

# 의사결정나무 및 랜덤포레스트 분류 모델을 이용한 교량 안전등급 예측

홍지수\* · 전세진\*\*

Hong, Jisu\*, Jeon, Se-Jin\*\*

## Prediction of Safety Grade of Bridges Using the Classification Models of Decision Tree and Random Forest

### ABSTRACT

The number of deteriorated bridges with a service period of more than 30 years has been rapidly increasing in Korea. Accordingly, the importance of advanced maintenance technologies through the predictions of age-induced deterioration degree, condition, and performance of bridges is more and more noticed. The prediction method of the safety grade of bridges was proposed in this study using the classification models of the Decision Tree and the Random Forest based on machine learning. As a result of analyzing these models for the 8,850 bridges located in national roads with various evaluation indexes such as confusion matrix, balanced accuracy, recall, ROC curve, and AUC, the Random Forest largely showed better predictive performance than that of the Decision Tree. In particular, random under-sampling in the Random Forest showed higher predictive performance than that of other sampling techniques for the C and D grade bridges, with the recall of 83.4%, which need more attention to maintenance because of the significant deterioration degree. The proposed model can be usefully applied to rapidly identify the safety grade and to establish an efficient and economical maintenance plan of bridges that have not recently been inspected.

**Keywords :** Machine learning, Decision tree, Random forest, Safety grade of bridges, Maintenance

### 초 록

국내에서 공용연수 30년 이상인 노후 교량의 수가 급증하고 있다. 이에 따라 교량 노후도, 상태 및 성능 예측을 바탕으로 한 첨단 유지관리 기술의 중요성이 점차 주목받고 있다. 이 연구에서는 머신러닝 기반의 의사결정나무 및 랜덤포레스트 분류 모델을 사용하여 교량의 안전등급을 예측하는 방법을 제안하였다. 일한국도상 교량 8,850개를 대상으로 해당 모델들을 혼동행렬, 균형 정확도, 재현율, ROC 곡선 및 AUC와 같이 여러 가지 평가 지표를 통해 분석한 결과 전반적으로 랜덤포레스트가 의사결정나무보다 더 나은 예측 성능을 보유했다. 특히 랜덤포레스트 중 랜덤 언더 샘플링 기법은 노후도가 비교적 커서 유지관리에 주의를 기울여야 하는 C, D등급 교량에 대해 재현율 83.4%로 다른 샘플링 기법들보다 예측 성능이 더 뛰어난 것으로 나타났다. 제안된 모델은 최근 점검이 실시되지 않은 교량들의 신속한 안전등급 파악 및 효율적이고 경제적인 유지관리 계획 수립에 유용하게 활용될 수 있을 것으로 기대된다.

**검색어 :** 머신러닝, 의사결정나무, 랜덤포레스트, 교량 안전등급, 유지관리

\* 정회원 · 아주대학교 건설시스템공학과, 공학석사 (Ajou University · hjs1201@ajou.ac.kr)

\*\* 종신회원 · 교신저자 · 아주대학교 건설시스템공학과 교수, 공학박사 (Corresponding Author · Ajou University · conc@ajou.ac.kr)

Received April 19, 2023/ revised May 10, 2023/ accepted May 10, 2023

### 1. 서론

시설물 공용연수가 증가함에 따라 노후화된 시설물의 안전성 확보를 위해 예방적 유지관리 시스템 구축에 대한 중요성이 강조되고 있다. 국토교통부 전국교량표준데이터(MOLIT, 2021b)에 따르면 Fig. 1과 같이 공용연수 30년 이상인 노후 교량의 수가 2030년에는 전체 교량의 42.5%를 차지하는 등 보수보강이 필요한 교량들이 점차 증가할 것으로 예상된다. 따라서 교량의 안전성 및 사용성 저하와 보수보강 비용 증가에 대비한 선제적 유지관리가 필요하다.

“시설물의 안전 및 유지관리에 관한 특별법”(이하 시안법)(MOLIT, 2021c)에 따라 제1~3종시설물에 해당하는 교량은 안전등급에 따라 점검 및 진단을 주기적으로 실시해야 한다. 점검 및 진단의 종류로는 정기안전점검(Periodic safety inspection), 정밀안전점검(Full safety inspection), 그리고 정밀안전진단(Full safety examination)이 있다. 정기안전점검은 외관조사 수준의 점검이며, 정밀안전점검은 면밀한 외관조사와 함께 간단한 측정 및 시험을 실시하는 점검이다. 한편, 정밀안전진단에서는 정밀한 외관조사와 더불어 각종 측정·시험장비를 이용하여 시설물의 상태 및 안전성 평가에 대한 상세 데이터를 확보한다. 교량의 안전등급은 점검 및 진단 결과에 따라 A(우수), B(양호), C(보통), D(미흡), E(불량)의 총 5단계로 분류되며, 안전등급에 따른 점검 및 진단 주기는 Table 1에 나타내었다(MOLIT, 2021a).

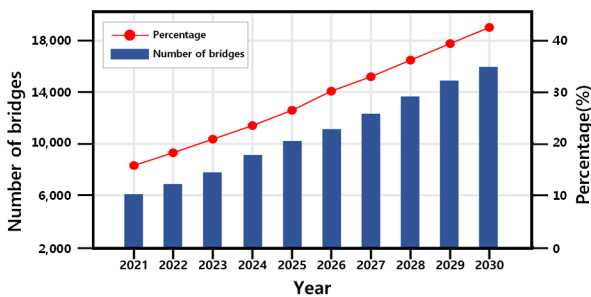


Fig. 1. Bridges with a Service Period of More than 30 Years

최근 정자교 붕괴 사고에서 보듯이 준공 후 30년 이상 경과된 노후 교량에 대한 우려가 커지면서 점검, 진단 및 보수보강을 포함한 유지관리 대책을 합리적으로 수립해야 할 필요성이 더욱 강조되고 있다(Yonhapnews, 2023). 그런데, 교량의 안전성과 사용성은 주로 안전등급을 통해 관리하고 있으므로, 신뢰할 수 있는 안전등급의 판정은 매우 중요하다. 또한, 주기적인 안전점검 및 진단 활동에 따른 보수보강 조치는 시설물의 안전도를 유지하거나 향상시키기 위해 필수적으로 요구되므로(Kang et al., 2016) 시설물 유지관리에 있어 점검 시기를 준수하는 것이 중요하다. 그러나 국내의 경우 사후 유지관리에 초점이 맞추어져 있어 관리시스템의 개발이 미흡하고, 현황 입력과 운영 상황 또한 부정확하게 관리되고 있다(Kim and Yoon, 2018). 2021년 기준으로 일반국도상에 위치한 시안법 적용 대상인 제1~3종시설물에 해당하는 교량 5,600개 중 점검 미 실시 및 점검 시기 미준수 교량은 101개로 1.8%에 해당된다. 또한 시안법 대상이 아닌 교량은 안전점검 및 진단이 필수적이지 않으므로 장기간 안전점검을 실시하지 않은 추가적인 교량들도 다수 존재한다. 이러한 교량들은 교량의 성능 및 상태 파악을 통한 선제적 보수보강 조치를 취하기 어렵기 때문에 안전성 및 내구성의 저하가 우려된다.

또한, 2018년 시안법 개정에 따라 기존에 시안법 대상이 아니었던 소규모 교량도 공용기간 10년이 경과되면 시안법 대상인 제3종 시설물로 새로이 지정되었다. 이로 인해 시안법에 따른 대상 교량이 증가하면서 안전점검 대상 교량이 점차 증가하고 있음에도 관리 인력 부족 및 유지관리 예산 부족 등의 문제로 적절하고 체계적인 안전관리가 이루어지지 않고 있는 실정이다(Lee et al., 2019a). 그리고 국내외 해외의 교량 점검 주기를 비교해서 살펴보면 국내의 점검 주기가 비교적 짧아서 이 또한 관리 인력 부족 현상을 심화시키고 있다(Lee and Kim, 2015). 이처럼 점검 대상 교량의 증가와 짧은 점검 주기에 따른 전문 인력 및 비용 부족으로 인해 합리적인 점검 및 진단이 어려운 실정이다(Kang, 2016).

이에 대한 대책으로 인력 대체 및 비용 절감을 위해 교량의 유지관리 체계에 인공지능과 확률론적 기법 등을 이용하여 교량의

Table 1. Status and Inspection/Examination Cycle According to Safety Grade

Grade	Status	Cycle		
		Periodic safety inspection	Full safety inspection	Full safety examination
A	The best condition without problems	More than every half a year	More than every 3 years	More than every 6 years
B	Minor damages in supplementary members		More than every 2 years	More than every 5 years
C	Minor damages in main members or extensive damages in supplementary members			
D	Major damages in main members	More than 3 times per year	More than every year	More than every 4 years
E	Serious damages in main members and immediate prohibition of usage of the bridge			

부재별 열화도를 예측하는 연구 등이 진행되고 있다.

국내에서는 교량 손상 정도를 예측하는 인공신경망 모델(Oh et al., 2010)과 베이지안 기법을 통한 교량의 부재 상태 예측 모델(Lee et al., 2018)이 제안된 바 있다. 해외의 연구로는 인공지능을 이용한 교량 부재별 상태등급 예측(Bektas et al., 2013; Nguyen and Dinh, 2019) 등이 있다. 이처럼 교량 부재별 상태 및 손상 정도를 예측하려는 연구는 다수 존재하지만, 부재별 상태는 안전등급과 같이 교량 전체의 안전성 및 내구성을 대표한다고 보기 어렵다. 한편 교량 상태 지수(BCI: Bridge Condition Index)의 예측 모델 연구(Martinez et al., 2020)도 존재하는데, 교량 상태 지수는 수치로 표현되어 있어 이에 적합한 머신러닝의 회귀기법을 사용한 바 있다. 하지만 국내 교량에서 사용하는 안전등급은 수치로 표현되지 않기 때문에 이러한 회귀기법을 적용하는 것은 적합하지 않다. 다만 국내에서 규칙 기반 분류 기법으로 교량의 안전등급을 추정하는 방법(Chung et al., 2016)도 시도된 바 있으며, 교량 등급에 있어 C등급 이하 교량을 P(Poor)로, A와 B등급은 G(Good)로 분류하여 두 가지로 통합하고, 이를 이진 분류를 통해 예측하였다. 그러나 A와 B등급은 각 등급에 해당되는 교량 수가 많으며, 교량

유지관리 측면에 있어 등급별로 적절한 유지관리 예산 배분 및 보수보강 조치가 요구되기 때문에 A와 B등급도 구분하여 예측하는 것이 바람직하다. 이는 대상을 2가지 종류로만 구분하는 비교적 단순한 이진 분류 알고리즘을 적용한 데 따르는 한계로 볼 수 있다.

따라서, 이러한 기존 연구의 한계점을 극복하고자 이 연구에서는 안전점검 미 실시 및 주기 미준수 교량의 조속한 안전등급 파악 및 특정 시기의 안전등급 예측을 통한 선제적이고 경제적인 유지관리 계획을 수립하기 위해 인공지능 기법을 활용하여 교량의 합리적인 안전등급 예측 모델을 구축하고자 하였다. 구체적으로 머신러닝의 의사결정나무 및 랜덤포레스트 알고리즘을 사용한 다중 분류 모델을 이용하여 교량의 안전등급을 A등급, B등급과 C, D등급으로 나누어 예측하고, 그러한 예측 성능을 각종 지표를 통해 종합적으로 평가하고 분석하였다. Fig. 2는 이 연구에서 교량 안전등급 예측 모델을 도출한 방법론을 요약하여 보여주고 있다.

## 2. 결정나무 기반 분류 모델 구축

### 2.1 의사결정나무 및 랜덤포레스트

이 연구에서는 2021년 상반기 기준 전국교량표준데이터(MOLIT, 2021b)와 시설물통합정보관리시스템(FMS, 2021)의 교량 정보를 사용하여 일반국도상 교량 8,850개를 대상으로 머신러닝의 결정나무 기반 분류 모델을 구축하였다. 여기서 머신러닝이란 인공지능의 한 분야로 컴퓨터가 데이터를 스스로 학습하여 해결책을 제안하는 기법이다(Géron, 2019). 전국교량표준데이터에는 특정한 범주로 나누어지는 범주형 데이터와 수치로 표현된 연속형 데이터가 모두 포함되어 있기 때문에 연속형과 범주형 데이터를 모두 사용할 수 있는 머신러닝 기법이 필요하다. 또한, 예측하고자 하는 목표인 안전등급은 A, B, C, D, E등급으로 나누어지므로 이진 분류 기법은 적절치 않으며 다중 분류가 가능해야 한다. 분류 알고리즘에는 연속형과 범주형 데이터 사용이 가능하며 이진 분류가 가능한 SVM(Support Vector Machine), 범주형 데이터의 사용만 가능하고 이진 분류를 할 수 있는 로지스틱 회귀, 연속형과 범주형 데이터 모두 사용 가능하며 다중 분류가 가능한 결정나무 기반 알고리즘 등이 있다(Scikit-learn developers, 2007-2022). 이 연구에서는 전국교량표준데이터에 가장 적합하다고 판단되는 결정나무 기반 알고리즘을 사용하였다.

결정나무 기반 알고리즘에도 몇 가지 종류가 있지만 이 연구에서는 가장 대표적인 의사결정나무(Decision Tree)와 랜덤포레스트(Random Forest)를 사용하였다. 두 알고리즘 모두 결정나무를 기반으로 일종의 규칙을 만들어 불순도가 낮아지는 방향으로 대상을 좁혀나가면서 분류한다. 이는 변수에 따른 분리 기준을 통해

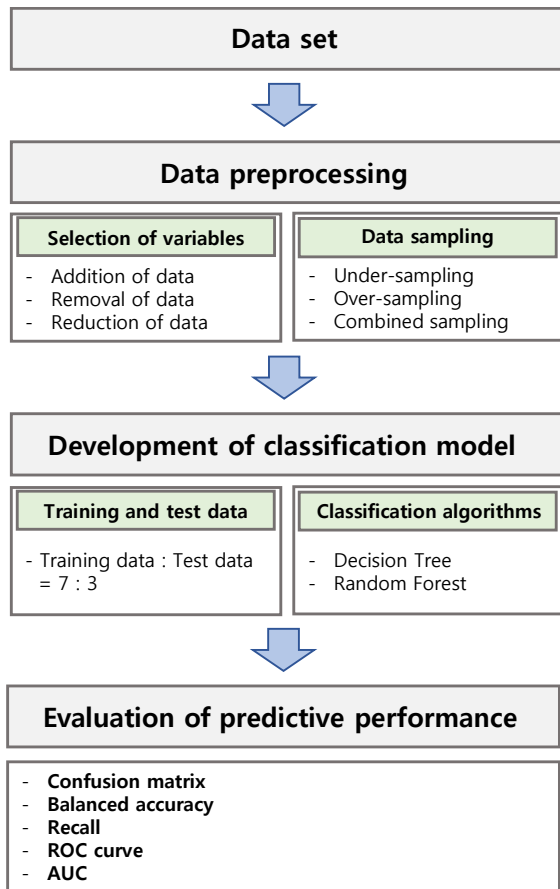


Fig. 2. Development and Evaluation of Classification Model

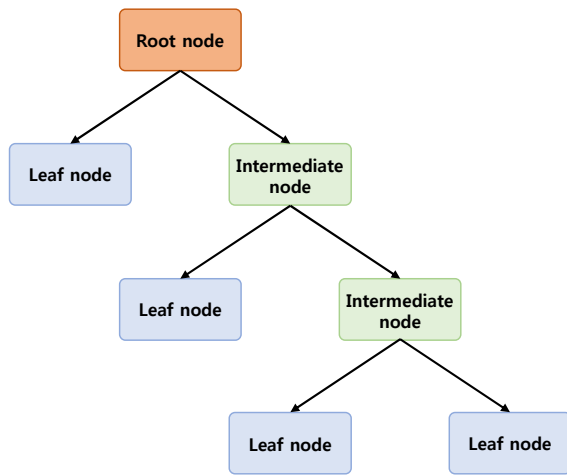


Fig. 3. Structure of Decision Tree-based Algorithm

불순도를 감소시키는 정도를 계산하여 분류 및 예측 작업에 효과적인 중요한 변수를 확인할 수 있어 유용하다(Kazemitabar et al., 2017). 결정나무의 구조는 Fig. 3과 같으며, 루트 노드(Root node)에서 분류가 시작되고 중간 노드(Intermediate node)에서 변수에 따른 분류가 진행된 후 마지막에 리프 노드(Leaf node)에서 최종적으로 분류 결과가 도출된다.

그러나 의사결정나무의 경우 단일 결정나무를 사용하므로 나무 깊이가 깊어지면 과적합(Overfitting)이 발생하여 예측력이 저하될 수 있다. 여기서 과적합이란 모델 생성 시 훈련 데이터(Training data)를 과도하게 학습하여 새로운 데이터를 예측하는 능력이 저하되는 현상을 의미한다. 이러한 의사결정나무의 단점을 보완한 랜덤포레스트는 앙상블 기법을 적용하여 다수의 결정나무를 생성함으로써 다양성을 확보하여 과적합을 방지한다. 일반적으로 앙상블 기법 중 배깅(Bagging: Bootstrap aggregating)을 사용하는데, 배깅은 데이터를 무작위로 선정하여 결정나무를 다양하게 생성하는 기법을 의미한다. 따라서 랜덤포레스트는 Fig. 4와 같이 배깅을 이용하여 다수의 결정나무를 무작위로 생성하고 다수결 또는 평균에 따라 분류하는 알고리즘으로서 과적합을 방지할 수 있어 대체로 의사결정나무보다는 성능이 우수하다. 그러나 데이터 수가 많아지면 분석 소요 시간이 길어지는 단점이 있다(Géron, 2019).

이때 의사결정나무와 랜덤포레스트 모두 데이터를 불순도가 최대한 감소하는 방향으로 분리하도록 학습을 진행하는데, 그 기준으로 엔트로피(Entropy)와 지니계수(Gini)를 사용한다. 엔트로피는 불순도를 수치화한 지표이며 확률 변수의 불확실성을 수치로 나타낸 것으로 식 (1)로 산정하며, 엔트로피의 수치가 1에 가까울수록 불순도가 높다. 지니계수는 확률분포가 어느 범주에 치우쳐 있는 정도를 계산하는 것으로 식 (2)로 도출할 수 있다. 엔트로피와 마찬가지로 지니계수가 1에 가까울수록 불순도가 높다.

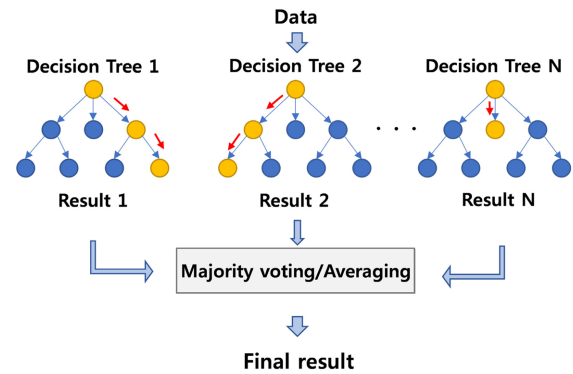


Fig. 4. Structure of Random Forest Algorithm

$$\text{Entropy}(A) = - \sum_{k=1}^m p_k \log_2(p_k) \quad (1)$$

$$\text{Gini}(A) = 1 - \sum_{k=1}^m p_k^2 \quad (2)$$

여기서,  $A$ 는 범주 전체,  $m$ 은 분류할 범주의 수,  $p_k$ 는 데이터가  $k$  범주에 속할 확률이다. 엔트로피와 지니계수는 큰 차이가 없으나, 지니계수는 계산 시 log를 사용하지 않아 속도가 빠르다. 또한 지니계수는 결정나무에서 가장 빈번한 범주로 분리하는 경향이 있지만, 엔트로피의 경우에는 조금 더 균형이 잡힌 결정나무를 생성하도록 하여 성능이 더 좋은 경우가 많다(Provost and Fawcett, 2013). 일반적으로 빠른 분석을 위해 지니계수를 많이 사용하기는 하지만, 이 연구에서는 교량의 안전등급을 좀 더 정확히 예측하기 위해 Scikit-learn의 파이선 라이브러리 중 매개변수들의 조합을 비교하는 GridSearchCV를 이용한 검증을 통하여 엔트로피와 지니계수의 두 알고리즘 중에서 더 뛰어난 성능을 보유한 것을 불순도 계산 기준으로 사용하였다(Scikit-learn developers, 2007-2022).

## 2.2 데이터 전처리

정제되지 않은 데이터에는 결측값이나 이상값이 존재할 수 있고, 불균형한 데이터 분포로 인하여 모델 개발 시 성능이 저하될 수도 있다. 따라서 분류 모델의 성능을 향상시키기 위해서는 결측값 제거, 변수 제거, 변수 축소 및 추가, 불균형 데이터의 샘플링과 같은 데이터 전처리 과정이 필요하다.

### 2.2.1 변수 추가, 제거, 축소 및 다중공선성

전국교량표준데이터에서 안전등급과 무관한 변수(관리기관명, 관리기관전화번호, 데이터기준일자) 및 결측값이 많은 변수(교량보수보강내역, 교량보수보강비용, 하부통과제한높이)는 제거하였다.

의미가 중복되는 변수들은 한 개의 변수로 대표하였는데, 차량통행 하중 및 설계활하중은 설계활하중으로, 시설물종별등급구분 및 점검유형은 시설물종별등급으로 통일하였다. 또한, 변수 중 준공연도와 점검일지는 두 변수의 차이를 계산하여 공용기간이라는 하나의 변수로 축소하였다.

한편, 고ؤ값이 한 개인 변수들은 모델 학습에 영향을 미치지 않으므로 제거하고, 고ؤ값이 과다하게 많은 변수들은 모델의 학습 능력을 저하시키기 때문에 제거하거나 축소해야 한다. 먼저, 고ؤ값이 한 개인 변수에 해당하는 내진설계적용여부와 내진성능확보여부는 모든 교량들이 해당 없음으로 데이터상에 표기되어 있어서 제거하였다. 최근 시공된 일부 교량들에는 내진설계가 적용되었을 것으로 예상되나 표기가 누락된 것으로 추측된다. 또한, 고ؤ값이 많은 변수(소재지도로명, 소재지지번, 시군구명, 시도명, 교량시작점경도, 교량시작점위도, 교량종료점경도, 교량종료점위도, 상부구조형식) 중에서 나머지 변수들은 제거하되, 상부구조형식은 고ؤ값 축소가 가능하여 유지하였다. Hur et al.(2010)은 안전등급 결정요인 중 교량 형식을 사용 재료에 따라 구분하여 평가하였는데, 이를 참고하여 상부구조형식의 고ؤ값을 RC(Reinforced Concrete)교, PSC(PreStressed Concrete)교, 강교, 기타로 축소하였다.

추가적으로 기존 자료에는 존재하지 않지만 교량 등급에 큰 영향을 미칠 것으로 예상되는 데이터인 교량별 평균일교통량(ADT: Average Daily Traffic)과 교량 위치 정보 데이터를 수집하여 변수에 추가하였다. 피로로 인한 교량의 손상은 교통하중 특성의 영향을 받기 때문에(Lee et al., 2010) 이와 연관된 교량별 평균일교통량을 도로 교량 및 터널 현황조사(MOLIT, 2021d)를 참고하여 확보하였다. 그리고, 탄산화 및 염해는 콘크리트와 강재의 내구성 저하현상을 가속화할 가능성이 있기 때문에 교량 위치를 도심지의 탄산화 및 해안가의 염해, 그리고 기타 지역으로 구분하여 변수에 반영하였다. 해안가의 염해의 경우 콘크리트표준시방서 해설(KCI, 2009)을 참고하여 서해안과 남해안의 해안선으로부터 250 m, 동해안의 해안선으로부터 1,000 m 거리 내에 존재하는 지역을 염해의 영향권으로 판단하였다. 또한, 탄산화는 행정구역 중 인구 규모가 큰 시 또는 자치구, 공업단지와 산업단지를 탄산화의 영향권으로 간주하였다.

한편, 전국표준교량데이터상의 보수보강내역 자료는 결측값이 많지만, 시설물통합정보관리시스템(FMS, 2021)의 보수보강내역은 비교적 결측값이 적기 때문에 이를 참고하여 보수보강과 관련된 변수를 추가하였다. 공용기간 동안 실시한 보수보강 활동의 종류가 온전하게 기록되지 않은 교량들이 많아 보수보강 횟수와 2년 이내 보수보강 내역 여부를 확인하여 변수로 추가하였다. 이때 비록 긴급하지는 않지만 결함의 존재로 인해 보수나 보강을 실시해야 하는 B와 C등급 교량의 정밀안전점검 주기가 2년이기 때문에

이 기간을 보수보강 여부의 기준으로 삼았다.

변수들은 상관관계가 거의 없는 독립변수를 사용해야 하며, 변수들 간 상관성이 높은 다중공선성이 발생하는 모델은 성능이 저하될 수 있어 바람직하지 않다(Dormann et al., 2013). 따라서 식 (3)과 같은 피어슨 상관계수(Pearson correlation coefficient)를 구하여 연속형 변수들의 상관관계를 산출하고, 높은 상관성을 가지는 변수들은 제거하였다.

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

$$= \frac{\sum_{i=1}^N (X_i - \mu_X)(Y_i - \mu_Y)}{\sigma_X \sigma_Y N}$$
(3)

여기서,  $\rho_{X,Y}$ 는 변수 간 피어슨 상관계수,  $X, Y$ 는 각 변수의 값,  $Cov(X, Y)$ 는 각 변수 간 공분산,  $\sigma_X, \sigma_Y$ 는 각 변수의 표준편차,  $\mu_X, \mu_Y$ 는 각 변수의 평균,  $N$ 은 변수의 개수이다. 변수 간 피어슨 상관계수를 시각화하면 Fig. 5와 같다. 통상적으로 피어슨 상관계수의 절댓값이 0.3 이상이면 변수 간에 뚜렷한 상관관계가 있다고 간주한다(Ratner, 2009). 따라서 피어슨 상관계수의 절댓값이 0.3 이상인 변수들은 교량폭, 교량연장과 공용기간으로 대표된다고 보고, 이 3가지 변수들을 제외하고는 제거하였다.

이상과 같은 절차를 거쳐 모델 구축에 사용하는 변수는 시설물종별등급구분, 교량연장, 교량폭, 상하행선분리여부, 상부구조형식, 공용기간, 교량별 평균일교통량, 교량위치와 2년 이내 보수보강

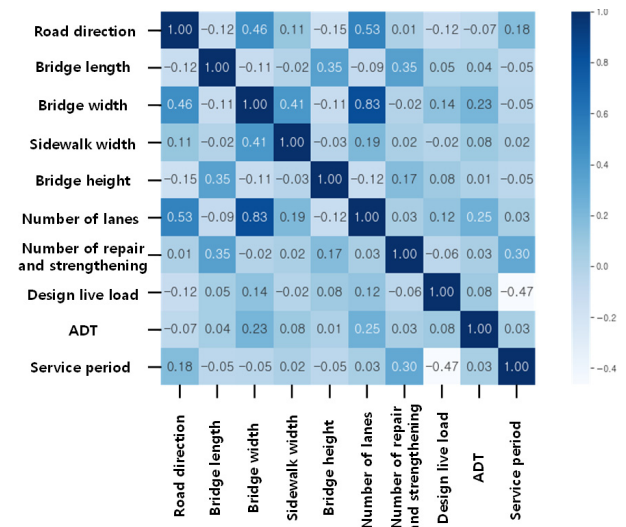


Fig. 5. Pearson Correlation Coefficient of Features

내역 여부와 같이 총 9개가 도출되었다. 따라서 모델의 구축 시 모델에서 예측하고자 하는 목표인 안전등급을 포함하면 총 10개의 데이터 종류를 사용하였다.

2.2.2 데이터 샘플링

머신러닝 모델의 개발을 위해서는 데이터를 훈련 데이터와 테스트 데이터(Test data)로 나눈 후 알고리즘이 훈련 데이터를 학습하여 모델을 구축하고, 테스트 데이터로 모델의 성능을 평가하는 절차를 따른다. 이때, 훈련 데이터와 테스트 데이터는 Fig. 2와 같이 7:3 비율 정도로 분할하는 것이 일반적이며, 이 연구에서도 이러한 비율을 취하였다.

일반국도상 교량 8,850개의 안전등급 분포는 Table 2와 같다. 일반국도 교량의 경우 E등급 교량은 없으며, C와 D등급 교량은 A와 B등급 교량에 비해 그 수가 현저히 적는데, 특히 D등급 교량의 수가 현저히 적다. 해당 데이터는 각 집단의 분포가 균등하지 않은 불균형 데이터이므로 이를 그대로 학습한다면 A와 B등급만 주로 학습하고, C와 D등급은 거의 학습하지 않는 문제가 발생하여 모델의 예측 성능이 저하될 수 있다. 특히 C와 D등급은 유지관리 측면에서 주의를 요하는 교량이므로 C와 D등급 교량을 올바르게

Table 2. Distribution of Safety Grade of Bridges Located in National Roads

Safety grade	Number of bridges	Percentage of bridges(%)
A	2,252	25.4
B	5,775	65.3
C	696	7.9
D	8	0.1
E	0	0
None (No inspection or examination)	119	1.3
Sum	8,850	100

예측할 수 있는 모델을 생성하도록 데이터를 학습시키는 것이 바람직하다. 따라서 이러한 불균형 문제를 개선하기 위해 C와 D등급을 한 집단으로 묶은 후 샘플링 기법을 통해 각 집단(A, B, C+D)의 분포를 균등하게 만들어 등급별 예측 성능을 높이고자 하였다.

이때 샘플링 기법에는 Fig. 6과 같이 언더 샘플링(Under-

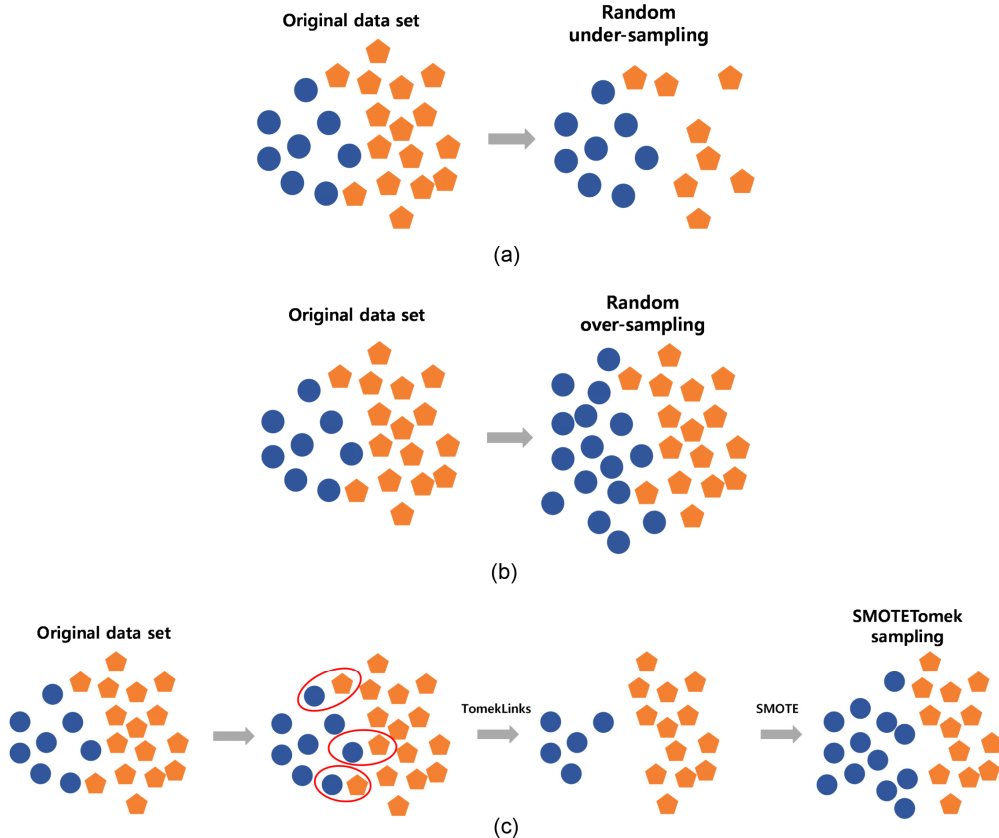


Fig. 6. Data Sampling Technique to Improve Imbalanced Data: (a) Random Under-sampling, (b) Random Over-sampling, (c) SMOTETomek Sampling

sampling), 오버 샘플링(Over-sampling) 및 복합 샘플링(Combined sampling)이 있다. 언더 샘플링이란 다수 집단의 데이터를 임의로 선택하여 소수 집단의 수에 맞도록 데이터 비중을 조절하여 모델링에 활용하는 방법이며, 유의미한 데이터로 축약시킬 수 있지만 중요 정보가 유실될 우려가 있다. 오버 샘플링은 언더 샘플링과는 반대로 소수 집단의 데이터를 다수 집단에 맞추어 증폭시키는 방법으로, 정보의 유실을 방지할 수 있지만 반복되거나 유사한 데이터의 증가로 과적합 현상이 일어나 모델의 성능을 저하시킬 우려가 있다. 한편, 언더 샘플링과 오버 샘플링의 단점을 보완하고자 두 가지 샘플링 기법을 결합한 복합 샘플링 기법도 있다(Lee et al., 2019b). 따라서 이 연구에서는 모델링에 언더 샘플링, 오버 샘플링 및 복합 샘플링 기법을 모두 적용하여 그 결과를 비교해 보았다. 언더 및 오버 샘플링 기법에도 몇 가지 방법이 있는데, 그 중 다수 집단 데이터를 소수 집단 데이터의 수만큼 무작위로 감소시키는 랜덤 언더 샘플링을, 소수 집단 데이터를 다수 집단 데이터의 수만큼 무작위로 증가시키는 랜덤 오버 샘플링을 사용하였다. 복합 샘플링의 경우 언더 샘플링의 TomekLinks 샘플링과 오버 샘플링의 SMOTE 샘플링을 결합한 SMOTETomek 샘플링을 사용하였다. 이는 소수 집단과 다수 집단에 속한 데이터 사이의 거리가 근접한 것을 한 집단으로 보고 이를 노이즈로 간주하여 제거한 후, 소수 집단에 해당하는 데이터를 가상으로 합성하고 증가시켜 데이터 불균형 문제를 해결하는 기법이다.

### 2.3 최적의 매개변수 적용 모델

의사결정나무 및 랜덤포레스트의 라이브러리를 그대로 사용하면 데이터에 적합한 매개변수가 지정되어 있지 않아 모델의 성능이 낮아진다. 이때 매개변수(Parameter)란 모델의 설정값을 의미하며 사용자가 직접 조절 가능한데, 예측 성능을 향상시키기 위해서는 최적의 매개변수를 선정해야 한다(Provost and Fawcett, 2013; Truică and Leordeanu, 2017). 엄밀히 말해 이처럼 사용자가 직접 값을 설정해 줄 수 있는 매개변수는 초매개변수(Hyper parameter, 하이퍼파라미터)로서 일반적인 매개변수와는 구별되어야 하지만, 여기에서는 기존 연구에서 통상적으로 그러하듯 매개변수로 지칭하였다.

이 연구에서는 Scikit-learn의 GridSearchCV를 이용하여 최적의 매개변수를 선정하였다. GridSearchCV는 분류에 사용하는 매개변수들의 순차적인 입력을 통해 지정한 매개변수 범위에서 모든 경우의 조합을 만들어 평가한다. 또한, 학습과 검증을 반복하여 과적합을 방지하고 일반화 성능을 높이는 교차검증을 동시에 진행하기 때문에 지정한 범위에서 가장 우수한 예측 성능을 가지는 매개변수값을 확인할 수 있다(Scikit-learn developers, 2007-2022). 의사결정나무와 랜덤포레스트에서 조정할 매개변수들을

Table 3에 나타내었으며, 여기서 max\_depth, min\_samples\_split 과 min\_samples\_leaf는 의사결정나무가 깊게 생성되는 것을 방지함으로써 과적합의 영향을 낮추어 모델의 성능을 확보하는 매개변수이므로 필수적으로 조정해야 한다. 각 샘플링별로 이 연구에서 도출된 의사결정나무 및 랜덤포레스트 알고리즘의 최적 매개변수는 각각 Table 4 및 5와 같다.

**Table 3.** Hyper Parameters in Decision Tree and Random Forest

Hyper parameter	Characteristics
criterion	The function to measure the quality of a split (entropy, gini)
max_depth	The maximum depth of the tree
min_samples_split	The minimum number of samples required to split an intermediate node
min_samples_leaf	The minimum number of samples required to be at a leaf node
max_features	The number of features to consider when looking for the best split (auto: $\sqrt{\text{all features}}$ , log: $\log_2(\text{all features})$ )
class_weight	Whether to apply the weight of each class (apply: balanced, non-apply: None)
splitter	The strategy used to choose the split at each node (the best method of splitting node for all features: best, the best method of splitting node after randomly extracting the features: random) *Only used for Decision Tree
bootstrap	Whether bootstrap samples are used when building trees *Only used for Random Forest

**Table 4.** Hyper Parameters in Decision Tree Model

Hyper parameter	Sampling		
	Random under-sampling	Random over-sampling	SMOTETomek sampling
criterion [entropy, gini]	gini	entropy	entropy
max_depth [1~10]	5	None	None
min_samples_split [1~15]	12	2	2
min_samples_leaf [1~15]	2	1	1
max_features [auto, log]	auto	auto	auto
class_weight [balanced, None]	balanced	None	None
splitter [best, random]	best	best	best

**Table 5.** Hyper Parameters in Random Forest Model

Hyper parameter	Sampling		
	Random under-sampling	Random over-sampling	SMOTETomek sampling
criterion [entropy, gini]	gini	gini	gini
max_depth [1~10]	9	9	9
min_samples_split [1~15]	4	7	2
min_samples_leaf [1~15]	1	1	1
max_features [auto, log]	auto	auto	auto
class_weight [balanced, None]	balanced	balanced	balanced
bootstrap [True, False]	False	False	False

### 3. 교량의 안전등급 예측 모델 성능 평가

이 연구에서 예측하고자 하는 안전등급의 범주는 A등급, B등급 및 C, D등급의 총 3가지로 구분되므로 다중 분류 모델에 해당된다. 해당 모델에서 테스트 데이터는 훈련 데이터와는 달리 샘플링하지 않기 때문에 범주별 분포가 불균형하다. 따라서 일반적인 정확도 (Accuracy)는 각 범주의 분포가 고려되지 않고 평균화되어 계산되므로 소수 범주와 다수 범주가 혼재하는 분류에서는 모델 성능의 적절한 평가 지표로 볼 수 없으며, 정확도 대신 다른 평가 지표가 사용되어야 한다(He and Garcia, 2009). 이 연구에서는 다수 범주인 A와 B등급뿐만 아니라 소수 범주인 C, D등급의 예측력도 우수한 모델을 선정하기 위해 여러 가지 평가 지표로서 혼동행렬 (Confusion matrix), 균형 정확도(Balanced accuracy), 재현율 (Recall), ROC 곡선(Receiver Operating Characteristic curve) 및 AUC(Area Under the Curve)를 사용하여 예측 성능을 평가하였다.

#### 3.1 혼동행렬

혼동행렬은 실제값과 예측값의 일치 여부를 보여주며, 이는 교량 등급 분류 모델의 예측 성능을 평가하는 기본 지표가 된다. Fig. 7에서 대상의 실제값과 예측값이 분류하고자 하는 특정 범주로 일치하면 TP(True Positive)이고, 실제값과 예측값이 모두 해당 범주가 아니면 TN(True Negative)이다. 또한 분류하고자 하는 범주에서 실제값은 특정 범주이나 예측값은 그 외 범주이면 FN(False Negative), 실제값은 특정 범주가 아니지만 예측값은

		Predicted class	
		Positive	Negative
Actual class	Positive	<b>TP</b> (True Positive)	<b>FN</b> (False Negative)
	Negative	<b>FP</b> (False Positive)	<b>TN</b> (True Negative)

**Fig. 7.** Confusion Matrix

그러한 특정 범주이면 FP(False Positive)를 의미한다. 즉, TP와 TN은 예측이 실제와 일치하는 정답이고, FP와 FN은 예측이 실제와 다른 오답이다. 특히, 실제로 C, D등급인 교량을 A와 B등급으로 상향하여 예측하는 경우인 FN은 교량의 유지관리가 비교적 시급한 C, D등급을 추출하지 못하여 구조 안전성에 큰 문제를 야기할 우려가 있으므로 그 개수를 최소화시키는 것이 바람직하다.

의사결정나무와 랜덤포레스트의 혼동행렬을 살펴보면 Fig. 8과 같이 정답인 TP와 TN의 개수는 모든 등급의 예측에 있어 랜덤포레스트가 의사결정나무보다 더 많은 것을 확인하였다. 또한, 샘플링 기법 중 랜덤 언더 샘플링을 적용한 모델이 의사결정나무와 랜덤포레스트 모두에서 실제 C, D등급인 교량을 A등급 또는 B등급으로 상향하여 예측하는 경우인 FN의 개수가 가장 적었다. 유지관리 측면에서는 C, D등급 교량의 예측 성능이 가장 중요하므로 의사결정나무와 랜덤포레스트 모두 랜덤 언더 샘플링 기법이 효과적이라고 판단된다. 추가적으로 C, D등급 교량의 예측에서 랜덤 오버 샘플링과 SMOTETomek 샘플링은 의사결정나무의 경우 TP보다 FN이 더 많아서 부적합했지만, 랜덤포레스트에서는 비록 랜덤 언더 샘플링의 수준에는 못 미쳤지만 FN보다 TP가 더 많은 것으로 나타났다. 한편, A등급의 예측력은 샘플링 기법에 관계없이 랜덤포레스트가 대체로 우수했고, B등급의 예측력은 분류 모델이나 샘플링 기법에 따라 차이가 있어 일률적으로 언급하기 어려웠다.

이처럼 혼동행렬을 통하여 의사결정나무와 랜덤포레스트 중 성능이 뛰어난 알고리즘을 판별할 수 있으나, 샘플링별 모델의 성능을 수치로 표현된 정량적인 기준으로 비교하기는 어렵다. 따라서 모델 예측 성능을 좀 더 명확하게 정량적으로 판단하기 위해 3.2절에서는 수치나 그래프로 표현될 수 있는 추가적인 지표들을 이용하여 분석하였다.

#### 3.2 정확도, C, D등급 재현율, ROC 곡선 및 AUC

일반적인 정확도는 전체에서 정답의 비율을 식 (4)로 구하며, 테스트 데이터의 범주별 분포가 균등하지 않으면 예측력을 과대 또는 과소평가할 우려가 있다. 따라서 이러한 정확도는 이 연구에서와



Random under-sampling			Random over-sampling			SMOTETomek sampling					
A grade	Predicted class		A grade	Predicted class		A grade	Predicted class				
	P	N		P	N		P	N			
Actual class	P	450	226	Actual class	P	399	277	Actual class	P	404	272
	N	506	1,438		N	270	1,674		N	329	1,615
B grade	Predicted class		B grade	Predicted class		B grade	Predicted class				
	P	N		P	N		P	N			
Actual class	P	661	1,702	Actual class	P	1,369	364	Actual class	P	1,201	532
	N	103	784		N	379	508		N	319	568
C, D grade	Predicted class		C, D grade	Predicted class		C, D grade	Predicted class				
	P	N		P	N		P	N			
Actual class	P	166	45	Actual class	P	66	145	Actual class	P	87	124
	N	734	1,675		N	137	2,272		N	280	2,129

(a)

Random under-sampling			Random over-sampling			SMOTETomek sampling					
A grade	Predicted class		A grade	Predicted class		A grade	Predicted class				
	P	N		P	N		P	N			
Actual class	P	513	163	Actual class	P	533	143	Actual class	P	538	138
	N	505	1,439		N	546	1,398		N	520	1,424
B grade	Predicted class		B grade	Predicted class		B grade	Predicted class				
	P	N		P	N		P	N			
Actual class	P	723	1,010	Actual class	P	873	860	Actual class	P	873	860
	N	99	788		N	137	750		N	118	769
C, D grade	Predicted class		C, D grade	Predicted class		C, D grade	Predicted class				
	P	N		P	N		P	N			
Actual class	P	176	35	Actual class	P	134	77	Actual class	P	150	61
	N	604	1,805		N	397	2,012		N	421	1,988

(b)

Fig. 8. Results of Confusion Matrix: (a) Decision Tree, (b) Random Forest

같이 테스트 데이터가 불균형한 경우에는 평가 지표로서 적절치 않다. 불균형한 테스트 데이터의 경우 식 (5)와 같이 재현율과 특이도의 산술평균으로 계산되는 균형 정확도를 평가 지표로 사용하는 것이 바람직하다. 이는 각 범주의 개수 차이의 영향을 줄임으로써

개수가 작은 범주의 예측 성능이 왜곡되지 않도록 하는 장점이 있다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Balanced accuracy} = \frac{1}{2} \left( \frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (5)$$

정밀도(Precision)는 식 (6)과 같이 특정 범주로 예측한 것 중에서 실제로 특정 범주에 존재하는 비율을 나타낸 것이며, 식 (7)과 같은 재현율은 실제 특정 범주에 속하는 것 중 예측이 맞는 비율을 의미한다(He and Garcia, 2009). 한편 F1-score는 식 (8)과 같이 정밀도와 재현율의 조화평균으로서 다중 분류 모델의 대표적인 평가 지표 중 하나이다(Grandini et al., 2020). 보통 재현율과 정밀도는 상대적인 관계를 가지고 있어 모든 범주의 예측 성능이 뛰어나지 않은 이상 재현율이 감소하면 정밀도가 상승하고, 정밀도가 감소하면 재현율이 상승하는 경향이 있다. F1-score는 조화평균의 특성상 재현율과 정밀도 중 하나가 저하되면 낮은 수치를 산출하게 된다. 따라서 모델의 목적에 따라 정밀도, 재현율과 F1-score 중 더 중요시되는 평가 지표를 선정해야 한다. 안전등급 예측 시 C, D등급 교량이 유지관리의 중점이 되므로 이를 올바르게 예측하는 것이 무엇보다 중요하다. 따라서 정밀도, 재현율, F1-score 중 C, D등급의 재현율을 주로 고려하여 예측 성능을 평가하였다.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{F1-score} = 2 \left( \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (8)$$

한편, ROC 곡선은 분류에서 비용에 해당되는 FPR(False Positive Rate)과 이득에 해당되는 TPR(True Positive Rate)의 비율을 시각적으로 나타내어 모델의 예측 성능을 쉽게 판단할 수 있도록 한 것이다. FPR은 실제 특정 범주에 존재하지 않는 값을 특정 범주에 존재한다고 예측하는 비율로 비용에 해당하며, TPR은 실제 특정 범주에 존재하는 값을 올바르게 예측하는 비율로 이득에 해당한다. ROC 곡선은 (0, 1)에 가까울수록 이득이 많이 발생해도 손실 발생이 적다는 것을 의미하므로 분류가 완벽에 가까움을 나타낸다(He and Garcia, 2009). AUC는 ROC 곡선의 아래쪽 면적을 나타낸 값으로 ROC 곡선의 성능을 수치로 비교할 때 유용하다. AUC는 1에 가까울수록 성능이 좋다는 것을 의미하고, 통상적으로 0.8 이상이면 성능이 뛰어난 분류기, 0.7 이상이면 성능이 준수한 분류기로 간주하며, 0.5 이하이면 분류의 의미가 없다는 것을 의미한다(Hosmer and Lemeshow, 2000). ROC 곡선 및 AUC는 분류 모델의 대표적인 평가 지표이며, 개별 범주에 따른 비용 및 손실을 계산하기 때문에 다른 범주의 영향을 적게

**Table 6.** Evaluation of Predictive Performance in Decision Tree

Evaluation index	Sampling		
	Random under-sampling	Random over-sampling	SMOTETomek sampling
Balanced accuracy(%)	61.1	56.4	56.8
Recall of C, D grade(%)	78.7	31.3	41.2
AUC	0.763	0.675	0.676

**Table 7.** Evaluation of Predictive Performance in Random Forest

Evaluation index	Sampling		
	Random under-sampling	Random over-sampling	SMOTETomek sampling
Balanced accuracy(%)	67.0	64.7	67.0
Recall of C, D grade(%)	83.4	63.5	71.1
AUC	0.823	0.834	0.834

받으므로 불균형 데이터의 분석에 적합하다고 판단하여 평가 지표로 사용하였다.

분석 결과 의사결정나무와 랜덤포레스트 모델의 균형 정확도, C, D등급 재현율, AUC를 Table 6 및 7에 나타내었고, ROC 곡선은 Fig. 9와 같다. 여기서 균형 정확도 및 AUC는 모든 등급에 대한 평균값이다. 의사결정나무의 경우 랜덤 언더 샘플링이 모든 측면에서 랜덤 오버 샘플링 및 SMOTETomek 샘플링보다 성능이 우수하였다. 특히, 랜덤 언더 샘플링은 AUC가 0.7 이상으로 성능이 준수하였고, 무엇보다 C, D등급의 재현율이 78.7%로 다른 두 가지 샘플링 기법보다 월등히 우수하였다. 랜덤포레스트의 경우 모든 샘플링에서 균형 정확도가 64~67% 수준으로 의사결정나무보다 좋은 성능을 보유했으며, AUC 또한 0.8 이상으로 성능이 뛰어난 분류기로 판단할 수 있었다. 다만 C, D등급의 재현율은 랜덤 언더 샘플링이 83.4%로 다른 두 샘플링보다 우수하였다. 이는 교량의 안전등급 판정과 관련된 기존의 이진 분류 모델 연구 (Chung et al., 2016)에서 C, D등급의 재현율이 67.3%였던 것과 비교할 때 재현율을 16.1%p 향상시킨 것으로서, 이 연구에서 적용한 분류 기법의 우수성을 보여주고 있다.

ROC 곡선의 경우 의사결정나무의 랜덤 언더 샘플링에서는 A등급 및 C, D등급의 예측 성능이 B등급보다 상대적으로 우수하게 나타났다. 하지만, 랜덤 오버 샘플링 및 SMOTETomek 샘플링은 등급에 관계 없이 곡선이 (0, 1)에서 비교적 멀리 떨어진 형상이기 때문에 성능이 낮았으며, 이러한 경향은 Table 6의 상대적으로 작은 AUC값에도 반영되어 있다. 반면 랜덤포레스트에서는 모든 샘플링 기법에서 ROC 곡선의 형태가 의사결정나무보다 우수한

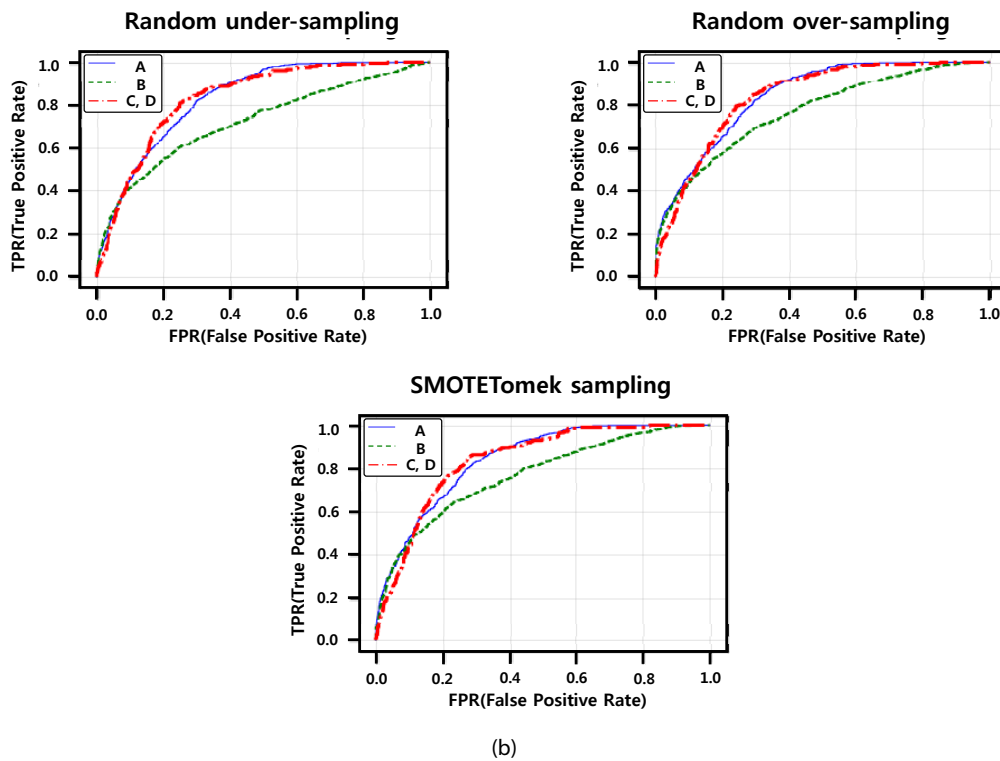
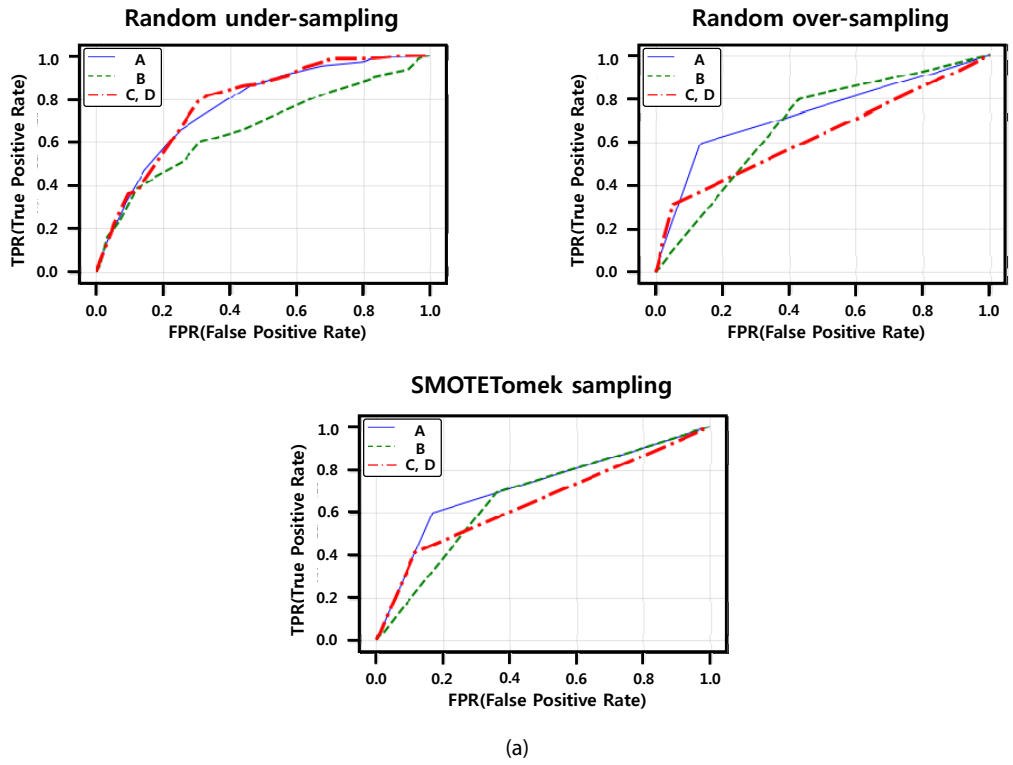


Fig. 9. Comparison of ROC Curves: (a) Decision Tree, (b) Random Forest

것으로 나타났으며, 이는 Table 7의 AUC값에서도 확인할 수 있다. 다만 B등급의 예측 성능이 A등급 및 C, D등급보다 떨어지는 경향은 의사결정나무의 랜덤 언더 샘플링과 유사하게 나타났다.

이러한 결과들을 바탕으로 의사결정나무와 랜덤포레스트의 성능을 전반적으로 비교하자면 균형 정확도, C, D등급 재현율, AUC, ROC 곡선 모두 랜덤포레스트가 더 우수한 것으로 나타났다. 특히, 랜덤포레스트의 샘플링 기법 중 C, D등급의 재현율이 월등히 뛰어난 랜덤 언더 샘플링이 부각되었다. 이상과 같이 교량의 안전등급 예측에 두 가지 분류 모델을 적용하여 분석한 결과 랜덤포레스트 모델의 랜덤 언더 샘플링이 C, D등급을 추출하는 예측력이 우수하여 권장될 수 있다.

### 3.3 활용 방안

이상과 같이 일반국도상 교량 8,850개에 대해서는 Table 4 및 5의 매개변수를 가지고 랜덤포레스트 모델의 랜덤 언더 샘플링 기법을 적용하는 것이 효과적인 것으로 나타났다. 데이터의 분포나 성질에 따라 최적의 분류 모델이나 샘플링 기법은 달라질 수 있으나, 일반국도 외에 고속국도나 지방도상의 교량들도 유사한 데이터 분포 및 성질을 가지고 있다고 가정한다면 동일한 기법을 적용하여 분석 가능하다.

제안된 기법을 활용하여 점검이 실시되지 못했거나 점검 시기가 초과된 교량의 성능을 예측할 때는 각종 데이터와 더불어 안전등급을 확인하고자 하는 연도를 입력하면 예상되는 교량 안전등급을 산출하게 된다. 특히, 제안 기법은 노후도가 큰 C, D등급 교량의 예측력이 우수하므로 교량의 적정 보수보강 시기 추정 및 유지관리 예산 산출에 유용하게 활용될 수 있을 것으로 기대된다.

## 4. 교량 안전등급의 영향 요인

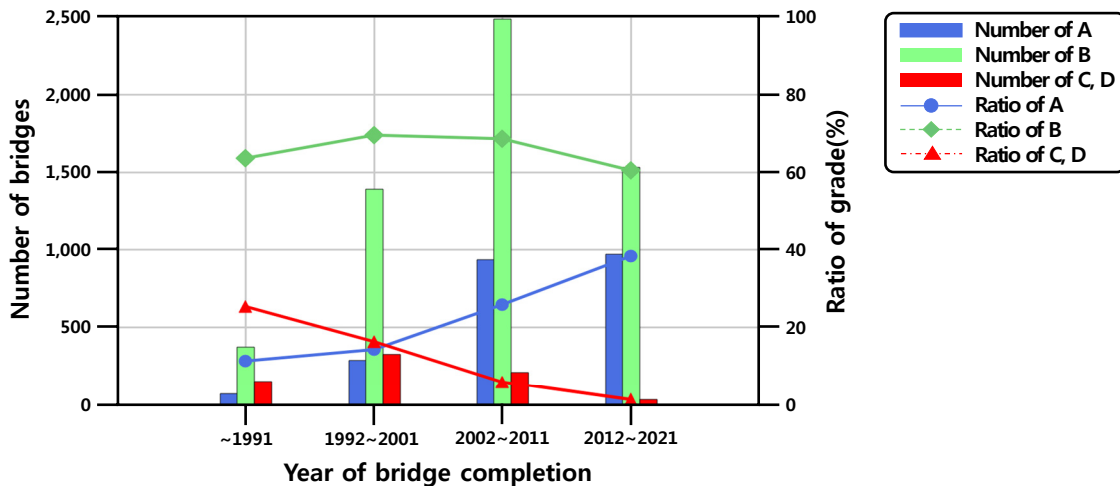
결정나무 기반 알고리즘으로 분류 모델들을 생성하였기 때문에 불순도가 감소하는 정도를 계산하여 모델을 생성하는 과정에서 변수들이 분류 성능에 영향을 미치는 정도인 변수 중요도(Variable importance)를 확인할 수 있다. 또한, 다른 알고리즘에도 사용하는 순열 변수 중요도(Permutation feature importance)도 계산할 수 있다. 순열 변수 중요도는 학습이 끝난 모델에서 변수를 하나씩 제거해가며 예측 성능을 많이 저하시키는 변수를 선정함으로써 파악할 수 있다(Scikit-learn developers, 2007-2022).

이 연구에서는 변수 중요도와 순열 변수 중요도를 사용하여 모델에 영향을 미치는 주요 요인을 도출하고 실제 경향과 비교하여 주요 변수와 안전등급의 관련성을 분석하였다. 이 연구에서 최적으로 나타난 랜덤 언더 샘플링을 적용한 랜덤포레스트 모델에서 변수 중요도와 순열 변수 중요도는 Table 8과 같으며, 공통적으로 나타난 중요 변수는 공용기간, 교량연장, 교통량, 시설물종별등급구분이다.

이하에서는 Table 8에 제시된 중요 변수들 중 대표적으로 공용기

**Table 8.** Variable Importance and Permutation Feature Importance of the Random Forest Using Random Under-sampling

Rank	Variable importance	Permutation feature importance
1	Service period	Service period
2	Bridge length	Bridge length
3	Average daily traffic	Average daily traffic
4	Facility class	Facility class
5	Bridge width	Separation of northbound and southbound lanes



**Fig. 10.** Safety Grade According to Completion Year of Bridges

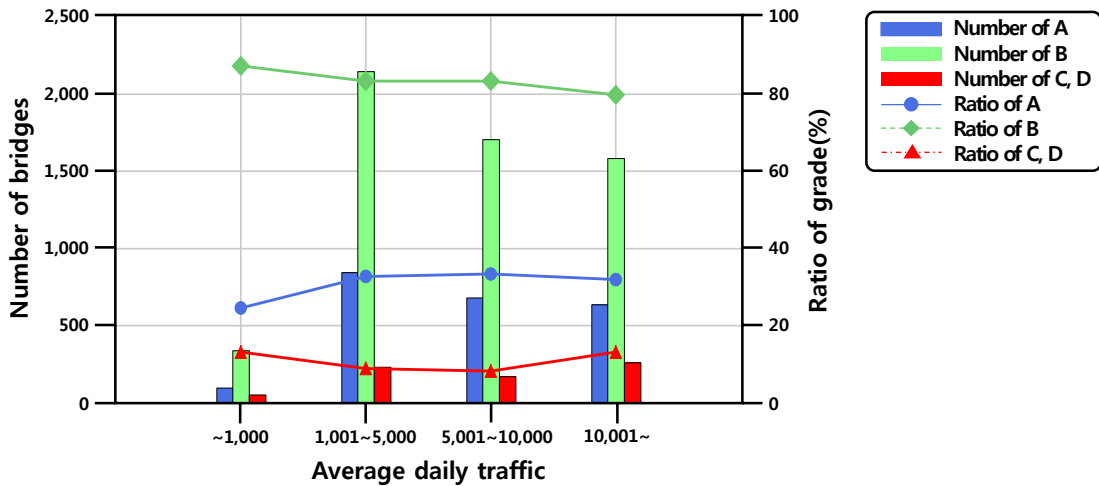


Fig. 11. Safety Grade According to Average Daily Traffic

간과 교통량이 교량 안전등급과 관련성이 큰 것으로 나타난 이유를 실제 상황에 비추어 분석하였다.

공용기간은 교량의 노후화와 직접적으로 관련된 시간적 요인으로 교량의 안전등급에 큰 영향을 준다. 교량 준공 후 시간이 많이 경과될수록 결함이 발생하고, 지속적인 사용으로 인하여 내구성과 안전성이 저하되기 때문에 교량의 노후화가 진행된다. Fig. 10에서 C, D등급 교량의 비율은 1991년 이전에 준공된 교량에서 25.6%로 가장 높았으며, 최근 10년 사이에 준공된 교량에서는 그 비율이 1.2%로 가장 낮았다. 또한, 최근 준공되어 공용기간이 짧은 교량일수록 노후화가 많이 진행되지 않아 결함이 적어서 A등급 교량의 비율이 증가하였다. 이처럼 A와 C, D등급의 교량은 공용기간에 따른 경향이 뚜렷하였다. 반면 B등급 교량의 경우 공용기간과의 관련성이 명확하지 않아 분류 모델에서 Fig. 9와 같이 B등급의 예측력이 저하된 일부 원인이 되었을 것으로 예상된다.

교통량은 차량하중에 의한 교량의 피로 현상과 관련하여 안전성 및 내구성에 영향을 미친다. 특히 사항으로는 Fig. 11과 같이 교통량이 하루 1,000대 이하이거나 10,000대 이상이면 C, D등급 교량의 비율이 다소 커지고 A등급 교량의 비율이 다소 작아지는 경향이 있었다. 특히, 교통량이 10,000대 이상으로 많은 경우 반복되는 차량하중으로 인해 피로가 누적되어 균열이나 처짐 등 교량의 사용성이 저하되고 노면의 파손 등과 같은 결함도 발생할 가능성이 크기 때문에 그러한 결과가 나타난 것으로 보인다. 이처럼 교통량이 매우 많거나 적은 경우 안전등급의 저하 경향이 나타났다.

### 5. 결론

이 연구에서는 교량의 안전등급 예측을 위해 머신러닝 알고리즘

인 의사결정나무와 랜덤포레스트를 이용하여 다중 분류 모델을 개발하였다. 수집된 교량 데이터에서 변수 추가, 제거, 축소 및 다중공선성 검토 과정을 거쳐 모델 구축에 필요한 최적의 변수들을 도출하였다. 개발된 다중 분류 모델의 예측 성능 평가 시 일반적인 평가 지표가 아닌 범주 간 불균형 데이터에 적합한 평가 지표들을 통해 모델이 준수한 예측 성능을 보유하고 있는지 판단하였다. 또한 훈련 데이터의 불균형 문제를 개선시키기 위해 랜덤 언더 샘플링, 랜덤 오버 샘플링, SMOTETomek 샘플링 기법을 각각 적용하여 결과를 비교하였다. 이 연구에서 도출된 주요 결론은 다음과 같다.

- (1) 결정나무 기반 알고리즘을 이용하여 교량의 안전등급 분류 모델을 개발하였기 때문에 모델 형성 과정에서 결과에 중요한 영향을 미치는 변수들을 파악할 수 있었다. 의사결정나무와 랜덤포레스트에서 공통적으로 공용기간, 교량연장, 교통량 및 시설물종별등급구분이 주요 인자로 확인되었다. 특히 교량의 노후화와 직접적으로 관련된 공용기간은 모델에서 가장 큰 영향을 미치는 것으로 나타났다. 이와 같이 변수 중요도를 통해 확인한 주요 변수들은 실제 예상되는 교량의 거동으로부터 분석해 보아도 역시 안전등급에 큰 영향을 미칠 것으로 예상되었다.
- (2) 일반적인 정확도로 다중 분류 모델의 성능을 평가하는 것은 왜곡된 결과를 산출할 수 있기 때문에 다중 분류 모델의 성능 평가에 적합한 혼동행렬 기반의 균형 정확도, 재현율, ROC 곡선 및 AUC와 같은 다양한 지표들을 활용하여 모델의 성능을 평가하였다. 그 결과 의사결정나무보다 랜덤포레스트를 적용한 다중 분류 모델이 전반적인 성능 평가 지표 측면에서 우수한 성능을 나타냈다. 따라서 랜덤포레스트가 의사결정나무보다 교량의 안전등급 예측에 더욱 적합한 알고리즘으로 판단되었다. 또한, 샘플링 기법에서는 두 가지 분류 모델 모두에서 랜덤

언더 샘플링이 대체로 우수한 예측력을 보였으며, 특히 노후화가 비교적 심하여 유지관리 측면에서 중요한 C, D등급의 재현율이 월등히 뛰어났다. 결론적으로 교량의 안전등급 예측을 위해서는 랜덤 언더 샘플링 기법이 적용된 랜덤포레스트 모델을 사용하는 것이 바람직하다고 판단된다.

- (3) C, D등급 교량을 C, D등급 그대로 분류하여 예측할 확률인 재현율은 랜덤 언더 샘플링 기법이 적용된 랜덤포레스트 모델에서 83.4%로, 기존의 이진 분류 모델에서의 67.3%보다 16.1%p 향상된 결과를 얻을 수 있었다. 이는 이 연구에 적용한 다중 분류 모델이 두 가지 분류만 가능한 이진 분류 모델과 비교할 때 더욱 다양한 교량 안전등급을 고려할 수 있다는 장점과 더불어 중요 등급 교량의 예측력이 우수함을 보여주고 있다.
- (4) 제안된 기법을 점검이 실시되지 못했거나 점검 시기가 초과된 교량의 데이터에 적용하면 현재 또는 특정 시기의 교량 안전등급을 예측할 수 있다. 특히, 제안된 기법은 C, D등급 교량의 예측력이 우수하므로 교량의 적절한 보수보강 시기 추정 및 유지관리 예산 산출에 유용하게 활용될 수 있을 것으로 기대된다. 이 연구는 일반국도상 교량의 분석에 집중했지만, 추후 고속국도나 지방도상 교량의 안전등급 분석에도 확장하여 적용할 수 있을 것으로 판단된다.

## References

- Bektas, B. A., Carriquiry, A. and Smadi, O. (2013). "Using classification trees for predicting national bridge inventory condition ratings." *Journal of Infrastructure Systems*, Vol. 19, No. 4, pp. 425-433, [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000143](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000143).
- Chung, S. H., Lim, S. R. and Chi, S. H. (2016). "Developing an estimation model for safety rating of road bridges using rule-based classification method." *Journal of KIBIM*, Vol. 6, No. 2, pp. 29-38, <https://doi.org/10.13161/kibim.2016.6.2.029> (in Korean).
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D. and Lautenbach, S. (2013). "Collinearity: A review of methods to deal with it and a simulation study evaluating their performance." *Ecography*, Vol. 36, No. 1, pp. 27-46, <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Facility Management System(FMS) (2021). <https://www.fms.or.kr/com/mainForm.do> (Accessed: September 12, 2021).
- Géron, A. (2019). *Hands-on machine learning with Scikit-learn, Keras, and TensorFlow*, 2nd Ed., O'Reilly Media, Inc.
- Grandini, M., Bagli, E. and Visani, G. (2020). *Metrics for multi-class classification: An overview*, arXiv:2008.05756, Available at: <https://arxiv.org/abs/2008.05756> (Accessed: July 25, 2022).
- He, H. and Garcia, E. A. (2009). "Learning from imbalanced data." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, No. 9, pp. 1263-1284, <https://doi.org/10.1109/TKDE.2008.239>.
- Hosmer, D. W. and Lemeshow, S. (2000). *Applied logistic regression*, 2nd Ed., John Wiley & Sons, Inc.
- Hur, Y. K., Lee, H. I., Shin, J. Y. and Park, C. H. (2010). "A research for the determinant factors of safety ratings in road-bridge." *Journal of the Korea Institute for Structural Maintenance and Inspection*, Vol. 14, No. 6, pp. 229-237, <https://doi.org/10.11112/jksmi.2010.14.6.229> (in Korean).
- Kang, B. H. (2016). "Suggestions for improving the competitiveness of safety inspection industry." *KSCE Magazine*, Vol. 64, No. 10, pp. 12-15 (in Korean).
- Kang, S. H., Choi, S. I., Kim, H. R. and Lee, J. S. (2016). "A study on performance evaluation of infrastructure safety and maintenance." *Korean Journal of Construction Engineering and Management*, Vol. 17, No. 2, pp. 80-89, <https://doi.org/10.6106/KJCEM.2016.17.2.080> (in Korean).
- Kazemitabar, S. J., Amini, A. A., Bloniarz, A. and Talwalker, A. (2017). "Variable importance using decision trees." *Proceedings of the 31st International Conference on Neural Information Processing Systems(NIPS 2017)*, pp. 425-434.
- Kim, S. J. and Yoon, M. O. (2018). "A study on the improvement program of bridge safety management through public-private governance." *Journal of the Korean Society of Hazard Mitigation*, Vol. 18, No. 1, pp. 145-156, <http://doi.org/10.9798/KOSHAM.2018.18.1.145> (in Korean).
- Korea Concrete Institute(KCI) (2009). Standard specification for concrete, KCI (in Korean).
- Lee, H. H., Kyung, K. S. and Jeon, J. C. (2010). "Fatigue life estimation method considering traffic properties for steel highway girder bridge." *Journal of Korean Society of Steel Construction*, Vol. 22, No. 3, pp. 209-218 (in Korean).
- Lee, H. H., Shin, B. G., Lee, Y. I. and Kim, Y. M. (2019a). "Suggestion of priority decision method for performance evaluation based on risk index for small and medium sized bridges." *Journal of the Korea Institute for Structural Maintenance and Inspection*, Vol. 23, No. 6, pp. 70-76, <https://doi.org/10.11112/jksmi.2019.23.6.70> (in Korean).
- Lee, I. K. and Kim, D. H. (2015). "Highway bridge inspection period based on risk assessment." *Journal of the Korea Institute for Structural Maintenance and Inspection*, Vol. 19, No. 3, pp. 64-72, <https://doi.org/10.11112/jksmi.2015.19.3.064> (in Korean).
- Lee, J. H., Lee, K. Y., Ahn, S. M. and Kong, J. S. (2018). "Proposal of maintenance scenario of feasibility analysis of bridge inspection using Bayesian approach." *Journal of the Korean Society of Civil Engineers*, Vol. 38, No. 4, pp. 505-516, <http://doi.org/10.12652/Ksce.2018.38.4.0505> (in Korean).
- Lee, K. N., Lim, J. T., Bok, K. S. and Yoo, J. S. (2019b). "Handling method of imbalance data for machine learning: Focused on sampling." *The Journal of the Korea Contents Association*, Vol. 19, No. 11, pp. 567-577, <https://doi.org/10.5392/JKCA.2019.19.11.567> (in Korean).

- Martinez, P., Mohamed, E., Mohsen, O. and Mohamed, Y. (2020). "Comparative study of data mining models for prediction of bridge future conditions." *Journal of Performance of Constructed Facilities*, Vol. 34, No. 1, 04019108, [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0001395](https://doi.org/10.1061/(ASCE)CF.1943-5509.0001395).
- Ministry of Land, Infrastructure and Transport(MOLIT) (2021a). *Guidelines for safety and maintenance of facilities*, MOLIT (in Korean).
- Ministry of Land, Infrastructure and Transport(MOLIT) (2021b). *National bridge standard data*, MOLIT, Available at: <https://www.data.go.kr/data/15081953/fileData.do> (Accessed: September 12, 2021) (in Korean).
- Ministry of Land, Infrastructure and Transport(MOLIT) (2021c). *Special act on the safety control and maintenance of establishments*, MOLIT (in Korean).
- Ministry of Land, Infrastructure and Transport(MOLIT) (2021d). *Yearbook of road bridge and tunnel statistics*, MOLIT, Available at: <https://bti.kict.re.kr/bti/publicMain/main.do> (Accessed: March 8, 2022) (in Korean).
- Nguyen, T. T. and Dinh, K. (2019). "Prediction of bridge deck condition rating based on artificial neural networks." *Journal of Science and Technology in Civil Engineering*, Vol. 13, No. 3, pp. 15-25, [https://doi.org/10.31814/stce.nuce2019-13\(3\)-02](https://doi.org/10.31814/stce.nuce2019-13(3)-02).
- Oh, S. T., Lee, D. J. and Lee, J. H. (2010). "A condition rating method of bridges using an artificial neural network model." *Journal of the Korean Society for Railway*, Vol. 13, No. 1, pp. 71-77 (in Korean).
- Provost, F. and Fawcett, T. (2013). *Data science of business: What you need to know about data mining and data-analytic thinking*, O'Reilly Media, Inc.
- Ratner, B. (2009). "The correlation coefficient: Its values range between +1/-1, or do they?" *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 17, pp. 139-142, <https://doi.org/10.1057/jt.2009.5>.
- Scikit-learn developers (2007-2022). <https://scikit-learn.org/>.
- Trucă, C. O. and Leordeanu, C. A. (2017). "Classification of an imbalanced data set using decision tree algorithms." *U.P.B. Scientific Bulletin, Series C: Electrical Engineering and Computer Science*, Vol. 79, Iss. 4, pp. 69-84.
- Yonhapnews (2023). *Most of the bridges in the first new town have passed more than 30 years...Citizens are worried about deterioration.* <https://www.yna.co.kr/view/AKR20230406092900061> (Accessed: April 12, 2023) (in Korean).