

기계학습 기법을 사용한 콘크리트 충전 강관 기둥의 강도 예측

이승혜¹ · 이유진² · 이기학^{3*} · 이재홍^{3*}

¹조교수, 세종대학교, 건축공학과, ²석사과정, 세종대학교, 건축공학과, ³교수, 세종대학교, 건축공학과

Compressive Strength Prediction of CFST Columns Using Machine Learning Methods

Lee, Seunghye¹, Lee, Yujin², Lee, Kihak^{3*}, Lee, Jaehong^{3*}

¹Assistant Professor, Dept. of Architectural Engineering, Sejong University, Seoul, 05006, Korea

²Graduate Student (Master's Course), Dept. of Architectural Engineering, Sejong University, Seoul, 05006, Korea

³Professor, Dept. of Architectural Engineering, Sejong University, Seoul, 05006, Korea

Abstract - Due to the complexities associated with the interaction between steel tube and concrete filling, more advanced regression models are demanded to predict the compressive strength of concrete-filled steel tubular (CFST) columns. Super learner models including various advanced ensemble learning methods, namely, random forest regression (RFR), an Adaptive Boosting (AdaBoost), Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Categorical Gradient Boosting (CatBoost) are presented to determine the ultimate axial strength as the output variable while the strength of material and geometry are the input data. A total of 1,730 tests, which is divided in two datasets, is trained and tested in the metamodels. The comparison of the present results and those from the design code (AISC 360-16, Eurocode 4, and AS/NZS 2327) indicates that the 10-fold cross validation results of the super learner methods show very high prediction accuracy in terms of four performance measures.

Keywords - CFST columns, Compressive strength prediction, Machine learning methods, Super learner, Deep learning

1. 서론

콘크리트 충전 강관(CFST: concrete-filled steel tubular) 기둥은 대칭적 형태를 띠는 기하학적 특성과 내부에 충전되어 구속되는 콘크리트로 인해 압축강도 향상, 에너지 흡수 성능 증가, 연성 증가 등의 많은 장점을 가지고 있다^[1]. 또한 강관은 콘크리트 타설 시 거푸집의 역할을 수행하며 그로 인해 수반되는 작업과정을 단축할 수 있으므로 전반적인 공사비용이 절감되어 초고층 빌딩, 장대교량, 지하구조물, 해

양 구조물 등 다양한 곳에 적용되고 있다^[2]. 특히 CFST 기둥의 공칭강도 예측에 대한 주제는 많은 연구자들에 의해 주목을 받아 왔으며 이들의 실험결과가 지속적으로 누적되어 동일 주제의 연구자들에게 큰 밑거름이 되고 있다. 이에 더하여 미국의 AISC 360-16^[3], 영국의 BS 5400^[4], 유럽의 Eurocode 4^[5], 일본의 AIJ 2008^[6], 중국의 GB 50936^[7], 호주/뉴질랜드의 AS/NZS 2327^[8] 등 다양한 기준에서 CFST 기둥 설계에 대한 지침을 제시하고 있다. 앞서 언급한 실험결과들은 각 기준에서 제시하고 있는 제안 식에 의한 결과와 비교하여 기준을 뒷받침, 혹은 보완하는 자료로 사용되고 있으며 특히 고강도 강재 혹은 고강도 콘크리트가 새롭게 개발됨에 따라 각 기준에서 허용하는 범위를 넘어서는 때 기준 개정작업 이전에 유효성을 확인하는 실험을 다방면으로 진행해오고 있다.

CFST 기둥의 설계에서는 공칭강도가 가장 중요한 요인이 되므로 앞에서 언급한 실험결과 이외에도 해석적 모델(analytical model), 유한요소모델(finite element model), 혹은

Note.-Discussion open until October 31, 2022. This manuscript for this paper was submitted for review and possible publication on March 07, 2022; revised on March 23, 2022; approved on March 24, 2022.

Copyright © 2022 by Korean Society of Steel Construction

*Corresponding author.

Tel. +82-2-3408-3286/+82-2-3408-3287

Fax. +82-2-3408-4331

E-mail. kihaklee@sejong.ac.kr/jhlee@sejong.ac.kr

은 역학모델(mechanical model) 등을 구축하여 공칭강도의 예측기법으로 사용하고 있다. 하지만 각 기준에서 제시하는 방법을 따를 경우 실험결과와 상당한 오차를 보이는 경우가 종종 발생한다⁹⁾.

공칭강도 예측의 주요 변수는 원형강관일 때 지름-두께비(diameter-to-thickness ratio) 그리고 높이-지름비(height-to-diameter ratio)이지만 이에 더하여 콘크리트의 압축강도, 강재의 인장강도 등이 변수로 작용한다. 입력변수와 CFST 기둥의 압축강도 간의 관계는 고도의 비선형이므로 계수를 활용한 수식만으로는 설명하기 힘든 거동이 존재하기 마련이다¹⁰⁾. 최근에는 이처럼 고도의 비선형 관계를 기계학습(machine learning)으로 예측하려는 노력이 폭발적으로 증가하고 있다.

한정적인 입력변수들을 사용하여 불확실하고 복잡한 대상 출력값을 예측하기 위해서는 기계학습 기법을 사용하여 메타모델(metamodel) 혹은 대리모델(surrogate model)을 구축할 수 있다. 고전적인 인공신경망(ANN: artificial neural network) 모델을 사용하여 CFST 기둥의 공칭강도를 예측한 기존 연구는 비조밀(non-compact)과 세장기둥¹¹⁾, 단주기둥¹²⁾, 고강도 콘크리트를 사용한 기둥¹³⁾ 등 다양한 대상을 다루어 왔다. 이외에도 GEP(gene expression programming) 기반 모델을 사용한 경우^{14),15)}나 ANFIS(adaptive network-based fuzzy inference system) 기법을 이용한 모델을 사용하여 CFST 기둥의 성능을 예측하는 기존 연구¹⁶⁾ 등을 다수 찾아볼 수 있다. 하지만 컴퓨터과학(computer science) 분야에서 지속적으로 발전된 기계학습 기법을 선보이고 있고 최근 정확성을 높이기 위한 혁신적인 기법들이 개발되고 있음에도 불구하고 이를 강구조 분야에 적용하는 움직임은 극히 소수에 불과하다¹⁷⁾.

본 연구에서는 기계학습의 앙상블 학습 기법(ensemble learning method) 중 하나인 슈퍼 러너(SL: super learner)를 사용하여 CFST 기둥의 압축강도를 예측하는 모델을 제안한다. 기존의 인공신경망 기법에서는 단일 모델을 사용하여 데이터를 훈련하는 방식이지만 앙상블 학습 기법은 다수의 모델을 결정 트리(decision tree) 형태로 결합하여 평가한 후 최적의 모델을 도출한다¹⁸⁾. 슈퍼 러너 기법은 이에 더하여 여러 앙상블 기법을 조합하여 고도의 비선형 관계를 규명하는 최적의 모델을 도출하는 강점을 가지고 있다¹⁹⁾. 제안된 학습 기법은 원형 CFST 기둥 실험결과를 통해 학습되었으며 총 1,730개의 데이터가 사용되었다. 다양한 평가방식을 통해 예측률을 확인하였으며 세 가지 기준(AISC 360-16,

Eurocode 4, AS/NZS 2327)을 통한 결괏값과 비교분석하였다. 제안된 학습모델은 추후 새로운 CFST 변수 데이터를 대입하여 최종적으로 압축강도를 예측할 수 있다.

2. 데이터셋

2.1 원형 CFST 기둥 데이터

본 연구에서는 원형 CFST 기둥을 대상으로 예측모델을 개발하였으며 총 1,730개의 데이터가 사용되었다. 본 데이터는 Goode²⁰⁾와 Denavit²¹⁾의 연구에서 발췌하였으며 해당 연구 또한 타 연구자들의 실험결과를 집대성한 모음이다. 이를 중심축에 하중을 받는 Dataset 1과 편심하중을 받는 Dataset 2로 나누었으며 각각의 개수는 Table 1에서 확인할 수 있다.

Table 1. Experimental specimens of circular CFST columns

Dataset	Loading type	Member type	No. of specimens
1	Concentric	Short column	659
		Slender column	586
2	Eccentric	Beam-column	485

Dataset 1의 short column과 slender column은 각각 $L/D \leq 4$ 와 $L/D > 4$ 의 경우를 의미하며 L 과 D 는 각각 시험편 높이와 원형 강관 단면의 바깥지름을 뜻한다(Fig. 1). Fig. 2에서는 Dataset 1의 L/D 분포를 나타내고 있다. $L/D \leq 4$ 인 경우의 short column은 모두 659개의 데이터수를 가지고 있으며 나머지 $L/D > 4$ 경우인 slender column의 분포는 4초 팻값부터 50까지 다양하게 분포되어있다.

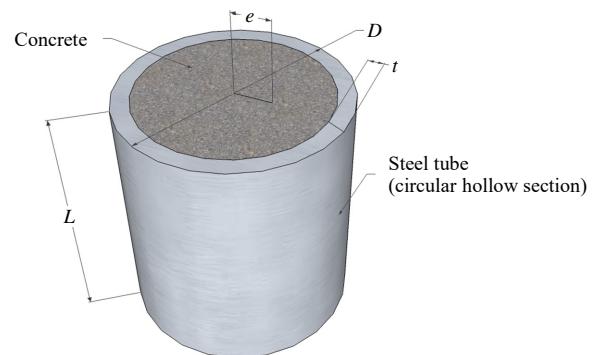


Fig. 1. Circular CFST column

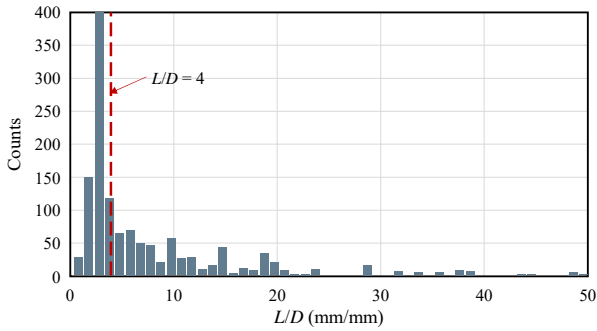


Fig. 2. Histogram of L/D from Dataset 1

Dataset 2는 편심하중으로 인하여 보-기둥(beam-column) 거동을 하는 데이터로 구성되었으며 Fig. 1의 그림에서 중심축으로부터 편심하중의 거리에 해당하는 변수 e 가 입력값에 추가된다. Table 2는 각 변수에 대한 최소, 최대, 평균, 표준편차를 나타내고 있으며 각 변수의 분포는 다음 절에서 상세히 서술한다.

Table 2. Summary of input variables in Dataset 1 and 2

Dataset		D (mm)	t (mm)	L (mm)	e (mm)	F_y (MPa)	f'_c (MPa)
1	Min.	44.5	0.5	152.4	-	178.3	7.6
	Max.	1,020.0	16.5	5,560.0	-	853.0	185.9
	Mean	158.8	4.3	1,055.1	-	335.7	47.5
	Standard deviation	104.8	2.5	1,000.3	-	91.0	30.7
2	Min.	76.0	0.9	284.5	4.0	185.7	18.4
	Max.	600.0	16.0	4,956.0	300.0	517.0	176.8
	Mean	143.3	4.2	1,758.5	39.9	327.7	51.3
	Standard deviation	55.8	2.0	1,042.3	34.3	59.0	26.8

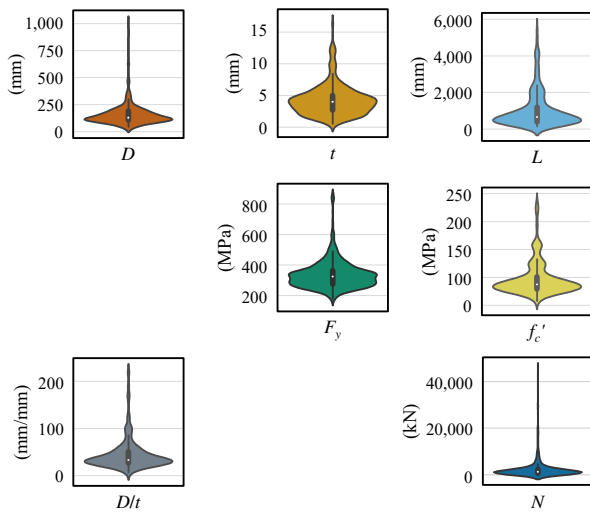


Fig. 3. Violin plots showing the distribution of Dataset 1

2.2 데이터 샘플의 통계적 분석

Fig. 3와 Fig. 4는 각각 Dataset 1과 Dataset 2 변수들의 바이올린 플롯(violin plot)을 나타내고 있다. 바이올린 플롯을 통해 각 입력변수들의 분포, 데이터의 확률밀도 등을 알아볼 수 있다.

Dataset 1은 원형강관의 바깥 지름(D)과 두께(t), 시험편의 높이(L), 강재의 항복강도(F_y), 콘크리트의 압축강도(f'_c), 직경-두께비(D/t)인 총 6개의 입력변수를 가지며 CFST 기둥의 공칭강도(N)를 결괏값으로 도출한다. Dataset 2는 이에 더하여 편심하중의 중심축으로부터의 거리(e)와 편심하중거리-직경비(e/D)를 추가하며 총 8개의 입력변수를 갖는다. 각각의 바이올린 플롯 내부의 검정 선으로부터 중앙값(median), 제1사분위(first quartile, Q1)와 제3사분위(third quartile, Q3)값을 확인할 수 있다. 모든 데이터가 실험을 통해 얻어졌고 여러 연구자들의 결과를 종합한 것이므로 그

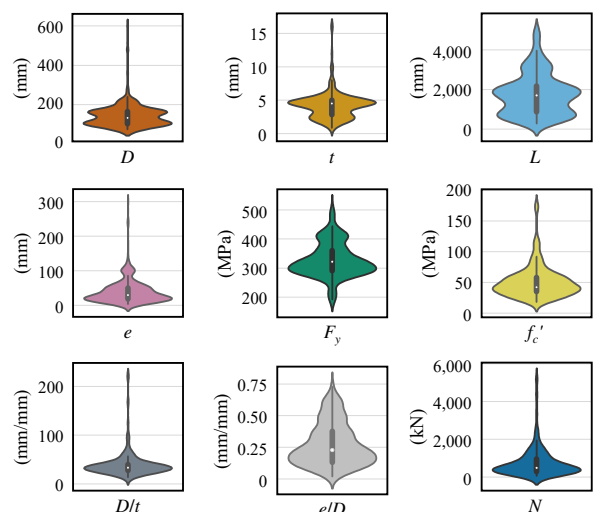


Fig. 4. Violin plots showing the distribution of Dataset 2

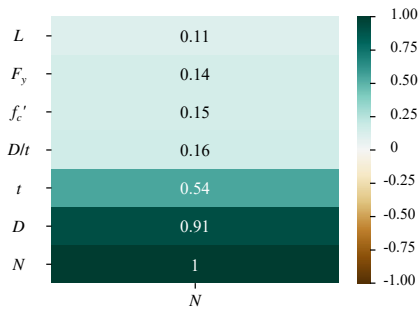


Fig. 5. Correlation matrix heat-map of the features in Dataset 1

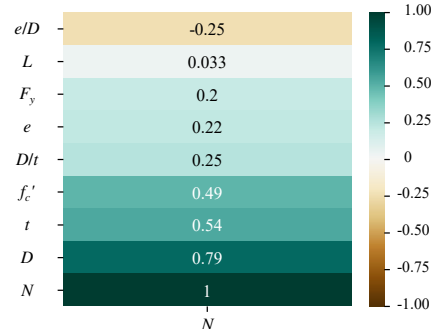


Fig. 6. Correlation matrix heat-map of the features in Dataset 2

분포가 일률적이지 않거나 중앙값을 훨씬 벗어나는 값을 보이는 경우가 있다.

Fig. 5와 Fig. 6는 각각 Dataset 1과 Dataset 2의 입력변수와 출력변수(N) 사이의 상관계수행렬(correlation matrix)을 나타내고 있다. 각 그림의 수치, 상관계수(correlation coefficient)는 두 변수 간의 연관된 정도를 나타낼 뿐 인과관계를 설명하는 것은 아니지만 두 변수 간의 인과관계를 설명하는 것은 아니지만 두 변수 간의 인과관계를 회귀 모델(regression model)을 통해 그 정도와 방향을 알아볼 수 있다. Fig. 5에서 Dataset 1의 경우를 살펴보면 강관의 바깥 지름과 두께에서 큰 상관 정도를 보이고 있다. 이에 더하여 Dataset 2의 보-기둥 구조는 콘크리트의 압축강도가 세 번째로 큰 상관관계를 보이고 있다. 이는 단순히 각각의 변수들과 출력변수 간의 상관 정도를 나타내는 것이므로 제안되는 예측모델을 통해 인과관계를 분석하고자 한다.

3. CFST 기둥을 위한 메타모델

3.1 수퍼 러너 구성을 위한 앙상블 모델

원형 CFST 기둥 데이터를 위한 메타모델을 구축하기 위해 수퍼 러너 기법을 선택하였다. Van der Laan *et al.*^[22]에 의해 제안된 본 기법은 후보 모델군에서 최적의 학습모델을 선택하며 이때 교차검증(cross validation) 기법을 사용하여 우열을 가린다. 수퍼 러너는 앙상블 학습 기법 내에서도 적층(stacking) 기법에 해당하며 다양한 모델을 선택 및 융합하여 예측율을 높이는 방법을 취한다. 본 연구에서는 앙상블 학습 기법 중 랜덤 포레스트(RFR: random forest regression)^[23], AdaBoost(Adaptive Boosting)^[24], GBM(Gradient Boosting Machine)^[25], XGBoost(eXtreme Gradient Boosting)^[26], LightGBM(Light Gradient Boosting Machine)^[27],

CatBoost(Categorical Gradient Boosting)^[28] 총 6가지 기법을 사용하였으며 예측률 상위 3개로 구성된 수퍼 러너 모델과 모든 앙상블 기법을 사용한 수퍼 러너 모델을 구성하여 비교하여 분석하였다.

랜덤 포레스트는 훈련 과정에서 구성한 다수의 결정 트리(tree)로부터 평균 예측치를 얻어 최적의 학습 기법을 도출하는 것이며 AdaBoost는 이에 발전하여 약한 학습기들을 수정할 수 있는 가중치 기법을 적용한 것이다. GBM은 앙상블 알고리즘 중 약한 학습기를 결합하여 강한 학습기를 만드는 부스팅(boosting) 기법의 기본 형태이며 XGBoost는 이에 더하여 병렬 학습을 통해 더욱 빠르게 연산할 수 있도록 개선된 기법이다. LightGBM은 모든 결정 트리를 균형적으로 발전시키지 않고 최대 손실값을 가지는 트리의 노드(node)를 반복분할하면서 예측 오류 손실을 최소화하는 기법이다. 기존의 부스팅 기법이 일반적으로 모든 훈련 데이터를 대상으로 잔차(residual)를 계산하는 반면 CatBoost는 훈련에 사용하는 데이터의 비율을 점차 증가시킴으로써 기존의 부스팅 기법이 갖고 있던 느린 학습 속도와 과적합(over-fitting) 문제를 개선할 수 있다.

3.2 수퍼 러너 모델

n개의 대상 데이터 $O_i = (X_i, Y_i)$ 에 대하여 실제값과 예측값 $\psi(X)$ 을 비교하는 손실(loss) $E[L(O, \psi)]$ 을 최소화하는 회귀 문제는 다음과 같이 구성할 수 있다.

$$\psi_o(X) = \arg \min_x E[L(O, \psi)] \quad (1)$$

k개의 교차검증 모듈을 갖는 테스트 데이터와 검증 데이터를 각각 $T(v)$ ($v = 1, 2, 3, \dots, k$)와 $V(v)$ ($v = 1, 2, 3, \dots, k$)라고 할 때 J개의 학습모델의 구성 $\hat{\Psi}_j$ ($j = 1, 2, 3, \dots, J$)은 다

음과 같이 표현된다.

$$\hat{\Psi}_{j,T(v)}(V(v)), \quad (j = 1, 2, 3, \dots, J) \quad (2)$$

본 연구에서는 총 6개의 학습모델이 사용되었으며 예측률 상위 3개로 구성된 수퍼 러너 모델도 추가로 구성하였다. 모든 앙상블 기법은 적층(stacked)되는 형식으로 수퍼 러너 안에 구성되며 각각의 학습모델로부터 얻은 예측값은 행렬식 $Z = \{\hat{\Psi}_{j,T(v)}(V(v))\}$ 로 구성되며 예측률에 따라 가중벡터 α 를 곱하여 수정된 조합은 아래와 같이 구성된다.

$$m(z|\alpha) = \sum_{j=1}^J \alpha_j \hat{\Psi}_{j,T(v)}(V(v)), \quad \sum_{j=1}^J \alpha_j = 1 \quad (3)$$

실제값 Y 와 예측값 사이의 손실이 최소가 되는 최적의 가중벡터 $\hat{\alpha}$ 를 얻은 후 최종적으로 아래와 같은 수퍼 러너 모델을 얻을 수 있다.

$$\hat{\Psi}_{SL}(X) = \sum_{j=1}^J \hat{\alpha}_j \hat{\Psi}_j(X) \quad (4)$$

Dataset 1과 Dataset 2의 값을 최적의 모델로 구성된 수퍼 러너 모델에 대입하여 효율성을 확인하였으며 다음 장에서 자세히 서술한다.

4. 결과 및 고찰

4.1 평가기법

제안된 회귀모델을 평가하기 위해 10겹 교차검증 횟수를 설정하였으며 평균 제곱근 오차(RMSE: root mean squared error), 평균 제곱 오차(MSE: mean squared error), 평균 절대 오차(MAE: mean absolute error), 결정계수(coefficient of determination, R^2)의 총 4가지 평가기법을 도입하였다. 특히 결정계수 R^2 지표는 예측값(t)이 입력값에 대해 얼마나 설명

력을 갖게 되는지를 의미하는 수치로 1에 가까운 수일수록 양의 상관관계를 의미한다. n 개의 데이터 샘플을 사용하였을 때 4가지 성능 평가 지표의 식은 아래와 같다.

- Coefficient of determination (R^2):

$$R^2 = \left(\frac{n \sum ty - (\sum t)(\sum y)}{\sqrt{n(\sum t^2) - (\sum t)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} \right)^2 \quad (5)$$

- Root mean square error (RMSE)

$$SE = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - y_i)^2} \quad (6)$$

- Mean square error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (t_i - y_i)^2 \quad (7)$$

- Mean absolute error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |t_i - y_i| \quad (8)$$

일반적으로 일정 비율의 학습 데이터와 테스트 데이터로 구분하여 수행하는 경우와 달리 k 겹 교차검증은 전체 데이터셋을 모델 검증에 활용할 수 있는 기법으로 테스트 집합을 순차적으로 바꿔가면서 모든 데이터에 대해 모형의 성과를 측정하는 검증방식이다. 본 연구에서는 10겹 교차검증으로 모델을 검증하였으며 전체 데이터셋을 10개로 분할한 후 각 분할 단위가 순차적으로 검증에 사용된다. 이렇게 함으로써 과적합을 피할 수 있으며 평가에 사용되는 데이터의 편중을 방지할 수 있다.

4.2 Dataset 1

Table 3는 Dataset 1에 대한 결과값을 보여주고 있다. 진한 글씨로 표시한 XGBoost, LightGBM, CatBoost의 결과는 앙상블 모델을 개별적으로 수행하였을 때 나온 결과 중 높

Table 3. Comparison between the performance of SL(super learner) with several ensemble methods (test results for Dataset 1)

Performance indicators	RFR	AdaBoost	GBM	XGBoost	LightGBM	CatBoost	SL-RAGXLC	SL-XLC
R^2	0.982	0.984	0.994	0.996	0.994	0.996	1.000	1.000
MSE	313,801.071	146,937.568	82,751.150	67,990.549	68,725.454	48,002.295	3,699.085	4,135.777
RMSE	513.015	371.535	271.501	244.758	256.215	208.634	57.613	61.600
MAE	211.942	246.093	109.842	112.948	122.462	100.551	19.706	30.701

은 예측률을 보이는 상위 3개의 알고리즘이다. 수퍼 러너는 6개의 모든 앙상블 모델을 조합한 경우 SL-RAGXLC와 상위 3개로 조합된 SL-XLC로 나누었으며 Table 3에서 결과값을 확인할 수 있다.

SL-RAGXLC와 SL-XLC의 경우는 결정계수 1.000의 값을 가지며 높은 예측률을 보였으며 다른 평가기법의 값 또한 개별적인 앙상블 모델을 수행하였을 경우와 비교하였을 때 향상된 양상을 보인다. 특히 모든 앙상블 기법을 조합한 SL-RAGXLC의 경우와 상위 3개 모델을 사용한 SL-XLC의 경우는 대동소이하며 SL-RAGXLC의 경우가 MSE = 3699.085, RMSE = 57.613, MAE = 19.706으로 더 낮은 오차값을 보였다. Fig. 7은 10번의 교차검증 과정을 수행한 R^2 결과값에 대한 박스-위스커 플롯(box-and-whisker plot)을 나타내고 있으며 수퍼 러너 모델들이 가장 높은 예측률을 보임을 알 수 있다.

Fig. 8은 세 가지 기준(AISC 360-16, Eurocode 4, AS/NZS 2327)을 통한 결과값과 SL-RAGXLC 결과값의 비교정규

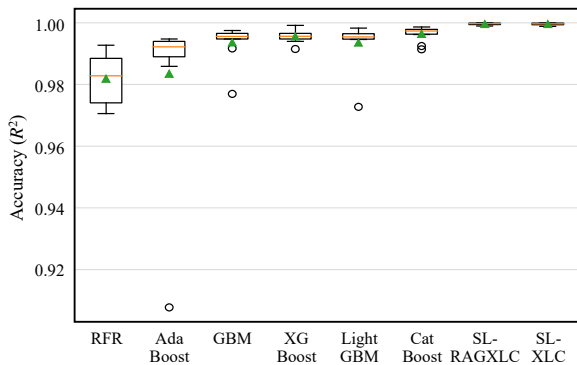


Fig. 7. Box-Whisker plots of the 10-fold cross validation for the super learner methods (SL-RAGXLC/SL-XLC) compared to standalone ensemble models using Dataset 1

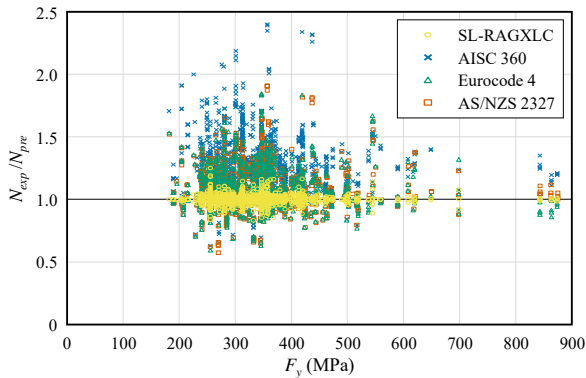


Fig. 9. Effect of steel yield strength on the predictions of SL-RAGXLC and code for Dataset 1

분포 분위수 대조도(Q-Q plot: quantile-quantile plot)를 나타내고 있다. 수퍼 러너(SL-RAGXLC)를 통한 결과값은 모든 범위에 걸쳐 좋은 예측을 보여주고 있다. 이는 각 기준에서 제시하는 식을 통한 결과 분포가 약간의 오차를 보이는 것과 대조적이다. 특히 AISC 360의 경우는 평가해야 할 극한 강도가 높아질수록 오차율이 증가하는 현상을 보이며 10,000 kN 이하의 분포에서는 세 가지 기준의 결과값 모두 약간의 오차분포를 보이는 것을 알 수 있다.

Fig. 9은 강재의 항복응력 분포에 따른 각 결과값의 예측 정확도 분포를 나타내고 있다. 제안된 수퍼 러너 기법을 통한 결과값은 강재의 항복응력 범위와 무관하게 높은 예측 정도를 보이고 있다. 이는 Fig. 10의 콘크리트의 압축응력 분포에 따른 각 결과값의 예측 정확도 분포도에서도 동일하게 확인할 수 있다.

Fig. 11은 Dataset 1 결과값에 대한 총괄적인 비교를 나타내고 있으며 R^2 값에 대한 모든 앙상블 모델의 순위를 쉽게 알 수 있다.

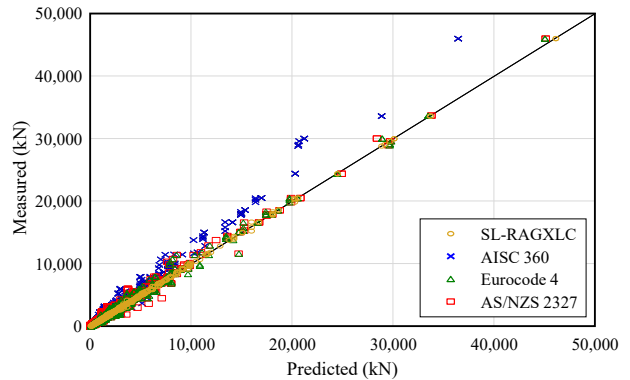


Fig. 8. Correlation (Q-Q plot: quantile-quantile plot) between measured and predicted output of Dataset 1 using SL-RAGXLC model

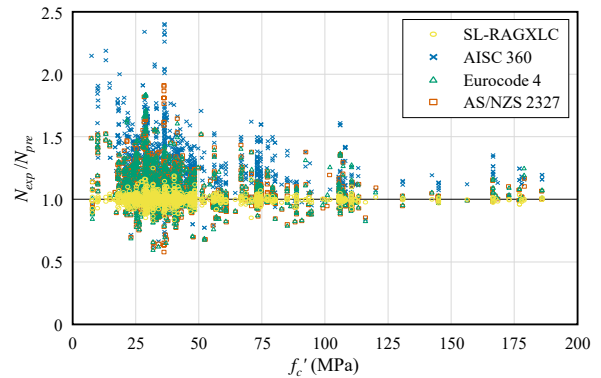


Fig. 10. Effect of concrete compressive strength on the predictions of SL-RAGXLC and code for Dataset 1

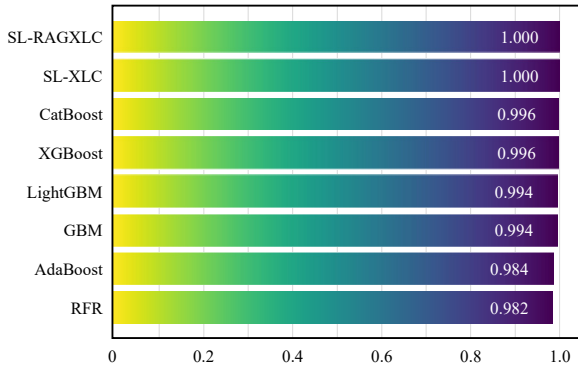


Fig. 11. Performance comparison of ensemble models using correlation coefficient (R^2) with 10-fold cross validation for Dataset 1

4.3 Dataset 2

Table 4는 Dataset 2에 대한 결괏값을 보여주고 있다. 진한 글씨로 표시한 AdaBoost, XGBoost, CatBoost의 결과는 앙상블 모델을 개별적으로 수행하였을 때 나온 결과 중 높은 예측률을 보이는 상위 3개의 알고리즘이다. 슈퍼 러너는 6개의 모든 앙상블 모델을 조합한 경우 SL-RAGXLC와 상위 3개로 조합된 SL-AXC로 나누어 수행하였다.

Table 4를 살펴보면 두 가지의 슈퍼 러너 모두 높은 R^2 값

0.999를 보이는 것을 알 수 있다. Dataset 2의 경우도 Dataset 1의 경우와 마찬가지로 6가지 앙상블 모델을 모두 사용한 SL-RAGXLC의 경우가 가장 낮은 MSE, RMSE, MAE 값을 보였다. Fig. 12는 10번의 교차검증 과정을 수행한 R^2 결괏값에 대한 박스-위스커 플롯을 보여주고 있으며 10번의 반복과정에서 SL-RAGXLC가 가장 높은 R^2 값을 보임을 확인할 수 있다.

Fig. 13은 Dataset 2에 적용한 세 가지 기준(AISC 360-16, Eurocode 4, AS/NZS 2327)을 통한 결괏값과 SL-RAGXLC 결괏값의 비교정규분포 분위수 대조도를 나타내고 있다. 슈퍼 러너(SL-RAGXLC)를 통한 결괏값은 모든 범위에 걸쳐 좋은 예측을 보여주고 있다. Dataset 2의 경우는 AS/NZS 2327의 결괏값이 평가해야 할 극한 강도가 높아질수록 오차율이 가장 크게 증가하는 현상을 보였다. Dataset 1의 경우보다 Dataset 2의 경우가 모든 기준값에 의한 결과가 더 큰 오차율을 보임을 알 수 있다. 이는 편심 정도를 나타내는 e 변수가 추가된 요인과 더불어 보-기동 거동을 하는 데이터 세트의 더 고도화된 비선형 관계가 복합적으로 작용한 것으로 풀이된다.

Fig. 14과 Fig. 15은 각각 강재의 항복응력 분포와 콘크리트의 압축응력 분포에 따른 각 결괏값의 예측 정확도 분포

Table 4. Comparison between the performance of SL(super learner) with several ensemble methods (test results for Dataset 2)

Performance indicators	RFR	AdaBoost	GBM	XGBoost	LightGBM	CatBoost	SL-RAGXLC	SL-XLC
R^2	0.940	0.954	0.932	0.958	0.948	0.963	0.999	0.999
MSE	32,364.409	31,753.936	39,044.021	26,347.094	34,365.872	24,467.605	393.512	415.452
RMSE	173.407	164.654	175.607	147.356	164.014	139.200	18.961	19.554
MAE	83.931	83.340	67.999	68.666	74.252	62.811	7.376	10.043

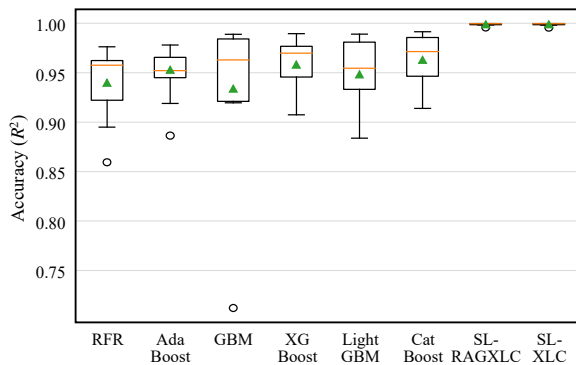


Fig. 12. Box-Whisker plots of the 10-fold cross validation for the super learner methods (SL-RAGXLC/SL-AXC) compared to standalone ensemble models using Dataset 2

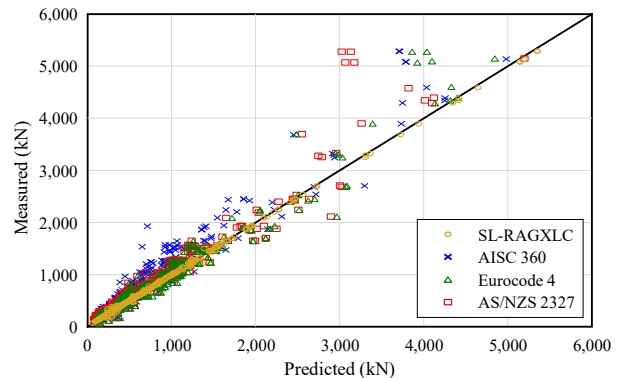


Fig. 13. Correlation (Q-Q plot: quantile-quantile plot) between measured and predicted output of Dataset 2 using SL-RAGXLC model

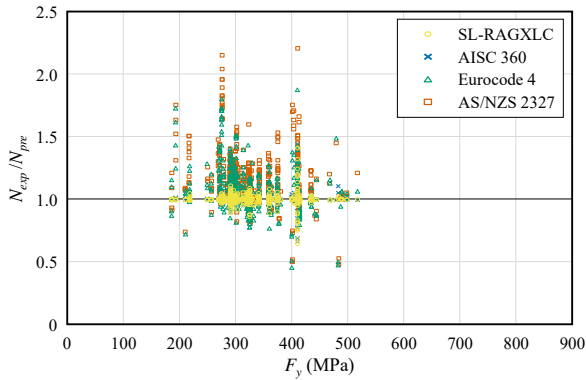


Fig. 14. Effect of steel yield strength on the predictions of SL-RAGXLC and code for Dataset 2

도를 나타내고 있다. 수퍼 러너를 통해 예측한 SL-RAGXLC의 결과는 전 범위 내에서 안정적인 예측정도를 보이고 있으며 이는 세 가지 기준에서 제시하는 식을 통한 결과값이 실제 실험값과의 차이를 보이는 점과 대조적이다. 전체적인 Dataset 2 결과값에 대한 총괄적인 비교는 Fig. 16에 나타내었다. 10겹 교차검증의 테스트값에 대한 R^2 결과를 비교할 수 있으며 수퍼 러너를 사용한 두 경우 모두 높은 예측률을 보이고 있음을 알 수 있다.

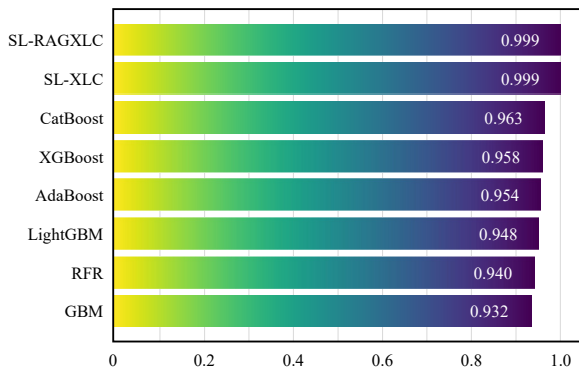


Fig. 16. Performance comparison of ensemble models using correlation coefficient (R^2) with 10-fold cross validation for Dataset 2

5. 결론

본 연구는 기계학습 기법을 사용한 CFST 기둥의 압축강도를 예측하는 모델을 제시하였다. 1,730개의 원형 CFST 기둥 실험결과를 학습 데이터로 사용하였으며 다양한 평가 방식을 통해 예측률을 확인하였다. 수퍼 러너 기법을 사용한 예측모델은 입력변수와 CFST 기둥의 압축강도 간의 고

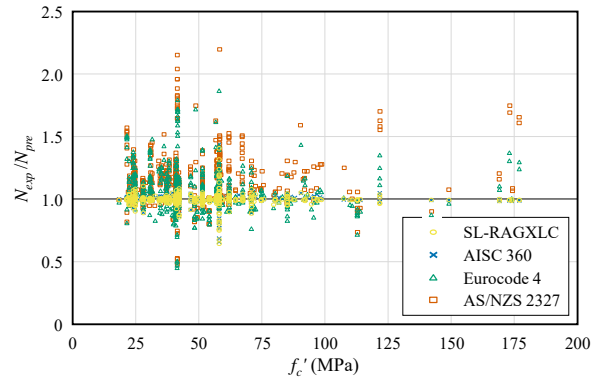


Fig. 15. Effect of concrete compressive strength on the predictions of SL-RAGXLC and code for Dataset 2

도의 비선형 관계를 비교적 정확하게 예측할 수 있음을 확인하였다. 이는 세 가지 기준(AISC 360-16, Eurocode 4, AS/NZS 2327)을 통한 결과값과 비교한 경우에도 우수함을 확인할 수 있었다. 기계학습을 사용한 CFST 기둥의 압축강도 예측모델은 추후 다양한 단면에 대해 확대해 갈 수 있으며 고강도 강재 혹은 고강도 콘크리트가 새롭게 개발되었을 때 그 유효성을 확인하는데 사용할 수 있다. 또한 기준에서 제시하는 식 이외의 경험적 모델(empirical mode)을 본 연구의 결과를 통해 새롭게 제시할 수 있을 것이다.

감사의 글

본 논문은 행정안전부장관의 지진방재내진분야 전문인력 양성사업으로 지원되었습니다.

참고문헌(References)

- [1] Lee, S.-H., Uy, B., Kim, S.-H., Choi, Y.-H., and Choi, S.-M. (2011) Behavior of High-Strength Circular Concrete-Filled Steel Tubular (CFST) Column Under Eccentric Loading, *Journal of Constructional Steel Research*, Elsevier, Vol.67, No.1, pp.1-13.
- [2] Wang, X., Fan, F., and Lai, J. (2022) Strength Behavior of Circular Concrete-Filled Steel Tube Stub Columns Under Axial Compression: A Review, *Construction and Building Materials*, Elsevier, Vol.322, 126144.
- [3] American Institute of Steel Construction (2016) *Specification for Structural Steel Buildings* (ANSI/AISC

- 360-16), USA.
- [4] British Standards Institution (2005) *Steel, Concrete and Composite Bridges: Code of Practice for Design of Composite Bridges* (BS 5400-1: 2005), UK.
- [5] European Committee for Standardization (2004) *Eurocode 4: Design of Composite Steel and Concrete Structures, Part 1-1: General Rules and Rules for Buildings* (EN 1994-1-1: 2004), Belgium.
- [6] Architectural Institute of Japan (2008) *Recommendations for Design and Construction of Concrete Filled Steel Tubular Structures* (2nd Ed.), AIJ, Japan (in Japanese).
- [7] Ministry of Housing and Urban-Rural Development (2014) *Technical Code for Concrete-Filled Steel Tubular Structures* (GB 50936-2014), China.
- [8] Standards Australia (2017) *Composite Structures – Composite Steel-Concrete Construction in Buildings* (AS/NZS 2327: 2017), Australia.
- [9] Lee, H.J., Park, H.G., and Choi, I.R. (2015) Axial Load Test on Rectangular CFT Column Using High-Strength Steel and Slender Section, *Journal of Korean Society of Steel Construction*, KSSC, Vol.27, No.2, pp.219-229 (in Korean).
- [10] Lee, C.H., Kang, K.Y., Kim, S.Y., and Koo, C.H. (2013) Review of Structural Design Provisions of Rectangular Concrete Filled Tubular Columns, *Journal of Korean Society of Steel Construction*, KSSC, Vol. 25, No.4, pp.389-398 (in Korean).
- [11] Ahmadi, M., Naderpour, H., and Kheyroddin, A. (2019) A Proposed Model for Axial Strength Estimation of Non-Compact and Slender Square CFT Columns, *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, Springer, Vol.43, No.1, pp.131-147.
- [12] Ren, Q., Li, M., Zhang, M., Shen, Y., and Si, W. (2019) Prediction of Ultimate Axial Capacity of Square Concrete-Filled Steel Tubular Short Columns Using a Hybrid Intelligent Algorithm, *Applied Sciences*, MDPI, Vol.9, No.14, 2802.
- [13] Ding, F.X., Luo, L., Wang, L., Cheng, S., and Yu, Z.-W. (2018) Pseudo-Static Tests of Terminal Stirrup-Confined Concrete-Filled Rectangular Steel Tubular Columns, *Journal of Constructional Steel Research*, Elsevier, Vol.144, pp.135-152.
- [14] Güneyisi, E.M., Gültekin, A., and Mermerdaş, K. (2016) Ultimate Capacity Prediction of Axially Loaded CFST Short Columns, *International Journal of Steel Structures*, KSSC, Vol.16, No.1, pp.99-114.
- [15] Mansouri, I., Chacón, R., and Hu, J.W. (2017) Improved Predictive Model to the Cross-Sectional Resistance of CFT, *Journal of Mechanical Science and Technology*, Springer, Vol.31, No.8, pp.3887-3895.
- [16] Basarir, H., Elchalakani, M., and Karrech, A. (2019) The Prediction of Ultimate Pure Bending Moment of Concrete-Filled Steel Tubes by Adaptive Neuro-Fuzzy Inference System (ANFIS), *Neural Computing and Applications*, Springer, Vol.31, No.2, pp.1239-1252.
- [17] Stulp, F., and Sigaud, O. (2015) Many Regression Algorithms, One Unified Model: A Review, *Neural Networks*, Elsevier, Vol.69, pp.60-79.
- [18] Ganaie, M.A., Hu, M., Tanveer, M.M., and Suganthan, P.N. (2021) Ensemble Deep Learning: A Review (preprint), *arXiv*, arXiv:2104.02395.
- [19] Young, S., Abdou, T., and Bener, A. (2018) Deep Super Learner: A Deep Ensemble for Classification Problems, *Advances in Artificial Intelligence (Proceedings of Canadian Conference on Artificial Intelligence)*, Springer, Canada, pp.84-95.
- [20] Goode, C.D. (2008) Composite Columns – 1819 Tests on Concrete-Filled Steel Tube Columns Compared with Eurocode 4, *The Structural Engineer*, The Institution of Structural Engineers, Vol.86, No.16, pp.33-38.
- [21] Denavit, M. (2005) *Steel-Concrete Composite Column Database*, <http://mark.denavit.me/Composite-Column-Database>.
- [22] Van der Laan, M.J., Polley, E.C., and Hubbard, A.E. (2007) Super Learner, *Statistical Applications in Genetics and Molecular Biology*, Walter de Gruyter, Vol.6, No.1.
- [23] Breiman, L. (1996) Bagging Predictors, *Machine Learning*, Springer, Vol.24, No.2, pp.123-140.
- [24] Freund, Y., and Schapire, R.E. (1996) Experiments with a New Boosting Algorithm, *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning (ICML '96)*, Morgan Kaufmann Publishers, USA, pp.148-156.
- [25] Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine, *Annals of Statistics*, Institute of Mathematical Statistics, Vol.29, No.5, 1189-1232.
- [26] Chen, T., and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22nd ACM SIGKDD International Conference on Know-*

- ledge Discovery and Data Mining (KDD 2016)*, Association for Computing Machinery, USA, pp.785-794.
- [27] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *Advances in Neural Information Processing Systems* (*Proceedings of the 31st conference on Neural Information Processing Systems*), Neural Information Processing Systems, USA, pp.3146-3154.
- [28] Dorogush, A.V., Ershov, V., and Gulin, A. (2018) CatBoost: Gradient Boosting with Categorical Features Support (Preprint), *arXiv*, arXiv:1810.11363.

요 약 : CFST(concrete-filled steel tubular) 기둥은 강관과 내부 콘크리트 재료 간의 복잡한 메커니즘으로 인하여 공칭강도 예측을 위해 더 진보한 회기 모델이 요구된다. CFST 기둥의 공칭강도 예측을 위해 다양한 앙상블 기법, 즉 랜덤 포레스트(RFR: Random Forest Regression), AdaBoost(Adaptive Boosting), GBM(Gradient Boosting Machine), XGBoost(eXtreme Gradient Boosting), LightGBM(Light Gradient Boosting Machine), CatBoost(Categorical Gradient Boosting) 기법을 결합한 수퍼 러너 모델이 제안되었다. 총 1,730개의 실험 데이터가 사용되었으며 재료 특성값과 기하학 치수가 입력변수로 사용되었다. 제안된 모델을 통해 예측된 결과를 비교하기 위해 AISC 360-16, Eurocode 4, AS/NZS 2327 세 개의 기준의 값과 비교하였다. 10경 교차검증을 통해 얻은 수퍼 러너 기법의 예측값은 매우 높은 예측률을 보였다.

핵심용어 : CFST 기둥, 압축강도 예측, 기계학습 기법, 수퍼 러너, 딥 러닝
