

# 기계학습 기반 해양 노출 환경의 콘크리트 교량 데이터를 활용한 염화물 확산계수 예측모델 개발

남우석<sup>1</sup>, 임홍재<sup>2\*</sup>

## Development of a Machine Learning-Based Model for the Prediction of Chloride Diffusion Coefficient Using Concrete Bridge Data Exposed to Marine Environments

Woo-Suk Nam<sup>1</sup>, Hong-Jae Yim<sup>2\*</sup>

**Abstract:** The chloride diffusion coefficient is a critical indicator for assessing the durability of concrete marine substructures. This study develops a prediction model for the chloride diffusion coefficient using data from concrete bridges located in marine exposure zones (atmospheric, splash, tidal), an aspect that has not been considered in previous studies. Chloride profile data obtained from these bridge substructures were utilized. After data preprocessing, machine learning models, including Random Forest (RF), Gradient Boosting Machine (GBM), and K-Nearest Neighbors (KNN), were optimized through hyperparameter tuning. The performance of these models was developed and compared under three different variable sets. The first model uses six variables: water-to-binder (W/B) ratio, cement type, coarse aggregate volume ratio, service life, strength, and exposure environment. The second model excludes the exposure environment, using only the remaining five variables. The third model relies on just three variables: service life, strength, and exposure environment factors that can be obtained from precision safety diagnostics. The results indicate that including the exposure environment significantly enhances model performance for predicting the chloride diffusion coefficient in concrete bridges in marine environments. Additionally, the three variable model demonstrates that effective predictions can be made using only data from precision safety diagnostics.

**Keywords:** Machine learning, Chloride diffusion coefficient, Marine exposure environment, Concrete bridge, Precision safety diagnosis data

### 1. 서론

염화물 확산은 해양 환경에 노출된 철근 콘크리트 구조물의 내구성을 저하시키는 주요 원인 중 하나이다. 염화물 이온이 콘크리트 내부로 침투하여 임계염화물량 초과 및 철근에 도달하면 부동태 피막이 파괴되어 부식을 유발한다. 이는 철근의 부피 팽창으로 인해 콘크리트에 균열 등의 손상을 초래하고, 결과적으로 구조적 손상에 의해 구조물의 성능 저하를 야기한다 (Metha et al., 2009). 이러한 염해는 해양 환경의 콘크리트 구조물에서 빈번히 발생하며, 이는 서비스 수명을 단축시키고 유지 보수 비용을 증가시킨다(Metha and Monteiro., 1993). 따라서 해양 환경의 철근 콘크리트 구조물의 내구성을 유지 및 향상시키기 위해서는 염화물 확산을 정확히 예측하는 것이 중요하다.

하지만 콘크리트의 불균질한 구조와 복합적 이온 전달 메커니즘(확산, 침투, 흡착)으로 인해 염화물 이온의 이동 현상을 추정하는 것은 어려움이 있다(Liu and Weyers, 1998).

염화물 확산계수는 콘크리트 구조물 내부에서 염화물 이온이 확산되는 속도에 대한 지표이다. 염화물 확산의 메커니즘은 Fick의 제2법칙을 기반으로 하며 이는 시간에 따른 염화물 농도 변화를 고려한 비정상상태(non-steady state)의 염화물 이동을 수학적으로 모델링 한 것이다.

염화물 확산계수는 다양한 방법론을 통해 발전해왔으며, 주로 실험실 테스트를 통한 Fick의 제2법칙을 기반으로 확산계수 모델이 제안되어 왔다. Bamforth(1999)는 실내실험 환경에서 물시멘트비 및 혼화재(플라이 애쉬, 고로슬래그, 실리카 폼)의 영향에 따른 모델을 제안하였다. Tang(1998)은 전기적으로 가속된 방법을 통해 콘크리트의 염화물 확산성을 평가하는 방법을 개발하였으며, 이는 기존의 자연 확산 방법보다 빠르게 결과를 도출할 수 있는 실용적 접근법이다. Thomas et al.(2002)는 시간의존성 확산계수를 포함한 지배방정식을 제안하였으나 이 모델은 정상상태 조건에서만 적용하도록 권고하고 있다. 추가로 Bentz(2003)는 불확실성을 고려하기 위해

<sup>1</sup>정회원, 부산대학교 사회환경시스템공학과 박사수료, 국토안전관리원 과정

<sup>2</sup>정회원, 부산대학교 사회환경시스템공학과 부교수, 교신저자

\*Corresponding author: yimhj@pusan.ac.kr

Department of Civil Engineering, Pusan National University, Pusan, 46241, Korea

•본 논문에 대한 토의를 2024년 11월 30일까지 학회로 보내주시면 2024년 12월호에 토론결과를 게재하겠습니다.

확률론적 방법인 MCS를 활용하여 변수들에 대한 변동성을 고려하였고 Polder(2007)는 성능 기반 지침을 통해 염화물 확산을 정량적으로 평가하려는 시도를 하였다. 그러나 이러한 선행 연구들에서는 실제 해양 환경의 다양한 변수들을 반영하지 못하는 한계가 있었다.

이러한 한계를 극복하기 위해 실제 해양 환경의 다양한 조건을 반영하고자 해안가에 직접 시편을 설치하여 데이터 취득과 염화물 확산계수를 예측하는 연구도 진행되었다. Kim(2013)은 해수폭로시험장(비말대, 간만대)을 만들어 600여개의 실험체를 폭로시키고 5년 동안 염화물 침투 프로파일을 조사하였으며, 그에 따라 실내실험과 현장실험의 표면염화물 결과를 통해 Fick의 법칙에 기반한 모델을 제안하였다. Yoon(2022)은 비말대, 간만대, 침치대를 구분하여 시험장을 만들었고 6개월~60개월 동안 기간이 증가함에 따라 확산계수의 변화를 관찰하였다. W/B, 이온 전하 이동, 야의 노출 기간, 플라이 애쉬 대체 비율의 변수로 비선형 회귀 분석을 통해 예측식을 제안하였다. 그러나 이러한 실제 환경의 연구에서도 제한된 변수만을 고려하는 한계가 있으며, 최근에는 컴퓨터 하드웨어와 기계학습 모델의 발전에 따라 다양한 변수를 고려할 수 있는 기계 학습 기법을 활용한 염화물 확산계수 예측 모델이 개발되고 있다.

Huiping(2024)은 KNN, SVM, ANN 세 가지 모델을 활용하고 실험실 기반의 데이터를 활용하여 W/B, 시멘트종류, 압축강도 등을 고려하여 높은 정확도로 염화물 확산계수를 예측하였다. Tran(2022)은 혼합재가 포함된 콘크리트의 염화물 확산계수를 예측하기 위해 KNN, SVM, ELM, LGB, XGB, RF, GBM, AdB 8 가지 기계 학습 알고리즘을 활용하였다. 6개의 논문에서 수집된 127개의 실험 데이터로 콘크리트 물성치인 W/B, 시멘트 함량, 플라이 애쉬 함량, GGBFS 함량, 실리카흙 함량, 골재 함량, 플라이 애쉬의 비표면적 및  $C_3A$  함량을 변수로 선정하였고 그 결과 GBM 모델이 가장 높은 성능을 보였다. Xuanrui(2024)은 실험실 기반에서 35일~235일 확산계수를 측정하고 W/B, 굵은골재 부피 비율, 시간, 온도, 습도의 변수를 고려하여 BP, DT, RF, LR, RR, KNN 기계학습 모델을 선정하여 모델에 적용하였고 0.608~0.874의  $R^2$ 으로 높은 예측 모델을 개발하였다. 하지만 선행 연구들은 실험실 기반의 데이터를 활용하고 있는데, 실험실 사용 염화물(NaCl)은 해수 노출에 비해 총 염화물 함량이 낮게 측정되기 때문에(De Weerd et al., 2023) 실제 현상과 오차가 발생할 수 있으며 해양 환경의 변수를 고려하지 못한다는 점에서 공용 중인 콘크리트 교량에 적용에는 한계를 가지고 있다.

따라서 본 논문에서는 해양 환경에 위치한 콘크리트 교량 12개교 하부구조(대기중, 비말대, 간만대)에서 취득한 107개 염화물 프로파일 데이터와 6개 변수의 데이터를 수집하였다. 수집된 염화물 프로파일을 통해 걸보기 확산계수를 산출하고 선행연구 등에서 검증된 기계학습 모델(RF, GBM, KNN)을

활용하여 염화물 확산계수 예측 모델을 개발하였다. 또한 콘크리트 물성치를 포함한 6개 변수(W/B, 시멘트 종류, 굵은골재 부피 비율, 공용연수, 압축강도, 노출 환경) 모델과 노출 환경을 제외한 5개 변수 모델, 정밀안전진단에서 취득 가능한 3개 변수(공용연수, 압축강도, 노출 환경) 모델의 성능을 비교하였다. 이를 통해 노출 환경 변수의 중요성을 검토하고, 정밀안전진단에서 취득 가능한 데이터만으로도 염화물 확산계수를 충분히 예측 가능성을 확인하였다.

## 2. 공용 중인 교량의 데이터 수집 전처리

### 2.1 공용 중인 교량의 데이터 수집

시설물안전법, 시설물 안전 및 유지관리 실시 세부지침에 의해 해안가에서 250m 이내의 콘크리트 시설물은 염화물 함량을 측정하여야 하며, 측정방법은 KS F 2713의 산-가용성 염화물 시험방법으로 규정하고 있다. 이에 시설물통합정보관리시스템(Facility Management System, FMS)의 정밀안전진단 데이터 중 총 12개교 하부구조(대기, 비말, 간만)에서 염화물 프로파일 데이터 107개를 취득하였다. 본 논문에서는 공용연수(9년~21년), W/B(38%~47%), 시멘트 종류(1종, 2종, 5종), 굵은골재 부피 비율(CV, 36%~57.2%)과 정밀안전진단으로 부터 확인된 압축강도(24.1Mpa~65.1Mpa), 해양 노출 환경(Atmosphere, Splash, Tidal)을 추가하여 총 6개의 변수를 선정하였다. W/B, 시멘트 종류, 굵은골재 부피 비율의 데이터를 수집하기 위해 각 교량의 준공지, 특별시방서 등에서 배합비 데이터를 취득하여 기계학습을 위한 데이터 세트를 구축하였고 12개교에 대한 상세 데이터는 Table 1에 정리하였다.

### 2.2 염화물 확산계수 산정 및 산점도

염화물 프로파일은 콘크리트 내부 깊이별로 취득한 콘크리트 염화물 함량을 말하며 자유 염화물과 결합된 총 염화물 함량을 나타낸다. 현장 추출된 콘크리트 코어에는 Wall effect, Leaching에 의해 표면이 아닌 표면 근처에서 염화물 함량이 최댓값을 가지는 Peak effect가 일어난다. Wall effect란 콘크리트를 타설 시 골재가 거꾸집과 마찰로 인해 중심부로 이동하여 표면 부분에 페이스트 함량이 증가하는 현상을 말한다.

Leaching은 해양 환경에 지속적으로 노출될 때 콘크리트 표면의 수화물이 물에 용해되어 염화물 결합 능력을 저하시키는 현상이다. 결과적으로 표면의 염화물 결합 능력은 Wall effect로 증가하고 Leaching으로 감소하는 복합적 현상이 일어난다. 이 때문에 Peak effect가 일어나면 프로파일을 정확히 평가하기 위해 수학적 모델에 따른 보정 작업이 요구된다(De Weerd et al., 2023). 주로 식 (1)과 같이 Fick의 제2법칙 이용

**Table 1** Detailed Data for Each Bridge

Bridge	Service life(year)	W/B(%)	Cement type	Strength(MPa)	Zone	Coarse Aggregate Volume fraction(%)
A	11, 15	45	1, 5	36.5	Atmosphere, Splash	43
B	10, 15	45	1, 5	24.1~42	Atmosphere, Splash, Tidal	42.5
C	10, 15	47	1	28.1~30.2	Atmosphere, Splash, Tidal	46.2
D	15, 20	45	2	28.5~32.5	Atmosphere, Splash	41.5
E	9, 14	40	2	37.1~50.2	Atmosphere, Splash	56
F	10	38	1	35.3~40	Atmosphere, Tidal	57
G	16, 20	45	1	24.5~27	Atmosphere, Splash, Tidal	40.9
H	16, 21	38	5	41.8~65.1	Atmosphere, Splash, Tidal	40.6
I	10	45	1	24.3~38.5	Atmosphere, Tidal	41
J	19	45	1, 5	25.1~34.7	Atmosphere	42.5
K	10, 15, 20	45	5	25.6~45.5	Atmosphere, Splash, Tidal	36
L	16	45	1	27.8~30.2	Atmosphere, Splash	41

하여 염화물 확산을 모델링 하며, 여기서  $C$ 는 염화물 농도,  $D_{app}$ 는 겔보기 확산계수,  $t$ 는 시간,  $x$ 는 침투 깊이를 나타낸다.

$$\frac{\partial C}{\partial t} = D_{app} \frac{\partial^2 C}{\partial x^2} \quad (1)$$

반무한 재료와 일정한 경계 조건에서 Fick의 제2법칙의 해는 보완 오차 함수(*erfc*)를 사용하여 식 (2)와 같이 표현 할 수 있으며, 여기서  $C(x,t)$ 는 깊이  $x$ 와 시간  $t$ 에서의 총 염화물 함량을 의미하며, 겔보기 확산계수( $D_{app}$ )와 표면염화물량( $C_s$ )은 피팅을 통해 산정되고 이 과정에서 수학적 보정이 적용된다 (Glasser et al., 2008). 피팅 과정은 오차 함수를 다항식으로 근사한 후, 식(2)를 선형화 하여 선형 회귀 분석을 통해 겔보기 확산계수 ( $D_{app}$ )를 산정한다. 구체적인 선형화 과정은 다음과 같다.

$$C(x,t) = C_s \operatorname{erfc}\left(\frac{x}{2\sqrt{D_{app}t}}\right) \quad (2)$$

식 (2)를 선형화하기 위해 다음과 같이 식(3)으로 변환을 수행한다.

$$\sqrt{C - C_i} = \sqrt{C_s - C_i} - \sqrt{\frac{C_s - C_i}{12D_{app}t}} \cdot x \quad (3)$$

여기서  $C_i$ 는 초기 염화물 농도를 나타내며 이 식은  $y = a + bx$  형태의 선형 방정식으로 표현될 수 있으며 여기서  $y, a, b$ 는 식 (4), (5), (6)과 같다.

$$y = \sqrt{C - C_i} \quad (4)$$

$$a = \sqrt{C_s - C_i} \quad (5)$$

$$b = -\sqrt{\frac{C_s - C_i}{12D_{app}t}} \quad (6)$$

이를 활용하여 최종적으로 표면염화물량과 겔보기 염화물 확산계수는 다음의 식 (7), (8)로 산정한다.

$$C_s = a^2 + C_i \quad (7)$$

$$D_{app} = \frac{(a/b)^2}{12 \cdot t} \quad (8)$$

이러한 피팅 과정을 통해 107개의  $D_{app}$ 를 산정하였고, 기계학습 모델의 성능 및 안정성 향상, 데이터의 이상치 영향 감소를 위해 정규성을 확보(Makin, 2022)하고자 검증(Shapiro-Wilk)을 실시하였다. 이 검증기법은 P-value가 0.05보다 크면 정규성을 가진다고 판단하며, Log, 제곱, 제곱근 변환 결과 P-value가 0.42로 산정된 제곱근 변환을 적용하였다.

본 논문에서는 겔보기 확산계수를 종속변수로 정하고, 모델의 성능 향상과 과소적합(Underfitting)을 방지하기 위해 107개 데이터를 증강(표준편차 0.01 적용)한 결과 500개의  $R^2$ 은 0.835, 1,000개의  $R^2$ 은 0.834, 2,000개의  $R^2$ 은 0.832로 성능이 유사하게 나타나 최소 증강인 500개로 데이터를 구축하였다. Fig. 1은 각 변수와 종속변수의 산점도를 나타내고 있으며 각 변수와 염화물 확산계수 간의 연관성을 시각적으로 확인할 수 있다.

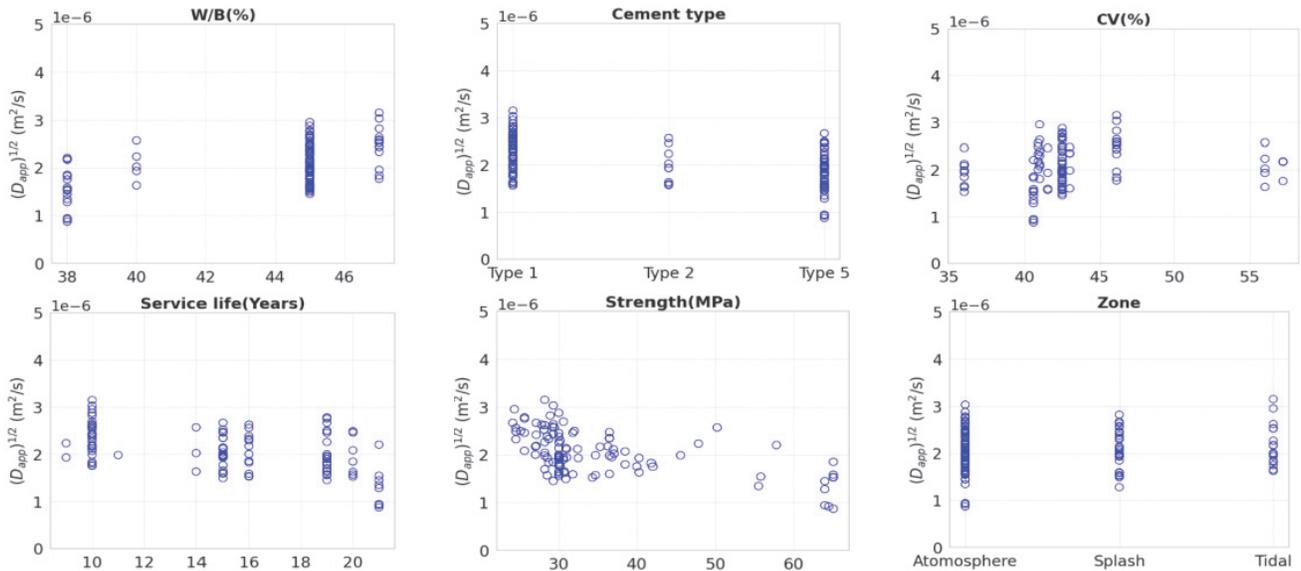


Fig. 1 Scatter Plot with the  $(D_{app})^{1/2}$

### 2.3 데이터 전처리

모델의 성능 향상을 위해 데이터 전처리를 수행하였으며 변수들 간 독립성 확보를 위해 다중공선성(Multicollinearity)에 대한 검토를 실시하였다. 다중공선성이 발생하면 변수 간 서로 종속적일 수 있으며, 이는 모델의 성능을 저하시킬 수 있다(Chan et al., 2022). 다중공선성을 확인하기 위해 2가지 검토 기법을 적용하였다. 첫 번째로 피어슨 상관계수는 두 변수 간의 선형 관계를 -1에서 1 사이의 값으로 나타내며, 절대값이 클수록 상관성이 높음을 의미한다(Benesty et al., 2009). 이에 피어슨 상관계수를 산정하여 강한 상관관계로 여겨지는 0.6 이상인 경우를 파악하였다. 두 번째로 분산팽창계수(Variance Inflation Factor, VIF)는 다중공선성의 정도를 수치화한 것으로 독립변수들 간의 선형관계의 정도를 나타낸 것이다(O'Brien, 2007). VIF를 산정하여 5 이상을 기준으로 각 변수들 간의 다중공선성을 검토하였다. Fig. 2는 피어슨 상관계수 산정 결과로 강도와 W/B의 상관 관계는 0.84로 다중공선성 가

능성이 있는 것으로 판단된다.

분산팽창계수 산정 결과 Fig. 3와 같이 W/B 5.44로 다중공선성이 상당한 수준으로 검토되었다. 또한 다중공선성 판단 기준인 5를 넘기지 않았지만 강도 결과도 4.65의 높은 수치로 산정 되었다. 일반적으로 W/B와 강도는 밀접한 연관성이 있다고 알려져 있으며, 이로 인해 높은 수준의 다중공선성이 발생할 것으로 예상될 수 있다. 그러나 본 연구에서는 서로 다른 환경에서 준공된 다양한 교량의 데이터를 다루고 있으므로, W/B와 강도가 각각 독립적으로 중요한 의미를 가질 수 있다.

따라서 본 연구에서는 W/B와 강도를 중요한 독립 변수로 고려하고 다중공선성으로 인한 모델의 성능 저하를 방지하기 위해 주성분분석법(Principal Component Analysis, PCA)을 적용하여 변수 간의 독립성을 확보하는 전처리를 수행하였다.

PCA란 다변량 통계 분석에서 사용하는 기법으로 데이터의 변동성을 보존하면서 고차원 데이터를 저차원으로 변환하여 주요 정보를 추출하고 변수 간의 독립성을 확보하는 기법이

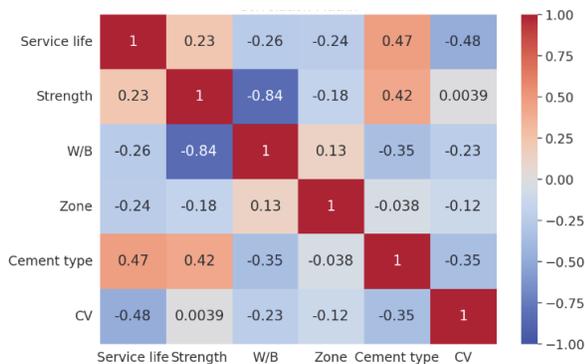


Fig. 2 Correlation Matrix of Features

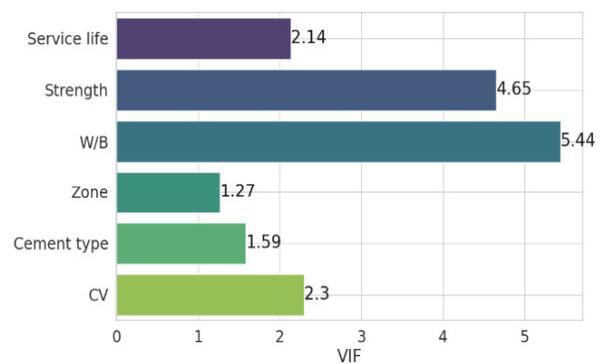


Fig. 3 VIF Analysis of Features

다. PCA가 수행되는 절차는 데이터 중심화, 공분산 행렬 계산, 고유값과 고유벡터 계산, 주성분 선택, 데이터 변환의 순으로 진행되며 설정한 주성분 축에 따라 변환된 데이터가 투영된 값으로 구성된다. PCA를 통한 차원 축소 과정에서 누적 설명 분산(Cumulative Explained Variance, CEV)은 원 데이터의 변동성의 정도를 나타내는 중요한 지표이다. 본 연구에서는 일반적으로 권장되는 70%~90%의 CEV기준(Jolliffe, 2002)과 기울기가 완만해지는 엘보우 포인트를 고려하여 차원 축소를 정하였다. Fig. 4와 같이 3개의 주성분(PC1: 40.2%, PC2: 26.4%, PC3: 16.9%)이 원 데이터 변동성의 83.5%를 설명하며 이에 따라 6차원의 데이터를 3차원 데이터로 축소하였다.

6차원의 데이터를 3차원으로 축소한 결과는 Fig. 5와 같고 (0, 0, 0)에서 멀어질수록 변동성이 큰 데이터가 포함된 것을 알 수 있으며, PCA 적용 후 데이터의 분포와 구조를 확인할 수 있다. Silhouette Score 산정 결과 0.5216(기준: 0.5 이상 군집화)으로 데이터가 군집 구조를 형성하고 있음을 알 수 있다.

PCA는 차원을 축소하며 변수들 간의 가중치가 산정되고, 이는 종속변수와 연계된 가중치가 아닌 변수들 간의 상대적 중요도를 나타내는 가중치다. 그 결과 6개의 변수에 대해서 공용연수 0.161, 강도 0.161, W/B 0.167, 노출 환경 0.125, 시

멘트 타입 0.138, 굵은골재 부피 비율 0.119의 가중치가 도출되었으며 이 가중치는 변수들의 선형 결합으로 구성된 PC(주성분)를 생성하는데 활용된다.

### 3. 기계학습 기법 및 최적화

#### 3.1 기계학습 기법

본 논문은 선행 연구들에서 염화물 확산계수 예측 성능이 검증된 모델을 선정 하였으며 3가지 모델은 Table 2과 같다.

Random Forest(RF)는 Breiman(2001)에 의해 제안된 앙상블 학습 방법으로, 여러 개의 Decision trees를 결합하여 예측하는 모델로 원본 데이터 세트에서 무작위로 샘플을 추출하여 다수의 서브 데이터 세트를 생성한다. 각 서브 데이터 세트는 개별 트리를 학습하는데 사용되고 개별 트리 결과의 평균을 통해 최종 예측을 한다. 이는 과적합을 방지하고 모델의 일

Table 2 Machine Learning Model

Model	Base
Random Forest	Decision Trees
Gradient Boosting Machine	Decision Trees
K-Nearest Neighbors	Instance

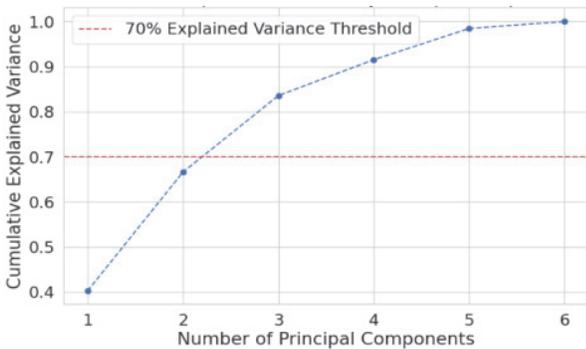


Fig. 4 Cumulative Explained Variance by PC

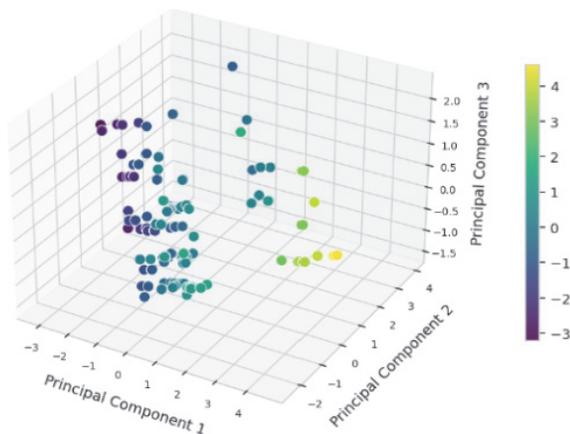


Fig. 5 3D Scatter Plot of PCA Results

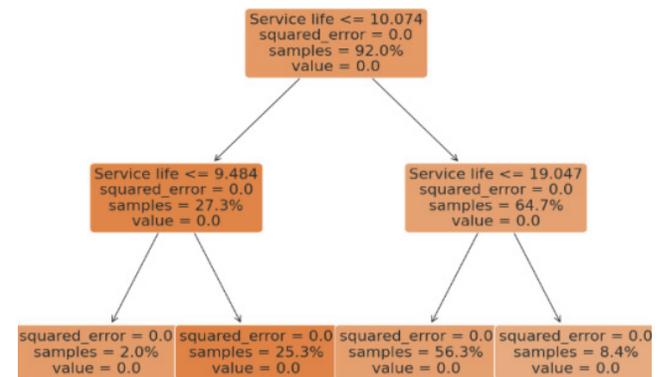


Fig. 6 RF Tree for Service Life(Partial)

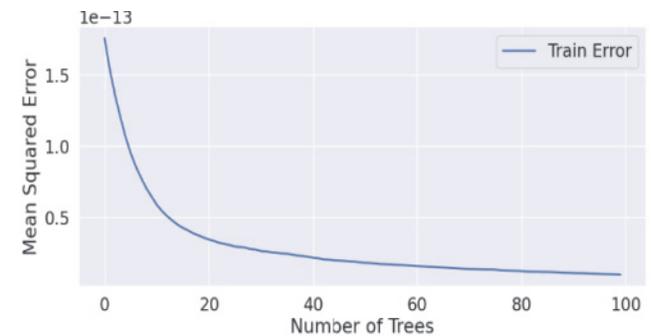


Fig. 7 Gradient Boosting Residuals over Iterations

반화 성능을 향상시킨다(Linjian, 2020). Fig. 6은 RF 수행에 따른 공용연수에 대한 트리 일부분이다.

Gradient Boosting Machine(GBM)은 Friedman(2001)에 의해 개발된 앙상블 학습 방법으로, 순차적으로 학습된 약한 학습기(weak learners)를 결합한 예측 모델이다. GBM은 각 학습 단계에서 이전 모델의 오류를 보정 하는 방식으로 작동하며, 각 단계에서 손실 함수(loss function)의 그라디언트를 사용하여 모델을 업데이트함으로써 복잡한 데이터 패턴을 효과적으로 모델링 할 수 있다. RF와 달리 과적합의 위험이 있어 데이터 전처리와 하이퍼파라미터 튜닝이 필요하다. Fig. 7은 GBM 모델에 대해 하이퍼파라미터 튜닝으로 도출된 최적 트리인 100까지 잔차(평균제곱오차)가 감소하는 결과를 보여준다.

K-Nearest Neighbors(KNN)은 비모수적(Non-parametric) Instance 기반의 방법으로 학습 대신 데이터 저장을 한다는 점에서 RF, GBM 모델과는 다르다. 새로운 데이터 포인트와 저장된 데이터의 각 포인트 간의 거리를 계산하고, 가장 가까운 K 개의 데이터를 선택하여 예측을 수행한다.

회귀 문제에서는 선택된 데이터들의 평균값으로 예측 값을 산정한다. Fig. 8은 KNN 모델에서 PC1과 PC2를 사용하여 Train 데이터와 Test 데이터의 분포를 보여준다. 또한, 새로운

데이터 포인트(Example Point)에 대해 K=2와 K=10일 때의 예측 과정을 예시로 나타내고 있다

### 3.2 하이퍼파라미터 튜닝(Hyperparameter Tuning)

기계 학습 모델의 매개변수(Hyperparameter)는 모델의 동작을 제어하고 조정하는 데 사용되는 설정값으로, 모델 학습 과정에서 학습되는 파라미터와는 구별된다. 이러한 매개변수는 모델의 구조와 학습 알고리즘의 동작 방식을 정의하며, 모델의 성능에 큰 영향을 미친다. 매개변수 최적화(하이퍼파라미터 튜닝)는 모델의 성능을 최대로 끌어올리기 위해 조정하는 과정이며, 적절한 매개변수 설정은 모델이 주어진 데이터에 대해 최적의 성능을 발휘할 수 있도록 하여 과적합(overfitting)이나 과소적합(underfitting)을 방지하는 데도 중요하다. 하이퍼파라미터 튜닝의 주요 목표는 모델 성능의 향상 및 모델이 새로운 데이터에 대해 높은 일반화 성능을 보이도록 하는 것이다.

Table 3는 RF의 하이퍼파라미터이며 max\_depth는 결정트리의 최대 깊이, leaf는 각 노드에 필요한 최소 샘플 수, split는

**Table 3** RF Hyperparameter Tuning Result

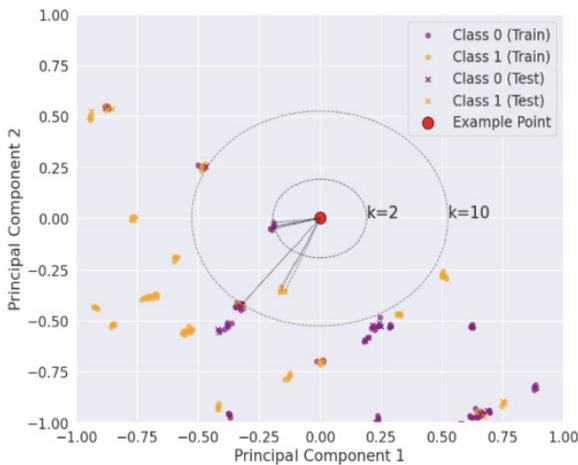
Hyperparameter			
max_depth	leaf	split	estimators
10	1	2	300

**Table 4** GBM Hyperparameter Tuning Result

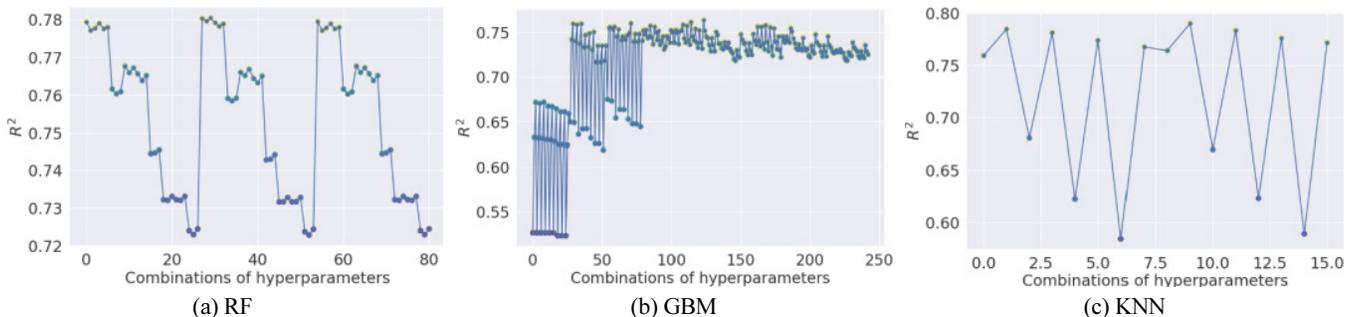
Hyperparameter				
max_depth	leaf	split	estimators	learning_rat
5	2	10	100	0.1

**Table 5** KNN Hyperparameter Tuning Result

Hyperparameter		
metric	n_neighbors	weights
manhattan	3	distance



**Fig. 8** KNN Result with PC1, PC2 (K=2, K=10)



**Fig. 9** Hyperparameter Tuning Result(RF, GBM, KNN)

각 노드를 분할하기 위해 필요한 최소 샘플 수, estimators는 RF 모델에서 사용할 트리의 수를 나타낸다.

Table 4에서 GBM의 하이퍼파라미터는 max\_depth, leaf, split, estimator은 RF와 동일하며 추가되는 지표 Learnin rate는 각 단계에서 모델이 학습하는 속도를 조절하는 역할로 0에서 1사이의 실수 값을 가진다.

Table 5의 KNN의 metric은 거리측정 방식을 결정하는 지표이고 n\_neighbors는 새로운 데이터 포인트의 회귀를 예측하기 위해 고려할 이웃의 수, weights는 각 이웃의 중요도를 설정하는 방식을 나타낸다.

각 모델의 하이퍼파라미터 튜닝 과정을 Fig. 9와 같이 볼 수 있으며, 이는 5-fold 교차 검증을 통한 성능 검증 데이터를 기준으로 가장 높은 R<sup>2</sup>에서 최적 성능의 하이퍼파라미터를 도출한 것이다. 최종적으로 선정된 하이퍼파라미터를 적용하여 테스트 데이터로 성능 평가한 결과 RF의 경우 R<sup>2</sup>= 0.859에서 0.864, GBM은 R<sup>2</sup>= 0.797에서 0.837, KNN은 R<sup>2</sup>= 0.851에서 0.870으로 최대 4.8%의 성능이 향상되었다.

#### 4. 염화물 확산계수 예측 결과

RF, GBM, KNN 3가지 모델에 대해 Train 데이터 80%와 Test 데이터 20%로 설정하여 모델 정확성을 검증하였다. 일반적으로 활용하는 회귀모델의 평가 지표인 식 (9)의 결정계수(R<sup>2</sup>)와 평균 제곱근 오차(RMSE) 및 평균 절대 오차(MAE)를 활용하여 모델을 평가하였다.

$$R^2 = 1 - \frac{MSE}{\frac{1}{n} \sum_{i=1}^n (y_i - y_{mean})^2} \quad (9)$$

평균 제곱근 오차(RMSE)는 예측 값과 실제 값 간의 차이를 제공하여 평균을 구하고, 그 평균의 제곱근을 취한 값으로 예측 오차의 크기를 측정하는 지표로 값이 작을수록 오차가 적

으며 식 (10)와 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{mean})^2} \quad (10)$$

평균 절대 오차(MAE)는 예측 값과 실제 값 사이의 절대 오차의 평균으로 예측 정확도를 측정하며, 값이 작을수록 더 정확한 모델임을 나타낸다(식 (11)).

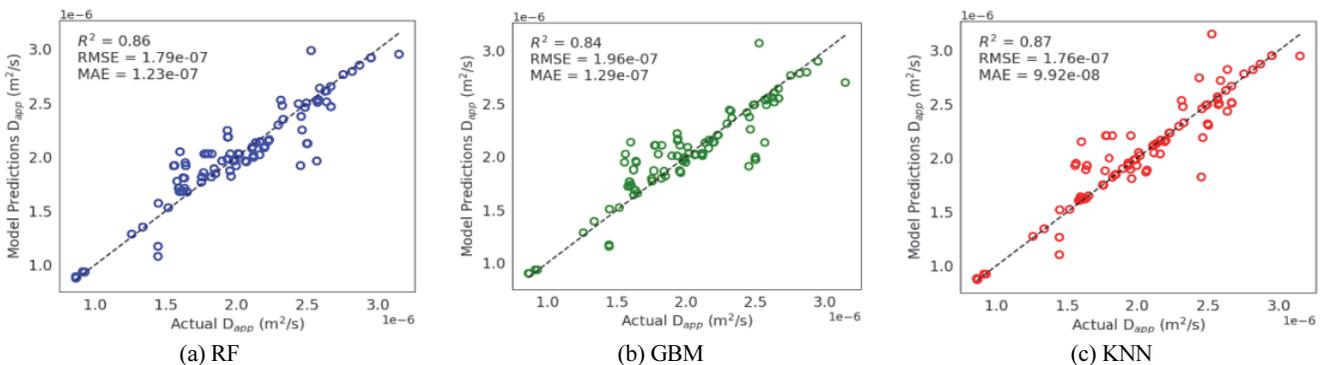
$$MAE = \frac{1}{n} |y_i - y_{predicted\ value}| \quad (11)$$

평가지표 선정 후 콘크리트 물성치를 모두 포함한 6개 변수(공용연수, 강도, W/B, 시멘트 타입, 굽은골재 부피 비율, 노출 환경)의 모델과 정밀안전진단에서 취득 가능한 3개 변수(공용연수, 강도, 노출 환경)에 대해 평가를 하였다. 먼저 6개 변수 모델의 결과 그래프는 Fig. 10과 같다. 6개 변수에 대한 평가 지표 결과는 Table 6과 같고 모델별 결과를 보면 R<sup>2</sup>이 가장 높고 RMSE, MAE 낮게 산정되어 KNN 모델이 가장 높은 예측 성능을 보여주었다. RF, GBM, KNN 평균 R<sup>2</sup>은 0.857로 산정되었다.

또한 노출환경의 변수에 대해 영향 정도를 알아보기 위해 노출 환경을 제외한 공용연수, 강도, W/B, 시멘트 타입, 굽은골재 부피 비율 5개 변수만으로도 모델을 예측하였다. 그 결과는 Fig. 11과 Table 7에 나타내었다. 6개 변수 모델과 노출 환경을 제외한 모델을 비교해 보면 RF의 경우 R<sup>2</sup>은 0.864에서

**Table 6** Model Results Including All Variables

	RF	GBM	KNN	Average
R <sup>2</sup>	0.864	0.837	0.870	0.857
RMSE	1.790e-07	1.958e-07	1.747e-07	1.83e-07
MAE	1.226e-07	1.289e-07	9.875e-08	1.17e-07



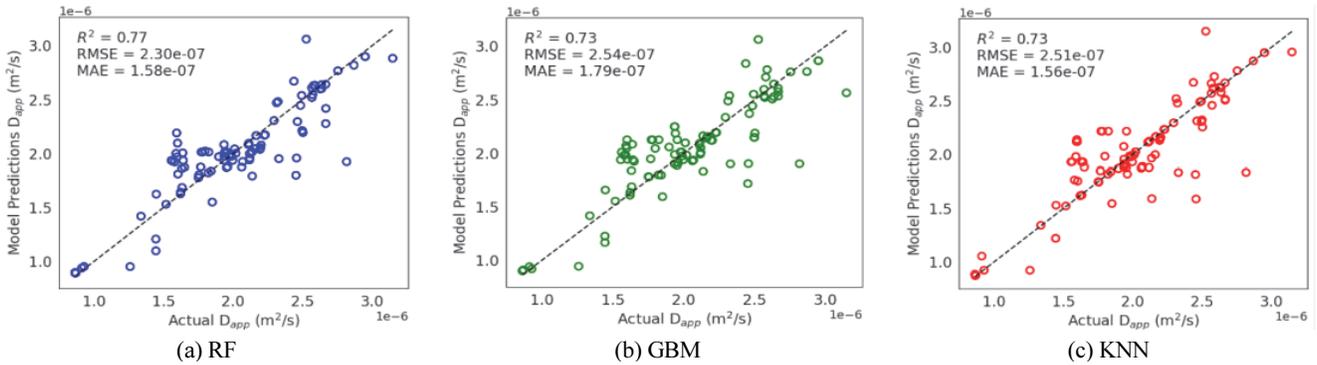
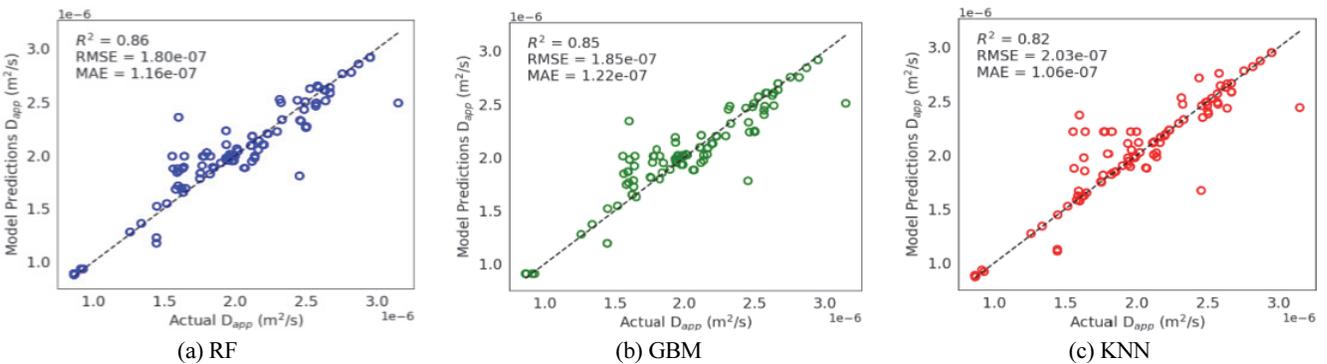
**Fig. 10** Model Results Including All Variables

**Table 7** Results of the model excluding exposure environment

	RF	GBM	KNN	Average
R <sup>2</sup>	0.775	0.726	0.731	0.744
RMSE	2.303e-07	2.537e-07	2.513e-07	2.45e-07
MAE	1.583e-07	1.785e-07	1.582e-07	1.65e-07

**Table 8** Model Results(Service life, Strength, Zone)

	RF	GBM	KNN	Average
R <sup>2</sup>	0.862	0.853	0.824	0.846
RMSE	1.796e-07	1.855e-07	2.033e-07	1.89e-07
MAE	1.163e-07	1.222e-07	1.056e-07	1.15e-07

**Fig. 11** Results of the model excluding exposure environment**Fig. 12** Model Results(Service life, Strength, Zone)

0.775(10.3% 감소), GBM의 R<sup>2</sup>은 0.837에서 0.726(13.2% 감소), KNN의 R<sup>2</sup>은 0.870에서 0.731(15.9% 감소)로 평균 R<sup>2</sup>은 0.857에서 0.744로 13.19%의 감소하였고, RMSE, MAE는 더 높은 값으로 오차 증가 및 정확도가 감소한다는 것을 알 수 있다. 이는 해양 환경에 위치한 콘크리트 교량의 경우 노출 환경 변수 포함이 염화물 확산계수 예측 모델 성능을 향상시킬 수 있음을 나타낸다.

정밀안전진단에서 취득 가능한 변수인 공용연수, 강도, 노출 환경 3개 변수 모델 결과의 시각화한 그래프는 Fig. 12, 평가 지표 결과는 Table 8과 같다. 모델별 결과를 보면 R<sup>2</sup>의 경우 RF 모델이 제일 높게 나왔고, RMSE도 RF 모델이 가장 낮게 산정되었다. MAE의 경우는 KNN의 모델이 가장 낮게 나왔으며, RF, GBM, KNN 평균 R<sup>2</sup>의 경우 0.846으로 산정되었다.

각 모델별 6개 변수 모델과 3개 변수 모델을 비교해보면 RF의 경우 비슷한 수준의 R<sup>2</sup>, RMSE, MAE를 보였다. 이는 여러

독립적인 의사결정 트리의 예측 결과를 평균 내는 방식으로 작동하고 각 트리가 서로 다른 변수 조합을 사용하므로 일부 변수가 제거 되더라도 다른 변수들로 보완이 가능하기 때문이다. GBM의 경우 3개 변수 적용 모델이 오히려 더 좋은 성능을 보였으며, 이는 순차적 학습 과정에서 적은 수의 변수에 집중하여 과적합의 위험이 줄어들고 더 효과적으로 학습할 수 있기 때문이다. KNN의 경우 3개 변수 모델에서 성능이 저하되는 결과를 확인 할 수 있으며 이는 포인트 간의 거리를 기반으로 예측하므로, 정보 손실이 발생할 경우 유사한 데이터 포인트를 구별하는 능력이 저하될 수 있기 때문으로 보인다. 또한, KNN은 다른 두 모델과 달리 데이터의 전역적 패턴을 학습하지 않고 지역적 정보에 의존하기 때문에 변수 감소에 따라 성능이 더 민감하게 반응 할 수 있다. 이러한 각 모델의 특성에 따라 변수에 따른 성능이 다르게 나타남을 알 수 있다.

콘크리트 물성치를 포함한 6개의 변수와 정밀안전진단에

서 취득 가능한 3개 변수를 비교해보면 6개 변수의  $R^2$ 은 0.857 이고 3개의 변수의  $R^2$ 은 0.846으로 약 1.3%의 성능 차이를 보이고 있다. 또한 RMSE, MAE 역시 유사한 결과를 보인다. 결과적으로 기존 염화물 확산계수 예측 시 중요 변수인 W/B를 포함한 콘크리트 물성치를 제외하고 정밀안전진단에서 취득 가능한 공용연수, 강도, 노출 환경의 데이터만으로도 염화물 확산계수를 충분히 예측 한 것으로 확인된다.

## 5. 결론

본 논문은 해양 노출 환경의 콘크리트 교량 하부구조에서 취득한 107개 염화물 프로파일 데이터를 통해 겉보기 확산계수를 산정하였고 각 교량의 W/B, 시멘트 종류, 굵은골재 부피 비율, 공용연수, 강도, 노출 환경의 데이터를 수집하여 염화물 확산계수 예측 모델을 개발하였다. 데이터 전처리 후 3가지 기계학습 모델(RF, GBM, KNN)을 선정하고 각 모델은 하이퍼파라미터 튜닝을 통해 최적화하여 6개 변수 모델(W/B, 시멘트 종류, 굵은골재 부피 비율, 공용연수, 강도, 노출 환경)과 노출 환경을 제외한 5개 변수 모델, 정밀안전진단에서 취득 가능한 3개 변수 모델(공용연수, 강도, 노출 환경)로 구분하여 예측 성능을 비교·검토하였다. 이에 따른 결론은 아래와 같다.

- (1) 변수 간 상관관계 분석 결과 W/B와 강도의 피어슨 상관 계수는 0.84, 분산팽창계수(VIF)는 W/B 5.2, 강도 4.65로 다중공선성이 존재함을 확인하였다. W/B와 강도는 연관성이 있어 다중공선성이 발생할 것으로 예상될 수 있으나 서로 다른 환경에서 준공된 교량에서는 W/B와 강도가 각각 독립적으로 중요한 의미를 가질 수 있어 PCA 수행으로 데이터의 독립성을 확보하였다.
- (2) 6개 변수의 모델에서는 KNN의  $R^2$ 이 0.870으로 가장 높은 성능을 보였으며, 3개 변수의 모델에서는 RF의  $R^2$ 이 0.862로 가장 높은 성능을 보였다. 변수가 감소함에 따라 RF는 유사한 정도의 성능을 보이고, GBM은 성능향상, KNN은 성능이 다소 저하되는 결과를 확인하였다.
- (3) 6개 변수를 고려한 RF, GBM, KNN의 평균  $R^2$ 은 0.857로 산정되었고, 노출환경(대기, 비말, 간만)을 제외한 5개 변수의 평균  $R^2$ 은 0.744로 13.19% 감소하는 결과를 보였다. 이는 해양 환경에 위치한 콘크리트 교량의 경우 노출 환경 변수를 포함하면 염화물 확산계수 예측 모델의 성능을 향상시킬 수 있음을 나타낸다.
- (4) 정밀안전진단에서 취득 가능한 3개 변수의  $R^2$  평균은 0.846으로 6개 변수의 평균  $R^2$ 과 1.3% 성능 차이를 보였으며, 이는 정밀안전진단에서 취득한 데이터만으로도 염화물 확산계수 예측이 충분히 가능할 수 있음을 나타낸다.

## 감사의 글

이 성과는 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.

## References

1. Metha, P. K., and Monteiro, P. N. (2009), Concrete-Structure, properties, and materials. 2nd edition, Prentice Hall, New-Jersey, 113-171.
2. Metha, P. K., and Monteiro, P. N. (1993), Concrete-structure, Properties, and Materials. 2nd edition. Prentice Hall, 1-7.
3. Liu, Y., and Weyers, R. E. (1998), Modeling the time-to-corrosion cracking in chloride contaminated reinforced concrete structures, *Cement and Concrete Research*, 28(9), 1357-1368.
4. Thamoas, M. D. A., and Bamforth, P. B. (1999), Modelling chloride diffusion in concrete Effect of fly ash and slag, *Cem. Concr. Res.*, 29, 487-495.
5. Tang, L., and Sorensen, H. E. (1998), Evaluation of the Rapid Test Methods for Measuring the Chloride Diffusion Coefficients of Concrete, Nordtest Project No. 1388-98.
6. Thomas, M. D. A., and Bentz, E. C. (2002), Computer Program for Predicting the Service Life and Life-cycle Costs of Reinforced Concrete Exposed to Chlorides (Life365 Manual), SFA, Love tsvile, 2002, 8-11.
7. Bentz, E. C. (2003), Probabilistic Modeling of Service Life for Structures subjected to Chlorides, *ACI Materials Journal*, Sep.-Oct., 391-397.
8. Polder, R. B., van der Wegen, G., and Boutz, M. (2007), Performance based guideline for service life design of concrete for civil engineering structures - A proposal discussed in the Netherlands. International RILEM Workshop on Performance Based Evaluation and Indicators for Concrete Durability, 31-39.
9. Kim, M. W. (2013), Prediction of Durability Life Cycle and Repair Cost Analysis using Chloride Diffusion Coefficient for Marine Concrete Structures. Doctoral dissertation, University of Seoul, Department of Civil Engineering.
10. Yoon, Y. S., Lee, J. S., Min, J. Y., and Kwon, S. J. (2022), Behavior of apparent chloride diffusion coefficient of fly ash concrete under long-term marine exposure, *Advances In Concrete Construction*, 146, 369-380 (in Korean).
11. HuiPing, Z., XiaoChao, L., Muhammad Nasir, A., Ahmed, A., Siyab, U. A., Fadi, A., and Ayaz, A. (2024), Analyzing chloride diffusion for durability predictions of concrete using contemporary machine learning strategies *Materials Today Communications*, 38.
12. Tran, V. Q. (2022), Machine learning approach for investigating chloride diffusion coefficient of concrete containing supplementary cementitious materials. *Construction and Building Materials*, 328, 127103.
13. Xuanrui, Y., Jiehong, L., Yang, Y., and Anxiang, S. (2024), Advancing service life estimation of reinforced concrete considering the coupling effects of multiple factors: Hybridized physical testing and machine learning approach, *Journal of Building Engineering*, 84.

14. KALIS (2023). Detailed Guidelines for Safety and Maintenance of the Implementation(Bridge).
15. Korean Standards Association (2017). KS F 2713: Test method for chloride content in concrete (in Korean).
16. Weerdt, K. D., Wilson, W., MACHner, A., and Georget, F. (2023), Chloride profiles – What do they tell us and how should they be used? *Cement and Concrete Research*, 173.
17. Makin, J. G. (2022), An Introduction to Modern Statistical Learning. arXiv preprint arXiv:2207.10185.
18. Glasser, F. P., Marchand, J., and Samson, E. (2008). Durability of Concrete - Degradation Phenomena Involving Detrimental Chemical Reactions, *Cement and Concrete Research*, 38(2), 226-246.
19. Chan, J. Y., Leow, S. M., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., and Chen, Y. L. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review Mathematics, *MDPI*, 10(8), 1283.
20. Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing*, Springer, Berlin, Heidelberg
21. O'Brien, R. M. (2007), A caution regarding rules of thumb for variance inflation factors, *Quality & Quantity*, 41(5), 673-690.
22. Jolliffe, I. T. (2002), *Principal Component Analysis*. Springer-Verlag New York.
23. Linjian, W. (2020), Study on Chloride Ion Diffusion Considering the Heterogeneity of Concrete Materials and the Blocking Effect of Reinforcement, Doctoral dissertation, Tianjin University.

Received : 08/09/2024

Revised : 09/11/2024

Accepted : 09/11/2024

---

**요 지** : 염화물 확산계수는 해양환경에 위치한 콘크리트 교량의 내구성 평가를 위한 중요한 지표 중 하나이다. 본 논문에서는 기존 연구에서 고려하지 않았던 해양 노출 환경(대기중, 비말대, 간만대)과 공용 중인 콘크리트 교량의 데이터를 활용해 염화물 확산계수 예측 모델을 개발하였다. 이를 위해 교량 하부구조에서 취득한 염화물 프로파일 데이터를 활용하였고 데이터 전처리 후 기계학습 모델인 RF, GBM, KNN을 하이퍼파라미터 튜닝을 통해 최적화 하였다. 콘크리트 물성치를 포함한 6개 변수(W/B, 시멘트 종류, 굵은골재 부피 비율, 공용연수, 강도, 노출 환경) 모델과 노출 환경을 고려하지 않은 5개 변수 모델, 정밀안전진단에서 취득 가능한 3개 변수(공용연수, 강도, 노출 환경) 모델을 개발하여 성능을 비교·검토 하였다. 그 결과 해양 환경에 위치한 콘크리트 교량의 경우 노출 환경을 고려함에 따라 염화물 확산계수 예측 모델의 성능을 향상시킬 수 있음을 확인하였으며, 또한 정밀안전진단 데이터만으로도 염화물 확산계수를 효과적으로 예측할 수 있음을 확인하였다.

**핵심용어** : 기계학습, 염화물 확산계수, 해양 노출 환경, 콘크리트 교량, 정밀안전진단 데이터

---