



머신러닝을 활용한 지반함몰 위험도 예측 회귀모델

Development of a Regression Model for Predicting Ground Subsidence Risk Using Machine Learning

이성열* · 김진영** · 공명식*** · 강재모****

Lee, Sungyeol*, Kim, Jinyoung**, Kong, Myeongsik***, and Kang, Jaemo****

Abstract

Ground subsidence occurs primarily in urban areas. The main causes of this damage are underground facilities and excavations. Ground subsidence presents significant risks to life and property, which causes social disruption and emphasizes the need for preventive measures. Recent studies have identified the causes of ground subsidence and developed effective risk management strategies. Because ground subsidence results from the complex interactions of multiple factors, a comprehensive approach to risk management is required. In this study, a machine learning-based regression model was developed to predict the risk of ground subsidence. The dataset included historical subsidence records, underground utility attributes, permeability coefficients, soil layer thicknesses, and elevation data for the study area. The collected data were preprocessed using the ArcGIS software suite. Additionally, condition-based datasets were constructed. These datasets were used to train the machine learning algorithms, and the performances of the models were evaluated using the root mean square error (RMSE) and coefficient of determination (R^2) to enable the selection of the best-performing model.

Key words : Ground Subsidence, Regression Model, Machine Learning, Risk Management

요 지

도시지를 중심으로 발생하는 지반함몰은 지하매설물의 손상 및 굴착공사 등이 주요 발생 원인으로 알려져 있다. 지반함몰 발생은 인명 및 재산의 피해를 발생시켜 사회적 혼란을 야기하므로 사전에 대비해야한다. 최근 들어 국내외에서 지반함몰의 원인을 규명하고 위험도 관리에 관한 연구가 수행되고 있다. 지반함몰은 다양한 원인이 복합적으로 작용하여 발생하는 현상이므로 다양한 접근방법으로 위험도 관리를 수행해야한다. 따라서 본 연구에서는 머신러닝을 활용한 지반함몰 위험도 예측 회귀모델을 개발하고자 대상지역의 지반함몰 이력정보, 지하매설물 속성정보, 투수계수, 지층 두께, 고도의 데이터를 수집하였다. 수집한 데이터를 ArcGIS를 통해 전처리하고 조건에 따른 데이터셋을 구축하였다. 구축한 데이터셋은 머신러닝 알고리즘에 학습하여 모델을 개발하였으며, RMSE와 R^2 를 활용하여 최적의 성능을 발휘하는 모델을 선정하였다.

핵심용어 : 지반함몰, 회귀분석, 머신러닝, 위험도 관리

1. 서 론

지반함몰은 지반내 토립자 유실에 따른 공동의 발생 및

확장으로 인해 지반상부가 붕괴되는 현상을 의미한다(Kim et al., 2017). 지반함몰은 주로 개발이 활발하게 이루어지는 도시지의 중심에서 발생하므로, 발생 시 사회적 혼란을 야기

*정회원, 한국건설기술연구원 지반연구본부 박사후연구원(E-mail: leesy@kict.re.kr)

Member, Postdoctoral Researcher, Department of Geotechnical Engineering Research Korea Institute of Civil Engineering and Building Technology

**정회원, 한국건설기술연구원 지반연구본부 수석연구원

Member, Senior Researcher, Department of Geotechnical Engineering Research Korea Institute of Civil Engineering and Building Technology

***한국건설기술연구원 지반연구본부 전임연구원

Research Specialist, Department of Geotechnical Engineering Research Korea Institute of Civil Engineering and Building Technology

****교신저자, 정회원, 한국건설기술연구원 지반연구본부 수석연구원(Tel: +82-31-910-0556, Fax: +82-31-995-0946, E-mail: jmkang@kict.re.kr)

Corresponding Author, Member, Senior Researcher, Department of Geotechnical Engineering Research Korea Institute of Civil Engineering and Building Technology

한다. 따라서 지반함몰 발생 원인을 파악하고 사고를 조기에 발견하여 피해를 최소화해야 한다.

지반함몰의 주요 발생원인은 지하매설물의 손상 및 인근 굴착 공사로 보고되었다(Seoul city, 2014). 침투수 및 지하수 위의 변동 등으로 지반내 물길이 형성되고, 물길을 따라 토립자가 지하매설물 내부로 유실되어 공동이 형성된다. 지반함몰은 형성된 공동이 이완과 팽창을 반복하여 확장되어 지반상부가 붕괴되는 메커니즘을 보이는 것으로 보고되었다(Kuwano et al., 2006). 이와 같이, 지반함몰의 발생 원인을 규명하고, 위험도를 예측할 수 있는 다양한 연구가 수행되었다.

Kim et al. (2017), Sato and Kuwano (2015) 등은 지반내 공동 발생 메커니즘 규명을 위해 일본과 한국 등에서 표준사를 활용하여 모형토조에 지반함몰 발생을 모사한 실내 모형 실험을 수행하였다. Mukunoki et al. (2009)는 모형토조 하부에 관로의 손상을 모사하여 지하수위 유출입에 의해 발생하는 공동의 메커니즘을 확인하고 시각화에 대한 연구를 수행하였다. 지반함몰의 직접적인 원인은 지반내 토립자의 유실이며, 이를 유발하는 인자를 도출하기 위한 다양한 연구가 수행되었다. Takeuchi et al. (2017)은 Decision Tree를 활용하여 지반함몰 위험도를 분석하였으며, Kim (2018)은 로지스틱 회귀분석을 통해 하수관로를 활용한 지반함몰 위험도 예측 모델을 개발하였다. 또한, Kim et al. (2021)은 관로의 밀집도를 활용하여 지반함몰 위험도를 예측한 연구를 발표하였다.

지하매설물 속성정보를 활용하여 지반함몰 위험도 예측 모델을 개발한 연구도 수행되었다. Lee et al. (2023a)은 지반함몰 발생여부와 지하매설물 속성정보 데이터를 활용하여 공간회귀분석을 통해서 지반함몰 위험도 예측 모델을 개발하여 발표하였다. 또한, Lee et al. (2023b)은 지반함몰 이력정보와 지하매설물 속성정보를 활용하여 머신러닝 기반의 지반함몰 위험도 분류 모델을 개발하고 위험지도를 가시화 하였다. 이와 같이, 지반함몰 사고를 방지하기 위해 위험도를 예측하는 연구가 다수 수행되었다.

지반함몰은 다양한 원인이 복합적으로 작용하여 발생하는 특성을 나타내며, 광범위한 지역에서 일부 발생하기 때문에 다양한 접근방법을 통해 위험도를 분석해야 한다. 따라서 본 연구에서는 머신러닝 기반의 회귀분석을 통한 지반함몰 위험도 예측 모델을 개발하였다. 이를 위해 대상지역의 지반함몰 이력정보와 지하매설물 속성정보, 투수계수, 지층 두께, 고도의 데이터를 수집하고 ArcGIS를 통해 전처리 과정을 수행하여 데이터셋을 구축하였다. 구축한 데이터셋은 상관분석을 통해 지반함몰 위험도와 유의한 영향인자를 선정하였다. 선정된 인자를 통해 OLS 분석과 머신러닝 기반의 Random Forest (RF), Linear Regression (LR), XGBoost (XGB), LightGBM (LGBM) 알고리즘에 학습시켜 모델을 개발하고 최적의 성능을 발휘하는 모델을 선정하였다. 선정된 모델을 통해 지반함몰 위험도 예측 분석 시 활용한 영향인자의 중요도를 제시하였다.

2. 머신러닝 알고리즘

2.1 랜덤포레스트(RandomForest, RF)

RF는 Tree를 기반으로 구성되어 있으며, 배깅(Bagging) 기법이 적용된 대표적인 알고리즘 중 하나이다. 배깅 기법은 다수의 결정 트리를 결합하여 모델의 성능을 향상시키는 앙상블 기법이다(Breiman et al., 1984).

RF를 통해 분류와 회귀 문제를 효과적으로 해결한 연구가 다수 발표되었으며(Pal, 2005; Park et al., 2019), 과적합의 위험이 상대적으로 낮아 인자 선정이 자유로운 특징을 갖고 있다.

2.2 Linear Regression (LR)

LR은 다수의 독립변수를 활용하여 종속변수와 관계를 설명하는 선형 회귀 분석 알고리즘이다(Holmes and Rinaman, 2014). LR은 각각의 독립변수에 개별적인 가중치를 산정하여 결과를 도출하며, 선형 회귀 모델이 기반이기 때문에 과적합으로부터 상대적으로 자유롭고 계산이 매우 빠른 장점을 갖고 있다. 또한, LR을 통해서 종속변수에 대한 회귀식을 도출할 수 있으며, p-value를 확인하여 모델의 신뢰성 평가가 가능하다(Alejandra et al., 2023).

2.3 XGBoost (XGB)

XGB는 부스팅(Boosting) 기법이 적용된 알고리즘이다. 단일 Tree를 활용하여 가중치를 산정하고 이를 다음 Tree에 순차적으로 가중치를 반영하는 방법으로 결과를 도출한다. 다수의 Tree를 모두 학습에 반영하여 모델의 성능은 탁월한 것으로 알려져 있다(Chen and Guestrin, 2016). 또한, 학습 속도가 상대적으로 빠르고 과적합을 방지 및 결측치 처리 등의 특징을 갖고 있다. 하지만 학습을 위한 비용이 크게 들고, 모델의 해석이 어려운 단점을 갖고 있어, Hyperparameter 튜닝이 까다롭다(Zhang and Haghani, 2015).

2.4 LightGBM (LGBM)

LGBM은 XGB와 마찬가지로 부스팅 기법이 적용된 알고리즘이며, Tree를 기반으로 하고 있다. LGBM은 리프를 분할하여 Tree를 확장하여 모델을 최적화한다(Ke et al., 2017). LGBM은 모델 학습 시 GPU의 활용이 가능하며, 대규모 데이터셋에 대한 학습속도가 빠르며, 높은 예측 성능을 보인다. 또한, 데이터의 별다른 전처리 없이 학습의 적용이 가능한 장점을 갖고 있다. 하지만, 소수의 데이터셋을 활용할 경우 과적합의 위험이 상주하며, 모델 해석의 어려움은 특징을 갖고 있다.

2.5 성능평가 지표

머신러닝 알고리즘을 활용한 회귀 모델의 평가를 위해서 RMSE (Root Mean Squared Error)과 R^2 를 활용하였다. RMSE

는 회귀 모델의 성능 평가에 주로 활용되며, 실제 값과 예측 값 사이의 차이를 평가한다(Davide et al., 2021). RMSE는 Eq. (1)을 통해 계산이 가능하다. Eq. (1)에서 N 은 데이터의 총 개수를 의미하고, y_i 는 실제 값, y' 는 예측 값을 의미한다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y')^2} \quad (1)$$

R^2 은 모델의 신뢰도를 평가할 수 있는 가장 일반적인 지표로 0~1사이의 값을 가지며, 1에 가까워질수록 모델의 성능이 뛰어나다.

3. 연구 방법 및 데이터

3.1 연구 방법

본 연구는 지반함몰 위험도 예측 분석을 위해 머신러닝 기반의 회귀분석 모델을 개발하고자 한다. 이를 위해서 대상 지역의 지반함몰 이력정보와 지하매설물 속성정보, 투수계수, 지층 두께, 고도의 데이터를 수집하였다. 수집한 데이터를 ArcGIS 프로그램을 통해 대상지역에 나타내고, 500 m × 500 m 크기로 총 2,002개의 Grid로 구분하였다. 각각의 Grid에는 수집한 데이터가 포함되어 있으며, 분석을 위해서 Grid 내부에 포함된 데이터를 추출하였다. 추출한 데이터에는 결측 및 오류값이 존재하며, 분석을 위한 데이터셋 구축을 위해 데이터 전처리를 수행하였다. 구축된 데이터셋을 80:20의 비율로 모델의 학습을 위한 Train 데이터셋과 모델의 평가를 위한 Test 데이터셋으로 분리하였다. 구축된 데이터셋을 활용하여 선형회귀모형인 OLS (Ordinary Least Square) 분석을 수행하고, Train 데이터셋을 머신러닝 알고리즘인 RF, LR, XGB, LGBM에 학습시키고 Test 데이터셋을 활용해 모델의 평가지표를 비교하여 최적의 결과를 나타내는 지반함몰 위험도 예측 회귀모델을 제시하였다. Fig. 1은 연구의 순서도를 나타낸 그림이다.

3.2 데이터

지반함몰 위험도 예측 회귀모델의 구축을 위해 대상지역의 지반함몰 이력정보와 지하매설물 속성정보, 투수계수, 지층 두께, 고도의 데이터를 수집하였다. 수집한 데이터에는 정보가 표기되어 있지 않거나, 잘못된 정보가 입력되어 있는 데이터가 일부 존재한다. 따라서 효과적인 분석을 위해 해당 데이터를 삭제하는 방법으로 전처리를 수행하였다.

3.2.1 지하매설물

지하매설물은 상수, 하수, 가스, 전기, 통신, 난방관에 대해서 공간정보와 관로의 직경, 매설년도, 길이, 심도, 기울기 등 다양한 정보가 포함되어 있으나, 기존연구(Lee et al., 2023b)을 통해 지반함몰과 유의미한 상관관계를 나타내

고 분석에 활용가능한 데이터로 구성되어 있는 관로의 직경, 매설년도, 길이 데이터만 추출하였다. 또한, 매설년도를 통해 관로의 활용년수를 산출하고, Grid 내부에 포함되어 있는 관로의 길이를 통해 밀집도의 데이터를 산출하였다.

효율적인 분석을 위해서 6종의 지하매설물 속성정보를 1종의 지하매설물 속성정보로 병합하였다. 지하매설물의 기본단위는 관로의 길이로 선정하였으며, 활용년수는 5년을 단위로 구분하고 직경은 50 mm 단위로 구분하여 해당 범위에 포함된 관로의 기본단위를 합계하여 데이터셋을 구성하였다.

3.2.2 지반환경 특성

지반함몰은 복합적인 원인에 의해서 지반내 토립자 유실에 따른 공동의 발생 및 확장으로 지반 상부가 붕괴되는 현상이다. 따라서 지반함몰 위험도 예측을 위해서는 토립자 유실에 영향을 미치는 인자를 활용한 분석이 필요하다. 기존 연구 결과, 투수계수와 지층 두께, 고도는 지하수의 흐름 및 지반침하에 영향을 미치는 것으로 알려져 있으므로 본 연구의 영향인자로 선정하였다.

대상지역의 투수계수와 지층 두께, 고도의 데이터는 지반 조사 데이터를 기반으로 지점의 값을 확보하였으며, 포인트 형태의 데이터를 광범위한 대상지역에 반영하기 위해서 Kriging 기법을 적용하여 보간하였다. Grid 내부에서도 지반환경 특성 데이터는 다양한 범위로 존재한다. 따라서, Grid의 정중앙값을 대푯값으로 선정하여 추출하였다.

3.2.3 지반함몰 이력정보

지반함몰 이력정보는 2010년~2015년 사이에 발생한 총 8,876개의 데이터이며, 공간정보와 발생 원인이 포함되어 있다. 지반함몰 이력정보를 바탕으로 지반함몰 위험도를 산정하였다. 지반함몰 위험도는 Grid 내부에 발생한 지반함

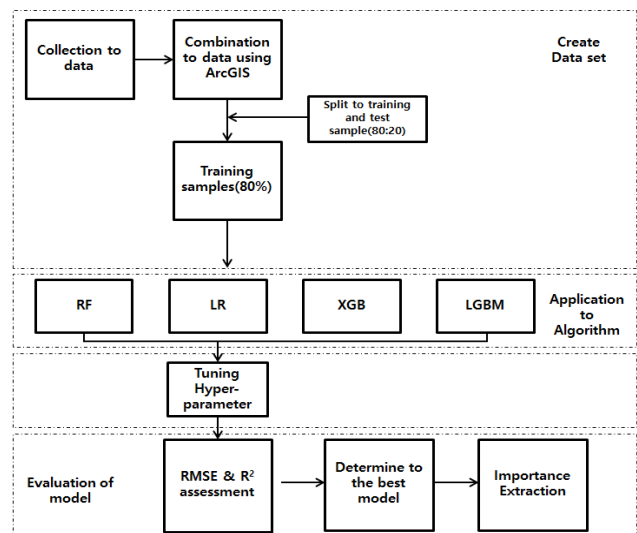


Fig. 1. Flow Chart

물 발생 횟수를 합계하여 산정하였으며, 위험도를 분류하는 명확한 기준이 없으므로 2, 3, 4단계로 구분하여 모델을 개발하고 평가지표를 비교하였다. 지반함몰 위험도를 2단계로 구분한 경우, 지반함몰 발생 여부를 기준으로 단계를 정의했다. 또한, 3단계로 구분한 경우에는 지반함몰 발생횟수가 5개 이상의 경우 3단계로 정의하고 1~4개의 경우 2단계, 0개의 경우 1단계로 정의하였다. 지반함몰 위험도를 4단계로 구분한 경우에는 지반함몰 발생횟수가 7개 이상인 경우 4단계, 3단계는 3~6개, 2단계는 1~2개, 1단계는 0개일 경우로 정의하였다. 지반함몰 위험도의 단계별 구분은 데이터의 불균형을 최소화할 수 있는 기준으로 선정하였다. 단계가 높을수록 위험도가 상대적으로 높다고 평가 가능하다. Table 1은 지반함몰 위험도 예측 회귀분석을 위한 데이터셋의 구성을 나타낸 것이다. Density는 지하매설물의 밀집도와 투수계수, 지층 두께, 고도는 숫자열 형태의 데이터로 해당 Grid의 중앙값을 추출해서 활용하였다. Year은 활용년도로 5년 단위로 활용된지 5년부터 50년까지로 구분하고 각각의 활용년수에 포함되는 관로의 길이를 합하여 활용하였다. Diameter는 관경으로 50 mm를 단위로 관로의 직경을 50~1,000 mm로 구분하고 각각의 관경에 포함되는 관로의 길이를 합하여 활용하였다. Table 2는 지반함몰 등급에 따른 발생개수를 정리한 표이며, Table 3은 지반함몰 위험도의 데이터 개수를 나타낸 것이다.

4. 연구 결과

4.1 상관분석

지반함몰 위험도 예측 회귀모델의 개발을 위해서 데이터

셋에 구성된 입력데이터(지하매설물 관경, 활용년수, 밀집도, 투수계수, 지층 두께, 고도)와 출력데이터(지반함몰 위험도) 사이의 Pearson 상관분석을 실시하였다. 지반함몰은 다양한 원인이 복합적으로 작용하여 발생하는 현상이므로, 단일 인자와 지반함몰 위험도 사이의 상관성은 낮을 것으로 예상된다. 따라서 본 연구에서는 지반함몰 위험도와 영향인자 사이의 p-value 값을 통해 인자의 유의성을 확인하였다. p-value가 0.05 이하일 경우, 지반함몰 위험도와 유의한 인자로 판단하였다. Table 2는 인자간 상관분석을 수행한 결과를 나타낸 것이다. Table 2에서 Y는 활용년수를 의미하고, DTR은 관경을 의미하며 숫자는 년도 및 관경의 범위를 의미한다. 예를 들어 Y_10은 5~10년 사이에 포함된 관로의 길이를 의미하고 DTR_100은 50~100 mm 사이에 포함된 관로의 길이를 의미한다. 상관분석 결과, 대부분의 인자가 지반함몰 위험도와 유의한 상관성을 보이는 것으로 나타났으나, 일부 p-value가 0.05 이상인 인자는 지반함몰 위험도와 유의하지 않은 인자로 판단하여 분석에서 제외하였다. 활용년수와 지반함몰 위험도와의 상관도를 살펴보면, 활용년수가 20년이 초과한 관로의 상관도가 상대적으로 높게 나타났다. 또한, 관경의 경우 550~600 mm 범위에 포함된 데이터가 높은 상관도를 보이는 것으로 나타났다.

밀집도, 투수계수, 지층 두께, 고도 인자의 상관분석 결과, 밀집도의 상관도가 높은 것으로 나타났으며, 밀집도와 지층 두께는 양의 상관관계를 보이고 고도와 투수계수는 음의 상관관계를 나타낸다. 이는 투수계수와 고도가 낮을수록 지반침하 등의 발생 위험이 높다는 기존 연구의 결과와 동일하다(Kim et al., 2012; Dayana et al., 2024).

Table 1. Condition of Datasets

Model	Density	Permeability	Thickness	Height	Year	Diameter (mm)	Risk level
1	Density	Permeability	Thickness	height	5	50	2
2							3
3							4

Table 2. Number of Ground Subsidence by Grade

Risk level	Ground subsidence count
2	Grade 1:0 case
	Grade 2:1 or more
3	Grade 1:0 case
	Grade 2:1 to 4 case
	Grade 3:5 or more
4	Grade 1:0 case
	Grade 2:1 to 2 case
	Grade 3:3 to 6 case
	Grade 4:7 or more

Table 3. Number of Data by Risk Level

Risk level	Number of data
2	Grade 1:689
	Grade 2:1,313
3	Grade 1:689
	Grade 2:667
	Grade 3:646
4	Grade 1:689
	Grade 2:415
	Grade 3:423
	Grade 4:475

4.2 지반함몰 위험도 예측 회귀모델

본 연구에서는 지반함몰 위험도 예측을 위한 회귀모델의 개발을 위해서 지반함몰 영향인자와 지반함몰 위험도 데이터를 활용하였다. 구축한 데이터셋을 머신러닝 알고리즘인 RF, LR, XGB, LGBM 알고리즘에 학습하여 모델의 성능을 평가하였다.

모델의 성능평가 지표로는 RMSE와 R^2 를 활용하였으며, Train 데이터셋과 Test 데이터셋에 대한 성능 지표 결과

값을 도출하였다. Train과 Test의 결과 값을 활용하여 두 값의 차이가 0.1 이내로 발생하게 되면 과적합을 회피한 것으로 판단하였다. Table 4는 모델의 성능평가 결과를 나타낸 표이다. Table 4를 살펴보면 모든 모델에서 과적합을 회피한 것으로 나타났다. 또한, 대체적으로 Model 3의 성능이 상대적으로 우수한 것으로 나타났으며, Model 3-XGB가 가장 높은 R^2 (0.61)가 도출되었다. Model 3-XGB에서 활용한 Hyperparameter는 Table 5와 같다.

Table 4. Results of Correlation Analysis

Model	1		2		3	
Factor	Corr	P-value	Corr	P-value	Corr	P-value
Y_5	0.119	0.000	0.137	0.000	0.135	0.000
Y_10	0.055	0.013	0.070	0.002	0.072	0.001
Y_15	0.196	0.000	0.224	0.000	0.219	0.000
Y_20	0.318	0.000	0.333	0.000	0.334	0.000
Y_25	0.467	0.000	0.517	0.000	0.533	0.000
Y_30	0.443	0.000	0.455	0.000	0.468	0.000
Y_35	0.378	0.000	0.397	0.000	0.398	0.000
Y_40	0.432	0.000	0.492	0.000	0.495	0.000
Y_45	0.466	0.000	0.546	0.000	0.551	0.000
Y_50	0.388	0.000	0.453	0.000	0.475	0.000
DTR_50	0.303	0.000	0.336	0.000	0.328	0.000
DTR_100	0.410	0.000	0.470	0.000	0.482	0.000
DTR_150	0.450	0.000	0.463	0.000	0.471	0.000
DTR_200	0.476	0.000	0.533	0.000	0.542	0.000
DTR_250	0.181	0.000	0.164	0.000	0.165	0.000
DTR_300	0.485	0.000	0.532	0.000	0.542	0.000
DTR_350	0.037	0.099	0.041	0.068	0.052	0.019
DTR_400	0.289	0.000	0.334	0.000	0.328	0.000
DTR_450	0.388	0.000	0.383	0.000	0.372	0.000
DTR_500	0.247	0.000	0.270	0.000	0.274	0.000
DTR_550	0.039	0.081	0.035	0.114	0.025	0.272
DTR_600	0.477	0.000	0.534	0.000	0.545	0.000
DTR_650	0.039	0.079	0.010	0.648	0.013	0.565
DTR_700	0.227	0.000	0.250	0.000	0.260	0.000
DTR_750	0.057	0.011	0.068	0.002	0.068	0.002
DTR_800	0.306	0.000	0.309	0.000	0.312	0.000
DTR_850	0.025	0.267	0.009	0.681	0.003	0.905
DTR_900	0.276	0.000	0.315	0.000	0.331	0.000
DTR_950	-0.035	0.115	-0.035	0.119	-0.035	0.117
DTR_1000	0.221	0.000	0.228	0.000	0.226	0.000
Density	0.374	0.000	0.410	0.000	0.406	0.000
Height	-0.280	0.000	-0.287	0.000	-0.290	0.000
Permeability	-0.094	0.000	-0.121	0.000	-0.133	0.000
Thickness	0.141	0.000	0.175	0.000	0.189	0.000

Table 5. Model Evaluation Results

Model 1	RMSE		R ²	
	Train	Test	Train	Test
RF	0.3314	0.3491	0.45	0.51
LR	0.3564	0.3708	0.44	0.38
XGB	0.3138	0.3335	0.57	0.50
LGBM	0.2972	0.3326	0.61	0.50
Model 2	RMSE		R ²	
	Train	Test	Train	Test
RF	0.4667	0.5402	0.67	0.57
LR	0.5516	0.5768	0.54	0.51
XGB	0.4604	0.5497	0.65	0.55
LGBM	0.4913	0.5461	0.63	0.56
Model 3	RMSE		R ²	
	Train	Test	Train	Test
RF	0.6558	0.7541	0.69	0.60
LR	0.7754	0.8106	0.56	0.53
XGB	0.6446	0.7494	0.69	0.61
LGBM	0.6901	0.7570	0.65	0.60

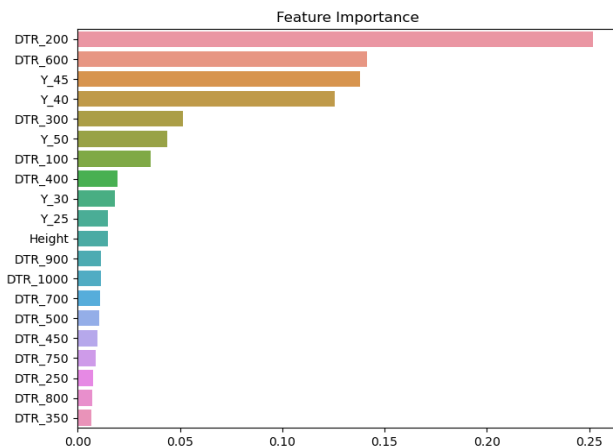


Fig. 2. Feature Importance

동일한 데이터셋을 활용하여 OLS 분석을 수행한 결과 Model 1의 R²는 0.435, Model 2는 0.543, Model 3는 0.569가 나타났다. 머신러닝 기반의 지반함몰 위험도 회귀모델 결과, 각 데이터셋에 따른 평가지표는 유의미한 차이를 보이지 않는 것으로 나타났다. 다만, 데이터의 불균형을 최소화한 Model 3의 조건에서 R²가 0.6 이상으로 나타났으며, 본 연구에서 제시한 분석 방법을 통해 지반함몰과 유의미한 상관성을 나타내는 영향인자를 추가한다면 신뢰도의 향상을 기대할 수 있을 것으로 판단된다.

최적의 성능을 도출하는 모델을 활용하여, 해당 모델이 지반함몰 위험도를 예측할 때 활용한 인자의 중요도를 Fig.

Table 6. Hyperparameters of the Model

Model	Hyperparameter
Model 3-XGB	Estimators (100)
	Learning_rate (0.15)
	Max_depth (2)

2에 나타났다. 관경이 151~200 mm 사이에 포함된 관로가 가장 높은 중요도를 도출하였으며, 활용년수가 35년 이상인 노후화된 관로의 중요도도 상대적으로 높게 나타났다. Table 6은 최적의 성능을 발휘하는 모델의 hyperparameter를 나타낸 표이다.

5. 결론 및 고찰

본 연구는 대상지역의 지반함몰 위험도를 2~4단계로 산정하고 위험도를 예측하기 위해서 머신러닝 기반의 회귀분석을 수행하였다. 지반함몰 사고 방지를 목적으로 지하매설물 속성정보를 활용한 지반함몰 위험도 예측모델을 분류 알고리즘 기반으로 개발한 연구가 수행된 바 있다. 해당 연구에서 개발된 모델은 지반함몰 특성을 감안했을 경우, 높은 신뢰도의 결과를 도출하였으나 대상지역을 3단계로만 구분하기 어렵기 때문에 동일 등급의 위험도에 대한 위험 순위의 산출이 어렵다. 본 연구에서 개발된 회귀모델의 경우 각각의 Grid마다 위험도가 수치로 예측되어 있어, 구간별로 위험 순위의 산출이 가능한 장점이 있다.

상관분석 결과, 노후화된 지하매설물 및 밀집도는 지반함몰 위험도와 높은 상관관계를 보이는 것으로 나타났다. 이는 지하매설물이 노후화 될수록 관로손상 가능성이 높아 나타난 결과로 해석할 수 있다. 또한 밀집도가 높은 지점은 지하매설물을 매설하기 위해 다수의 굴착이 수행되었을 수 있으며, 이로 인한 지반의 교란 등에 의해 나타난 결과로 판단된다 (Kim et al., 2021). 고도와 투수계수, 지층 두께도 지반함몰 위험도와 유의한 상관성을 보이는 것으로 나타났다. 이는 해당 인자들이 직·간접적으로 영향을 미치는 것으로 해석할 수 있다.

본 연구에서 개발된 Model 3-XGB는 R²가 0.61로 나타났다. 이는 일반적인 수행하는 OLS의 R² (0.569)보다 높게 나타났다. 머신러닝 기반의 회귀모델이 지반함몰 위험도 예측에 상대적으로 효과적인 것으로 판단할 수 있다. 지반함몰은 광범위한 지역에서 불규칙하게 발생하는 현상으로 정확한 위험도를 예측하는 것은 매우 어려운 일이다 (Lee et al., 2023b). 본 연구에서 개발된 모델의 성능(R² 0.6 이상)을 고려하면 대상지역의 지반함몰 위험지도의 가시화가 가능할 것이다. 또한, 지반함몰 위험 관리를 위한 GPR 등의 장비를 활용한 공동탐사 시 우선순위 결정에 도움이 될 것으로 기대된다.

광범위한 지역에서 다양한 원인이 복합적으로 작용하여 발생하는 지반함몰 특성상 다양한 접근방법으로 지반함몰 위험도를 예측하는 연구가 꾸준히 수행되어야 한다. 추후에는 지하철과 지하터널 등의 데이터를 추가적으로 활용하여 높은 신뢰도의 지반함몰 위험도 예측모델 개발이 필요하다.

감사의 글

본 연구는 과학기술정보통신부 한국건설기술연구원 연구개발적립금 자체사업으로 수행되었습니다(과제번호 20240306-001, 지하 생활기반시설 입체적 위험평가를 위한 재해·재난 DB구축(1/1)).

References

- Alejandra, E.P.H., Victor, F., Luis, G.E., and Carlos, A.D.B.P. (2023). Adaptive Significance Levels in Tests for Linear Regression Models: The e-Value and P-Value Cases. *Entropy*, Vol. 25, No. 1, 19.
- Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and regression trees*. New York: Taylor & Francis.
- Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD'16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.
- Davide, C., Matthijs, J.W., and Giuseppe, J. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, Vol. 7, e623.
- Dayana, C.C., Edgar, Q.B., and Mehrah, M. (2024). Land Subsidence Due to Groundwater Exploitation in Unconfined Aquifers: Experimental and Numerical Assessment with Computational Fluid Dynamics. *Water*, Vol. 16, No. 3, 467.
- Holmes, W., and Rinaman, W. (2014). Multiple Linear Regression. In W. Holmes, and W. Rinaman (Eds.), *Statistical Literacy for Clinical Practitioners* (pp. 367-396). Cham: Springer.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Part of Advances in Neural Information Processing Systems 30*, pp. 3146-3154.
- Kim, B.J., Jung, J.H., Jang, Y.S., and Chun, B.S. (2012). A Study on the Lining Stability of Old Tunnel Using Groundwater Flow Modelling and Coupled Stress-Pore Water Pressure Analysis. *Journal of the Korean Geotechnical Society*, Vol. 28, No. 4, pp. 101-113 (In Korean).
- Kim, J.Y., Kang, J.M., Choi, C.H., and Park, D.H. (2017). Correlation Analysis of Sewer Integrity and Ground Subsidence. *Journal of the Korean Geo-Environmental Society*, Vol. 18, No. 6, pp. 31-37 (In Korean).
- Kim, J.Y., Kang, J.M., and Choi, C.H. (2021). Correlation Analysis of the Occurrence of Ground Subsidence According to the Density of Underground Pipelines. *Journal of the Korean Geo-Environmental Society*, Vol. 22, No. 11, pp. 23-29 (In Korean).
- Kim, K.Y. (2018). *Susceptibility Model for Sinkholes Caused by Damaged Sewer Pipes Based on Logistic Regression*. Master's thesis, Seoul National University (In Korean).
- Kuwano, R., Horii, T., Kohashi, H., and Yamauchi, K. (2006). Defects of sewer pipes causing cave-in's in the road. *Proc. 5th International Symposium on New Technologies for Urban Safety of Mega Cities in Asia*, Phuket, Thailand, pp. 347-353.
- Lee, S.Y., Kang, J.M., and Kim, J.Y. (2023a). Analysis of Ground Subsidence Vulnerability in Urban Areas Using Spatial Regression Analysis. *Applied Sciences*, Vol. 13, No. 15, 8603.
- Lee, S.Y., Kang, J.M., and Kim, J.Y. (2023b). Prediction Modeling of Ground Subsidence Risk Based on Machine Learning Using the Attribute Information of Underground Utilities in Urban Areas in Korea. *Applied Sciences*, Vol. 13, No. 9, 5566.
- Mukunoki, T., Kuwano, N., Otani, J., and Kuwano, R. (2009). Visualization of three dimensional failure in sand due to water inflow and soil drainage from defected underground pipe using X-ray CT. *Soils and Foundations*, Vol. 49, No. 6, pp. 959-968.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, Vol. 26, pp. 217-222.
- Park, E.J., Park, J.H., and Kim, H.H. (2019). Mapping Species-Specific Optimal Plantation Sites Using Random Forest in Gyeongsangnam-do Province, South Korea. *Journal of Agriculture & Life Science*, Vol. 53, No. 6, pp. 65-74 (In Korean).
- Sato, M., and Kuwano, R. (2015). Influence of location of subsurface structures on development of underground cavities induced by internal erosion. *Soils and Foundations*, Vol. 55, No. 4, pp. 829-840.

Seoul Seokchon-dong Cavity Cause Investigation Committee (Seoul city). (2014). *Cause Analysis of Cavity at Seokchon Underground Roadway and Road Cavity* (In Korean).

Takeuchi, D., Fukatani, W., Miyamoto, T., and Yokota, T. (2017). Using decision tree analysis to extract factors affecting road subsidence. *Journal of the Japan Sewage Works Association*, Vol. 54, No. 657, pp. 124-133.

Zhang, Y., and Haghani, A. (2015). A gradient boosting method

to improve travel time prediction. *Transportation Research Part C. Emerging Technologies*, Vol. 58, pp. 308-324.

<i>Received</i>	■	October 10, 2024
<i>Revised</i>	■	October 10, 2024
<i>Accepted</i>	■	October 23, 2024