Integrating automated machine learning and interpretability analysis in architecture, engineering and construction industry: A case of identifying failure modes of reinforced concrete shear walls

Dong Liang[1] and Fan Xue[2*]

This is the peer-reviewed post-print version of the paper:
Liang, D. & Xue, F. (2023). Integrating automated machine learning and interpretability analysis in architecture, engineering and construction industry: A case of identifying failure modes of reinforced concrete shear walls. *Computers in Industry*, 145, 103883. Doi: 10.1016/j.compind.2023.103883
The final version of this paper is available at https://doi.org/10.1016/j.compind.2023.103883

Abstract

Machine learning (ML) has been recognized by researchers in the architecture, engineering, and construction (AEC) industry but undermined in practice by (i) complex processes relying on data expertise and (ii) untrustworthy 'black box' models. As a result, ML results of complex non-linear AEC problems, such as the failure mechanism of reinforced concrete (RC) shear walls, are not comparable with empirical and mechanics-based models. This paper aims to integrate automated ML (AutoML) and interpretability analysis to study the failure mechanism of RC shear walls. In this study, we collected a dataset of 351 comprehensive samples for the failure mode identification of RC shear walls. First, the AutoML model trained using the dataset outperformed a set of conventional ML methods in terms of the F_1 accuracy score. Then, three model-agnostic interpretability analysis methods confirmed the trustworthiness of the AutoML model. The contribution of this paper is three-fold. First, AutoML sheds light on the automatic identification of failure modes of RC sheer walls. Second, the interpretability analysis can validate 'black-box' ML models against long-established domain knowledge in solving non-linear AEC problems. Third, for AEC industrial practitioners, the whole process is automatic, accurate, less reliant on data expertise, and interpretable.

Highlights

- Integration of automated Machine Learning (AutoML) and interpretable analysis for accurate and trustworthy ML
- A comprehensive dataset of 351 tests of failure mode identification of RC shear walls
- A 'black box' AutoML model outperformed five conventional ML algorithms in terms of F_1 accuracy score
- Complementing local and global interpretability analysis methods for explaining and validating the AutoML model
- Trustworthiness of the AutoML model confirmed with domain knowledge via interpretability analysis results

[1] Dong LIANG, MSc, PhD student. leodong@connect.hku.hk , : https://orcid.org/0000-0002-2918-7898
Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; Email:

[2] Fan XUE, PhD, Assistant Professor. xuef@hku.hk , : http://orcid.org/0000-0003-2217-3693
Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; Email:
[*]: Corresponding author, Tel: +852 3917 4174, Fax: +852 2559 9457; Email: xuef@hku.hk

Keywords

Machine learning; Interpretability analysis; RC shear walls; Failure mode identification; AutoML

1 Introduction

Architecture, engineering, and construction (AEC) researchers have been increasingly interested in complex non-linear problems (Su et al. 2021). In civil engineering, for example, traditional physics-based models often cannot achieve accurate results and efficiency at the same time. A typical problem is the failure mode identification of reinforced concrete (RC) shear walls, a vital lateral-force-resisting component for buildings in wind-prone or earthquake-prone areas (Farrar et al. 1993). The failure mechanism, involving various geometrical configurations and material parameters, is too complex to formalize empirical or mechanics-based models. The failure mode identification of RC shear walls used to depend on computation-intensive and time-consuming continuum-based finite element models (Dashti et al. 2014); the high costs restricted its capability in many cases, such as building vulnerability analysis and regional risk assessment (Mangalathu et al. 2020).

Machine learning (ML) has been used to analyze AEC problems (Singh 2021) due to the data availability (Vadyala et al. 2022). In ML, target variables are predicted using data science assumptions without references to rigorous physical theoretical inference (Montáns et al. 2019). Example applications are structural behavior prediction (Luo & Paal 2022) and structural health monitoring (Zhang et al. 2020). In fact, advanced ML techniques provide possible alternatives to physics-based models for computationally demanding numerical modeling, as well as dangerous and expensive experimental testing (Deka 2019).

However, ML applications in AEC have lagged behind expectations (Deka 2019) due to two groups of possible reasons (Kuwajima et al. 2020): (i) complex data-expertise-reliant modeling process and (ii) untrustworthy 'black box' models (Feng et al. 2021). First, developing a high-quality ML model is time-consuming and requires multiple steps and expert knowledge (Hidalgo-Mompeán et al. 2021). Furthermore, most advanced ML methods, such as artificial neural network (ANN), support vector machines (SVM), bagging, boosting, and deep learning, are often 'black box' models, as opposed to 'white box' models based on mechanics or empirical regression (Feng et al. 2021). The missing or uninterpreted step-by-step details in these 'black box' ML models lead to untrustworthiness for industrial practitioners and regulators who are sensitive to risks in healthcare, legislation (e.g., European Union's GDPR), and critical engineering (Linardatos et al. 2020).

Two new emerging technologies, i.e., automated ML (AutoML) and interpretability analysis, have the potential to solve the two issues in complex AEC problems. AutoML is a technique of automating the time-consuming and expertise-reliant procedures in selecting and training ML models; it has been successfully applied in many science and engineering areas such as medical science (Tran et al. 2021; Anwar 2021), public service (Li et al. 2019), finance (Vakhrushev et al. 2021), materials (Bangaru et al. 2019), and the construction domain (Zhao

et al. 2023). Interpretability analysis refers to developing human-readable explanations to facilitate practitioners' comprehension of why an ML model makes certain decisions or predictions (Molnar 2020; Guo et al. 2022).

This study aims to integrate AutoML and interpretability analysis to study failure mode identification of RC shear walls. This research involves a dataset of 351 shear wall failure samples collected from the literature, 15 low-level ML algorithms in AutoML, and three interpretability analysis methods for validating the trained AutoML model. The contribution of this paper is three-fold. First, AutoML sheds light on the automatic identification of the failure modes of RC sheer walls. Second, interpretability analysis can validate the 'black-box' models with long-established AEC domain knowledge. Third, for AEC industrial practitioners, the whole process is automatic, accurate, less reliant on data expertise, and interpretable.

2 Overview of AutoML and interpretability analysis

2.1 AutoML

AutoML is the process of automating the time-consuming and iterative procedures involved in building ML models (Feurer et al. 2020). Figure 1 shows the full process of developing a high-quality ML model. Traditionally, appropriate ML algorithms, extensive computational resources, time, and domain expertise are essential for training or building an ML model (Gijsbers et al. 2019). Even professional data scientists may be confused in each phase by many factors, such as unbalanced class, parameter noise, and selection of performance measures. As a result, there is an increasing interest in automating and democratizing the processes in Figure 1.

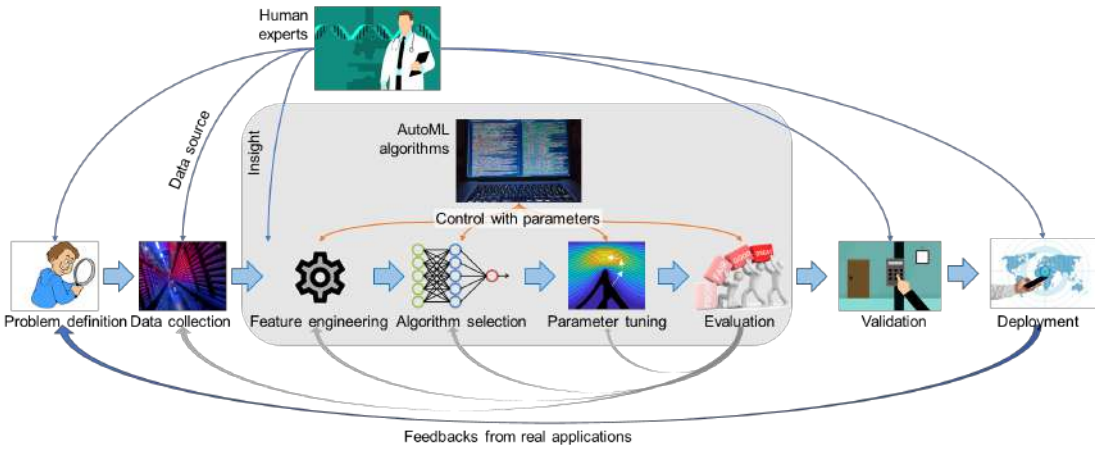


Figure 1 Workflow of machine learning, where the gray box indicates the scope of AutoML

AutoML, as shown in the gray area in Figure 1, automatically completes up to four tasks, including feature engineering, algorithm selection, parameter tuning, and evaluation (Elshawi et al. 2019). Thus, AutoML is more efficient and productive while maintaining the ML model quality. For example, advanced AutoML algorithms outperformed human experts in five out of six construction management datasets (Zhao et al. 2023). In the software market, there exist commercial AutoML platforms, such as Google Cloud AutoML and Microsoft Azure, and open-source libraries, such as Auto-sklearn, MLBox, and TPOT (Elshawi et al. 2019). Yet, the 'black box' problem of ML models remains regardless of the use of AutoML;

to a certain extent, an AutoML model is even harder to explain due to the sophisticated parameter controls in AutoML algorithms.

2.2 Interpretability analysis and applications in civil engineering

Interpretability analysis aims to explain the rationale behind the decisions and outcomes of a computerized model so that a human can comprehend it (Gilpin et al. 2018). There are two categories of interpretability analysis: model-specific and model-agnostic (Molnar 2020). Model-specific interpretation tools are limited to intrinsically interpretable ML models, such as a decision tree model of project-related waste generation behaviors (Yang et al. 2021). In comparison, a model-agnostic interpretation can be applied to any ML model after the training process. Furthermore, model-agnostic interpretable methods can be divided into two groups: global methods and local methods. Global model-agnostic methods describe the average behavior of an ML model (Friedman 2001), while local model-agnostic methods, such as local surrogate (LIME) (Ribeiro et al. 2016), explain a subset of prediction results.

Model-agnostic interpretability analysis has been applied to ML models in the AEC industry in recent years. Table 1 summarizes the application problems, ML task, selected interpretability analysis methods, sample size, contribution, and limitations in seven related studies. In short, existing studies focused on the global model-agnostic interpretability of the regression model's partial dependent plot (PDP) and coalitional-game-theory-based SHapley Additive exPlanations (SHAP) are popular methods. Yet, disadvantages were observed as well. For example, PDP was criticized for the feature interaction that can produce wrong interpretation results (Molnar 2020). SHAP has three limitations in unintuitive feature attributions, possibly biased and misleading interpretations, and unclear estimation thresholds. Furthermore, local model-agnostic methods received less attention. Thus, there is a need to gauge and compare global and local interpretability analysis methods. In this paper, two global model-agnostic interpretation methods: PDP and ALE, and one local method, LIME, are studied for the failure mode identification problem.

Table 1 List of recent model-agnostic interpretability work of ML models in civil engineering

Targets of ML	ML task	Local / Global*	Tools#	Source	Sample size	Contribution	Limitations
Ultimate load of rectangular CFST columns	Regression	G	PD analysis	Le & Phan (2020)	99	PD analysis of each input variable' influence on the ultimate load	1) PD analysis ignores feature dependence.
Shear Strength Prediction of Squat RC Walls	Regression	L & G	SHAP	Feng et al. (2021)	434	Coalitional game-theory based SHAP to explain predictions from both local and global aspects	1) TreeSHAP (a model-agnostic subclass of SHAP method) leading to unintuitive attributions, 2) Possibly problematic explanations by SHAP, 3) Unanswered thresholds by the SHAP.
Ground motion characteristics (PGV and PGA)	Regression	L & G	SHAP	Somala et al. (2021)	4,000	SHAP methods to explain the influence factors for estimating PGA&PGV	Ditto.

Auxetic behavior of cementitious cellular composites	Regression	G	SHAP	Lyngdoh et al. (2022)	850	SHAP to analyze the influence of each input variable on the Poisson's ratios	1) No local interpretability analysis involved, 2) Lack validation of interpretation with expert knowledge.
Magnitude of moment capacity at failure; and the failure mode/causes	Regression and Classification	G	PDP, SHAP, surrogate, & feature interaction	Naser et al. (2021)	103	An explainability case study on reinforced concrete (RC) beams strengthened with fiber-reinforced polymer (FRP) composite laminates	1) Many interpretable methods without cross-references, 2) Feature interaction not considered in other methods such as PDP, 3) No local interpretability analysis involved.
Structural performance of GFRP elastic gridshells	Regression	G	PDP, ALE, & SHAP	Kookalani et al. (2022)	400	Three interpretable methods to analyze the influences of variables on the stress prediction	1) No local interpretability analysis involved.
Seismic response of R/C buildings	Regression	G	PDP & SHAP	Demertzis et al. (2023)	5,850	Use two methods to analyze the influence of Housner intensity on the maximum interstory drift ratio	1) Lacking of comprehensive interpretability analysis on all parameters, 2) No local interpretability analysis involved.

*: L: Local; G: Global.
#: PD: Partial dependence; SHAP: SHapley Additive exPlanations; PDP: PD plot; ALE: accumulated local effects.

3 The problem of failure mode identification of RC shear walls

RC shear walls are the primary lateral load-carrying structure elements in many high-rise buildings due to superior high lateral strength and stiffness (Paulay 1975). For the desired ductile failure modes, RC walls are designed to have a 'strong shear and weak flexure,' which is commonly accomplished by restricting the axial compressive force ratio and adding boundary components to the structure (Lefas et al. 1990). According to the aspect ratio (shear span length divided by wall length), shear walls are typically classified as either squat (low-rise) or slender (tall, high-rise) (Birely 2012). A ductile failure mechanism characterized by flexural yielding toward the base is more likely to occur in slender walls. Squat walls are more prone to a nonductile shear-controlled failure mechanism, which is best defined as an abrupt reduction of stiffness and strength under seismic loads. According to ASCE/SEI 41-17 (ASCE 2017), walls with an aspect ratio larger than 3.0 (often controlled by flexure) are regarded as slender. In contrast, those with an aspect ratio less than 1.5 (mostly controlled by shear) are considered short and squat. The behavior of walls having an aspect ratio between 1.5 and 3.0 is influenced by both shear and flexure. It is also suggested in FEMA 306 (ATC 1998) that well-designed slender shear walls (with aspect ratios of more than 3.0) are more likely to experience ductile flexural failure.

Ductile flexural failure is caused by concrete crushing or a longitudinal reinforcement fracture at the plastic hinge zone, as shown in Figure 2(I). Generally, flexural failure is uncommon in squat shear walls, particularly those with aspect ratios below 1.0. However,

depending on the reinforcing characteristics, flexural failure is possible in such walls, and shear failure may also be seen. The nonductile shear failure mechanism of squat shear walls was divided into three categories by Paulay and Priestley (1992) sliding shear failure, diagonal tension failure, and diagonal compression failure, as shown in Figures 2(II) and 2(III). The presence of one or more corner-to-corner diagonal cracks, as seen in Figure 2(II), is diagnostic of diagonal tension failure (T), which may happen when a wall does not have enough horizontal shear reinforcement. The concrete crushes under this compression develop widespread crack patterns in diagonal compression failure mode (C), as illustrated in Figure 2(III). Sliding shear failure (S) may be caused by two different things: (1) several major cracks at the wall base and (2) the buckling of rebars and the crushing of concrete in a narrow band along the base of the wall, as illustrated in Figure 2(IV), following considerable flexural reinforcement yielding. In RC design and retrofitting, designers and engineers are more concerned about the ductile or non-ductile failure mode of shear walls than detailed classification. Therefore, this study divides the classification work into two phases: the first stage, classifying ductile and non-ductile failure modes, and the second stage, classifying three detailed non-ductile failure modes; the first stage often satisfies most engineering needs.

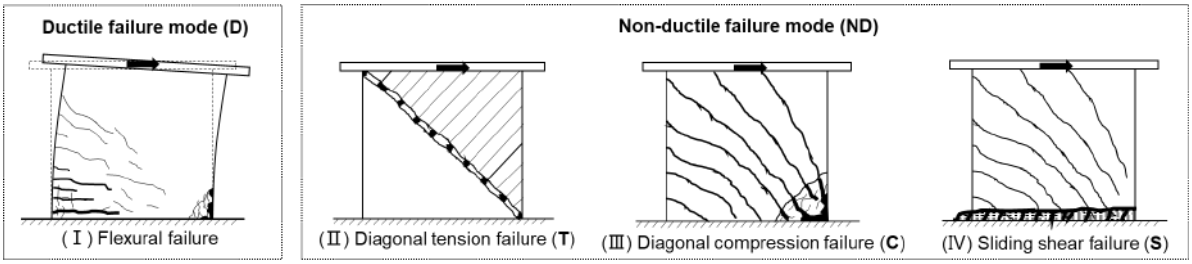


Figure 2 Typical failure modes of RC shear walls, where dashed boxes indicate the first-stage classification

4 Research methods

Figure 3 briefs the three-step flowchart: dataset preparation, AutoML using Auto-sklearn, and interpretability analysis in failure mode identification of the RC shear wall in this study.

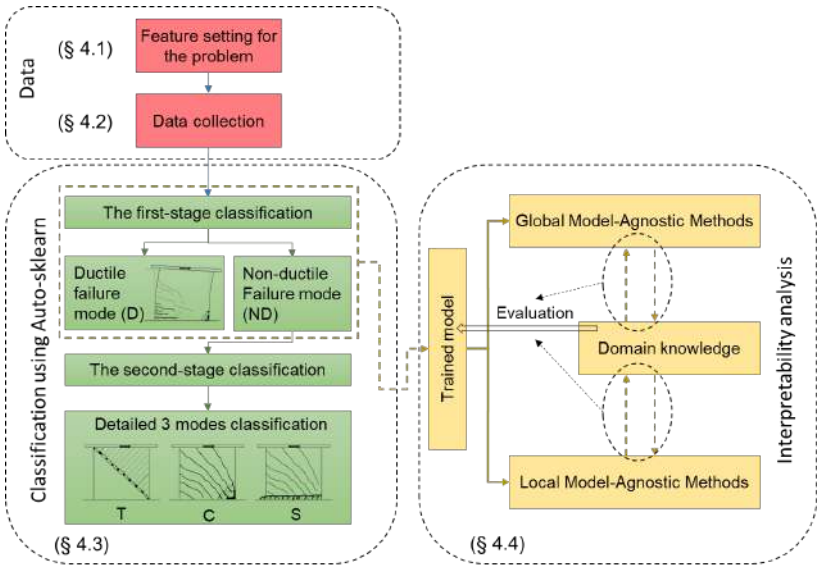


Figure 3 The proposed workflow for failure mode identification of RC shear wall

4.1 Feature settings

Since there are few studies on the failure mode identification of RC shear walls using data-driven methods, this study mainly refers to national regulations about design criteria of structural concrete, including JGJ 3-2010 (China code) (MOHURD 2010), ACI 318-19 (U.S. code) (ACI 2019), EC-8 (Europe code) (CEN TC-250 2004), mechanics theories (Vecchio & Collins 1986), and an existing study by (Mangalathu et al. 2020) to set up features. Among these existing references, a total of four types of features, material property ($f_{y,vw}$, $f_{y,hw}$, $f_{y,cb}$, f_c), reinforcement details (ρ_{vw}, ρ_{hw}, ρ_{vb}, ρ_{hb}), geographic details (l_w/t_w, M/Vl_w, A_b/A_g, CS), and load (P/f_cA_g), were considered to influence the seismic behavior of RC shear wall designs. Specifically, 13 features were selected for classifying the failure mode of RC shear walls, and a description of each feature is given in Table 2. The study by Salonikios et al. (1999) described more details on the sensors and methods for measuring these features in laboratory tests. These 13 features are also considered essential to influence the seismic behavior of other load-bearing structures in many studies related to data-driven structure behavior prediction (Sujith Mangalathu 2018; Hosein Naderpour 2021).

Table 2 Feature setting (excluding classification target) for the problem

Group	Symbol	Unit	Feature description
Material property	$f_{y,vw}$	MPa	Yield stresses of vertical bars in web
	$f_{y,hw}$	MPa	Yield stresses of horizontal reinforcement in web
	$f_{y,cb}$	MPa	Yield stress of confinement reinforcement in boundaryelement
	f_c	MPa	Concrete compressive strength
Reinforcement details	ρ_{vw}	10^3	Web vertical reinforcement ratio
	ρ_{hw}	10^3	Web horizontal reinforcement ratio
	ρ_{vb}	10^3	Boundary region vertical reinforcement ratio
	ρ_{hb}	10^3	Boundary region horizontal reinforcement ratio
Geometric characteristics	l_w/t_w	10^3	Proportion of length and thickness
	M/Vl_w	10^3	Aspect ratio
	A_b/A_g		Proportion of the area of the boundary element and the gross sectional area
	CS		Cross-section type, i.e., Rectangular (R), Barbell (B), and Flanged (F)
Load	P/f_cA_g		Axial load ratio

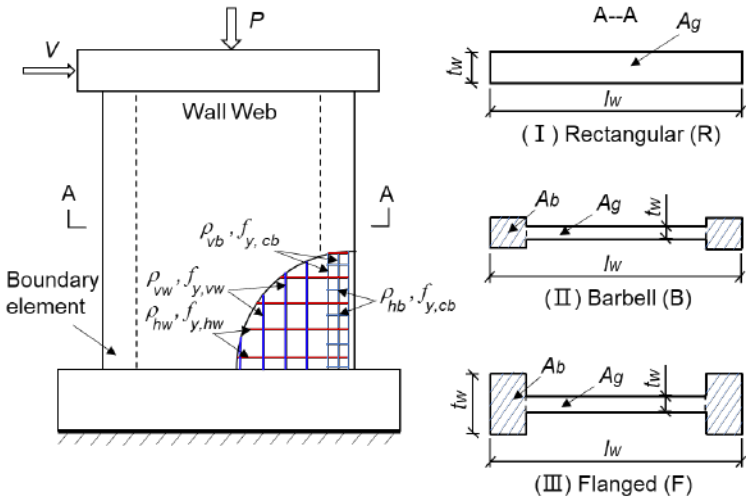


Figure 4 Schematic diagram of RC shear wall tests with three types of cross-sections

4.2 Data collection

In the process of collecting data, the data needs to first meet the requirement of containing all 13 features listed in Table 2. Some extra rules are followed to reduce the impact of irregular data for the main purpose of this study. Specifically, the cross sections of all selected shear walls are one-story, one-bay, symmetric, and without openings. In addition, all selected walls have continuous longitudinal reinforcement without lap splice, as well as deformed and straight reinforcement. Finally, the aspect ratio (M/Vl_w or h_w/l_w) is considered the dominant feature for the seismic behavior of RC shear walls according to the ASCE/SEI 41-17 (ASCE 2017). According to the aspect ratio, RC shear walls are conventionally classified into slender walls (aspect ratio > 3.0), moderate walls (1.5 ≤ aspect ratio ≤ 3.0), and squat walls (aspect ratio < 1.5). Therefore, the data collected in this study equally covers all three types of walls to ensure the applicability of the proposed method.

A total of 351 records were collected from laboratory tests reported in the literature. For example, the first 242 records were from the ACI 445B Shear Wall Database (Usta et al. 2017), with the criterion that each data record covers all features in Table 2. The rest of the 109 records were also collected from published experiments, including Chile (Massone et al. 2009), Canada (Abdulridha & Palermo 2017; Mohamed et al. 2014; Palermo et al. 2002), China (Chiou et al. 2004; Deng et al. 2008; Hwang et al. 2004; Jiang et al. 2013; Kuang & Ho 2008; Lu et al. 2018; Ren et al. 2018; Su & Wong 2007; Wen-yuan et al. 2018; Zhou et al. 2010), the United States (Terzioglu et al. 2018), Australia and New Zealand (Dabbagh 2005; Gebreyohaness et al. 2014; Tripathi et al. 2019; Yanez et al. 1991), Japan (Kabeyasawa & Matsumoto 1992; Kimura et al. 1996; Nakamura et al. 2009; Tokunaga & Nakachi 2012), Slovenia (Tomazevic et al. 1996), and Greece (Christidis & Trezos 2017). The existing open-source library—ACI 445B Shear Wall Database was also used by Mangalathu et al. (2020) for verifying the feasibility of identifying failure modes of shear walls by ML methods. A total of eight algorithms—Naïve Bayes, K-Nearest Neighbors, Decision Tree, Random Forest, AdaBoost, XGBoost, LightGBM, CatBoost, and XGBoost—are used, and the highest accuracy was 86% from Random Forest (Mangalathu et al. 2020). The quality and reliability of existing experiment-based data were initially confirmed.

Table 3 lists an excerpt of the collected dataset, while the full data table is available in the supplemental materials of this paper. Figure 5 briefs the distributions and statistical details of the dataset, respectively. The dataset has a wide range of distribution, and common engineering values are covered. Figure 5 (m) shows that three types of cross-sections accounted for 69.2% (R), 18.9% (B), and 11.9% (F), respectively. The distribution of failure modes for two-stage classification is shown in Figure 5(n)(o).

Table 3: Excerpt of the collected dataset with 351 records of RC shear walls failure tests

No.	*fy,vw*	*fy,hw*	M/Vl_w	…	*CS*	…	Failure mode (1st)	Failure mode (2nd)	Source
1	585	610	1.00	…	R	…	ND	S	(Salonikios et al. 1999)
2	585	610	1.00	…	R	…	ND	S	(Salonikios et al. 1999)
3	610	610	1.00	…	R	…	ND	S	(Salonikios et al. 1999)
…	…	…	…	…	…	…	…	…	…
351	345	345	2.5	…	R	…	ND	S	(Zhou et al. 2010)

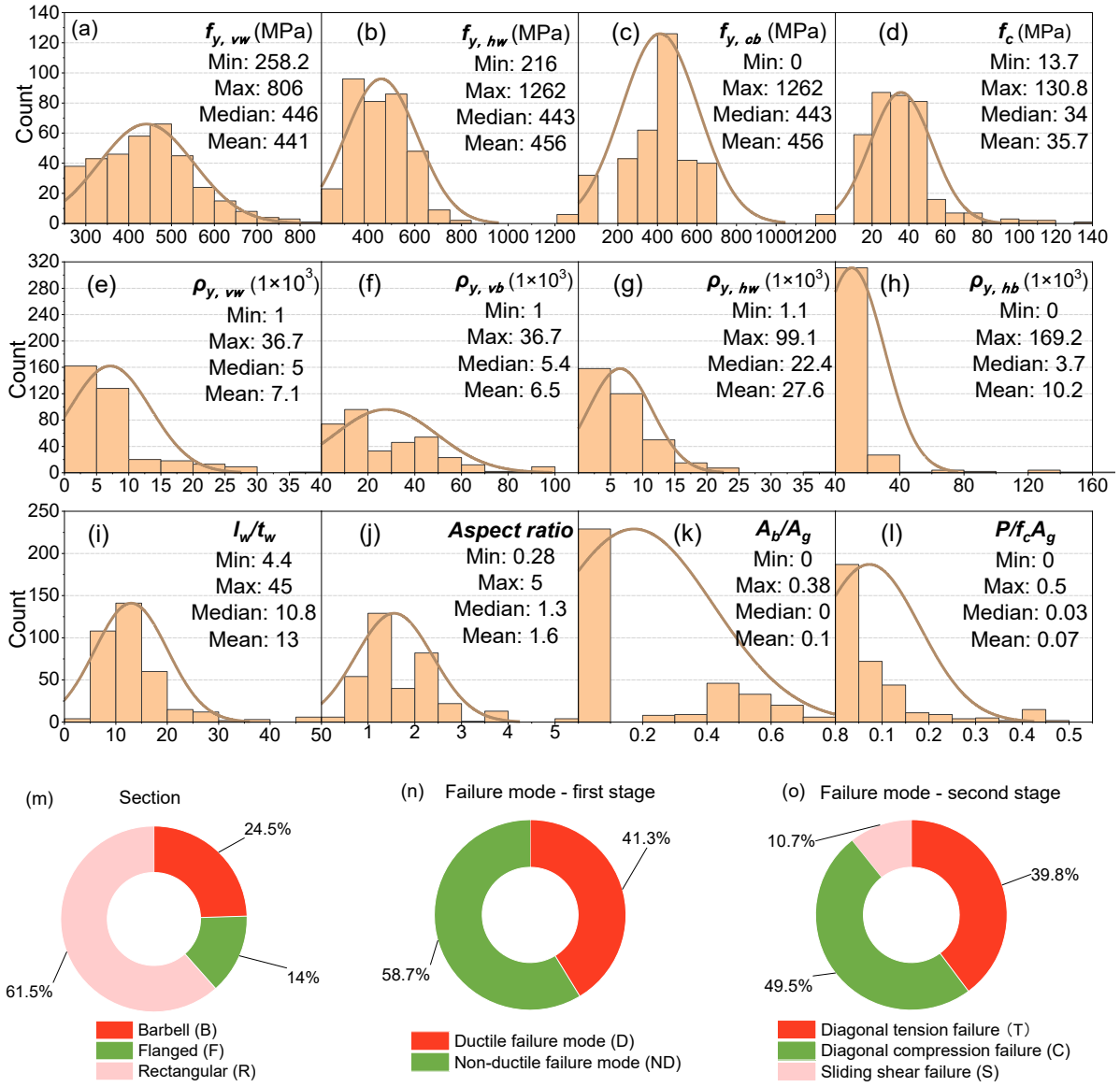


Figure 5: The distributions of the features in the dataset. (a) – (l): Numeric features; (m) – (o) categorical features and targets

Furthermore, Figure 6 shows the matrix of Pearson's linear correlations between the features. Statistically, the circles in Figure 6 indicate one close relation among $f_{y,hw}$, $f_{y,cb}$, and $f_{y,vw}$, and another between *CS* and A_b/A_g. The latter's close relation is in line with domain knowledge—A_b/A_g—which highly depends on the cross-section type. The imbalanced cross-section type shown in Figure 5 (n) might also lead to the first relation since $f_{y,hw}$, $f_{y,cb}$, and $f_{y,vw}$

are not highly linearly correlated in theory. It should be noted that the circled relations might break the independence assumptions of some interpretability analysis methods, such as PDP.

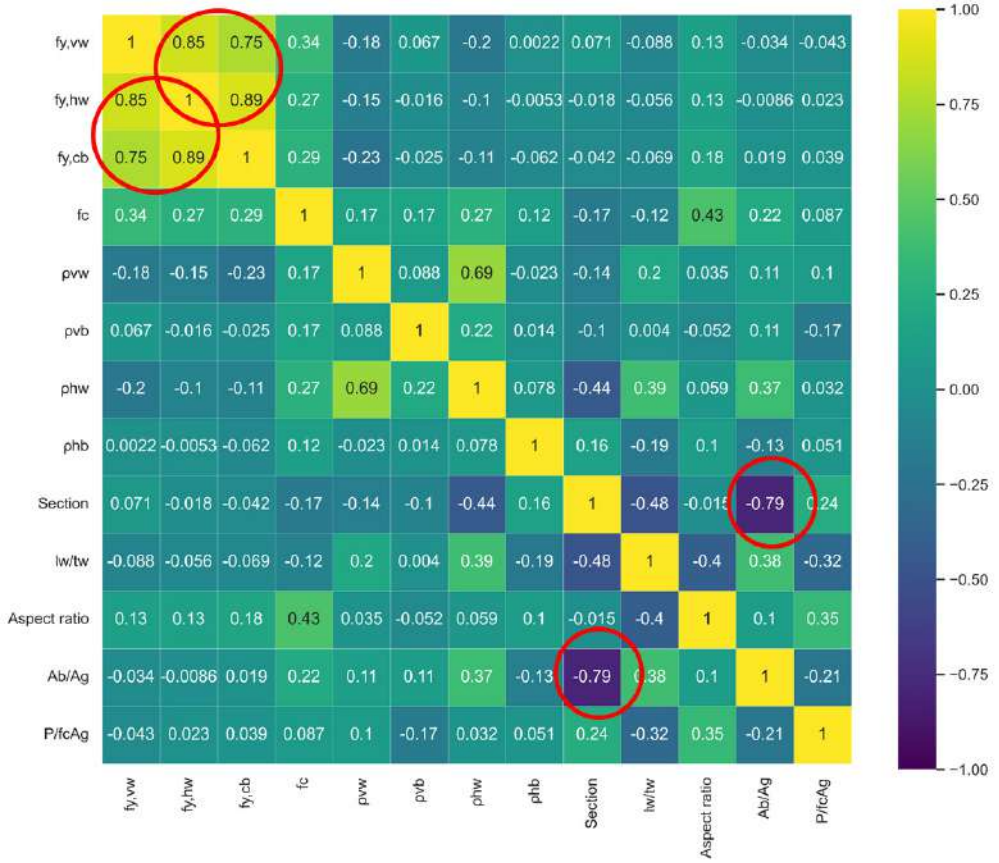
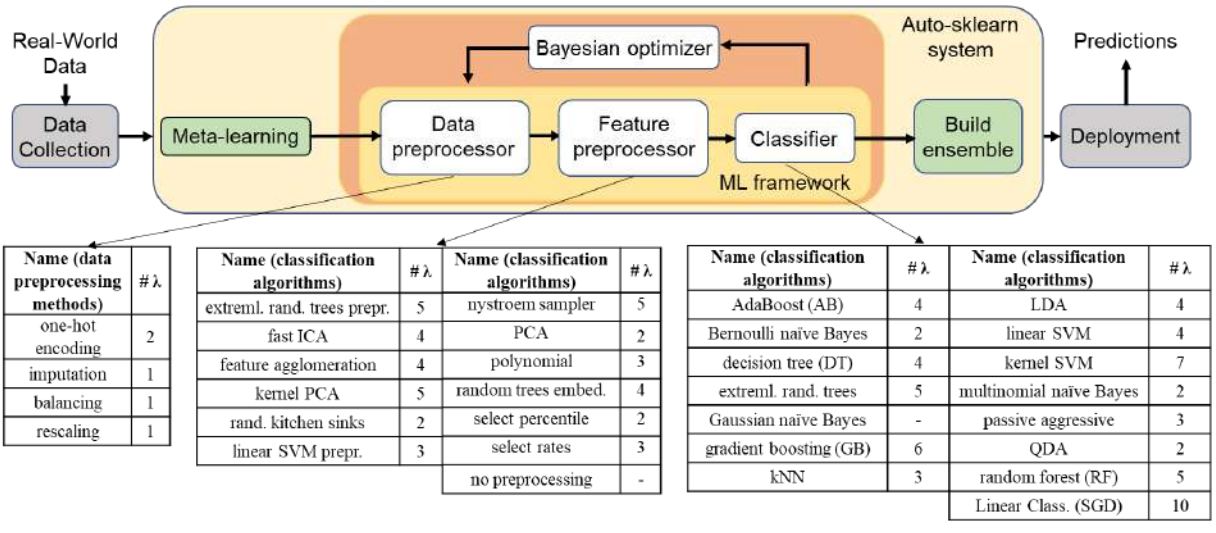


Figure 6 Pearson's correlation matrix of the 13 features

4.3 Classification using Auto-sklearn

This study applies Auto-sklearn, an open-source AutoML library (Feurer et al. 2020), to the problem of RC shear wall failure mode identification. Auto-sklearn has been tested successfully in a large study based on 209 metadata sets and 39 datasets (Feurer et al. 2020). As shown in Figure 7, Auto-sklearn comprises 15 low-level ML algorithms, 14 preprocessing methods, and four data preprocessing techniques. Auto-sklearn employs Bayesian optimization to automate the processes of feature engineering, algorithm selection, parameter tuning, and the entire modeling process shown in Figure 7. Therefore, AutoML can save human experts from heavy reliance on data expertise (Elshawi et al. 2019). More details of Auto-sklearn can be found in Feurer et al. (2015b).



Name (data preprocessing methods)	# λ
one-hot encoding	2
imputation	1
balancing	1
rescaling	1

Name (classification algorithms)	# λ	Name (classification algorithms)	# λ
extreml. rand. trees prepr.	5	nystroem sampler	5
fast ICA	4	PCA	2
feature agglomeration	4	polynomial	3
kernel PCA	5	random trees embed.	4
rand. kitchen sinks	2	select percentile	2
linear SVM prepr.	3	select rates	3
		no preprocessing	-

Name (classification algorithms)	# λ	Name (classification algorithms)	# λ
AdaBoost (AB)	4	LDA	4
Bernoulli naïve Bayes	2	linear SVM	4
decision tree (DT)	4	kernel SVM	7
extreml. rand. trees	5	multinomial naïve Bayes	2
Gaussian naïve Bayes	-	passive aggressive	3
gradient boosting (GB)	6	QDA	2
kNN	3	random forest (RF)	5
		Linear Class. (SGD)	10

Figure 7 Structural diagram and process details in Auto-sklearn

The training target is set to F_1 accuracy, which is an unbiased performance indicator:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{2}$$

$$\text{Macro F1} = \left(\sum_{i=1}^{n} 2 \times \frac{\text{Recall}(i) \times \text{Precision}(i)}{\text{Recall}(i) + \text{Precision}(i)}\right) / n, \tag{3}$$

where *TP*, *TN*, *FP*, and *FN* indicate true positives, true negatives, false positives, and false negatives for the binary shear wall failure mode classification problem, respectively. It should be noted that F1 is able to measure performance objectively when the class balance is skewed (Li et al. 2017) and is widely adopted in the latest ML-related studies (Humphrey et al. 2022; Chen & Xue 2023). After the AutoML model is trained, five manual ML algorithms—Naive Bayes, Random Forest, SVM, Artificial Neural Network, and gradient boosting—that are widely adopted in the literature are compared with the AutoML model according to the F_1 value. In addition, 10-fold cross-validation is used to validate ML models for a limited sample of data. Such a method is superior to split sample validation (Ellis & Mookim 2013; Wong & Yeh 2020; Yuan et al. 2020) because using only a single split of full samples is sensitive to the particular split sample created.

4.4 Interpretability analysis

Three interpretability analysis methods are implemented to explain the trained AutoML from quantitative and qualitative aspects in this study. Two global model-agnostic interpretability analysis methods are PDP and ALE, and a local model-agnostic interpretability analysis method is LIME. Overall, the collective results of the three independent methods can jointly verify and confirm the explainability of the trained AutoML model.

4.4.1 Partial dependent plot (PDP)

PDP illustrates the marginal impact of one or two features on an ML model's predicted output (Friedman 2001). PDP demonstrates whether a result and a feature have a linear, monotonic, or complicated relationship. The partial dependence function is defined as (Molnar 2020):

$$\hat{f}_S(x_S) = E_{X_C}\left[\hat{f}(x_S, X_C)\right] = \int \hat{f}(x_S, X_C) d(X_C). \quad (4)$$

The x_S are features related to the partial dependency function, whereas the X_C are other features utilized in the ML model $\hat{f}$. The features are handled as random variables. Partial dependency operates by marginalizing the output of the ML model across the distribution of features in set *C*, such that the function reveals the link between set *S* and the predicted result. By marginalizing the features, one can get a dependency function on *S* features, including interactions with other features. The PDP method is intuitive, clear, and easy to implement. However, the PDP presumes the marginal feature set is uncorrelated with the feature set *S*. If the assumption does not hold, the results of the PDP may be unreliable (Molnar 2020).

Figure 8 illustrates the estimation of PDPs. With the feature 'aspect ratio' as an example, the estimating process is as follows:

1) Choose a feature, such as 'Aspect ratio.'
2) Create a grid.
3) For each grid value:
    a. Replace the feature with the grid value, such as setting all aspect ratios to 1.5 in Figure 8(b).
    b. Calculate the individual conditional expectation (light blue curves) for each single data piece).
    c. Average the predictions of an individual conditional expectation.
4) Complete a curve, shown as the black curve in Figure 8(a).

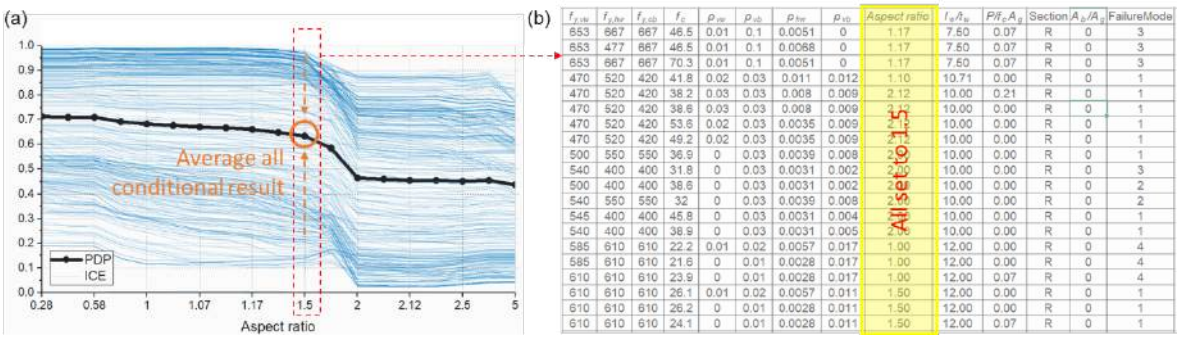


Figure 8 Example PDP analysis of the 'aspect ratio' feature. (a) PDP series plot; (b) each point in (a) indicating the individual expectations with a condition of aspect ratio = 1.5

4.4.2 Accumulated local effects (ALE)

ALE plots average the changes in predictions (Apley & Zhu 2020), which could block the effect of other features (Grömping 2020). Therefore, ALE plot is a less biased substitute for PDP when the features are correlated. In ALE, the features are divided into intervals. Essentially, ALE plots take the average of the changes in the predictions and aggregate them throughout the intervals to evaluate local effects:

$$\hat{f}_{S,ALE}(x_S)=\int_{z_{0,S}}^{x_S}\left(\int_{x_C}\frac{\partial \hat{f}(x_S,x_C)}{\partial x_S}d(X_C \mid X_S=z_S)d\right)dz_S-\text{constant} \quad (5)$$

where z_s denotes the interval value of the concerned feature of ALE, and other symbols are the same as those in Eqn. (1).

Figure 9 shows ALE estimation of two correlated features X_1 and X_2. First, the feature X_1 is divided into intervals, indicated as vertical dash lines. The predicted data points are replaced by the difference in each interval. The differences are accumulated and centered, resulting in the final ALE curve. Overall, the ALE method is more complex and less intuitive than PDP; therefore, ALE is often used together with the PDP method.

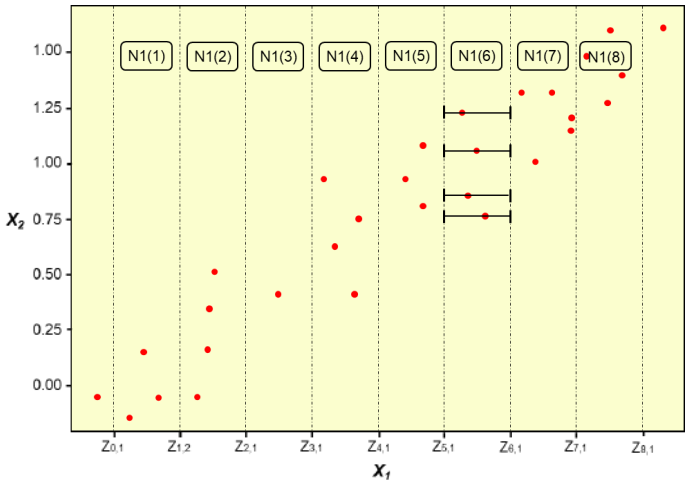


Figure 9 ALE results of two correlated features (X_1 and X_2)

4.4.3 Local interpretable model-agnostic explanations (LIME)

LIME (Ribeiro et al. 2016) trains local surrogate models, as opposed to global surrogate models in PDP and ALE, to approximate the predictions of the underlying black box model. LIME aims to fully understand the rationale behind the predictions of an ML model. First, LIME creates a new dataset of samples' perturbations and their predictions, then develops an interpretable model of a weighted new dataset according to the closeness of the samples to the instance of interest. Commonly interpretable ML models for LIME are linear regression, logistic regression, and decision tree. The predictions from the ML model should be reasonably approximated by the learned model at the local level.

Figure 10 shows the LIME algorithm's sampling and local model training. The whole process is based on a classification example using tabular data:

a) Prepare the ML prediction results based on features X_1 and X_2, such as the predicted classes 1 (green) and 0 (orange) in Figure 10(a).
b) Select an instance of interest (big dot with cross sign) and small dots sampled.
c) Weigh the points that are located closer to the instance of interest.
d) Learn decision boundary Locally from the weighted samples, shown as the dashed line in Figure 10(d).

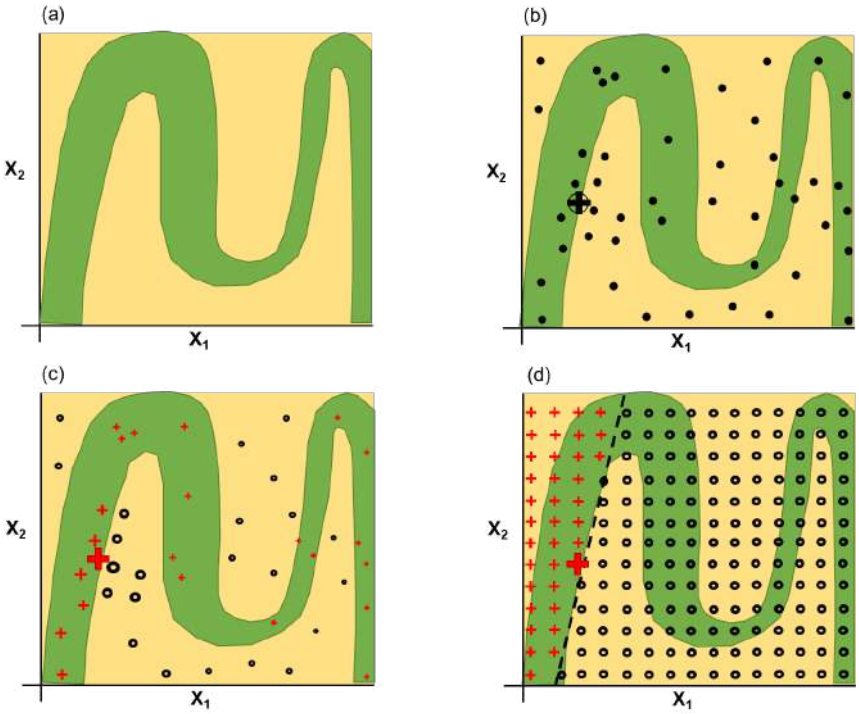


Figure 10 LIME algorithm for local interpretability. (a) ML prediction results; (b) sampling; (c) local model training; (d) local interpretability results

In summary, PDP and ALE are appropriate and complementary methods for explaining the general mechanisms in the trained AutoML model. PDP is intuitive but better for less correlated features, while ALE can work with correlated features. LIME is also an intuitive local method and can explain individual predictions locally.

5 Experimental results

5.1 Experimental settings

All software was operated in single threading mode on a desktop computer with Intel i7-10700 2.9 GHz CPU, 32 GB memory, Python 3.7, and Ubuntu 18.04. The Auto-sklearn toolkit (ver. 0.14.7) was used, while five baseline ML algorithms were from OrangeML (ver. 3.32) (Demsar et al. 2013). Three open-source tools—*PDPbox* (ver. 0.2), *Alibi* (ver. 0.8), and *LIME* library (ver. 0.2.0)—were selected as efficient implementations of PDP, ALE, and LIME, respectively.

5.2 Classification results of AutoML

Figure 11 compares the confusion matrices of AutoML against the baseline ML. In the classification of 'ductile (D)' and 'non-ductile (ND)' modes, the Auto-sklearn model was the best with an F_1 = 90.6%, while the Auto-sklearn model was also the best (F_1 = 84.5%) for classifying the three types of 'non-ductile (ND)' modes. It should be noted that the next best model in Figure 11a and the third best model in Figure 11b are Random Forest, which echoed the findings in Mangalathu et al. (2020). The comparison results confirmed that AutoML is automatic (saving manual interventions that relied on data expertise) and effective (less erroneous) for the failure mode identification of RC shear walls.

Classification algorithm	Naive Bayesian	Random Forest	SVM	Neural Network	Gradient Boosting	AutoML
Precision	79.7%	86.5%	87.8%	87.4%	88.2%	90.6%
Recall	80.0%	90.8%	87.5%	87.3%	87.8%	90.6%
Accuracy	80.3%	88.6%	88.0%	87.7%	88.6%	90.9%
Macro F1	79.9%	88.2%	87.6%	87.4%	87.9%	90.6%

Classification algorithm	Naive Bayesian	Random Forest	SVM	Neural Network	Gradient Boosting	AutoML
Macro Precision	67.7%	86.7%	86.2%	84.5%	81.5%	89.6%
Macro Recall	69.8%	76.8%	86.7%	75.5%	80.3%	82.1%
Accuracy	74.3%	84.5%	80.1%	84.5%	85.0%	88.3%
Macro F1	67.9%	80.0%	69.4%	78.5%	80.8%	84.9%

Figure 11: Comparison of the confusion matrices of four manual ML and AutoML models through 10-fold cross-validation of training: (a) first-stage classification; (b) second-stage classification

The details in the Auto-sklearn training revealed more insights. In the model, an ensemble model was built for each fold of prediction, in which different algorithms with different hyperparameters are combined to achieve a better prediction performance. Figure 12 shows one ensemble model. In Figure 12, the 'radom_forest' classifier with a setting 'max_features = 3, N_estimators = 512' was one of the best low-level ML algorithms; the 'gradient_boosing' low-level classifier with a setting 'early_stopping = true, L2_regularization = 8.06, learning_rate = 0.092, max_iter = 128' was another of the best low-level algorithms. Human experts can, therefore, replicate the AutoML model manually using the tracked training record.

5.3 Results of interpretability analyses of AutoML

5.3.1 Partial dependent plot (PDP)

Figure 13 shows the PDP curves of the trained AutoML model regarding all 11 assumingly independent features. The *x*-axes in Figure 13 denote the features, while the quantiles of the feature are the dash lines. The *y*-axes indicate the probabilities of non-ductile (ND) failure modes, whereas a low value (e.g., $y = 0.3$) represents the ductile (D) failure mode. The individual conditional expectation is illustrated with a light blue curve, while the PDP curve is drawn as a thick black curve.

Model_ID	Rank	Ensemble_weight	Type	Cost	Duration
2	1	0.08	Random_forest	0.145	1.347
11	2	0.06	Gradient_boosting	0.145	1.331
20	3	0.06	Random_forest	0.145	1.639
21	4	0.06	Random_forest	0.161	1.522
23	5	0.08	Mlp	0.177	1.534
13	6	0.06	Random_forest	0.194	2.096
7	1	0.02	Random_forest	0.194	1.463
18	8	0.06	Gradient_boosting	0.210	1.950
14	9	0.04	Extra_trees	0.210	1.294
22	10	0.02	Random_forest	0.226	1.442
24	11	0.02	Random_forest	0.242	1.537
3	14	0.02	Liblinear_svc	0.242	1.082
8	15	0.08	Random_forest	0.242	1.469
9	12	0.06	Gradient_boosting	0.242	0.877
12	13	0.02	Extra_trees	0.242	1.279
5	16	0.12	Gradient_boosting	0.258	1.551
6	17	0.04	Random_forest	0.306	1.375
16	18	0.04	Mlp	0.306	0.887
17	19	0.04	Lda	0.516	0.856
19	20	0.02	Passive_aggressive	0.548	0.712

RandomForestClassifier	
Max_features	3
N_estimators	512
N_jobs	1
Random_state	1
Warm_start	true

GradientBoostingClassifier	
Early_stopping	true
L2_regularization	8.06
Learning_rate	0.092
Max_iter	128
Max_leaf_nodes	200
N_iter_no_change	18
Random_state	1
Validation_fraction	0.143
Warm_start	true

Figure 12: Details of the ensemble models built in a fold of AutoML training

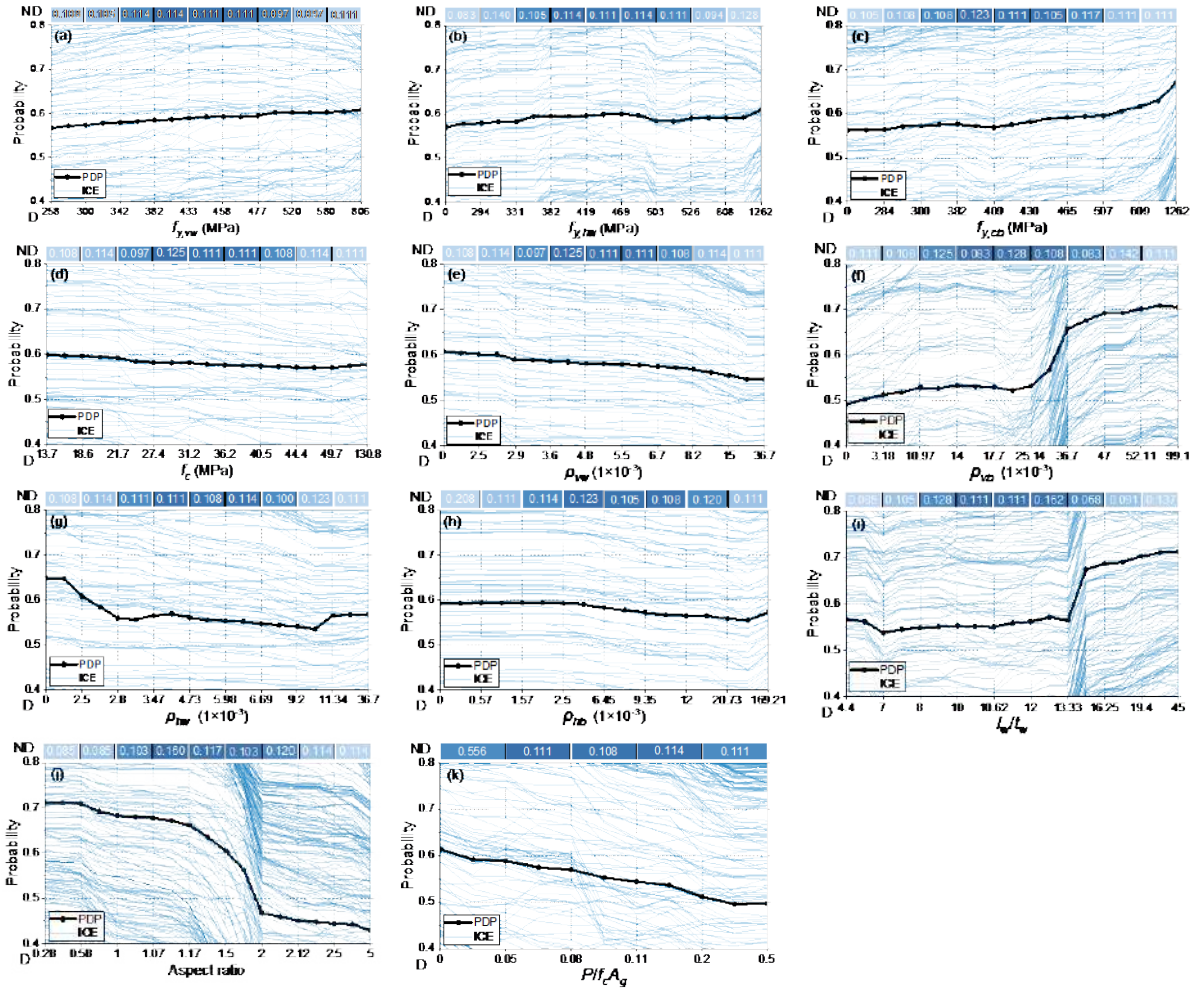


Figure 13 PDP curves of 11 independent features regarding the trained AutoML model

Figure 13 (f) (i-k) shows that predictions of the AutoML model are significantly influenced by four features, namely the boundary region vertical reinforcement ratio (ρ_{vb}), l_w/t_w, aspect ratio, and axial load ratio (P/f_cA_g). First, the probability of non-ductile failure mode increases with the increase in ρ_{vb}. The evidence of this phenomenon is that a larger ρ_{vb} leads to higher flexural strength, while shear strength ensures stability. The shear walls are more likely to fail in a non-ductile (shear-damage-based) mode with larger flexural strength (Gulec 2009).

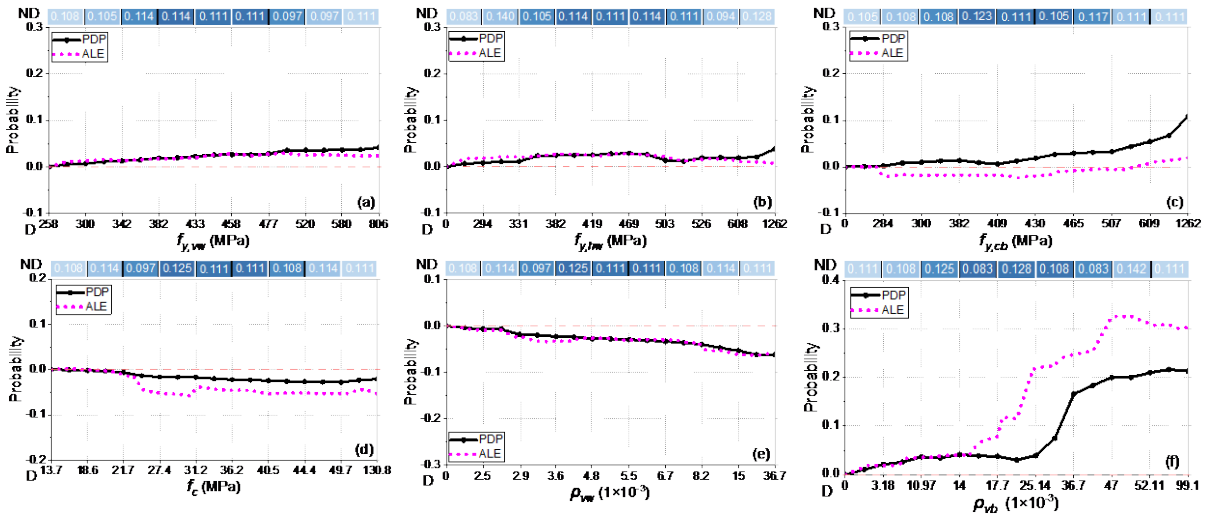
In addition, Figure 13 (i) shows that the probability of non-ductile failure mode obviously increases when l_w/t_w increases, especially when exceeding 13.3. l_w/t_w increases the flexural strength due to the increase in vertical reinforcement moment with a stable cross-section of shear walls, but the shear strength does not change significantly as the area and reinforcement amount of shear walls do not change (Lefas et al. 1990). When the threshold of flexural strength exceeds shear strength, the probability of non-ductile failure mode may increase obviously.

Figure 13 (j) illustrates that non-ductile failures are more likely to occur when the aspect ratio is < 1.5; the probability of ductile failure is high when the aspect ratio exceeds 2. This observation is generally consistent with the theory in (ASCE 2017), which stated that the aspect ratio 1.5 (< 1.5 mostly controlled by shear damage (non-ductile) and 3.0 (>3.0 often controlled by flexure damage [ductile]) are regarded as the thresholds of shear and flexure failure modes.

Finally, Figure 13(k) shows that the higher the P/f_cA_g, the larger probability of flexure (ductile) failure mode predicted by the AutoML model. The trend can be explained by common knowledge about the RC structure component design theory. That is, a proper axial load could be helpful for the improvement of shear strength, and the flexural failure mode is more likely to happen (Moehle 2015).

5.3.2 Accumulated Local Effects (ALE)

Figure 14 shows the ALE curves of the trained AutoML model, with background PDP curves that were adjusted to start at zero. Overall, the trends of the PDP and ALE curves are similar for most features. The interpretations of specified features by PDP curves could be explained by the ALE. The remainder of this section focuses on two more features—A_b/A_g and CS—than those verified by domain knowledge in Sect. 5.3.1.



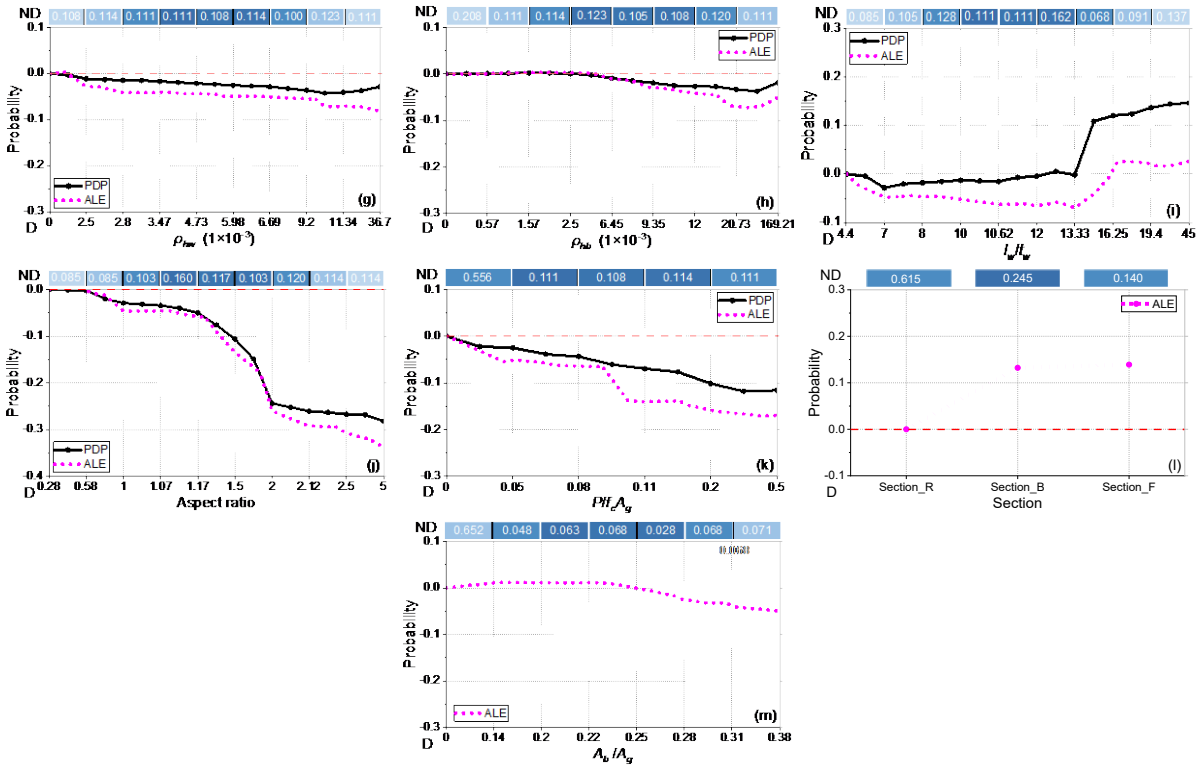


Figure 14 Comparison of PDP and ALE curves of all features

Figure 14(l) shows that barbell and flanged walls are more prone to fail in a non-ductile mode according to the trained AutoML model. Specifically, compared to RC walls with rectangular cross-sections, walls containing boundary components—barbells or flanges—are more vulnerable to diagonal compression failure, which is a non-ductile failure mode. This is because flanged and barbell walls may accommodate a larger amount of reinforcement at the ends of the wall, which, in turn, improve the wall's flexural strength and raise the shear demands in the wall web (Gulec 2009).

Surprisingly, Figure 14(m) shows that the probability trend does not change much when A_b/A_g increases from 0 to 0.38, which is not anticipated. According to civil engineering knowledge, larger A_b/A_g leads to a higher probability of non-ductile failure since non-ductile-failure-controlled flanged and barbell walls correspond to the larger A_b/A_g than rectangular walls in practice.

5.3.3 Local interpretable model-agnostic explanations (LIME)

Figure 15 shows the interpretation of the trained AutoML model by LIME for two typical instances (Nos. 275 and 219). The No. 219 instance is a ductile failure mode, while No. 275 is a non-ductile failure mode. In contrast with the two global methods, LIME focuses on individual prediction results. Figure 15 lists the overall contributions of all features of the instance according to their importance, while the partial effects are illustrated. The tornado charts in Figure 15 illustrate the trained Auto-sklearn model's key decision features' partial effects and thresholds.

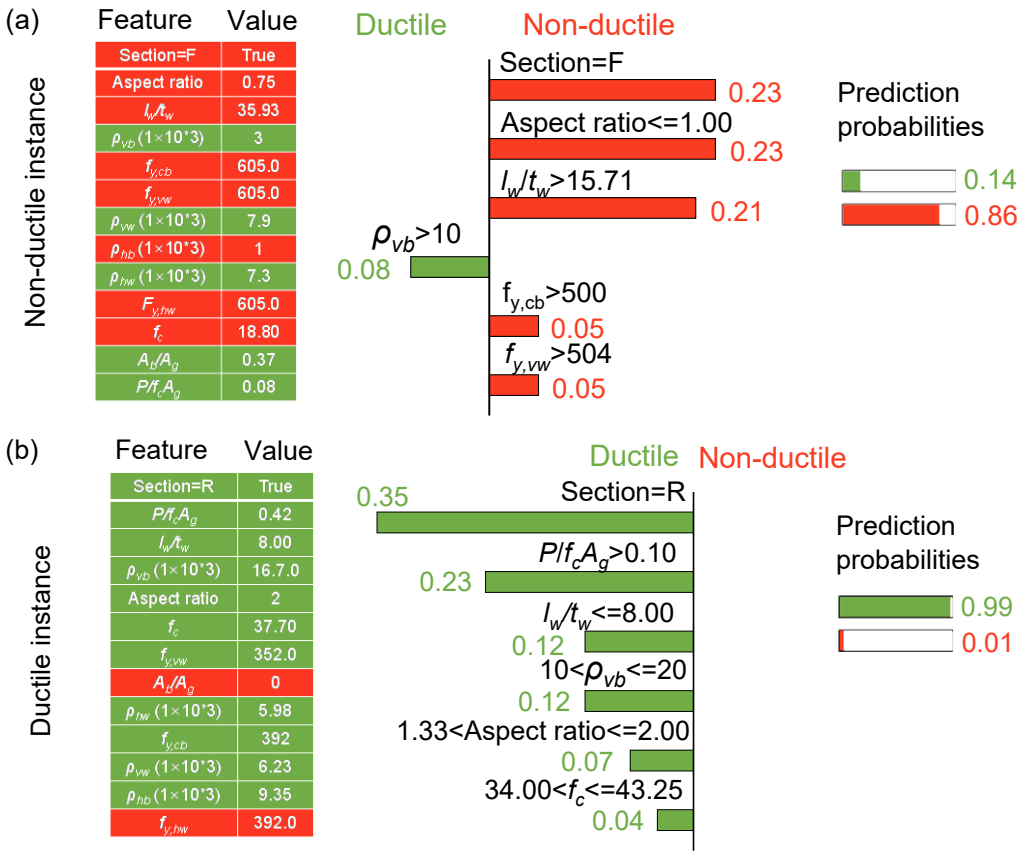


Figure 15 LIME explanations for two typical instances in the dataset. (a) A non-ductile instance; (b) a ductile instance.

In Figure 15(a), the final classification was a non-ductile failure mode due to the high prediction probability—0.86 > 0.50—corresponding to the non-ductile failure mode. Among the top six features, the section, aspect ratio, l_w/t_w, $f_{y,cb}$, and $f_{y,vw}$ contributed to the non-ductile failure mode with scores at 0.23, 0.23, 0.21, 0.05, and 0.05, respectively. In contrast, ρ_{vb} contributed to the ductile failure mode with a score of 0.08. The flanges section contributes most to the non-ductile failure mode classification result. The evidence is that diagonal compression failure, one type of non-ductile failure mode, is more likely to occur in walls with boundary elements (such as barbells or flanges) (Gulec 2009). Meanwhile, the aspect ratio <= 1.00 contributes to the non-ductile failure mode, and it corresponds to the theory that the shear walls with an aspect ratio < 1.5 are mostly controlled by shear damage (non-ductile) (ASCE 2017). The evidence of other features, including l_w/t_w and ρ_{vb}, can be found in the explanation of Figure 13(i) and (f), respectively, in Sect. 5.3.1.

In Figure 15(b), all top six features (i.e., section, P/f_cA_g, l_w/t_w, ρ_{vb}, aspect ratio, and f_c) contribute to the ductile failure mode, of which the probability is at 0.99, and the corresponding scores are at 0.35, 0.23, 0.12, 0.12, 0.07, and 0.04, respectively. The related explanation is also consistent with existing knowledge. For instance, the rectangular section contributes most to the ductile failure mode since it has smaller flexural strength and is more prone to flexural failure mode (ductile mode).

In summary, PDP interpretability analysis indicates four uncorrelated features influencing predictions: boundary region vertical reinforcement ratio (ρ_{vb}), l_w/t_w, aspect ratio, and axial load ratio (P/f_cA_g). In detail, the probability of non-ductile failure mode increases with the increase of ρ_{vb}. Non-ductile failures are more likely to occur when the aspect ratio is < 1.5, but the probability of ductile failure is high when the aspect ratio exceeds 2. The higher P/f_cA_g leads to a larger probability of flexure (ductile) failure mode of shear walls. In addition, ALE curves exhibit similar trends as that in PDP analysis for most uncorrelated features. For uncorrelated features, ALE shows that barbell and flanged walls are more prone to fail in a non-ductile mode according to the trained AutoML model. Finally, the LIME provides a detailed explanation for individual prediction, including importance contributions, partial effects, and corresponding thresholds of all features.

6 Discussion

6.1 Application

As the major lateral load-carrying components in structure systems, shear walls have received much attention due to the higher structural lateral resistance ability requirement for high-rise buildings. However, the existing design process or retrofitting evaluation relies on computing-intensive simulation and expensive experimental testing. The fast and reliable failure mode identification of the shear walls and related explanations for prediction results can help improve the existing practices in designing or retrofitting RC shear walls. As shown in Figure 16, designers and engineers can first use AutoML tools to examine the failure mode of shear walls after acquiring a preliminary design plan based on architectural requirements and structural standards. Generally, the ductile failure mode is preferred to achieve better seismic performance. Therefore, if they get non-ductile failure results in the first step, they can use the interpretation to acquire direct information about how to reverse the failure mode by input factors modifications. Especially if the tendency, contribution ratio, and threshold of each feature are detailed and illustrated in LIME, designers could modify the feature according to the interpretation result. This modification could be an iterative process until satisfactory results are obtained, meaning the ductile failure mode of the shear wall. A similar iterative modification process based on LIME interpretation analysis results is applied in the automotive industry (Pana & Stark 2022).

The iterative RC wall design modification process in Figure 16 can thus take advantage of the trained AutoML model and interpretability analysis. Compared to the conventional computing-intensive simulation and expensive experiment, the novel iterative process can improve traditional simulation-based and test-based design processes in terms of efficiency and costs of money, manpower, and simulation and experiment resources (Chiarello et al. 2021).

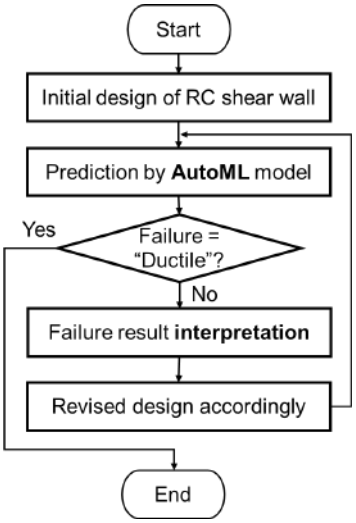


Figure 16 An iterative RC wall design process based on the AutoML model and interpretability analysis

6.2 Implications

By resolving the complex data-expertise-reliant modeling process and the untrustworthy 'black box' of ML application in failure mode identification, the proposed integration of automated ML and interpretability analysis has the potential to solve more complex AEC problems using 'black box' ML techniques. Example problems are the prediction of the behaviors of structural elements and systems (Huang et al. 2022), structure design optimization (Fang et al. 2022), structure health monitoring (Yang et al. 2022), off-site modular building production (Li et al. 2021b), tower crane optimization (Huang et al. 2021), site work packaging (Li et al. 2021a), project-related solid waste generation behaviors (Yang et al. 2021), the semantic enrichment of buildings and the city (Xue et al. 2021; Li et al. 2022), and the prediction of the mechanical properties of the materials used in AEC (Nguyen et al. 2020).

6.3 Pros and Cons of this Study

This study has three advantages:

1. AutoML automates and improves the ML model's establishment, including algorithm selection, parameter tuning, and model evaluation. Auto-sklearn, the AutoML selected in this study, outperformed other 5 manual algorithms in the prediction of the ML of walls' failure mode according to Sect. 5.2. In addition, Auto-sklearn is open source for academia and industry.
2. The global interpretation methods tested in Sects. 5.3.1 and 5.3.2 visualized the input–output correlations and sensitivities of the trained AutoML model. Civil engineers can validate (or reject) the AutoML model against the features with domain knowledge. For example, the interpretation of AutoML indicates that aspect ratios 1.5 (< 1.5 mostly controlled by shear damage (non-ductile)) and 3.0 (> 3.0 often controlled by flexure damage (ductile)) are regarded as the threshold of shear and flexure failure mode, which is consistent with existing knowledge such as code (ASCE 2017). The local interpretation could help users gain a deeper understanding of individual classification results.

3. Last but not least, the interpretation presented in this study triangulates the learned ML models with domain knowledge in civil engineering. Then, the mechanisms of the case AEC problem can be interpreted in more comprehensive (against all features), visual, and quantitative ways. For example, the quantified threshold of ρ_{vb} of RC shear walls between the ductile and non-ductile failure modes was detected in PDP in Sect. 5.3.1, whereas the only trend that greater ρ_{vb} leads to a greater probability of non-ductile failure was known in past civil engineering knowledge.

However, there also exist four limitations in this study:

1. Compared with general AEC knowledge, the interpretation of the ML model focuses on the marginal effect (i.e., the learned discrimination "areas" in the feature space). The interpretability in this study may thus be unable to explain the whole feature space.
2. Although some interpretability methods, such as ALE, could be used to interpret the marginal effect of correlated features, the correlation among different features may influence the accuracy of the interpretability result. For example, the significance of A_b/A_g was not captured in Sect. 5.3.2. One possible reason may be that too much weight has been given to the section, which shares a similar meaning with the A_b/A_g.
3. Due to the small data samples, the time cost of different ML models is limited and, therefore, is not taken into consideration. Time cost could be a good performance indicator when meeting more complex AEC problems and larger datasets.
4. Several assumptions, such as balanced classes or continuous and average feature distribution, were required for the interpretability analyses in this study. Features of interest unable to meet the assumptions may result in unexpected curves or trends in PDP or ALE plots. For example, in the ALE analysis of the feature A_b/A_g, the result was not in line with the domain knowledge, where a possible reason was that half of the instances in the dataset had low values (0) for feature A_b/A_g.

7 Conclusion

Although ML showed success in terms of error metrics in civil engineering, two fatal problems—the complex data-expertise-relied modeling process and the untrustworthy 'black box' nature—have undermined ML adoptions and applications due to critiques from civil engineers and legislators. This paper presents an integration of the AutoML algorithm and interpretability analysis to solve the two fatal problems with a complex AEC problem—the identification of failure modes of RC shear walls. A training dataset of 351 cyclic tests of RC shear walls was collected in this study. Based on the training dataset, 10-fold cross-validation was utilized for training and validating the AutoML model. As a result, the trained AutoML model outperformed the conventional ML algorithms. Regarding interpretability, two global model-agnostic methods—PDP and ALE—and a local model-agnostic method, LIME, were adopted to explain the trained AutoML model per feature and per instance, respectively. The interpretability results confirmed that most observed trends and correlations in the trained AutoML model were in line with domain knowledge. Overall, the proposed integration of the

AutoML algorithm and interpretability analysis is confirmed to address the two fatal issues for the case problem.

The following conclusions can be drawn from this paper:

1) As a high-performance, easy-to-use, and open-source AutoML software, Auto-sklearn was confirmed to be effective and fast in the failure identification of RC shear walls. Evidence is that Auto-sklearn's output outperformed five manual ML algorithms.
2) PDP analysis based on the partial effects confirmed a list of significant features, including boundary region vertical reinforcement ratio (ρ_{vb}), l_w/t_w, aspect ratio, and axial load ratio (P/f_cA_g). The features' significance was consistent with long-established domain knowledge. For example, low values (<1.5) of the aspect ratio led to a high probability of non-ductile failure mode, and high values (>2) contributed to a high probability of ductile failure mode.
3) ALE found similar partial effects with PDP for uncorrelated features, of which two have different estimation mechanisms. For correlated features, such as A_b/A_g and section, interpretation results show that flanged and barbell RC shear walls are more prone to fail in a non-ductile mode.
4) LIME provides a local interpretation for an individual prediction. The partial effect of each feature and related detailed evidence is given, which could provide detailed design/retrofitting suggestions for a single building case.

The contribution to knowledge from this study lies mainly in structural engineering, but also in applied computing. For the former, a theory-based, comprehensive data sampling is provided. More importantly, integration of an AutoML method and interpretability analysis provides an automatic and interpretable problem solution, which resolves the two fatal problems—meaning the complex data-expertise-relied modeling process and the untrustworthy 'black box' nature—of the application of ML in AEC industry. For the latter, empirical evidence is provided to the real-world performance and applicability of AutoML and three interpretability methods for a complex AEC problem. In the future, the proposed integrated workflow can be applied to extracting engineering information and knowledge from complex AEC problems. In addition, big data and other advanced interpretability methods can be adopted for interpreting complex AEC informatics with sophisticated ML methods.

Acknowledgment

This study was supported by Hong Kong Research Grants Council (RGC) (Nos. T22-504/21-R, C7080–22GF)) and in part by Shenzhen Science Technology and Innovation Commission (SZSTI) (No. SGDX20201103093600002). We appreciate the open data sources offered in the literature listed in section 4.2.

References

Abdulridha, A. & Palermo, D. (2017). Behaviour and modelling of hybrid SMA-steel reinforced concrete slender shear wall. *Engineering Structures, 147*(15), 77-89. doi:10.1016/j.engstruct.2017.04.058

ACI. (2019). *Building Code Requirements for Structural Concrete (ACI 318-19) and Commentary.* Farmington Hills, MI, USA: American Concrete Institute. doi:10.14359/51716937

Anwar, T. (2021). COVID19 diagnosis using AutoML from 3D CT scans. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (pp. 203-507). ICCV. doi:10.1109/ICCVW54120.2021.00061

Apley, D. W. & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82*(4), 1059-1086. doi:10.1111/rssb.12377

ASCE. (2017). *Seismic Evaluation and Retrofit of Existing Buildings (ASCE/SEI 41-17).* Reston, Virginia, USA: ASCE. doi:10.1061/9780784414859

ATC. (1998). *Evaluation of Earthquake Damaged Concrete and Masonry Buildings: Basic Procedures Manual (FEMA 306).* Washington, DC: Federal Emergency Management Agency.

Bangaru, S. S., Wang, C., Hassan, M., Jeon, H. W. & Ayiluri, T. (2019). Estimation of the degree of hydration of concrete through automated machine learning based microstructure analysis – A study on effect of image magnification. *Advanced Engineering Informatics, 42*, 100975. doi:10.1016/j.aei.2019.100975

Birely, A. C. (2012). *Seismic Performance of Slender Reinforced Concrete Structural Walls.* Ann Arbor: University of Washington (Doctoral dissertation).

CEN TC-250. (2004). *Eurocode-8: Design of Structures for Earthquake Resistance—Part 1: General Rules, Seismic Actions and Rules for Buildings.* Brussels: CEN.

Chen, S.-H. & Xue, F. (2023). Automatic BIM detailing using deep features of 3D views. *Automation in Construction, 148*, 104780. doi:10.1016/j.autcon.2023.104780

Chiarello, F., Belingheri, P. & Fantoni, G. (2021). Data science for engineering design: State of the art and future directions. *Computers in Industry, 129*, 103447. doi:10.1016/j.compind.2021.103447

Chiou, Y. J., Mo, Y. L., Hsiao, F. P., Liou, Y. W. & Sheu, M. S. (2004). Behavior of high seismic performance walls. *13th World Conference on Earthquake Engineering*, (p. No. 3180). Vancouver.

Christidis, K. I. & Trezos, K. G. (2017). Experimental investigation of existing non-conforming RC shear walls. *Engineering Structures, 140*(1), 26-38. doi:10.1016/j.engstruct.2017.02.063

Dabbagh, H. (2005). *Strength and Ductility of High-strength Concrete Shear Walls Under Reversed Cyclic Loading.* Sydney, Australia: University of New South Wales, Doctoral dissertation.

Dashti, F., Dhakal, R. & Pampanin, S. (2014). Numerical simulation of shear wall failure mechanisms. *The 2014 New Zealand Society for Earthquake Engineering Conference.* Auckland, New Zealand. Retrieved from https://www.nzsee.org.nz/db/2014/oral/22_Dashti.pdf

Deka, P. C. (2019). *A Primer on Machine Learning Applications in Civil Engineering.* New York: CRC Press. doi:10.1201/9780429451423

Demertzis, K., Kostinakis, K., Morfidis, K. & Iliadis, L. (2023). An interpretable machine learning method for the prediction of R/C buildings' seismic response. *Journal of Building Engineering, 63*(Part A), 105493. doi:10.1016/j.jobe.2022.105493

Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbontar, J.,

Zitnik, M. & Zupan, B. (2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research, 14*(1), 2349-2353.

Deng, M., Liang, X. & Yang, K. (2008). Experimental study on seismic behavior of high performance concrete shear wall with new strategy of transverse confining stirrups. *14th World Conference on Earthquake Engineering.* Beijing.

Ellis, R. P. & Mookim, P. G. (2013). *K-Fold Cross-Validation is Superior to Split Sample Validation for Risk Adjustment Models.* Boston University-Department of Economics.

Elshawi, R., Maher, M. & Sakr, S. (2019). Automated machine learning: State-of-the-art and open challenges. *arXiv preprint*, arXiv:1906.02287. https://arxiv.org/abs/1906.02287

Fang, C., Ping, Y., Gao, Y., Zheng, Y. & Chen, Y. (2022). Machine learning-aided multi-objective optimization of structures with hybrid braces–Framework and case study. *Engineering Structures, 269*, 114808. doi:10.1016/j.engstruct.2022.114808

Farrar, C. R., Reed, J. W. & Salmon, M. W. (1993). Failure modes of low-rise shear walls. *Journal of Energy Engineering, 119*(2), 119-138. doi:10.1061/(ASCE)0733-9402(1993)119:2(119)

Feng, D.-C., Wang, W.-J., Mangalathu, S. & Taciroglu, E. (2021). Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. *Journal of Structural Engineering, 147*(11), 04021173. doi:10.1061/(ASCE)ST.1943541X.0003115

Feurer, M., Eggensperger, K., Falkner, S., Lindauer, M. & Hutter, F. (2020). Auto-sklearn 2.0: The next generation. *arXiv preprint*, arXiv:2007.04074, 24. https://arxiv.org/abs/2007.04074, 24

Feurer, M., Springenberg, J. T. & Hutter, F. (2015b). Initializing Bayesian hyperparameter optimization via metalearning. *Proceedings of the AAAI Conference on Artificial Intelligence. 29*, pp. 1128-1135. Palo Alto, California USA: AAAI Press. doi:10.1609/aaai.v29i1.9354

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189-1232. doi:10.1214/aos/1013203451

Gebreyohaness, A., Clifton, C., Butterworth, J. & Ingham, J. (2014). Experimental assessment of inadequately detailed reinforced concrete wall components. *ACI Structural Journal, 111*(2), 279-290. doi:10.14359/51686520

Gijsbers, P., LeDell, E., Thomas, J., Poirier, S., Bischl, B. & Vanschoren, J. (2019). An open source AutoML benchmark. *arXiv preprint*, arXiv:1907.00909. https://arxiv.org/abs/1907.00909

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M. & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *IEEE 5th International Conference on data science and advanced analytics (DSAA)* (pp. 80-89). IEEE. doi:10.1109/DSAA.2018.00018

Grömping, U. (2020). *Model-Agnostic effects plots for interpreting machine learning models.* Berlin: Beuth University of Applied Sciences Berlin.

Gulec, C. K. (2009). *Performance-based Assessment and Design of Squat Reinforced Concrete Shear Walls.* Buffalo, NY, USA: State University of New York at Buffalo.

Guo, H., Zhang, Y. & Zhu, K. (2022). Interpretable deep learning approach for tool wear monitoring in high-speed milling. *Computers in Industry, 138*, 103638. doi:10.1016/j.compind.2022.103638

Hidalgo-Mompeán, F., Fernández, J. F., Cerruela-García, G. & Márquez, A. C. (2021). Dimensionality analysis in machine learning failure detection models. A case study

with LNG compressors. *Computers in Industry, 128*, 103434. doi:10.1016/j.compind.2021.103434

Hosein Naderpour, M. M. (2021). Failure mode prediction of reinforced concrete columns using machine learning methods. *Engineering Structures, 248*, 113263. doi:10.1016/j.engstruct.2021.113263

Huang, C., Li, W., Lu, W., Xue, F., Liu, M. & Liu, Z. (2021). Optimization of multiple-crane service schedules in overlapping areas through consideration of transportation efficiency and operational safety. *Automation in Construction, 127*, 103716. doi:10.1016/j.autcon.2021.103716

Huang, C., Li, Y., Gu, Q. & Liu, a. J. (2022). Machine learning–based hysteretic lateral force-displacement models of reinforced concrete columns. *Journal of Structural Engineering, 148*(3), 04021291. doi:10.1061/(ASCE)ST.1943-541X.0003257

Humphrey, A., Kuberski, W., Bialek, J., Perrakis, N., Cools, W., Nuyttens, N., Elakhrass, H. & Cunha, P. A. (2022). Machine-learning classification of astronomical sources: estimating F1-score in the absence of ground truth. *Monthly Notices of the Royal Astronomical Society: Letters, 517*(1), L116–L120. doi:10.1093/mnrasl/slac120

Hwang, S.-J., Tu, Y.-S., Yeh, Y.-H. & Chiou, T.-C. (2004). Reinforced concrete partition walls retrofitted with carbon fiber reinforced polymer. *ANCER Annual Meeting: Networking of Young Earthquake Engineering Researchers and Professionals*.

Jiang, H., Wang, B. & Lu, X. (2013). Experimental study on damage behavior of reinforced concrete shear walls subjected to cyclic loads. *Journal of Earthquake Engineering, 17*(7), 958-971. doi:10.1080/13632469.2013.791895

Kabeyasawa, T. & Matsumoto, K. (1992). Test and analyses of ultra-high strength reinforced concrete shear walls. *Earthquake Engineering, 10th World Conference.* Balkema.

Kimura, H., Sugano, S. & Nagashima, T. (1996). Seismic behavior of reinforced concrete columns using ultra-high-strength concrete under high axial load. *4th Int. Sym. on the Utilization of HS/HPC*, (p. 184). Paris.

Kookalani, S., Cheng, B. & Torres, J. L. (2022). Structural performance assessment of GFRP elastic gridshells by machine learning interpretability methods. *Frontiers of Structural and Civil Engineering*, 1-18. doi:10.1007/s11709-022-0858-5

Kuang, J. & Ho, Y. (2008). Seismic behavior and ductility of squat reinforced concrete shear walls with nonseismic detailing. *ACI Structural Journal, 105*(2), 225-231. doi:10.14359/19738

Kuwajima, H., Yasuoka, H. & Nakae, T. (2020). Engineering problems in machine learning systems. *Machine Learning, 109*(5), 1103-1126. doi:10.1007/s10994-020-05872-w

Le, T.-T. & Phan, H. C. (2020). Prediction of ultimate load of rectangular CFST columns using interpretable machine learning method. *Advances in Civil Engineering*, 8855069. doi:10.1155/2020/8855069

Lefas, I. D., Kotsovos, M. D. & Ambraseys, N. N. (1990). Behavior of reinforced concrete structural walls: Strength, deformation characteristics, and failure mechanism. *Structural Journal, 87*(1), 23-31. doi:10.14359/2911

Li, D.-C., Hu, S. C., Lin, L.-S. & Yeh, C.-W. (2017). Detecting representative data and generating synthetic samples to improve learning accuracy with imbalanced data sets. *Plos One, 12*(8), e0181853. doi:10.1371/journal.pone.0181853

Li, M., Xue, F., Wu, Y. & Yeh, A. G. (2022). A room with a view: Automatic assessment of window views for high-rise high-density areas using City Information Models and deep

transfer learning. *Landscape and Urban Planning, 226*, 104505. doi:10.1016/j.landurbplan.2022.104505

Li, X., Chi, H. L., Lu, W., Xue, F., Zeng, J. & Li, C. Z. (2021a). Federated transfer learning enabled smart work packaging for preserving personal image information of construction worker. *Automation in Construction, 128*, 103738. doi:10.1016/j.autcon.2021.103738

Li, X., Wu, L., Zhao, R., Lu, W. & Xue, F. (2021b). Two-layer Adaptive Blockchain-based Supervision model for off-site modular housing production. *Computers in Industry, 128*, 103437. doi:10.1016/j.compind.2021.103437

Li, Z., Guo, H., Wang, W. M., Guan, Y., Barenji, A. V., Huang, G. Q., McFall, K. S. & Chen, X. (2019). A Blockchain and AutoML Approach for Open and Automated Customer Service. *IEEE Transactions on Industrial Informatics, 15*(6), 3642-3651. doi:10.1109/TII.2019.2900987

Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy, 13*(1), 18. doi:10.3390/e23010018

Lu, X., Yang, B. & Zhao, B. (2018). Shake-table testing of a self-centering precast reinforced concrete frame with shear walls. *Earthquake Engineering and Engineering Vibration, 17*(2), 221-233. doi:10.1007/s11803-018-0436-y

Luo, H. & Paal, S. G. (2022). Data-driven seismic response prediction of structural components. *Earthquake Spectra, 38*(2), 1382-1416. doi:10.1177/87552930211053345

Lyngdoh, G. A., Kelter, N.-K., Doner, S., Krishnan, N. A. & Das, S. (2022). Elucidating the auxetic behavior of cementitious cellular composites using finite element analysis and interpretable machine learning. *Materials & Design, 213*, 110341. doi:10.1016/j.matdes.2021.110341

Mangalathu, S., Jang, H., Hwang, S.-H. & Jeon, J.-S. (2020). Data-driven machine-learning-based seismic failure mode identification of reinforced concrete shear walls. *Engineering Structures, 208*, 110331. doi:10.1016/j.engstruct.2020.110331

Massone, L. M., Orakcal, K. & Wallace, J. W. (2009). Modelling of Squat Structural Walls Controlled by Shear. *ACI Structural Journal, 106*(5), 646-655. doi:10.14359/51663105

Moehle, J. (2015). *Seismic Design of Reinforced Concrete Buildings.* New York: McGraw-Hill Education.

Mohamed, N., Farghaly, A. S., Benmokrane, B. & Neale, K. W. (2014). Experimental investigation of concrete shear walls reinforced with glass fiber–reinforced bars under lateral cyclic loading. *Journal of Composites for Construction, 18*(3), A4014001. doi:10.1061/(asce)cc.1943-5614.0000393

MOHURD. (2010). *Technical Specification for Concrete Structures of Tall Building (JGJ 3-2010) (in Chinese).* Beijing: Ministry of Housing and Vrban-Rural Development, PRC.

Molnar, C. (2020). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.* Victoria, Canada: Leanpub.

Montáns, F., Chinesta, F., Gomez-Bombarelli, R. & Kutz, J. N. (2019). Data-driven modeling and learning in science and engineering. *Comptes Rendus Mécanique, 347*(11), 845-855. doi:10.1016/j.crme.2019.11.009

Nakamura, N., Tsunashima, N., Nakano, T. & Tachibana, E. (2009). Analytical study on energy consumption and damage to cylindrical and I-shaped reinforced concrete shear walls subjected to cyclic loading. *Engineering Structures, 31*(4), 999-1009. doi:10.1016/j.engstruct.2008.12.013

Naser, M. (2021). An engineer's guide to eXplainable Artificial Intelligence and Interpretable Machine Learning: Navigating causality, forced goodness, and the false perception of inference. *Automation in Construction, 129*, 103821. doi:10.1016/j.autcon.2021.103821

Nguyen, T.-D., Tran, T.-H. & Hoang, N.-D. (2020). Prediction of interface yield stress and plastic viscosity of fresh concrete using a hybrid machine learning approach. *Advanced Engineering Informatics, 44*, 101057. doi:10.1016/j.aei.2020.101057

Palermo, D., Vecchio, F. J. & Solanki, H. (2002). Behavior of three-dimensional reinforced concrete shear walls. *ACI Structural Journal, 99*(1), 81-89. doi:10.14359/11038

Pana, Y. & Stark, R. (2022). An interpretable machine learning approach for engineering change management decision support in automotive industry. *Computers in Industry, 138*, 103633. doi:10.1016/j.compind.2022.103633

Paulay, T. & Priestley, M. J. (1992). *Seismic Design of Reinforced Concrete and Masonry Buildings* (Vol. 768). New York: Wiley.

Paulay, T. (1975). Design aspects of shear walls for seismic areas. *Canadian Journal of Civil Engineering, 2*(3), 321-344. doi:10.1139/l75-030

Ren, F., Chen, J., Chen, G., Guo, Y. & Jiang, T. (2018). Seismic behavior of composite shear walls incorporating concrete-filled steel and FRP tubes as boundary elements. *Engineering Structures, 168*(1), 405-419. doi:10.1016/j.engstruct.2018.04.032

Ribeiro, M. T., Singh, S. & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *The 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). ACM. doi:10.1145/2939672.2939778

Ribeiro, M. T., Singh, S. & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. arXiv:1606.05386. https://arxiv.org/abs/1606.05386

Salonikios, T. N., Kappos, A. J., Tegos, L. A. & Penelis, G. G. (1999). Cyclic load behavior of low-slenderness reinforced concrete walls: Design basis and test results. *ACI Structural Journal, 96*(4), 649-660. doi:10.14359/703

Singh, P. (2021). Data Mining techniques and its application in civil engineering—A review. In A. K. Verma & P. K. Kapur, *Asset Analytics* (pp. 175-183). Singapore: Springer. doi:10.1007/978-981-16-0037-1_15

Somala, S. N., Chanda, S., Karthikeyan, K. & Mangalathu, S. (2021). Explainable Machine learning on New Zealand strong motion for PGV and PGA. *Structures, 34*, 4977-4985. doi:10.1016/j.istruc.2021.10.085

Su, M., Peng, H. & Li, S. (2021). A visualized bibliometric analysis of mapping research trends of machine learning in engineering (MLE). *Expert Systems with Applications, 186*, 115728. doi:10.1016/j.eswa.2021.115728

Su, R. & Wong, S. (2007). Seismic behaviour of slender reinforced concrete shear walls under high axial load ratio. *Engineering Structures, 29*(8), 1957-1965. doi:10.1016/j.engstruct.2006.10.020

Sujith Mangalathu, J.-S. J. (2018). Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques. *Engineering Structures, 160*, 85-94. doi:10.1016/j.engstruct.2018.01.008

Terzioglu, T., Orakcal, K. & M.Massone, L. (2018). Cyclic lateral load behavior of squat reinforced concrete walls. *Engineering Structures, 160*(1), 147-160. doi:10.1016/j.engstruct.2018.01.024

Tokunaga, R. & Nakachi, T. (2012). Experimental study on edge confinement of reinforced concrete core walls. *15th World Conference on Earthquake Engineering.* Lisbon.

Tomazevic, M., Lutman, M., Capuder, F. & Petkovic, L. (1996). Seismic behaviour of RC shear-walls: an experimental study. *11th World conference on Earthquake Engineering*, (p. No.381). Acapulco, Mexico.

Tran, N. K., Howard, T., Walsh, R., Pepper, J., Loegering, J., Phinney, B., Salemi, M. R. & Rashidi, H. H. (2021). Novel application of automated machine learning with MALDI-TOF-MS for rapid high-throughput screening of COVID-19: A proof of concept. *Scientific Reports, 11*, 8219. doi:10.1038/s41598-021-87463-w

Tripathi, M., Dhakal, R. P. & Dashti., F. (2019). Bar buckling in ductile RC walls with different boundary zone detailing: Experimental investigation. *Engineering Structures, 198*, 109544. doi:10.1016/j.engstruct.2019.109544

Usta, M., Pujol, S., 445B, A. S., Puranam, A., Song, C. & Wang, Y. (2017). ACI 445B Shear Wall Database. Purdue University Research Reposito. doi:https://doi.org/10.4231/R7HH6H39

Vadyala, S. R., Betgeri, S. N., C.Matthews, J. & Matthews, E. (2022). A review of physics-based machine learning in civil engineering. *Results in Engineering, 13*, 100316. doi:10.1016/j.rineng.2021.100316

Vakhrushev, A., Ryzhkov, A., Savchenko, M., Simakov, D., Damdinov, R. & Tuzhilin, A. (2021). LightAutoML: AutoML solution for a large financial services ecosystem. arXiv:2109.01528v2. https://arxiv.org/abs/2109.01528v2

Vecchio, F. J. & Collins, M. P. (1986). The modified compression-field theory for reinforced concrete elements subjected to shear. *ACI Journal, 83*(2), 219-231. doi:10.14359/10416

Wen-yuan, Z., Ke, W., Qiang, W., Yong, C., Yu, Z. & Yu-kun, D. (2018). Aseismic behavior of composite shear wall with stiffened double steel plates and infilled concrete. *Engineering Mechanics, 35*(11), 125-133. doi:10.6052/j.issn.1000-4750.2017.06.0482

Wong, T.-T. & Yeh, P.-Y. (2020). Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering, 32*(8), 1586-1594. doi:10.1109/TKDE.2019.2912815

Xue, F., Wu, L. & Lu, W. (2021). Semantic enrichment of building and city information models: A ten-year review. *Advanced Engineering Informatics, 47*, 101245. doi:10.1016/j.aei.2020.101245

Yanez, F., Park, R. & Paulay, T. (1991). Seismic behavior of reinforced concrete structural walls with irregular openings. *Pacific Conference of Earthquake Engineering*, (pp. 67-78). New Zealand.

Yang, X., Gao, Y., Fang, C., Zheng, Y. & Wang, W. (2022). Deep learning-based bolt loosening detection for wind turbine towers. *Structural Control and Health Monitoring, 29*(6), e2943. doi:10.1002/stc.2943

Yang, Z., Xue, F. & Lu, W. (2021). Handling missing data for construction waste management: machine learning based on aggregated waste generation behaviors. *Resources, Conservation and Recycling, 175*, 105809. doi:10.1016/j.resconrec.2021.105809

Yuan, L., Guo, J. & Wang, Q. (2020). Automatic classification of common building materials from 3D terrestrial laser scan data. *Automation in Construction, 110*, 103017. doi:10.1016/j.autcon.2019.103017

Zhang, M., Lia, M., Zhang, J., Liu, L. & Li, H. (2020). Onset detection of ultrasonic signals for the testing of concrete foundation piles by coupled continuous wavelet transform

and machine learning algorithms. *Advanced Engineering Informatics, 43*, 101034. doi:10.1016/j.aei.2020.101034

Zhao, R., Yang, Z. & Xue, F. (2023). Automated Machine Learning for the construction industry: Outperformance in classification and regression benchmarking tests with a practical guideline. *Building Research & Information*, (under review).

Zhou, Y., Lu, X. & Dong, Y. (2010). Seismic behaviour of composite shear walls with multi-embedded steel sections. Part I: experiment. *The Structural Design of Tall and Special Buildings, 19*(6), 618-636. doi:10.1002/tal.597