), 임령(Age)은 임분의 생산성과 생육 상태를 평가하는 대표적인 지표로 활용되어 왔다. 그러나 이러한 절대 성장량(absolute growth metrics)은 수종, 경쟁 상태, 기후 조건 및 산림 관리 이력 등 다양한 환경적 요인의 영향을 동시에 받기 때문에 지반 상태를 직접적으로 설명하는 변수로 활용하기에는 한계가 있다.

이에 본 연구에서는 절대 성장량 대신 상대 성장 효율(relative growth efficiency)의 개념에 주목하였다. 상대 성장 효율은 동일한 생육 기간 동안 수목이 달성한 성장 수준을 의미하며, 생육 환경의 질적 차이를 보다 민감하게 반영할 수 있다. 이를 정량화하기 위하여 다음과 같은 임상파생변수를 새롭게 정의하였다.

$$-HIAC = \text{Height} / \text{Age}$$

$$-HIAC = \text{Height} / \text{Age Class}$$

$$-DBHAge = \text{DBH} / \text{Age}$$

$$-HIDBH = \text{Height} / \text{DBH}$$

이들 변수는 단순한 성장량이 아니라 수목이 주어진 시간 동안 얼마나 효율적으로 성장하였는지를 나타내는 상대적 성장 지표(relative growth indicators)로 해석할 수 있다. 생태지질학(Eco-Geology) 관점에서 기반암의 풍화 정도는 토심(soil depth), 수분 저장 능력(water-holding capacity), 양분 공급 능력(nutrient availability) 및 근권(root zone)의 발달 가능성을 결정하며, 이는 궁극적으로 식생의 성장 속도와 생산성에 영향을 미친다.

예를 들어 풍화가 충분히 진행된 기반암 지역에서는 토양층이 상대적으로 깊고 수분 이용 가능성이 높기 때문에 동일한 임령에서도 수목의 성장 효율이 증가할 가능성이 크다. 반면 경암이 지표면 가까이에 분포하는 지역에서는 토양 발달과 수분 공급이 제한되어 상대적으로 낮은 성장 효율을 나타낼 수 있다.

따라서 HIAC, HIAC, HIDBH 및 DBHAge와 같은 임상파생변수는 단순한 산림생장 변수가 아니라 지하 기반암의 풍화 상태를 지상부 식생의 성장 반응으로 환산하여 표현하는 생태지질학적 대리지표(Eco-Geological Proxy)로 해석될 수 있다. 이러한 변수는 기존의 지형·지질 변수만

으로는 설명하기 어려운 식생-기반암 상호작용 정보를 포함하며, 본 연구가 제안하는 생태지질학 기반 암반등급 예측모델의 핵심 입력 변수로 활용된다.

특히 본 연구의 차별성은 기존 암반등급 예측 연구가 주로 지형 및 지질 정보에 의존한 것과 달리, 식생의 성장 효율 정보를 이용하여 지하 기반암 상태를 간접적으로 추론하는 새로운 접근을 제시하였다는 점에 있다. 이는 산림생태학과 지반공학을 연결하는 융합적 예측 프레임워크로서 학문적·실무적 확장 가능성을 제공한다.

3. EDA and Modeling Design

기계학습 모델 구축에 앞서 데이터의 분포 특성, 변수 간 관계 및 모델링 적합성을 평가하기 위하여 탐색적 데이터 분석(Exploratory Data Analysis: EDA)을 수행하였다[17][18]. EDA는 데이터의 전반적인 품질을 점검하고 잠재적인 이상치(outlier), 클래스 불균형(class imbalance), 변수 간 상관구조 및 비선형적 관계를 사전에 확인함으로써 후속 모델링의 타당성을 확보하기 위한 절차로 구성되었으며, 주요 분석 결과는 Figure 2에 제시하였다.

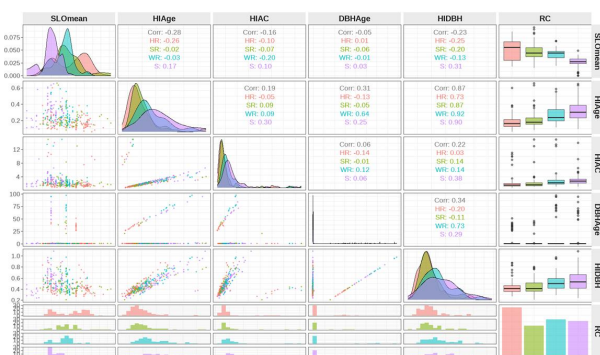


Fig. 2. Dependent variable distribution and scatterplot and variable characteristic distribution graphs.

먼저 종속변수인 암반등급(Rock Class: RC)의 빈도와 구성 비율을 확인하여 특정 클래스에 심각한 불균형이 존재하는지를 검토하였다. 연속형 독립변수의 분포 특성과 자료 범위는 히스토그램과 산점도 행렬(scatterplot matrix)을 이용하여 확인하였으며, 이를 통해 변수별 분포 형태, 이상치의 존재 가능성 및 변수 간 잠재적인 선형·비선형 관계를 탐색하였다[17].

또한 본 연구에서 제안한 임상파생변수인 HIAge, HIAC, HIDBH 및 DBHAge는 수고, 흉고직경, 임령 및 지위수 등 기존 임상변수의 조합을 통해 산출된 비율형 변수이므로, 원변수 및 파생변수 간 상관관계와 다중공선성(multicollinearity)의 가능성을 추가로 검토하였다. 분석

결과 일부 임상변수와 파생변수 사이에서 구조적으로 높은 상관관계가 확인되어, 후속 모델링 단계에서는 변수 간 종속성과 각 모델의 다중공선성에 대한 민감도를 함께 고려하였다[18].

전체 330개 관측자료는 클래스별 구성 비율의 차이로 인한 학습 편향을 줄이기 위하여 층화 무작위 추출(stratified random sampling)을 적용하여 70:30의 비율로 분할하였다[19]. 이에 따라 232개 자료는 훈련 집합으로, 98개 자료는 독립 시험 집합으로 구성하였으며, 두 집합에서 원자료의 4개 암반등급인 경암(Hard Rock), 연암(Soft Rock), 풍화암(Weathered Rock) 및 토사(Soil)의 구성 비율이 가능한 한 유지되도록 하였다. 종속변수인 암반등급은 범주형 변수로 처리하였다.

모델의 과적합(overfitting)을 완화하고 최적의 하이퍼파라미터(hyperparameter)를 탐색하기 위하여 훈련 집합에 5-Fold 교차검증을 3회 반복하는 반복 교차검증(Repeated Cross Validation)을 적용하였다[19]. 각 반복에서 훈련자료를 5개의 부분집합으로 구분한 후 일부를 검증자료로 순환 활용함으로써 특정 자료 분할에 따른 성능 변동과 평가 편향을 줄이고, 모델의 일반화 성능을 안정적으로 추정하고자 하였다. 자료 전처리와 하이퍼파라미터 튜닝은 독립 시험자료의 정보가 학습 과정에 포함되지 않도록 훈련자료와 교차검증 과정 내에서 수행하였다[18][19].

예측 성능을 객관적으로 비교하기 위하여 학습 원리가 서로 다른 7개의 지도학습 알고리즘을 선정하였다. 적용 모델은 트리 앙상블 기반의 Random Forest(RF)[11]과 XGBoost[13], 신경망 기반의 NN, 단일 은닉층 다층퍼셉트론(MLP-Single), 다중 은닉층 다층퍼셉트론(MLP-Multi), 커널 기반의 Support Vector Machine(SVM)[12] 및 거리 기반의 K-Nearest Neighbor(KNN)로 구성하였다[18][19]. 이와 같은 다중 알고리즘 비교는 특정 모델 구조에 따른 분석 편향을 줄이고, 산림 공간정보와 임상파생변수의 특성에 적합한 예측 모델을 상대적으로 평가하기 위하여 수행하였다.

최종 모델의 성능은 모델 학습과 하이퍼파라미터 조정에 사용되지 않은 독립 시험 집합을 대상으로 Accuracy, AUC 및 F1-score 등의 평가 지표를 이용하여 비교하였다[18][19]. 이를 바탕으로 각 알고리즘의 암반등급 분류 성능과 일반화 가능성을 종합적으로 검토하고, 산림 공간정보를 활용한 암반등급 예측에 상대적으로 적합한 기계학습 모델을 선정하였다. R 4.0 환경에서 수행한 기계학습 기반 암반등급 분류예측모델의 구축 및 도출 과정은 Figure 3에 제시하였다.

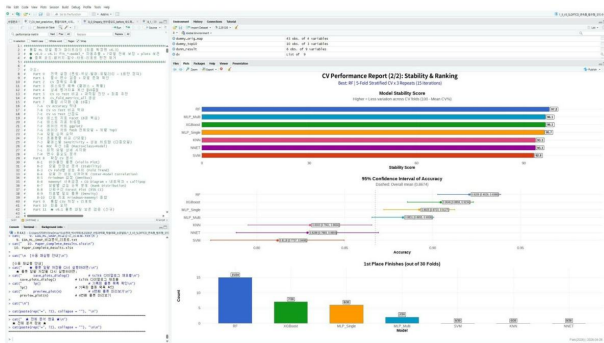


Fig. 3. Derivation of ML-based classification prediction model analysis. (R4.0 program)

IV. Conclusions

본 연구는 전남 동부지역 임도 절토사면에서 조사된 산림·지형·지질 및 임상 자료를 바탕으로 산림 공간정보와 기계학습 기법을 활용한 암반등급 예측모델을 구축하고, 임도 노선계획과 사전 지반조사를 지원하는 보조적 의사 결정 도구로서의 적용 가능성을 검토하였다. 주요 연구 결과는 다음과 같다.

독립 시험 집합(Test Set, n=98)을 대상으로 7개 기계 학습 분류 알고리즘의 예측 성능을 비교·평가하였다. 이 가운데 주요 4개 모델의 Accuracy, AUC 및 F1-score를 Table 2에 제시하였으며, 전체 7개 모델의 AUC 비교 결과는 Figure 4에 나타내었다. 또한 각 모델의 암반등급별 ROC 곡선은 Figure 5에 제시하였다.

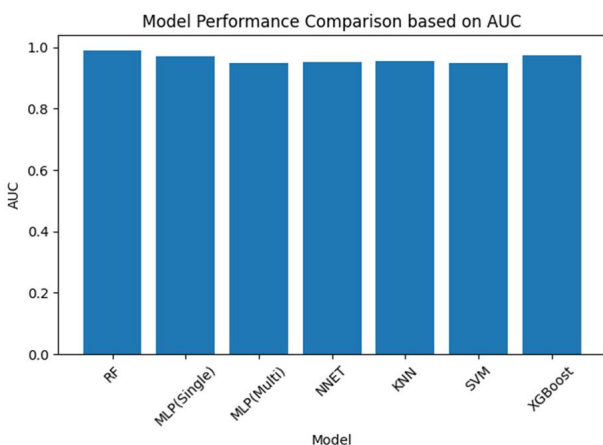


Fig. 4. AUC-based model performance comparison graph

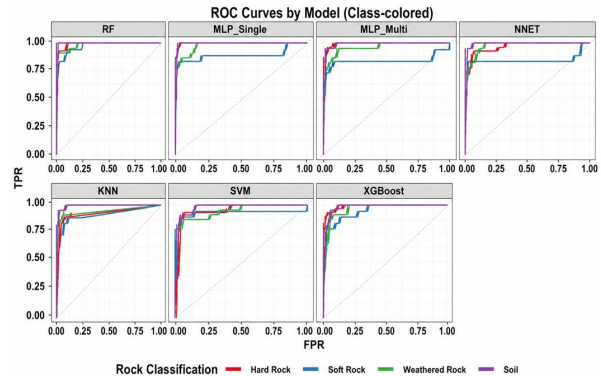


Fig. 5. Comparing ROC curves and AUC values by Model.

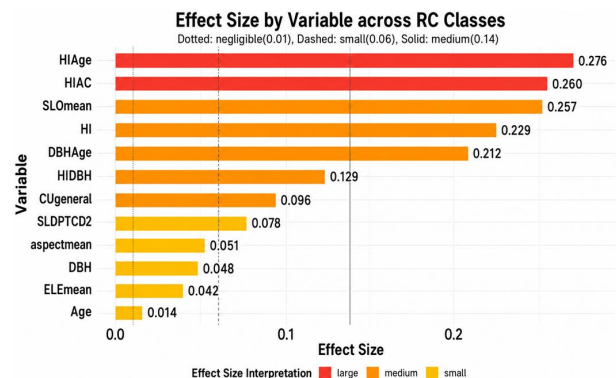


Fig. 6. Effect Size by Variable across RC Classes

Table 2에서 Random Forest(RF)는 Accuracy 93.88%, AUC 0.9895 및 F1-score 0.9374로 나타나 주요 비교 모델 가운데 가장 우수한 성능을 보였다. MLP_Single은 Accuracy 92.86%, AUC 0.9695 및 F1-score 0.9221로 나타나 RF에 이어 두 번째로 높은 정확도와 F1-score를 보였다. NNET의 Accuracy는 86.73%, AUC는 0.9508, F1-score는 0.8664였으며, XGBoost는 Accuracy 83.67%, AUC 0.9730 및 F1-score 0.8249로 나타났다(Table 2).

Table 2. Model Performance

MODEL	Accuracy	AUC	F1-score
Random Forest (RF)	93.88%	0.9895	0.9374
MLP (Single)	92.86%	0.9695	0.9221
NNET	86.73%	0.9508	0.8664
XGBoost	83.67%	0.9730	0.8249

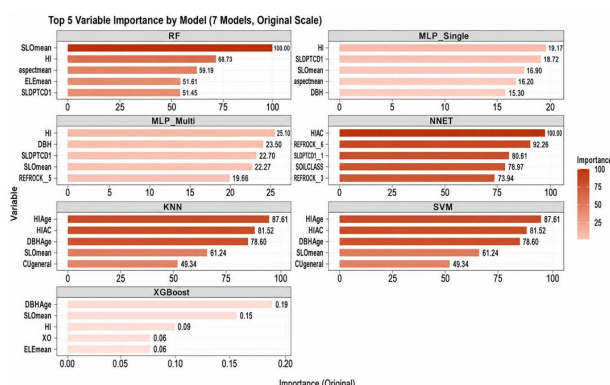


Fig. 7. Top 5 Variable Importance by Model(7 Models, Original Scale)

전체 7개 모델의 AUC를 비교한 결과, RF가 가장 높은 값을 보였으며, XGBoost와 MLP_Single도 상대적으로 높은 AUC를 나타내었다(Figure 4). 반면 Accuracy와 F1-score에서는 모델 간 성능 차이가 AUC보다 비교적 뚜렷하게 나타났다. 특히 XGBoost는 AUC가 0.9730으로 높게 나타났으나 Accuracy와 F1-score는 각각 83.67%와 0.8249로 상대적으로 낮아, 순위 판별능력과 최종 클래스 분류성능 사이에 차이가 있음을 확인할 수 있었다(Table 2; Figure 4).

모델별·클래스별 ROC 곡선을 비교한 결과, 대부분의 모델에서 ROC 곡선이 좌측 상단에 위치하여 전반적으로 양호한 클래스 판별능력을 보였다(Figure 5). RF의 클래스별 ROC 곡선은 네 개 암반등급에서 비교적 안정적으로 좌측 상단에 분포하였으며, 이는 RF가 특정 클래스에 편중되지 않고 전반적으로 균형 있는 판별성능을 나타냈음을 보여준다. MLP_Single, MLP_Multi 및 NNET에서도 대체로 높은 ROC 성능이 확인되었으나, 일부 클래스에서는 곡선 형태와 판별성능에 차이가 나타났다(Figure 5).

이상의 결과를 종합하면 다음과 같다.

첫째, RF는 Accuracy, AUC 및 F1-score에서 비교적 일관되게 높은 값을 나타내어 본 연구자료의 암반등급 분류에 상대적으로 적합한 모델로 평가되었다(Table 2; Figure 4; Figure 5). 이러한 결과는 RF가 산림·지형·지질 변수에 포함된 비선형 관계와 변수 간 상호작용을 비교적 안정적으로 학습하였기 때문으로 판단된다. 다만 본 연구의 성능평가는 구축된 표본과 독립 시험 집합을 기준으로 이루어졌으므로, 다른 지역의 독립자료를 활용한 외부검증을 통해 모델의 공간적 일반화 가능성을 추가로 확인할 필요가 있다.

둘째, 임상 파생변수는 Kruskal-Wallis 검정에서 연속형 변수 중 상대적으로 큰 효과크기를 나타내어 암반등급 간 분포 차이를 구분하는 데 유용한 통계적 정보를 제공하였다.

특히 HI0Age와 HI0C는 각각 0.276과 0.260의 효과크기를 보여 주요 변별변수로 확인되었다(Figure 6). 또한 이들 변수는 RF를 비롯한 여러 기계학습 모델의 변수 중요도 분석에서도 HI(수고), DBH(흉고직경) 등 주요 임상변수와 함께 상위권에 위치하여 암반등급 예측에 기여하는 주요 인자로 나타났다(Figure 7). 다만 변수 중요도의 크기는 모델별 산정 방식과 변수 간 상관구조마다 상이한 영향을 받을 수 있으므로, 특정 변수의 독립적인 인과효과로 해석하기보다는 예측 모델 내에서의 상대적 기여도로 이해하는 것이 적절하다.

HI(수고), DBH(흉고직경), HI0Age(임령대비수고)와 HI0C(영급대비수고)는 각각 상대적 생장 수준을 나타내므로, 유사한 생육조건에서 나타나는 수목의 생장 효율 차이를 정량화한 지표로 볼 수 있다. 이들 변수가 암반등급 간 통계적 차이를 보이고 기계학습 모델에서도 주요 변수로 평가된 결과는 기반암의 종류와 풍화 상태, 토양 발달 및 수분환경과 같은 지하 조건이 수목의 장기적인 생장 특성에 일정 부분 반영될 수 있다는 생태지질학적 가설을 정량적으로 뒷받침한다.

그러나 본 연구에서 확인된 결과는 임상 파생변수와 암반등급 사이의 통계적 연관성과 예측 기여도를 제시한 것이며, 기반암의 풍화 상태가 수목생장에 미치는 직접적인 인과관계를 확정한 것은 아니다. 따라서 향후에는 토양수분, 유효토심, 토양물리성, 뿌리 발달 및 암반 풍화도에 대한 현장조사를 병행하여 임상 파생변수와 지반 조건 사이의 생태적·지질학적 기작을 추가로 검증할 필요가 있다.

이러한 점을 고려할 때 임상 파생변수와 RF 기반 예측 모델은 시추조사 또는 현장 지반조사를 직접 대체하는 수단이라기보다, 임도 계획단계에서 예상 암반구간을 사전에 선별하고 현장조사 및 시추조사의 우선순위를 설정하는 비파괴적 사전평가 자료로 활용될 가능성이 있다.

ACKNOWLEDGEMENT

This work was supported by the Glocal University Project of Sunchon National University.

REFERENCES

- [1] D. R. Montgomery, "Road surface drainage, channel initiation, and slope instability," *Water Resources Research*, Vol. 30, No. 6, pp. 1925-1932, 1994. DOI: 10.1029/94WR00538

- [2] F. C. Dai and C. F. Lee, "Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong," *Geomorphology*, Vol. 42, No. 3-4, pp. 213-228, 2002. DOI: 10.1016/S0169-555X(01)00087-3
- [3] N. Barton, R. Lien, and J. Lunde, "Engineering classification of rock masses for the design of tunnel support," *Rock Mechanics*, Vol. 6, No. 4, pp. 189-236, 1974. DOI: 10.1007/BF01239496
- [4] A. T. C. Goh, "Back-propagation neural networks for modeling complex systems in geotechnical engineering," *Journal of Geotechnical Engineering*, Vol. 121, No. 3, pp. 245-251, 1995. DOI: 10.1061/(ASCE)0733-9410(1995)121:3(245)
- [5] Z. T. Bieniawski, *Engineering Rock Mass Classifications: A Complete Manual for Engineers and Geologists in Mining, Civil, and Petroleum Engineering*, John Wiley & Sons, pp. 1-251, 1989.
- [6] Ministry of Land, Infrastructure and Transport (MOLIT), *Standard Estimating System for Construction Works*, MOLIT, 2024.
- [7] Korea Expressway Corporation (KEC), *Road Design Manual*, KEC, 2021.
- [8] D. U. Deere and D. W. Deere, "The rock quality designation (RQD) index in practice," *ASTM Special Technical Publication*, Vol. 984, pp. 91-101, 1988.
- [9] T. R. Stacey, "Slope stability in hard rock open pit mines," *International Journal of Rock Mechanics and Mining Sciences*, Vol. 46, No. 5, pp. 983-993, 2009. DOI: 10.1016/j.ijrmms.2009.03.005
- [10] A. Asteris, P. G. Kolovos, and K. G. Tsiambaos, "Machine learning techniques for engineering geological characterization of rock masses," *Bulletin of Engineering Geology and the Environment*, Vol. 80, No. 2, pp. 1209-1225, 2021. DOI: 10.1007/s10064-020-01973-5
- [11] L. Breiman, "Random forests," *Machine Learning*, Vol. 45, No. 1, pp. 5-32, 2001. DOI: 10.1023/A:1010933404324
- [12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, Vol. 20, No. 3, pp. 273-297, 1995. DOI: 10.1007/BF00994018
- [13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016. DOI: 10.1145/2939672.2939785
- [14] X. Zhang et al., "Landslide susceptibility mapping using random forest and extreme gradient boosting: A case study of Fengjie, Chongqing," *Geological Journal*, Vol. 58, No. 6, pp. 2372-2387, 2023. DOI: 10.1002/gj.4683
- [15] J. Pohjankukka, T. Pahikkala, P. Nevalainen, and J. Heikkonen, "Estimating the prediction performance of spatial models via spatial k-fold cross validation," *International Journal of Geographical Information Science*, Vol. 31, No. 10, pp. 2001-2019, 2017. DOI: 10.1080/13658816.2017.1346255
- [16] W. J. Hahm, C. S. Riebe, C. E. Lukens, and S. Araki, "Bedrock composition regulates mountain ecosystems and landscape evolution," *Proceedings of the National Academy of Sciences*, Vol. 111, No. 9, pp. 3338-3343, 2014. DOI: 10.1073/pnas.1315667111
- [17] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, 2nd ed., Springer, pp. 15-46, 2021.
- [18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, pp. 219-260, 2009.
- [19] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer, pp. 181-220, 2013.
- [20] Jeon, H.-J., Lee, J.-U., Kim, J.-H., Park, K.-J., & Lee, J.-H. *Analysis of Topographic Factors for Estimating Rock Area in Forest Road Planning*, 2025.

Authors



Gwang-Jin Park received the B.E. degree in Agricultural Engineering from the College of Life and Environmental Sciences, Konkuk University, Seoul, Korea, in 2003. He completed the integrated M.S.-Ph.D. program

in Forest Resources, Division of Forest Resources and Landscape Architecture, Sunchon National University, Sunchon, Korea, where he is currently a Ph.D. Candidate. Since 2003 he has worked in the forestry sector, including the Goheung-gun Forestry Cooperative, forest project construction, and the National Forestry Cooperative Federation, and he is currently working as a forest engineering consultant. He is interested in forest road planning and design, forest engineering, forest spatial information, and machine learning-based prediction modeling.



Moon Su Park received the B.S. degree in Forestry from Sunchon National University, Korea, in 1986, the M.S. degree in Forestry from Chonnam National University, Korea, in 1988, and the Ph.D. degree in Forest

Resources from Gyeongsang National University, Korea, in 1996. He joined the faculty of the Department of Forest Resources, Sunchon National University, Sunchon, Korea, in 1997, and is currently a Professor in the same department. He also served as Dean of the Division of Forest Resources and Landscape Architecture and as Head of the Forest Resources major from 2009 to 2010. He is interested in revegetation of disturbed lands such as cut slopes, erosion control and environmental conservation engineering, and floristic studies.