

RESEARCH ARTICLE

# 데이터 스케일링과 분할 방식에 따른 예측모델의 영향 분석 - 그라운드 앵커가 설치된 흙막이 벽체 대상

신준우<sup>1</sup> · 한희수<sup>2\*</sup>

<sup>1</sup>국립금오공과대학교 토목공학과 박사과정, <sup>2</sup>국립금오공과대학교 토목공학과 교수

## Analysis of the Impact on Prediction Models Based on Data Scaling and Data Splitting Methods - For Retaining Walls with Ground Anchors Installed

Jun Woo Shin<sup>1</sup> · Heui Soo Han<sup>2\*</sup>

<sup>1</sup>Ph.D. Student, Department of Civil Engineering, Kumoh National Institute of Technology

<sup>2</sup>Professor, Department of Civil Engineering, Kumoh National Institute of Technology

### Abstract

Recently, there has been a growing demand for underground space, leading to the utilization of earth retaining walls for deep excavations. Earth retaining walls are structures that are susceptible to displacement, and their measurement and management are carried out in accordance with the standards established by the Ministry of Land, Infrastructure, and Transport. However, managing displacement through measurement can be considered similar to post-processing. Therefore, in this study, we not only predicted the horizontal displacement of a retaining wall with ground anchors installed using machine learning, but also analyzed the impact of the prediction model based on data scaling and data splitting methods while learning measurement data using machine learning. Custom splitting was the most suitable method for learning and outputting measurement data. Data scaling demonstrated excellent performance, with an error within 1 and an R-squared value of 0.77 when the anchor tensile force and water pressure were standardized. Additionally, it predicted a negative displacement compared to a model that without scaling.

**Keywords:** SVR, displacement predicting, machine learning, retaining wall, data scaling, data splitting

 OPEN ACCESS

\*Corresponding author: Heui Soo Han  
E-mail: hanhs@kumoh.ac.kr

Received: 8 November, 2023

Revised: 19 December, 2023

Accepted: 19 December, 2023

© 2023 The Korean Society of Engineering  
Geology



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

### 초 록

최근 지하공간에 대한 수요가 많아지면서 대심도 굴착으로 흙막이 벽체가 활용되고 있다. 흙막이 벽체는 변위에 취약한 구조물로써 국토교통부에서 제시한 관리기준으로 계측관리를 수행하고 있으나 계측 관리를 통한 대비는 후처리에 가깝다고 할 수 있다. 따라서 본 연구에서는 머신러닝을 이용한 그라운드 앵커가 설치된 흙막이 벽체의 수평변위 예측뿐만 아니라 머신러닝으로 계측데이터를 학습하는 과정에서 데이터 스케일링과 분할 방법에 따른 예측모델의 영향을 분석하였다. 이 연구의 분석결과에 따르면 사용자 지정분할방법이 적합하지 않은 경우도 있었으나 일정 조건하에서 데이터 학습 및 출력에 가장 우수한 성능을 나타냈다. 데이터 스케일링은 앵커의 인장력, 수압을 표준화하였을 때 오차 1 이내, R-

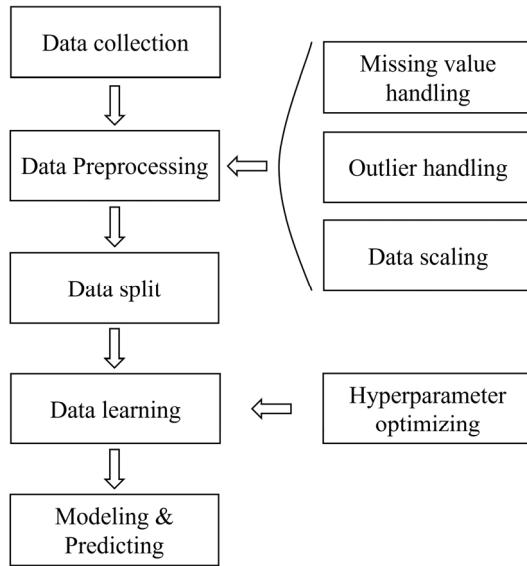
Squared 값이 0.77로 우수한 성능을 나타냈으며 스케일링을 적용하지 않은 모델 대비 음의 변위를 예측하는 결과를 보였다.

**주요어:** SVR, 변위 예측, 머신러닝, 흙막이 벽체, 데이터 스케일링, 데이터 분할

## 서론

급격한 경제성장과 가속화되는 토지개발로 대한민국 영토의 개발용지가 줄어들어 지상만을 활용한 토지개발에는 한계가 있어 최근 지하공간으로 눈을 돌려 이를 활용하고자 많은 시도들이 시행되고 있다. 이에 따라 굴착된 지반을 지지하기 위해 주로 흙막이 벽체가 사용된다. 흙막이 벽체는 변위에 취약한 구조물로써 국토교통부에서 제시한 구조물기초설계기준, 지하안전영향평가서 표준메뉴얼 등을 통해 관리기준을 세워 이를 토대로 현장에서는 붕괴 및 변형에 대한 계측관리를 수행하고 있다. 그러나 계측관리를 통한 흙막이 벽체의 변위에 대한 대비는 계측 이후에 대비하여 후처리에 가깝다고 할 수 있다. 이에 자동화 계측기술에 대한 많은 연구가 진행되고 개발이 되었으나, 실제 적용성에 대한 문제가 제기되고 있다. 변위를 확인하는 시점을 앞당기고자 변위를 예측하는 방법에 대해 다양한 연구들이 시작되었다. 흙막의 벽체의 변위에 대한 예측은 경험식에서부터 시작되었다. 이러한 경험식은 단순한 모델링을 바탕으로 사용하기 쉬운 이점이 있지만, 단순한 모델링은 적용성에 대한 문제가 있다(Peck, 1969; Lambe et al., 1970; Goldberg et al., 1976; O'Rourke, 1976, 1981; Moormann, 2004).

벽체변위를 예측하기 위해 여러 변수를 고려할 필요성이 대두되고 컴퓨터 수치해석에 대한 개발이 활발히 이루어지면서 수치해석을 통한 변위예측에 대한 다양한 연구가 수행되었다(Park et al., 2009, 2016, Ann et al., 2014; Goh et al., 2017; Chung et al., 2018; Yun, 2019; Kim and Jung, 2020; Park and Joung, 2020). 그러나 수치해석을 통한 변위예측도 많은 시간과 비용이 들 뿐만 아니라 자동화가 어렵고 현장조건을 모사하는데 한계가 있다. 이러한 선행연구들의 문제점에 대한 인식과 인공지능이 지반공학적 문제에 적용되면서 최근들어 굴착에 따른 흙막이 벽체의 변위예측에 대한 연구가 활발히 수행되었다. Kim et al.(2000)은 흙막이 벽체의 변위 예측 시스템을 개발하기 위해 다층 퍼셉트론(Multi-layer perceptron)을 이용한 인공신경망 모델을 구축하였다. Ji et al.(2014)은 옹벽의 동적 측면 변형과 지반 침하를 예측하기 위한 LSSVR(Least square support vector regression) 기반 시계열 방법을 제안하였다. An et al.(2019)은 차분진화 알고리즘의 지반공학 분야(터널 및 흙막이 벽체의 변위)에 적용성을 분석하고자 Sharan 이론 및 Blum의 방법에 대해 수치해석을 수행하였다. Yoo and Yang(2020)은 도심지 지하굴착 및 터널현장의 변위 및 침하를 예측할 수 있는 인공신경망을 개발하였다. Liu et al.(2021)은 인공 신경망(Artificial neural network), 랜덤 포레스트(Random forest) 및 서포트 벡터 머신(Support vector machine)을 통해 깊이에 따른 벽체의 수평변위를 맵핑(mapping)하기 위한 머신러닝 모델을 개발하였고, Lee et al.(2022)은 차분진화 알고리즘으로 지반의 정보를 예측하고, 디지털 트윈(Digital twin)을 통한 흙막이 벽체의 안정성을 판단하는 기법을 제시하였다. Song(2023)은 SVM(Support vector machine)을 이용하여 앵커가 설치된 흙막이 벽체의 수평변위를 예측하는 연구를 수행하였고, Lee et al.(2023)은 굴착 중인 지반의 물성치를 차분진화 알고리즘으로 추정하고, 벽체의 안정성을 평가할 수 있는 DNN(Deep neural network) 모델을 개발하는 연구를 수행하였다. Seo and Chung(2022)은 흙막이 벽체 변위 예측시스템을 개발하기 위한 딥러닝 기법을 활용한 예측모델을 구축하였고, 이후 Seo and Chung(2023)은 시공 중 흙막이 벽체 수평변위 예측을 위한 양상을 모델을 구축하였다. 이러한 머신러닝 연구에서 예측, 분류 등의 모델을 구축하기 위해서는 Fig. 1과 같은 데이터 학습과정을 거쳐야 한다.



**Fig. 1.** Data learning process.

수집된 데이터는 가공되지 않은 데이터로 머신러닝 모델의 종류, 설명변수의 공학적, 역학적, 재료적 특성에 따라 단위와 크기 및 성질이 각각 다르다. 수집된 데이터만으로는 머신러닝 모델의 학습과정에서 오차와 이상치, 유효치를 구분할 수 없기 때문에 연구자의 판단하에 데이터전처리(Preprocessing)를 수행해야하고 학습데이터와 검증데이터를 구분하는 데이터세트 분할을 수행해야 한다. 이를 고려하지 않은 학습과정을 거칠 경우 데이터세트에 이상치가 포함될 수 있고 모델의 과적합(Overfitting)의 우려가 있어 예측모델의 성능에 영향을 끼칠 우려가 있다. 따라서 본 연구에서는 시계열 비선형 데이터를 기반으로 한 머신러닝 학습과정에서 데이터전처리와 데이터분할에 따른 각 모델의 성능 비교평가를 수행하여 지반공학적 관점에서의 머신러닝 모델구축을 위한 데이터 학습과정의 가이드라인을 제시하고자 한다.

## 머신러닝 모델

### SVR

본 연구에서 사용한 머신러닝 모델은 비선형 데이터의 예측에 적합한 SVR(Support vector regression)을 사용하였다. SVR은 주어진 데이터를 기반으로 선형 회귀뿐만 아니라 비선형 관계를 모델링할 수 있고, 마진(margin)을 최대화하는 특징을 활용하여 오차를 최소화할 수 있는 특징이 있다.

### Kernel Trick

본 연구에 사용된 데이터는 시계열 기반의 계측데이터로써 깊이에 따른 변위가 비선형으로 나타나는 데이터이다. 비선형 분석을 수행하기 위해서는 커널 트릭(Kernel Trick)을 통해 데이터를 고차원 공간으로 맵핑(mapping)하여 비선형관계를 모델링할 수 있다. 커널 트릭은 대표적으로 다항식(Polynomial), 시그모이드(Sigmoid), 방사 기저 함수(Radial Basis Function, RBF) 커널이 있는데 그 중 비선형 분석에 효과적인 RBF 커널을 사용하였다. RBF 커널을 사용할 때는  $C$ ,  $\gamma$ ,  $\epsilon$  세 가지의 하이퍼파라미터(초매개변수)를 사용한다.  $C$ 는 오차 허용 범위를 조절하여 모델이 학습 데이터에 얼마나 엄격하게 모델링할지를 제어하는 하이퍼파라미터이다.  $\gamma$ 는 RBF 커널의 모양을 제어하여 데이터 포인트 간의 유사도를 조절

하는 하이퍼파라미터이다.  $\epsilon$ 은 모델의 오차 허용 범위로 모델이 허용하는 오차의 최대 크기를 의미한다. 일반적으로  $\epsilon$ 은 모델의 정확도와 정밀도를 조정하는데 사용되는 하이퍼파라미터로써 매우 작은 값으로 본 연구에서는 0.01로 고정하여 설정하였다. 하이퍼파라미터는 설정값에 따라 예측성능을 높이고 일반화에 큰 영향을 미칠 수 있으나 반대로 과적합(Overfitting)의 우려가 있을 수 있어 최적화방법을 통해 설정한다. 본 연구에서는 시간, 비용적 측면에서 다소 비효율적일 수 있으나 설정된 범위내에서의 모든 경우의 수를 검토할 수 있는 그리드 서치(Grid search) 방법을 채택하였다. 최적화방법을 통해 산정된 하이퍼파라미터는 Table 1과 같다.

**Table 1.** Applied hyperparameters value

Hyperparameter	Grid search range
$C$	0.1 : 100
$\gamma$	0.001 : 1
$\epsilon$	0.01

## 데이터 세트 구성

데이터 세트는 앵커가 설치된 흙막이 벽체 현장의 계측데이터를 수집하여 구성하였다. 수집된 데이터는 연구를 위해 엄격히 통제된 현장이 아닌 실제 시공현장으로 경북지역의 10개소에서 약 13,000개 데이터를 수집하였다.

가시설현장은 시공과정에서 지반의 변형, 날씨로 인한 지반 포화 등 다양한 변수상황이 발생한다. 따라서 각 현장을 설명할 수 있는 독립변수가 많을수록 현장의 특징을 설명하기 유리하다. 그러나 예측을 위한 모델을 구성하는 과정에서 독립변수의 개수가 많을수록 과적합(Overfitting)의 우려가 있을 수 있기 때문에 현장의 공통된 특징을 모사하기 위한 최소한의 독립변수로 계측조건과 지반조건, 구조특성을 고려하여 Table 2와 같이 구성하였다. 구성된 변수들의 기본 통계량은 Table 3과 같다.

**Table 2.** Types of variables configured in the prediction model

Type	Variable	Unit	Note
Dependent variable	Displacement	mm	-
	Elapsed time	Day	Measure condition
	Depth	m	
	Anchor tensile force	Ton	Structural character
	Water level	m	
Independent variable	Water pressure	t/m <sup>2</sup>	Ground condition
	Earth pressure	t/m <sup>2</sup>	
	Cohesion	t/m <sup>2</sup>	
	Inner friction angle	Deg (°)	
	Unit weight	t/m <sup>3</sup>	

**Table 3.** Basic statistical analysis for each variable

Variables	Count	Mean	Standard deviation	Minimum	Median	Maximum
Displacement		6.701	7.83	-4.94	3.69	62.14
Elapsed time		92.67	83.22	0	64	357
Depth		6.04	3.97	0.5	5.5	24
Anchor tensile force		14.29	4.27	7.04	13.97	25.37
Water level	12,479	-8.57	2.81	-13.25	-8.57	-2.3
Water pressure		0.79	1.94	0	0	14.38
Earth pressure		5.18	6.91	0.19	2.28	20
Cohesion		1.07	0.99	0	1	3
Inner friction angle		30.24	4.07	24	32	35
Unit weight		1.94	0.18	1.8	1.9	2.4

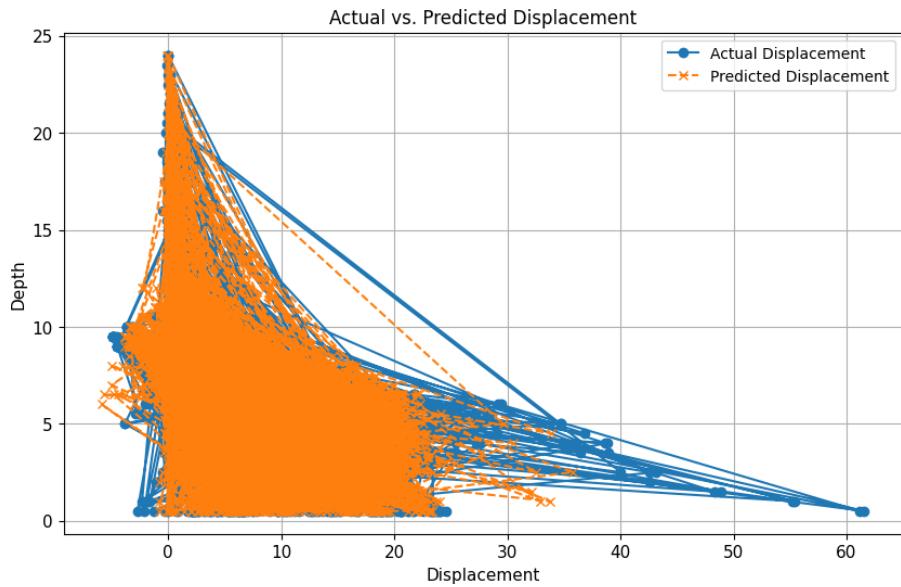
## 학습과정에 따른 예측비교

데이터 전처리와 데이터 분할에 따른 모델링의 영향을 분석하기 위해 모델의 예측성능을 비교하였다. 성능 평가지표로는 MSE(Mean squared error), MAE(Mean absolute error), R-Squared를 적용하였다. MSE와 MAE는 예측값과 실측값의 차이를 평균내어 산정하는 지표로 그 차이를 제곱하거나 절대값을 취하는 것에 차이가 있다. 두 평가지표는 0에 가까울수록 오차범위가 좁다는 것을 의미한다. R-Squared는 모델이 종속 변수의 분산을 얼마나 설명하는지를 확인할 수 있는 회귀모델의 성능을 판단하는 대표적인 평가지표로 0부터 1까지의 범위로 1에 가까울수록 우수한 예측 성능을 나타낸다.

## 데이터 분할에 따른 예측결과

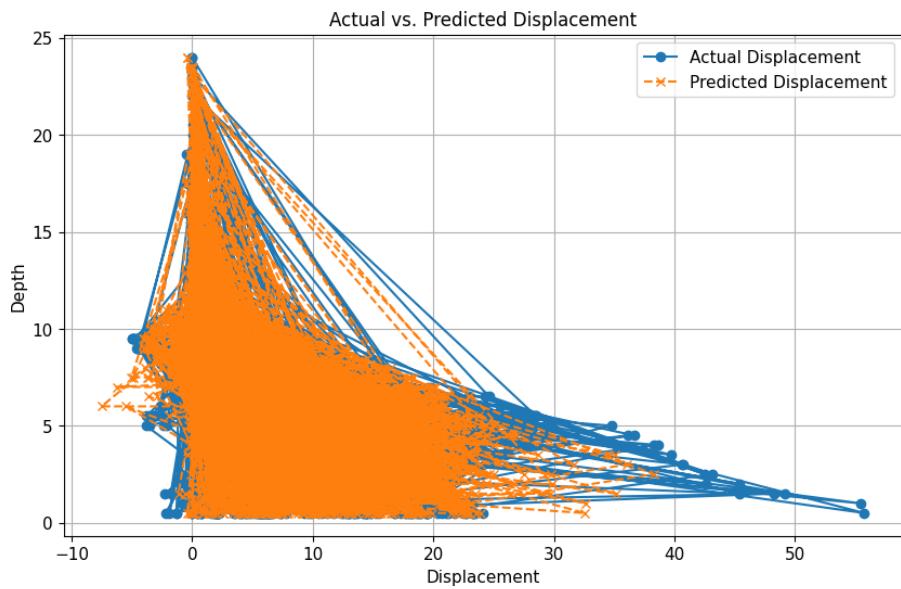
데이터 분할은 머신러닝에서 반드시 거쳐야 하는 과정으로 전체데이터를 학습데이터와 검증데이터로 분할하는 과정이다. 이를 통해 모델의 과적합을 방지하고 일반화를 검증하며 모델의 성능 평가를 수행할 수 있다. 데이터 분할은 머신러닝 모델에 따라 K-폴드 교차 검증, 계층화 데이터 분할 등 다양한 방법들이 적용될 수 있으나 이러한 방법들은 데이터 특성을 고려한 방법으로 결국 분할 방법은 난수 시드 분할(Random seed), 랜덤 분할(Random), 사용자 지정 분할(Custom)으로 구분할 수 있다. 따라서 본 연구에서는 위 세 가지 분할방법을 대상으로 모델링에 끼치는 영향분석을 수행하였다. 또한 이외의 변수로 인한 영향을 배제하기 위해 본 연구에서는 데이터 스케일링을 적용하지 않았으며 학습데이터와 검증데이터의 비율은 8:2로 설정하였다.

난수 시드 분할은 랜덤 분할 방법의 일종으로 데이터 분할을 난수 시드를 기반으로 무작위 분할하는 방법으로 결과 재현성이거나 하이퍼파라미터 튜닝, 모델의 성능 평가에 유리한 특징이 있다. 난수 시드 분할을 통한 모델링 결과, Fig. 2에서 확인할 수 있듯 검증데이터에 비해 소극적으로 예측한 결과가 나타났으나 추세는 비교적 일치하는 것으로 나타났다.



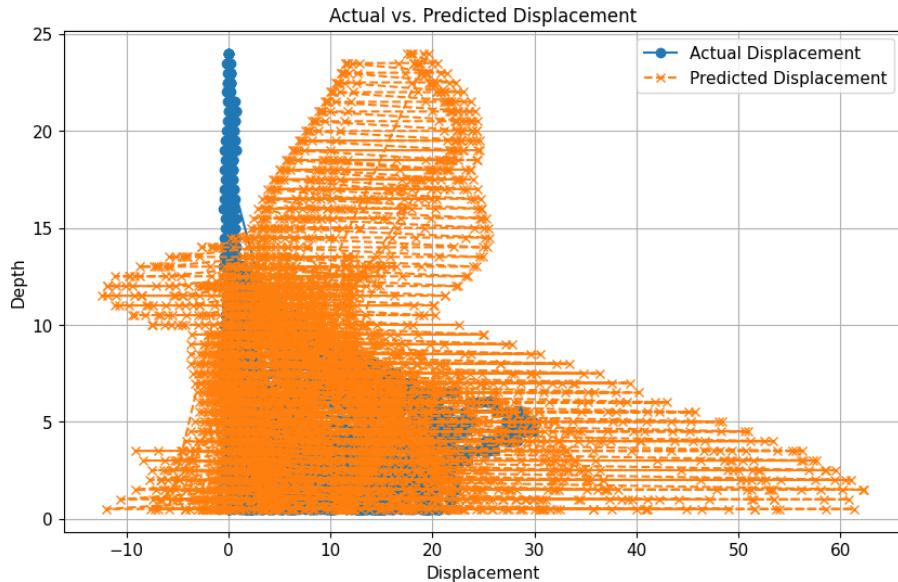
**Fig. 2.** Total site prediction results by random number seed splitting.

랜덤 분할은 데이터를 무작위로 분할하는 방법으로 모델의 일반화 능력을 평가하는데 유리한 특징이 있으나 해석마다 무작위 분할마다 다른 예측모델이 나타날 수 있다. 랜덤 분할을 통한 모델링 결과, Fig. 3과 같이 난수 시드 분할결과와 마찬가지로 검증데이터에 비해 소극적으로 예측한 결과가 나타났으나 추세는 비교적 일치하는 것으로 나타났다.



**Fig. 3.** Total site prediction results by random splitting.

사용자 지정 분할은 데이터분할을 사용자가 직접 지정하여 분할하는 방법으로 시계열 데이터와 같은 범주 전체를 학습해야 할 때 유리한 특징이 있다. 사용자 지정 분할을 통한 모델링 결과, Fig. 4와 같이 일정 깊이에서 음의 변위를 예측하였으며, 예측결과가 과다하게 해석되어 검증데이터 대비 불일치하는 것으로 나타났다.



**Fig. 4.** Total site prediction results by custom splitting.

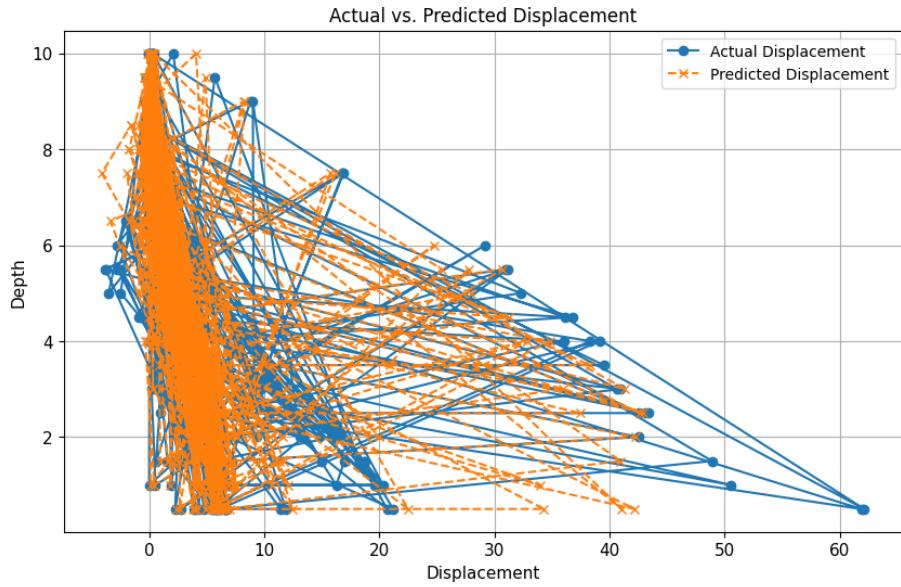
각 데이터 분할방법에 따른 예측모델의 평가지표는 Table 4와 같이 난수 시드 분할과 랜덤 분할의 경우 R-Squared 값이 0.8 이상으로 추세는 활용할 만한 수준이나, 오차(MSE)가 10 이상으로 큰 값을 나타내었다. 사용자 지정 분할은 R-Squared 값이 음수로 산정되어 활용이 어려운 평가지표를 보였다.

**Table 4.** Total site model evaluation results according to splitting type

Model type \ Evaluation	MSE	MAE	R-Squared
Random seed	12.3272	1.8214	0.8079
Random	10.6084	1.7060	0.8190
Custom	127.682	7.7073	-1.0454

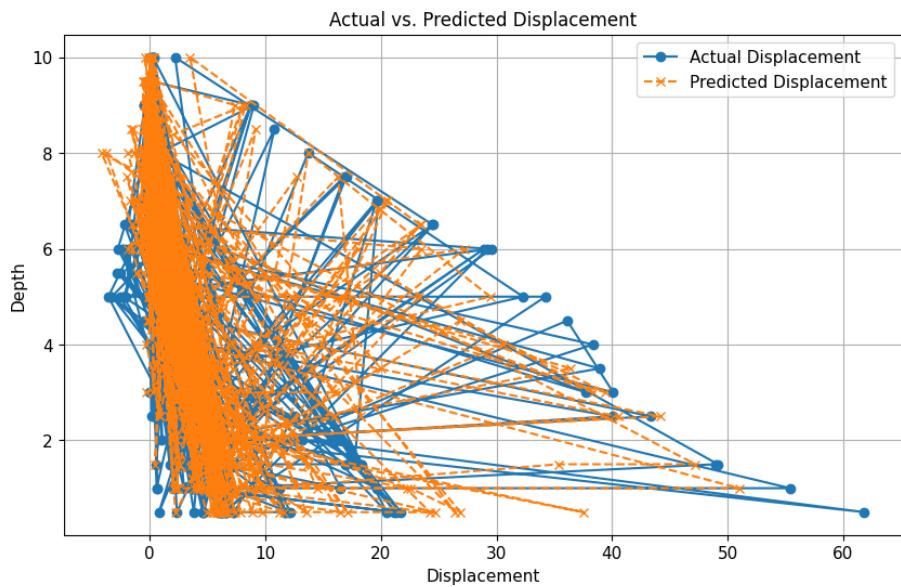
본 연구에 사용된 데이터셋은 10개의 현장의 계측데이터가 종합되어 데이터 학습시 다른 현장의 특성이 반영된 예측 결과가 나타난 것으로 판단된다. 따라서 해당 현장의 변위를 예측하기 위해서는 해당 현장의 특성만을 고려하기 위해 단일 현장의 데이터를 학습하는 것이 적합하다고 판단하여, 계측된 약 2,000개의 데이터 값을 단일현장별로 각각 구분하여 분할방법에 따른 예측결과를 재분석하였다.

난수 시드 분할을 통한 모델링 결과, Fig. 5와 같이 검증데이터에 비해 소극적인 변위를 예측하여 최대변위에 대한 정확도가 낮게 나타났으나, 예측결과는 비교적 일치하였다.



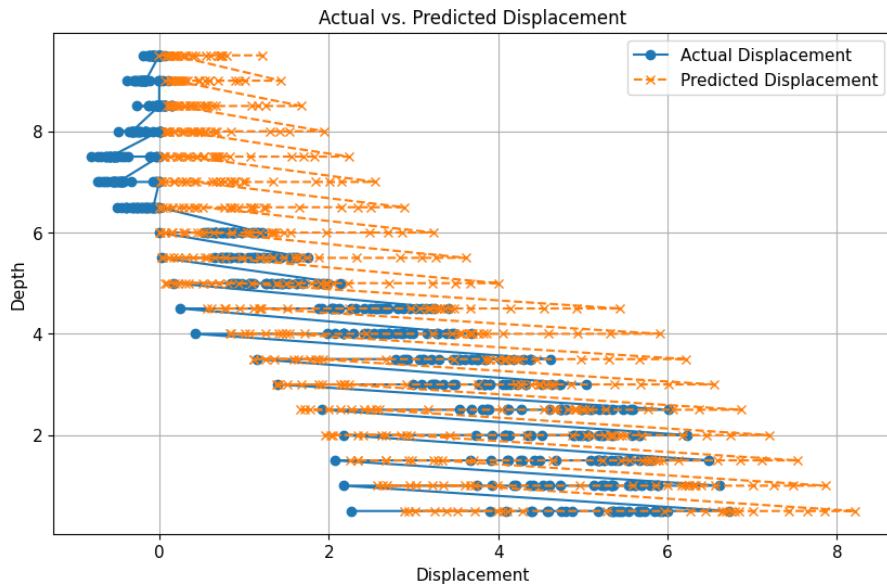
**Fig. 5.** Single site prediction results by random seed splitting.

랜덤 분할을 통한 모델링 결과, Fig. 6과 같이 난수 시드 분할과 마찬가지로 검증데이터에 비해 소극적인 변위를 예측하여 최대변위에 대한 정확도가 낮으나 추세는 비교적 일치하였다.



**Fig. 6.** Single site prediction results by random splitting.

사용자 지정 분할을 통한 모델링 결과, Fig. 7과 같이 오차의 범위가 극소한 것을 확인할 수 있다. 주목할 만한 점은 일정 깊이에서 음의 변위를 예측하지 못하고, 전체적으로 검증데이터에 비해 최대변위가 과다하게 예측되었다.



**Fig. 7.** Single site prediction results by custom splitting.

각 데이터 분할방법에 따른 단일 현장 예측모델의 평가지표 결과는 Table 5와 같이 세가지 분할 방법 모두 R-squared 값이 0.77~0.84로 높은 성능을 나타내었다. 난수 시드, 랜덤 분할에서 오차(MSE, MAE)가 비교적 큰 반면 사용자 지정 분할에서는 오차가 약 1 정도로 나타났다. 주목할 만한 것은 난수 시드 분할과 랜덤 분할은 최대변위에 대한 소극적인 예측으로 인한 오차가 크고 데이터가 무작위로 출력되기 때문에 시계열 데이터에 활용하기는 어렵다. 반면 사용자 지정 분할에서는 R-squared가 난수시드 분할과 랜덤분할 방법보다 비교적 낮은 예측 성능을 보였으나 오차에서는 우수한 성능을 보였다. 또한 시계열 데이터의 특성에 맞게 깊이와 경과 시간에 따라 데이터가 출력되는 특징을 보였다.

**Table 5.** Single site model evaluation results according to splitting type

Model type \ Evaluation	MSE	MAE	R-Squared
Random seed	11.7118	1.3966	0.8449
Random	10.6511	1.4714	0.8268
Custom	1.0853	0.8439	0.7734

### 데이터 전처리에 따른 예측결과

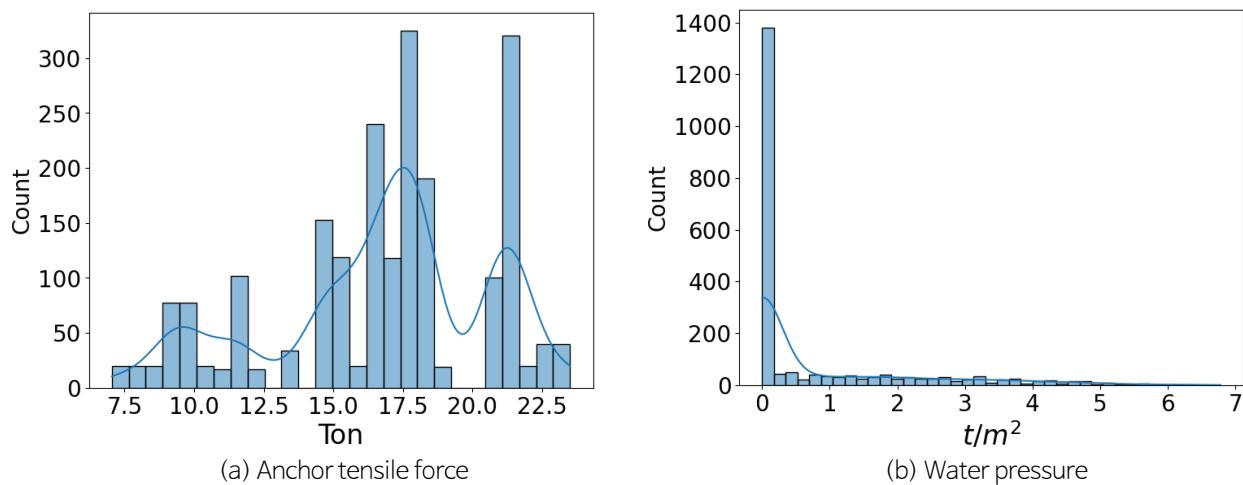
데이터 전처리는 원시데이터의 노이즈, 누락, 이상치 등의 문제를 처리하고 데이터의 일관성을 유지하고 표준화를 통하여 모델의 성능을 향상시키는 중요한 과정이다. 데이터 전처리에서는 결측치 처리, 이상치 처리, 데이터 스케일링으로 구분된다. 본 연구에 쓰인 계측데이터는 실제 시공현장의 계측데이터로 학술연구를 위한 엄격히 통제된 조건의 계측데이터가 아니기 때문에 결측치와 이상치가 존재할 수 있다. 결측치가 발견된 현장은 데이터세트에서 제외할 수 있으나 이상치의 경우 지반의 불균질한 특성과 가시설 현장의 여러 변수로 인해 제외 혹은 대체할 경우 유효데이터를 제외 혹은 대체하는 역설적인 상황이 발생할 수도 있다. 따라서 본 연구에서 검출된 결측치는 제외하고 이외의 데이터는 무결성 데이터라고 가정하였다.

데이터 자체가 의미를 가지는 설명변수인 경과시간과 깊이, 지하수위, 점착력, 내부마찰각, 단위중량은 스케일링을 적용할 경우 오히려 모델의 학습 및 성능이 저하될 수 있기 때문에 데이터를 스케일링하지 않았다.

앵커의 인장력은 설계축력이 현장에 따라 다르므로 Fig. 8a와 같이 데이터의 범위가 넓다. 데이터의 범위가 다른 변수 대비 넓은 경우 모델의 학습 및 성능 저하를 유발할 수 있기 때문에 스케일링이 필요하다.

수압은 지하수위에 따라 발생하는데, 깊이에 따라 수압이 발생하지 않는 구간이 대부분으로 데이터는 Fig. 8b와 같이 0이 최빈값으로 분포한다. 이에 따라 발생한 수압은 오히려 이상치로 판단될 가능성이 있기 때문에 스케일링이 필요하다.

따라서 본 연구에서는 설명변수 중 수압과 앵커의 인장력이 데이터 스케일링의 대표적인 방법인 표준화와 정규화에 따른 모델링에 끼치는 영향 분석을 수행하였다.



**Fig. 8.** Data distribution of variables.

앵커의 인장력만 표준화한 모델은 스케일링을 적용하지 않은 모델에 비해 전체적인 예측성능이 저하된 것을 Table 6의 평가지표에서 알 수 있다. 그러나 Fig. 9에서 음의 범위를 소극적으로 예측하였다.

**Table 6.** Single site model evaluation results by data preprocessing

Model type	Evaluation	MSE	MAE	R-Squared
Without scaling		1.0853	0.8439	0.7734
Anchor tensile force standardization		1.2220	0.9185	0.7449
Anchor tensile force normalization		2.3022	1.2319	0.5194
Water pressure standardization		1.0950	0.8301	0.7714
Water pressure normalization		1.0972	0.8611	0.7709
Both variables standardization		1.2742	0.9372	0.7340
Both variables normalization		2.2774	1.2131	0.5245
Anchor tensile force standardization & Water pressure normalization		3.0001	1.4198	0.3735
Anchor tensile force normalization & Water pressure standardization		2.5228	1.3053	0.4733

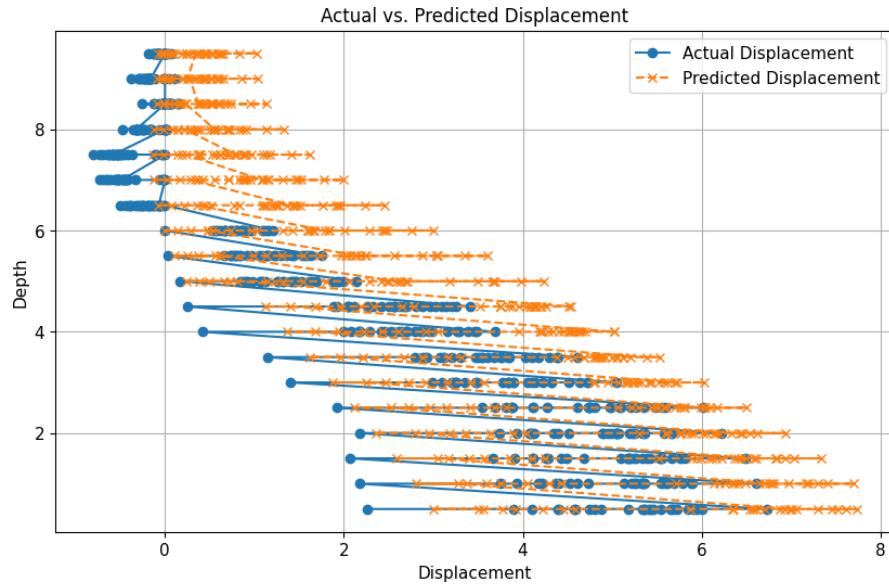


Fig. 9. Anchor tensile force standardization.

앵커의 인장력을 정규화한 모델은 표준화한 모델에 비해 MSE, MAE는 증가하고 R-squared 값이 큰 폭으로 감소한 것으로 나타났으나 Fig. 10에서 음의 변위 예측성능이 보다 우수하였다.

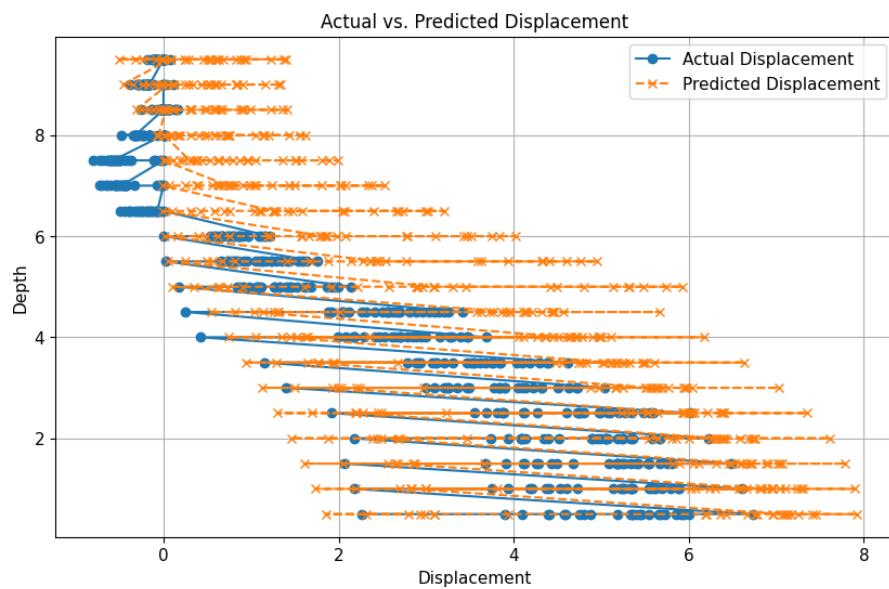


Fig. 10. Anchor tensile force normalization.

수압을 표준화한 모델은 스케일링을 적용하지 않은 모델과 평가지표상 차이가 거의 없으나 주목할 만한 점은 Fig. 11에서 음의 변위를 소극적으로 예측하였다.

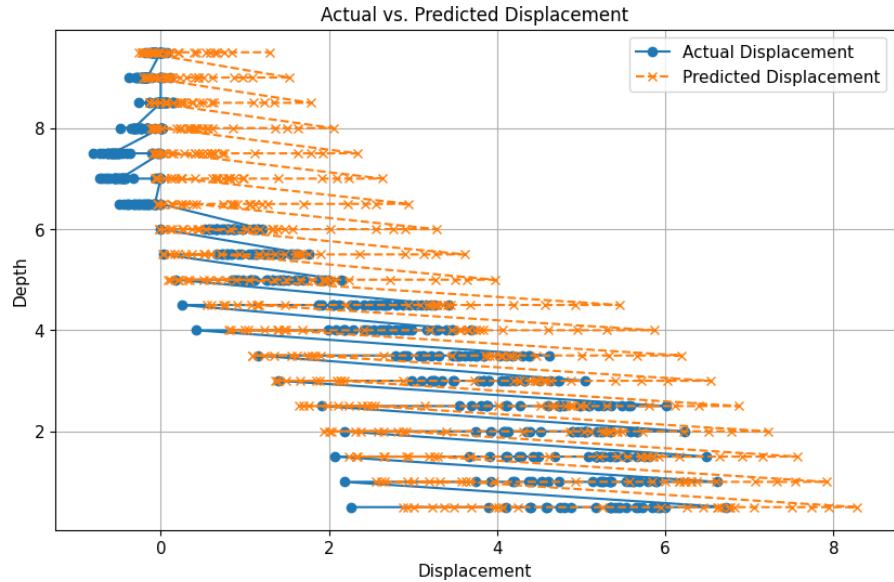


Fig. 11. Water pressure standardization.

수압을 정규화한 모델은 스케일링을 적용하지 않은 모델과 평가지표상 차이가 거의 없으나 Fig. 12에서 표준화한 모델에 비해 음의 변위를 예측하지 못한 결과를 나타내었다.

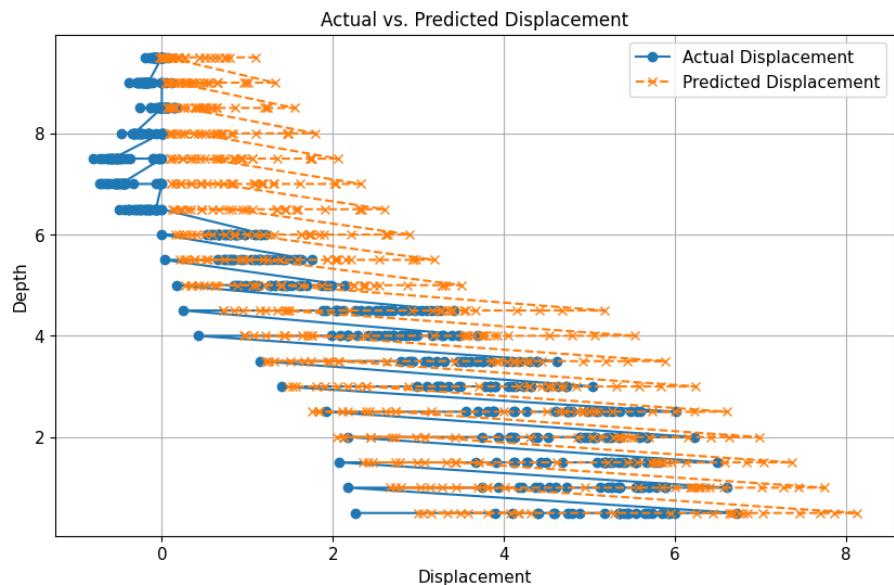


Fig. 12. Water pressure normalization.

앵커의 인장력과 수압을 표준화한 모델은 스케일링을 적용하지 않은 모델에 비해 MAE, MSE가 증가하고 R-squared 값이 감소한 것으로 나타났으나 Fig. 13에서 모든 모델과 비교하였을 때 음의 변위를 가장 우수하게 예측하였다.

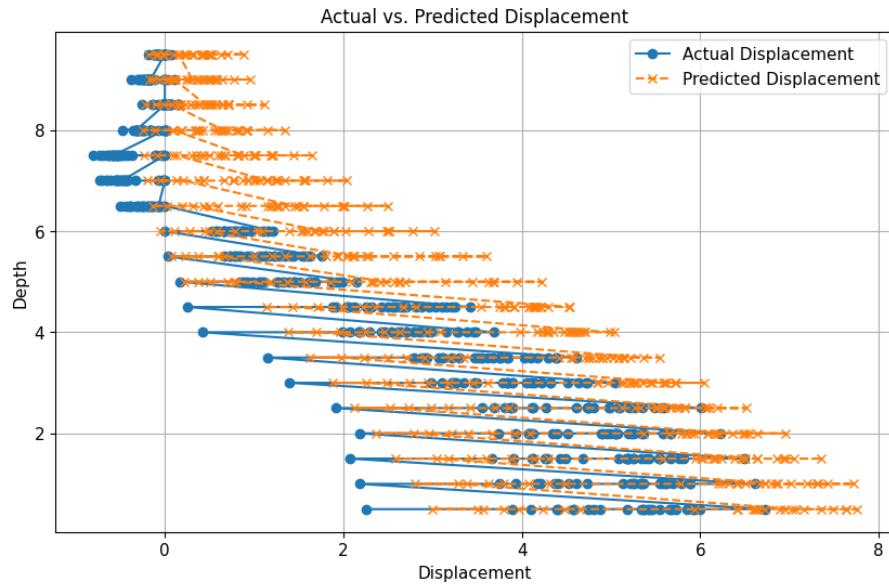


Fig. 13. Anchor tensile force & Water pressure standardization.

앵커의 인장력과 수압을 정규화한 모델은 스케일링을 적용하지 않은 모델에 비해 MAE, MSE가 증가하고 R-squared 값이 감소하였고, Fig. 14에서 음의 변위를 예측하는 성능이 가장 우수하게 향상되었다.

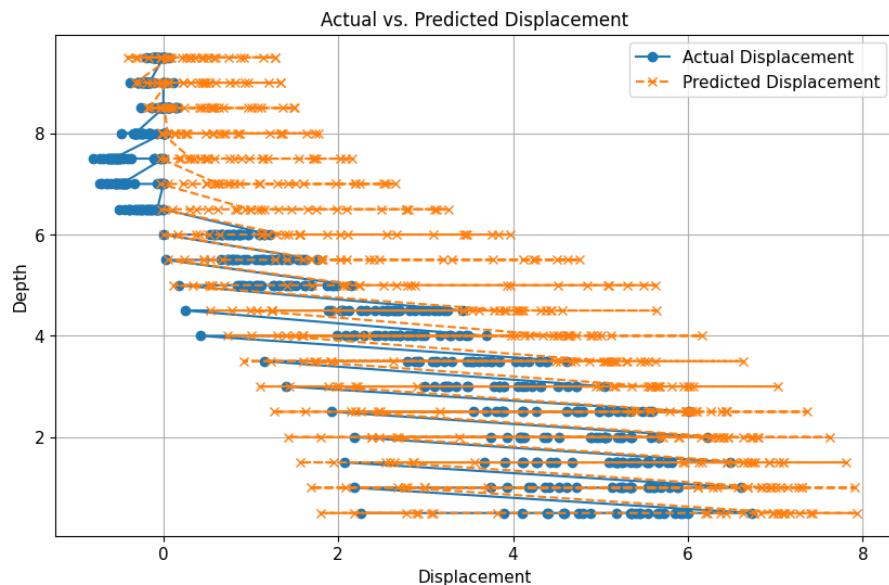
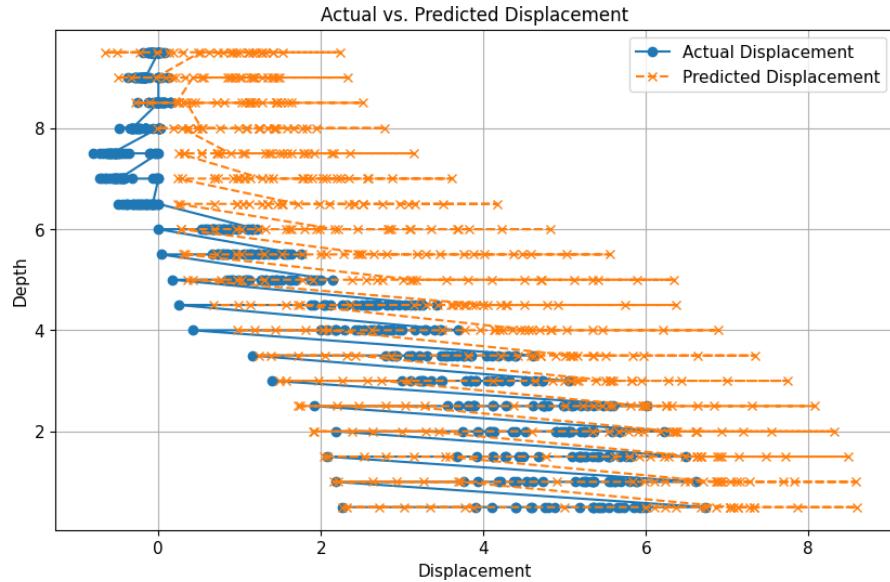


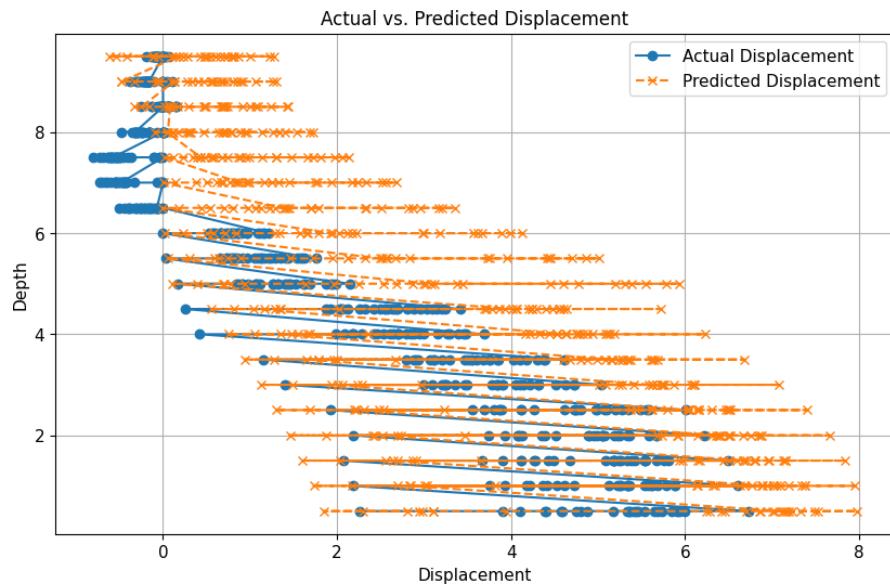
Fig. 14. Anchor tensile force & Water pressure normalization.

앵커의 인장력을 표준화, 수압을 정규화한 모델은 전체 모델 중 가장 낮은 평가지표를 나타냈으며 Fig. 15에서 음의 변위 뿐 아니라 최대변위를 검증데이터에 비해 과다하게 예측하였다.



**Fig. 15.** Anchor tensile force standardization & Water pressure normalization.

앵커의 인장력을 정규화, 수압을 표준화한 모델은 스케일링을 적용하지 않은 모델에 비해 MSE, MAE가 증가하고 R-squared값이 큰 폭으로 감소하였고, Fig. 16에서 앵커의 인장력 표준화, 수압을 정규화한 모델에 비해 최소 변위는 과다하게 예측하고 최대 변위는 소극적으로 예측되었다.



**Fig. 16.** Anchor tensile force normalization & Water pressure standardization.

## 결론

본 연구에서는 시계열 비선형 데이터를 기반으로 한 머신러닝 학습과정에서 데이터 전처리와 분할에 따른 모델링 성능에 끼치는 영향 분석을 수행하기 위해 앵커가 설치된 가시설이 시공된 경북지역의 10개소 현장의 계측데이터를 바탕으로 SVR을 이용하여 흙막이 벽체의 변위를 예측할 수 있는 모델을 구축하였고, 그 결과는 다음과 같다.

- (1) 데이터 분할 방법에 따른 모델링의 성능 영향 분석 결과, 난수 시드 분할, 랜덤 분할은 오차는 크게 나타났으나 추세는 활용할 만한 수준이었다. 그러나 무작위로 분할된 데이터가 무작위로 출력되어 깊이, 시간에 따른 변위를 확인할 수 없었다. 사용자 지정 분할의 성능지표는 활용이 불가능한 수준으로 나타났다.
- (2) 사용자 지정 분할에서 성능지표가 매우 저하된 이유는 10개소 현장의 계측데이터를 통합한 데이터세트를 학습하여 모델이 과적합되어 겹중데이터의 현장특성을 반영하지 못하였다. 따라서 단일현장을 대상으로 데이터 분할 방법에 따른 모델링 성능 영향을 재분석한 결과, 난수 시드 분할, 랜덤 분할은 전체 현장의 해석결과와 같은 결과를 나타냈으나 사용자 지정 분할은 MSE, MAE 값은 1 이하, R-squared 값은 0.77로 우수한 수준의 성능을 나타냈다.
- (3) 이를 통해 데이터세트를 구성할 때 현장의 특성을 반영하기 위해서는 단일 현장의 계측데이터로 구성하는 것이 유리하며, 통합 현장의 계측데이터로 구성할 경우에는 각 현장의 특성을 반영할 수 있는 설명변수를 구성하는 것이 모델링의 성능 향상에 도움이 될 것으로 판단된다. 또한 시계열 데이터로 계측되는 각 현장의 개별특성을 학습시키기 위해서는 사용자 지정 분할이 적합한 분할 방법으로 판단된다.
- (4) 데이터 스케일링에 따른 모델링의 성능 영향 분석 결과, 데이터 스케일링을 수행하지 않은 모델에 비해 성능은 다소 저하된 것을 알 수 있다. 그러나 스케일링을 수행하지 않은 모델에서는 예측하지 못한 음의 변위를 예측할 수 있었다. 특히, 두 설명변수(수압, 앵커의 인장력)를 모두 표준화한 모델은 평가지표에서 미소한 차이를 보였으며 음의 변위 예측에 우수한 성능을 보였다.
- (5) 이를 통해 각 설명변수의 데이터 특성, 분포에 따라 데이터 스케일링을 적용해야 하며 모델을 평가할 때 평가지표 뿐만 아니라 시각화하여 판단해야 할 필요성을 확인하였다.

추후 데이터 스케일링 적용에 따른 음의 변위 예측 성능 향상과 관련한 인과관계에 대해 충분한 연구가 이루어져 데이터의 특성과 분포에 따른 스케일링 적용에 대한 가이드라인을 제시할 수 있을 것으로 기대된다.

## 사사

이 연구는 금오공과대학교 학술연구비로 지원되었음(2021).

## References

- An, J.S., Kang, K.N., Kim, S.H., Song, K.I., 2019, Analysis for applicability of differential evolution algorithm to geotechnical engineering field, Journal of the Korean Geotechnical Society, 35(4), 27-35 (in Korean with English abstract).
- Ann, S.Y., Jung, G.S., Choi, H.S., Choi, E.J., 2014, A study on the development for the programmable numerical formular using the horizontal displacements by the inclinometer, Proceedings of the 2014 Fall Geosynthetics Conference, Jeonju, 29-32 (in Korean with English abstract).
- Chung, J.S., Shin, Y.W., Kim, M.H., Kook, Y.M., Jeong, K.Y., Kim, P.S., Lee, S.H., 2018, A study on the displacement

- behavior according to the analysis model of ground excavation, *Journal of the Korean Geo-Environmental Society*, 19(4), 27-32 (in Korean with English abstract).
- Goh, A.T.C., Zhang, F., Zhang, W., Zhang, Y., Liu, H., 2017, A simple estimation model for 3D braced excavation wall deflection, *Computers and Geotechnics*, 83(2017), 106-113.
- Goldberg, D.T., Jaworski, W.E., Gordon, M.D., 1976, Lateral support systems and underpinning, Report No. FHWA-RD-75-128 (PB-257210), Federal Highway Administration, Washington DC, 51-87.
- Ji, Z., Wang, B., Deng, S., You, Z., 2014, Predicting dynamic deformation of retaining structure by LSSVR-based time series method, *Neurocomputing*, 137, 165-172.
- Kim, H.T., Park, S.W., Kwon, Y.H., Kim, J.H., 2000, Development of a system predicting maximum displacements of earth retaining walls at various excavation stages using artificial neural network, *Journal of the Korean Geotechnical Society*, 16(1), 83-97 (in Korean with English abstract).
- Kim, T.S., Jung, Y.H., 2020, Evaluation of soil stiffness and excavation support wall deformation at deep excavation site using inverse analysis, *Journal of the Korean Geo-Environmental Society*, 12(12), 5-10 (in Korean with English abstract).
- Lambe, T.W., Wolfskill, L.A., Wong, I.H., 1970, Measured performance of braced excavation, *Journal of the Soil Mechanics and Foundations Division*, 96(3), 817-836.
- Lee, D.G., Kang, K.N., Song, K.I., 2022, Multi-layered ground back analysis of retaining wall using differential evolution algorithm : Basic research of digital twin, *Journal of the Korean Geo-Environmental Society*, 23(1), 25-30 (in Korean with English abstract).
- Lee, D.G., Yu, J.Y., Choi, J.Y., Song, K.I., 2023, Development of stability evaluation algorithm for C.I.P. retaining walls during excavation, *Journal of the Korean Geotechnical Society*, 39(9), 13-24 (in Korean with English abstract).
- Liu, D., Lin, P., Zhao, C., Qiu, J., 2021, Mapping horizontal displacement of soil nail walls using machine learning approaches, *Acta Geotechnical*, 16(12), 4027-4044.
- Moormann, C., 2004, Analysis of wall and ground movements due to deep excavations in soft soil based on a new worldwide database, *Soil and Foundations*, 44(1), 87-98.
- O'Rourke, T.D., 1976, The ground movements related to braced excavation and their influence on adjacent buildings, Report No. DOT-TST 76T-23, US Department of Transportation, 7-46.
- O'Rourke, T.D., 1981, Ground movements caused by braced excavations, *Journal of the Geotechnical Engineering Division*, 107(9), 1159-1178.
- Park, C.S., Joung, S.M., 2020, An analytical study on the horizontal displacement and excavation width of H-pile earth retaining wall, *Proceedings of the KSCE 2020 Convention*, Seogwipo, 1177-1178 (in Korean with English abstract).
- Park, C.S., Park, D.W., Lee, Y.K., 2009, A study on the displacement analysis of earth retaining wall, *Proceedings of the Civil Expo 2009*, Hoengseong, 2059-2062 (in Korean with English abstract).
- Park, D.H., Song, S.M., Park, K.C., Ahn, C.Y., 2016, Numerical simulation of braced excavation using nonlinear constitutive model, *Proceedings of the KSCE 2016 Convention*, Seogwipo, 270-271 (in Korean with English abstract).
- Peck, B.B., 1969, Deep excavations and tunneling in soft ground, State of the Art Report, Proceedings of the 7th International Conference on Soil Mechanics and Foundation Engineering, Mexico, 4, 225-290.
- Seo, S.H., Chung, M.K., 2022, Engineering application of artificial intelligence to predict displacement of retaining walls in excavation sites, *Proceedings of the KSCE 2022 Convention*, Busan, 7-8 (in Korean with English abstract).
- Seo, S.H., Chung, M.K., 2023, Development of an ensemble prediction model for lateral deformation of retaining wall under construction, *Journal of the Korean Geotechnical Society*, 39(4), 5-17 (in Korean with English abstract).
- Song, J.H., 2023, Prediction of anchored retention wall horizontal displacement using SVM, MSc Thesis, Kumoh National Institute of Technology, 19-27 (in Korean with English abstract).

Yoo, C.S., Yang, J.W., 2020, A study on development of artificial neural network (ANN) for preliminary design of urban deep Excavation and tunnelling, Journal of the Korean Geosynthetics Society, 19(1), 11-23 (in Korean with English abstract).

Yun, Y.M., 2019, Development of site management system for temporary facility construction using back analysis, Journal of the Society of Disaster Information, 15(4), 570-577 (in Korean with English abstract).